

Genome-wide eQTL Mapping Identifies Regulatory Variants Underlying Growth Traits in Rainbow Trout

Ridwan O. Ahmed

University of Maryland

Guglielmo Raymo

University of Maryland

Ali Ali

Auburn University

Rafet Al-Tobasei

Middle Tennessee State University

Timothy D. Leeds

National Center for Cool and Cold Water Aquaculture

Mohamed Salem

`mosalem@umd.edu`

University of Maryland

Research Article

Keywords:

Posted Date: September 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10745635/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. Mohamed Salem is a member of the editorial board of BMC Genomics. Otherwise, the authors declare that they have no competing interests.

Genome-wide eQTL Mapping Identifies Regulatory Variants Underlying Growth Traits in Rainbow Trout

Ridwan O. Ahmed¹, Guglielmo Raymo¹, Ali Ali^{1,2}, Rafet Al-Tobasei³, Timothy D. Leeds⁴
Mohamed Salem^{1*}

¹Department of Animal and Avian Sciences, University of Maryland, College Park, MD, 20742-231, USA

²School of Fisheries, Aquaculture and Aquatic Sciences, Auburn University, Auburn, AL 36849, United States

³Computational Science Program, Middle Tennessee State University, Murfreesboro, TN, 37132, USA

⁴National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV, USA

*Corresponding author; mosalem@umd.edu

Abstract

Background: Growth traits in rainbow trout are complex and influenced by numerous genetic variants, yet the regulatory mechanisms underlying phenotypic variation remain poorly characterized.

Results: We integrated low-coverage whole-genome sequencing, RNA sequencing, differential gene expression analysis, and genome-wide expression quantitative trait locus (eQTL) mapping to identify regulatory variation associated with body weight, muscle yield, and condition factor. Genome-wide eQTL mapping identified 234,630 significant cis-SNP-gene associations involving 2,308 genes. We subsequently focused on 6,275 significant TSS-proximal cis-eQTL associations located within 10 kb of transcription start sites and integrated these with white-muscle differential expression, genotype, and phenotypic data. Candidate genes were prioritized using convergent evidence from complementary analyses of genotype frequencies, genotype-dependent muscle gene expression, differential expression, and quantitative phenotypes. This integrative approach identified candidate regulatory variants associated with body weight, muscle yield, and condition factor and revealed substantial cis-regulatory differentiation between two selectively bred genetic lines. Several candidates showed concordant associations among genotype, muscle gene expression, and phenotype. Candidate SNP association analysis using genomic mixed models in the broader population provided additional support for variants near LOC118940393 and within neurofilament medium polypeptide-like associated with condition factor.

Conclusions: This study provides the first comprehensive eQTL resource in rainbow trout and demonstrates the utility of combining genomic and transcriptomic data to elucidate regulatory variation influencing growth. These findings offer valuable targets for functional validation and potential incorporation into genomics-assisted selective breeding programs.

1.0 Background

Growth is a complex and highly polygenic trait regulated by interactions among genetic, environmental, and physiological factors. Understanding the molecular mechanisms underlying growth variation is essential for advancing functional genomics and improving predictive models of phenotypes from genotype. In vertebrates, including salmonids, growth is influenced by a network of pathways related to cellular proliferation, energy metabolism, endocrine signaling, and structural development.

Rainbow trout (*Oncorhynchus mykiss*) is a well-established model for studying complex traits in teleosts, owing to its fully sequenced genome, availability of genomic tools, and diverse phenotypic datasets. Previous studies have shown that traits related to growth—such as body weight, muscle yield, and condition factor—are governed by numerous genetic variants, each contributing modest effect [1-4]. Despite numerous genome-wide association studies (GWAS), most associated loci explain only a small proportion of phenotypic variance, limiting their utility for trait prediction and interpretation. Notably, many of these variants lie in non-coding regions of the genome, where their functional roles are less obvious, making it challenging to determine how they influence gene regulation and, consequently, phenotypic traits.

Transcriptomic analyses have provided additional insights into the molecular basis of growth. For instance, differential expression studies have implicated genes involved in the growth hormone/insulin-like growth factor (GH/IGF) axis, PI3K-Akt, JAK-STAT, and MAPK pathways in muscle and liver tissues of fast-growing versus slow-growing rainbow trout families[5-9]. These studies also highlighted roles for protein metabolism, myogenesis, lipid processing, redox regulation, and regulatory contributions from long non-coding RNAs.

However, connecting expression profiles to specific causal variants remains a challenge. Expression quantitative trait loci (eQTL) mapping bridges this gap by integrating gene expression and genotype data to identify variants that regulate gene expression. eQTLs can act locally (cis-eQTLs), typically within or near the gene they influence, or distantly (trans-eQTLs), affecting genes on different loci or chromosomes[10]. These regulatory variants can illuminate the mechanisms through which genetic differences influence complex traits by modulating transcriptional activity.

To date, there has been limited eQTL mapping in rainbow trout, with the only prior effort using microarray data and a small number of genetic markers to explore acclimation to seawater [11]. With the advent of high-throughput RNA-sequencing and dense genome-wide SNP datasets, it is now feasible to conduct comprehensive eQTL analyses in this species.

In this study, we performed genome-wide eQTL mapping and integrated TSS-proximal cis-eQTLs with white-muscle differential expression, genotype, and phenotypic data to prioritize regulatory variants associated with body weight, muscle yield, and condition factor in rainbow trout.

2.0 Methodology

2.1 Ethical statement

Husbandry practices and experimental procedures at the facility were approved by the University of Maryland, College Park, IACUC animal study protocol, protocol number 1593175-6.

2.2 Rainbow trout population, rearing, and harvest

The fish population and rearing conditions have been described previously [12]. Briefly, the study population originated from the rainbow trout muscle-yield selection line developed at the National Center for Cool and Cold Water Aquaculture (NCCCWA). The line was initially selected for growth for five generations beginning in 2002, after which the breeding objective shifted to increased muscle yield in 2014, as detailed by Cleveland et al. [13]. Fish used in the present study belonged to the 2020 year class and represented third-generation families following the transition to muscle-yield selection. A total of 442 females from 40 families were transferred from NCCCWA to the University of Maryland, College Park's Crane Aquaculture Facility at 322 days post-hatch and maintained in a recirculating aquaculture system under closely monitored water-quality conditions. Phenotypic measurements and tissue samples were collected between 450 and 485 days post-hatch. The fish were taken off feed a day before harvesting. Fish were euthanized using physical stunning through a blow to the skull with a blunt wooden instrument immediately followed by exsanguination. At harvest, whole-body weight, fillet (muscle) yield, and body length measurements were obtained from individual fish. The condition factor was calculated as $(10^5 \times \text{Body weight} / \text{Body length}^3)$.

2.3 RNA extraction

Total RNA was extracted from tissue samples of fish used in this study. The fish used for differential gene expression analysis were classified into “High” and “Low” groups for each trait based on their phenotypic values. RNA extraction was performed using RNazol reagent following the procedure described by Ahmed et al. [12]. The differential expression analyses included 24 fish from four families for body weight, 23 fish from four families for muscle yield, 48 fish from eight families for condition factor, and 94 fish from 17 families for genetic line (Table 1).

Table 1. The number of fish samples and families per trait

	Body weight	Muscle yield	Condition Factor	Genetic Line
High	12 (2 families)	12 (2 families)	24 (4 families)	46 (8 families)
Low	12 (2 families)	11 (2 families)	24 (4 families)	41 (8 families)

2.4 Library preparation and sequencing

Library preparation was performed individually for each sample following the protocol described by Ahmed et al. [12]. Libraries were generated using the Integrated DNA Technologies (IDT) xGen kit with the NEB poly(A) selection module. The procedure involved RNA

fragmentation, random priming, first-strand cDNA synthesis by reverse transcription, 3'-end tailing and adapter ligation, followed by PCR amplification. The resulting libraries were sequenced at the Oklahoma Medical Research Foundation NGS Core, USA, using an Illumina NovaSeq S4 platform with paired-end 150-cycle sequencing.

2.5 Differential Gene Expression Analyses

Raw paired-end RNA-seq reads (2×150 bp) used for differential gene expression and downstream integrated analyses were independently processed outside the CLC workbench workflow used for the main eQTL analysis (described below). Read quality was assessed using FastQC (v0.12) [14], followed by adapter and quality trimming with Trimmomatic (v0.39) [15] using the parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and MINLEN:36. Trimmed reads were aligned to the rainbow trout reference genome (*Oncorhynchus mykiss* USDA_OmykA_1.1, GCF_013265735.2) using HISAT2 (v2.2.x) [16] with the --dta option for downstream transcriptome assembly. Alignments were sorted and indexed using SAMtools (v1.1.x) [17], and paired-end gene-level fragment counts were generated with featureCounts (Subread v2.0.x) [18] using the corresponding NCBI RefSeq annotation (GCF_013265735.2_USDA_OmykA_1.1_genomic.gtf). Thus, the expression data used in the differential expression and integrated analyses were generated independently of the CLC workbench-based processing of the 120 samples used for the primary eQTL mapping (described below).

Differential expression analysis was performed in R (v4.3.x) [19] using two complementary methods, edgeR [20] and DESeq2 [21], with phenotypic group and RNA-sequencing batch included in the statistical models. Lowly expressed genes were filtered using filterByExpr() in edgeR and independent filtering in DESeq2. In edgeR, raw counts were TMM-normalized, dispersion parameters were estimated, and quasi-likelihood F-tests were performed using a generalized linear model with the design $\sim$ batch + condition. DESeq2 used the same model structure, with normalization based on estimated size factors and statistical significance assessed using Wald tests. Genes with an FDR-adjusted P -value < 0.05 and $|\log_2$ fold change > 1 were considered significantly differentially expressed.

2.6.1 DNA Extraction, Genotyping, and Quality Control (QC)

Genomic DNA was extracted from fish tissues (muscle, liver, or fin clips) using a modified salting-out protocol [22]. DNA quality was assessed using gel electrophoresis, while DNA quantity was measured using Qubit fluorometry (Catalog# Q32853).

2.6.2 DNA Library Preparation and Sequencing

Low-coverage whole-genome sequencing libraries were prepared using a modified Illumina Nextera DNA Flex Library Preparation protocol. Paired-end libraries had an average insert size of approximately 350 bp, and library quality was evaluated using an Agilent TapeStation. Final

libraries were quantified by fluorometry, pooled, and sequenced using standard Illumina adapters on an Illumina NovaSeq 6000 platform with 2×150 -bp paired-end reads.

Samples were sequenced at an average genome coverage of approximately $1\times$ for genotyping, followed by genotype imputation using L-Impute v0.18 (Gencove, Inc., New York, NY), based on the imputation approach described by Rubinacci et al. [23]. Low-coverage sequencing has been demonstrated to provide a cost-effective alternative to genotyping arrays while achieving comparable genotyping quality [24].

Sequencing reads were processed through the Gencove imputation workflow, which integrates custom and publicly available software to impute low-coverage sequence data against a haplotype reference panel. Briefly, the pipeline uses a modified implementation of GLIMPSE2 [23], which extends the Li and Stephens model [25], to infer genome-wide genotypes by combining sequence information from each sample with a haplotype reference panel representing the genetic diversity of the target population or species. Posterior probabilities were assigned to imputed genotype calls to facilitate quality control and downstream analysis. Following imputation, VCF files were combined and variants with genotype probabilities below 0.90 were excluded.

2.6.3 Creation of Haplotype Reference Database

The rainbow trout reference genome USDA_OmykA_1.1 (NCBI RefSeq accession GCF_013265735.3) was obtained from NCBI and used to construct the haplotype reference panel. Alternative contigs shorter than 100 kb were removed, after which the genome was indexed using BWA (bwa index -a bwtsw). A FASTA index and sequence dictionary were additionally generated using samtools faidx and Picard CreateSequenceDictionary, respectively. The reference population comprised 160 individuals with high-coverage paired-end sequence data ($10\times$ – $30\times$), representing both domesticated and wild rainbow trout. These samples included publicly available datasets from BioProjects PRJNA386519 ($n = 61$) and PRJNA402066 ($n = 60$), together with 39 individuals from private collections.

Sequence reads were aligned to the reference genome using the BWA-MEM implementation in Sentieon Driver. Variants were identified for each individual using Sentieon Haplotyper to generate gVCFs, followed by joint genotyping with GVCFTyper. A maximum of 12 alternative alleles was retained at each genomic position, while non-polymorphic sites were maintained in the resulting VCF. The jointly genotyped variants were filtered to exclude singletons and variants failing quality thresholds based on QD, FS, MQ, MQRankSum, ReadPosRankSum, and SOR metrics. Residual missing genotypes were subsequently imputed and phased using Beagle to generate the final Gencove haplotype reference panel.

The predictive accuracy of genotype imputation was evaluated using leave-one-out cross-validation. Five reference individuals were sequentially excluded from the reference panel, downsampled to approximately $1\times$ sequencing coverage, and imputed using the remaining reference individuals. Accuracy was assessed using genotype concordance at biallelic SNPs. Mean

concordance across all sites was 94.6%, increasing to 96.5% for high-confidence genotype calls with posterior probabilities > 0.90 .

2.6.4 Quality control

Genotype quality control was performed using VCFtools [26]. Variants were filtered using a minor allele frequency threshold of 0.05 (`--maf 0.05`), a Hardy–Weinberg equilibrium threshold of 1×10^{-6} (`--hwe 1e-06`), restriction to biallelic variants (`--max-alleles 2`), and a maximum missing genotype rate of 0.5 (`--max-missing 0.5`). Following quality control, 14,115,938 SNPs were retained for downstream eQTL analysis.

2.7 Expression Quantitative Loci (eQTL) Mapping

Genome-wide eQTL mapping was performed using RNA-seq data from 120 tissue samples, comprising muscle ($n = 90$), liver ($n = 20$), and pyloric cecum ($n = 10$). Each tissue sample was obtained from a unique fish; therefore, no individual contributed more than one tissue sample to the eQTL analysis. This multi-tissue dataset was used as a discovery resource to identify genetic variants associated with variation in gene expression across tissues. The rainbow trout genome annotation was obtained from NCBI, and RNA-seq reads were processed on CLC Genomic Workbench (version 23) to generate gene-level expression estimates. The parameters for each procedure step in the CLC genomics workbench can be found in Supplementary file 1. Raw counts were converted to transcripts per kilobase million (TPM), and genes with a minimum TPM of 0.01 in at least 90% of samples were retained for analysis. Expression values were quantile normalized across samples.

Associations between SNP genotypes and normalized gene expression were tested using the Matrix eQTL package with the general linear model option [27]. The model included RNA-sequencing batch, tissue type, fish family, and high- or low-phenotypic family status as covariates to account for major sources of variation in gene expression. Cis-eQTLs were defined as significant SNP-gene associations involving variants located within ± 1 Mb of the transcription start site (TSS) of the corresponding gene, whereas all other significant SNP-gene associations were classified as trans-eQTLs. Multiple testing was controlled using the false discovery rate (FDR), with $q < 0.05$ considered statistically significant [10, 28].

Because the growth-related phenotypes investigated in this study were primarily associated with muscle growth and body composition, significant cis-eQTLs identified in the multi-tissue discovery analysis were subsequently integrated with white-muscle differential gene expression results to prioritize trait-relevant candidate genes and regulatory variants. Thus, the genome-wide eQTL analysis was used to identify regulatory associations across tissues, whereas downstream genotype-expression-phenotype analyses of prioritized candidates were evaluated using white-muscle expression data to provide tissue-relevant support for their potential roles in growth-related traits.

2.8 Candidate SNP and Gene Prioritization

To prioritize cis-eQTL variants associated with host traits, we considered cis-eQTL variants within genes that show differential expression for each trait. Candidate genes are defined as genes that were significantly differentially expressed between the high- and low-phenotype groups using both DESeq2 and edgeR and that also possessed at least one significant cis-eQTL. Thus, although initial eQTL discovery was performed across multiple tissues, downstream genotype-expression-phenotype analyses of prioritized growth-associated cis-eQTLs were conducted using white-muscle expression, providing tissue-relevant support for the candidate regulatory variants.

For each candidate gene, the lead cis-eQTL variant within that gene was selected and used for downstream integrated analysis. The lead variant is the variant with the lowest FDR P-value. This resulted in 6, 34, 24 and 37 selected variants for body weight, muscle yield, condition factor and genetic line, respectively. Genotype calls for the selected cis-eQTL variants with a minimum of 60% non-missing genotypes in the samples used for the differential expression analysis were extracted and used for downstream analysis. This resulted in five, twenty-nine, eighteen, and thirty-seven SNP variants for body weight, muscle yield, condition factor and genetic line, respectively. Hard-called genotypes were converted to alternate-allele dosages of 0, 1, or 2, representing homozygous reference, heterozygous, and homozygous alternate genotypes, respectively.

Five complementary questions were evaluated separately for each gene–eQTL pair. First, genotype distributions were compared between high and low fish within each trait. Pearson’s chi-square test was used when expected cell counts were sufficient; otherwise, Fisher’s exact test was used. Second, logistic regression was used to estimate the change in the odds of belonging to the high phenotypic group per additional alternate allele. Third, the effect of alternate-allele dosage on variance-stabilized (VST-transformed) expression of the corresponding gene was evaluated using a linear model that included RNA-sequencing batch and phenotypic group as covariates. This model tested whether the cis-eQTL remained associated with expression after accounting for the expression difference between high and low phenotypic group and using VST-transformed expression values. The variance-stabilizing transformation (VST) was applied to the normalized count matrix to produce approximately homoscedastic expression measurements suitable for linear-model analyses. The transformed expression values were used here as continuous outcomes in the regression models. Fourth, the difference in VST-transformed expression between high and low fish was evaluated using a linear model adjusted for RNA-sequencing batch. Fifth, continuous phenotype was regressed on alternate-allele dosage to test association between eQTL alleles and quantitative phenotype.

These questions support the proposed biological chain:

eQTL genotype → gene expression → phenotypic variation

P-values from each family of gene-level tests per trait were adjusted for multiple testing independently using the Benjamini–Hochberg procedure. Associations with FDR < 0.05 were regarded as statistically significant, whereas results with FDR values slightly above 0.05 were interpreted as suggestive.

2.9 Candidate genetic variant association analysis using genomic mixed models

To complement our results, candidate SNP association analysis was conducted to evaluate variants previously prioritized through the integration of cis-eQTL, differential gene expression, differential allele and genotype frequencies for each trait. Candidate variants were grouped according to the trait or biological contrast from which they were originally prioritized, producing four predefined candidate panels. The body-weight panel comprised five candidate SNPs, the muscle-yield panel comprised 29 candidate SNPs, the condition-factor panel comprised 18 candidate SNPs, and the genetic-line panel comprised 37 candidate SNPs.

Genome-wide variants from lcWGS data (10,975,457) were processed independently of the candidate panels to construct a genomic relationship matrix. Genome-wide SNP quality control was performed using PLINK by excluding variants with genotype missingness greater than 5%, minor allele frequency below 0.01, or Hardy–Weinberg equilibrium P-values below 1×10^{-6} . The quality-controlled genome-wide markers (2,462,145) were subsequently pruned for linkage disequilibrium using a sliding window of 50 SNPs, a step size of five SNPs, and an (r^2) threshold of 0.5. The remaining 200,638 LD-pruned markers were used to construct a genomic relationship matrix in GCTA.

Associations were evaluated for five quantitative phenotypes (body weight, muscle-yield percentage, muscle weight, condition factor, and visceral weight) using 441 fish. Each candidate panel was tested against all five phenotypes to evaluate both trait-specific validation and potential cross-trait associations. Mixed linear model association analyses were conducted using the GCTA MLMA procedure [29]. The primary model included tank, genetic line, and harvest date as fixed covariates, the candidate SNP genotype as a fixed marker effect, and a polygenic random effect defined by the genome-wide genomic relationship matrix. The model can be expressed as:

$$y = Xb + x_j\beta_j + Zu + e$$

where: y is the vector of phenotypic observations; X is the incidence matrix relating observations to the fixed effects; b is the vector of fixed effects, including tank, genetic line, and harvest date; x_j is the vector of genotype dosages for candidate SNP j , coded according to the number of copies of the effect allele; β_j is the additive substitution effect of candidate SNP j ; Z is the incidence matrix relating observations to the random genomic effects; u is the vector of random polygenic effects captured by the genome-wide genomic relationship matrix; and e is the vector of residual effects.

The random genomic and residual effects were assumed to follow: $u \sim N(0, G\sigma^2_g)$ and $e \sim N(0, I\sigma^2_e)$, where G is the genome-wide genomic relationship matrix, I is an identity matrix, σ^2_g is the genomic variance, and σ^2_e is the residual variance.

Multiple-testing correction was performed separately for each candidate SNP panel and phenotype combination. Benjamini–Hochberg false discovery rate (FDR) Multiple-testing correction was applied using the number of candidate SNPs successfully tested within each panel.

3.0 RESULTS

3.1 eQTL Analysis

The eQTL analysis investigated associations between SNPs and gene expression levels for 542,578,507 cis- and 608,927,160,581 trans-potential SNP-gene pairs. The QQplot in Fig. 1 showed that the distribution of the observed p-values of cis-eQTL displayed an earlier departure from the diagonal than that of trans-eQTL, indicating stronger enrichment of significant associations among cis-eQTL tests than among trans-eQTL tests. In total, we identified 234,630 significant cis-SNP-gene associations ($FDR < 0.05$) and 5,300,774 trans-SNP-gene associations ($FDR < 0.05$) (Supplementary file 1 and 2). These associations involved 2,308 and 28,645 unique genes for cis- and trans-eQTLs, respectively. The cis-eQTLs are clustered around the transcription start site (Figure 1).

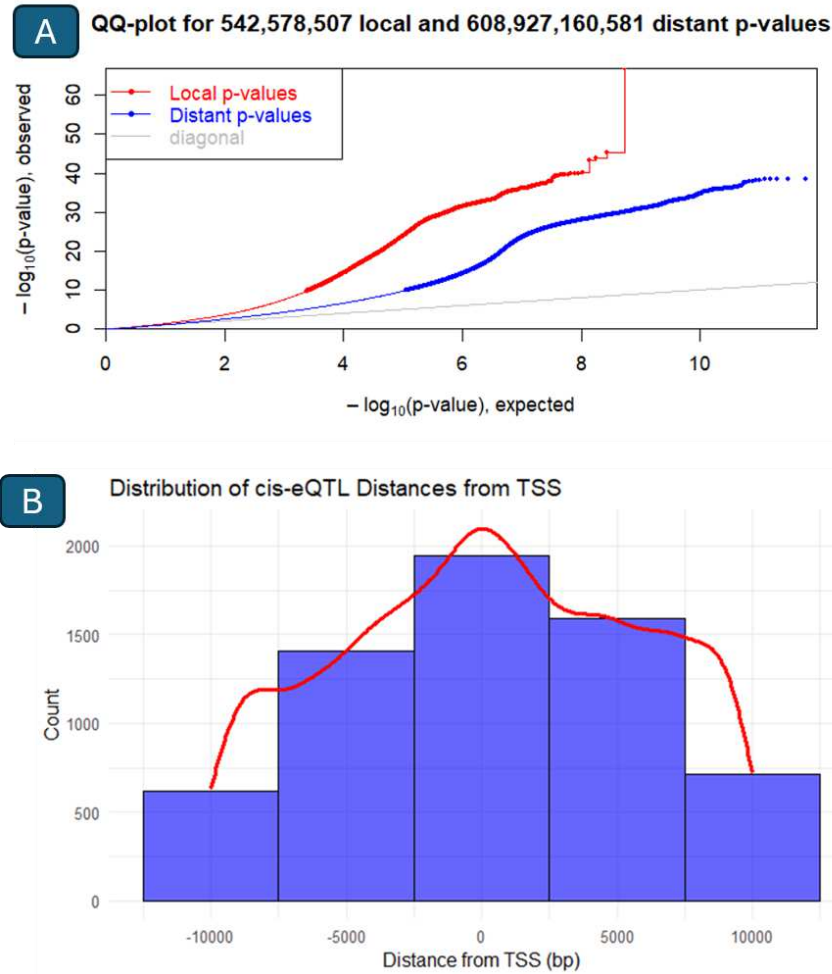


Figure 1. (A) QQ-plot of $-\text{Log}_{10}(P\text{-value})$ of eQTL analysis using MatrixEQTL, and (B) the distribution of cis-eQTLs 10kb around the transcription start site (TSS)

After filtering, we identified 6,275 cis-eQTL variants located within 10 kb of the transcription start site, encompassing the promoter regions of their associated genes. Candidate genes were selected from this refined eQTL set. To further prioritize regulatory variants and target genes, we integrated differential gene expression analysis related to growth and muscle yield traits in rainbow trout as described below.

3.2 Differential Gene Expression Analysis

3.2.1 Phenotypic difference between the High and Low families

The average phenotype in the high and low families for each trait is shown in Figure 2. The difference between the high and low groups is statistically significant ($p < 0.01$).

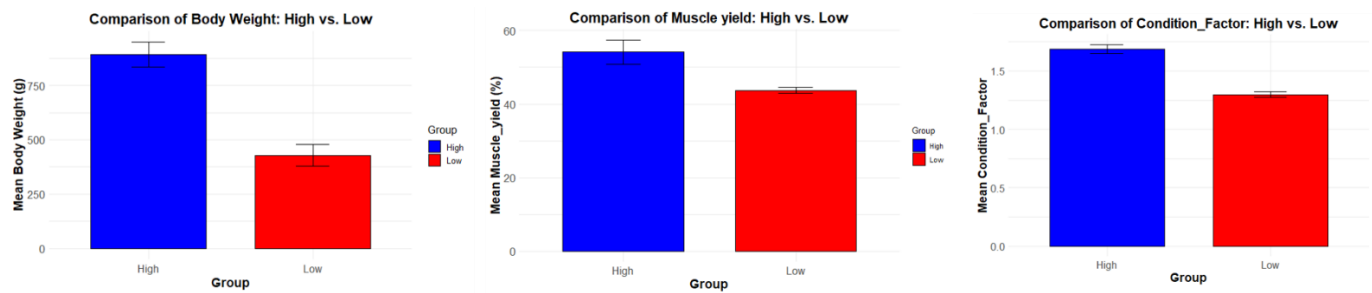


Figure 2. Mean phenotypic values of fish in the high and low groups for body weight, muscle yield, and condition factor. Sample sizes for the high and low groups were 12 vs. 12 for body weight, 12 vs. 11 for muscle yield, and 24 vs. 24 for condition factor, respectively. The genetic-line comparison included 46 ARS-FY-H and 41 ARS-FY-L fish.

3.2.2 Differentially expressed genes (DEGs)

3.2.2.1 Differential gene expression associated with body weight

Differential gene expression between the High and Low phenotypic groups was evaluated using both DESeq2 and edgeR. Using an adjusted significance threshold of $FDR < 0.05$, DESeq2 identified 435 significantly differentially expressed genes (DEGs), whereas edgeR identified 80 significant DEGs for body weight. Although the total number of significant genes differed between the two statistical methods, comparison of the two DEG sets demonstrated substantial concordance. A total of 79 genes were identified as significantly differentially expressed by both DESeq2 and edgeR. The direction and magnitude of differential expressions were generally consistent between the two methods for the overlapping genes.

Among the most significantly differentially expressed genes consistently identified by both methods, several genes related to muscle development and differentiation were upregulated in the high body weight group, including paired box gene 7b (*pax7b*), small muscle protein X-linked (*smpx*), and acetylcholine receptor subunit alpha (LOC110496997). Genes involved in cell signaling and growth regulation included regulator of G-protein signaling 5 (LOC110524579), G protein pathway suppressor 2 (LOC110500675), atypical chemokine receptor 3-like (LOC118965241), and angiopoietin-1 (LOC110507402). Differentially expressed genes associated with cell adhesion, extracellular matrix organization, and tissue remodeling included cadherin-5 (LOC110506352), hyaluronan and proteoglycan link protein 3 (LOC110506823), and serine protease 23 (LOC110495406). Other notable genes were involved in RNA regulation and protein turnover, including CUGBP Elav-like family member 2 (LOC110524948) and ankyrin repeat domain-containing protein 13C (LOC110497762); energy and cellular metabolism, including ATP synthase membrane subunit c locus 3b (*atp5mc3b*) and dehydrogenase/reductase SDR family member 7C-like (LOC110538581); and immune regulation and DNA maintenance, including T-cell activation inhibitor, mitochondrial (LOC110507551) and MutL homolog 1 (*mlh1*),

respectively. All of these genes were upregulated in the high body weight group. The complete list of differentially expressed genes is provided in Supplementary File 1.

3.2.2.2 Differential gene expression associated with Muscle yield

Differential expression analysis revealed extensive transcriptional differences between fish with high and low muscle yield. Using an adjusted false discovery rate (FDR) ≤ 0.05 , DESeq2 identified 2,255 differentially expressed genes (DEGs), whereas edgeR identified 1,625 DEGs. Of the DESeq2 DEGs, 1,219 genes were upregulated in high muscle yield and 1,036 genes were upregulated in low muscle yield. Similarly, edgeR identified 807 genes upregulated in high muscle yield and 818 genes upregulated in low muscle yield.

A high degree of concordance was observed between the two methods, with 1,576 genes identified as significant by both DESeq2 and edgeR, indicating robust and reproducible differential expression signals despite differences in the underlying statistical frameworks.

Among the most strongly differentially expressed genes consistently identified by both methods, several genes related to cell signaling, growth, and differentiation were upregulated in the high muscle yield group, including voltage-dependent calcium channel γ -5 subunit-like (LOC110485983), neurogenic differentiation factor 2-like (LOC110486238), ectoderm-neural cortex protein 1 (ENC1; LOC110509579), and fibroblast growth factor 4B-like (LOC110524584). Other genes upregulated in high muscle yield group associated with transcriptional regulation and cellular development included AN1-type zinc finger protein 6 (LOC110506882), and genes involved in extracellular matrix organization and growth-related protein processing included hyaluronan and proteoglycan link protein 3 (LOC110506144) and LDL receptor-related protein chaperone MESD (LOC110526714). U2 spliceosomal RNA (LOC118944197), involved in RNA processing and pre-mRNA splicing, was also upregulated in the high muscle yield group. Conversely, the low muscle yield group showed higher expression of several genes associated with ribosomal function and protein synthesis, including 60S ribosomal protein L39 (LOC110488974), 60S ribosomal protein L37 (LOC110523343), ribosomal protein L38 (*RPL38*), 60S ribosomal protein L29 (*RL29*), and 40S ribosomal protein S29 (LOC110529849). In addition, cytochrome b-c1 complex subunit 6, mitochondrial (LOC110524069), which is involved in mitochondrial electron transport and energy metabolism, was upregulated in the low muscle yield group.

Overall, these results indicate that high and low muscle yield in rainbow trout are associated with distinct transcriptomic profiles. High muscle yield fish exhibit upregulation of genes involved in cell signaling, transcriptional regulation, extracellular matrix organization, and developmental processes, while low muscle yield fish show increased expression of genes encoding ribosomal and mitochondrial proteins, suggesting differences in protein synthesis and cellular metabolic activity between the two phenotypic groups.

3.2.2.3 Differential gene expression associated with Condition factor

Using an adjusted false discovery rate (FDR) ≤ 0.05 , DESeq2 identified 2,117 differentially expressed genes (DEGs), whereas edgeR identified 1,817 DEGs. Of the DESeq2 DEGs, 1,050 genes were upregulated in high condition factor fish and 1,067 genes were upregulated in low condition factor fish. Similarly, edgeR identified 651 genes upregulated in high condition factor fish and 1,166 genes upregulated in low condition factor fish.

A total of 1,599 genes were identified as significant by both DESeq2 and edgeR, indicating strong agreement between the two methods and demonstrating the robustness and reproducibility of the differential expression results. The full list of the DEGs can be found in Supplementary file 1

Among the most significantly differentially expressed genes consistently identified by both methods, several genes associated with transcriptional and post-transcriptional regulation and RNA processing were upregulated in the high condition factor group, including RNA guanine-7 methyltransferase activating subunit (*ramac*), cytoplasmic polyadenylation element-binding protein 3 (LOC110526678), zinc finger protein 36 C3H type-like 1b (*zfp3611b*), and transcription factor 4 (*tcf4*). Other genes upregulated in high condition factor group include those involved in protein processing, trafficking, and turnover including protein ERGIC-53 (LOC110523995) and inactive ubiquitin carboxyl-terminal hydrolase 54 (LOC110498959), and phosphatidylinositol glycan anchor biosynthesis class P (*pigp*) associated with GPI-anchor biosynthesis and membrane protein processing. Others include genes related to calcium signaling, cytoskeletal regulation, and cell signaling including plasma membrane calcium-transporting ATPase 1 (LOC110493206), target of Nesh-SH3 (LOC110527193), and calpain-1 catalytic subunit (LOC110499567). In addition, glypican-1 (LOC110521536), involved in cell-surface signaling and extracellular interactions, and coagulation factor X (LOC110511683), involved in hemostatic processes, were upregulated in the high condition factor group. Conversely, the low condition factor group showed higher expression of several genes associated with ribosomal function and protein synthesis, including 60S ribosomal protein L37 (LOC110523343), 60S ribosomal protein L39 (LOC110488974), ribosomal protein L38 (*rpl38*), and ribosomal protein S28 (LOC110508919). Other genes upregulated in the low condition factor group included forkhead box J2 (*foxj2*), associated with transcriptional regulation, epidermal growth factor receptor pathway substrate 15 (*eps15*), involved in receptor-mediated endocytosis and cellular signaling, and the uncharacterized gene LOC118936815.

Overall, these results reveal that fish with high and low condition factor exhibit distinct transcriptomic profiles. High condition factor fish show upregulation of genes involved in protein processing and trafficking, transcriptional regulation, calcium homeostasis, ubiquitin-mediated protein turnover, cell signaling, extracellular matrix interactions, and growth-related pathways. In contrast, low condition factor fish are enriched for genes related to ribosomal function and protein synthesis, suggesting differences in translational activity and cellular metabolism between the two phenotypic groups.

3.2.2.4 Differential gene expression associated with genetic line (ARS-FY-H vs ARS-FY-L)

Using an adjusted false discovery rate ($FDR \leq 0.05$), DESeq2 identified 4,481 differentially expressed genes (DEGs), whereas edgeR identified 4,001 DEGs. Of the DESeq2 DEGs, 2,297 genes were upregulated in ARS-FY-H and 2,184 genes were upregulated in ARS-FY-L. Similarly, edgeR identified 1,998 genes upregulated in ARS-FY-H and 2,003 genes upregulated in ARS-FY-L, indicating a high degree of agreement between the two methods in both the number and direction of differentially expressed genes.

Among the most significantly differentially expressed genes identified by both methods, genes associated with growth, differentiation, and cellular signaling were upregulated in the ARS-FY-H line, including voltage-dependent calcium channel γ -5 subunit-like (LOC110485983), neurogenic differentiation factor 2-like (LOC110486238), fibroblast growth factor 4B-like (LOC110524584), ectoderm-neural cortex protein 1 (LOC110509579), nitric oxide synthase 1 (*nos1*), and inositol monophosphatase 1 (*imp1*). Genes involved in lysosomal function and cellular stress/metabolism, including lysosomal integral membrane protein type-2 (*scarb2c*) and heme oxygenase 2b (*hmox2b*), were also upregulated in ARS-FY-H, together with the uncharacterized genes LOC110512232 and LOC110503618. Conversely, the ARS-FY-L line showed higher expression of immune-related genes, particularly complement component 9 (*c9*) and C5a receptor (*c5ar*), together with the uncharacterized genes LOC110506892, LOC110536805, LOC110536956, and LOC110507330.

Overall, these results demonstrate substantial transcriptomic divergence between the two selectively bred genetic lines. The ARS-FY-H line is characterized by upregulation of genes involved in neuronal signaling, calcium transport, growth factor signaling, oxidative stress response, lysosomal function, and cellular metabolism, while the ARS-FY-L line shows increased expression of genes related to the complement system, innate immune response, and other metabolic pathways. These findings suggest that long-term selection has led to broad differences in physiological and molecular processes between the two breeding lines.

3.3 Integration of cis-eQTL and differential expression analyses

To determine whether differentially expressed genes were under local genetic regulation, significant cis-eQTLs were integrated with the differential expression results.

The cis-eQTL analysis identified 6,275 significant SNP-gene associations ($FDR < 0.05$), representing 468 unique genes with evidence supporting local genetic regulation. For body weight, comparison of these genes with the differential expression results showed that six genes were supported simultaneously by significant cis-eQTLs and significant differential expression in both DESeq2 and edgeR for body weight. These six genes therefore represent high-priority expression-regulated candidate genes whose transcript abundance is influenced by local genetic variation while also differing significantly between the high and low body weight phenotypic groups. The lead cis-eQTLs regulating these genes were located on chromosomes 12, 17, 18, 21, 22 and 27.

The genes are glutamine aminotransferase class 1 domain containing 3, Y+L amino acid transporter 2, G protein pathway suppressor 2, 28S ribosomal protein S23, mitochondrial, dehydrogenase/reductase SDR family member 7C-B-like, and uncharacterized LOC118940393.

In addition to body weight, a total of thirty-four (34), twenty-four (24), and forty-five (45) candidate genes were both differentially expressed between the high and low phenotypic groups and contained significant lead cis-eQTL variants for muscle yield, condition factor and genetic line, respectively. These genes/SNPs were selected for downstream integrative analysis. It should be noted that although all candidate SNPs were located within 10 kb of the transcription start site (TSS) of their associated genes, only nine were located within 1 kb of the TSS (supplementary file 1). Thus, most variants were in proximal gene-regulatory regions rather than the core or proximal promoter region.

A few genes overlapped across two or more traits. GATD3A (glutamyl amidotransferase domain-containing protein 3A) is the only gene identified across all four traits (body weight, condition factor, genetic line, and muscle yield). Deoxythymidylate kinase (DTYMK), small integral membrane protein 8, and BRCA1-associated protein were common for condition factor, genetic line, and muscle yield. Several genes were shared between body weight and muscle yield, including Y+L amino acid transporter 2, mitochondrial 28S ribosomal protein S23 (MRPS23), dehydrogenase/reductase SDR family member 7C-B-like (DHRS7C-like), and LOC118940393. Candidate genes common to condition factor and muscle yield include biphenyl hydrolase-like, neurofilament medium polypeptide-like, annexin A1, mitogen-activated protein kinase kinase 22 (MAP3K22), and sorting nexin 22, whereas centrosomal protein 290 (CEP290), GRIP1-associated protein 1, translation machinery-associated protein 7, MAPK/MAK/MRK overlapping kinase (MOK), and CTD small phosphatase-like protein 3 were shared between genetic line and muscle yield.

3.4 Integrated analysis supporting candidate genes for growth traits in rainbow trout

The integrated analyses addressed five complementary questions: (i) whether genotype frequencies for a specific gene differed between high and low phenotypic group for each trait (ii) whether eQTL genotype dosage predicted phenotypic-group membership, (iii) whether genotype dosage influenced gene expression, (iv) whether VST-transformed gene expression differed between high and low phenotypic group, and (v) whether genotype dosage was associated with continuous phenotype. We repeated these analyses for each trait as follows.

3.4.1 Body weight

We focused on lead cis-eQTL variants identified within genes that were differentially expressed between the high and low body weight groups. Five lead cis-eQTL variants were successfully recovered from the genotype dataset after QC and were evaluated using complementary analyses examining genotype distributions, genotype–phenotype associations, genotype–expression relationships and expression differences between body-weight groups.

3.4.1.1 Genotype distributions differ between high and low body weight fish

Genotype distributions differed significantly ($FDR < 0.05$) between the high and low body weight fish groups for four of the five evaluated loci (Figure 3, Supplementary file 1). The strongest difference was observed for the mitochondrial 28S ribosomal protein S23 locus (27_28482591_A_C), in which the C/C genotype predominated in low body weight fish whereas the A/C genotype predominated in high body weight fish. Similarly, significant genotype-frequency differences were observed for LOC118940393, dehydrogenase/reductase SDR family member 7C-B-like, and Y+L amino acid transporter 2. In contrast, genotype frequencies at the G protein pathway suppressor 2 locus did not differ significantly after multiple-testing correction, although the alternate allele remained more frequent in the high body weight fish.

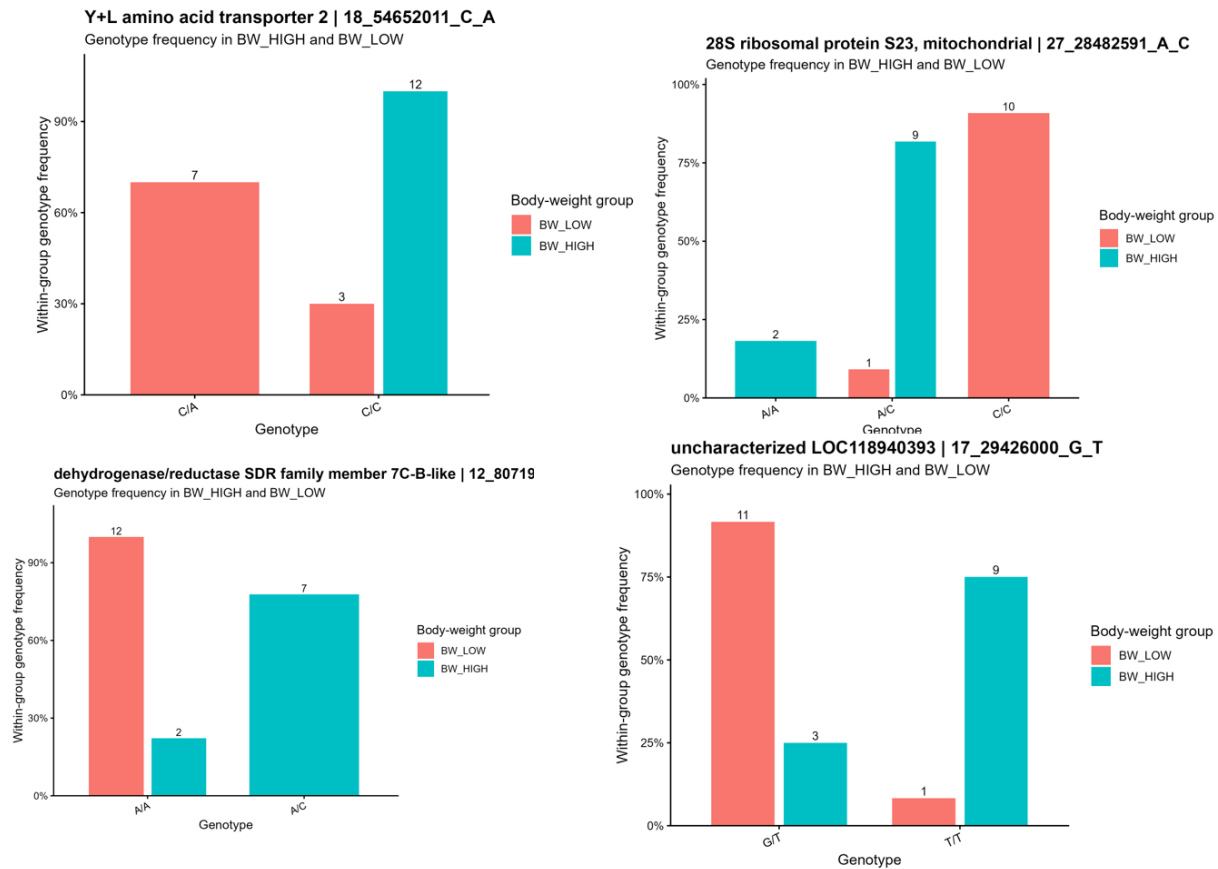


Figure 3. Genotype distributions and alternate (ALT) allele frequencies of lead cis-eQTL variants located within differentially expressed genes associated with body weight. BW_HIGH = high body weight and BW_LOW = low body weight fish.

3.4.1.2 Association between eQTL genotype dosage and body-weight group

This question investigates whether alternative allele dosage increases the probability of belonging to the high body weight fish group. Logistic regression demonstrated that the alternate T allele of LOC118940393 (17_29426000_G_T) significantly increased the probability of belonging to the high weight group (OR = 33.0, FDR = 0.0239) (Supplementary file 1). Although the remaining loci showed similar directional trends, those variants exhibited quasi-complete separation because individual genotype classes occurred almost exclusively within one body-weight group. Consequently, their odds-ratio estimates were unstable despite the significant genotype-frequency differences identified above.

3.4.1.3 eQTL genotype dosage predicts gene expression

Four of the five candidate genes exhibited significant genotype-dependent differences in VST-transformed expression (Figure 4, Supplementary file 1). Increasing T-allele dosage at LOC118940393 significantly reduced transcript abundance ($\beta = -1.16$, FDR = 7.4×10^{-5}), whereas increasing C-allele dosage at the SDR family member 7C-B-like locus significantly increased expression ($\beta = 1.78$, FDR = 0.011). Similar dosage-dependent reductions were observed for G protein pathway suppressor 2 and mitochondrial 28S ribosomal protein S23. The expression of Y+L amino acid transporter 2 increased in heterozygous fish (FDR = 0.053).

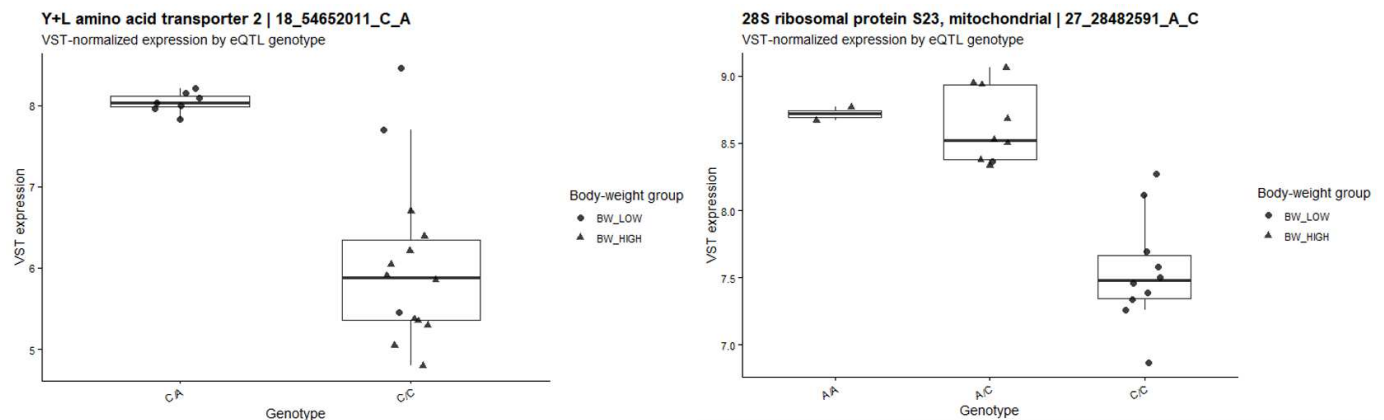


Figure 4: eQTL variants influence VST-normalized expression of candidate genes

3.4.1.4 Candidate-gene expression differs between body-weight groups

Because raw RNA-seq counts are heteroscedastic, gene expression was represented using variance-stabilized (VST) values generated by DESeq2. Unlike the original differential-expression analysis, which modelled raw count data using a negative-binomial distribution, this analysis used continuous expression values to visualize and quantify expression differences between high and low body weight fish.

All five candidate genes exhibited significant expression differences between body-weight groups (Figure 5, Supplementary file 1). Expression of LOC118940393, G protein pathway suppressor 2 and Y+L amino acid transporter 2 was significantly lower in high body weight fish, whereas expression of SDR family member 7C-B-like and mitochondrial 28S ribosomal protein S23 was significantly higher in high body weight fish.

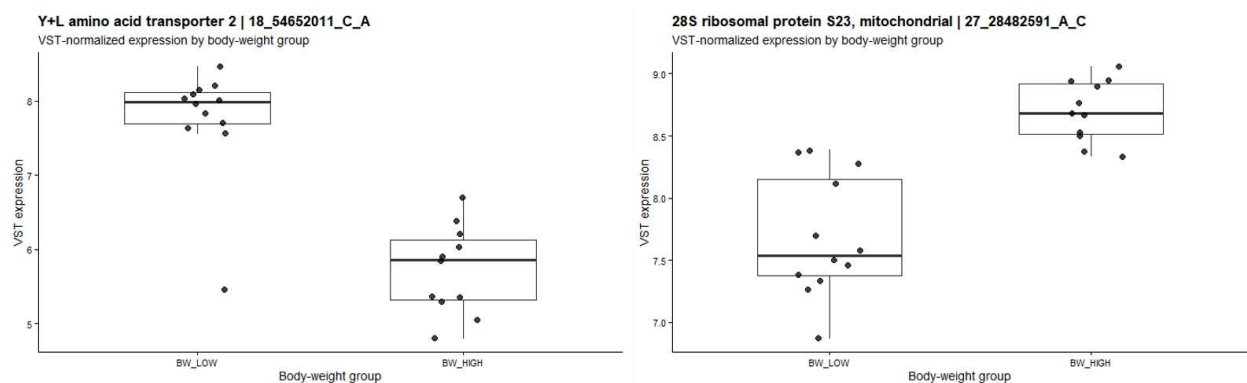


Figure 5: VST gene expression significantly differs between high and low body-weight groups

3.4.1.5 eQTL genotype dosage predicts continuous body weight

Significant associations between alternate-allele dosage and continuous body weight were detected for all five variants (Figure 6, Supplementary file 1). Each additional T allele at LOC118940393 increased body weight by approximately 316 units, whereas each additional C allele at the SDR family member 7C-B-like locus increased body weight by approximately 367 units. Conversely, the alternate alleles at mitochondrial 28S ribosomal protein S23 and Y+L amino acid transporter 2 were associated with significant reductions in body weight.

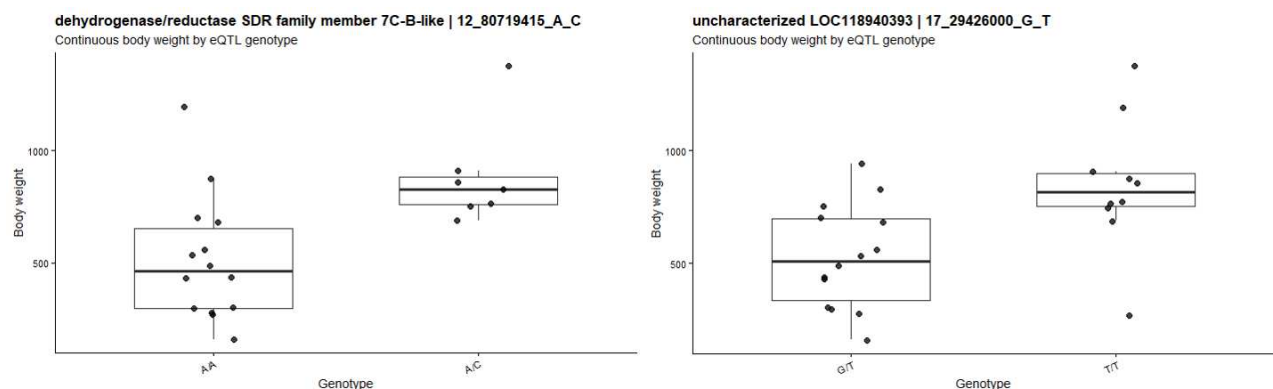


Figure 6: Association between candidate genotypes and body weight

Overall, the integrated genotype–expression–phenotype analyses identified five strong candidate genes exhibiting consistent regulatory effects on body weight. The alternate alleles of LOC118940393 and G protein pathway suppressor 2 were associated with reduced gene expression but increased body weight, whereas the alternate allele of dehydrogenase/reductase SDR family member 7C-B-like increased gene expression and was also associated with higher body weight. In contrast, the alternate alleles of mitochondrial 28S ribosomal protein S23 and Y+L amino acid transporter 2 were associated with reduced body weight, with the former decreasing gene expression and the latter increasing transporter expression. Overall, these findings reveal highly coherent genotype–expression–phenotype relationships, suggesting that regulatory variation in these genes may contribute to body-weight differences in rainbow trout.

3.4.2 Muscle yield

The same analytical framework described above for body weight was subsequently applied to muscle yield, condition factor, and genetic line. Results for the complete set of genotype–frequency, genotype–phenotype, genotype–expression, and expression–phenotype analyses are provided in Supplementary file 1. The results presented here focus on some of the most significant associations and candidate genes.

3.4.2.1 Integrated regulatory candidates associated with muscle yield

The same analytical framework described above for body weight was applied to muscle yield. Thirty-four genes that were both differentially expressed between high- and low-muscle-yield fish and contained significant cis-eQTLs were initially prioritized, of which 29 lead variants passed quality-control criteria and were evaluated in the integrated analyses. Fifteen of these variants showed significant differences in genotype distribution between high- and low-muscle-yield fish ($FDR < 0.05$), 19 showed significant genotype-dependent effects on VST-adjusted gene expression, and 27 corresponding genes remained significantly differentially expressed between muscle-yield groups (Supplementary Figures S1–S3, Supplementary file 1). Five variants were significantly associated with continuous muscle yield after multiple-testing correction ($FDR < 0.05$) (Figure 7, Supplementary file 1).

The strongest integrated candidates were LOC118965296, LOC118940393, LOC110527523, GRIP1-associated protein 1, and decaprenyl-diphosphate synthase subunit 2. LOC118965296 (7_52700228_T_G) exhibited one of the clearest genotype–expression–phenotype relationships: the G allele was enriched in high-muscle-yield fish, increased gene expression, and was positively associated with continuous muscle yield. Similarly, the T allele of LOC118940393 (17_29424191_C_T) was enriched in high-muscle-yield fish and positively associated with muscle yield, while increasing T-allele dosage reduced expression and expression was lower in high-muscle-yield fish, supporting an inverse regulatory relationship between expression and muscle yield. LOC110527523 (7_22684234_C_T) showed the strongest quantitative phenotype association, with the T/T genotype occurring exclusively in high-muscle-yield fish and T-allele dosage positively associated with muscle yield. For GRIP1-associated protein 1

(16_67040408_C_T), the C/C genotype occurred exclusively in high-muscle-yield fish, whereas increasing T-allele dosage was associated with reduced muscle yield. Decaprenyl-diphosphate synthase subunit 2 also showed marked genotype differentiation between muscle-yield groups and a positive association between alternate-allele dosage and continuous muscle yield.

Several additional genes, including LOC110486628, neurofilament medium polypeptide-like, Y+L amino acid transporter 2, CTD small phosphatase-like protein 3, sorting nexin 22, and RhoQ, showed directionally consistent relationships among genotype, gene expression, and muscle yield, although their continuous-phenotype associations did not remain significant after FDR correction. Collectively, these results identify a subset of cis-regulated genes for which regulatory genetic variation is associated with both transcript abundance and muscle-yield variation, providing strong candidates for further investigation of the genetic regulation of muscle yield in rainbow trout.

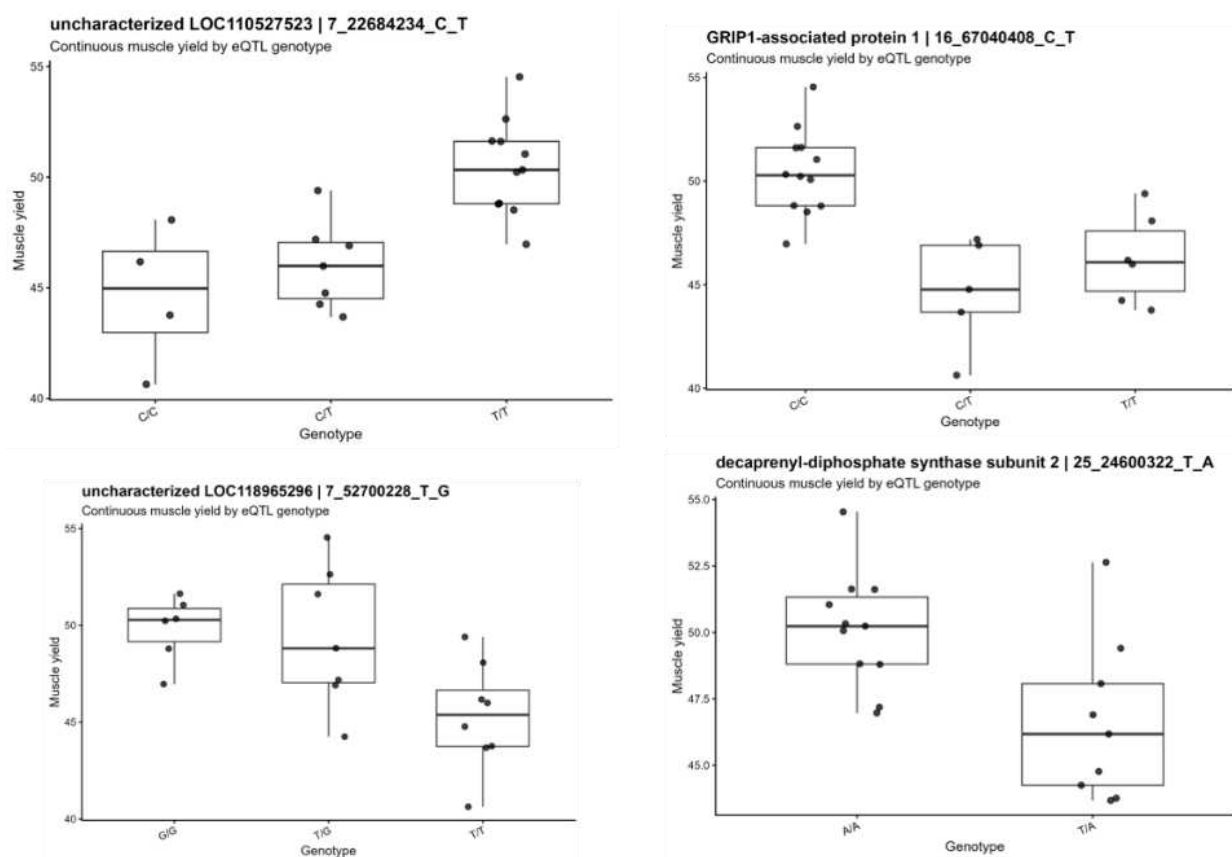


Figure 7: Association between candidate genotypes and muscle yield

3.4.3 Condition Factor

3.4.3.1 Integrated regulatory candidates associated with condition factor

The same analytical framework described above for body weight was applied to condition factor. Twenty-four genes that were both differentially expressed between high- and low-condition-factor fish and contained significant cis-eQTLs were initially prioritized, of which 18 lead variants passed quality-control criteria and were evaluated in the integrated analyses. Eleven variants showed significant genotype-frequency differences between condition-factor groups, eight were significantly associated with group membership by logistic regression, 14 showed significant genotype-dependent effects on gene VST expression, and 17 corresponding genes remained significantly differentially expressed between condition-factor groups ($FDR < 0.05$) (Supplementary Figures S4–S7; Supplementary file 1). Importantly, 11 variants were also significantly associated with continuous condition factor after multiple-testing correction (Figure 8, Supplementary file 1).

The strongest integrated candidates were small integral membrane protein 8, neurofilament medium polypeptide-like, ras guanyl-releasing protein 3, and remodeling and spacing factor 1, which were supported across all five analyses. Neurofilament medium polypeptide-like (1_82180236_A_T) exhibited a particularly coherent positive relationship, whereby the alternate allele (T) increased gene expression, increased the probability of belonging to the high-condition-factor group, and was associated with higher continuous condition factor. In contrast, alternate alleles of small integral membrane protein 8 (4_12550641_G_A), ras guanyl-releasing protein 3 (23_20671281_G_A), and remodeling and spacing factor 1 (27_17480695_T_G) were associated with reduced probability of high-condition-factor membership and lower continuous condition factor.

Several additional genes showed strong evidence across multiple analyses. Zinc finger CCCH domain-containing protein 3 exhibited the strongest quantitative association, with increasing alternate-allele dosage associated with an approximately 0.23-unit increase in condition factor. Lateral signaling target protein 2 homolog, BRCA1-associated protein, and intracellular-type I interferon receptor were also supported by multiple genotype, expression, and phenotype associations. Overall, these results identify a subset of cis-regulated genes showing strong associations between genetic variation, transcript abundance, and condition-factor variation in rainbow trout.

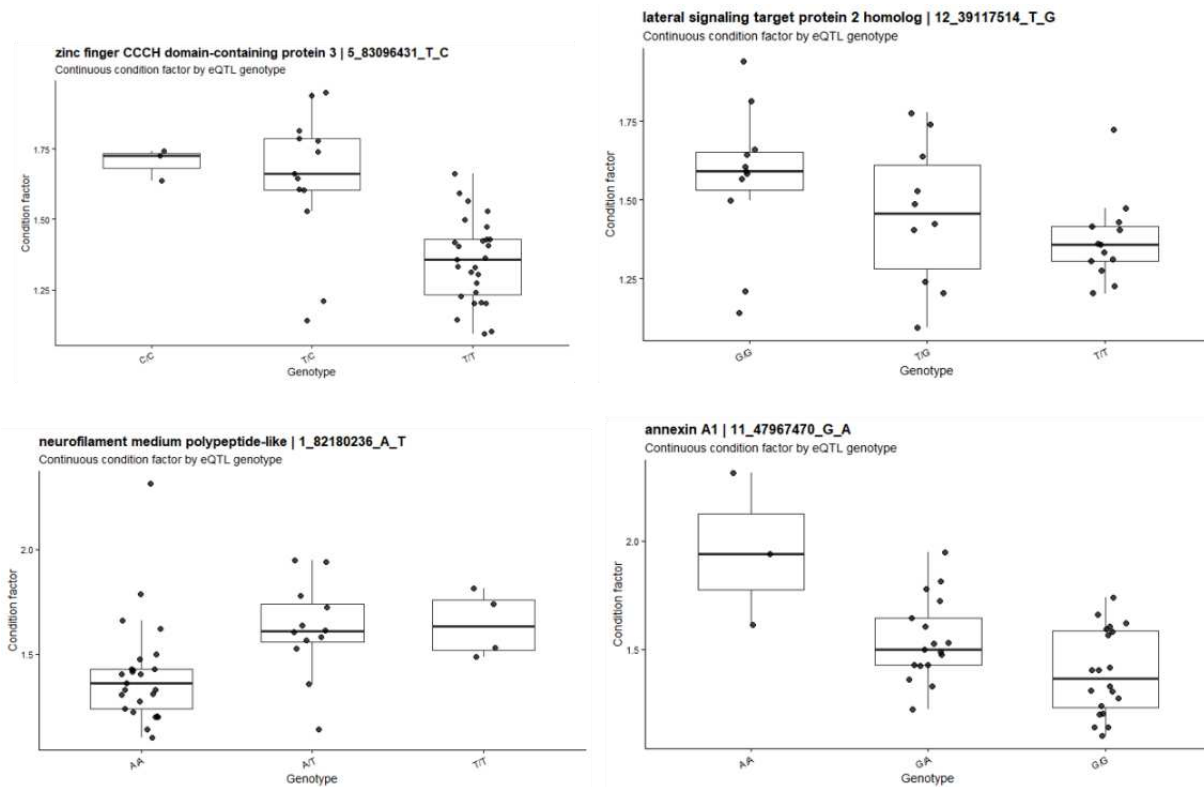


Figure 8: Association between candidate genotypes and condition factor

3.4.4 Genetic line

3.4.4.1 Integrated regulatory differentiation between genetic lines

The same analytical framework described above for body weight was applied to genetic line using 87 fish, comprising 46 ARS-FY-H and 41 ARS-FY-L individuals. Forty-five genes that were both differentially expressed between the two genetic lines and contained significant cis-eQTLs were initially prioritized, of which 37 lead variants passed quality-control criteria and were evaluated in the integrated analyses. Extensive regulatory differentiation was observed between the two lines: 25 of the 37 variants showed significant genotype-frequency differences, 20 were significantly associated with genetic-line membership by logistic regression, 35 showed significant genotype-dependent effects on gene VST expression, and all 37 corresponding genes remained significantly differentially expressed between ARS-FY-H and ARS-FY-L ($FDR < 0.05$) (Supplementary Figures S8–S10; Supplementary file 1).

The strongest integrated candidates included Integrator complex subunit 13, serine/threonine-protein kinase Chk2, LCK2, intraflagellar transport 122 homolog, integral membrane protein 2C, src-like-adaptor 2, small integral membrane protein 8, and BRCA1-associated protein. For several loci, alternate-allele frequencies and dosage effects were strongly differentiated between the two lines. In particular, the alternate alleles of Integrator complex subunit 13 (15_61754869_G_T),

src-like-adaptor 2 (16_67518028_C_T), LCK2 (25_16203173_A_C), intraflagellar transport 122 homolog (16_59382302_G_T), serine/threonine-protein kinase Chk2 (6_17485031_G_C), and integral membrane protein 2C (24_28869061_T_C) were associated with increased probability of ARS-FY-H membership, whereas alternate alleles associated with small integral membrane protein 8 (4_12550641_G_A), chromodomain-helicase-DNA-binding protein 5 (17_67045255_C_T), iron-sulfur cluster scaffold protein b (5_37147993_G_T), translation machinery-associated protein 7 (17_88184784_T_C), calphotin (3_77851477_C_T), GRIP1-associated protein 1 (16_67040408_C_T), and eukaryotic translation initiation factor 4E-binding protein 2 (1_81619984_T_C) were associated with reduced probability of ARS-FY-H membership. These genotype differences were accompanied by strong genotype-dependent and line-dependent differences in transcript abundance.

Because the two genetic lines represent divergent selection for muscle yield, the final analysis tested whether the line-associated regulatory variants were also associated with continuous muscle yield across individuals. In contrast to the extensive genotype and expression differentiation between the two lines, none of the 37 lead cis-eQTL variants was significantly associated with continuous muscle yield after FDR correction (Supplementary file 1). Thus, the identified variants primarily characterize cis-regulatory and transcriptional divergence between ARS-FY-H and ARS-FY-L rather than individual variation in muscle yield. Collectively, these results demonstrate substantial cis-regulatory differentiation between the two selectively bred rainbow trout lines and identify regulatory loci that may contribute to their broader molecular divergence.

3.5 Candidate SNP variants association across the fish population

The candidate SNP association analysis identified a strong association in the chromosome 17 region located near the uncharacterized locus LOC118940393, with two nearby variants (17_29426000_G_T and 17_29424191_C_T; 1.8kb apart) producing significant ($FDR < 0.05$) associations with fish condition factor (Supplementary file 1). These SNPs were prioritized in the muscle yield and body weight candidate SNP panels. In addition, the 1_82180236_A_T variant within the Neurofilament medium polypeptide-like, prioritized in both the condition-factor SNP panel and the muscle-yield SNP panel is associated with the condition factor phenotype ($FDR < 0.05$). Another SNP on chromosome 21 (21_51152905_C_A) prioritized from the body weight panel shows suggestive association with fish condition factor ($FDR = 0.06$) (Supplementary file 1). The results support a cross-trait architecture in which variants originally prioritized for body weight or muscle yield may influence related measures of body condition and composition.

4. Discussion

4.1 Integrating differential gene expression and cis-eQTL analyses improves identification of functional candidate genes

Complex production traits such as body weight, muscle yield, and condition factor are controlled by numerous loci of small to moderate effect together with environmental, physiological, and developmental influences. Consequently, no single genomic approach is

expected to fully explain their underlying genetic architecture. Genome-wide association studies have identified numerous genomic regions associated with economically important traits in aquaculture species [30-33]; however, many associated variants reside within non-coding regions of the genome, making it difficult to identify the genes and molecular mechanisms responsible for phenotypic variation. Expression quantitative trait locus (eQTL) analysis provides an important bridge between DNA sequence variation and gene function by identifying regulatory variants that influence transcript abundance. A comprehensive eQTL study in rainbow trout is lacking.

Here, we performed eQTL analysis using 120 samples and 14,115,938 SNP variants, which could serve as a mini catalogue of rainbow trout eQTLs. Results from differential gene expression analysis, cis-expression quantitative trait locus (cis-eQTL) mapping, whole-genome sequence-derived genotypes, categorical and quantitative phenotypes were integrated to investigate the molecular basis of three economically important production traits in rainbow trout: body weight, muscle yield, and condition factor. We evaluated whether regulatory variants identified through cis-eQTL mapping were consistently associated with transcript abundance, genotype frequencies, and phenotypic variation. Such integrative analyses provide substantially stronger evidence for biological function than differential expression or association analyses performed independently because they directly connect inherited DNA variation with transcriptional regulation and organismal performance.

4.2 Expression QTL Mapping

We conducted eQTL analysis to understand the global genotypic regulation of gene expression and its influence on growth traits in rainbow trout. The gene expression level is a molecular phenotype that bridges the gap between the gene and its phenotype. Expression quantitative traits analysis (eQTL) revealed 234,630 cis and 5,300,774 trans-SNP-gene associations for rainbow trout muscle yield and quality traits. We focused more on the cis-acting QTLs since they tend to have a larger effect size and have a more localized effect on gene expression, making them easier to pinpoint and link to specific phenotypic changes [34, 35].

As shown in Figure 1B, most of the cis-eQTL clustered around their respective genes' transcription start site in line with observations from other eQTL studies from cattle [36] and human [37]. We further filtered the cis-eQTLs to focus on only those variants within 10kb of the transcription start sites. These are known to be enriched for promoter and enhancer regulatory elements [38] with a higher probability of affecting the expression of their eQTL genes.

4.3 Differential expression reveals biological pathways underlying growth traits in rainbow trout

Differential expression analyses identified extensive transcriptional differences between contrasting phenotypic groups for all evaluated traits.

For body weight, DESeq2 identified 435 significant differentially expressed genes whereas edgeR identified 80 genes, with 79 genes detected by both methods. Several of the most significant genes are involved in skeletal muscle development, lipid metabolism, and mitochondrial energy

production; similar to those identified by Claveland et al. [5]. In particular, PAX7B, a key regulator of skeletal muscle satellite cells, plays an essential role in myogenesis, muscle regeneration, and postnatal muscle growth [39]. Differential expression of FABP3, CPT1A, ACADVL, ADIPOR2, GOS2, and delta-6 desaturase further suggests that variation in intracellular fatty-acid transport, mitochondrial β -oxidation, adiponectin signaling, and lipid mobilization contributes to differences in body weight. Similar metabolic pathways have been implicated previously in growth regulation across salmonids and other aquaculture species [40-46].

Muscle-yield comparisons revealed even more extensive transcriptional differences, with 2,255 significant genes identified by DESeq2 and 1,625 by edgeR, including 1,576 genes shared between both methods with complete agreement in direction of effect. Fish exhibiting high muscle yield showed increased expression of genes involved in muscle growth, tissue development, lipid utilization, and cellular organization, including follistatin (FST), lipoprotein lipase (LPL), fibroblast growth factor 4B-like (FGF4B-like), and sex hormone-binding globulin beta (SHBG). Conversely, fish with low muscle yield displayed elevated expression of genes associated with innate immunity, complement activation, acute-phase responses, and lipid transport, including LECT2, C9, CFH, HPXA, HPXB, A2M, APOA1, APOA2, and APOB. These results suggest that greater muscle yield is associated with enhanced anabolic processes, whereas reduced muscle yield may be accompanied by increased inflammatory and immune-related activity. Our findings are consistent with those of a recent transcriptomic study of rainbow trout selectively bred for divergent fillet yield [47], which also reported that differences in muscle yield are associated with transcriptional changes involving muscle development, extracellular matrix remodeling, growth factor signaling, and metabolic regulation. Although the specific DEGs differed between studies, the convergence of these biological processes suggests conserved molecular mechanisms underlying variation in muscle yield across rainbow trout populations.

Condition-factor comparisons similarly identified genes involved in growth regulation, endocrine signaling, metabolism, and immune function. Among the most prominent differentially expressed genes were follistatin (FST), which regulates muscle growth through inhibition of myostatin signaling [48], and the glucocorticoid receptor (NR3C1), a central regulator of energy balance, stress physiology, and metabolic homeostasis [49]. Differential expression of GATA3, CEBPB, and corticotropin-releasing factor-binding protein (CRF-BP) further implicate transcriptional regulation, endocrine function, and lipid metabolism in determining body condition. In contrast, elevated expression of LECT2, an important mediator of innate immunity, suggests that inflammatory pathways may also contribute to variation in condition factor. Collectively, these findings indicate that body condition reflects the balance between growth-promoting pathways, endocrine regulation, metabolic efficiency, and immune function.

The transcriptomic differences observed between the ARS-FY-H and ARS-FY-L genetic lines are consistent with those reported by Mankiewicz et al. [47], who also demonstrated that divergent selection for fillet yield is accompanied by widespread changes in gene expression associated with growth factor signaling, muscle development, and metabolic regulation. Although the specific differentially expressed genes differed between studies, the convergence of these biological processes suggests that long-term selection for improved fillet yield drives coordinated

transcriptional changes in pathways regulating growth, energy metabolism, and tissue development.

4.4 Concordance among genotypes, expression and phenotypic groups strengthens evidence for candidate genes

Candidate genes for growth traits in rainbow trout discussed here are genes harboring significant cis-eQTL variants within 10 kb of their TSS and are also differentially expressed for that phenotype. Candidate genes are also supported simultaneously by genotype-frequency analyses, genotype-dependent expression, differential expression, categorical and quantitative phenotype analyses. This makes them particularly compelling candidates for functional validation because they are supported by independent statistical approaches.

For body weight, LOC118940393, LOC110538581 (dehydrogenase/reductase SDR family member 7C-B-like), LOC110507901 (mitochondrial 28S ribosomal protein S23), LOC110496550 (Y+L amino acid transporter 2), and LOC110500675 (G protein pathway suppressor 2) consistently demonstrated agreement among regulatory genotype, transcript abundance, and body-weight phenotypic variation. Several of the body-weight candidate genes have biological functions consistent with growth regulation. Y+L amino acid transporter 2 (SLC7A6) is involved in cellular amino acid uptake required for protein synthesis and muscle accretion [50], whereas dehydrogenase/reductase SDR family member 7C-like contributes to skeletal muscle energy metabolism and cellular homeostasis [51, 52]. Mitochondrial ribosomal protein S23 is essential for mitochondrial protein synthesis and ATP production [53], processes fundamental for tissue growth, while G protein pathway suppressor 2 (GPS2) regulates metabolic and transcriptional pathways [54, 55] influencing energy utilization and tissue development.

For muscle yield, particularly strong evidence was observed for variants within GRIP1-associated protein 1, LOC110527523, LOC118965296, LOC118940393, and decaprenyl-diphosphate synthase subunit 2, which demonstrated evidence across genotype-frequency analyses, expression regulation, and muscle-yield phenotypic associations. The decaprenyl-diphosphate synthase subunit 2 participates in coenzyme Q biosynthesis and mitochondrial respiratory-chain activity [56, 57], processes required for energy production and muscle metabolism. Similarly, condition-factor analyses identified neurofilament medium polypeptide-like, small integral membrane protein 8, RAS guanyl-releasing protein 3, and remodeling and spacing factor 1 as high-priority regulatory candidates because they consistently showed significant associations across nearly all analytical approaches. RAS guanyl-releasing protein 3 has established functions in Ras–MAPK and AKT signaling and cellular proliferation [58], whereas remodeling and spacing factor 1 participates in chromatin remodeling, developmental specification, and regulation of cell proliferation [59]. Neurofilament medium polypeptide-like is involved primarily in neuronal cytoskeletal organization and axonal stability [60, 61]. Additional evidence supporting functional importance was observed for zinc finger CCCH domain-containing protein 3, a regulator of RNA processing and transport [62] and BRCA1-associated protein which has recognized roles in cell-cycle control, transcriptional regulation, and cellular metabolism [63], both of which exhibited strong genotype-expression-phenotype relationships. Collectively, these candidates implicate

growth-signaling, chromatin regulation, RNA processing, mitochondrial or membrane function, and neuromuscular organization in variation in condition factor.

Candidate genes for genetic line with strongest evidence are Src-like-adaptor 2 (SLA2), Protein tyrosine kinases LCK2 (LOC100136288), Integrator complex subunit 13 (INTS13), Serine/threonine-protein kinase (CHEK2), Intraflagellar transport 122 homolog (IFT122), BRCA1-associated protein, Integral membrane protein 2C (IMP2C), Solute carrier family 35 member D2 (SLC35D2), CTD small phosphatase-like protein 3 (CTDSPL3), Iron–sulfur cluster scaffold protein b (ISCUB) and Ribonuclease inhibitor-like. These genes harbor cis-regulatory variants possibly underlying molecular and genetic divergence between ARS-FY-H and ARS-FY-L. Their known functions encompass immune and kinase signaling, transcription and RNA processing, DNA-damage and cell-cycle control, ciliary developmental signaling, membrane and Golgi transport, protein glycosylation, and mitochondrial iron–sulfur cluster biogenesis [64-67]. In particular, ISCUB provides a plausible connection to mitochondrial metabolism [64, 68] and muscle energetic capacity, while SLA2 and LCK2 suggest divergence in immune-signaling pathways [65, 69, 70], and INTS13, CTDSPL3, and CHEK2 indicate differences in transcriptional and cell-cycle regulation [66, 67].

The consistency across analytical approaches demonstrates the value of integrating functional genomics for studying complex quantitative traits. By prioritizing candidates supported by multiple complementary lines of evidence rather than individual statistical associations, this framework increases confidence that the identified loci contribute to phenotypic variation through biologically meaningful regulatory mechanisms and reduces the likelihood of chance findings. Notably, several candidates were shared across traits, suggesting common regulatory mechanisms underlying growth and body composition. GATD3A (glutamyl amidotransferase domain-containing protein 3A) was the only gene identified across all four traits—body weight, condition factor, genetic line, and muscle yield—making it a particularly notable candidate. DTYMK, small integral membrane protein 8, and BRCA1-associated protein were shared among condition factor, genetic line, and muscle yield. Genes common to body weight and muscle yield included Y+L amino acid transporter 2, MRPS23, DHRS7C-like, and LOC118940393, while biphenyl hydrolase-like, neurofilament medium polypeptide-like, annexin A1, MAP3K22, and sorting nexin 22 were shared between condition factor and muscle yield. CEP290, GRIP1-associated protein 1, translation machinery-associated protein 7, MOK, and CTD small phosphatase-like protein 3 were shared between genetic line and muscle yield. Collectively, these recurrent candidates may represent core regulators connecting multiple economically important growth and body-composition traits through shared or interconnected biological pathways.

Additional support for selected candidates was obtained from genomic mixed-model analyses conducted in the broader fish population. Variants in the chromosome 17 region near LOC118940393 and within the neurofilament medium polypeptide-like locus remained significantly associated with condition factor after accounting for population relatedness using a genome-wide genomic relationship matrix. These findings provide additional evidence linking these regulatory regions to variation in body condition.

4.5 Biological and breeding implications and future directions

Integrating cis-eQTL mapping with differential gene expression and phenotypic data provides a biologically informed approach for prioritizing loci underlying economically important traits by connecting DNA sequence variation with transcriptional regulation and phenotypic variation. Regulatory variants supported across multiple analytical layers may therefore represent stronger candidates for functional validation and, if independently validated, could contribute to genomic selection, marker-assisted breeding, and improved genomic prediction in rainbow trout.

Genome-wide eQTL discovery was performed using expression data pooled across muscle, liver, and pyloric cecum, with tissue type included as a covariate. Because genotype-by-tissue interactions were not explicitly modeled, the genome-wide eQTL catalog should not be interpreted as representing tissue-specific regulatory effects. However, prioritized growth-associated cis-eQTLs were subsequently evaluated using white-muscle expression, and the genotype-dependent expression observed for many candidate genes provides muscle-specific support for their regulatory effects. Future tissue-stratified eQTL analyses with larger sample sizes will be valuable for characterizing tissue-specific regulatory architecture.

However, the present findings were derived from a single experimental population, and the observed genotype–expression–phenotype associations do not establish causality. Validation in independent breeding populations and functional studies, including CRISPR-mediated genome editing, will therefore be important for confirming the effects and mechanisms of prioritized variants. Evaluation across additional tissues, developmental stages, and genetic backgrounds will further establish their broader biological relevance and determine whether incorporating validated regulatory variants can improve genomic prediction of economically important traits.

4.6 Conclusions

Overall, this study demonstrates that integrating differential gene expression, cis-expression quantitative trait locus mapping, whole-genome sequence-derived genotypes, qualitative and quantitative phenotypes provides a powerful framework for identifying functional regulatory variants underlying economically important production traits in rainbow trout. The concordance observed among regulatory genotype, transcript abundance, and phenotype strengthens confidence in the identified candidate genes and suggests that regulatory variation may be an important component of the genetic architecture of body weight, muscle yield, and condition factor. These findings provide biologically informed candidate loci for future functional validation and represent promising targets for incorporation into genomic selection strategies aimed at improving growth performance and fillet quality in rainbow trout.

Data Availability

All datasets generated for this study are included in the manuscript and/or the Additional Files. All raw sequence data generated in this study has been deposited in NCBI under BioProject accession number: PRJNA974903, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA974903>. The raw VCF

files for the lcWGS data generated for this project have been deposited in NCBI under the BioProject accession number: PRJEB109732

Abbreviations

DEGs: Differentially Expressed Genes, **NCCCWA:** USDA National Center of Cool and Cold Water Aquaculture, **eQTL:** Expression quantitative trait loci, **SNP:** Single nucleotide polymorphism, **VST:** Variance Stabilizing Transformation, **VCF:** Variant Call Format.

Funding

Funding was obtained for this study from the United States Department of Agriculture, National Institute of Food and Agriculture, competitive grants No. 2023-67015-39742, and 2026-67015-45723.

Author information

Authors and Affiliations

1. Department of Animal and Avian Sciences, University of Maryland, College Park, MD, 20742, USA

Ridwan O. Ahmed, Guglielmo Raymo, Ali Ali & Mohamed Salem

2. Computational Science Program, Middle Tennessee State University, Murfreesboro, TN, 37132, USA

Rafet Al-Tobasei

3. Department of Agriculture Kearneysville, National Center for Cool and Cold Water Aquaculture, United States, Agricultural Research Service, Kearneysville, WV, 25430, USA

Tim Leeds

Contributions

MS and RA conceived and designed the experiments. RA, AA, TL, and MS performed the experiments and analyzed the data. RA wrote the paper. All authors reviewed and approved the publication.

Corresponding author

Correspondence to Mohamed Salem (mosalem@umd.edu).

Ethics declarations

Ethics approval and consent to participate

Husbandry practice and experimental procedures at the facility were approved by the IACUC animal study protocol of the University of Maryland, College Park, protocol number 1593175-6. All methods were carried out in accordance with relevant guidelines and regulations. All methods were carried out in accordance with ARRIVE guidelines.

Consent for publication

Not applicable.

Competing interests

Mohamed Salem is a member of the editorial board of BMC Genomics. Otherwise, the authors declare that they have no competing interests.

Supplementary files:

Supplementary File 1. Detailed parameters used for RNA-seq processing in CLC Genomics Workbench, significant cis-eQTL associations, differential gene expression results for body weight, muscle yield, condition factor, and genetic line, integrated candidate SNP-gene analyses, candidate SNP association results from genomic mixed models, positions of candidate cis-eQTLs relative to transcription start sites, and supplementary figures.

Supplementary File 2: Complete significant trans-eQTL SNP-gene associations

References

1. Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M: **Genome-wide identification of loci associated with growth in rainbow trout**. *BMC Genomics* 2020, **21**(1).
2. Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Brett Kenney P, Vallejo RL, Palti Y, Leeds TD: **Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*)**. *Frontiers in Genetics* 2016, **7**(NOV).
3. Hu G, Gu W, Bai QL, Wang BQ: **Estimation of genetic parameters for growth traits in a breeding program for rainbow trout (*Oncorhynchus mykiss*) in China**. *Genetics and Molecular Research* 2013, **12**(2):1457-1467.
4. Reis Neto RV, Yoshida GM, Lhorente JP, Yáñez JM: **Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*)**. *Molecular Genetics and Genomics* 2019, **294**(3):563-571.
5. Cleveland BM, Gao G, Leeds TD: **Transcriptomic Response to Selective Breeding for Fast Growth in Rainbow Trout (*Oncorhynchus mykiss*)**. *Marine Biotechnology* 2020, **22**(4):539-550.
6. Codina M, Montserrat N, De La Serrana DG, Rallièrre C, Gabillard JC, Navarro I, Rescan P, Gutiérrez J: **Role of insulin and IGFs in fish muscle development and quality**. In: *Archiv fur Tierzucht: 2006*; 2006: 33-38.
7. Danzmann RG, Kocmarek AL, Norman JD, Rexroad CE, Palti Y: **Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients**. *BMC Genomics* 2016, **17**(1).

8. Hou ZS, Xin YR, Zeng C, Zhao HK, Tian Y, Li JF, Wen HS: **GHRH-SST-GH-IGF axis regulates crosstalk between growth and immunity in rainbow trout (*Oncorhynchus mykiss*) infected with *Vibrio anguillarum*.** *Fish and Shellfish Immunology* 2020, **106**:887-897.
9. Reindl KM, Kittilson JD, Bergan HE, Sheridan MA: **Growth hormone-stimulated insulin-like growth factor-1 expression in rainbow trout (*oncorhynchus mykiss*) hepatocytes is mediated by ERK, PI3K-AKT, and JAK-STAT.** *American Journal of Physiology - Regulatory Integrative and Comparative Physiology* 2011, **301**(1):R236-R243.
10. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S: **Large-scale cis-and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression.** *Nature genetics* 2021, **53**(9):1300-1310.
11. Le Bras Y, Dechamp N, Montfort J, Le Cam A, Krieg F, Quillet E, Prunet P, Le Roy P: **Acclimation to seawater in rainbow trout: a QTL/eQTL approach.** In: *9 World Congress on Genetics Applied to Livestock Production: 2010*: German Society for Animal Science; 2010: np.
12. Ahmed RO, Ali A, Leeds T, Salem M: **RNA-Seq analysis of the pyloric caecum, liver, and muscle reveals molecular mechanisms regulating fillet color in rainbow trout.** *BMC genomics* 2023, **24**(1):579.
13. Cleveland BM, Radler LM, Leeds TD: **Growth, fillet yield, and muscle quality traits are not affected by a genotype by diet interaction in rainbow trout consuming diets that differ in lipid content.** *Journal of the World Aquaculture Society* 2023.
14. Andrews S: **FastQC: a quality control tool for high throughput sequence data.** In.: Cambridge, United Kingdom; 2010.
15. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30**(15):2114-2120.
16. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.** *Nat Biotechnol* 2019, **37**(8):907-915.
17. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM *et al*: **Twelve years of SAMtools and BCFtools.** *Gigascience* 2021, **10**(2):giab008.
18. Liao Y, Smyth GK, Shi W: **The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads.** *Nucleic acids research* 2019, **47**(8):e47-e47.
19. Team RC: **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> 2016.
20. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26**(1):139-140.
21. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol* 2014, **15**(12):550.
22. Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A, Rexroad CE, 3rd, Yao J: **RNA-Seq identifies SNP markers for growth traits in rainbow trout.** *PLoS One* 2012, **7**(5):e36264.
23. Rubinacci S, Hofmeister RJ, Sousa da Mota B, Delaneau O: **Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes.** *Nature Genetics* 2023, **55**(7):1088-1090.
24. Wasik K, Berisa T, Pickrell JK, Li JH, Fraser DJ, King K, Cox C: **Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics.** *BMC Genomics* 2021, **22**(1):197.
25. Li N, Stephens M: **Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data.** *Genetics* 2003, **165**(4):2213-2233.
26. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST: **The variant call format and VCFtools.** *Bioinformatics* 2011, **27**(15):2156-2158.
27. Shabalín AA: **Matrix eQTL: ultra fast eQTL analysis via large matrix operations.** *Bioinformatics* 2012, **28**(10):1353-1358.

28. Grieve IC, Dickens NJ, Pravenec M, Kren V, Hubner N, Cook SA, Aitman TJ, Petretto E, Mangion J: **Genome-wide co-expression analysis in multiple tissues.** *PLoS One* 2008, **3**(12):e4033.
29. Yang J, Lee SH, Goddard ME, Visscher PM: **GCTA: a tool for genome-wide complex trait analysis.** *The American journal of human genetics* 2011, **88**(1):76-82.
30. Salem M, Al-Tobasei R, Ali A, Lourenco D, Gao G, Palti Y, Kenney B, Leeds TD: **Genome-wide association analysis with a 50K transcribed gene SNP-chip identifies QTL affecting muscle yield in rainbow trout.** *Frontiers in genetics* 2018, **9**:387.
31. Ahmed RO, Ali A, Al-Tobasei R, Leeds T, Kenney B, Salem M: **Weighted Single-Step GWAS Identifies Genes Influencing Fillet Color in Rainbow Trout.** *Genes (Basel)* 2022, **13**(8):1331.
32. Al-Tobasei R, Ali A, Garcia ALS, Lourenco D, Leeds T, Salem M: **Genomic predictions for fillet yield and firmness in rainbow trout using reduced-density SNP panels.** *BMC Genomics* 2021, **22**(1):92.
33. Salem M, Al-Tobasei R, Ali A, Lourenco D, Gao G, Palti Y, Kenney B, Leeds TD: **Genome-Wide Association Analysis With a 50K Transcribed Gene SNP-Chip Identifies QTL Affecting Muscle Yield in Rainbow Trout.** *Front Genet* 2018, **9**:387.
34. Westra H-J, Franke L: **From genome to function by studying eQTLs.** *Biochimica et Biophysica Acta (BBA)-molecular basis of Disease* 2014, **1842**(10):1896-1902.
35. Wang L, Babushkin N, Liu Z, Liu X: **Trans-eQTL mapping in gene sets identifies network effects of genetic variants.** *Cell Genomics* 2024, **4**(4).
36. Cai W, Zhang Y, Chang T, Wang Z, Zhu B, Chen Y, Gao X, Xu L, Zhang L, Gao H: **The eQTL colocalization and transcriptome-wide association study identify potentially causal genes responsible for economic traits in Simmental beef cattle.** *Journal of Animal Science and Biotechnology* 2023, **14**(1):78.
37. Consortium G, Ardlie KG, Deluca DS, Segrè AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T: **The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.** *Science* 2015, **348**(6235):648-660.
38. Consortium G: **The GTEx Consortium atlas of genetic regulatory effects across human tissues.** *Science* 2020, **369**(6509):1318-1330.
39. Rahman NIA, Lam CL, Sulaiman N, Abdullah NAH, Nordin F, Ariffin SHZ, Yazid MD: **PAX7, a key for myogenesis modulation in muscular dystrophies through multiple signaling pathways: a systematic review.** *International Journal of Molecular Sciences* 2023, **24**(17):13051.
40. Chen G, Zhou Y, Yu X, Wang J, Luo W, Pang M, Tong J: **Genome-Wide Association study reveals SNPs and candidate genes related to growth and body shape in Bighead carp (*Hypophthalmichthys nobilis*).** *Marine Biotechnology* 2022, **24**(6):1138-1147.
41. Torstensen B, Nanton D, Olsvik P, Sundvold H, Stubhaug I: **Gene expression of fatty acid-binding proteins, fatty acid transport proteins (cd36 and FATP) and β -oxidation-related genes in Atlantic salmon (*Salmo salar* L.) fed fish oil or vegetable oil.** *Aquaculture Nutrition* 2009, **15**(4):440-451.
42. Kolditz C, Borthaire M, Richard N, Corraze G, Panserat S, Vachot C, Lefèvre F, Médale F: **Liver and muscle metabolic changes induced by dietary energy content and genetic selection in rainbow trout (*Oncorhynchus mykiss*).** *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 2008, **294**(4):R1154-R1164.
43. Chen B, Xiao W, Zou Z, Zhu J, Li D, Yu J, Yang H: **Comparing transcriptomes reveals key metabolic mechanisms in superior growth performance Nile tilapia (*Oreochromis niloticus*).** *Frontiers in Genetics* 2022, **13**:879570.
44. Bou Mira M: **New insights into lipid and carbohydrate metabolism in teleost fish: transcriptional and functional characterization of adipocytes.** 2016.

45. Ali A, Al-Tobasei R, Kenney B, Leeds TD, Salem M: **Integrated analysis of lncRNA and mRNA expression in rainbow trout families showing variation in muscle growth and fillet quality traits.** *Scientific Reports* 2018, **8**(1):12111.
46. Jin Y, Datsomor AK, Olsen RE, Vik JO, Torgersen JS, Edvardsen RB, Wargelius A, Winge P, Grammes F: **Targeted mutagenesis of $\Delta 5$ and $\Delta 6$ fatty acyl desaturases induce dysregulation of lipid metabolism in Atlantic salmon (*Salmo salar*).** *BMC genomics* 2020, **21**(1):805.
47. Mankiewicz JL, Gao G, Leeds T, Cleveland BM: **Transcriptomic Profiles of Rainbow Trout (*Oncorhynchus mykiss*) Selectively Bred for High and Low Fillet Yield.** *Marine Biotechnology* 2025, **27**(4):102.
48. Gilson H, Schakman O, Kalista S, Lause P, Tsuchida K, Thissen J-P: **Follistatin induces muscle hypertrophy through satellite cell proliferation and inhibition of both myostatin and activin.** *American Journal of Physiology-Endocrinology and Metabolism* 2009, **297**(1):E157-E164.
49. Garabedian MJ, Harris CA, Jeanneteau F: **Glucocorticoid receptor action in metabolic and neuronal function.** *F1000Research* 2017, **6**:1208.
50. Taylor P: **Role of amino acid transporters in protein metabolism.** In: *The Molecular Nutrition of Amino Acids and Proteins*. Elsevier; 2016: 49-64.
51. Arai S, Ikeda M, Ide T, Matsuo Y, Fujino T, Hirano K, Sunagawa K, Tsutsui H: **Functional loss of DHR57C induces intracellular Ca^{2+} overload and myotube enlargement in C2C12 cells via calpain activation.** *American Journal of Physiology-Cell Physiology* 2017, **312**(1):C29-C39.
52. Yu C, Liu C, Liu B, Feng J, Deng Y, Yuan J: **Bridging metabolism and immuno-inflammation: a novel framework to characterize dilated cardiomyopathy subtype.** *Journal of Translational Medicine* 2025, **23**(1):1145.
53. Gopisetty G, Thangarajan R: **Mammalian mitochondrial ribosomal small subunit (MRPS) genes: A putative role in human disease.** *Gene* 2016, **589**(1):27-35.
54. Barilla S, Liang N, Mileti E, Ballaire R, Lhomme M, Ponnaiah M, Lemoine S, Soprani A, Gautier J-F, Amri E-Z: **Loss of G protein pathway suppressor 2 in human adipocytes triggers lipid remodeling by upregulating ATP binding cassette subfamily G member 1.** *Molecular metabolism* 2020, **42**:101066.
55. Barilla S: **Role of GPS2 in the regulation of adipocyte fate and function: a multi-omics approach.** Karolinska Institutet (Sweden); 2020.
56. Quinzii CM, López LC, Von-Moltke J, Naini A, Krishna S, Schuelke M, Salviati L, Navas P, DiMauro S, Hirano M: **Respiratory chain dysfunction and oxidative stress correlate with severity of primary CoQ10 deficiency.** *The FASEB journal: official publication of the Federation of American Societies for Experimental Biology* 2008, **22**(6):1874.
57. Artuch R, Salviati L, Jackson S, Hirano M, Navas P: **Coenzyme Q10 deficiencies in neuromuscular diseases.** In: *Inherited Neuromuscular Diseases: Translation from Pathomechanisms to Therapies*. Springer; 2009: 117-128.
58. Yang D, Tao J, Li L, Kedei N, Toth ZE, Czap A, Velasquez JF, Mihova D, Michalowski AM, Yuspa SH: **RasGRP3, a Ras activator, contributes to signaling and the tumorigenic phenotype in human melanoma.** *Oncogene* 2011, **30**(45):4590-4600.
59. Zhang Z, Jones AE, Wu W, Kim J, Kang Y, Bi X, Gu Y, Popov IK, Renfrow MB, Vassilyeva MN: **Role of remodeling and spacing factor 1 in histone H2A ubiquitination-mediated gene silencing.** *Proceedings of the National Academy of Sciences* 2017, **114**(38):E7949-E7958.
60. Barry D, Carpenter C, Yager C, Golik B, Barry K, Shen H, Mikse O, Eggert L, Schulz D, Garcia M: **Variation of the neurofilament medium KSP repeat sub-domain across mammalian species: implications for altering axonal structure.** *Journal of Experimental Biology* 2010, **213**(1):128-136.

61. Devarakonda SS, Basha S, Pithakumar A, LB T, Mukunda DC, Rodrigues J, Biswas S, Pai AR, Belurkar S, Mahato KK: **Molecular mechanisms of neurofilament alterations and its application in assessing neurodegenerative disorders.** *Ageing Research Reviews* 2024, **102**:102566.
62. Wang Y, Yu Y, Pang Y, Yu H, Zhang W, Zhao X, Yu J: **The distinct roles of zinc finger CCHC-type (ZCCHC) superfamily proteins in the regulation of RNA metabolism.** *RNA biology* 2021, **18**(12):2107-2126.
63. Di L-J, Fernandez AG, De Siervi A, Longo DL, Gardner K: **Transcriptional regulation of BRCA1 expression by a metabolic switch.** *Nature structural & molecular biology* 2010, **17**(12):1406-1413.
64. Saha PP, Kumar SP, Srivastava S, Sinha D, Pareek G, D'Silva P: **The presence of multiple cellular defects associated with a novel G50E iron-sulfur cluster scaffold protein (ISCU) mutation leads to development of mitochondrial myopathy.** *Journal of Biological Chemistry* 2014, **289**(15):10359-10377.
65. Holland SJ, Liao XC, Mendenhall MK, Zhou X, Pardo J, Chu P, Spencer C, Fu A, Sheng N, Yu P: **Functional cloning of Src-like adapter protein-2 (SLAP-2), a novel inhibitor of antigen receptor signaling.** *The Journal of experimental medicine* 2001, **194**(9):1263-1276.
66. Welsh SA, Gardini A: **Genomic regulation of transcription and RNA processing by the multitasking Integrator complex.** *Nature Reviews Molecular Cell Biology* 2023, **24**(3):204-220.
67. Cai Z, Chehab NH, Pavletich NP: **Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase.** *Molecular cell* 2009, **35**(6):818-829.
68. Stiban J, So M, Kaguni LS: **Iron-sulfur clusters in mitochondrial metabolism: multifaceted roles of a simple cofactor.** *Biochemistry (Moscow)* 2016, **81**(10):1066-1080.
69. Zhang Y, Li X, Xiang R: **SLA2 is Associated With Immune evasion and Exhaustion of CD8+ T Cells in Gastric Cancer.** *Journal of Cellular and Molecular Medicine* 2026, **30**(9):e71164.
70. Woessner NM, Uleri V, Stepanek O, Minguet S: **The TCR and LCK: foundations for T-cell activation and therapeutic innovation.** *Frontiers in immunology* 2025, **16**:1737013.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile2.txt](#)
- [SupplementaryFile1.xlsx](#)