

# Genome-wide identification of loci associated with growth in rainbow trout

**Ali Ali**

University of Maryland College Park

**Rafet Al-Tobasei**

Computational science program, Middle Tennessee State University

**Daniela Lourenco**

Department of animal and dairy science, University of Georgia, Athens

**Tim Leeds**

USDA-ARS National Center for Cool and Cold Water Aquaculture

**Brett Kenney**

Division of animal and nutritional sciences, University of West Georgia

**Mohamed Salem** (✉ [mosalem@umd.edu](mailto:mosalem@umd.edu))

University of Maryland College Park Official Bookstore <https://orcid.org/0000-0003-2142-6716>

---

## Research article

**Keywords:** Body weight, fish, genomic selection, QTL, GWAS, WssGBLUP.

**Posted Date:** February 21st, 2020

**DOI:** <https://doi.org/10.21203/rs.2.17035/v4>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on March 5th, 2020. See the published version at <https://doi.org/10.1186/s12864-020-6617-x>.

# Abstract

**Background:** Growth is a major economic production trait in aquaculture. Improvements in growth performance will reduce time and cost for fish to reach market size. However, genes underlying growth have not been fully explored in rainbow trout.

**Results:** A previously developed 50K gene-transcribed SNP chip, containing ~21K SNPs showing allelic imbalances potentially associated with important aquaculture production traits including body weight, muscle yield, was used for genotyping a total of 789 fish with available phenotypic data for bodyweight gain. Genotyped fish were obtained from two consecutive generations produced in the NCCCWA growth-selection breeding program. Weighted single-step GBLUP (WssGBLUP) was used to perform a genome-wide association (GWA) analysis to identify quantitative trait loci (QTL) associated with bodyweight gain. Using genomic sliding windows of 50 adjacent SNPs, 247 SNPs associated with bodyweight gain were identified. SNP-harboring genes were involved in cell growth, cell proliferation, cell cycle, lipid metabolism, proteolytic activities, chromatin modification, and developmental processes. Chromosome 14 harbored the highest number of SNPs ( $n = 50$ ). An SNP window explaining the highest additive genetic variance for bodyweight gain (~6.4%) included a nonsynonymous SNP in a gene encoding inositol polyphosphate 5-phosphatase OCRL-1. Additionally, based on a single-marker GWA analysis, 33 SNPs were identified in association with bodyweight gain. The highest SNP explaining variation in bodyweight gain was identified in a gene coding for thrombospondin-1 (THBS1) ( $R^2 = 0.09$ ).

**Conclusion:** The majority of SNP-harboring genes, including OCRL-1 and THBS1, were involved in developmental processes. Our results suggest that development-related genes are important determinants for growth and could be prioritized and used for genomic selection in breeding programs.

## Background

Aquaculture is a growing agribusiness that enhances food security and increases economic opportunities worldwide [1]. A key challenge for this industry is to sustain the increasing consumer demand for seafood [2]. Salmonid species have been extensively studied as cultured fish species due to their economic and nutritional value [3]. Growth performance, particularly the efficiency of converting feed to bodyweight gain, is one of the most economically important traits [3]. Growth is a complex trait controlled by environmental and genetic factors. Despite the multi-environmental factors that may affect growth, quantitative genetics studies revealed moderate to high levels of growth rate heritability [4, 5]. Thus, artificial selection for growth is plausible, allowing potential improvement through selective breeding programs [5].

Selective breeding improves heritable traits, taking advantage of existing genetic variation between individuals/families. Previous studies showed that selective breeding programs can improve animals' bodyweights, thereby contributing to increased aquaculture production [6, 7]. Selection on harvest weight can improve growth rate [8] and flesh color, and reduce production cost [9]. Successful genetic programs

depend on the establishment of a base population with natural genetic variation, which helps to achieve a long-term response to selection. A family-based selection line for growth was established in 2002 at the USDA National Center for Cool and Cold Water Aquaculture (NCCCWA). Five generations of selection yielded a 10% gain in bodyweight per generation [10] at harvest. More efforts are required to understand the genetic basis of bodyweight gain for genetically improved strains to achieve fast/efficient production [2].

QTL mapping has been extensively applied in plants and farmed animals to determine the genetic architecture of the complex traits. Several QTL mapping studies were performed to assess the genetic basis of growth in Atlantic salmon, Coho salmon, and rainbow trout [3]. For instance, a significant QTL for body weight was co-localized with another moderate-effect QTL for maturation timing in the linkage group RT-27 in rainbow trout [11-13]. In addition, QTL for body weight and condition factor were co-localized on linkage group RT-9 and RT-27 [4]. However, classical QTL mapping has some limitations. Linkage analysis is time-consuming and depends on the segregation of alleles within a family, limiting the power to detect associations between markers and phenotypes of interest [5]. In addition, the identified QTL encompasses several megabases that contain hundreds, if not thousands, of genes, making it challenging to identify the causal gene in a QTL [14].

Genomic resources have been developed for rainbow trout, including the release of the first genome assembly draft [15] and a newly assembled genome (GenBank assembly, NCBI accession GCA\_002163495, RefSeq assembly accession GCF\_002163495). New sequencing technologies have identified SNPs that are widely distributed throughout the genome; this SNP distribution enabled the construction of high-density genetic maps [16, 17]. About 90% of the genetic variation comes from SNPs that are highly adaptable to large-scale genotyping and, therefore, most suitable for GWA studies [8]. The rainbow trout genome was successfully used for calling variants [18], and these variants have been used to build a 50K transcribed gene SNP chip suitable for association mapping [19]. GWA studies have been employed to test the association between SNP markers spread throughout the genome and complex quantitative traits of interest [20]. Owing to the drastic reduction in cost and time required for genotyping a large number of markers, GWA studies are replacing QTL linkage mapping [21]. SNP markers in linkage disequilibrium (LD) with QTL associated with the trait of interest could be identified from GWA analyses and prioritized in selective breeding programs [20]. Many GWA studies conducted on livestock species led to the identification of genes and mutations associated with economic traits [20]. Recently, a few GWA studies have been implemented in aquaculture species [20], including rainbow trout. These studies aimed to identify markers associated with bodyweight [22], fillet quality [19, 22], and disease resistance [23]. Growth traits are controlled by small-effect variants in the farmed Atlantic salmon [24]. In addition, a recent GWA study using a 57K SNP array identified QTL explaining a small proportion of additive genetic variance for body weight in rainbow trout. A single window on chromosome 5 was responsible for 1.4 and 1.0% of the additive genetic variance in body weight at 10 and 13 months post-hatching, respectively [22].

In this study, we used a 50K transcribed gene SNP chip, recently developed in our laboratory, to perform GWA analyses [19]. The chip has 21K SNPs of potential associations with muscle growth, fillet quality, and disease resistance traits. In order to randomize SNP distribution in this chip, 29K additional SNPs were added to the chip following a strategy of 2 SNPs per each SNP-harboring gene. The SNP chip has been successfully used to identify QTL associated with muscle yield [19], and fillet firmness and protein content [25] in rainbow trout. The objective of this study was to use the 50K SNP array to identify large-effect QTL associated with the growth rate that could be applied in genomic selection.

## Results And Discussion

Growth performance defines fish production, and therefore, it affects aquaculture industry profitability. Progress in growth-related traits could lead to reductions in time and cost to market size [26]. Traditional selection, based on the phenotype, has been applied to select for growth traits resulting in approximately 10% gain in body weight per generation [10]. The economic significance of growth to aquaculture encouraged several studies aimed at understanding the genetic basis/mechanisms underlying the phenotype [26]. Genomic approaches have the potential to expedite genetic gains compared to traditional selection. SNPs account for 90% of sequence variants in humans [27]; therefore, SNPs are most suitable for genetic evaluation of breeding candidates in selection programs. The fish population used for the current GWA analysis had an average bodyweight gain per day of  $3.27 \pm 0.96$  (g). Variations in bodyweight gain among 789 fish used for the current GWA analysis are shown in Figure 1. The estimated heritability for bodyweight gain in rainbow trout was  $0.30 \pm 0.05$ . In this study, a 50K SNP chip was used to identify genomic regions associated with bodyweight gain, based on 50 SNP sliding windows and single-marker association analysis. It is worth mentioning that a total of 90 fish from YC2010 were used in our previous study [18] to identify putative SNPs associated with muscle growth and quality traits (WBW, muscle yield, fat content, shear force, and whiteness index). The putative SNPs showing allelic imbalance (7.9K SNPs) with the five growth and quality traits were included in the SNP chip [19]. To make sure those fish do not interfere with the GWAS results, those 90 fish were excluded from the analysis in this study.

## Identifying QTL associated with bodyweight gain using WssGBLUP

WssGBLUP-based GWA analysis identified a total of 247 SNPs associated with additive genetic variance in bodyweight gain. These SNPs exist in 107 protein-coding genes, 6 lncRNAs, and 36 intergenic regions. SNPs were identified in windows explaining at least 2% (arbitrary value) of the additive genetic variance for bodyweight gain (Table S1). The genomic regions that harbor SNPs were clustered on 7 chromosomes (2, 4, 8, 9, 13, 14, and 18) (Figure 2). Chromosome 14 had the most significant peaks associated with bodyweight gain (up to 6.37%) and the highest number of SNPs ( $n = 50$ ) in windows explaining additive genetic variance for the studied trait (Table S1, Figure 2). Many of the SNPs ( $n = 100$ ) were located within the 3'UTR of their genes suggesting a role of these SNPs in microRNA, post-

transcriptional regulation of gene expression. All QTLs associated with bodyweight gain are listed in Table (S1). To gain understandings into the biological significance of the identified QTL, we annotated SNP-harboring genes and followed this annotation by gene enrichment analysis. Functional annotation analysis showed that SNP-harboring genes were involved in cell growth, cell cycle, cell proliferation, lipid metabolism, proteolytic activities, developmental processes, and chromatin modification. Enriched terms included lysosomal proteins/enzymes and fatty acid biosynthesis (Table S2).

## **SNPs in genes regulating cell growth, cell cycle and cell proliferation**

Coordinated hypertrophy and hyperplasia are essential for growing organisms [28]. Five chromosomes (2, 4, 9, 13, and 14) had SNPs regulating cell growth, cell cycle, and cell proliferation (Table 1). Chromosome 2 had 14 SNPs in 6 genes coding for caveolin-1 (CAV-1), testin (TES), eukaryotic translation initiation factor 4 gamma 2 (EIF4G2), sodium-dependent neutral amino acid transporter B (0) AT2 (SLC6A15), kinesin-like protein KIF21A (KIF21A), and G1/S-specific cyclin-D1 (CCND1). Six SNPs spanning ~1.8 Kb were identified in CAV-1. The latter has a role in inhibiting the activity of TGF- $\beta$ , probably by enfolding TGF- $\beta$  receptors in membrane invaginations [29]. Knockdown of CAV-1 had a tumor-suppressing effect by inhibiting cell proliferation [30], arresting cells in the G0/G1 phase, and inhibiting the expression of cell cycle-related proteins such as cyclin D1 [30]. Two SNPs were identified in each of TES and EIF4G2. TES negatively regulates cell proliferation and inhibits tumor cell growth [31, 32], whereas eIF4G2 positively regulates cell growth and proliferation, prevents autophagy, and releases cells from nutrient-sensing control by mTOR [33]. Each of SLC6A15 and KIF21A had a single SNP. Depletion of SLC6A15 attenuates leucine's effects in reducing weight gain associated with a high-fat diet [34]. KIF21A has been identified in association with growth in pigs [35]. We identified 2 SNPs in the CCND1 gene. This cyclin is expressed during the G1 phase to signal initiation of DNA synthesis; it is suppressed during the S phase to allow DNA synthesis [36]. Cancer cell proliferation [37] and the growth of multifocal dysplastic lesions [38] were regulated through CCND1.

A total of 21 SNPs were identified on chromosomes 4, 9, and 13. Chromosome 4 had 9 SNPs in 3 genes coding for transcription factor AP-1 (AP-1), protein PRRC2C (PRRC2C), and myocilin (MYOC). Transcription factor AP-1 transduces growth signals to the nucleus, mediated by upregulation of positive cell cycle regulators [39], which enhance the expression of genes involved in growth [40]. Whereas PRRC2C regulates the cell cycle and cell proliferation, and it controls the growth of lung cancer cells *in vitro* [41]. MYOC had 4 nonsynonymous SNPs. Transgenic mice, with 15-fold over-expressed MYOC, exhibited skeletal muscle hypertrophy with an approximate 40% increase in muscle weight [42]. We identified 2 SNPs on chromosome 9 in the gene coding for protein RCC2 homolog. RCC2 is a crucial regulator of cell cycle progression during the interphase [43]. There were ten SNPs in 3 genes on chromosome 13. Four SNPs, spanning 2.3 Kb, were localized in a gene coding for prohibitin (PHB). This protein suppresses cell growth by controlling E2F transcriptional activity [44]. Four SNPs spanned a gene coding for cyclin-dependent kinase 12 (CDK12). Depletion of CDK12 revealed increased numbers of accumulated cells at the G2/M phase and supported a role for CDK12 in maintaining genomic stability

[45]. STAT3 had two SNPs in the 3'UTR. Knockdown of STAT3 inhibits cell proliferation and leads to irreversible growth arrest [46].

Chromosome 14 had 11 SNPs in seven genes coding for prominin-1-A (PROM1A), fibroblast growth factor-binding protein 1 (FGFBP1), cyclin A2 (CCNA2), re-initiation and release factor (MCTS1), septin-6 (SEPT6), tenomodulin (TNMD), and 60S ribosomal protein L36a (RPL36A). PROM1A has a role in cell proliferation and differentiation [47]. FGFBP1 promotes fibroblast growth factor2 (FGF2) signaling during angiogenesis, tissue repair, and tumor growth [48]. A single SNP was identified in the CCNA2 gene. This gene has a crucial role in cell cycle by regulating the initiation and progression of DNA synthesis [49]. The untranslated regions of a gene coding for MCTS1 had two SNPs in windows explaining up to ~6.4% of the additive genetic variance for bodyweight gain. Overexpression of MCTS1 promotes lymphoid tumor development leading to increased growth rates and protection against apoptosis [50]. In addition, MCTS1 is involved in cell cycle progression by decreasing the length of the G1 phase without a reciprocal increase in other phases [51]. Each of SEPT6 and RPL36A had 2 SNPs in windows associated with the additive genetic variance for bodyweight gain. Knockdown of SEPT6 leads to loss of cell polarity as a result of nuclear accumulation of the adaptor protein NCK, which arrests the cell cycle [52]. Overexpression of RPL36A leads to rapid cell cycling which enhances cell proliferation [53]. Of note, TNMD had an SNP in a window explaining 5.5% of the additive genetic variance. TNMD is essential for tenocyte proliferation and collagen fibril maturation [54]. Thirty-one genes involved in cell growth, cell cycling, and cell proliferation were differentially expressed (DE) in fish families (year class "YC" 2010), exhibiting divergent whole-body weight (WBW) phenotype. Of these genes, CAV was downregulated in families of high WBW relative to those of low WBW [55]. Our results indicate a role for increased biomass and cell numbers in explaining variations in body weight.

## **SNPs in genes regulating cell growth, cell cycle and cell proliferation**

Coordinated hypertrophy and hyperplasia are essential for growing organisms [28]. Five chromosomes (2, 4, 9, 13, and 14) had SNPs regulating cell growth, cell cycle, and cell proliferation (Table 1). Chromosome 2 had 14 SNPs in 6 genes coding for caveolin-1 (CAV-1), testin (TES), eukaryotic translation initiation factor 4 gamma 2 (EIF4G2), sodium-dependent neutral amino acid transporter B (0) AT2 (SLC6A15), kinesin-like protein KIF21A (KIF21A), and G1/S-specific cyclin-D1 (CCND1). Six SNPs spanning ~1.8 Kb were identified in CAV-1. The latter has a role in inhibiting the activity of TGF- $\beta$ , probably by enfolding TGF- $\beta$  receptors in membrane invaginations [29]. Knockdown of CAV-1 had a tumor-suppressing effect by inhibiting cell proliferation [30], arresting cells in the G0/G1 phase, and inhibiting the expression of cell cycle-related proteins such as cyclin D1 [30]. Two SNPs were identified in each of TES and EIF4G2. TES negatively regulates cell proliferation and inhibits tumor cell growth [31, 32], whereas eIF4G2 positively regulates cell growth and proliferation, prevents autophagy, and releases cells from nutrient-sensing control by mTOR [33]. Each of SLC6A15 and KIF21A had a single SNP. Depletion of SLC6A15 attenuates leucine's effects in reducing weight gain associated with a high-fat diet [34]. KIF21A has been identified in association with growth in pigs [35]. We identified 2 SNPs in the CCND1 gene. This cyclin is expressed

during the G1 phase to signal initiation of DNA synthesis; it is suppressed during the S phase to allow DNA synthesis [36]. Cancer cell proliferation [37] and the growth of multifocal dysplastic lesions [38] were regulated through CCND1.

A total of 21 SNPs were identified on chromosomes 4, 9, and 13. Chromosome 4 had 9 SNPs in 3 genes coding for transcription factor AP-1 (AP-1), protein PRRC2C (PRRC2C), and myocilin (MYOC). Transcription factor AP-1 transduces growth signals to the nucleus, mediated by upregulation of positive cell cycle regulators [39], which enhance the expression of genes involved in growth [40]. Whereas PRRC2C regulates the cell cycle and cell proliferation, and it controls the growth of lung cancer cells *in vitro* [41]. MYOC had 4 nonsynonymous SNPs. Transgenic mice, with 15-fold over-expressed MYOC, exhibited skeletal muscle hypertrophy with an approximate 40% increase in muscle weight [42]. We identified 2 SNPs on chromosome 9 in the gene coding for protein RCC2 homolog. RCC2 is a crucial regulator of cell cycle progression during the interphase [43]. There were ten SNPs in 3 genes on chromosome 13. Four SNPs, spanning 2.3 Kb, were localized in a gene coding for prohibitin (PHB). This protein suppresses cell growth by controlling E2F transcriptional activity [44]. Four SNPs spanned a gene coding for cyclin-dependent kinase 12 (CDK12). Depletion of CDK12 revealed increased numbers of accumulated cells at the G2/M phase and supported a role for CDK12 in maintaining genomic stability [45]. STAT3 had two SNPs in the 3'UTR. Knockdown of STAT3 inhibits cell proliferation and leads to irreversible growth arrest [46].

Chromosome 14 had 11 SNPs in seven genes coding for prominin-1-A (PROM1A), fibroblast growth factor-binding protein 1 (FGFBP1), cyclin A2 (CCNA2), re-initiation and release factor (MCTS1), septin-6 (SEPT6), tenomodulin (TNMD), and 60S ribosomal protein L36a (RPL36A). PROM1A has a role in cell proliferation and differentiation [47]. FGFBP1 promotes fibroblast growth factor2 (FGF2) signaling during angiogenesis, tissue repair, and tumor growth [48]. A single SNP was identified in the CCNA2 gene. This gene has a crucial role in cell cycle by regulating the initiation and progression of DNA synthesis [49]. The untranslated regions of a gene coding for MCTS1 had two SNPs in windows explaining up to ~6.4% of the additive genetic variance for bodyweight gain. Overexpression of MCTS1 promotes lymphoid tumor development leading to increased growth rates and protection against apoptosis [50]. In addition, MCTS1 is involved in cell cycle progression by decreasing the length of the G1 phase without a reciprocal increase in other phases [51]. Each of SEPT6 and RPL36A had 2 SNPs in windows associated with the additive genetic variance for bodyweight gain. Knockdown of SEPT6 leads to loss of cell polarity as a result of nuclear accumulation of the adaptor protein NCK, which arrests the cell cycle [52]. Overexpression of RPL36A leads to rapid cell cycling which enhances cell proliferation [53]. Of note, TNMD had an SNP in a window explaining 5.5% of the additive genetic variance. TNMD is essential for tenocyte proliferation and collagen fibril maturation [54]. Thirty-one genes involved in cell growth, cell cycling, and cell proliferation were differentially expressed (DE) in fish families (year class "YC" 2010), exhibiting divergent whole-body weight (WBW) phenotype. Of these genes, CAV was downregulated in families of high WBW relative to those of low WBW [55]. Our results indicate a role for increased biomass and cell numbers in explaining variations in body weight.

## **SNPs in genes regulating proteolytic activities**

A total of 19 SNPs involved in proteolytic activities were identified in 12 genes (Table 3). Out of them, 9 SNPs were located on 4 genes involved in the KEGG lysosome pathway; lysosomal associated membrane protein 2 (LAMP2), V-type proton ATPase subunit H (ATP6V1H), galactosidase alpha (GLA), and neuraminidase 1 (NEU1). Five SNPs in LAMP2 have been identified in windows explaining the highest genetic variation (~6%) in this category. LAMP2 is essential during autophagy for the fusion of autophagosomes with lysosomes [68]. ATP6V1H is a vacuolar (H<sup>+</sup>)-ATPase, which is required to acidify the phagosome/lysosome for proper processing [69]. GLA and NEU1 are lysosomal acid hydrolases (glycosidases) required to breakdown glycoproteins [70]. NEU1 was associated with suppression of ovarian carcinoma [71]. In addition, 9 SNPs were identified in 4 genes engaged in the phagosome pathway. These genes are encoding ras-related protein Rab-5C (RAB5C), ATP6V1H, LAMP2, and integrin beta-3 (ITGB3). An SNP on chromosome 4 was located in a gene coding for OMA1 zinc metalloproteinase (OMIM). The OMIM is a protease essential for mitochondrial inner membrane proteostasis maintenance [72], and its deficiency leads to increased body weight and obesity [73]. Plectin had two SNPs. Mutation in plectin results in muscular dystrophy [74]. In addition, we identified 5 SNPs located on 4 genes exhibiting peptidase activity; trypsin-3, carboxypeptidase A1, carboxypeptidase B2 (CPB2), and high choriolytic enzyme 2. Forty-three genes have functions related to protein metabolic processes and were DE in fish families (YC 2010) showing substantial variation in BW [55]. These results support a role for protein turnover in determining body weight.

## **SNPs in genes regulating developmental process and chromatin modification**

Forty-five SNPs were identified in 21 genes involved in development and chromatin remodeling (Table 4 & Table S1). Chromosome 4 had 12 SNPs in five genes coding for phosphatidylinositol glycan anchor biosynthesis class C (PIGC), SUN domain-containing ossification factor (SUCO), transmembrane emp24 domain-containing protein 5 (TMED5), histone H2A deubiquitinase MYSM1 (MYSM1), and biogenesis of lysosome-related organelles complex-1 subunit 2 (BLOS2). PIGC encodes an endoplasmic reticulum membrane protein that has been linked to embryonic lethality [75]. Mutagenesis of SUCO leads to failure of osteoblast maturation, a decrease in the synthesis of type I collagen, and eventually catastrophic defects in skeletal development [76]. The gene encoding TMED5 has GO terms belonging to chromatin binding [77]. Knockdown of MYSM1, a histone H2A deubiquitinase, led to embryonic lethality and growth retardation [78]. BLOS2 harbored 6 SNPs in windows explaining up to 4.9% of the additive genetic variance. BLOS2 is a negative regulator of the Notch system, and lack of BLOS2 in mice was embryonic lethal and led to developmental defects [79]. We identified 6 SNPs on chromosomes 8 and 9. SNPs spanned three genes (2 SNPs/gene) encoding NADH dehydrogenase [ubiquinone] flavoprotein 2 (NDUFV2), ralA binding protein 1 (RALBP1), and short-chain dehydrogenase/reductase 3 (DHRS3). NDUFV2 is involved in nervous system development [77], whereas RALBP1 was involved in the regulation of actin dynamics during embryogenesis [80]. Knockdown of DHRS3 led to a phenotype with

underdeveloped head structure and perturbed somitogenesis [81]. Chromosome 13 harbored the highest number of SNPs ( $n = 19$ ) in this category. These SNPs were located in genes coding for methyltransferase-like protein 2-A (METTL2A), telethonin (TCAP), synaptonemal complex protein SC65 (SC65), peptidyl-prolyl cis-trans isomerase FKBP10 (FKBP10), 2',3'-cyclic-nucleotide 3'-phosphodiesterase (CNP), and histone acetyltransferase KAT2A (KAT2A). METTL2A has GO terms belonging to methyltransferase activity [77]. Four SNPs were identified in TCAP. TCAP-null mice exhibit abnormal myofiber size variation and increased levels of TCAP binding protein, myostatin [82]. SC65 had two SNPs; whereas, FKBP10 had 4 SNPs. SC65 is expressed during skeletal development and acts as a regulator of bone mass homeostasis. Lack of SC65 leads to a progressive osteopenia [83]. Loss of function mutations in FKBP10 resulted in mice that were not able to survive birth, and embryos exhibited a growth delay and tissue fragility [84]. CNP had the highest number of SNPs on chromosome 13. This protein regulates blood supply to the developing embryo [85]. KAT2A encodes a protein that acts as a histone H3 succinyltransferase and exhibits a role in tumor cell proliferation and development [86]. KAT2A is involved in the regulation of developmental processes by mediating acetylation of TBX5 [87]. Six SNPs were identified on chromosome 14 in genes coding for Rap guanine nucleotide exchange factor 2 (RAPGEF2), glutathione S-transferase P (GSTP1), inositol polyphosphate 5-phosphatase OCRL-1 (OCRL), ETS-related transcription factor Elf-1 (ELF1), and mediator of RNA polymerase II transcription subunit 12 (MED12). OCRL was located in a window explaining the highest genetic variation in bodyweight gain (~6.4%), followed by ELF1 (~5.5%). Lacking both OCRL and its paralog (Inpp5b) led to the early lethality of mice embryos [88]. ELF1 has a role in maintaining cell polarity during development [89]. In addition, chromosome 18 had 2 SNPs in a gene encoding double-strand-break repair protein rad21 homolog (RAD21) (Table S1), which is involved in chromatin binding [77]. Sixty-three genes involved in development were DE in fish families (YC 2010) exhibiting divergent WBW phenotypes [55]. In agreement with a recent GWA study in rainbow trout [90], our results suggest a major role for genes involved in development in regulating genetic variation in bodyweight gain.

## Single marker association analysis

Genotyped SNPs were filtered out at a minor allele frequency (MAF)  $< 0.05$  and Hardy–Weinberg equilibrium (HWE) ( $p < 0.001$ ) yielding 29,451 filtered SNPs. In order to identify single SNP markers associated with bodyweight gain, filtered SNPs were subjected to a general linear regression analysis which allows accounting for multiple fixed effects but does not account for familial correlation. Next, residuals of the regression model were regressed on the genetic factors using QFAM, available in PLINK [91], which corrects for the family structure through a special permutation procedure. A total of 738 SNPs were significantly associated with the bodyweight gain (empirical  $p$ -value  $< 0.001$ ) following 20,000 permutations. However, a two-stage analysis that calculates residual-outcome from the regression of the outcome on multiple covariates then uses the adjusted-outcome for downstream analysis, showed bias and loss of power in genetic association studies [92, 93]. Therefore, we performed a family-based association analysis using a generalized score test which allows for multiple covariates. A total of 42 SNPs were identified associated with the bodyweight gain after accounting for multiple comparisons

(Bonferroni-corrected  $p$  "BONF"  $< 1.70E-06$ ). In order to avoid false positives, the common SNPs between the two-stage and generalized score tests were considered significantly associated with the variation in bodyweight gain (Table S3). In this study, we have identified 33 common SNPs spread over 13 chromosomes with a potential impact on the bodyweight gain (Bonferroni-corrected  $p$  "BONF"  $< 1.70E-06$ ; Table S3 & Figure 3). One-third of the identified SNPs (33.33%) spanned chromosome 15. SNP-harboring genes were involved in development, cell growth, cell proliferation, and proteolysis. Genes explaining the highest variation in bodyweight gain are coding for thrombospondin-1 (THBS1), microtubule-associated protein 4 (MAP4), D-3-phosphoglycerate dehydrogenase (PHGDH), calyntenin-1, nucleolar protein 16 (NOP16), and butyrophilin subfamily 1 member A1 (BTN1A1) (Table 5). THBS1 and MAP4, ranked at the top of the list, explaining ~9% and 6% of the variation in bodyweight gain, respectively. THBS1 is involved in complex biological processes, including angiogenesis and tissue development [94]. Mutation in THBS1 was associated with vascular permeability, accounting for embryonic lethality [75]. Interestingly, seven SNPs spanning ~21Kb on chromosome 15, were identified in the gene coding for MAP4. In mice, blocking the expression of muscle-specific MAP4 transcript didn't affect the myoblast growth, but rather severely perturbed the myotube formation indicating a critical role in myogenesis [95]. PHGDH was upregulated in fully differentiated myotubes relative to myoblasts [96]. In addition, three synonymous SNPs were identified in calyntenin-1, NOP16, and BTN1A1. Each SNP explained ~3% of the variation in bodyweight gain. Two intronic SNPs were previously identified in the calyntenin-1 gene affecting the genetic variance for fillet yield and weight in rainbow trout [22]. NOP16 regulates rRNA production and ribosomal biogenesis. Knockdown of NOP16 dramatically reduced tumor cell growth [97]. BTN1A1 has a function in cell proliferation and development [98].

Three missense mutations were identified in genes coding for collagenase-3 (MMP13), elongation factor 2 (eEF2), and basic leucine zipper and W2 domain-containing protein 1-A (Table 5). Each SNP explained ~2% of the variation in bodyweight gain. MMP13 plays a critical role in skeletal system development [99]. eEF2 is a key component in the translation machinery. Inactivation of eEF2 terminates protein synthesis and causes cellular death during mouse embryonic development [100]. An SNP was identified in a gene encoding death-associated protein kinase 3 (Table 5). This protein is involved in the regulation of autophagy [101]. Notably, five 3'UTR mutations were identified in a gene coding for polymerase I and transcript release factor (PTRF/cavin-1) (Table 5). Lack of cavin-1 in mice and humans caused muscular dystrophy [102]. Cavin-1 supports cell proliferation and migration in humans and shows downregulated expression during myogenic differentiation [103]. The remaining SNPs associated with the variation in bodyweight gain are listed in Additional Table (S3).

Single SNP GWA analysis provided an additional set of SNPs, potentially regulating variation in bodyweight gain. In the current study, dividing the genome into chromosomal segments/windows, defined by 50 adjacent markers, outperformed the single-marker analysis in identifying a larger number of SNPs (247 vs. 33 SNPs, respectively) describing the genetic architecture of the studied trait. On chromosome 13, there was a single common significant SNP detected by the two GWA approaches in a gene coding for synaptic vesicle membrane protein VAT-1 homolog (VAT-1). This protein interacts with Talin-1; the key driver of cell migration [104]. Similar results have been previously reported in rainbow

trout [25]. Compared to p-value based peaks, the highest peaks based on the variance explained depends on allele frequency which means a high-effect SNP of low frequency reduces the variance explained [105]. The WssGBLUB method has been proven to be optimal in livestock populations with a large number of phenotyped animals with a long history of pedigree recording, but lacking genotype data [105]. The two GWA approaches adopted in the current study revealed significant roles of genes related to developmental process in regulating bodyweight gain. Routine use of single-SNP and multi-marker for GWA analysis has been recommended to take advantage of the complete genotype information [106].

Consistent with our data, a previous GWA study in rainbow trout identified small-effect QTL on chromosome 9 that affected additive genetic variance for bodyweight [90]. However, QTL associated with growth rate varied between the studies, and this discrepancy may be due to testing of different populations and gene-by-environment interactions. A 57K genomic SNP panel has been exploited for GWA analysis, using the same fish population as the current study; the study identified one window on chromosome 5 with small effects on the additive genetic variance for body weight. The window explained 1.38 and 0.95% of the additive genetic variance for body weight at 10 and 13 months, respectively [22]. However, this window was not identified in our study, perhaps, because we considered only windows explaining 2% of the additive genetic variance or more. Several markers, each explaining less than 0.1% of the variance, were identified to be associated with body weight in a GWA study for Atlantic salmon [20]. Fish population, marker density, LD, and size of adjacent SNP windows may, partially, explain the discrepancies in the results obtained from the different studies. In addition, SNPs used in the current study were identified from fish families of extreme phenotypes and thus, perhaps, are more informative for the current GWA analysis [19]. In agreement with previous GWA studies, growth is multifactorial in nature, and growth-related genes regulate development, cell proliferation, energy metabolism, and growth [90, 107]. Overall, the current study further describes the genetic architecture of the studied trait and provides putative markers for breeding candidates that can be used for selection purposes.

## Conclusions

The current GWA study identified growth-related QTL and novel genes associated with the growth rate in rainbow trout. Compared to previous GWA studies in Atlantic salmon and rainbow trout, this work revealed relatively large-effect QTL associated with growth, which appears to be a polygenic trait in nature controlled by many genes on multiple chromosomes. Chromosomes 4 and 14 had the most significant peaks that explained a reasonable proportion of the additive genetic variance for bodyweight gain. The majority of SNP were within genes involved in developmental processes. Intriguingly, the gene harboring the most significant nonsynonymous SNP was previously reported to encode a protein vital to embryonic development. These findings provide a genetic basis that will enhance our understanding of the molecular mechanisms regulating growth in teleost fish as well as provide putative markers that could be prioritized when estimating genomic breeding values for growth rate.

# Methods

## *Fish population, tissue sampling, and phenotype*

Fish population was previously described [19, 25]. Briefly, fish bodyweight data were collected from two consecutive generations (YC 2010 & 2012) produced from the NCCCWA growth-selection breeding program. The NCCCWA breeding program was established in 2002 and has continued for 5 generations of selection producing full-sib families as previously described [10]. Fish used in the current study were harvested from their respective families to allow for measuring other lethal phenotypes, as we previously described [19, 25]. Fish were euthanized with an overdose of MS-222 at a concentration of 300 mg/L. Breeding, hatching, and feeding schedules were previously reported in detail [18].

A total of 789 fish representing 98 families from YC 2010 and 99 families from YC 2012 were phenotyped. For fish sampling of each generation, a single fish from each family was randomly assigned to one of five collection groups (~100 fish each) over five consecutive weeks (one group/week). The YC 2010 fish were collected between 410- and 437-days post-hatch with a mean bodyweight of 985 g (SD = 239 g). Fish from the YC 2012 were collected between 446- and 481-days post-hatch with a mean bodyweight of 1,803 g (SD = 305 g). The bodyweight gain was calculated as the fish body weight in grams divided by the fish age in days. The pedigree-based heritability  $h^2$  ( $h^2_{ped}$ ) for growth was estimated according to Zaitlen et al., [108].

## *SNP genotyping and quality control*

The 50K transcribed gene SNP-chip used in this study was recently developed and used to identify QTL associated with muscle yield [19], fillet firmness and protein content [25]. Sources of all SNPs used to build the current SNP chip were previously described [18].

As described before, a total of 1,728 fish from the NCCCWA growth- and Bacterial Cold Water Disease (BCWD)-selection lines [19] were used to assess the quality of this Affymetrix SNP chip. The SNP chip and sample metrics were reported in our previous publication [19]. Assessment of quality control (QC) and filtration of samples/genotypes have been performed using the Affymetrix SNPolar software at the default parameters [109]. A total of 789 genotyped fish had available phenotypic data for bodyweight gain and passed the QC; those were used for the current GWA analyses.

## *Fifty-SNP window GWA analysis*

The Weighted single-step GBLUP (WssGBLUP) has been used to perform GWA analysis, as we previously described [19, 25]. WssGBLUP allows genotyped and ungenotyped animals to be used at the same time, and integrates phenotype, genotype and pedigree information using a mixed model for single-trait analysis as previously described [19, 25]:

**See formula 1 in the supplemental files.**

where  $y$  is the vector of the phenotypes,  $b$  is the vector of fixed effects including fish data-collection group and hatch-year,  $a$ ,  $w$ , and  $e$  are the vectors of direct additive genetic (i.e., animal effect), random family, and residual effects, respectively. The matrices  $X$ ,  $Z_1$ , and  $Z_2$  are incidence matrices for the effects contained in  $b$ ,  $a$ , and  $w$ , respectively.

This model combines all the relationship information based on pedigree and genotypes into a single matrix ( $H^{-1}$ ):

**See formula 2 in the supplemental files.**

where  $H^{-1}$  is the inverse of the realized relationship matrix ( $H$ ),  $H^{-1}$  is the inverse of the relationship matrix based on pedigree information,  $H^{-1}$  is the inverse of the pedigree relationship matrix for genotyped animals only, and  $H^{-1}$  is the inverse of the genomic relationship matrix.

A modified REMLF90 (AIREMLF90) [110] was used to estimate variances using the Average-Information algorithm. The inbreeding value, accounted for the construction of the inverse of the pedigree relationship matrix, was previously calculated using INBUPGF90 [19, 111]. Pedigree data of 63,808 fish produced from the NCCCWA growth-selection line over five consecutive generations, were fed to INBUPGF90 to calculate the inbreeding value. Using PREGSF90 [111], 35,322 SNPs (70.6%) passed the QC at the following settings; MAF > 0.05, call rate for SNP and samples > 0.90, and HWE < 0.15.

Similar to our previous WssGBLUP analyses [19, 25], two iterations were used in the current analysis where all SNPs were equally weighted (i.e., weight = 1.0) during the first iteration. POSTGSF90 [111] was used to compute SNP effects and weights using sliding windows of 50 adjacent SNPs. The qqman package in R was used to plot the proportion of additive genetic variance explained by every 50 SNPs-genomic window [112].

## Single marker GWA analysis

Two different algorithms were used to perform family-based association analysis of the SNP genotypes with bodyweight gain, and detect signals robust for population stratification. First, QFAM in PLINK version 1.07 [91] was used to perform the family-based association analysis using permutations. QFAM does not allow accounting for the significant contribution of the variables (such as fish data-collection groups and YC) to the predictive power of bodyweight gain model. Therefore, the outcome was adjusted in a linear model in an R package to account for fixed effects (data-collection group and YC) and population stratification using the first two principal components. In the linear model of association using QFAM, the adjusted-outcome was regressed on allele count and the family structure was corrected using 20,000 permutations. Second, a family-based association analysis was performed using a generalized score test [113]. This test accounts for familial correlation using a kinship matrix and allows for multiple covariates. P-values were adjusted by Bonferroni correction to account for multiple testing. The qqman package was used to generate a Manhattan plot showing  $-\log_{10}$  (observed p-value) obtained from the GWA analysis.

# Gene annotation and enrichment analysis

SNPs bed file and the rainbow trout genome gff file were provided to Bedtools to annotate the SNPs as previously described [19, 114]. To perform gene enrichment analysis, SNP-harboring genes were uploaded to the Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 [115, 116]. In order to avoid counting duplicated genes, Fisher Exact statistics were calculated based on DAVID gene IDs, which remove redundancies in the original IDs. The list of annotation terms and their associated genes were filtered out based on Fisher Exact  $< 0.05$ .

## Declarations

## Acknowledgments

The authors acknowledge J. Everson, M. Hostuttler, K. Jenkins, J. Kretzer, J. McGowan, K. Melody, T. Moreland, and D. Payne for technical assistance. The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the USDA or the ARS of any product or service to the exclusion of others that may be suitable. USDA is an equal opportunity provider and employer.

## Funding

This study was supported by a competitive grant No. 2014-67015-21602 from the United States Department of Agriculture, National Institute of Food and Agriculture (MS), and by the USDA, Agricultural Research Service CRIS Project 1930-31000-010 “Utilizing Genetics and Physiology for Enhancing Cool and Cold Water Aquaculture Production”. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agents. R.A.T trainee’s projects are supported by Grant Number T32HL072757 from the National Heart, Lung, and Blood Institute.

## Availability of data and materials

All datasets generated for this study are included in the manuscript and/or the Additional Files. The genotypes (ped and .map files) and phenotypes are available in our previous publication [25].

## Authors’ contributions

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

### *Ethics approval and consent to participate*

Institutional Animal Care and Use Committee of the United States Department of Agriculture, National Center for Cool and Cold Water Aquaculture (Leetown, WV) specifically reviewed and approved all husbandry practices used in this study (IACUC protocol #056).

### ***Consent for publication***

Not applicable.

### ***Competing interests***

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that they have no competing interests

## **Abbreviations**

**WssGBLUP:** Weighted single-step GBLUP

**GWA:** Genome-wide association

**SNP:** Single nucleotide polymorphism

**QTL:** Quantitative trait loci

**NCCCWA:** USDA National Center of Cool and Cold Water Aquaculture

**LD:** Linkage disequilibrium

**DE:** Differentially expressed

**YC:** Year class

**BCWD:** Bacterial Cold Water Disease

**DAVID:** Database for Annotation, Visualization and Integrated Discovery

**MAF:** Minor allele frequency

**HWE:** Hardy–Weinberg equilibrium

**UTR:** untranslated region

**GO:** Gene ontology

**QC:** Quality control

# References

1. Burbridge, Hendrick, Roth, Rosenthal: **Social and economic policy issues relevant to marine aquaculture**. *Journal of Applied Ichthyology* 2001, **17**(4):194-206.
2. Fornshell G: **Rainbow Trout – Challenges and Solutions**. In., vol. 10; 2002: 545–557.
3. Tsai HY, Hamilton A, Guy DR, Tinch AE, Bishop SC, Houston RD: **The genetic architecture of growth and fillet traits in farmed Atlantic salmon (*Salmo salar*)**. *BMC Genet* 2015, **16**:51.
4. Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG: **Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*)**. *BMC Genet* 2010, **11**:63.
5. Gutierrez AP, Yanez JM, Fukui S, Swift B, Davidson WS: **Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*)**. *PLoS One* 2015, **10**(3):e0119730.
6. Gjedrem T: **Breeding plans for rainbow trout**. In: *The Rainbow Trout: Proceedings of the First Aquaculture-sponsored Symposium held at the Institute of Aquaculture, University of Sterling, Scotland*. Edited by Gall GAE, vol. 100; 1992: 73–83.
7. Gjedrem T: **Genetic variation in quantitative traits and selective breeding in fish and shellfish**. *Aquaculture* 1983, **33**(1-4):51-72.
8. 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.
9. Dufflocq P, Lhorente JP, Bangera R, Neira R, Newman S, Yáñez JM: **Correlated response of flesh color to selection for harvest weight in coho salmon (*Oncorhynchus kisutch*)**. *Aquaculture* 2017, **472**:38-43.
10. Leeds TD, Vallejo RL, Weber GM, Pena DG, Silverstein JS: **Response to five generations of selection for growth performance traits in rainbow trout (*Oncorhynchus mykiss*)**. *Aquaculture* 2016, **465**:341-351.
11. Drew RE, Schwabl H, Wheeler PA, Thorgaard GH: **Detection of QTL influencing cortisol levels in rainbow trout (*Oncorhynchus mykiss*)**. *Aquaculture* 2007, **272**(1):S183-S194.
12. Haidle L, Janssen JE, Gharbi K, Moghadam HK, Ferguson MM, Danzmann RG: **Determination of quantitative trait loci (QTL) for early maturation in rainbow trout (*Oncorhynchus mykiss*)**. *Mar Biotechnol (NY)* 2008, **10**(5):579-592.
13. O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM: **Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes**. *J Hered* 2003, **94**(4):273-284.
14. Price AH: **Believe it or not, QTLs are accurate!** *Trends Plant Sci* 2006, **11**(5):213-216.
15. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B, Bento P, Da Silva C, Labadie K, Alberti A *et al*: **The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates**. *Nat Commun* 2014, **5**:3657.

16. Al Tobasei R, Palti Y, Wiens GD, M. S: **Identification of SNPs with allelic imbalances in rainbow trout genetic lines showing different susceptibility to infection with *Flavobacterium psychrophilum***. In: *PAG-XXV Plant & Animal Genomes Conference: January 14-18 2017;; San Diego, California.;* 2017;.
17. Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E, Yao J, Thorgaard GH, Salem M, Rexroad CE, 3rd: **A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids**. *Mol Ecol Resour* 2014, **14**(3):588-596.
18. Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B, Salem M: **Identification of SNPs associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-Seq samples in rainbow trout**. *BMC Genomics* 2017, **18**(1):582.
19. 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):387.
20. Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, Matika O, Bishop SC, Houston RD: **Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array**. *BMC Genomics* 2015, **16**:969.
21. Schielzeth H, Husby A: **Challenges and prospects in genome-wide quantitative trait loci mapping of standing genetic variation in natural populations**. *Ann N Y Acad Sci* 2014, **1320**:35-57.
22. Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, 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*)**. *Front Genet* 2016, **7**:203.
23. Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE, 3rd, Wiens GD: **Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing**. *PLoS One* 2015, **10**(9):e0138435.
24. Yoshida GM, Lhorente JP, Carvalheiro R, Yanez JM: **Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (*Salmo salar L.*)**. *Anim Genet* 2017, **48**(6):698-703.
25. Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M: **Genome-Wide Association Study Identifies Genomic Loci Affecting Filet Firmness and Protein Content in Rainbow Trout**. *Front Genet* 2019, **10**(386):386.
26. Li N, Zhou T, Geng X, Jin Y, Wang X, Liu S, Xu X, Gao D, Li Q, Liu Z: **Identification of novel genes significantly affecting growth in catfish through GWAS analysis**. *Mol Genet Genomics* 2018, **293**(3):587-599.
27. Collins FS, Brooks LD, Chakravarti A: **A DNA polymorphism discovery resource for research on human genetic variation**. *Genome Res* 1998, **8**(12):1229-1231.
28. Goranov AI, Cook M, Ricicova M, Ben-Ari G, Gonzalez C, Hansen C, Tyers M, Amon A: **The rate of cell growth is governed by cell cycle stage**. *Genes Dev* 2009, **23**(12):1408-1422.
29. Le Saux CJ, Teeters K, Miyasato SK, Hoffmann PR, Bollt O, Douet V, Shohet RV, Broide DH, Tam EK: **Down-regulation of caveolin-1, an inhibitor of transforming growth factor-beta signaling, in acute allergen-induced airway remodeling**. *J Biol Chem* 2008, **283**(9):5760-5768.

30. Wang R, Li Z, Guo H, Shi W, Xin Y, Chang W, Huang T: **Caveolin 1 knockdown inhibits the proliferation, migration and invasion of human breast cancer BT474 cells.** *Mol Med Rep* 2014, **9**(5):1723-1728.
31. Wang Z, Wei H, Yu Y, Sun J, Yang Y, Xing G, Wu S, Zhou Y, Zhu Y, Zhang C *et al*: **Characterization of Ceap-11 and Ceap-16, two novel splicing-variant-proteins, associated with centrosome, microtubule aggregation and cell proliferation.** *J Mol Biol* 2004, **343**(1):71-82.
32. Tobias ES, Hurlstone AF, MacKenzie E, McFarlane R, Black DM: **The TES gene at 7q31.1 is methylated in tumours and encodes a novel growth-suppressing LIM domain protein.** *Oncogene* 2001, **20**(22):2844-2853.
33. Ramirez-Valle F, Braunstein S, Zavadil J, Formenti SC, Schneider RJ: **eIF4GI links nutrient sensing by mTOR to cell proliferation and inhibition of autophagy.** *J Cell Biol* 2008, **181**(2):293-307.
34. Drgonova J, Jacobsson JA, Han JC, Yanovski JA, Fredriksson R, Marcus C, Schioth HB, Uhl GR: **Involvement of the neutral amino acid transporter SLC6A15 and leucine in obesity-related phenotypes.** *PLoS One* 2013, **8**(9):e68245.
35. Fernandez AI, Perez-Montarelo D, Barragan C, Ramayo-Caldas Y, Ibanez-Escriche N, Castello A, Noguera JL, Silio L, Folch JM, Rodriguez MC: **Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip.** *BMC Genet* 2012, **13**:41.
36. Yang K, Hitomi M, Stacey DW: **Variations in cyclin D1 levels through the cell cycle determine the proliferative fate of a cell.** *Cell Div* 2006, **1**:32.
37. Okabe H, Lee SH, Phuchareon J, Albertson DG, McCormick F, Tetsu O: **A critical role for FBXW8 and MAPK in cyclin D1 degradation and cancer cell proliferation.** *PLoS One* 2006, **1**:e128.
38. Williams TM, Cheung MW, Park DS, Razani B, Cohen AW, Muller WJ, Di Vizio D, Chopra NG, Pestell RG, Lisanti MP: **Loss of caveolin-1 gene expression accelerates the development of dysplastic mammary lesions in tumor-prone transgenic mice.** *Mol Biol Cell* 2003, **14**(3):1027-1042.
39. Shaulian E, Karin M: **AP-1 as a regulator of cell life and death.** *Nat Cell Biol* 2002, **4**(5):E131-136.
40. Shen Q, Uray IP, Li Y, Krisko TI, Strecker TE, Kim HT, Brown PH: **The AP-1 transcription factor regulates breast cancer cell growth via cyclins and E2F factors.** *Oncogene* 2008, **27**(3):366-377.
41. de Miguel FJ, Sharma RD, Pajares MJ, Montuenga LM, Rubio A, Pio R: **Identification of alternative splicing events regulated by the oncogenic factor SRSF1 in lung cancer.** *Cancer Res* 2014, **74**(4):1105-1115.
42. Lynch JM, Dolman AJ, Guo C, Dolan K, Xiang C, Reda S, Li B, Prasanna G: **Mutant myocilin impacts sarcomere ultrastructure in mouse gastrocnemius muscle.** *PLoS One* 2018, **13**(11):e0206801.
43. Yenjerla M, Panopoulos A, Reynaud C, Fotedar R, Margolis RL: **TD-60 is required for interphase cell cycle progression.** *Cell Cycle* 2013, **12**(5):837-841.
44. Wang S, Faller DV: **Roles of prohibitin in growth control and tumor suppression in human cancers.** *Transl Oncogenomics* 2008, **3**:23-37.
45. Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, Cimermancic P, Ule J, Peterlin BM: **The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression**

- of DNA damage response genes.** *Genes Dev* 2011, **25**(20):2158-2172.
46. Sherry MM, Reeves A, Wu JK, Cochran BH: **STAT3 is required for proliferation and maintenance of multipotency in glioblastoma stem cells.** *Stem Cells* 2009, **27**(10):2383-2392.
47. Mak AB, Pehar M, Nixon AM, Williams RA, Uetrecht AC, Puglielli L, Moffat J: **Post-translational regulation of CD133 by ATase1/ATase2-mediated lysine acetylation.** *J Mol Biol* 2014, **426**(11):2175-2182.
48. Tassi E, Al-Attar A, Aigner A, Swift MR, McDonnell K, Karavanov A, Wellstein A: **Enhancement of fibroblast growth factor (FGF) activity by an FGF-binding protein.** *J Biol Chem* 2001, **276**(43):40247-40253.
49. Arsic N, Bendris N, Peter M, Begon-Pescia C, Rebouissou C, Gadea G, Bouquier N, Bibeau F, Lemmers B, Blanchard JM: **A novel function for Cyclin A2: control of cell invasion via RhoA signaling.** *J Cell Biol* 2012, **196**(1):147-162.
50. Levenson AS, Thurn KE, Simons LA, Veliceasa D, Jarrett J, Osipo C, Jordan VC, Volpert OV, Satcher RL, Jr., Gartenhaus RB: **MCT-1 oncogene contributes to increased in vivo tumorigenicity of MCF7 cells by promotion of angiogenesis and inhibition of apoptosis.** *Cancer Res* 2005, **65**(23):10651-10656.
51. Prosniak M, Dierov J, Okami K, Tilton B, Jameson B, Sawaya BE, Gartenhaus RB: **A novel candidate oncogene, MCT-1, is involved in cell cycle progression.** *Cancer Res* 1998, **58**(19):4233-4237.
52. Kremer BE, Adang LA, Macara IG: **Septins regulate actin organization and cell-cycle arrest through nuclear accumulation of NCK mediated by SOCS7.** *Cell* 2007, **130**(5):837-850.
53. Kim JH, You KR, Kim IH, Cho BH, Kim CY, Kim DG: **Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma.** *Hepatology* 2004, **39**(1):129-138.
54. Docheva D, Hunziker EB, Fassler R, Brandau O: **Tenomodulin is necessary for tenocyte proliferation and tendon maturation.** *Mol Cell Biol* 2005, **25**(2):699-705.
55. 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.** *Sci Rep* 2018, **8**(1):12111.
56. Lin R, Tao R, Gao X, Li T, Zhou X, Guan KL, Xiong Y, Lei QY: **Acetylation stabilizes ATP-citrate lyase to promote lipid biosynthesis and tumor growth.** *Mol Cell* 2013, **51**(4):506-518.
57. Chen JW, Dodia C, Feinstein SI, Jain MK, Fisher AB: **1-Cys peroxiredoxin, a bifunctional enzyme with glutathione peroxidase and phospholipase A2 activities.** *J Biol Chem* 2000, **275**(37):28421-28427.
58. Bou M, Montfort J, Le Cam A, Ralliere C, Leuret V, Gabillard JC, Weil C, Gutierrez J, Rescan PY, Capilla E *et al.*: **Gene expression profile during proliferation and differentiation of rainbow trout adipocyte precursor cells.** *BMC Genomics* 2017, **18**(1):347.
59. Puig-Oliveras A, Revilla M, Castello A, Fernandez AI, Folch JM, Ballester M: **Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat.** *Sci Rep* 2016, **6**:31803.

60. Aguan K, Scott J, See CG, Sarkar NH: **Characterization and chromosomal localization of the human homologue of a rat AMP-activated protein kinase-encoding gene: a major regulator of lipid metabolism in mammals.** *Gene* 1994, **149**(2):345-350.
61. Miklos R, Xu X, Lametsch R: **Application of pork fat diacylglycerols in meat emulsions.** *Meat Sci* 2011, **87**(3):202-205.
62. Luong A, Hannah VC, Brown MS, Goldstein JL: **Molecular characterization of human acetyl-CoA synthetase, an enzyme regulated by sterol regulatory element-binding proteins.** *J Biol Chem* 2000, **275**(34):26458-26466.
63. Das AK, Uhler MD, Hajra AK: **Molecular cloning and expression of mammalian peroxisomal trans-2-enoyl-coenzyme A reductase cDNAs.** *J Biol Chem* 2000, **275**(32):24333-24340.
64. Soccio RE, Breslow JL: **StAR-related lipid transfer (START) proteins: mediators of intracellular lipid metabolism.** *J Biol Chem* 2003, **278**(25):22183-22186.
65. Olsen RK, Olpin SE, Andresen BS, Miedzybrodzka ZH, Pourfarzam M, Merinero B, Frerman FE, Beresford MW, Dean JC, Cornelius N *et al*: **ETFDH mutations as a major cause of riboflavin-responsive multiple acyl-CoA dehydrogenation deficiency.** *Brain* 2007, **130**(Pt 8):2045-2054.
66. Ritchie MK, Johnson LC, Clodfelter JE, Pemble CWt, Fulp BE, Furdul CM, Kridel SJ, Lowther WT: **Crystal Structure and Substrate Specificity of Human Thioesterase 2: INSIGHTS INTO THE MOLECULAR BASIS FOR THE MODULATION OF FATTY ACID SYNTHASE.** *J Biol Chem* 2016, **291**(7):3520-3530.
67. Leeds T, Kenney P, Manor M: **Genetic parameter estimates for feed intake, body composition, and fillet quality traits in a rainbow trout population selected for improved growth.** In: *International Symposium on Genetics in Aquaculture*. Auburn, AL; 2012: 259.
68. Hubert V, Peschel A, Langer B, Groger M, Rees A, Kain R: **LAMP-2 is required for incorporating syntaxin-17 into autophagosomes and for their fusion with lysosomes.** *Biol Open* 2016, **5**(10):1516-1529.
69. Trombetta ES, Ebersold M, Garrett W, Pypaert M, Mellman I: **Activation of lysosomal function during dendritic cell maturation.** *Science* 2003, **299**(5611):1400-1403.
70. Winchester B: **Lysosomal metabolism of glycoproteins.** *Glycobiology* 2005, **15**(6):1R-15R.
71. Chahal HS, Wu W, Ransohoff KJ, Yang L, Hedlin H, Desai M, Lin Y, Dai HJ, Qureshi AA, Li WQ *et al*: **Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma.** *Nat Commun* 2016, **7**:12510.
72. Rainbolt TK, Lebeau J, Puchades C, Wiseman RL: **Reciprocal Degradation of YME1L and OMA1 Adapts Mitochondrial Proteolytic Activity during Stress.** *Cell Rep* 2016, **14**(9):2041-2049.
73. Quiros PM, Ramsay AJ, Sala D, Fernandez-Vizarra E, Rodriguez F, Peinado JR, Fernandez-Garcia MS, Vega JA, Enriquez JA, Zorzano A *et al*: **Loss of mitochondrial protease OMA1 alters processing of the GTPase OPA1 and causes obesity and defective thermogenesis in mice.** *EMBO J* 2012, **31**(9):2117-2133.

74. Natsuga K, Nishie W, Shinkuma S, Arita K, Nakamura H, Ohyama M, Osaka H, Kambara T, Hirako Y, Shimizu H: **Plectin deficiency leads to both muscular dystrophy and pyloric atresia in epidermolysis bullosa simplex.** *Hum Mutat* 2010, **31**(10):E1687-1698.
75. Shamseldin HE, Tulbah M, Kurdi W, Nemer M, Alsahan N, Al Mardawi E, Khalifa O, Hashem A, Kurdi A, Babay Z *et al.*: **Identification of embryonic lethal genes in humans by autozygosity mapping and exome sequencing in consanguineous families.** *Genome Biol* 2015, **16**:116.
76. Sohaskey ML, Jiang Y, Zhao JJ, Mohr A, Roemer F, Harland RM: **Osteopotencia regulates osteoblast maturation, bone formation, and skeletal integrity in mice.** *J Cell Biol* 2010, **189**(3):511-525.
77. Gaudet P, Livstone MS, Lewis SE, Thomas PD: **Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium.** *Brief Bioinform* 2011, **12**(5):449-462.
78. Huo Y, Li BY, Lin ZF, Wang W, Jiang XX, Chen X, Xi WJ, Yang AG, Chen SY, Wang T: **MYSM1 Is Essential for Maintaining Hematopoietic Stem Cell (HSC) Quiescence and Survival.** *Med Sci Monit* 2018, **24**:2541-2549.
79. Zhou W, He Q, Zhang C, He X, Cui Z, Liu F, Li W: **BLOS2 negatively regulates Notch signaling during neural and hematopoietic stem and progenitor cell development.** *Elife* 2016, **5**.
80. Boissel L, Fillatre J, Moreau J: **Identification and characterization of the RLIP/RALBP1 interacting protein Xreps1 in *Xenopus laevis* early development.** *PLoS One* 2012, **7**(3):e33193.
81. Kam RK, Shi W, Chan SO, Chen Y, Xu G, Lau CB, Fung KP, Chan WY, Zhao H: **Dhrs3 protein attenuates retinoic acid signaling and is required for early embryonic patterning.** *J Biol Chem* 2013, **288**(44):31477-31487.
82. Markert CD, Meaney MP, Voelker KA, Grange RW, Dalley HW, Cann JK, Ahmed M, Bishwokarma B, Walker SJ, Yu SX *et al.*: **Functional muscle analysis of the Tcap knockout mouse.** *Hum Mol Genet* 2010, **19**(11):2268-2283.
83. Gruenwald K, Castagnola P, Besio R, Dimori M, Chen Y, Akel NS, Swain FL, Skinner RA, Eyre DR, Gaddy D *et al.*: **Sc65 is a novel endoplasmic reticulum protein that regulates bone mass homeostasis.** *J Bone Miner Res* 2014, **29**(3):666-675.
84. Lietman CD, Rajagopal A, Homan EP, Munivez E, Jiang MM, Bertin TK, Chen Y, Hicks J, Weis M, Eyre D *et al.*: **Connective tissue alterations in Fkbp10<sup>-/-</sup> mice.** *Hum Mol Genet* 2014, **23**(18):4822-4831.
85. Cameron VA, Aitken GD, Ellmers LJ, Kennedy MA, Espiner EA: **The sites of gene expression of atrial, brain, and C-type natriuretic peptides in mouse fetal development: temporal changes in embryos and placenta.** *Endocrinology* 1996, **137**(3):817-824.
86. Wang Y, Guo YR, Liu K, Yin Z, Liu R, Xia Y, Tan L, Yang P, Lee JH, Li XJ *et al.*: **KAT2A coupled with the alpha-KGDH complex acts as a histone H3 succinyltransferase.** *Nature* 2017, **552**(7684):273-277.
87. Ghosh TK, Aparicio-Sanchez JJ, Buxton S, Ketley A, Mohamed T, Rutland CS, Loughna S, Brook JD: **Acetylation of TBX5 by KAT2B and KAT2A regulates heart and limb development.** *J Mol Cell Cardiol* 2018, **114**:185-198.
88. Bernard DJ, Nussbaum RL: **X-inactivation analysis of embryonic lethality in *Ocrl wt<sup>-/-</sup>; Inpp5b<sup>-/-</sup>* mice.** *Mamm Genome* 2010, **21**(3-4):186-194.

89. Mishra L, Cai T, Levine A, Weng D, Mezey E, Mishra B, Gearhart J: **Identification of elf1, a beta-spectrin, in early mouse liver development.** *Int J Dev Biol* 1998, **42**(2):221-224.
90. Reis Neto RV, Yoshida GM, Lhorente JP, Yanez JM: **Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*).** *Mol Genet Genomics* 2019, **294**(3):563-571.
91. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**(3):559-575.
92. Demissie S, Cupples LA: **Bias due to two-stage residual-outcome regression analysis in genetic association studies.** *Genet Epidemiol* 2011, **35**(7):592-596.
93. Che R, Motsinger-Reif AA, Brown CC: **Loss of power in two-stage residual-outcome regression analysis in genetic association studies.** *Genet Epidemiol* 2012, **36**(8):890-894.
94. Resovi A, Pinessi D, Chiorino G, Taraboletti G: **Current understanding of the thrombospondin-1 interactome.** *Matrix Biol* 2014, **37**:83-91.
95. Mangan ME, Olmsted JB: **A muscle-specific variant of microtubule-associated protein 4 (MAP4) is required in myogenesis.** *Development* 1996, **122**(3):771-781.
96. Cui Z, Chen X, Lu B, Park SK, Xu T, Xie Z, Xue P, Hou J, Hang H, Yates JR *et al*: **Preliminary quantitative profile of differential protein expression between rat L6 myoblasts and myotubes by stable isotope labeling with amino acids in cell culture.** *Proteomics* 2009, **9**(5):1274-1292.
97. Zhang C, Yin C, Wang L, Zhang S, Qian Y, Ma J, Zhang Z, Xu Y, Liu S: **HSPC111 governs breast cancer growth by regulating ribosomal biogenesis.** *Mol Cancer Res* 2014, **12**(4):583-594.
98. Smith IA, Knezevic BR, Ammann JU, Rhodes DA, Aw D, Palmer DB, Mather IH, Trowsdale J: **BTN1A1, the mammary gland butyrophilin, and BTN2A2 are both inhibitors of T cell activation.** *J Immunol* 2010, **184**(7):3514-3525.
99. Inada M, Wang Y, Byrne MH, Rahman MU, Miyaura C, Lopez-Otin C, Krane SM: **Critical roles for collagenase-3 (Mmp13) in development of growth plate cartilage and in endochondral ossification.** *Proc Natl Acad Sci U S A* 2004, **101**(49):17192-17197.
100. Liu S, Bachran C, Gupta P, Miller-Randolph S, Wang H, Crown D, Zhang Y, Wein AN, Singh R, Fattah R *et al*: **Diphthamide modification on eukaryotic elongation factor 2 is needed to assure fidelity of mRNA translation and mouse development.** *Proc Natl Acad Sci U S A* 2012, **109**(34):13817-13822.
101. Bialik S, Kimchi A: **The death-associated protein kinases: structure, function, and beyond.** *Annu Rev Biochem* 2006, **75**:189-210.
102. Ding SY, Liu L, Pilch PF: **Muscular dystrophy in PTFR/cavin-1 null mice.** *JCI Insight* 2017, **2**(5):e91023.
103. Faggi F, Chiarelli N, Colombi M, Mitola S, Ronca R, Madaro L, Bouche M, Poliani PL, Vezzoli M, Longhena F *et al*: **Cavin-1 and Caveolin-1 are both required to support cell proliferation, migration and anchorage-independent cell growth in rhabdomyosarcoma.** *Lab Invest* 2015, **95**(6):585-602.

104. Gleissner CM, Pyka CL, Heydenreuter W, Gronauer TF, Atzberger C, Korotkov VS, Cheng W, Hacker SM, Vollmar AM, Braig S *et al*: **Neocarzilin A Is a Potent Inhibitor of Cancer Cell Motility Targeting VAT-1 Controlled Pathways**. *ACS Cent Sci* 2019, **5**(7):1170-1178.
105. Aguilar I, Legarra A, Cardoso F, Masuda Y, Lourenco D, Misztal I: **Frequentist p-values for large-scale single step genome-wide association, with an application to birth weight in American Angus cattle**. *Genet Sel Evol* 2019, **51**(1):28.
106. Lorenz AJ, Hamblin MT, Jannink JL: **Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley**. *PLoS One* 2010, **5**(11):e14079.
107. Peng W, Xu J, Zhang Y, Feng J, Dong C, Jiang L, Feng J, Chen B, Gong Y, Chen L *et al*: **Erratum: An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (*Cyprinus carpio*)**. *Sci Rep* 2016, **6**:30101.
108. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL: **Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits**. *PLoS Genet* 2013, **9**(5):e1003520.
109. Liu S, Vallejo RL, Palti Y, Gao G, Marancik DP, Hernandez AG, Wiens GD: **Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout**. *Front Genet* 2015, **6**:298.
110. Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A, Vitezica Z: **Manual for BLUPF90 family of programs**. Univ. Georg., Athens, USA; 2018.
111. **BLUPF90 and related programs (BGF90) [WWW Document]**, in **Proceeding of 7th World Congress on Genetics Applied to Livestock Production (Montpellier)** [<http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=28-07.pdf>]
112. Turner SD: **qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots**. *bioRxiv* 2014.
113. Song YE, Lee S, Park K, Elston RC, Yang HJ, Won S: **ONETOOL for the analysis of family-based big data**. *Bioinformatics* 2018, **34**(16):2851-2853.
114. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics* 2010, **26**(6):841-842.
115. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic Acids Res* 2009, **37**(1):1-13.
116. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nat Protoc* 2009, **4**(1):44-57.

## Tables

Table 1. Genomic sliding windows of 50 SNPs explaining at least 2% of the additive genetic variance for bodyweight gain by affecting growth, cell cycle, and cell proliferation. A color gradient on the left indicates differences in additive genetic variance explained by windows containing the representative SNP marker (green is the highest and red is the lowest). SNPs are sorted according to their chromosome positions.

Variance (%)	CHR	SNP position	Strand	Gene ID	Function	Gene annotation	Region/effect
2.17	2	74563933	-	LOC110499492	Growth proliferation cell cycle	caveolin-1	3'UTR
2.17	2	74563958	-	LOC110499492	Growth proliferation cell cycle	caveolin-1	3'UTR
2.17	2	74564091	-	LOC110499492	Growth proliferation cell cycle	caveolin-1	3'UTR
2.18	2	74564100	-	LOC110499492	Growth proliferation cell cycle	caveolin-1	3'UTR
2.22	2	74564246	-	LOC110499492	Growth proliferation cell cycle	caveolin-1	3'UTR
2.20	2	74565775	-	LOC110499492	Growth proliferation cell cycle	caveolin-1	CDS/nonsyn
2.20	2	74685800	-	LOC110502335	Growth proliferation	testin	3'UTR
2.19	2	74685894	-	LOC110502335	Growth proliferation	testin	3'UTR
2.17	2	75370941	+	LOC110499615	Growth	eukaryotic translation initiation factor 4 gamma 2	3'UTR
2.05	2	75387737	-	LOC110499615	Growth	eukaryotic translation initiation factor 4 gamma 2	5'UTR
2.09	2	75915086	+	LOC110499724	Growth	sodium-dependent neutral amino acid transporter B/OJAT2	3'UTR
2.08	2	76020999	+	LOC110502363	Growth	kinesin-like protein KIF21A	CDS/syn
2.71	2	76926159	-	LOC110499952	Growth cell cycle	G1/S-specific cyclin-D1	3'UTR
2.71	2	76935360	-	LOC110499952	Growth proliferation cell cycle	G1/S-specific cyclin-D1	CDS/syn
2.67	4	22804020	-	LOC110521616	Growth proliferation cell cycle	transcription factor AP-1	CDS/syn
5.08	4	23074540	+	LOC110521622	Growth proliferation cell cycle	protein PRRC2C	3'UTR
5.09	4	23115313	+	LOC110521624	Cell growth	myocilin	CDS/syn
5.09	4	23115457	+	LOC110521624	Cell growth	myocilin	CDS/nonsyn
5.12	4	23115513	+	LOC110521624	Cell growth	myocilin	CDS/nonsyn
5.12	4	23126838	+	LOC110521624	Cell growth	myocilin	CDS/nonsyn
5.12	4	23126883	+	LOC110521624	Cell growth	myocilin	CDS/nonsyn
5.11	4	23127016	+	LOC110521624	Cell growth	myocilin	3'UTR
5.08	4	23127090	+	LOC110521624	Cell growth	myocilin	3'UTR
2.91	9	34968537	-	LOC110532120	Cell cycle	protein RCC2 homolog	3'UTR
2.94	9	34968872	-	LOC110532120	Cell cycle	protein RCC2 homolog	3'UTR
2.20	13	33264383	-	LOC110486224	Growth proliferation	prohibitin	3'UTR
2.51	13	33264877	-	LOC110486224	Growth proliferation	prohibitin	CDS/syn
2.91	13	33266711	-	LOC110486224	Growth proliferation	prohibitin	CDS/syn
3.10	13	33266714	-	LOC110486224	Growth proliferation	prohibitin	CDS/syn
3.57	13	33709713	-	LOC110486239	Cell cycle	cyclin-dependent kinase 12	Intronic
3.60	13	33710394	-	LOC110486239	Cell cycle	cyclin-dependent kinase 12	CDS/syn
3.46	13	33710428	-	LOC110486239	Cell cycle	cyclin-dependent kinase 12	CDS/syn
3.40	13	33723093	-	LOC110486239	Cell cycle	cyclin-dependent kinase 12	CDS/syn
2.49	13	34487347	-	rbt stat 3	Cell proliferation	Stat3	3'UTR
2.47	13	34488012	-	rbt stat 3	Cell proliferation	Stat3	3'UTR
4.05	14	61190135	+	LOC110488945	Cell proliferation	prominin-1-A	3'UTR
4.05	14	61190693	+	LOC110488945	Cell proliferation	prominin-1-A	3'UTR
4.68	14	61198245	+	LOC110488947	Growth	fibroblast growth factor-binding protein 1	CDS/nonsyn
4.68	14	61274423	-	LOC110488948	Cell cycle	cyclin-A2	3'UTR
6.35	14	62297716	-	mcts1	Growth cell cycle	MCTS1, re-initiation and release factor	3'UTR
6.03	14	62306737	-	mcts1	Growth cell cycle	MCTS1, re-initiation and release factor	5'UTR
5.53	14	62441647	+	LOC110488975	Cell cycle	septin-6	CDS/syn
5.54	14	62451734	+	LOC110488975	Cell cycle	septin-6	CDS/syn
5.51	14	62565647	-	LOC110488980	Cell proliferation	tenomodulin	3'UTR
4.90	14	64142816	-	LOC110488986	Cell proliferation cell cycle	60S ribosomal protein L36a	CDS/syn
3.69	14	64145397	-	LOC110488986	Cell proliferation cell cycle	60S ribosomal protein L36a	CDS/syn

Table 2. Genomic sliding windows of 50 SNPs explaining at least 2% of the additive genetic variance for bodyweight gain and involved in lipid metabolism. A color gradient on the left indicates differences in additive genetic variance explained by windows containing the representative SNP marker (green is the highest and red is the lowest). SNPs are sorted according to their chromosome positions.

Variance (%)	CHR	SNP position	Strand	Gene ID	Function	Gene annotation	Region/effect
3.33	4	22956257	-	prdx6	Lipid metabolism	peroxiredoxin 6	3'UTR
4.47	4	22956370	-	prdx6	Lipid metabolism	peroxiredoxin 6	3'UTR
4.87	4	22957625	-	prdx6	Lipid metabolism	peroxiredoxin 6	CDS/syn
4.87	4	22973619	+	plpp6	Lipid metabolism	phospholipid phosphatase 6	5'UTR
5.08	4	23103208	+	vamp4	Lipid metabolism	vesicle associated membrane protein 4	3'UTR
5.11	4	23191213	-	pigc	Lipid metabolism	phosphatidylinositol glycan anchor biosynthesis class C	CDS/syn
5.12	4	23583365	+	LOC110521633	Lipid metabolism	disabled homolog 1-like	3'UTR
5.53	4	23612742	-	prkaa2	Lipid metabolism	protein kinase AMP-activated catalytic subunit alpha 2	3'UTR
5.14	4	23614045	-	prkaa2	Lipid metabolism	protein kinase AMP-activated catalytic subunit alpha 2	CDS/syn
5.15	4	23621492	-	prkaa2	Lipid metabolism	protein kinase AMP-activated catalytic subunit alpha 2	CDS/syn
5.12	4	23673384	+	LOC110521634	Lipid metabolism	phospholipid phosphatase 3	CDS/syn
5.14	4	23673839	+	LOC110521634	Lipid metabolism	phospholipid phosphatase 3	3'UTR
5.15	4	23674164	+	LOC110521634	Lipid metabolism	phospholipid phosphatase 3	3'UTR
5.04	4	23674244	+	LOC110521634	Lipid metabolism	phospholipid phosphatase 3	3'UTR
4.93	4	23674341	+	LOC110521634	Lipid metabolism	phospholipid phosphatase 3	3'UTR
2.28	8	81708445	-	LOC110530856	Lipid metabolism	acetyl-coenzyme A synthetase, cytoplasmic	3'UTR
2.30	8	81731169	-	LOC110530856	Lipid metabolism	acetyl-coenzyme A synthetase, cytoplasmic	CDS/syn
3.68	8	82494156	-	pecr	Lipid metabolism	peroxisomal trans-2-enoyl-CoA reductase	3'UTR
3.89	13	33596203	-	LOC110486236	Lipid metabolism	stAR-related lipid transfer protein 3	3'UTR
3.88	13	33596584	-	LOC110486236	Lipid metabolism	stAR-related lipid transfer protein 3	3'UTR
2.83	13	33884980	-	LOC110486250	Lipid metabolism	ATP-citrate synthase	CDS/syn
2.20	14	60291342	+	etfdh	Lipid metabolism	electron transfer flavoprotein dehydrogenase	CDS/nonsyn
2.99	14	60307455	+	etfdh	Lipid metabolism	electron transfer flavoprotein dehydrogenase	CDS/syn
3.98	14	60307821	+	etfdh	Lipid metabolism	electron transfer flavoprotein dehydrogenase	3'UTR
4.05	14	60307829	+	etfdh	Lipid metabolism	electron transfer flavoprotein dehydrogenase	3'UTR
4.05	14	60310085	-	ppid	Lipid metabolism	peptidylprolyl isomerase D	CDS/nonsyn
5.52	14	64135868	+	gla	Lipid metabolism	galactosidase alpha	CDS/nonsyn
2.17	18	21016416	-	LOC110495960	Lipid metabolism	5'-AMP-activated protein kinase subunit gamma-1	CDS/syn
2.18	18	28251973	+	olah	Lipid metabolism	oleoyl-ACP hydrolase	3'UTR

Table 3. Genomic sliding windows of 50 SNPs explaining at least 2% of the additive genetic variance for bodyweight gain and involved in proteolytic activities. A color gradient on the left indicates differences in additive genetic variance explained by windows containing the representative SNP marker (green is the highest and red is the lowest). SNPs are sorted according to their chromosome positions.

Variance (%)	CHR	SNP position	Strand	Gene ID	Function	Gene annotation	Region/effect
2.10	2	74081579	+	LOC110499430	Proteolysis	high choriolytic enzyme 2	CDS/nonsyn
2.17	2	74723434	+	LOC110499523	Proteolysis	carboxypeptidase A1	CDS/syn
5.21	4	23313642	+	oma1	Proteolysis	OMA1 zinc metallopeptidase	3'UTR
3.74	8	82356552	+	LOC110531054	Lysosome   phagosome	V-type proton ATPase subunit H	CDS/syn
3.82	8	82378763	+	LOC110531054	Lysosome   phagosome	V-type proton ATPase subunit H	3'UTR
3.89	13	33511088	-	LOC110486231	Phagosome	integrin beta-3	3'UTR
2.58	13	34233553	-	LOC110486260	Phagosome	ras-related protein Rab-5C	5'UTR
5.99	14	62343545	+	lamp2	Lysosome   phagosome	lysosomal associated membrane protein 2	CDS/nonsyn
5.98	14	62344131	+	lamp2	Lysosome   phagosome	lysosomal associated membrane protein 2	CDS/nonsyn
5.73	14	62346342	+	lamp2	Lysosome   phagosome	lysosomal associated membrane protein 2	Intronic
5.73	14	62346648	+	lamp2	Lysosome   phagosome	lysosomal associated membrane protein 2	Intronic
5.71	14	62347227	+	lamp2	Lysosome   phagosome	lysosomal associated membrane protein 2	Intronic
5.52	14	64135868	+	gla	Lysosome	galactosidase alpha	CDS/nonsyn
2.22	18	20850725	-	LOC110495951	Proteolysis	carboxypeptidase B2	CDS/syn
2.20	18	20850779	-	LOC110495951	Proteolysis	carboxypeptidase B2	CDS/syn
2.02	18	27653954	+	LOC110496062	Proteolysis	plectin	3'UTR
2.12	18	27654825	+	LOC110496062	Proteolysis	plectin	3'UTR
2.07	18	28883011	-	LOC110496097	Proteolysis	trypsin-3	CDS/syn
2.10	18	29044410	-	neu1	Lysosome	neuraminidase 1	CDS/syn

Table 4. Genomic sliding windows of 50 SNPs explaining at least 2% of the additive genetic variance in bodyweight gain and involved in the development and chromatin modification.

A color gradient on the left indicates differences in additive genetic variance explained by windows containing the representative SNP marker (green is the highest and red is the lowest). SNPs are sorted according to their chromosome positions.

Variance (%)	CHR	SNP position	Strand	Gene ID	Function	Gene annotation	Region/effect
5.11	4	23191213	-	plgc	Development	phosphatidylinositol glycan anchor biosynthesis class C	CDS/syn
5.11	4	23200735	-	LOC110521628	Development	SUN domain-containing ossification factor	CDS/non-syn
5.13	4	23209006	-	LOC110521628	Development	SUN domain-containing ossification factor	CDS/non-syn
5.23	4	23279962	+	LOC110521629	Chromatin modification	transmembrane emp24 domain-containing protein 5	3'UTR
5.21	4	23280369	+	LOC110521629	Chromatin modification	transmembrane emp24 domain-containing protein 5	3'UTR
5.21	4	23292496	+	LOC110521630	Chromatin modification	histone H2A deubiquitinase MYSM1 like	3'UTR
4.86	4	23788993	-	LOC110521636	Development	biogenesis of lysosome-related organelles complex 1 subunit 2	3'UTR
4.32	4	23789059	-	LOC110521636	Development	biogenesis of lysosome-related organelles complex 1 subunit 2	3'UTR
3.70	4	23789076	-	LOC110521636	Development	biogenesis of lysosome-related organelles complex 1 subunit 2	3'UTR
3.18	4	23789224	-	LOC110521636	Development	biogenesis of lysosome-related organelles complex 1 subunit 2	3'UTR
2.78	4	23789293	-	LOC110521636	Development	biogenesis of lysosome-related organelles complex 1 subunit 2	3'UTR
2.25	4	23789299	-	LOC110521636	Development	biogenesis of lysosome-related organelles complex 1 subunit 2	3'UTR
2.31	8	81744839	+	LOC110530857	Development	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	CDS/non-syn
2.31	8	81758124	+	LOC110530857	Development	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	3'UTR
3.34	8	82154725	+	ralbp1	Development	ralA binding protein 1	5'UTR
3.33	8	82162987	+	ralbp1	Development	ralA binding protein 1	CDS/syn
2.93	9	35051335	+	LOC110532123	Development	short chain dehydrogenase/reductase 3	5'UTR
2.93	9	35058654	+	LOC110532123	Development	short chain dehydrogenase/reductase 3	CDS/syn
3.78	13	33503527	-	LOC110486230	Chromatin modification	methyltransferase-like protein 2 A	CDS/non-syn
3.90	13	33571191	-	LOC110486234	Development	teletinonin	3'UTR
3.89	13	33572642	-	LOC110486234	Development	teletinonin	3'UTR
3.87	13	33572680	-	LOC110486234	Development	teletinonin	3'UTR
3.89	13	33573086	-	LOC110486234	Development	teletinonin	CDS/non-syn
3.29	13	33854555	-	LOC110486245	Development	synaptonemal complex protein SC65	3'UTR
3.16	13	33860250	-	LOC110486245	Development	synaptonemal complex protein SC65	CDS/syn
3.16	13	33861791	+	LOC110486247	Development	peptidyl prolyl isomerase FKBP10	CDS/syn
3.21	13	33863127	+	LOC110486247	Development	peptidyl prolyl isomerase FKBP10	CDS/syn
2.93	13	33864468	+	LOC110486247	Development	peptidyl prolyl isomerase FKBP10	CDS/syn
2.93	13	33866204	+	LOC110486247	Development	peptidyl prolyl isomerase FKBP10	CDS/syn
2.87	13	33914921	-	LOC110486251	Development	2',3' cyclic nucleotide 3' phosphodiesterase	3'UTR
2.86	13	33914958	-	LOC110486251	Development	2',3' cyclic nucleotide 3' phosphodiesterase	3'UTR
2.86	13	33915493	-	LOC110486251	Development	2',3' cyclic nucleotide 3' phosphodiesterase	3'UTR
2.86	13	33916293	-	LOC110486251	Development	2',3' cyclic nucleotide 3' phosphodiesterase	CDS/syn
2.86	13	33918316	-	LOC110486251	Development	2',3' cyclic nucleotide 3' phosphodiesterase	CDS/non-syn
2.87	13	33918333	-	LOC110486251	Development	2',3' cyclic nucleotide 3' phosphodiesterase	CDS/syn
2.87	13	33918394	-	LOC110486251	Development	2',3' cyclic nucleotide 3' phosphodiesterase	CDS/non-syn
2.58	13	34218617	-	LOC110486259	Development chromatin modification	histone acetyltransferase KAT2A	CDS/syn
4.06	14	60473623	+	raggef2	Development	Rap guanine nucleotide exchange factor 2	3'UTR
4.86	14	61835067	-	LOC110488957	Development	glutathione S-transferase P	3'UTR
4.86	14	61841933	-	LOC110488957	Development	glutathione S-transferase P	CDS/syn
6.37	14	62242007	-	LOC110488962	Development	inositol polyphosphate 5-phosphatase OCRL 1	CDS/non-syn
5.49	14	62558595	+	LOC110488979	Development	ETS-related transcription factor E1f 1	3'UTR
2.53	14	64208910	-	LOC110488993	Development	mediator of RNA polymerase II transcription subunit 12	CDS/syn
2.10	18	29258476	-	LOC110496110	Chromatin modification	double-strand break repair protein rad21 homolog	CDS/non-syn
2.04	18	29258567	-	LOC110496110	Chromatin modification	double-strand break repair protein rad21 homolog	CDS/syn

Table 5. A subset of SNP markers significantly associated with bodyweight gain using two family-based association analyses. A color gradient on the left indicates phenotypic variability explained by each single SNP marker (green is the highest and red is the lowest). SNPs were sorted according to their chromosome positions. Note: EMP1 is pointwise empirical p-value estimated using QFAM, whereas P\_RAO is the estimated p-value using a generalized score test.

R <sup>2</sup>	CHR	SNP	EMP1	P_RAO	position	Region	Strand	Gene ID	Gene annotation
0.04	3	AX-171639447	0.0011	2.94E-07	52195881	CDS syn	+	LOC110520091	D-3-phosphoglycerate dehydrogenase-like
0.09	4	AX-171630496	0.00005	8.13E-07	79714730	CDS nonsyn	+	LOC110522588	thrombospondin-1-like
0.01	8	AX-171630631	0.00005	6.60E-09	77245636	CDS syn	-	LOC110530779	death-associated protein kinase 3-like
0.03	9	AX-171616254	0.00015	2.66E-08	7679223	CDS syn	-	LOC110531523	butyrophilin subfamily 1 member A1-like
0.03	9	AX-171599860	0.00005	7.57E-08	44719009	CDS syn	-	dstn1	calsyntenin 1
0.02	12	AX-89917012	0.00005	8.36E-09	27438119	CDS nonsyn	-	LOC110537437	collagenase 3-like
0.03	12	AX-171621474	0.0001	3.62E-07	32632594	CDS syn	-	LOC110537518	nucleolar protein 16-like
0.01	13	AX-171609128	0.00145	1.56E-07	34514689	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.01	13	AX-171609127	0.0012	3.28E-07	34514726	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.01	13	AX-171609121	0.00115	2.64E-07	34515780	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.01	13	AX-171609117	0.00105	2.35E-07	34516041	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.01	13	AX-171609115	0.0014	3.37E-07	34516555	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.06	15	AX-171609715	0.0001	1.30E-07	17706212	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.06	15	AX-171609716	0.0001	3.67E-07	17706284	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.06	15	AX-171609717	0.0001	8.75E-08	17706417	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.05	15	AX-171609719	0.00005	1.70E-07	17706541	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.05	15	AX-171609720	0.00005	1.84E-07	17706838	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.06	15	AX-171609722	0.00005	9.37E-07	17717951	CDS nonsyn	-	LOC110489787	microtubule-associated protein 4-like
0.06	15	AX-171606782	0.00005	7.29E-08	17728093	5'UTR	N/A	LOC110489788	microtubule-associated protein 4-like
0.02	18	AX-171611008	0.00035	1.66E-06	18981359	CDS nonsyn	+	LOC110495923	basic leucine zipper and W2 domain-containing protein 1-A-like
0.06	20	AX-171644862	0.00005	2.46E-07	19516087	lncRNA	+	LOC110499159	uncharacterized LOC110499159
0.06	20	AX-171612012	0.0013	6.40E-07	20747206	Intergenic	N/A	N/A	N/A
0.02	28	AX-171616325	0.00145	1.22E-06	7285183	CDS nonsyn	-	LOC110508425	elongation factor 2-like

## Figures

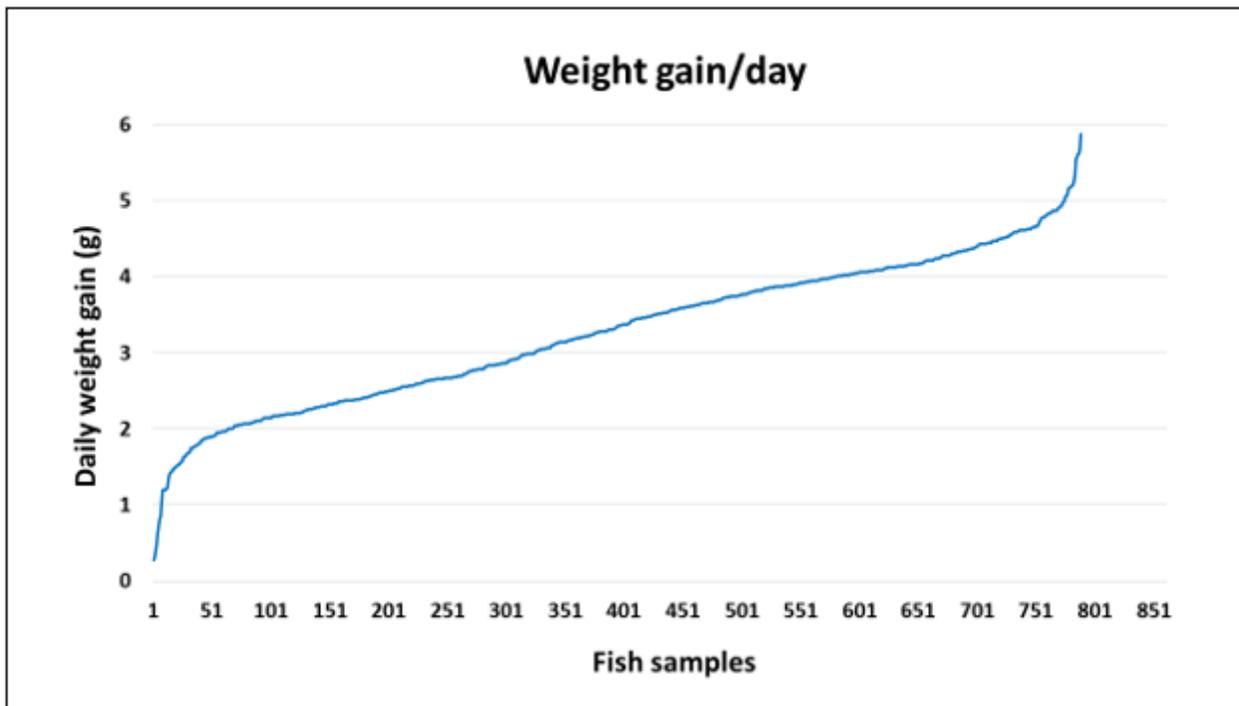
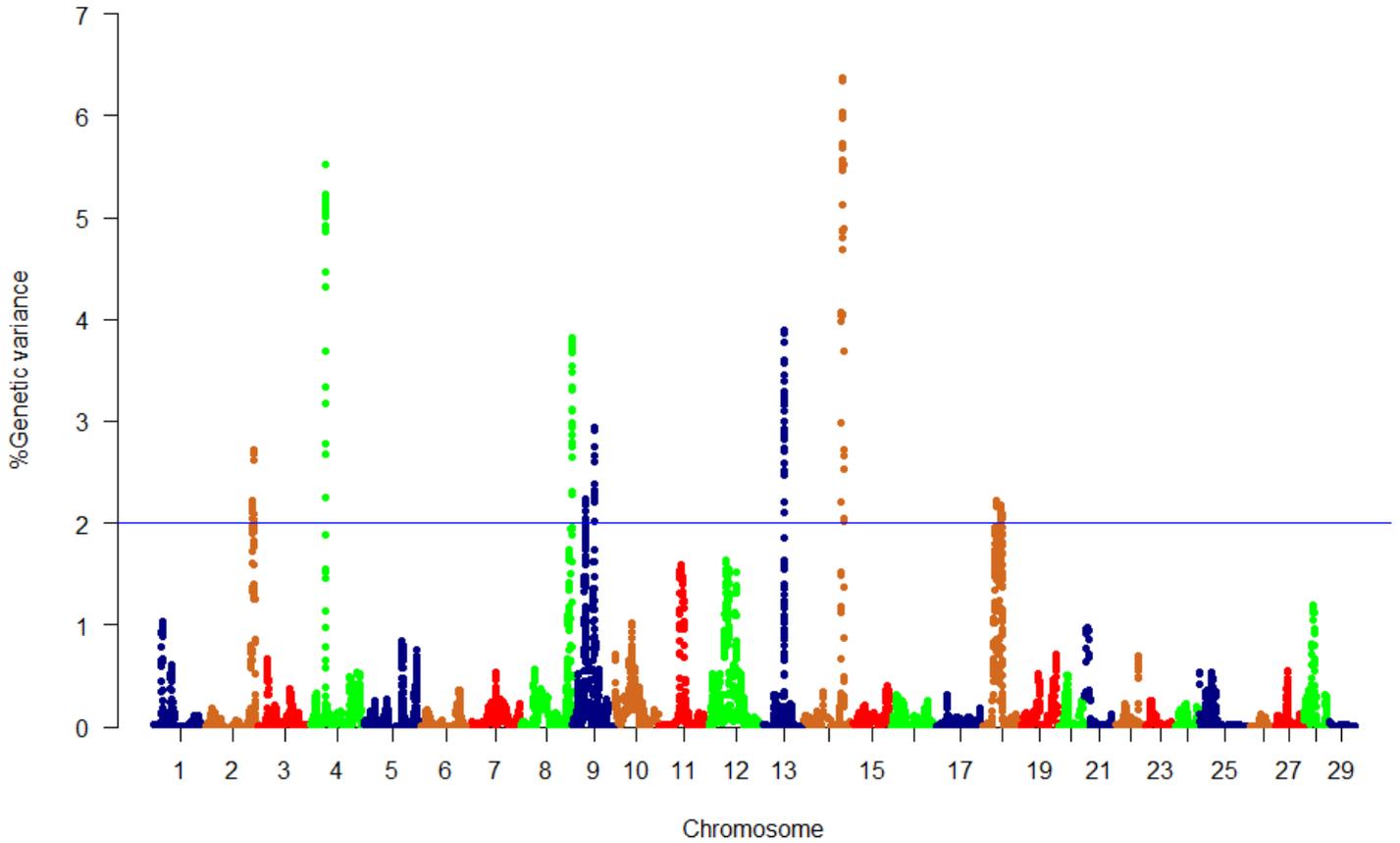


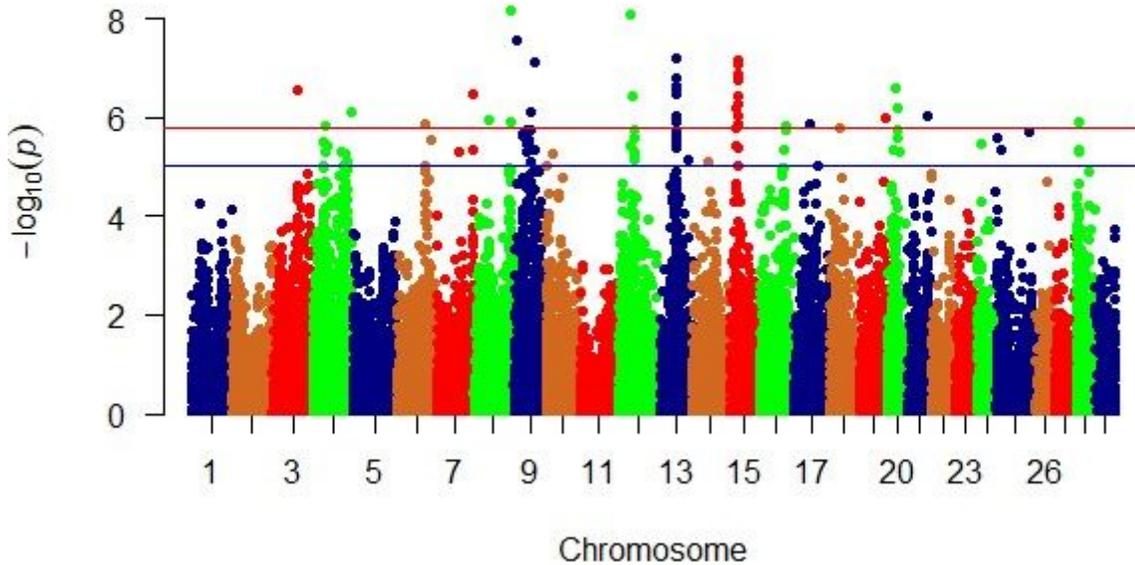
Figure 1

Variations in bodyweight gain among fish samples used in GWA analysis.



**Figure 2**

Manhattan plot displaying the association between genomic sliding windows of 50 SNPs and bodyweight gain. Chromosome 14 showed the highest peaks with genomic loci explaining up to 6.37% of the additive genetic variance. The blue line represents 2% of additive genetic variance explained by SNPs.



**Figure 3**

Manhattan plot displaying single SNP markers associated with variations in bodyweight gain using a family-based association analysis (generalized score test). Suggestive and significance threshold p-values of  $1e-05$  and  $1.70e-06$  are represented by blue and red horizontal lines represent, respectively.

## Supplementary Files

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

- [formulas.docx](#)
- [SupplementaryTables13.xlsx](#)