GENOME-WIDE ANALYSES OF GENES AFFECTING GROWTH, MUSCLE ACCRETION AND FILLET QUALITY TRAITS IN RAINBOW TROUT

by

Ali Reda Eid Ali

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy in Molecular Biosciences

Middle Tennessee State University

August 2019

Dissertation committee:

Dr. Mohamed (Moh) Salem, Chair

Dr. Anthony L. Farone

Dr. Jason Jessen

Dr. Mary Farone

Dr. Elliot Altman

I dedicate this research to my parents, my wife and kids for their unwavering love and endless support throughout this long trip. Without their great selflessness, sacrifices, and prayers, I would not have been able to fulfill and achieve this dream of mine

ABSTRACT

Growth performance is one of the most economic important traits for the aquaculture industry. In addition, fillet quality attributes are among the primary determinants of consumer acceptability. To study the genetic architecture of these traits, phenotypic characterization of whole body weight (WBW), muscle yield, fat content, shear force, moisture, protein content, and whiteness were measured in ~500 fish representing 98 families from a growth-selected line.

RNA-Seq was used to sequence the muscle transcriptome of different families exhibiting divergent phenotypes for each trait. In total, 240 and 1,280 differentially expressed (DE) protein-coding genes and long noncoding RNAs (lncRNAs), respectively, were identified in fish families exhibiting contrasting phenotypes. Expression of many DE lncRNAs (n = 229) was positively correlated with overlapping, neighboring or distantly located protein-coding genes (n = 1,030), resulting in 3,392 interactions. Three DE antisense lncRNAs were co-expressed with sense genes known to impact muscle quality traits.

A 50K SNP chip has been developed and used across four genome-wide association (GWA) studies to identify quantitative trait loci (QTL) explaining variations in fish growth and fillet quality attributes. In the first study, QTL explaining up to 28.40% of the additive genetic variance for muscle yield were identified; particularly on chromosomes 14 and 16. In the second study, 247 QTL associated with bodyweight gain were identified. Most SNPs affecting muscle yield and bodyweight gain exist in genes that act as major regulators of developmental processes. In the third study, the additive genetic variance for fillet firmness and protein content was investigated where RYR3 harbored most SNPs affecting the two

traits. In the fourth study, sixty-one common SNPs on chromosomes 19 and 29 affecting the muscle fat and moisture content were identified. SNP-harboring genes, in the common QTL, were mainly involved in lipid metabolic process and cytoskeleton remodeling. Presence of common QTL associated with multiple phenotypes suggests common mechanisms underlying those phenotypes in fish.

The present work identified DE genes and genetic markers explaining variations in growth and fillet quality phenotypes in selectively bred trout populations. Such markers could be used for marker-assisted and genomic selection in breeding programs of rainbow trout.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER V

CHAPTER IV

CHAPTER V

LIST OF APPENDICES

INTRODUCTION

Aquaculture and genomic selection

Natural fisheries are being depleted because of overfishing and overexploitation and thus cannot meet the increasing demand for seafood. Imports of seafood have increased in the USA from $5 billion to $14 billion (USDA 2005, USDC 2008) during the last two decades without a significant change in exports. The existing unbalance in importing and exporting seafood in USA led to an expanding trade deficit over two decades (1990-2008) (USDC 2008). In the meantime, seafood demand is expected to increase by about 4.4 billion pounds by 2025. Aquaculture is a growing agribusiness that has potential to enhance food security and increase economic opportunities worldwide (Burbridge et al 2001). Salmonid species have been extensively studied as aquaculture species because of their economic and nutritional values (Tsai et al 2015a). A key challenge for aquaculture industry is to sustain the increasing consumer demand for seafood (Fornshell 2002). Additional efforts are needed on the development of genetically improved strains to achieve fast/efficient production of diseases-resilient fish, with high fillet quality (Fornshell 2002). Despite the multi-environmental factors that may affect growth and other quality traits, quantitative genetic studies showed moderate to high levels of heritability for most of the economically important traits (Ali et al 2018, Gutierrez et al 2015, Wringe et al 2010). Thus, artificial selection is plausible in fish allowing potential improvement through selective breeding programs (Gutierrez et al 2015).

Growth performance and disease resistance are the most economic important traits for the aquaculture industry (Tsai et al 2015a). In addition, fillet quality attributes, such as fat content, color and texture, are among the primary determinants of consumer acceptability

(Rasmussen 2001). Selective breeding represents an approach to control for phenotypic traits of interest. Breeding programs in fish, including Atlantic salmon and rainbow trout, have focused on growth rate, fat content, color, and disease resistance (Ali et al 2018, Sodeland et al 2013). Selective breeding improves heritable traits through the existing genetic variation between individuals/families. Gjedrems' studies (Gjedrem 1983, Gjedrem 1992) showed that selective breeding programs can generate animals of improved body weight and resistance to infectious diseases which contributes to increased aquaculture production. Family-based selection lines for growth and disease resistance toward *Flavobacterium psychrophilum* (Fp) were established in 2002 and 2005, respectively, at the USDA National Center of Cool and Cold Water Aquaculture (NCCCWA). Five generations of selection yielded a genetic gain of approximately 10% in harvest body weight per generation (Leeds et al 2016). It was reported that selection based on harvest weight can improve growth rate (Salem et al 2012) and flesh color, and reduce production cost (Dufflocqa et al 2016). Recently, genetic selection was introduced in rainbow trout to improve flesh quality (Hu et al 2013, Kause et al 2007). A two generation program of selection based on fat was initiated in rainbow trout to produce lean and fat lines where the fat content increased by 15 to 31% in the fat line (Florence et al 2015). These lines represented unique models to study the effect of muscle fat content on the fillet quality (Florence et al 2015). Selection on fat content affects color and fillet texture (Florence et al 2015), and this approach also proved to be successful in improving the feed conversion ratio (FCR) and protein-retention efficiency (Kause et al 2016).

Traditional genetic improvement programs to determine animals with genetic merit use statistical analyses of phenotypes and pedigree information (Dang et al 2014). Only family-

specific estimated breeding values (EBVs) are used for breeding candidates in traditional breeding programs. Identifying markers that are associated with muscle growth and quality, and disease resistance can be applied to improve these traits through marker-assisted and genomic selections. Marker-assisted selection has been recently applied for infectious pancreatic necrosis virus (IPNV) resistance in Atlantic salmon (Houston et al 2008, Moen et al 2009) to improve breeding for phenotypes with large-effect QTL. Genomic selection tools increase the efficiency of genetic improvement in livestock relative to pedigree-based selective breeding methods (Taylor et al 2016) and has been demonstrated for BCWD resistance in rainbow trout (Vallejo et al 2017). The ability to use genomic selection will allow further within-family selection, and thus is anticipated to increase the accuracy of genetic predictions and selection response. However, genetic architecture of growth and fillet quality traits remains uncertain in fish. Understanding the genetic basis of the phenotypic traits in question and development of fish strains of improved genetic gain will enhance the efficiency of breeding programs, and influence the aquaculture industry profitability and consumer satisfaction (Ali et al 2018).

Long noncoding RNAs are potential regulators of fish growth and muscle accretion

The ENCODE project showed that only 1-2% of the human genome encodes for proteins; a major portion of the transcribed genome represents noncoding RNAs that include miRNAs and lncRNAs (Al-Tobasei et al 2016). LncRNAs are at least 200 nucleotides (nt) long and lack protein-coding potential. Previous efforts to identify lncRNAs used expressed sequence tags (ESTs) and full-length cDNA of long read length that produced few transcripts of high accuracies (Ravasi et al 2006). Recently, RNA-Seq

has been extensively utilized, using different tissues and biological conditions, to identify various novel lncRNAs. Although the biological contributions of most of lncRNAs have not yet been established, lncRNAs attracted very much attention as the functionally characterized ones, through knockdown experiments, exhibited crucial roles in various biological processes including cell cycle, apoptosis, differentiation and development, diseases and immunity (Cabianca et al 2012, Kambara et al 2014, Li and Rana 2014, Mathieu et al 2014, Paneru et al 2016, Wei et al 2014, Wu et al 2014, Yang et al 2014, Zhou et al 2015, Zhu et al 2013). Genome-wide studies identified lncRNAs exhibiting differential expression during skeletal muscle differentiation (Li et al 2012, Lu et al 2013, Zhao et al 2015); some of them were experimentally validated as participants in regulation of myogenesis, such as H19, malat1, MyoD upstream ncRNA (MUNC), lncMyoD, developmental pluripotency-associated 2 (Dppa2) Upstream binding Muscle lncRNA (DUM), and Linc-MD1 (Cesana et al 2011, Dey et al 2014, Mueller et al 2015, Watts et al 2013).

LncRNA molecular sponging of miRNA (i.e. sequestration of miRNA) has recently been reported as an important mode of action of many lncRNAs in mouse and human. For example, Linc-MD1 and malat1 have been identified as miR-133 sponges; miR-133 regulates mRNA abundance of important myogenic transcription factors such as serum response factor (SRF) and myocyte enhancer factor 2C (Mef2C) (Cesana et al 2011, Han et al 2015). Additionally, lncRNAs can act in *cis-* or in *trans-* manner to regulate neighboring and/or distant genes in mouse and human myoblasts. For example, lncMyoD is located away from MyoD locus, and binds to IGF2-mRNA-binding protein 2 (IMP2) which controls genes promoting cell cycle arrest. Depletion of lncMyoD enhances the

activity of IMP2 and impairs myogenesis (Gong et al 2015). MyoD upstream ncRNA (MUNC) is 5 kb upstream of MyoD locus and regulates the MyoD activity by enhancing the 5 kb region (distal regulatory region) (Mueller et al 2015). Dppa2 Upstream Binding Muscle lncRNA (DUM) which is located near Dppa2 gene, has been reported as a regulator of myogenesis by recruiting DNA methyl-transferase (Dnmt) family members to repress neighboring genes (Wang et al 2015). Previous studies suggested involvement of miRNAs, transcription factors and other regulatory molecules in controlling muscle growth and fillet quality traits (Esau et al 2004, Huang et al 2012, Yan et al 2013). However, the role of lncRNA in regulating these traits is still not well understood. Therefore, we sought to identify the interplay between lncRNAs and protein-coding genes in families with contrasting muscle growth and fillet quality phenotypes.

SNP arrays and genome-wide association studies

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 important production traits, such as growth, in rainbow trout. For instance, a significant QTL for body weight was co-localized with another moderate-effect QTL for maturation timing in linkage group RT-27 in rainbow trout (Drew et al 2007, Haidle et al 2008, O'Malley et al 2003). In addition, QTL for body weight and condition factor were co-localized on linkage group RT-9 and RT-27 (Wringe et al 2010). Other mapping studies reported QTL associated with temperature tolerance (Somorjai et al 2003), osmoregulation capacities (Le Bras et al 2011), stress response (Vallejo et al 2009), development rate (Nichols et al 2007), natural killer cell-like activity (Zimmerman et al 2004), BCWD (Vallejo et al 2010) and whirling disease resistance

(Baerwald et al 2011). However, classical QTL mapping has a number of limitations. Linkage analysis is time consuming and depends on association within a family limiting the power to detect associations between markers and phenotypes of interest (Gutierrez et al 2015). In addition, the identified QTL encompasses several megabases which contain hundreds, if not thousands, of genes making it difficult to identify the causal gene in a QTL (Price 2006).

Genomic resources have been developed for rainbow trout, including release of the first genome assembly draft (Berthelot et al 2014) and a newly assembled genome (GenBank assembly, NCBI accession GCA_002163495, RefSeq assembly accession GCF_002163495). New sequencing technologies facilitated identification of SNPs that are widely distributed throughout the genome, and construction of high density genetic maps (Al-Tobasei et al 2017, Palti et al 2014). About 90% of the genetic variation comes from SNPs that are highly adaptable to large-scale genotyping and therefore, most suitable for genome-wide association studies (Salem et al 2012). The rainbow trout genome was successfully used for calling variants (Al-Tobasei et al 2017), and these variants have been used to build a 50K transcribed gene SNP chip suitable for association mapping (Salem et al 2018). Over the past decade, GWA analyses have been extensively used, in mammals including human, to facilitate the investigation of hundreds of thousands to millions of variants association with complex phenotypic traits and diseases (Hindorff et al 2009). Owing to the drastic reduction in cost and time required for genotyping a large number of markers, GWA studies are replacing QTL linkage mapping (Schielzeth and Husby 2014). SNP markers in linkage disequilibrium (LD) with QTL affecting the trait of interest could be identified from GWA analyses and applied in selective breeding programs (Tsai et al

2015b). In fish, a limited number of GWA analyses have been conducted in Atlantic salmon (Tsai et al 2015b), catfish (Geng et al 2016), orange-spotted grouper (Yu et al 2018), and rainbow trout (Gonzalez-Pena et al 2016, Salem et al 2018). The studied traits in fish included growth (Tsai et al 2015b, Yu et al 2018), muslce fat content (Sodeland et al 2013, Zheng et al 2016), disease resistance (Palti et al 2015), head size (Geng et al 2016), heat stress (Jin et al 2017), low oxygen tolerance (Zhong et al 2017), and muscle yield (Gonzalez-Pena et al 2016, Salem et al 2018). GWA studies showed that growth-related traits are controlled by small-effect variants in the farmed Atlantic salmon (Yoshida et al 2017). Four significant SNPs affecting genetic variance for fat content in salmon were identified on chromosomes 9 and 10, based on 5,650 genome-wide distributed SNPs (Sodeland et al 2013). A high density 250K SNP array in common carp revealed only eight genome-wide suggestive significance SNPs related to muscle fat content, whereas there were no SNPs that surpassed the genome-wide significance level (Zheng et al 2016). In rainbow trout, GWA analysis revealed QTL associated with BCWD resistance (Liu et al 2015, Palti et al 2015). In addition, a recent GWA study identified QTL with small effects on additive genetic variance for body weight, fillet yield and fillet weight using a 57K SNP array. 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, whereas two windows on chromosome chromosome 9 explained only 1.0–1.5% of genetic variance for fillet yield and fillet weight (Gonzalez-Pena et al 2016). No GWA studies have been conducted in rainbow trout to identify the genetic architecture of many important quality traits including fillet firmness, protein content, fat content, and moisture.

In the current project, we aimed to improve and utilize a recently assembled lncRNA reference of rainbow trout and investigate the interplay between lncRNAs and protein-coding genes which may explain the variation in growth and fillet quality attributes in a rainbow trout population developed by the USDA/NCCCWA selective breeding program. In addition, the project aimed to explore the genome-wide architecture and markers associated with genetic variance for the most important growth-related and muscle quality-related attributes such as bodyweight gain, muscle yield, fillet firmness, protein content, fat content, and moisture content. Candidate genetic markers could be prioritized and used for marker-assisted and genomic selection in breeding programs of rainbow trout.

OBJECTIVES

GENERAL OBJECTIVES

- To develop invaluable genomic resources suitable for future genetics and genomic studies in rainbow trout.

- To identify genetic markers, associated with growth and muscle quality attributes, that could be considered for marker-assisted and genomic selection in breeding programs of rainbow trout.

SPECIFIC OBJECTIVES

- To improve a recently assembled lncRNA reference of rainbow trout by identifying long non-coding RNAs from strand-specific RNA-Seq datasets generated from muscle and gill tissues.

- To identify the interplay between lncRNAs and protein-coding genes in fish families exhibiting contrasting muscle growth and fillet quality phenotypes.

- To develop a 50K SNP chip suitable for genome-wide association analyses.

- To identify QTL associated with genetic variance for muscle yield, fillet firmness, protein content, bodyweight gain, fat content and moisture content through GWA analyses.

CHAPTER I

INTEGRATED ANALYSIS OF LNCRNA AND MRNA EXPRESSION IN RAINBOW TROUT FAMILIES SHOWING VARIATION IN MUSCLE GROWTH AND FILLET QUALITY TRAITS

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

ABSTRACT

Muscle yield and quality traits are important for the aquaculture industry and consumers. Genetic selection for these traits is difficult because they are polygenic and result from multifactorial interactions. To study the genetic architecture of these traits, phenotypic characterization of whole body weight (WBW), muscle yield, fat content, shear force and whiteness were measured in ~500 fish representing 98 families from a growth-selected line. RNA-Seq was used to sequence the muscle transcriptome of different families exhibiting divergent phenotypes for each trait. We have identified 240 and 1,280 differentially expressed (DE) protein-coding genes and long noncoding RNAs (lncRNAs), respectively, in fish families exhibiting contrasting phenotypes. Expression of many DE lncRNAs (n = 229) was positively correlated with overlapping, neighboring or distantly located protein-coding genes (n = 1,030), resulting in 3,392 interactions. Three DE antisense lncRNAs were co-expressed with sense genes known to impact muscle quality traits. Forty-four DE lncRNAs had potential sponge functions to miRNAs that affect muscle quality traits. This study 1) defines muscle quality associated protein-coding and

noncoding genes and 2) provides insight into non-coding RNAs involvement in regulating growth and fillet quality traits in rainbow trout.

INTRODUCTION

Aquaculture is the fastest growing agribusiness with potential to improve food security and expand economic opportunities worldwide (Burbridge et al 2001). Regarding the US Trout Industry, a key challenge for this industry is expansion to meet increasing consumer demand (Fornshell 2002). Additional effort should focus on development of genetically improved strains to achieve fast/efficient production of fish with high fillet yields and quality (Fornshell 2002).

Fillet is the most nutritional and economic important part of the fish; it is high in protein and, depending on the species, is relatively low in fat (Bugeon et al 2010). Muscle yield and flesh quality greatly affect fish processing profitability (Rora et al 2001). Several variables impact salmonid muscle yield such as harvest weight and age endpoint, animal nutrition (Einen et al 1998, Einen et al 1999), sexual status (Paaver et al 2004), and genetic factors (Smith et al 1988). Fillet quality attributes, such as fat content, color and texture, are the primary determinants of consumer acceptability (Rasmussen 2001). Recently, genetic selection was introduced in rainbow trout to improve flesh quality (Hu et al 2013, Kause et al 2007). Selection on fat content affected color and fillet texture (Florence et al 2015), and this approach improved feed conversion ratio (FCR) and protein-retention efficiency (Kause et al 2016) in rainbow trout. Moreover, selection on harvest weight can improve growth rate (Salem et al 2012) and flesh color, in addition to reducing production cost (Dufflocqa et al 2016). Selective breeding improves heritable traits through the existing genetic variation between individuals/families. If flesh qualities are incorporated

in selection programs, family selection can result in progress toward enhancing these traits (Gjedrem 1997). Gjedrem's studies showed advances in body weight using selection over three generations (Gjedrem 1992). A family-based selection line for growth was established in 2002 at the USDA National Center of Cool and Cold Water Aquaculture (NCCCWA); five generations of selection yielded a genetic gain of approximately 10% in harvest body weight per generation (Leeds et al 2016). Identifying markers that are associated with muscle growth and quality will improve these traits through selective breeding. In this study, we examined variation of muscle yield and quality traits in hatch year 2010 (third-generation families) of the NCCCWA growth-selected line. In addition, we profiled transcriptome expression of fish families showing contrasting phenotypes in whole body weight (WBW), muscle yield, muscle fat content, shear force, and whiteness index.

The ENCODE project showed that only 1-2 % of the human genome encodes for proteins; a major category of the transcribed part represents noncoding RNAs that include miRNAs and lncRNAs. Expression of lncRNAs is regulated according to physiological demands of the cell, suggesting a role for lncRNAs as key regulators of gene expression. Recent studies have demonstrated that lncRNAs contribute to regulation of various cellular processes including cell cycle, apoptosis, differentiation and development, diseases and immunity (Cabianca et al 2012, Kambara et al 2014, Li and Rana 2014, Mathieu et al 2014, Paneru et al 2016, Wei et al 2014, Wu et al 2014, Yang et al 2014, Zhou et al 2015, Zhu et al 2013). Genome-wide studies identified lncRNAs exhibiting differential expression during skeletal muscle differentiation (Li et al 2012, Lu et al 2013, Zhao et al 2015). Some lncRNAs were experimentally validated as participants in regulation of myogenesis

including H19, malat1, MyoD upstream ncRNA (MUNC), lncMyoD, developmental pluripotency-associated 2 (Dppa2), Upstream binding Muscle lncRNA (DUM), and Linc-MD1 (Cesana et al 2011, Dey et al 2014, Mueller et al 2015, Watts et al 2013).

LncRNA molecular sponging, or sequestration of miRNA, has been reported recently as an important mode of action of many lncRNAs including H19 (Dey et al 2014), Linc-MD1 and malat1 (Cesana et al 2011, Han et al 2015). H19 harbors miR-675 (Dey et al 2014) and Let-7 family (Kallen et al 2013) target sites permitting participation in regulation of myogenesis. Linc-MD1 and malat1 have been reported as miR-133 sponges. miR-133 regulates mRNA abundance of important myogenic transcription factors such as serum response factor (SRF) and myocyte enhancer factor 2C (Mef2C) (Cesana et al 2011, Han et al 2015). However, sponging is not the only mechanism of action of lncRNAs in regulating skeletal muscle differentiation. LncRNAs can act in *cis-* or in *trans-* configurations to regulate neighboring or distant genes. For example, lncMyoD, located away from MyoD locus, binds to IGF2-mRNA-binding protein 2 (IMP2) that controls genes promoting cell cycle arrest. Knockdown of lncMyoD resulted in increased activity of IMP2 and impaired myogenesis (Gong et al 2015). MyoD upstream ncRNA (MUNC) is 5 kb upstream of MyoD locus and regulates the activity of the latter by enhancing the 5 kb region (distal regulatory region) (Mueller et al 2015). Dppa2 Upstream Binding Muscle lncRNA (DUM), located near Dppa2 gene, has been reported as a regulator of myogenesis by recruiting DNA methyl-transferase (Dnmt) family members to repress neighboring genes (Wang et al 2015).

Previous studies suggested involvement of miRNAs, transcription factors and other regulatory molecules in controlling muscle growth and fillet quality traits (Esau et al 2004,

Huang et al 2012, Yan et al 2013). However, role of lncRNA in regulating these traits is still not well understood. Therefore, the objective of this study was to identify the interplay between lncRNAs and protein-coding genes in families with contrasting muscle growth and fillet quality phenotypes. We identified hundreds of protein-coding genes that were co-expressed with DE lncRNAs. Moreover, we found lncRNAs acting as natural sponges for microRNAs, and searched for common miRNA target sites in co-expressed protein-coding and lncRNA genes. We identified co-expressed protein-coding and lncRNA genes harboring binding sites to *cis* regulatory elements of transcription factors involved in myogenesis. This study improves our understanding of the role of protein-coding genes and lncRNAs in 1) muscle growth, and 2) mechanisms underlying variations in phenotypes studied. Additionally, this work will help identify genetic markers for genomic selection in development of improved germplasm for aquaculture.

MATERIAL AND METHODS

Ethics statement

Fish were maintained at the USDA National Center of Cool and Cold Water Aquaculture (NCCCWA) and all experimental protocols and animal procedures were approved and carried out in accordance with the guidelines of NCCCWA Institutional Animal Care and Use Committee Protocols #053 and #076.

Tissue sampling and phenotypic data collection

Fish population and sampling were described in detail in our previous publication (Al-Tobasei et al 2017). Briefly, phenotypic data and muscle samples were collected from ~500 female fish representing 98 families (~5 fish/family) from the growth-selected line at NCCCWA (year class 2010) (Al-Tobasei et al 2017, Salem et al 2012). Full-sib families

were produced from single-sire×single-dam matings over a 6-week period. Eggs were reared in spring water, and incubation temperature was manipulated between 7 and 13°C so that all families hatched within a 3-week period. Each family was reared at ambient water temperature (~12.5°C) in a separate 200-L tank to retain pedigree information and were fed a commercial fishmeal-based diet (Zeigler Bros Inc., Gardners, PA) using a programmable robotic feeding system (Arvotec, Huutokoski, Finland). At ~5-months post-hatch, fish were given unique identification by tagging with a passive integrated transponder (Avid Identification Systems Inc., Norco, CA) in the left-side dorsal musculature, and tagged fish were combined and reared in 800-L communal tanks supplied with partially-recirculated spring water (ambient temperature ~13°C) until harvest at ~13 months post-hatch. Fish were fed a commercial fishmeal-based diet using automatic, programmable feeders (Arvotec, Huutokoski, Finland). The initial daily feeding rate in young fish was approximately 2.5% of body weight, and the daily feeding rate was gradually decreased to approximately 0.75% of body weight as fish grew. This feeding schedule is generally similar to that described previously (Hinshaw 1999). Fish were starved for 5 days prior to harvest.

Fish sampling was performed as we previously described (Al-Tobasei et al 2017). Briefly, WBW was measured in all fish belonging to 98 fish families then families were ranked descendingly based on their WBW. For muscle sampling, the 2nd or 3rd fish from each family was chosen to adjust the distribution of WBW around the median of the family. Selected fish were randomly assigned to one of five harvest groups (one fish/ family/harvest group). The five harvest groups were sampled at ~13 month old over 5 consecutive weeks (one group per each week, mean body weight = 985 g; SD = 239 g).

Muscle yield and fillet quality analyses were performed as previously described (Salem et al 2013). In brief, muscle yield was assessed as a percent of muscle weight relative to WBW. A muscle section (40×80 mm) was separated from the dorsal musculature for texture analysis. The Soxhlet solvent extractor with petroleum ether was used to analyze crude fat. Fillet texture was assessed using a five-blade, Allo-Kramer shear cell connected with a Texture Analyzer (Model TA-HDi®; Texture Technologies Corp., Scarsdale, NY), provided with a 50 kg load cell and at a crosshead speed of 127 mm/min. Texture Expert Exceed software (version 2.60; Stable Micro Systems Ltd., Surrey, U.K.) was used to record and analyze force-deformation graphs. Peak shear force (g/g sample) was recorded then families were ranked in a descending order.

In each of five consecutive weeks, approximately 100 fish (i.e., 1 fish per full-sib family per week) were anesthetized in approximately 100 mg/L of tricaine methane sulfonate (Tricaine-S, Western Chemical, Ferndale, WA) weighed, slaughtered, and eviscerated. A muscle sample was excised from the left dorsal musculature approximately midway between the head and dorsal fin and frozen in liquid nitrogen. Head-on gutted carcasses were packed in ice, transported to the West Virginia University Muscle Foods Processing Laboratory (Morgantown, WV), and stored overnight. The next day, carcasses were manually processed into trimmed, skinless fillets by a trained faculty member and weighed. Fresh fillet surface color was measured with a Chroma Meter (Minolta, Model CR-300; Minolta Camera Co., Osaka, Japan), calibrated using a standard white plate No. 21333180 (CIE Y 93.1; x 0.3161; y 0.3326). L* (lightness), a* (redness), and b* (yellowness) values were recorded at three locations above the lateral line along the long axis of the right fillet, and these values were used to calculate a fillet whiteness index

according to the following equation; Whiteness = 100 – [(100 – L)2 + a2 + b2]1/2 (Institute 1991). The left-side fillet was frozen for subsequent proximate analysis, and a 4×8 cm fillet section was cut from the right side for subsequent cooked texture analysis. Details of the proximate and cooked texture analyses were previously described (Manor et al 2015).

cDNA library construction and sequencing

For RNA-Seq analyses, 98 fish families were ranked in a descending order according to the collected phenotypic data for each trait. An average of eight different families (~5 fish each) showing opposite phenotypes for each of the five phenotypes were selected (4 high-ranked families versus 4 low-ranked families for each trait) as we previously described (Al-Tobasei et al 2017). Each family represents a full-sib family from the growth-selected line. Fillet tissue was collected from each fish and flash frozen in liquid nitrogen, and these tissues were then stored at −80°C until total RNA isolation. Total RNA was isolated from each sample using TRIzol™ (Invitrogen, Carlsbad, CA). Quantity of total RNA was assessed by Qubit then the quality and integrity were checked by gel electrophoresis and the bioanalyzer 2100 (Aglient, CA). Total RNA from 5 samples of each family was used for RNA sequencing. Equal masses of total RNA from samples of each family were pooled and used for RNA sequencing. cDNA libraries were prepared and sequenced on Illumina HiSeq (single-end, 100bp read length) using multiplexing standard protocols. Because some fish families were common between the traits, the total number of selected families for RNA-Seq was 22 families. Briefly, first-strand was synthesized with a random hexamer and SuperScript II (Life Technologies). Double stranded DNA was blunt-ended, 3'-end A-tailed, and ligated to indexed adaptors. The adaptor-ligated double-stranded cDNA was PCR-amplified for 10 cycles with the Kapa

HiFi polymerase (Kapa Biosystems, Woburn, MA). The final libraries were Qubit-quantitated (Life Technologies, Grand Island, NY), and an Agilent bioanalyzer DNA7500 DNA chip (Agilent Technologies, Wilmington, DE) was used to determine the average size. Indexed libraries were pooled in equimolar concentration before sequencing using TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA).

RNA-Seq expression analyses

Sequencing reads were trimmed using the trimmomatic to remove the adaptors followed by FastQC quality control checks. Trimmed reads used for downstream analyses had quality score of Q30 or higher. For gene expression analysis, three references were combined for mapping the reads. Combined reference consisted of mRNAs identified in the rainbow trout genome reference (Berthelot et al 2014) and the newly annotated mRNA and lncRNA transcriptome references (Al-Tobasei et al 2016, Salem et al 2015). Albeit, the lncRNA reference needed to be improved in order to determine the orientation of lncRNA relative to the overlapping protein-coding genes. For this purpose, strand-specific RNA-Seq libraries from muscle (is being submitted to NCBI) and gill (GenBank Acc#SRP035242) were used to improve the lncRNA reference assembly basically as we previously described (Al-Tobasei et al 2016). In brief, reads from muscle and gill tissues were mapped to the reference genome (Berthelot et al 2014) using TopHat. Cufflink was used to assemble the mapped reads into transcripts. Transcripts longer than 200nt, and without coding potential and similarity to other noncoding RNA classes were considered as putative lncRNAs. Any single-exon lncRNA, adjacent to a protein-coding gene within 500nt and in sense direction, was removed. All 25,516 newly identified, non-redundant lncRNA transcripts were merged with the old lncRNA reference yielding a total of ~51k

lncRNA transcripts (available at https://www.animalgenome.org/repository/pub/MTSU2017.1228/).

For quantification of expression of protein-coding genes and lncRNAs, sequencing reads from selected families were mapped to the reference using CLC genomics workbench (https://www.qiagenbioinformatics.com/). The CLC built-in, RNA-Seq analysis tool was used to generate expression tracks for transcripts. Statistical analysis using edgeR (Robinson et al 2010) was performed on the expression values (Transcripts Per Kilobase Million; TPM) produced from RNA-Seq analyses to identify the DE genes. Genes with FDR < 0.05 and fold change value ≥ 2 or ≤ -2 were considered as significant DE genes. The sequencing data is being submitted to the NCBI SRA database.

Validation of DE genes by qPCR

To verify results obtained from RNA-Seq analyses, twelve DE protein-coding genes and five DE lncRNAs were chosen for validation by qPCR. In addition, qPCR was used to validate association of eight transcripts with the reported phenotypes across 90 random samples. Primer3 (Untergasser et al 2012) was used to design primers listed in Appendix A and Supplementary Table, S1. In order to get rid of genomic DNA, RNA was treated with Optimize™ DNAase I (Fisher Bio Reagents, Hudson, NH) according to the manufacturer's protocol. Reverse transcription reaction was performed to synthesize the first strand cDNAs via a Verso cDNA Synthesis Kit (Thermo Scientific, Hudson, NH) according to the manufacturer guidelines. qPCR was carried out by CFX96™ Real Time System (Bio Rad, Hercules, CA). Each qPCR reaction contained 2.5 µL template (100 ng/ µL), 1 µL (10 µM working solution) forward and reverse primers, 5 µL SYBR Green master mix (Bio-Rad, Hercules, CA 94547), and 1.5 µL nuclease free water. A negative

control reaction, without template, was performed for each primer to make sure that RNAs were free of genomic DNA. Sample analyses were replicated 3 times. *β*- actin gene was used as a control for normalization of expression. Only primers showing efficiency between 90 and 110% were used for qPCR. The PCR conditions for all reactions were 95°C for 30 sec followed by 40 cycles. Each cycle started with 95°C for 15 sec, followed by the appropriate annealing temperature for each primer for 30 sec, and completed at 60°C for 30 sec. The delta delta Ct ($\Delta\Delta$Ct) method (Paneru et al 2016, Schmittgen and Livak 2008) was used to quantify gene expression using qPCR data.

Gene clustering and physical genomic localization

Expression values, TPM, of lncRNAs and protein-coding genes were used to generate gene clusters. Briefly, a scaling method in CLC genomics workbench was used to normalize the expression values of all transcripts. Normalized expression values of all transcripts in 22 trout families were uploaded to the Multi-experiment Viewer (MeV) program (Eisen et al 1998, Howe et al 2011, Saeed et al 2003) to cluster protein-coding genes and lncRNAs at a minimum correlation threshold (R) of 0.85. In-house Perl scripts were used to classify lncRNAs according to relative location to their neighboring protein-coding genes on the rainbow trout genome (Al-Tobasei et al 2016, Berthelot et al 2014).

Functional annotation and gene enrichment analysis

For functional annotation, Gene Ontology (GO) analysis of DE protein-coding genes was performed by Blast2GO (Lagirand-Cantaloube et al 2008, Li et al 2007) and basic local alignment search against the KEGG database through KAAS-KEGG server Ver. 1.67x (Moriya et al 2007), as we previously described (Ali et al 2014). Additionally, the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8

(Huang et al 2009a, Huang et al 2009b) was used to perform gene enrichment analysis (Fisher Exact < 0.05) for protein-coding genes that are neighboring to and/or co-expressed with DE lncRNAs. The functional annotation chart of co-expressed genes was uploaded to EnrichmentMap plugin (Merico et al 2010) within the cytoscape (Lopes et al 2010) for visualization. The EnrichmentMap organizes the gene-sets (nodes), including pathways and Gene Ontology (GO) terms, into a network "enrichment map". Overlapping gene-sets were clustered together with an FDR cutoff < 0.05 and overlap coefficient cutoff set at 0.5.

Computational prediction of miRNA and lncRNA targets

For consensus miRNA target prediction, the DE lncRNAs and 3' UTR of their co-expressed protein-coding genes were uploaded to the small RNA analysis server (sRNAtoolbox) (Rueda et al 2015). The server has a pipeline for consensus animal miRNA target prediction. The pipeline uses three prediction tools for this purpose; miRanda, PITA, and TargetSpy. In addition to these three tools, "RNA22 version 2.0" was independently used as a fourth prediction tool to generate more reliable results. We considered the miRNA target when it had been predicted by at least three tools. The minimum free energy threshold of the microRNA: target hybridization was set at -13 Kcal/mol for all the tools. For lncRNA targets, DE lncRNAs and their co-expressed protein-coding genes were provided to a locally installed LncTar program (Li et al 2015). The normalized deltaG (ndG) cutoff was set at -0.10.

Identification of putative transcription factor binding sites (TFBS)

Promoter regions of DE lncRNA and their co-expressed protein-coding genes were scanned for putative TFBS of 26 transcription factors that are known to be involved in skeletal muscle development. These transcription factors are myogenin, MyoD, NF-AT1,

c-Fos, c-Jun, JunB, FOXO4, CREB, Elk-1, E47, MAZ, MEF-2C, GATA-2, NFI/CTF, NF-Y, VDR, Smad3, Smad4, PEA3, SRF, Sp1, Sp3, YY1, p53, GR, and AR. An in-house Perl script was applied to retrieve 500 nucleotide upstream sequences of DE lncRNA and their co-expressed protein-coding genes. Extracted promoter sequences were uploaded to the ALGGEN server to find TFBS using PROMO software (Farré et al 2003, Messeguer et al 2002). Parameters used were a maximum dissimilarity rate of 5% and a RE query (expectation of finding each of the matching motifs in a random sequence) < 0.05.

RESULTS AND DISCUSSION

Phenotypic variation in population

Genes involved in controlling fish muscle growth and quality were explored by characterizing global gene expression of mRNA and lncRNA in rainbow trout families revealing variations in WBW, muscle yield, and fillet quality traits (fat content, shear force, and whiteness index). In this study, variations in muscle yield and quality traits were characterized in fish from a growth-selected line at NCCCWA breeding program (after three generations of selection). To account for effects of WBW as a variable that may contribute to our interpretation of muscle yield and quality data, we performed a multivariable regression analysis using a mixed model. The model included random family effect, fixed sex effect and harvest dates and WBW was included as a linear covariate. Muscle yield and fat content showed moderate regression coefficient (R^2) values of 0.56 and 0.50 with WBW, respectively (Figure 1, A&B). On the other hand, shear force and fillet whiteness had low coefficient values of 0.18 and 0.01, respectively (Figure 1, C&D). Previous studies indicated potential genetic association between fast growth and increased muscle yield, paler fillets, and firmer texture (Leeds et al 2012). Fillet paleness increases

with increasing fat content. Moreover, moderate to high heritability for muscle yield, muscle weight, carcass weight, fat percentage, shear force, and fillet color were previously estimated; these estimated heritabilities imply existence of substantial additive genetic variation for growth and carcass traits in the population (Gonzalez-Pena et al 2016, Leeds et al 2012).

To ensure that the gene expression association detected in muscle yield and fat content was not confounded by WBW, phenotypic values of the muscle yield and fat content were corrected for WBW, and the corrected values were used to select families with contrasting variations in muscle yield (49.5% of BW ± 1.1 vs. 44.8% of BW ± 1.7) and fat content (9.6% ± 0.9 vs. 6.3% ± 1.0). Phenotypic variation in growth and muscle quality traits in 98 families (~5 fish each) are shown in Figure 2. Mean, standard deviation and phenotypic coefficient of variation for families showing divergent phenotypes (Average 4 high- vs. 4 low-ranked families) in each trait are listed in Appendix B and Table S2. Divergent phenotypic differences were statistically significant (P < 0.01).
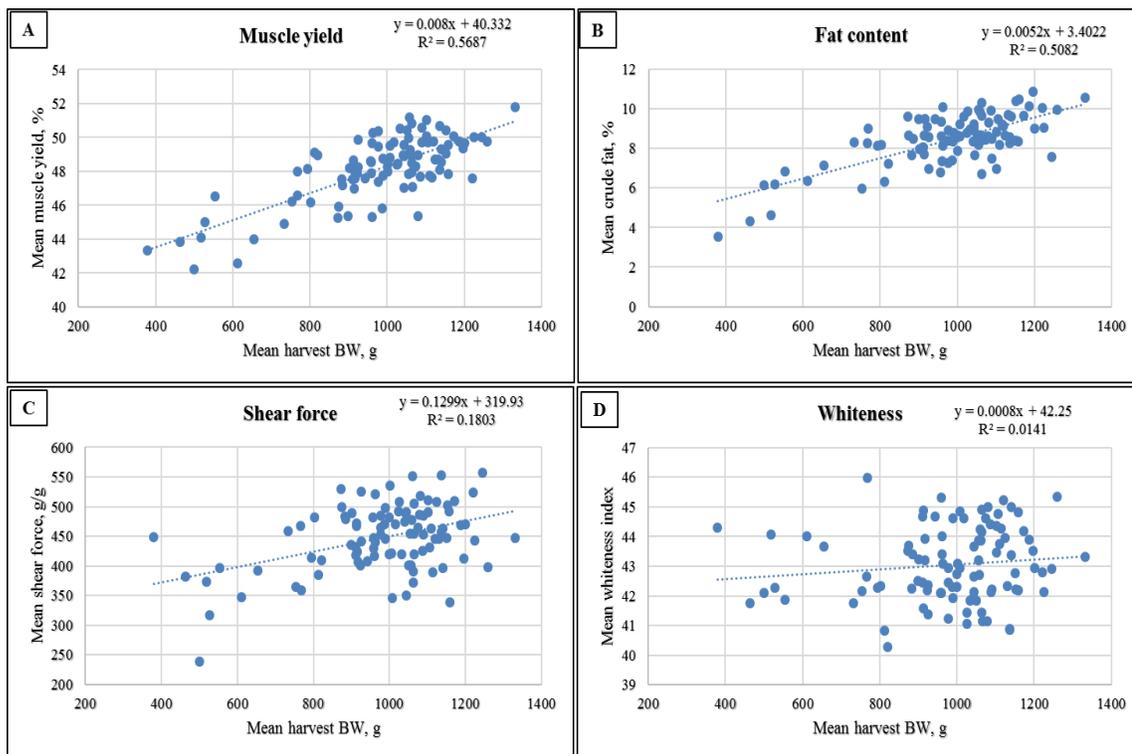
Figure 1. Effect of WBW as a variable on muscle yield and other quality traits (fat content, shear force and whiteness index) variations. WBW showed moderate regression coefficient (R^2) values of 0.56 and 0.50 with muscle yield and fat content, respectively. Fillet whiteness and shear force had low coefficient values of 0.18 and 0.01 with WBW, respectively.
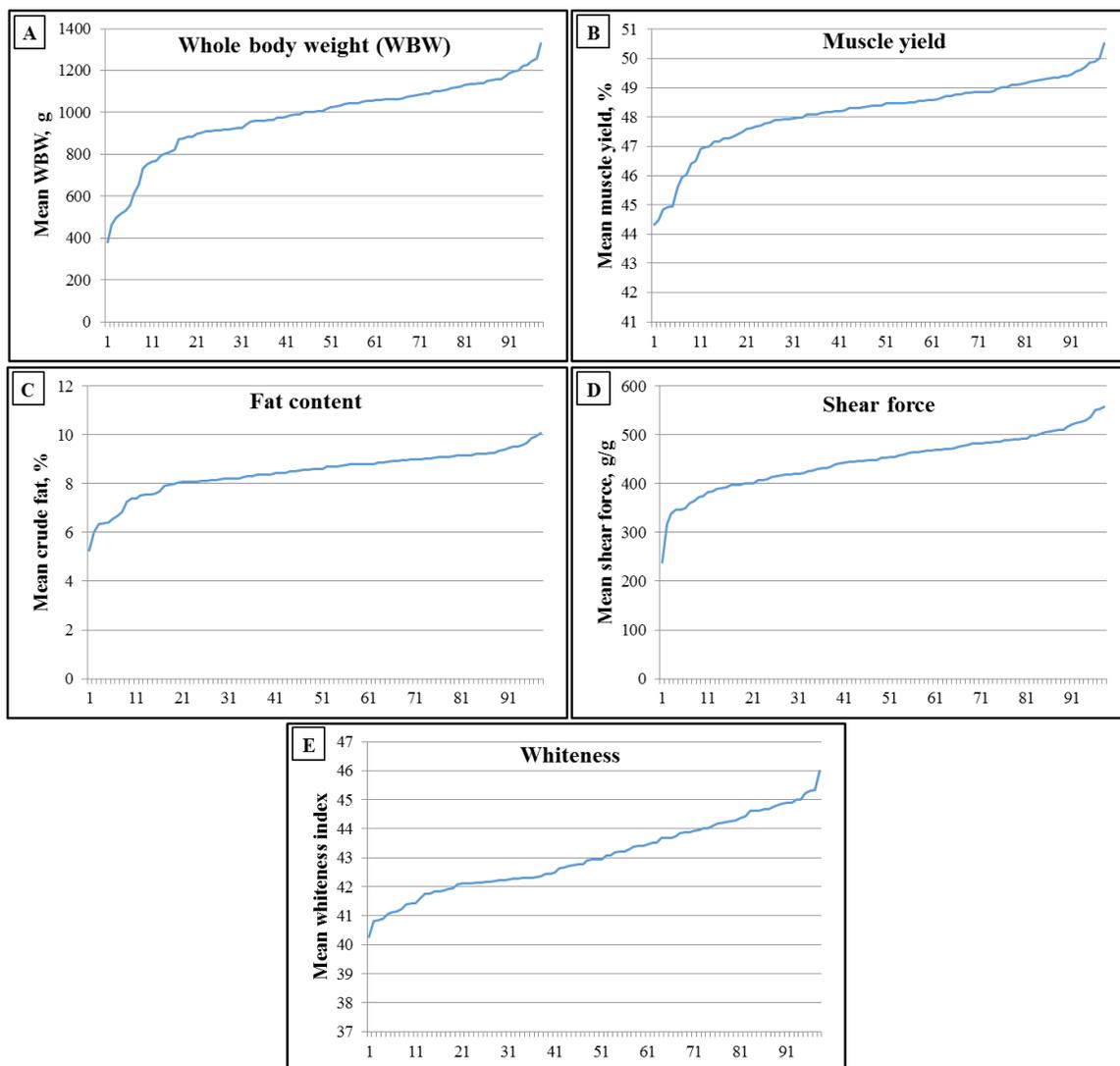
Figure 2. Phenotypic variation in growth and muscle quality traits in 98 families (~5 fish each).

Analysis of RNA-Seq data and identification of DE protein-coding and lncRNA genes

RNA sequencing of 22 fish families yielded a total of 259,634,620 raw reads (average of 11,801,573 reads/ family) with ~ 9.3 depth of coverage. Sequencing reads were trimmed to generate 250,303,394 high quality reads. A total of 219,459,206 (87.68%) trimmed

reads were mapped to the reference with ~ 7.3 depth of coverage. Quality and mapping statistics of sequencing reads were provided in Appendix C and Table (S3). To identify DE genes, high- and low-ranked families for each trait were subjected to unpaired-comparisons using the CLC Genomics workbench. A total of 240 and 1,280 non-redundant protein-coding and lncRNA genes, respectively, were DE in all studied traits with FDR < 0.05 and a minimum fold change value ≥ 2 or $\leq$ -2. Out of 1,280 DE lncRNAs, 1,061 novel lncRNA transcripts were identified in this study (Table S4). In agreement with a previous work, a higher variability in the expression of lncRNA compared with the protein-coding genes was observed (Kornienko et al 2016, Paneru et al 2016). Regression analysis showed a high Pearson correlation coefficient ($R^2 = 0.89$) between number of DE protein-coding genes and DE lncRNAs in fish families showing contrasting phenotypes. Figure 3 (A&B) shows Venn diagrams of the DE genes; WBW, muscle yield and fat content exhibited a large number of common, DE protein-coding genes (n = 41) and DE lncRNAs (n = 220). These results are consistent with previously reported pleiotropic and epistatic effects of genomic loci on fat and muscle weight controlling WBW as a composite trait (Brockmann et al 2009). Whereas, whiteness and shear force exhibited a large number of unshared DE protein-coding and lncRNA genes. Shear force and whiteness displayed fewer DE protein-coding genes and DE lncRNAs compared to the other traits. Detailed information about DE genes are shown in Figure (3C) and Supplementary Tables (S4-5). There was a significant correlation between transcript fold-change values determined by RNA-Seq and qPCR ($R^2 = 0.89$ for protein-coding genes and 0.81 for lncRNAs; Figures 4A & 4B, respectively).
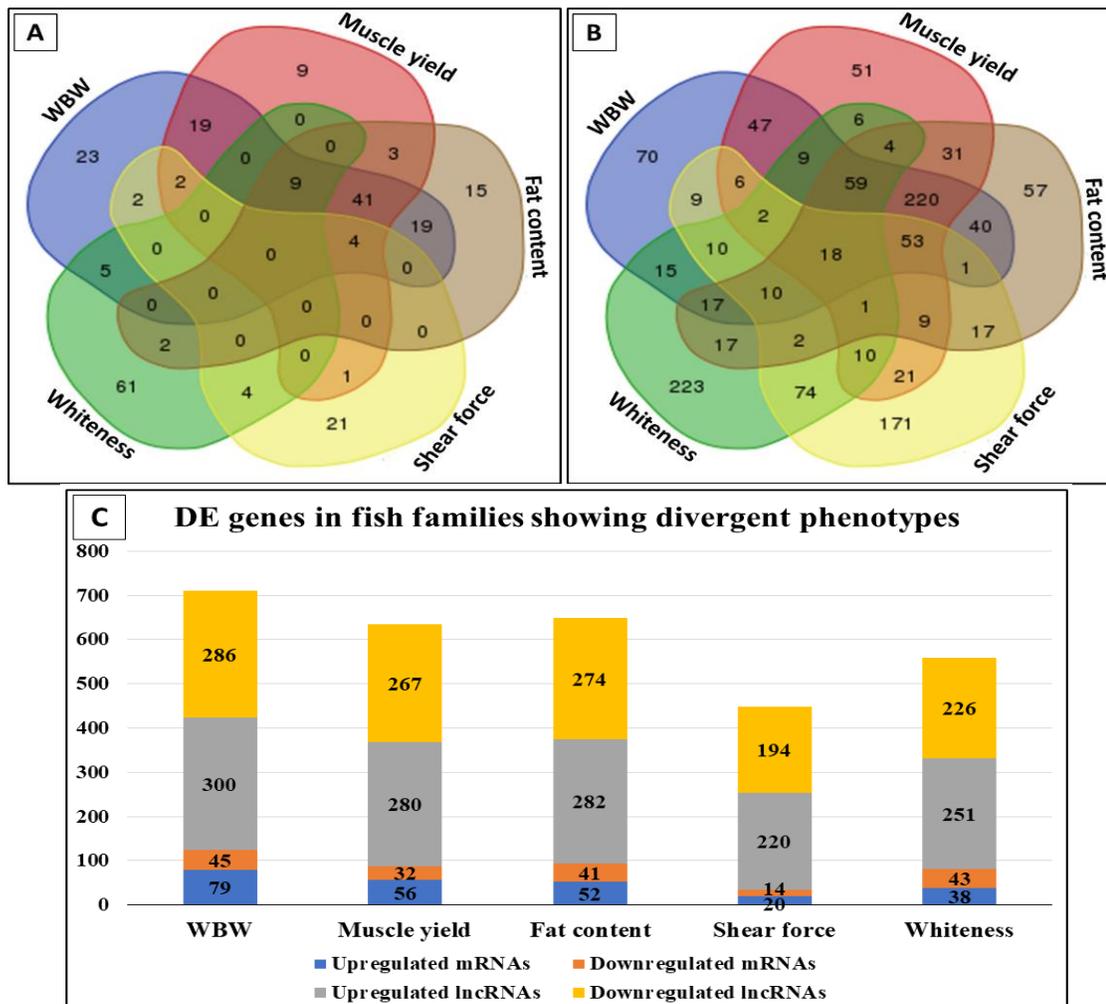
Figure 3. Venn diagram showing unique DE protein-coding genes (A) and DE lncRNAs (B) for each trait in addition to common genes between different traits. Number of DE genes for each trait is shown in (C) (FDR < 0.05 and fold change ≥ 2 or ≤ -2).

Annotation of DE protein-coding genes in families showing divergent phenotypes

GO analysis showed that DE protein-coding genes were highly represented in macromolecule metabolic process (n = 119) (Table S6). Enrichment analysis of those genes (Fisher Exact < 0.05) was provided in Table (S7). In addition, several transcripts were involved in growth-associated mechanisms such as protein metabolism (n = 90), muscle structure development (n = 52), lipid metabolism (n = 34), and oxidation-reduction processes (n = 22) (Table S6). Similar GO annotations were previously observed in genes associated with muscle growth and quality traits in rainbow trout (Danzmann et al 2016, Salem et al 2006, Salem et al 2010, Salem et al 2012).

In the current study, several genes (n = 46) that are involved in either proteolytic processes or cell growth were significantly DE in fish families showing contrasting phenotypes (Appendix D & Table S5). Out of them, 17 transcripts had functions related to protein ubiquitination, autophagy, proteolysis, or lysosome activity (Appendix D & Table S5). For example, we identified five transcripts encoding F-box only 32 (Atrogin-1), an E3 ubiquitin ligase, were upregulated in families showing low WBW (Table 1). High levels of atrogin-1 stimulate protein degradation and suppress protein synthesis (Wang et al 2010b). All Atrogin-1 transcripts showed a similar pattern of expression. By aligning the five transcripts to the trout genome, it turned out that they are partial sequences mapped to a single genomic locus indicating transcript misassembly. On the other hand, 14 out of 46 transcripts were categorized as transcription regulators (Appendix D & Table S5); of them, ankyrin repeat domain-containing protein 1 (*ANKRD1*) and gremlin-1 showed the highest fold changes (Table 1). *ANKRD1* was downregulated whereas gremlin-1 was upregulated in fish families with high WBW and fat content. Previously, gremlin-1 has

been reported as a regulator of proliferation and differentiation of myogenic progenitors in skeletal muscle (Frank et al 2006). The list of this category also includes connective tissue growth factor (CTGF) which enhances proliferation/differentiation, and proteasome-associated ECM29 homolog that enhances proteolysis; both showed DE in fish families exhibiting high whiteness index (Table 1). More details about other DE transcripts involved in cell growth/ proliferation were provided in Appendix D and Supplementary Table (S5).

Many DE genes (n = 30) were associated with fat metabolism. Of them, 17 transcripts were DE in families with variations in fat content (Appendix D & Table S5). Caveolin-3, 5-AMP-activated protein kinase subunit gamma-2 (*AAKG2*), and endophilin-b1 were downregulated, and seventeen other transcripts, including lymphocyte g0 g1 switch protein 2 (*G0S2*), hormone-sensitive lipase, perilipin-1 (*PLIN1*), lipoprotein lipase (*LIPL*), apolipoprotein A-IV, fatty acid-binding heart (*FABPH*), adipocyte plasma membrane-associated protein (*APMAP*), and diacylglycerol o-acyltransferase 2 were upregulated in families exhibiting high fat content (Table 1). Based on our KEGG pathway analysis, *CAV3* was involved in the focal adhesion pathway. Cell junction-related pathways that include focal adhesion and preserve tissue integrity were enriched along with lipid metabolism pathways in fast- and slow-growing chicken breeds suggesting a role in intramuscular fat deposition (Cui et al 2012). Previous studies showed association of *APMAP* (De clercq et al 1997, Kestin et al 1993), diacylglycerol O-acyltransferase 1 (*DGAT1*) (Yuan et al 2013) and *PLIN1* (Bickel et al 2009, Gandolfi et al 2011, Gol et al 2016, Kern et al 2004, Londos et al 2005, Zhang et al 2013, Zhang et al 2015) with adiposity and carcass traits. Anti-*APMAP* antibodies was used to decrease the backfat and

increase the lean meat percentage in pigs and other animals (De clercq et al 1997, Kestin et al 1993). It is worth mentioning that seven of these fat metabolism associated genes were DE in fish families of contrasting whiteness index suggesting a role of fat in determining fillet color (Appendix D & Table S5).

Several structural genes (n = 50) showed differential expression in families of divergent phenotypes. For example, 14 transcripts encoding myosin heavy, fast and slow chains in addition to two transcripts encoding myosin light chain-3 and -4 exhibited differential expression in fish families of contrasting phenotypes (Appendix D & Table S5). In order to investigate whether transcripts of myosin heavy chain represent different isoforms generated as a result of the alternative splicing events, we mapped them to the trout genome. The transcripts were mapped to of six unique genomic loci, some represent partial/ incomplete myosin sequences as a result of transcriptome misassembly. To emphasize the correlation between myosin heavy chain (*MYSS*) (GSONMT00032573001) and muscle yield, we quantified the abundance of *MYSS* transcripts in 90 fish using qPCR; these fish were randomly chosen from the population sample evaluated in this study. *MYSS* had an R^2 value of 0.12 (p-value = 0.001), suggesting a significant role of *MYSS* in explaining variation in muscle yield (Figure 4C). Other transcripts necessary for the muscle mass and muscle contraction were also included in the DE list. For example, seven transcripts encoding three regulatory subunits of troponin complex (Troponin I, C, and T) were DE. Two transcripts encoding troponin I, fast (*TNNI2*) were significantly upregulated in families showing high WBW and muscle yield whereas two other *TNNI2* transcripts were downregulated in these traits (Table 1).

Table 1. A subset of differentially expressed protein-coding genes (FDR < 0.05) in fish families showing contrasting phenotypes. Fold change, in families of high phenotype relative to those of low phenotypes, ≥ 2 or ≤ 2 are shown.

Feature ID	Annotation	Fold change/trait (high/low)	FDR/trait	Trait
Genes involved in protein ubiquitination and growth				
TCONS_00058870	F-box only protein 32	-3.61|-3.16|-3.39	2.5E-02|4.0E-02|4.3E-02	WBW|Mus%|Fat%
TCONS_00058871	F-box only protein 32	-3.14	4.90E-02	WBW
TCONS_00098636	F-box only protein 32	-2.71	4.60E-02	WBW
GSONMT00016768001	F-box only protein 32	-2.36	3.20E-09	WBW
GSONMT00031929001	F-box only protein 32	-2.19	2.10E-06	WBW
GSONMT00065477001	Ankyrin repeat domain-containing protein 1	-6.32|-4.99|-6.63	2.6E-06|2.4E-05|3.2E-06	WBW|Mus%|Fat%
GSONMT00041801001	Gremlin-1	4.63|4.32	3.9E-02|3.9E-02	WBW|Fat%
GSONMT00011893001	Connective tissue growth factor	-9.19	3.10E-06	Whiteness
TCONS_00152966	Proteasome-associated ECM29 homolog	6.02	2.10E-02	Whiteness
Genes involved in fat metabolism				
GSONMT00070016001	Caveolin-3	-3.29|-3.25|-3.73	2.00E-04|2.00E-04|1.04E-04	WBW|Mus%|Fat%
GSONMT00000701001	5-AMP-activated protein kinase subunit γ-2	-4.14|-3.68|-3.86	3.5E-02|3.5E-02|3.5E-02	WBW|Mus%|Fat%
GSONMT00038618001	Endophilin-b1 isoform x1	-3.86	6.40E-04	Fat%
GSONMT00017014001	Lymphocyte g0 g1 switch protein 2	4.47|3.66|3.84	3.8E-06|4.5E-04|2.2E-04	WBW|Mus%|Fat%
GSONMT00019603001	Hormone-sensitive lipase isoform x2	2.6	2.70E-02	Fat%
GSONMT00076211001	Perilipin-1	2.62|2.36|2.64	5.0E-11|3.8E-08|4.5E-11	WBW|Mus%|Fat%
GSONMT00054194001	Lipoprotein lipase	2.62|2.19|2.46	2.2E-06|1.0E-03|3.9E-05	WBW|Mus%|Fat%
GSONMT00079455001	Apolipoprotein A-IV	3.51|3.29|3.27|2.95	2.4E-12|1.2E-10|1.3E-10|2.9E-10	WBW|Mus%|Fat%|Whiteness
GSONMT00000920001	Fatty acid-binding heart	2.45|2.23|2.25	1.9E-39|8.9E-31|3.2E-31	WBW|Mus%|Fat%
GSONMT00080511001	Adipocyte plasma membrane-associated protein	2.69|2.23|2.58|2.03	9.2E-09|5.9E-05|1.0E-07|2.4E-05	WBW|Mus%|Fat%|Whiteness
GSONMT00075321001	Diacylglycerol o-acyltransferase 2	2.85|2.38|2.73	2.3E-03|4.3E-02|5.3E-03	WBW|Mus%|Fat%
Structural genes				
GSONMT00032573001	Myosin heavy chain	43.33|43.40	3.1E-02|3.1E-02	WBW|Mus%
GSONMT00065900001	Troponin fast skeletal muscle	99.04|72|49.87|2.99	2.6E-10|4.0E-07|1.3E-04|2.8E-02	WBW|Mus%|Fat%|Shear force
TCONS_00057247	Troponin fast skeletal muscle	12.82|13.64	3.0E-07|5.9E-08	WBW|Mus%
GSONMT00023675001	Troponin fast skeletal muscle	-2.38|-2.17|-2.1	6.3E-101|3.8E-85|2.4E-82	WBW|Mus%|Fat%
GSONMT00065895001	Troponin fast skeletal muscle	-2.75|-2.71|-2.27|-2.25	0.0E+00|0.0E+00|2.2E-236|5.9E-192	WBW|Mus%|Fat%|Shear force
Genes involved in calcium metabolism				
GSONMT00012525001	Stanniocalcin	23.75	1.40E-06	Shear force
TCONS_00012355	Stanniocalcin	12.82	2.20E-05	Shear force
GSONMT00027968001	Stanniocalcin	11.08	1.84E-08	Shear force
GSONMT00063580001	Parvalbumin	-4.2|-4.89|-3.97|-6.02	1.8E-104|5.1E-119|3.4E-99|3.1E-136	WBW|Mus%|Fat%|Shear force
Genes involved in oxidative stress				
GSONMT00020998001	Thioredoxin	3.25|2.66|3.05	2.0E-15|2.7E-09|3.4E-13	WBW|Mus%|Fat%
GSONMT00070684001	Glutathione peroxidase 1	2.62|2.62|2.81	3.7E-24|5.5E-24|1.4E-28	WBW|Mus%|Fat%

Troponin isoforms have an impact on the muscle fiber characteristics and could be used to improve quality traits in selection programs (Choi and Kim 2009). In order to test if troponin could be used as a potential biomarker for muscle quality traits in rainbow trout, we quantified the highly upregulated transcript of *TNNI2* (GSONMT00065900001) across 90 random fish samples. *TNNI2* (GSONMT00065900001) showed significant association with WBW ($R^2 = 0.10$; p-value = 0.008), muscle yield ($R^2 = 0.06$; p-value = 0.02), and fat content ($R^2 = 0.15$; p-value = 0.0003). Skeletal muscle *TNNI* was suggested to be used as a biomarker to identify fat adulteration (Park et al 2014). More details about other troponin subunits were provided in Appendix D and Supplementary Table (S5). Further, 27 other structural transcripts were DE; 10 of them were upregulated in high WBW and muscle yield fish families. These transcripts included type II keratin E3, nebulin, PDZ and LIM domain 5, tropomyosin alpha-3 chain, slow myotomal muscle tropomyosin, and type I cytoskeletal-13 (Appendix D & Table S5).

Interestingly, three transcripts of stanniocalcin (*STC*; GSONMT00027968001, GSONMT00012525001 & TCONS_00012355) were highly overexpressed in families of high shear force (Table 1). The three transcripts were mapped to a single genomic locus in the trout genome. *STC* is the main regulatory hormone of Ca^{+2} homeostasis in fish (Verma and Alim 2014). Calcium is essential in regulating post-mortem muscle tenderization, at least partially, through activating the Ca^{2+}-dependent cysteine proteases (Calpains) (Lian et al 2013). Similar to *MYSS* (GSONMT00032573001) and *TNNI2* (GSONMT00065900001), qPCR regression analysis showed that *STC* had R^2 value of 0.05 (p-value = 0.03) indicating a potential role of *STC* in explaining shear force variation (Figure 4D). However, Ca^{+2} analyses showed insignificant difference (p-value = 0.09) in

muscle tissue of fish from the 4th generation (an average of 60.83 umol/g dry versus 48.71 umol/g dry in families of high and low shear force, respectively, data not shown). Here we reported 19 other DE protein-coding genes that bind to Ca^{+2} ions such as parvalbumin (*PV*) (Appendix D & Table S5). Opposite to *STC*, *PV* (GSONMT00063580001) was the highly downregulated transcript in fish of high shear force and muscle yield (Table 1). *PV* was previously suggested as a biomarker for muscle mass and tenderness (Cassar-Malek and Picard 2016, Picard et al 2010).

Thioredoxin (*THIO*) and glutathione peroxidase-1 (*GPXl*) were upregulated in families of high WBW, muscle yield and fat content (Table 1). These results agree with previous studies reported that *THIO* has antioxidant and antiapoptotic properties and induces autocrine cell growth (Yoshioka et al 2006) while *GPXl* prevents fat oxidation that deteriorates fillet flavor and color (Suryanti et al 2014). We also identified other DE transcripts involved in oxidation-reduction reactions and having oxidoreductase activity. These transcripts are retinol dehydrogenase 11 (*RDH11*), dimethylaniline monooxygenase (*FMO5*), cytochrome c oxidase subunit VIb isoform 1 (*CX6B1*), very-long-chain enoyl-CoA reductase, and NADH-cytochrome b5 reductase 3 (*NB5R3*) (Appendix D & Table S5).

Correlation between DE lncRNAs and protein-coding genes

Some lncRNA annotations are available only for human and other model species. LncRNAs are poorly conserved among species (Derrien et al 2012) and this characteristic makes it hard to directly annotate lncRNAs, and consequently, anticipate their impact on the muscle growth and quality phenotypes. To help in identifying potential annotations of lncRNA in rainbow trout, the correlation between the DE lncRNAs and protein-coding

genes were studied according to their physical locations in the genome and expression correlation. According to the physical location, out of 1,280 DE lncRNAs, there were 368 genic and 912 intergenic lncRNAs. The genic lncRNAs were further subdivided as exonic (n = 112) or intronic (n = 256). More information about this classification is provided in Table (S8). Based on these criteria, DE lncRNAs and mRNAs were classified into the following two categories.
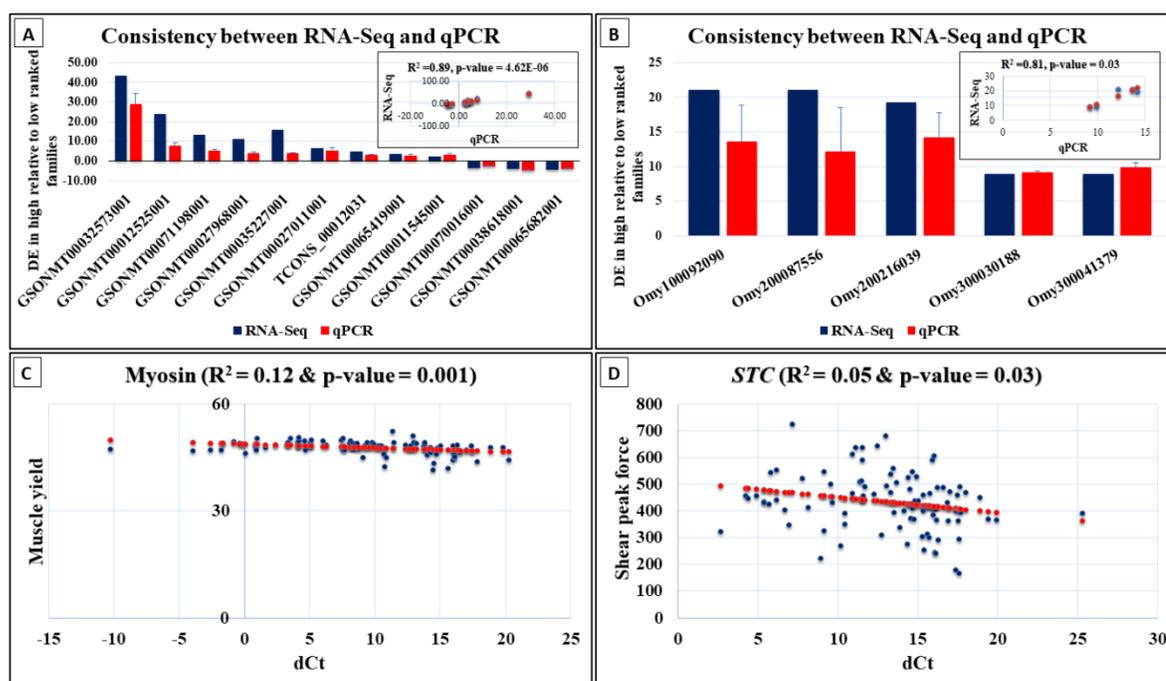
Figure 4. Consistency between RNA-Seq and qPCR measurements of DE protein-coding genes (A) and DE lncRNAs (B) (R^2 = 0.89 & 0.81, respectively). Correlations between expression pattern of *MYSS* (C) and *STC* (D) with muscle yield and shear force, respectively.

1- *Correlated and overlapped DE lncRNAs and protein-coding genes*

Many lncRNAs act in *cis* configuration to regulate expression of their adjacent genes (Tian et al 2010, Ørom et al 2010). To identify the potential *cis*-acting regulatory lncRNAs in this study, we first identified the lncRNAs that were overlapping with protein-coding genes. There were 368 genic DE lncRNAs either fully (n = 13) or partially overlapped (n = 355) with protein-coding genes in sense or antisense orientation (Table S8). Second, to identify the probable relationships between the DE lncRNAs and their overlapping protein-coding genes, we compared their expression pattern across the 22 different families considered in this study. Normalized expression values (TPM) were used to generate gene clusters between DE lncRNA and protein-coding genes with a correlation coefficient value of R ≥ 0.85 (Table S9). Six DE lncRNAs were correlated in expression with six overlapping protein-coding genes (Table 2 & Figure 5). Orientation of each lncRNA relative to its overlapping protein-coding locus was confirmed by strand-specific PCR. Association between two out of six pairs of DE lncRNA-mRNA and the phenotypes was validated by qPCR (Table 3).
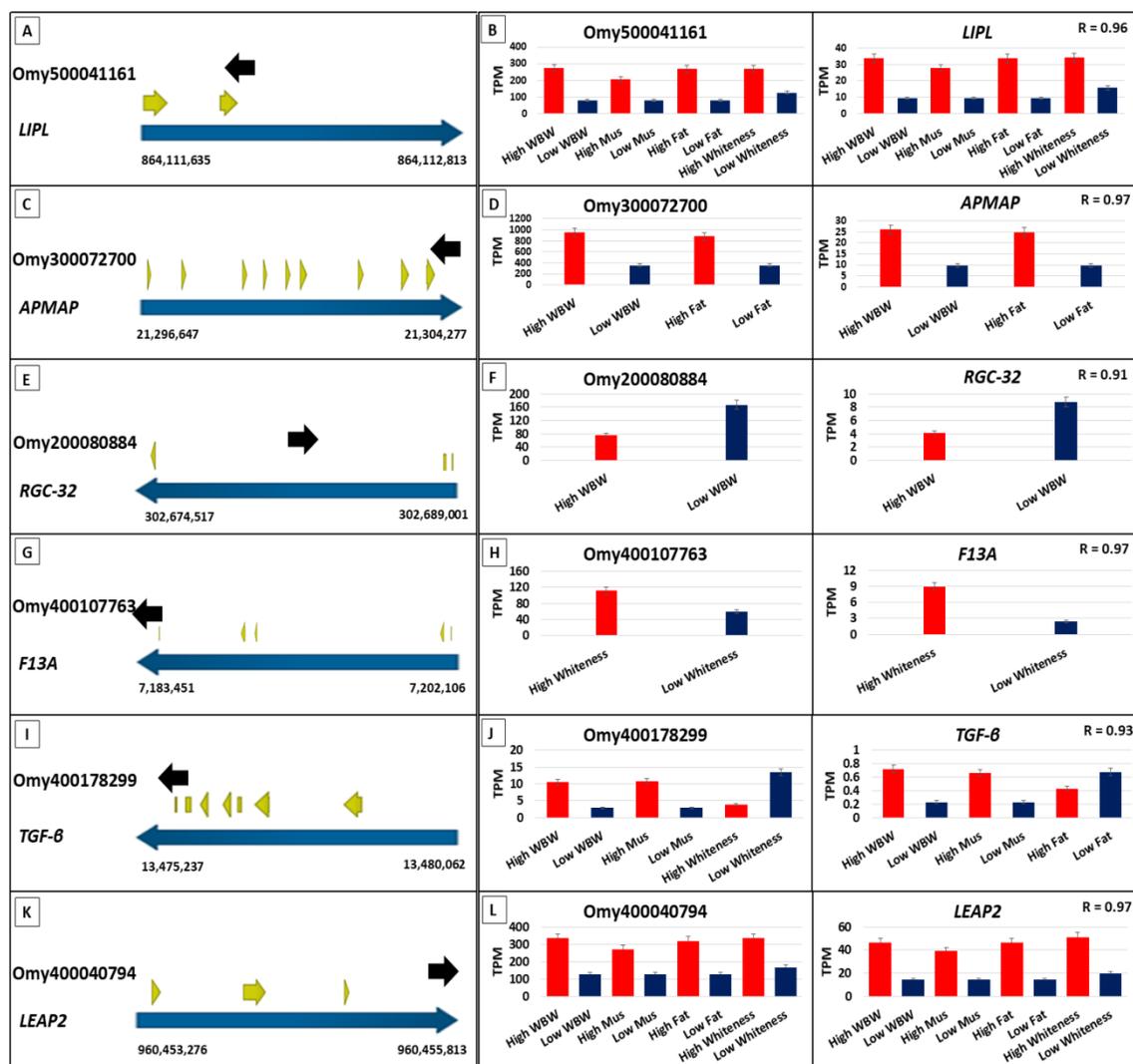
Figure 5. (A-L) Orientation of lncRNAs relative to the overlapping protein-coding loci (on the left) and comparison of the expression patterns (TPM) of DE lncRNAs with the overlapping DE protein-coding genes across families showing significant phenotypic variations (on the right).

Interestingly, in five (out of six) lncRNA-mRNA pairs, lncRNA and its counterpart protein-coding gene were DE in association with specific traits (FDR < 0.05). In the first pair, lncRNA Omy500041161 was partially overlapped with the second exon of the *LIPL* gene (GSONMT00039165001) in an antisense orientation, and their expressions were positively correlated (R = 0.96); both transcripts were upregulated in families exhibiting high WBW, muscle yield, fat content, and whiteness (Figure 5, A&B). qPCR analysis revealed a significant association between the first pair (Omy500041161/ *LIPL*) and the fat phenotype across 90 random fish samples (Table 3). LncRNA target prediction analysis showed that lncRNA Omy500041161 targets the overlapping *LIPL* gene with high-confidence cutoff values (free energy [dG] of -396.84, and normalized binding free energy [ndG] = -198.42). Furthermore, lncRNA Omy500041161 and its overlapping gene (*LIPL*) shared transcription factor binding sites for androgen receptor (*AR*) and vitamin D receptor (*VDR*). The *AR* and *VDR* contribute to skeletal muscle development (Zhao et al 2008) and function (Girgis et al 2014). In agreement with our work, *LIPL* was upregulated in fast-versus slow-growing chickens (Claire D'Andre et al 2013). Hydrolysis of circulating triglycerides and very low-density lipoproteins by *LIPL* produce free fatty acids that could be stored as neutral lipids in adipose tissue or used as an energy source by skeletal muscle (Albalat et al 2007). In the second pair, lncRNA Omy300072700 was partially overlapped with the last 3′ exon of the *APMAP* gene (GSONMT00080511001) in an antisense orientation and exhibited positively correlated expressions (R = 0.97). These two transcripts were upregulated in families with high WBW and fat content (Figure 5, C&D). Interestingly, the antisense lncRNA (Omy300072700) was predicted to compete with its co-expressed/overlapping target *APMAP* (GSONMT00080511001) in binding mir-26a and

mir-4185 (Table S10). In the third pair, lncRNA Omy200080884 was completely overlapped in antisense direction with the first 3′ intron of the gene coding for response gene to complement 32 protein (*RGC-32;* GSONMT00034829001). The two transcripts showed positively correlated downregulation in families with high WBW (R = 0.91) (Figure 5, E&F). *RGC-32* is a downstream target of transforming growth factor-beta (*TGF-β*) (Zhu et al 2012). In the fourth pair, lncRNA Omy400107763 was completely overlapped in sense orientation with the first 3′ exon and part of first 3′ intron of coagulation factor XIII A chain gene (GSONMT00082197001). Both transcripts showed strong positive expression correlation (R = 0.97). The transcripts were upregulated in families showing high whiteness index (Figure 5, G&H). In the fifth pair, lncRNA Omy400040794 was partially overlapped with the 3′ UTR of the gene coding for liver-expressed antimicrobial peptide 2 (*LEAP2*) (GSONMT00033306001) in sense direction with strong positive expression correlation (R = 0.97) (Figure 5K). These transcripts were upregulated in high versus low families of WBW, muscle yield, fat content, and whiteness (Figure 5L). It has been also reported that the use of antimicrobial compounds increases the shelf life and quality of the fillet (Lucera et al 2012). Further studies are needed to explore the roles of these lncRNAs in regulating their protein-coding counterparts.

Also, DE lncRNAs correlated with overlapping non-DE protein-coding genes. DE lncRNA Omy400178299 was positively correlated (R = 0.93) with non-DE *TGF-β* (GSONMT00041090001) across the 22 families. Overlap was in a sense orientation of the *TGF-β* sixth and seventh exons (Figure 5I&J). The microRNAs (mir-10b and mir-181d) were predicted to target Omy400178299 and its co-expressed/overlapping *TGF-β gene* (GSONMT00041090001) (Table S10). Elevated levels of *TGFβ* and its downstream

mediators (Smad 2, 3 and 4) were correlated with high levels of miR-181d (Wang et al 2010a) and mir-10b (Ouyang et al 2014) in cancer cells. Remarkably, qPCR analysis showed a significant association between Omy400178299/ *TGF-β* pair and WBW phenotype across 90 individual fish selected randomly ($\sim R^2 = 0.09$; Table 3). In addition, DE lncRNA Omy400037611 was correlated with overlapping non-DE protein-coding gene (R = 0.85) that codes for GTPase IMAP family member 4 isoform (GSONMT00033945001).

2- *Correlated and non-overlapping DE lncRNAs and protein-coding genes*

 DE lncRNAs within this category have been subdivided into two groups:

a. *Cis*-acting DE lncRNAs

As mentioned earlier, 912 out of 1,280 DE lncRNAs were categorized as intergenic, and they did not overlap with protein-coding genes. To identify lncRNAs with a potential *cis*-regulatory effect to non-overlapping neighboring genes, we searched for DE lncRNAs with protein-coding genes on both sides within a distance of 50 kb.

The 912 DE lncRNAs had 841 protein-coding genes within 50 kb. Gene enrichment analysis of the neighboring genes (Fisher Exact < 0.05) revealed that 111 genes (13.2%) were enriched in 10 KEGG pathways. These pathways include focal adhesion, insulin-signaling, ERbB signaling, phosphatidylinositol signaling, FoxO signaling, JAK-STAT signaling, and mTOR signaling. Further, epidermal growth factor domain and insulin-like growth factor binding protein were enriched (Table S11). These signaling pathways are involved in regulating skeletal muscle growth/mass (Egerman and Glass 2014, Wackerhage and Ratkevicius 2008).

Interestingly, 11 DE lncRNAs were co-expressed with twelve neighboring protein-coding genes of a potential importance to muscle growth and quality. Transcripts encoding *STC* (GSONMT00012525001) and Omy100092090 were about 1.5 kb away from each other, and their expression was positively correlated (R = 0.94) (Table 2). LncRNA Omy100092090 and *STC* were DE in association with muscle shear force. Additionally, a transcript coding for DNA (cytosine-5)-methyltransferase 3A (*DNMT3a*) (GSONMT00061222001), located within 7.9 kb from DE lncRNA (Omy400016750), exhibited positive expression correlation (R = 0.88). These two transcripts were upregulated in families with a high whiteness index (Table 2). Previous studies showed that the DNA methylation gene, *DNMT3a,* was highly expressed in skeletal muscle and significantly associated with quality traits (Guo et al 2012, Liu et al 2012, Liu et al 2015).

A single transcript encoding *FBXO32* (Atrogin-1_ (GSONMT00031929001) and DE lncRNA Omy400042056, located within 1.3 kb, were positively correlated (R = 0.88; Table 2). *FBXO32* and Omy400042056 were downregulated in fish families of high WBW. This result agrees with a previous study in rainbow trout that indicated upregulation of *FBXO32* was associated with muscle atrophy (Cleveland and Evenhuis 2010). Remaining and neighboring co-expressed DE lncRNAs and muscle relevant protein-coding genes are shown in Table 2.

Table 2. Differentially expressed lncRNAs (FDR < 0.05) showing correlation with overlapping and neighboring protein-coding genes existing within 50 kb.

LncRNA	Protein-coding gene	Annotation	Overlap	Direction	R	Trait
Overlapping protein-coding genes & LncRNAs						
Omy500041161	GSONMT00039165001	Lipoprotein lipase	Partial/Exonic	Antisense	Positive (0.96)	WBW/Mus%/Fat%/Whiteness
Omy300072700	GSONMT00080511001	Adipocyte plasma membrane-associated	Partial/Exonic	Antisense	Positive (0.97)	Fat%/WBW
Omy200080884	GSONMT00034829001	Response gene to complement 32 protein	Complete/Intronic	Antisense	Positive (0.91)	WBW
Omy400107763	GSONMT00082197001	Coagulation factor XIII A chain	Exonic	Sense	Positive (0.97)	Whiteness
Omy400178299	GSONMT00041090001	Transforming growth factor-beta	Exonic	Sense	Positive (0.93)	WBW/Mus%/Whiteness
Omy400040794	GSONMT00033306001	Liver-expressed antimicrobial peptide 2	Exonic	Sense	Positive (0.97)	WBW/Mus%/Fat%/Whiteness
Correlating protein-coding genes & LncRNAs lie within 50Kb			**Distance (Kb)**			
Omy100092090	GSONMT00012525001	Stanniocalcin	1.456	Antisense/Intergenic	Positive (0.94)	Shear force
Omy400016750	GSONMT00061222001	DNA (cytosine-5)-methyltransferase 3A	7.931	Antisense/Intergenic	Positive (0.88)	Whiteness
Omy500090683	GSONMT00056166001	Nuclear protein localization protein 4 homolog	1.162	Antisense/Intergenic	Positive (0.85)	WBW/Fat%
Omy200231682	GSONMT00079455001	Apolipoprotein A-IV	1.786	Unknown/Intergenic	Positive (0.90)	WBW/Mus%/Fat%/Whiteness
Omy500089619	GSONMT00002133001	s-adenosylmethionine synthase isoform type-1	0.67	Antisense/Intergenic	Positive (0.97)	WBW/Mus%/Fat%
Omy100162939	GSONMT00017975001	Myosin-6 isoform x1	18.744	Unknown/Intergenic	Positive (0.96)	Shear force
Omy100162939	GSONMT00017978001	Slow myosin heavy chain 1	38.362	Unknown/Intergenic	Positive (0.90)	Shear force
Omy500072095	GSONMT00067129001	Autophagy-related protein 9A isoform x1	1.875	Antisense/Intergenic	Positive (0.85)	WBW/Fat%
Omy500086794	GSONMT00007952001	Ras-related protein Rab-1A	1.569	Antisense/Intergenic	Positive (0.96)	WBW
Omy500084299	GSONMT00066744001	Triadin- partial	19.644	Antisense/Intergenic	Positive (0.93)	WBW
Omy400001433	GSONMT00041695001	Calcium-binding and coiled-coil domain-containing protein 1	3.742	Sense/Intergenic	Positive (0.99)	WBW/Whiteness
Omy400042056	GSONMT00031929001	F-box only protein 32	1.26	Sense/Intergenic	Positive (0.88)	WBW

Table 3. Association between expression of two overlapping and co-expressed, DE lncRNA-mRNA pairs (Omy500041161-*LIPL* and Omy400178299-*TGF-β*) and muscle quality traits validated by qPCR across 90 randomly selected individual fish. * indicates a significant p-value <0.05.

Transcript ID	WBW (R^2, p-value)	Muscle% (R^2, p-value)	Fat% (R^2, p-value)	Shear force (R^2, p-value)	Whiteness (R^2, p-value)
Omy500041161	0.02 (0.24)	0.01 (0.43)	0.11 (0.01)*	0.0044 (0.59)	0.00067 (0.83)
GSONMT00039165001 (*LIPL*)	0.21 (0.000086)*	0.07 (0.03)*	0.24 (0.000019)*	0.01 (0.4)	0.02 (0.23)
Omy400178299	0.09 (0.01)*	0.04 (0.1)	0.01 (0.37)	0.03 (0.18)	0.01 (0.58)
GSONMT00041090001 (*TGF-β*)	0.08 (0.01)*	0.04 (0.08)	0.05 (0.06)	0.00021 (0.9)	0.01 (0.48)

b. *Trans*-**acting DE lncRNAs**

It has been shown that lncRNAs can work in both *cis* and *trans* configuration (Jeon and Lee 2011, Schmitz et al 2010) to regulate protein-coding genes located distantly on the same or a different chromosome. To determine their expression correlation, DE lncRNAs and protein-coding genes were clustered together according to their expression values across 22 families. Several clusters, with expression correlation (R) $^{\geq}$ 0.85, have been identified between DE lncRNAs and all protein-coding genes, including DE protein-coding genes, that are distantly distributed in the trout genome (Table S9). These clusters include correlations between single lncRNAs and several different protein-coding genes as previously reported (Carpenter et al 2013, Paneru et al 2016). Cytoscape platform was used to visualize molecular interaction networks among the whole set of co-expressed genes and the phenotypic traits. We detected that, among all the 229 co-expressed lncRNAs, lncRNA (Omy500089619) exhibited the highest negative correlation with WBW, muscle yield, and fat content phenotypes across 22 families. qPCR analysis, across 90 individuals, revealed a significant association between Omy500089619 and the WBW, muscle yield, and fat content phenotypes ($R^2 = 0.09$ (p-value = 0.02), 0.09 (p-value = 0.02), and 0.15 (p-value = 0.002), respectively). In addition, four protein-coding genes, co-expressed with Omy500089619, showed the highest negative correlations with the aforementioned three phenotypic traits (Figure 6A). These protein-coding genes are *CAV3* (GSONMT00070016001), very-long-chain enoyl-CoA reductase (GSONMT00029837001; TECR), s-adenosylmethionine synthase isoform type-1 (GSONMT00002133001; METK1), and *AAKG2* (GSONMT00000701001). The *AAKG2* acts as a metabolic master switch that turns on fatty acid oxidation by acetyl-CoA

carboxylase-2 phosphorylation and turns off fatty acid synthesis by acetyl-CoA carboxylase-1 phosphorylation (Hardie and Pan 2002). The list also includes the lncRNAs Omy500041161 and Omy300072700 that were upregulated in families showing variations in WBW, muscle yield, and fat content. These DE lncRNAs were co-expressed across 22 families with distantly located and upregulated protein-coding genes that impact fillet quality; these upregulated protein-coding genes are *APMAP, G0S2, THIO, DGAT2, FABPH, PLIN1, LEAP2,* vitamin K-dependent protein S (*PROS*), serotransferrin (*TRFE*), and LIPL* (Figure 6B).
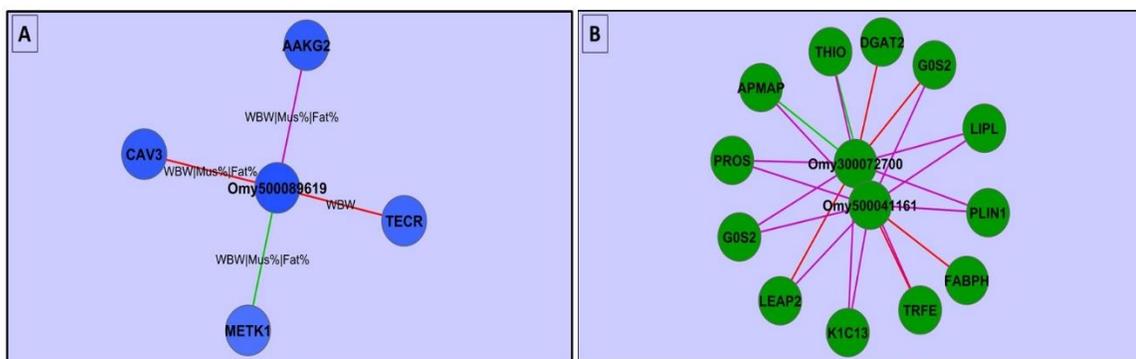
Figure 6. (A) Correlation between single DE lncRNA and co-expressed distant protein-coding genes in addition to their correlations with the phenotypes (WBW, muscle%, and fat%). The color intensity of the nodes reflects the correlations between the genes and phenotypes. The color of the edges reflects the correlation between the protein-coding genes and DE lncRNA where red denotes $0.90 > R \geq 0.85$, purple color denotes $0.95 > R \geq 0.90$, and green color means $R \geq 0.95$. (B) Correlation between two antisense DE lncRNAs and their co-expressed distant protein-coding genes that are significantly upregulated in fish families showing variations in WBW, muscle%, and fat%.

Gene enrichment analysis of the lncRNA co-expressed protein-coding genes

Co-expression analysis across 22 families, with Pearson's correlation coefficients of R ≥ 0.85, was performed to predict the probable targets of lncRNAs in *cis/trans*-regulatory relationships. A total of 3,392 positive correlations (edges) were successfully detected. Gene enrichment analysis using DAVID (Huang et al 2009a, Huang et al 2009b) was performed to determine the probable functions of co-expressed genes and infer the mechanism of gene regulation by lncRNAs (Figure 7). Functional annotation revealed 279 genes with enriched GO terms under the biological process. These genes are involved in cell adhesion, the ubiquitin-dependent protein catabolic process, development, and ATP synthesis-coupled proton transport. When categorized according to molecular function, 273 genes have enriched terms; these functions include ATP binding, catalytic activity, oxidoreductase activity, actin binding, lipid binding, and electron carrier activity. Additionally, 271 co-expressed genes (34.7%) were enriched in 20 KEGG pathways including metabolic pathways, oxidative phosphorylation, focal adhesion, tight junction, and PPAR signaling pathway. The PPAR signaling pathway is a nuclear hormone receptor containing pathway that plays a role in lipid metabolism. Genes involved in PPAR signaling pathway were associated with the intramuscular fat content (Lim et al 2015). Our results showed that most of the enriched GO terms belonged to lipid metabolism, energy production and conversion, and protein posttranslational modification and turnover. Genes with similar annotations have been previously reported to contribute to muscle growth and quality traits in rainbow trout (Danzmann et al 2016, Salem et al 2006, Salem et al 2010, Salem et al 2012). Thus, DE lncRNAs may contribute significantly to muscle growth and thereby impact muscle characteristics through their interaction with genes affecting muscle

food quality traits. Furthermore, the results support the regulatory mechanism of lncRNAs through mediation of cellular energy responses (Liu et al 2016).
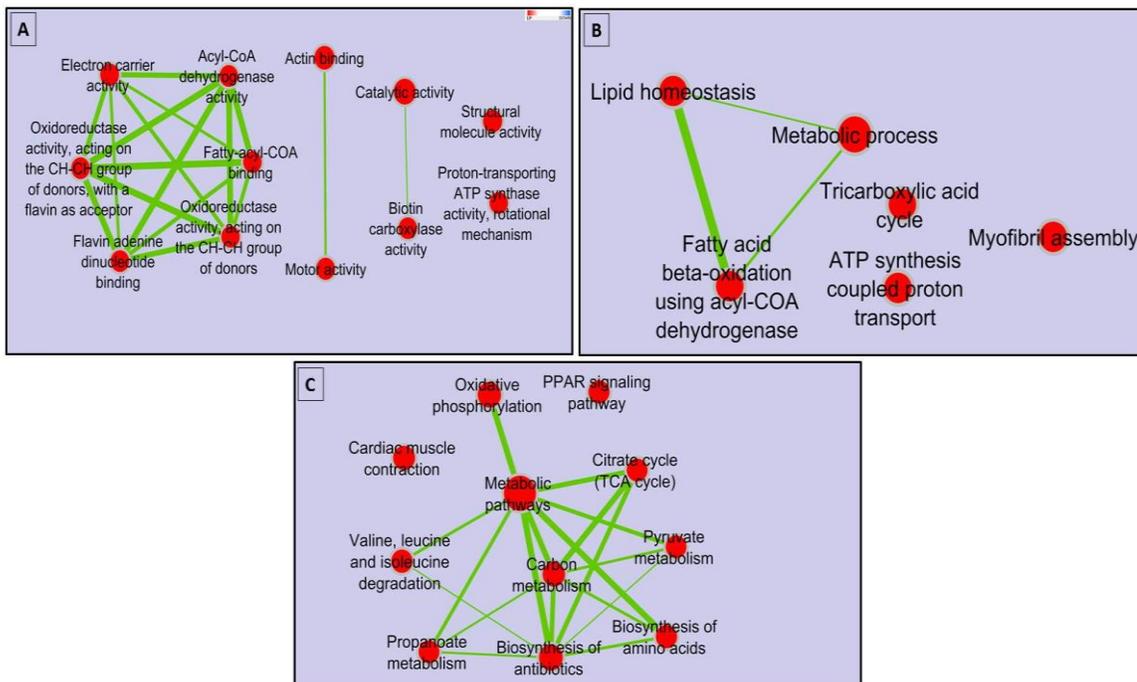
Figure 7. Gene enrichment analysis of protein-coding genes co-expressed with DE lncRNAs. Enriched gene-sets are represented as red nodes connected according to their GO/ KEGG pathway relations. Color intensity of the node represents the fold enrichment while node size represents number of genes in the gene-set. Enriched terms belonging to metabolic pathways, energy and growth-related mechanisms were predominant.

Transcription factor binding sites (TFBS) in promoter regions of DE lncRNAs and their co-expressed protein-coding genes

In this study, we scanned 500 nts upstream of 229 DE lncRNAs and their co-expressed protein-coding genes to predict TFBS in their promoter regions. Co-expressed genes,

exhibiting similar expression patterns, are expected to be controlled by similar regulatory mechanisms and are likely regulated by the same transcription factors (Allocco et al 2004, Yu et al 2003). A total of 26 binding motifs that have a role in skeletal muscle development were used for scanning the promoter region of co-expressed protein-coding and noncoding genes. A total of 209 (91.3%) DE lncRNAs and 946 (91.8%) co-expressed, protein-coding genes harboring the same TFBS have been identified. Seventy-one DE lncRNAs and 340 co-expressed protein-coding genes had putative TFBS for myogenin, while 52 DE lncRNAs and 246 co-expressed, protein-coding genes had putative TFBS for myoD gene. Myogenin and myoD control determination and terminal differentiation of skeletal muscle cells. Supplementary file Table S12 contains all pairs of co-expressed genes and their common TFBS. Previous studies focused on the pivotal role of miRNAs in the regulation of major myogenic pathways (Braun and Gautel 2011). The current results propose a potential post-transcriptional regulatory role for lncRNAs in myogenesis.

CONCLUSIONS

Muscle yield and quality traits are determinants of the aquaculture industry profitability and consumers' satisfaction. These traits result from multifactorial interactions and given that the largest part of the transcriptome is noncoding and the role of lncRNA in regulating myogenesis is increasing, an integrated analysis of mRNA and lncRNA in fish families showing divergent phenotypes for muscle yield and quality traits has been performed. DE protein-coding genes in families of contrasting phenotypes were identified. Of them, *MYSS* and *TNNI2* isoforms, and *STC* were explaining part of the phenotypic variations, suggesting them as potential markers for WBW and muscle yield, and shear force, respectively. However, the lncRNA showed higher variability in terms of expression

between divergent families. Given the fact that lncRNAs are poorly conserved, we identified networks/ hubs between DE lncRNAs and their overlapping, neighboring, or distantly located on the genome based on expression correlation analysis. For example, the overlapping Omy500041161/ *LIPL* and Omy400178299/ *TGF-β* gene pairs revealed significant association with fat content and WBW, respectively. Additionally, lncRNA (Omy500089619) exhibited significant correlations with WBW, muscle yield, and fat content. These genes are good candidates for future knockdown studies to verify their roles in controlling the phenotypic variations. This study revealed complex microRNA sponge effects for lncRNA that may contribute to fast/efficient growth rates by controlling genes belonging to protein catabolic/anabolic pathways. Further, promoter regions of DE lncRNAs and their co-expressed protein-coding genes harbored similar TFBS necessary for muscle development, suggesting common transcription initiation mechanisms. Therefore, the current study highlights the possible regulatory interactions exerted by noncoding RNAs to control expression of protein-coding genes that impact muscle quality traits, and adds additional layers of complexity that may help in understanding the molecular network of muscle development.

Author contributions

M.S., B. K. and T. D. L. conceived and designed the experiments. M.S., B. K. and T. D. L. performed the experiment. A.A., R. A. T. and M.S. analyzed the data. A.A. and M.S. wrote the paper. All authors reviewed and approved the manuscript.

REFERENCES

Al-Tobasei R, Paneru B, Salem M (2016). Genome-Wide Discovery of Long Non-Coding RNAs in Rainbow Trout. *PLoS One* **11:** e0148940.

Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B *et al* (2017). Identification of SNPs associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-Seq samples in rainbow trout. *BMC Genomics* **18:** 582.

Albalat A, Saera-Vila A, Capilla E, Gutiérrez J, Pérez-Sánchez J, Navarro I (2007). Insulin regulation of lipoprotein lipase (LPL) activity and expression in gilthead sea bream (Sparus aurata). *Comp Biochem Physiol B Biochem Mol Biol* **148:** 151-159.

Ali A, Rexroad CE, Thorgaard GH, Yao J, Salem M (2014). Characterization of the rainbow trout spleen transcriptome and identification of immune-related genes. *Front Genet* **5:** 348.

Allocco DJ, Kohane IS, Butte AJ (2004). Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics* **5:** 18.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B *et al* (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications* **5:** 3657.

Bickel PE, Tansey JT, Welte MA (2009). PAT proteins, an ancient family of lipid droplet proteins that regulate cellular lipid stores. *Biochim Biophys Acta* **1791:** 419-440.

Braun T, Gautel M (2011). Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. *Nat Rev Mol Cell Biol* **12:** 349-361.

Brockmann G, Tsaih S-W, Neuschl C, Churchill G, Li R (2009). Genetic factors contributing to obesity and body weight can act through mechanisms affecting muscle weight, fat weight, or both. *Physiol Genomics* **36:** 114–126.

Bugeon J, Lefevre F, Cardinal M, Uyanik A, Davenel A, Haffray P (2010). Flesh quality in large rainbow trout with high or low fillet yield. *Journal of Muscle Foods*. pp 702-721.

Burbridge P., Hendrick V., Roth E., H. R (2001). Social and economic policy issues relevant to marine aquaculture. *J Appl Ichthyol* **17:** 194–206.

Cabianca DS, Casa V, Gabellini D (2012). A novel molecular mechanism in human genetic disease: a DNA repeat-derived lncRNA. *RNA biology* **9:** 1211-1217.

Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL *et al* (2013). A long noncoding RNA mediates both activation and repression of immune response genes. *Science* **341:** 789-792.

Cassar-Malek I, Picard B (2016). Expression Marker-Based Strategy to Improve Beef Quality. *ScientificWorldJournal* **2016:** 2185323.

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M *et al* (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147:** 358-369.

Choi YM, Kim BC (2009). Muscle fiber characteristics, myofibrillar protein isoforms, and meat quality. *Livestock Science* **122:** 105–118.

Claire D'Andre H, Paul W, Shen X, Jia X, Zhang R, Sun L *et al* (2013). Identification and characterization of genes that control fat deposition in chickens. *J Anim Sci Biotechnol* **4:** 43.

Cleveland BM, Evenhuis JP (2010). Molecular characterization of atrogin-1/F-box protein-32 (FBXO32) and F-box protein-25 (FBXO25) in rainbow trout (Oncorhynchus mykiss): Expression across tissues in response to feed deprivation. *Comp Biochem Physiol B Biochem Mol Biol* **157:** 248-257.

Cui HX, Liu RR, Zhao GP, Zheng MQ, Chen JL, Wen J (2012). Identification of differentially expressed genes and pathways for intramuscular fat deposition in pectoralis major tissues of fast-and slow-growing chickens. *BMC Genomics* **13:** 213.

Danzmann RG, Kocmarek AL, Norman JD, Rexroad CE, Palti Y (2016). Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics* **17:** 60.

De clercq L, Mourot J, Genart C, Davidts V, Boone C, Remacle C (1997). An anti-adipocyte monoclonal antibody is cytotoxic to porcine preadipocytes in vitro and depresses the development of pig adipose tissue. *J Anim Sci* **75:** 1791-1797.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H *et al* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22:** 1775-1789.

Dey BK, Pfeifer K, Dutta A (2014). The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev* **28:** 491-501.

Dufflocqa P, Lhorenteb JP, Bangerab R, Neirac R, Newmand S, Yáñeza JM (2016). Correlated response of flesh color to selection for harvest weight in coho salmon (Oncorhynchus kisutch). *Aquaculture*.

Egerman MA, Glass DJ (2014). Signaling pathways controlling skeletal muscle mass. *Crit Rev Biochem Mol Biol* **49:** 59-68.

Einen O, Waagan B, Thomassen MS (1998). Starvation prior to slaughter in Atlantic salmon (Salmo salar): I. Effects on weight loss, bodyshape, slaughter- and fillet-yield, proximate and fatty acid composition. *Aquaculture* **166:** 85–104.

Einen O, Mørkøre T, Rørå AMB, Thomassen MS (1999). Feed ration prior to slaughter— a potential tool for managing product quality of Atlantic salmon (*Salmo salar*). *Aquaculture* **178:** 149–169.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95:** 14863-14868.

Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV *et al* (2004). MicroRNA-143 regulates adipocyte differentiation. *J Biol Chem* **279:** 52361-52365.

Farré D, Roset R, Huerta M, Adsuara JE, Roselló L, Albà MM *et al* (2003). Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res* **31:** 3651-3653.

Florence L, Mireille C, Jérôme B, Laurent L, Françoise M, Edwige Q (2015). Selection for muscle fat content and triploidy affect flesh quality in pan-size rainbow trout, Oncorhynchus mykiss *Aquaculture* **448:** 569-577.

Fornshell G (2002). Rainbow Trout — Challenges and Solutions. pp *545–557*.

Frank NY, Kho AT, Schatton T, Murphy GF, Molloy MJ, Zhan Q *et al* (2006). Regulation of myogenic progenitor proliferation in human fetal skeletal muscle by BMP4 and its antagonist Gremlin. *J Cell Biol* **175:** 99-110.

Gandolfi G, Mazzoni M, Zambonelli P, Lalatta-Costerbosa G, Tronca A, Russo V *et al* (2011). Perilipin 1 and perilipin 2 protein localization and gene expression study in skeletal muscles of European cross-breed pigs with different intramuscular fat contents. *Meat Sci* **88:** 631-637.

Girgis CM, Mokbel N, Cha KM, Houweling PJ, Abboud M, Fraser DR *et al* (2014). The vitamin D receptor (VDR) is expressed in skeletal muscle of male mice and modulates 25-hydroxyvitamin D (25OHD) uptake in myofibers. *Endocrinology* **155:** 3227-3237.

Gjedrem T (1992). Breeding plans for rainbow trout. In: Gall GAE (ed). *The Rainbow Trout: Proceedings of the First Aquaculture-sponsored Symposium held at the Institute of Aquaculture, University of Sterling, Scotland.* pp 73–83.

Gjedrem T (1997). Flesh quality improvement in fish through breeding. Aquaculture International. pp 197-206.

Gol S, Ros-Freixedes R, Zambonelli P, Tor M, Pena RN, Braglia S *et al* (2016). Relationship between perilipin genes polymorphisms and growth, carcass and meat quality traits in pigs. *J Anim Breed Genet* **133:** 24-30.

Gong C, Li Z, Ramanujan K, Clay I, Zhang Y, Lemire-Brachat S *et al* (2015). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev Cell* **34:** 181-191.

Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB *et al* (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (Oncorhynchus mykiss). *Front Genet* **7:** 203.

Guo X, Liu X, Xu X, Wu M, Zhang X, Li Q *et al* (2012). The expression levels of DNMT3a/3b and their relationship with meat quality in beef cattle. *Mol Biol Rep* **39:** 5473-5479.

Han X, Yang F, Cao H, Liang Z (2015). Malat1 regulates serum response factor through miR-133 as a competing endogenous RNA in myogenesis. *FASEB J* **29:** 3054-3064.

Hardie DG, Pan DA (2002). Regulation of fatty acid synthesis and oxidation by the AMP-activated protein kinase. *Biochem Soc Trans* **30:** 1064-1070.

Hinshaw JM (1999). Trout production: Feeds and feeding methods. . *Southern Regional Aquaculture Center* **223**.

Howe EA, Sinha R, Schlauch D, Quackenbush J (2011). RNA-Seq analysis in MeV. *Bioinformatics* **27:** 3209-3210.

Hu G, Gu W, Bai QL, Wang BQ (2013). Estimation of genetic parameters for growth traits in a breeding program for rainbow trout (Oncorhynchus mykiss) in China. *Genet Mol Res* **12:** 1457-1467.

Huang CW, Li YH, Hu SY, Chi JR, Lin GH, Lin CC *et al* (2012). Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia (Oreochromis niloticus). *J Anim Sci* **90:** 4266-4279.

Huang dW, Sherman BT, Lempicki RA (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4:** 44-57.

Huang dW, Sherman BT, Lempicki RA (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37:** 1-13.

Institute NF (1991). A Manual of Standard Methods for Measuring and Specifying the Properties of Surimi. In: National Fisheries Institute W, DC (ed).

Jeon Y, Lee JT (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146:** 119-133.

Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L *et al* (2013). The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell* **52:** 101-112.

Kambara H, Niazi F, Kostadinova L, Moonka DK, Siegel CT, Post AB *et al* (2014). Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res* **42:** 10668-10680.

Kause A, Paananen T, Ritola O, Koskinen H (2007). Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. *J Anim Sci* **85:** 3218-3227.

Kause A, Kiessling A, Martin SA, Houlihan D, Ruohonen K (2016). Genetic improvement of feed conversion ratio via indirect selection against lipid deposition in farmed rainbow trout (Oncorhynchus mykiss Walbaum). *Br J Nutr* **116:** 1656-1665.

Kern PA, Di Gregorio G, Lu T, Rassouli N, Ranganathan G (2004). Perilipin expression in human adipose tissue is elevated with obesity. *J Clin Endocrinol Metab* **89:** 1352-1358.

Kestin S, Kennedy R, Tonner E, Kiernan M, Cryer A, Griffin H *et al* (1993). Decreased fat content and increased lean in pigs treated with antibodies to adipocyte plasma membranes. *J Anim Sci* **71:** 1486-1494.

Kornienko AE, Dotter CP, Guenzl PM, Gisslinger H, Gisslinger B, Cleary C *et al* (2016). Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. *Genome Biol* **17:** 14.

Lagirand-Cantaloube J, Offner N, Csibi A, Leibovitch MP, Batonnet-Pichon S, Tintignac LA *et al* (2008). The initiation factor eIF3-f is a major target for atrogin1/MAFbx function in skeletal muscle atrophy. *EMBO J* **27:** 1266-1276.

Leeds T, Kenney P, Manor M (2012). Genetic parameter estimates for feed intake, body composition, and fillet quality traits in a rainbow trout population selected for improved growth. *International Symposium on Genetics in Aquaculture*: Auburn, AL. p 259.

Leeds TD, Vallejo RL, Weber GM, Pena DG, Silverstein JS (2016). Response to five generations of selection for growth performance traits in rainbow trout (Oncorhynchus mykiss). *Aquaculture* **465:** 341-351.

Li HH, Willis MS, Lockyer P, Miller N, McDonough H, Glass DJ *et al* (2007). Atrogin-1 inhibits Akt-dependent cardiac hypertrophy in mice via ubiquitin-dependent coactivation of Forkhead proteins. *J Clin Invest* **117:** 3211-3223.

Li J, Ma W, Zeng P, Wang J, Geng B, Yang J *et al* (2015). LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinform* **16:** 806-812.

Li T, Wang S, Wu R, Zhou X, Zhu D, Zhang Y (2012). Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics* **99:** 292-298.

Li Z, Rana TM (2014). Decoding the noncoding: prospective of lncRNA-mediated innate immune regulation. *RNA biology* **11:** 979-985.

Lian T, Wang L, Liu Y (2013). A New Insight into the Role of Calpains in Post-mortem Meat Tenderization in Domestic Animals: A review. *Asian-Australas J Anim Sci* **26:** 443-454.

Lim D, Chai HH, Lee SH, Cho YM, Choi JW, Kim NK (2015). Gene Expression Patterns Associated with Peroxisome Proliferator-activated Receptor (PPAR) Signaling in the Longissimus dorsi of Hanwoo (Korean Cattle). *Asian-Australas J Anim Sci* **28:** 1075-1083.

Liu X, Guo XY, Xu XZ, Wu M, Zhang X, Li Q *et al* (2012). Novel single nucleotide polymorphisms of the bovine methyltransferase 3b gene and their association with meat quality traits in beef cattle. *Genet Mol Res* **11:** 2569-2577.

Liu X, Usman T, Wang Y, Wang Z, Xu X, Wu M *et al* (2015). Polymorphisms in epigenetic and meat quality related genes in fourteen cattle breeds and association with beef quality and carcass traits. *Asian-Australas J Anim Sci* **28:** 467-475.

Liu X, Xiao ZD, Han L, Zhang J, Lee SW, Wang W *et al* (2016). LncRNA NBR2 engages a metabolic checkpoint by regulating AMPK under energy stress. *Nat Cell Biol* **18:** 431-442.

Londos C, Sztalryd C, Tansey JT, Kimmel AR (2005). Role of PAT proteins in lipid metabolism. *Biochimie* **87:** 45-49.

Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD (2010). Cytoscape Web: an interactive web-based network browser. *Bioinformatics* **26:** 2347-2348.

Lu L, Sun K, Chen X, Zhao Y, Wang L, Zhou L *et al* (2013). Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. *EMBO J* **32:** 2575-2588.

Lucera A, Costa C, Conte A, Del Nobile MA (2012). Food applications of natural antimicrobial compounds. *Front Microbiol* **3:** 287.

Manor ML, Cleveland BM, Kenney PB, Yao J, Leeds T (2015). Differences in growth, fillet quality, and fatty acid metabolism-related gene expression between juvenile male and female rainbow trout. *Fish Physiology and Biochemistry* **41:** 533-547.

Mathieu EL, Belhocine M, Dao LT, Puthier D, Spicuglia S (2014). [Functions of lncRNA in development and diseases]. *Medecine sciences : M/S* **30:** 790-796.

Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5:** e13984.

Messeguer X, Escudero R, Farré D, Núñez O, Martínez J, Albà MM (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18:** 333-334.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35:** W182-185.

Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR *et al* (2015). MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. *Mol Cell Biol* **35:** 498-513.

Ouyang H, Gore J, Deitz S, Korc M (2014). microRNA-10b enhances pancreatic cancer cell invasion by suppressing TIP30 expression and promoting EGF and TGF-β actions. *Oncogene* **33:** 4664-4674.

Paaver T, Gross R, Ilves P (2004). Growth rate, maturation level and flesh quality of three strains of large rainbow trout (Oncorhynchus mykiss) reared in Estonia. . *Aquacult Int* **12:** 33-45.

Paneru B, Al-Tobasei R, Palti Y, Wiens GD, Salem M (2016). Differential expression of long non-coding RNAs in three genetic lines of rainbow trout in response to infection with Flavobacterium psychrophilum. *Scientific reports* **6:** 36032.

Park BS, Oh YK, Kim MJ, Shim WB (2014). Skeletal Muscle Troponin I (TnI) in Animal Fat Tissues to Be Used as Biomarker for the Identification of Fat Adulteration. *Korean J Food Sci Anim Resour* **34:** 822-828.

Picard B, Berri C, Lefaucheur L, Molette C, Sayd T, Terlouw C (2010). Skeletal muscle proteomics in livestock production. *Brief Funct Genomics* **9:** 259-278.

Rasmussen RS (2001). Quality of farmed salmonids with emphasis on proximate composition, yield and sensory characteristics *Aquac Res* **32: 767–786**.

Robinson MD, McCarthy DJ, Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139-140.

Rora AMB, Morkore T, Einen O (2001). Primary processing (Evisceration and filleting). In: Kestin SC, Warriss PD (eds). *Farmed Fish Quality*. Blackwell Science, Oxford. pp 249-260.

Rueda A, Barturen G, Lebrón R, Gómez-Martín C, Alganza Á, Oliver JL *et al* (2015). sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res* **43:** W467-473.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N *et al* (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34:** 374-378.

Salem M, Kenney PB, Rexroad CE, Yao J (2006). Microarray gene expression analysis in atrophying rainbow trout muscle: a unique nonmammalian muscle degradation model. *Physiol Genomics* **28:** 33-45.

Salem M, Kenney PB, Rexroad CE, Yao J (2010). Proteomic signature of muscle atrophy in rainbow trout. *J Proteomics* **73:** 778-789.

Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A *et al* (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS One* **7:** e36264.

Salem M, Manor ML, Aussanasuwannakul A, Kenney PB, Weber GM, Yao J (2013). Effect of sexual maturation on muscle gene expression of rainbow trout: RNA-Seq approach. *Physiol Rep* **1:** e00120.

Salem M, Paneru B, Al-Tobasei R, Abdouni F, Thorgaard GH, Rexroad CE *et al* (2015). Transcriptome Assembly, Gene Annotation and Tissue Gene Expression Atlas of the Rainbow Trout. *PLoS ONE* **10:** e0121778.

Schmittgen TD, Livak KJ (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols* **3:** 1101-1108.

Schmitz KM, Mayer C, Postepska A, Grummt I (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* **24:** 2264-2269.

Smith RR, L. KH, M. RJ, L. RG (1988). Growth, carcass composition, and taste of rainbow trout of different strains fed diets containing primarily plant or animal protein *Aquaculture* **70:** 309–321.

Suryanti U, Bintoro V, Atmomarsono U, Pramono Y, AM L (2014). Antioxidant activity of Indonesian endogenous duck meat marinated in ginger (*Zingiber officinale* Roscoe) extract. *Int J Poult Sci* **13:** 102–107.

Tian D, Sun S, Lee JT (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143:** 390-403.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M *et al* (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40:** e115.

Verma SK, Alim A (2014). Differential activity of stanniocalcin in male and female fresh water teleost Mastacembelus armatus (Lacepede) during gonadal maturation. *PLoS One* **9:** e101439.

Wackerhage H, Ratkevicius A (2008). Signal transduction pathways that regulate muscle growth. *Essays Biochem* **44:** 99-108.

Wang B, Hsu SH, Majumder S, Kutay H, Huang W, Jacob ST *et al* (2010a). TGFbeta-mediated upregulation of hepatic miR-181b promotes hepatocarcinogenesis by targeting TIMP3. *Oncogene* **29:** 1787-1797.

Wang H, Liu D, Cao P, Lecker S, Hu Z (2010b). Atrogin-1 affects muscle protein synthesis and degradation when energy metabolism is impaired by the antidiabetes drug berberine. *Diabetes* **59:** 1879-1889.

Wang L, Zhao Y, Bao X, Zhu X, Kwok YK, Sun K *et al* (2015). LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Res* **25:** 335-350.

Watts R, Johnsen VL, Shearer J, Hittel DS (2013). Myostatin-induced inhibition of the long noncoding RNA Malat1 is associated with decreased myogenesis. *American journal of physiology Cell physiology* **304:** C995-1001.

Wei N, Pang W, Wang Y, Xiong Y, Xu R, Wu W *et al* (2014). Knockdown of PU.1 mRNA and AS lncRNA regulates expression of immune-related genes in zebrafish Danio rerio. *Dev Comp Immunol* **44:** 315-319.

Wu Z, Liu X, Liu L, Deng H, Zhang J, Xu Q *et al* (2014). Regulation of lncRNA expression. *Cellular & molecular biology letters* **19:** 561-575.

Yan B, Guo JT, Zhu CD, Zhao LH, Zhao JL (2013). miR-203b: a novel regulator of MyoD expression in tilapia skeletal muscle. *J Exp Biol* **216:** 447-451.

Yang X, Gao L, Guo X, Shi X, Wu H, Song F *et al* (2014). A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS One* **9:** e87797.

Yoshioka J, Schreiter ER, Lee RT (2006). Role of thioredoxin in cell growth through interactions with signaling molecules. *Antioxid Redox Signal* **8:** 2143-2151.

Yu H, Luscombe NM, Qian J, Gerstein M (2003). Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet* **19:** 422-427.

Yuan Z, Li J, Gao X, Gao H, Xu S (2013). Effects of DGAT1 gene on meat and carcass fatness quality in Chinese commercial cattle. *Mol Biol Rep* **40:** 1947-1954.

Zhang HL, Fan HJ, Liu XL, Wu Y, Hou SS (2013). Molecular cloning of the perilipin gene and its association with carcass and fat traits in Chinese ducks. *Genet Mol Res* **12:** 1582-1592.

Zhang L, Zhu Q, Liu Y, Gilbert ER, Li D, Yin H *et al* (2015). Polymorphisms in the Perilipin Gene May Affect Carcass Traits of Chinese Meat-type Chickens. *Asian-Australas J Anim Sci* **28:** 763-770.

Zhao W, Pan J, Wang X, Wu Y, Bauman WA, Cardozo CP (2008). Expression of the muscle atrophy factor muscle atrophy F-box is suppressed by testosterone. *Endocrinology* **149:** 5449-5460.

Zhao W, Mu Y, Ma L, Wang C, Tang Z, Yang S *et al* (2015). Systematic identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development. *Scientific reports* **5:** 8957.

Zhou M, Wang X, Li J, Hao D, Wang Z, Shi H *et al* (2015). Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. *Molecular bioSystems* **11:** 760-769.

Zhu J, Fu H, Wu Y, Zheng X (2013). Function of lncRNAs and approaches to lncRNA-protein interactions. *Science China Life sciences* **56:** 876-885.

Zhu L, Qin H, Li PY, Xu SN, Pang HF, Zhao HZ *et al* (2012). Response gene to complement-32 enhances metastatic phenotype by mediating transforming growth factor beta-induced epithelial-mesenchymal transition in human pancreatic cancer cell line BxPC-3. *J Exp Clin Cancer Res* **31:** 29.

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G *et al* (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143:** 46-58.

APPENDICES

Target gene	Forward primer	Reverse primer	Trait	Amplicon size
GSONMT00035227001	5-ATTGAGAACCAGGCTGCTGT-3	5-GATTGATGTCCACTGGGAAT-3	WBW	80-100 bp
GSONMT00011545001	5-GGAATGGATCCTATGGCAGA-3	5-AGGAGGAGCTGTGAAAAGCA-3	Fat%	80-100 bp
GSONMT00027968001	5-GCCTCCTTCCTTCAAGATCC-3	5-TCTCTCTCCAATGCCTCCAC-3	Shear force	80-100 bp
GSONMT00071198001	5-TAGCTCCGGGGAATACACAG-3	5-GGATCCAAGTCCCTGACAGA-3	Shear force	80-100 bp
GSONMT00065682001	5-CCCCAAAGTCAAAGAGTGGA-3	5-TCTGTTGGCCTAATGCCTTT-3	Fat%	80-100 bp
GSONMT00070016001	5-GATCAACAGAGACCCCAAGC-3	5-AGACTGTGAGTCCCGTCAGG-3	Fat%	80-100 bp
GSONMT00038618001	5-GCAGAGATGGAGCTGAGGAT-3	5-GGGCAGCATGTGTAGTGTTG-3	Fat%	80-100 bp
GSONMT00032573001	5-TGATCACCACCAACCCCTAC-3	5-GGCAATGAACTCGTCCACAT-3	Muscle%	80-100 bp
GSONMT00065419001	5-AGTTCTTCCTGACCCTGCAA-3	5-ATTCGGGGATGAGTGTTCTG-3	Fat%	80-100 bp
GSONMT00027011001	5-CCAAGCTACTGGAACGGAGA-3	5-CTCAAAATGGTCTCGGCAGT-3	Fat%	80-100 bp
GSONMT00012525001	5-CTGGGGCCAGACATGGAGAC-3	5-GCGGAGTTAGCACCTTGGTT-3	Shear force	80-100 bp
TCONS_00012031	5-GGTGCCAGGCGATAAGAATA-3	5-TACCGACCAGCCATGTATGA-3	Fat%	80-100 bp
Omy200087556	5-CATCAGCCCTGGTCAAAGTT-3	5-CGGAACGTTTTCTCCTATGG-3	Fat%	80-100 bp
Omy100092090	5-GGGAAGCGTCGATCTCTAAC-3	5-CCACATTCAACGTCCAGCTA-3	Shear force	80-100 bp
Omy300030188	5-AACAGGAGCCAAGGTTAAGG-3	5-CACAAGTTCTCAGCGGTCTG-3	Muscle%	80-100 bp
Omy300041379	5-GTCTGGTCATCCCTTTTGGA-3	5-TCCCACACACCGGTAGAAAT-3	Muscle%	80-100 bp
Omy200216039	5-TAACCCACCGGACTCTTCAG-3	5-TCTTGGGGACCAGTATCTTCA-3	Fat%	80-100 bp
Omy400178299	5-AAAATGCTTGTCCCTTGGAA-3	5-GAGACACCACCCAAACACAA-3	90 samples	80-100 bp
Omy500041161	5-CGTTGGCTGAAATCCTGACT-3	5-CCCCATACGATGACCCTCTA-3	90 samples	80-100 bp
GSONMT00065900001	5-AGGATTTGCTCGAAGCTGAA-3	5-ATCCTTGGAGTGGGGAATCT-3	90 samples	80-100 bp
GSONMT00039165001	5-GGAAGTTTGCTCACCTGGTC-3	5-ATTCTCTCATGCGTCCTGCT-3	90 samples	80-100 bp
GSONMT00041090001	5-GAGCTGGTGAAGAGGAAACG-3	5-TCCCTCCTGGTCAATCTCTG-3	90 samples	80-100 bp
Omy500089619	5-CTAGCCCTGCCTCTGGTAAT-3	5-TTGGAAAGAGACACACGACTTC-3	90 samples	80-100 bp

Appendix A: The list of primers used for qPCR analyses.

Trait	Family #	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Mean	SD	CV
Low WBW	1	754.00	118.00	380.90	640.00	694.50	517.48	264.82	0.51
	2	594.00	633.00	369.80	632.70	406.70	527.24	128.53	0.24
	3	753.80	276.50	494.50	412.60	381.20	463.72	179.98	0.39
	4	476.70	208.80	813.80	N/A	N/A	499.77	303.16	0.61
High WBW	1	1360.60	1289.20	1389.90	1249.70	934.80	1244.84	182.04	0.15
	2	1239.80	1146.40	1028.60	1251.30	1018.10	1136.84	111.36	0.10
	3	1665.90	1126.10	1270.40	1152.10	1079.60	1258.82	238.21	0.19
	4	1031.30	1073.70	1094.40	1269.00	1327.70	1159.22	130.69	0.11
Low muscle%	1	42.48	44.43	45.09	45.11	47.14	44.85	1.67	0.04
	2	43.48	43.79	45.76	45.97	45.76	44.95	1.21	0.03
	3	41.49	43.54	46.05	46.36	47.22	44.93	2.36	0.05
	4	42.06	45.00	45.88	N/A	N/A	44.31	2.00	0.05
High muscle%	1	48.38	48.94	49.59	49.77	49.87	49.31	0.63	0.01
	2	48.36	49.83	50.49	50.56	50.71	49.99	0.97	0.02
	3	47.51	49.55	49.22	49.11	52.47	49.57	1.80	0.04
	4	49.26	49.00	49.09	49.67	N/A	49.25	0.30	0.01
Low fat%	1	4.55	5.69	7.23	7.22	7.10	6.36	1.20	0.19
	2	5.32	5.99	5.45	6.26	7.02	6.01	0.68	0.11
	3	5.64	5.89	7.01	6.56	6.52	6.32	0.55	0.09
	4	4.99	6.17	8.54	N/A	N/A	6.57	1.81	0.28
High fat%	1	8.43	9.52	8.79	9.50	10.54	9.35	0.81	0.09
	2	7.97	9.89	9.60	10.70	10.21	9.67	1.04	0.11
	3	9.54	8.89	9.06	9.97	12.17	9.93	1.32	0.13
	4	8.66	8.89	9.57	9.77	N/A	9.22	0.53	0.06
Low shear force	1	339.84	400.71	368.77	299.75	323.54	346.52	39.34	0.11
	2	415.19	351.46	312.89	430.81	180.75	338.22	100.14	0.30
	3	240.64	244.08	275.15	459.88	363.01	316.55	94.11	0.30
	4	166.98	.	310.48	N/A	N/A	238.73	101.47	0.43
High shear force	1	471.13	492.90	393.24	682.74	638.92	535.79	121.01	0.23
	2	526.52	561.10	511.88	614.19	N/A	553.42	45.46	0.08
	3	605.36	500.83	545.55	467.09	N/A	529.71	59.80	0.11
Low whiteness	1	38.85	44.36	41.45	41.17	41.31	41.43	1.96	0.05
	2	39.79	39.59	40.47	41.63	42.99	40.89	1.42	0.03
	3	40.55	40.87	42.50	42.99	41.88	41.76	1.04	0.02
	4	43.05	39.28	42.03	39.06	N/A	40.86	1.99	0.05
High whiteness	1	42.83	44.44	43.97	46.40	46.64	44.86	1.63	0.04
	2	44.34	44.18	46.73	42.72	46.14	44.82	1.62	0.04
	3	45.46	45.31	46.58	45.33	43.96	45.33	0.93	0.02
	4	50.12	41.03	44.81	43.33	45.71	45.00	3.37	0.07
	5	45.57	40.10	45.86	41.36	N/A	43.22	2.93	0.07

Appendix B: Variations in mean, standard variation and phenotypic coefficient for families showing divergent phenotypes.

Family	Raw reads	Trimmed reads	Mapped reads	Unique mapped reads
F1	11,463,752	11,072,326	9,598,340	8,359,665
F2	11,501,736	11,111,392	9,792,945	8,485,417
F3	10,265,281	9,885,924	8,753,482	7,661,743
F4	12,061,693	11,630,088	10,249,808	8,927,833
F5	10,771,278	10,406,390	9,069,517	7,905,686
F6	11,186,848	10,752,131	9,542,654	8,221,035
F7	10,550,389	10,154,005	9,004,224	7,803,555
F8	11,583,491	11,185,994	9,810,265	8,488,121
F9	11,501,241	11,050,540	9,769,383	8,352,365
F10	10,904,096	10,497,696	9,119,865	7,994,655
F11	10,639,704	10,237,197	9,023,906	7,882,743
F12	11,062,200	10,633,054	9,334,973	8,165,824
F13	14,170,919	13,695,400	11,980,050	10,413,623
F14	11,433,311	11,041,530	9,664,721	8,355,389
F15	12,912,221	12,421,871	11,032,928	9,432,616
F16	12,534,185	12,109,622	10,668,895	9,188,361
F17	13,039,308	12,549,276	11,052,150	9,569,354
F18	10,832,313	10,417,604	9,234,856	8,134,466
F19	12,710,308	12,284,239	10,815,942	9,354,612
F20	13,734,875	13,272,610	11,589,030	10,217,936
F21	13,224,624	12,732,116	10,604,171	9,179,738
F22	11,550,847	11,162,389	9,747,101	8,441,320
Total	259,634,620	250,303,394	219,459,206	190,536,057

Appendix C: Quality and mapping statistics of sequencing reads.

Genes involved in anabolic and catabolic processes (n = 46)

Name	Symbol		Annotation	Traits	Fold change	FDR
TCONS_00058870	FBX32	Protein ubiquitination-related	F-box only protein 32	WBW|Muscle%|Fat%	-3.61|-3.16|-3.39	2.5E-02|4.0E-02|4.3E-02
TCONS_00058871	FBX32		F-box only protein 32	WBW	-3.14	4.90E-02
GSONMT00016768001	FBX32		F-box only protein 32	WBW	-2.36	3.20E-09
GSONMT00031929001	FBX32		F-box only protein 32	WBW	-2.19	2.10E-06
TCONS_00098636	FBX32		F-box only protein 32	WBW	-2.71	4.60E-02
GSONMT00048131001	KLHDB		kelch-like protein diablo-like isoform x2	WBW|Muscle%|Fat%	2.6|2.77|2.66	3.6E-05|3.3E-06|1.5E-05
GSONMT00043239001	KLH20		kelch-like protein 20-like	WBW|Muscle%|Fat%	2.5|2.95|3.05	4.6E-02|3.7E-03|2.7E-03
TCONS_00037768	ECM29		proteasome-associated ECM29 homolog	Whiteness	3.29	4.20E-02
TCONS_00152966	ECM29		Proteasome-associated ECM29 homolog	Whiteness	6.02	2.10E-02
GSONMT00060609001	K0408	Autophagy-related	protein soga3-like isoform x2	WBW|Muscle%	-2|-2.33	3.9E-02|1.6E-02
GSONMT00058681001	VWF		von willebrand factor	WBW|Fat%	2.69|2.93	4.9E-02|3.2E-02
GSONMT00062737001	MMP9	Protease-related activity	matrix metalloproteinase 9	Shear force	-5.21	4.60E-02
GSONMT00075197001	CBPC1		cytosolic carboxypeptidase 1-like isoform x2	WBW|Muscle%|Fat%	2.13|2.17|2.11	2.69E-02|2.71E-02|3.69E-02
GSONMT00000383001	LXN		latexin precursor	Fat%	-2.13	4.10E-03
GSONMT00005930001	SCRB2	Lysosome-related	lysosome membrane protein 2-like isoform x2	WBW|Muscle%|Fat%	2.51|2.3|2.58	3.5E-03|1.7E-02|2.2E-03
C20247_c1_seq1_	LYAG		Lysosomal alpha-glucosidase	Shear force	2.66	4.70E-03
GSONMT00012076001	SCRB2		lysosome membrane protein 2-like isoform x2	Muscle%	2	1.00E-04
GSONMT00065477001	ANKR1	Transcription regulators	ankyrin repeat domain-containing protein 1-like	WBW|Muscle%|Fat%	-6.32|-4.99|-6.63	2.6E-06|2.4E-05|3.2E-06
GSONMT00041801001	GREM1		gremlin-1-like	WBW|Fat%	4.63|4.32	3.9E-02|3.9E-02
GSONMT00034829001	RGCC		response gene to complement 32 protein	WBW	-2.3	0.031949
GSONMT00063334001	DS22B		dual specificity protein phosphatase 22-b-like	Muscle%|Fat%	-2.06|-2.1	3.2E-02|3.3E-02
GSONMT00060133001	CACO1		calcium-binding and coiled-coil domain-containing protein 1-like	WBW|Whiteness	-2|2.28	1.3E-15|3.2E-24
GSONMT00065198001	ANKR1		ankyrin repeat domain-containing protein 1	Shear force	3.46	1.20E-04
GSONMT00057355001	FZD7B		atrial natriuretic peptide-converting enzyme-like	WBW|Fat%	2.27|2.38	2.3E-02|1.1E-02
GSONMT00036628001	CEBPA		ccaat enhancer-binding protein alpha-like	WBW|Fat%	2.38|2.36	1.0E-02|1.1E-02
GSONMT00011545001	TBX1		T-box transcription factor TBX1 isoform X1	Fat%	2.1	4.18E-03
GSONMT00029001001	BHE40		class e basic helix-loop-helix protein 40-like	Whiteness	-2.08	3.10E-02
GSONMT00069458001	POLY		retrotransposon-like family member (retr-1)-like	WBW|Fat%	2.85|2.77	2.1E-02|3.2E-02
GSONMT00041695001	CACO1		calcium-binding and coiled-coil domain-containing protein 1-like	WBW|Whiteness	-2.03|2.36	1.1E-19|1.5E-30
GSONMT00061226001	RTF1		rna polymerase-associated protein rtf1 homolog	Muscle%|Shear force	2.16|2.1	3.7E-03|3.7E-02
TCONS_00046480	GAS6		Growth arrest-specific 6	Whiteness	2.5	4.30E-02
GSONMT00011893001	CTGF	Others involved in growth/ proliferation	connective tissue growth factor	Whiteness	-9.19	3.10E-06
GSONMT00055672001	GSDF		gonadal soma derived factor	WBW|Muscle%	2.73|2.33	2.6E-03|3.4E-02
GSONMT00010157001	SRPK		srsf protein kinase 3-like	WBW	2.13	1.10E-02
GSONMT00070016001	CAV1		caveolin-3-like	WBW|Muscle%|Fat%	-3.29|-3.25|-3.73	2.00E-04|2.00E-04|1.04E-04
GSONMT00011010001	TPM1		tropomyosin alpha-3 chain	WBW|Muscle%	2.23|2.23	2.1E-10|2.3E-10
GSONMT00045295001	PROS		vitamin k-dependent protein s	WBW|Muscle%|Fat%|Whiteness	3.36|2.73|3.18|2.04	3.2E-19|1.1E-11|6.6E-17|1.5E-08
GSONMT00031199001	CILP1		cartilage intermediate layer protein 1-like	Whiteness	2.23	5.30E-06
GSONMT00037697001	TSP4B		cartilage oligomeric matrix protein	WBW|Fat%	2.08|2.07	2.4E-02|2.7E-02
GSONMT00045692001	ITA7		integrin alpha-7-like	Whiteness	-2.3	4.50E-02
GSONMT00076980001	MLRV		myosin regulatory light chain ventricular cardiac muscle isoform-like	Whiteness	2.19	1.30E-10
TCONS_00003078	PDLIM5		PDZ and LIM domain 5-like isoform X4	WBW|Muscle%|Fat%	2.55|2.13|2.41	6.5E-25|1.6E-14|5.2E-21
TCONS_00015827	PDLIM5		PDZ and LIM domain 5	WBW|Fat%	2.27|2.31	6.1E-04|5.4E-04
TCONS_00058866	PDLIM5		PDZ and LIM domain 5	WBW|Fat%	2.39|2.3	5.3E-10|9.2E-09
TCONS_00116788	ACT1		Actin-1	WBW|Shear force	2.51|-2.03	3.0E-06|1.0E-04
TCONS_00130305	CO6A1		Collagen alpha-1(VI) chain	Whiteness	2.16	2.30E-02

Genes involved in fat metabolism (n = 30)

Name	Symbol		Annotation	Traits	Fold change	FDR
GSONMT00070016001	CAV1	DE genes in fish families of contrasting fat content	caveolin-3-like	WBW|Muscle%|Fat%	-3.29|-3.25|-3.73	2.00E-04|2.00E-04|1.04E-04
GSONMT00000701001	AAKG2		5 -amp-activated protein kinase subunit gamma-2- partial	WBW|Muscle%|Fat%	-4.14|-3.68|-3.86	3.5E-02|3.5E-02|3.5E-02
GSONMT00038618001	SHLB1		endophilin-b1 isoform x1	Fat%	-3.86	0.000639555
GSONMT00017014001	G0S2		lymphocyte g0 g1 switch protein 2	WBW|Muscle%|Fat%	4.47|3.66|3.84	3.8E-06|4.5E-04|2.2E-04
GSONMT00019603001	LIPS		hormone-sensitive lipase isoform x2	Fat%	2.6	2.70E-02
GSONMT00076211001	PLIN1		perilipin-1	WBW|Muscle%|Fat%	2.62|2.36|2.64	5.0E-11|3.8E-08|4.5E-11
GSONMT00054194001	LIPL		lipoprotein lipase	WBW|Muscle%|Fat%	2.62|2.19|2.46	2.2E-06|1.0E-03|3.9E-05
GSONMT00079455001	APOA4		apolipoprotein a-iv	WBW|Muscle%|Fat%|Whiteness	3.51|3.29|3.27|2.95	2.4E-12|1.2E-10|1.3E-10|2.9E-10
GSONMT00000920001	FABPH		fatty acid-binding heart	WBW|Muscle%|Fat%	2.45|2.23|2.25	1.9E-39|8.9E-31|3.2E-31
GSONMT00080511001	APMAP		adipocyte plasma membrane-associated protein	WBW|Muscle%|Fat%|Whiteness	2.69|2.23|2.58|2.03	9.2E-09|5.9E-05|1.0E-07|2.4E-05
GSONMT00075321001	DGAT2		diacylglycerol o-acyltransferase 2	WBW|Muscle%|Fat%	2.85|2.38|2.73	2.3E-03|4.3E-02|5.3E-03
GSONMT00029700001	G0S2		lymphocyte g0 g1 switch protein 2	WBW|Muscle%|Fat%	3.32|2.91|3.03	1.3E-20|4.1E-15|1.4E-16
GSONMT00079925001	G0S2		lymphocyte g0 g1 switch protein 2	WBW|Muscle%|Fat%	3.1|2.6|2.73	1.3E-65|1.2E-42|3.2E-48
GSONMT00060484001	G0S2		lymphocyte g0 g1 switch protein 2	WBW|Muscle%|Fat%	2.71|2.33|2.38	5.6E-04|1.4E-02|1.1E-02
GSONMT00076212001	PLIN2		perilipin-2 isoform x3	WBW|Fat%	2.22|2.1	8.0E-15|1.8E-12
GSONMT00039165001	LIPL		lipoprotein lipase	WBW|Muscle%|Fat%|Whiteness	3.53|2.93|3.53|2.17	3.0E-11|3.7E-07|3.9E-11|7.4E-06
TCONS_00012031	LIPL		Lipoprotein lipase	WBW|Muscle%|Fat%|Whiteness	4.53|3.97|4.72|2.39	2.4E-06|8.8E-05|1.2E-06|3.1E-03
GSONMT00064603001	APMAP		adipocyte plasma membrane-associated protein	WBW|Fat%	2.22|2.03	6.3E-05|1.5E-03
GSONMT00036628001	CEBPA		ccaat enhancer-binding protein alpha-like	WBW|Fat%	2.38|2.36	1.0E-02|1.1E-02
GSONMT00065419001	CIDEC		cell death activator cide-3	WBW|Fat%	3.71|3.58	2.0E-02|3.2E-02
GSONMT00011893001	CTGF		connective tissue growth factor	Whiteness	-9.19	3.10E-06
GSONMT00000536001	APOEB		apolipoprotein e	Whiteness	2.81	1.00E-08
GSONMT00016296001	SAA5		serum amyloid a	Whiteness	4.92	2.40E-02
GSONMT00067189001	PLIN2		perilipin-2 isoform x2	WBW	2.2	0.032
GSONMT00077214001	AAKG2		5 -amp-activated protein kinase subunit gamma-2-like	WBW|Muscle%	-5.21|-3.36	1.3E-02|3.5E-02
GSONMT00029837001	TECR		very-long-chain enoyl- reductase-like	Muscle%	-3.61	1.00E-02
GSONMT00018640001	TECR		very-long-chain enoyl- reductase-like	WBW	-2.04	3.70E-02
GSONMT00028411001	FABPI		fatty acid-binding intestinal	WBW|Muscle%	2.16|2.13	4.3E-31|8.6E-30
GSONMT00058883001	RET7		retinoid-binding protein 7	WBW|Muscle%	2.23|2.01	8.0E-04|9.5E-03
GSONMT00055312001	RET7		retinoid-binding protein 7	WBW	2.51	2.60E-02

Gene ID	Symbol	Family	Description	Trait	Fold change	FDR
Structural genes (n = 50)						
GSONMT00032573001	MYSS	Myosin	myosin heavy chain	WBW|Muscle%	43.33|43.40	3.1E-02|3.1E-02
GSONMT00011437001	MYSS		myosin heavy chain	WBW|Muscle%|Fat%|Whiteness	-4.26|-2.69|-2.6|-3.48	2.1E-10|6.0E-06|1.2E-05|3.3E-10
GSONMT00057539001	MYSS		myosin heavy chain	Muscle%	-2.51	1.60E-73
GSONMT00002223001	MYSS		myosin heavy chain	Muscle%	-2.36	5.80E-04
GSONMT00021873001	MYSS		myosin heavy chain	Muscle%|Fat%	-2.1|-2.33	9.6E-14|4.5E-17
GSONMT00011435001	MYSS		myosin heavy chain	Fat%	-2.01	2.40E-02
GSONMT00017687001	MYSS		myosin heavy chain	Whiteness	2.13	2.20E-02
GSONMT00037087001	MYSS		myosin heavy chain	Shear force	3.63	1.50E-05
GSONMT00049811001	MYH4		myosin heavy fast skeletal muscle- partial	Muscle%|Fat%	-2.81|-3.1	4.5E-02|2.3E-02
GSONMT00065372001	MYSS		myosin heavy fast skeletal muscle- partial	Shear force	2.53	1.90E-02
GSONMT00065375001	MYH8		myosin heavy partial	WBW|Muscle%|Shear force	2.25|2.28|2.69	2.38E-02|2.40E-02|2.86E-02
GSONMT00065371001	MYH7		myosin-7- partial	WBW|Muscle%|Shear force	2.45|2.43|2.69	2.0E-02|2.2E-02|2.1E-02
GSONMT00065370001	MYH7		myosin-7b-like isoform x3	WBW|Shear force	2.36|2.64	4.7E-02|2.1E-02
GSONMT00059652001	MYH7		slow myosin heavy chain 1	Whiteness	2.45	1.40E-02
GSONMT00067106001	MYL3		myosin light chain 3-like	Shear force	2.45	5.00E-18
GSONMT00075636001	MYL4		myosin light chain 4-like	Whiteness	2.19	1.50E-05
GSONMT00065900001	TNNI2	Troponin	troponin fast	WBW|Muscle%|Fat%|Shear force	99.04|72|49.87|2.99	2.6E-10|4.0E-07|1.3E-04|2.8E-02
TCONS_00057247	TNNI2		troponin fast skeletal muscle-like	WBW|Muscle%	12.82|13.64	3.0E-07|5.9E-08
GSONMT00065895001	TNNI2		troponin fast skeletal muscle-like	WBW|Muscle%|Fat%|Shear force	-2.75|-2.71|-2.27|-2.25	0.0E+00|0.0E+00|2.2E-236|5.9E-192
GSONMT00023675001	TNNI2		troponin fast skeletal muscle-like	WBW|Muscle%|Fat%	-2.38|-2.17|-2.1	6.3E-101|3.8E-85|2.4E-82
GSONMT00006413001	TNNC2		troponin skeletal muscle-like	WBW|Muscle%|Fat%	-2.81|-2.6|-2.45	7.2E-109|1.3E-96|2.6E-87
GSONMT00014443001	TNNT1		troponin slow skeletal muscle-like	Fat%	-2.79	8.50E-04
GSONMT00043840001	TNNI3		troponin slow skeletal muscle-like	Whiteness	-2.19	4.70E-02
GSONMT00043393001	NEBU	Upregulated in families of high WBW & muscle%	nebulin-like	WBW|Muscle%	2.46|2.48	3.33E-02|3.36E-02
GSONMT00011012001	TPM1		slow myotomal muscle tropomyosin	WBW|Muscle%|Fat%	2.46|2.48|2.11	1.0E-19|7.5E-20|4.8E-12
GSONMT00011010001	TPM1		tropomyosin alpha-3 chain	WBW|Muscle%	2.23|2.23	2.1E-10|2.3E-10
GSONMT00072046001	MYBPC3		myosin-binding cardiac-type-like	WBW|Muscle%	2.04|2.01	1.6E-04|2.9E-04
TCONS_00003078	PDLIM5		PDZ and LIM domain 5-like isoform X4	WBW|Muscle%|Fat%	2.55|2.13|2.41	6.5E-25|1.6E-14|5.2E-21
GSONMT00076552001	ACTC		alpha cardiac muscle 1	WBW|Muscle%|Fat%	3.61|5.24|4.59	8.5E-03|9.3E-06|1.6E-04
GSONMT00075410001	K1C13		type i cytoskeletal 13-like	WBW|Muscle%|Fat%	2.46|2.11|2.31	2.6E-49|3.9E-31|1.5E-41
GSONMT00049606001	K1C13		type i enveloping like	WBW|Muscle%	3.05|2.97	7.4E-03|1.2E-02
GSONMT00040664001	K2C8		type ii keratin e3	WBW|Muscle%|Fat%|Shear force	4|3.89|2.39|-3.18	1.4E-08|3.8E-06|1.7E-02|3.5E-05
GSONMT00010161001	K2C8		type ii keratin e3	WBW|Muscle%	3.66|3.63	9.9E-04|1.6E-03
GSONMT00070016001	CAV1		caveolin-3-like	WBW|Muscle%|Fat%	-3.29|-3.25|-3.73	2.00E-04|2.00E-04|1.04E-04
TCONS_00116788	ACT1		Actin-1	WBW|Shear force	2.51|-2.03	3.0E-06|1.0E-04
GSONMT00037697001	TSP4B		cartilage oligomeric matrix protein	WBW|Fat%	2.08|2.07	2.4E-02|2.7E-02
TCONS_00130305	CO6A1		Collagen alpha-1(VI) chain	Whiteness	2.16	2.30E-02
GSONMT00062170001	NFH		collagen triple helix repeat protein	Shear force	-4.56	3.90E-02
TCONS_00011906	DESM		Desmin	WBW	-2.17	2.00E-04
TCONS_00014415	FSD2L		fibronectin type III and SPRY domain-containing 2	Shear force	-2.22	0.0401
GSONMT00051870001	FLNC		filamin-c-like isoform x2	Whiteness	2.16	1.20E-03
GSONMT00045692001	ITA7		integrin alpha-7-like	Whiteness	-2.3	4.50E-02
GSONMT00010163001	K2C8		keratin 8	WBW|Fat%	2.19|2.11	1.4E-56|3.7E-50
GSONMT00045963001	MFAP4		microfibril-associated glycoprotein 4-like	WBW|Whiteness	3.2|3.27	3.7E-02|4.2E-02
GSONMT00010164001	K1C18		type i cytoskeletal 18-like	WBW|Fat%	2.19|2.14	5.6E-15|6.3E-14
GSONMT00076980001	MLRV		myosin regulatory light chain ventricular cardiac muscle isoform-like	Whiteness	2.19	1.30E-10
GSONMT00007147001	MYPC2		myosin-binding protein fast-type-like	Fat%	-2.08	2.30E-02
TCONS_00208113	MYOZ2		Myozenin-2	Whiteness	-2	4.10E-03
TCONS_00015827	PDLI5		PDZ and LIM domain 5	WBW|Fat%	2.27|2.31	6.1E-04|5.4E-04
TCONS_00058866	PDLI5		PDZ and LIM domain 5	WBW|Fat%	2.39|2.3	5.3E-10|9.2E-09
Genes involved in calcium metabolism (n = 22)						
GSONMT00027968001	STC		Stanniocalcin	Shear force	11.08	1.84E-08
GSONMT00012525001	STC		Stanniocalcin	Shear force	23.75	1.40E-06
TCONS_00012355	STC		Stanniocalcin	Shear force	12.82	2.20E-05
GSONMT00063580001	PRVA		parvalbumin	WBW|Muscle%|Fat%|Shear force	-4.2|-4.89|-3.97|-6.02	1.8E-104|5.1E-119|3.4E-99|3.1E-136
GSONMT00043621001	PRVB2		parvalbumin beta 2	WBW|Muscle%|Fat%	-2.48|-2.17|-2.19	1.9E-246|1.0E-190|1.1E-192
GSONMT00025610001	PRV7		parvalbumin-7-like isoform x1	Shear force	2.11	5.40E-05
GSONMT00025642001	PRVA		parvalbumin-7-like isoform x1	Whiteness	2.35	4.60E-05
GSONMT00006413001	TNNC2		troponin skeletal muscle-like	WBW|Muscle%|Fat%	-2.81|-2.6|-2.45	7.2E-109|1.3E-96|2.6E-87
GSONMT00053755001	FZD7B		atrial natriuretic peptide-converting enzyme-like	WBW|Fat%	2.27|2.38	2.3E-02|1.1E-02
TCONS_00048015	RCAN1		Calcipressin-1	Muscle%	2.53	2.80E-02
GSONMT00041695001	CACO1		calcium-binding and coiled-coil domain-containing protein 1-like	WBW|Whiteness	-2.03|2.36	1.1E-19|1.5E-30
GSONMT00060133001	CACO1		calcium-binding and coiled-coil domain-containing protein 1-like	WBW|Whiteness	-2|2.28	1.3E-15|3.2E-24
GSONMT00037697001	TSP4B		cartilage oligomeric matrix protein	WBW|Fat%	2.08|2.07	2.4E-02|2.7E-02
GSONMT00070016001	CAV1		caveolin-3-like	WBW|Muscle%|Fat%	-3.29|-3.25|-3.73	2.00E-04|2.00E-04|1.04E-04
GSONMT00011893001	CTGF		connective tissue growth factor	Whiteness	-9.19	3.10E-06
GSONMT00045692001	ITA7		integrin alpha-7-like	Whiteness	-2.3	4.50E-02
GSONMT00067106001	MYL3		myosin light chain 3-like	Shear force	2.45	5.00E-18
GSONMT00075636001	MYL4		myosin light chain 4-like	Whiteness	2.19	1.50E-05
GSONMT00076980001	MLRV		myosin regulatory light chain ventricular cardiac muscle isoform-like	Whiteness	2.19	1.30E-10
TCONS_00045156	TRDN		triadin isoform X3	Whiteness	-2.23	5.00E-02
GSONMT00066744001	TRDN		triadin- partial	Shear force|Whiteness	2.04|-2.03	1.3E-02|1.2E-02
GSONMT00045295001	PROS		vitamin k-dependent protein s	WBW|Muscle%|Fat%|Whiteness	3.36|2.73|3.18|2.04	3.2E-19|1.1E-11|6.6E-17|1.5E-08
Genes involved in oxidative stress (n = 8)						
GSONMT00020998001	THIO		thioredoxin	WBW|Muscle%|Fat%	3.25|2.66|3.05	2.0E-15|2.7E-09|3.4E-13
GSONMT00070684001	GPX1		glutathione peroxidase 1	WBW|Muscle%|Fat%	2.62|2.62|2.81	3.7E-24|5.5E-24|1.4E-28
GSONMT00035467001	WWOX		retinol dehydrogenase 11-like	Whiteness	-2.93	6.60E-03
GSONMT00011651001	FMO5		dimethylaniline monooxygenase	WBW|Muscle%|Fat%	-2.23|-2.41|-2.22	4.75E-02|2.82E-02|4.79E-02
GSONMT00078333001	CX6B1		cytochrome c oxidase subunit vib isoform 1	Shear force	2.17	3.60E-16
GSONMT00029837001	TECR		very-long-chain enoyl-CoA reductase-like	Muscle%	-3.61	1.00E-02
GSONMT00018640001	TECR		very-long-chain enoyl-CoA reductase-like	WBW	-2.04	3.70E-02
GSONMT00045968001	NB5R3		nadh-cytochrome b5 reductase 3-like	WBW	2.04	3.90E-04

Appendix D: Selected list of DE protein-coding genes (FDR < 0.05 and fold change [high/low] ≥ 2 or ≤ -2) identified in the five studied traits.

CHAPTER II

GENOME-WIDE ASSOCIATION ANALYSIS WITH A 50K TRANSCRIBED GENE SNP-CHIP IDENTIFIES QTL AFFECTING MUSCLE YIELD IN RAINBOW TROUT

ABSTRACT

Detection of coding/functional SNPs that change the biological function of a gene may lead to identification of putative causative alleles within QTL regions and discovery of genetic markers with large effects on phenotypes. This study has two-fold objectives, first to develop, and validate a 50K transcribed gene SNP-chip using RNA-Seq data. To achieve this objective, two bioinformatics pipelines, GATK and SAMtools, were used to identify ~21K transcribed SNPs with allelic imbalances associated with important aquaculture production traits including body weight, muscle yield, muscle fat content, shear force, and whiteness in addition to resistance/susceptibility to bacterial cold-water disease (BCWD). SNPs ere identified from pooled RNA-Seq data collected from ~620 fish, representing 98 families from growth- and 54 families from BCWD-selected lines with divergent phenotypes. In addition, ~29K transcribed SNPs without allelic-imbalances were strategically added to build a 50K Affymetrix SNP-chip. SNPs selected included two SNPs per gene from 14K genes and ~5K non-synonymous SNPs. The SNP-chip was used to genotype 1728 fish. The average SNP calling-rate for samples passing quality control (QC;

1,641 fish) was $\geq$ 98.5%. The second objective of this study was to test the feasibility of using the new SNP-chip in GWA (Genome-wide association) analysis to identify QTL explaining muscle yield variance. GWA study on 878 fish (representing 197 families from 2 consecutive generations) with muscle yield phenotypes and genotyped for 35K polymorphic markers (passing QC) identified several QTL regions explaining together up to 28.40% of the additive genetic variance for muscle yield in this rainbow trout population. The most significant QTLs were on chromosomes 14 and 16 with 12.71% and 10.49% of the genetic variance, respectively. Many of the annotated genes in the QTL regions were previously reported as important regulators of muscle development and cell signaling. No major QTLs were identified in a previous GWA study using a 57K genomic SNP chip on the same fish population. These results indicate improved detection power of the transcribed gene SNP-chip in the target trait and population, allowing identification of large-effect QTLs for important traits in rainbow trout.

INTRODUCTION

Aquaculture provides sustainable production of food fish with high protein/low-saturated fat to satisfy increasing U.S. and worldwide demand. To enable increased production by the aquaculture industry and to meet the ever-growing demand for fish, we need fast/efficient growth and high-quality fillets. However, a major constraint to increasing production efficiency is the lack of genetically improved strains of fish for aquaculture (Gjerde 2006, WorldFish Center 2009). Development of tools that will enable genomic selection for improved aquaculture production traits will greatly benefit the aquaculture industry.

Fast/efficient muscle growth is a major trait affecting profitability of the aquatic muscle food industry. The genetic basis of muscle growth traits is not well studied in fish. Understanding molecular mechanisms of fish muscle growth can facilitate broodstock selection decisions. Skeletal muscle is the most abundant tissue and edible portion of fish and typically constitutes about 50-60% of the fish weight (Salem et al 2006). Growth, development and quality traits of muscle are governed by organized expression of genes encoding contractile and regulatory proteins (Gerrard and Grant 2003).

Genetic maps, characterizing the inheritance patterns of traits, and markers have been developed and used for a wide range of species, including fish. These tools target the discovery of allelic variation affecting traits with an ultimate goal of identifying DNA sequences underlying phenotypes (Rexroad et al 2008). Markers have been identified with a variety of molecular techniques. Single nucleotide polymorphisms (SNPs) are abundant and distributed genome-wide, therefore, they are most suitable for high-throughput association studies (Gonzalez-Pena et al 2016, Wang et al 2008). Marker-assisted selection (MAS) can be used to improve breeding for phenotypes with large-effect QTLs. This method has been recently applied for the trait of infectious pancreatic necrosis virus (IPNV) resistance in Atlantic salmon (Houston et al 2008, Moen et al 2009). Genomic selection (GS) tools have been developed to increase the efficiency of genetic improvement in livestock compared to conventional pedigree-based selective breeding methods(Taylor et al 2016). This concept has been recently demonstrated for bacterial cold-water disease (BCWD) resistance in rainbow trout aquaculture (Vallejo et al 2017a). SNPs located within or near coding sequences, cSNPs, are especially important because they have the potential to change protein function (Al-Tobasei et al 2017a, Brookes 2007, Salem et al 2012).

Therefore, cSNPs are particularly useful as genetic markers with large-effect on phenotypes, allowing MAS and improved accuracy of whole-genome selection. Because the muscle yield trait targeted in this study requires lethal sampling to measure the phenotype, only family-specific EBVs are available for breeding candidates in traditional breeding programs. The ability to use genomic selection or MAS will allow further within-family selection for the muscle yield trait, and thus is anticipated to increase the accuracy of genetic predictions and selection response.

Recently, we used an RNA-Seq approach to identify putative SNPs with allelic imbalances associated with total body weight, muscle yield, muscle fat content, shear force, and whiteness (Al-Tobasei et al 2017a, Salem et al 2012). Similarly, RNA-Seq data were used to identify SNPs with allelic imbalances in fish families showing variations in resistance to *Flavobacterium psychrophilum,* the etiological agent of BCWD in rainbow trout (Al-Tobasei et al 2017b, Marancik et al 2014). Together about 50K and 229K transcribed SNPs were identified in the two studies, respectively. Of them, ~21K SNPs had allelic-imbalances in families with contrasting phenotypes. The first objective of this study was to design, develop, and validate a 50K transcribed gene SNP-chip. The chip content includes the 21K transcribed SNPs with allelic-imbalances associated with the aforementioned traits and ~29K SNPs without allelic-imbalances that were strategically added to achieve more even genome-wide distribution. The new SNP-Chip is available from Affymetrix. The second objective of this study was to test the feasibility of using the new SNP-chip in GWA analysis to identify QTL explaining muscle yield variance in the USDA/NCCCWA rainbow trout growth-selected line. The results were compared with

a previous GWA study for the same trait in the same population that we have previously conducted with a genomic-based 57K SNP chip (Gonzalez-Pena et al 2016).

MATERIALS AND METHODS

Ethics statement

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 and experimental procedures used in this study (Protocols #056 and 076).

Source and selection of SNPs for the chip

Recently, we used RNA-Seq and two bioinformatic pipelines, GATK and SAMtools, for discovering coding/functional SNPs from 98 rainbow trout fish families (5 fish each) showing variations in whole-body weight, muscle yield, muscle fat content, shear force, and whiteness (Al-Tobasei et al 2017a). GATK detected 59,112 putative SNPs and SAMtools detected 87,066 putative SNPs. The two datasets contained approximately 50K non-redundant common SNPs; of which, 30,529 mapped to protein-coding genes (with 7.7% non-synonymous SNPs) and 4,386 mapped to lncRNAs. A total of 7,930 non-redundant SNPs had allelic imbalances between the low- and high-ranked families for the phenotypes. Validation of a subset of 92 SNPs revealed 1) 86.7-93.8% success rate in identifying polymorphic SNPs and 2) 95.4% consistent matching between DNA and cDNA genotypes, indicating a high rate of identifying SNPs using RNA-Seq. This SNP data set was recently published, Al-Tobasei, et al., (Al-Tobasei et al 2017a) .and is available through the NCBI dbSNP database (accession numbers ss#2711191806-2711287038 in addition to ss#2137497773).

Similarly, we identified transcribed gene SNPs in two genetic lines, ARS-Fp-R (resistant) and ARS-FP-S (susceptible), that were created by selective breeding to exhibit divergent resistance to BCWD. RNA-Seq analysis of pooled RNA samples was used to identify SNPs from the resistant and susceptible genetic lines. Fish belonging to resistant and susceptible genetic lines were collected on day 1 and day 5 post-challenge with Fp versus PBS injection (Al-Tobasei et al 2017b, Marancik et al 2014). Using GATK bioinformatics pipelines, ~229K transcribed SNPs were identified (Al-Tobasei et al 2017b). The total number of SNPs with allelic imbalance, after removing redundant SNPs, was 7,951.

The SNPs identified in the previous two studies were used as a source to build the SNP array described in this study. About 21K transcribed SNPs with allelic-imbalances associated with the above-listed traits were included in the chip. These SNPs were identified from pooled RNA-Seq data collected from ~620 fish, representing 98 families from the ARS growth-selected line and 54 families from the ARS-Fp-R and -S lines. In addition, about 29K transcribed SNPs without allelic-imbalances were selected from all the putative SNPs and were strategically added to the chip with the aim of achieving even distribution of SNPs along the rainbow trout 29 chromosomes. The additional SNPs were selected to represent as many genes as possible in the genome: two SNPs were selected per gene from 14K genes with available SNPs. The chip includes ~5K non-synonymous SNPs. The chip has probe sets for a total number of 50,006 SNPs.

Chip genotyping quality assessment

The SNP-chip was used in genotyping 1,728 fish from the USDA-ARS genetic lines. The Affymetrix SNPolisher software was used to calculate the chip SNP- and sample-

metrics and assess QCs and filter samples/genotypes at the default setting (Palti et al 2015). Forty-seven SNPs previously genotyped by a Fluidigm PCR-based assay (Al-Tobasei et al 2017a) were used to check quality of Affymetrix chip genotyping using 120 samples genotyped by both the chip and Fluidigm SNP assays. In addition, we confirmed the quality of the SNPs and the order of the samples included in the genotyping panel through pedigree check. Among the fish genotyped we included previously confirmed parental-pairs of nine families with 470 offspring and confirmed an average of 99.4% matching between offspring SNP genotypes and the genotypes of the expected parents.

SNP genomic distribution and annotation

SNPs used in building the chip were identified using the first draft of the rainbow trout reference genome (Berthelot et al 2014). To update genomic coordinates according to the newly released genome assembly (GenBank assembly Accession GCA_002163495, RefSeq assembly accession GCF_002163495) (Gao et al 2018), SNPs were mapped by BLASTing the SNP probe sequences (70 nt) to the new genome sequence. Sequences with 100% identity match and no gap with single hits were assigned to the new genome position. Sequences with multiple hits were re-Blasted using probe size of 150 nt by adding 40 nt flanking sequence in both direction. A total of 45K SNPs out of 50K SNPs were successfully assigned to the new genome and were used for the GWA analyses.

SNPeff program was used to classify and annotate functional effects of the SNPs (Cingolani et al 2012). The gff file of the new rainbow trout genome reference was used to determine position of the SNPs in a gene i.e. located within mRNA start and end positions (genic), within a CDS, 5'UTR or 3'UTR. SNPs not within start and end positions of mRNA were considered intergenic. Upstream/ downstream intergenic SNPs were determined if

located within 5 Kb of an mRNA. SNPs within lncRNAs were determined using gtf file of our previously reported lncRNA reference (Al-Tobasei et al 2016). SNP annotation was performed by intersecting the SNPs bed file with the gff/gtf file using Bedtools software (Quinlan and Hall 2010).

Rainbow trout population and phenotypes used for GWA analysis

Genome-wide association analysis was carried out using fish from a growth-selected line that has been previously described (Leeds et al 2016). Briefly, this synthetic line is a 2-yr-old winter/spring-spawning population that was developed beginning in 2002, became a closed population in 2004, and since then has gone through 5 generations of genetic selection for improved growth performance. Fish from two consecutive generations (i.e., the third and fourth generations of growth selection) were included in this study. Phenotypic data and DNA samples were collected from 878 fish (406 fish representing 98 families from year-class (YC) 2010 and 472 fish representing 99 families from YC 2012). Among the 878 fish genotyped for GWAS, 40 fish were previously used for the discovery of the muscle yield associated SNP as described above (Al-Tobasei et al 2017a). The aforementioned SNP array was used for GWAS. Methods used to sample fish from each nucleus family and to characterize muscle yield have been described previously (Gonzalez-Pena et al 2016). Eggs were hatched in spring water at 7-13°C to synchronize hatch times. Each family was stocked separately in 200-L tanks and hand-fed a commercial fishmeal-based diet beginning at swim-up. Neomales were developed from a subset of alevins from the previous year class by feeding 2 mg/kg of 17α-methyltestosterone for 60 d post-swim-up, and the masculinized females were used as sires for the following generation. At 5-months old, fish were uniquely tagged by inserting a passive integrated transponder, and

tagged fish were combined and reared in 1,000-L communal tanks. Fish were fed a commercial fishmeal-based diet using automatic feeders. EBV were computed based on a two-trait model, 10-mo BW and thermal growth coefficient (TGC), using MTDFREML (Boldman et al 1995). Each generation, EBV was used as selection criterion and mating decisions were made to maximize genetic gain while constraining the inbreeding rate to ≤1% per generation using EVA evolutionary algorithm (Berg et al 2006). Data from masculinized fish were not used in the growth analysis.

Fish were harvested between 410 and 437 days post-hatch (mean body weight = 985 g; SD = 239 g), between 446 and 481 days post-hatch (mean body weight = 1803 g; SD = 305 g), for the 2010, and 2012 hatch years, respectively. Individual body weight data were recorded at harvesting. Fish were taken off feed 5 days before harvesting. For measurement of muscle yield when harvested at each of five consecutive weeks, approximately 100 fish (i.e., 1 fish per full-sib family per week) were anesthetized in approximately 100 mg/L of tricaine methane sulfonate (Tricaine-S, Western Chemical, Ferndale, WA) slaughtered, and eviscerated. Head-on gutted carcasses were packed in ice, transported to the West Virginia University Muscle Foods Processing Laboratory (Morgantown, WV), and stored overnight. The next day, carcasses were manually processed into trimmed, skinless fillets by a trained faculty member and weighed; muscle yield was calculated as a percent of total body weight (Salem et al 2013). The fish used in GWA had an average muscle yield of 48.91% (SD=2.42).

GWA analysis

Weighted single-step GBLUP (WssGBLUP) was used to perform GWA analysis as implemented in previous studies (Misztal et al 2014, Wang et al 2012, Wang et al 2014).

In addition to phenotypic data, wssGBLUP integrates genotype and pedigree information to increase estimation precision and detection power (Wang et al 2012) in a combined analysis that is executed by the BLUPF90 software (Misztal et al 2002).

The following mixed model was used for single trait analysis:

$y = \mathrm{X}b + \mathrm{Z}_1a + \mathrm{Z}_2w + e$

where y is the vector of the phenotypes, b is the vector of fixed effects including harvest group and hatch year, a is the vector of additive direct genetic effects (i.e., animal effect), w is the vector of random family effect, and e is the residual error. The matrices X, Z_1, and Z_2 are incidence matrices for the effects contained in b, a, and w, respectively. The additive direct genetic effect is a correlated effect with covariance structure given by $\mathbf{H}\sigma_a^2$, where σ_a^2 is the additive direct genetic variance and $\mathbf{H}$ is the realized relationship matrix that combines pedigree and genomic relationships(Legarra et al 2009). In the WssGBLUP mixed model equations, the inverse of $\mathbf{H}$ is used (Aguilar et al 2010).

$$H^{-1}=A^{-1}+\begin{bmatrix} 0 & 0 \\ 0 & G^{-1}-A_{22}^{-1} \end{bmatrix}$$

where $\mathbf{A}^{-1}$ is the inverse of the pedigree relationship matrix and has the dimension of the number of animals in the pedigree; $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree relationship matrix among genotyped animals and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix; both $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ have the dimension of the number of genotyped animals. The random family effect is uncorrelated and just accounts for the fact the animals within the same family were raised in a common environment, and the covariance structure is given by $\mathbf{I}\sigma_w^2$, where $\mathbf{I}$ is an identity matrix and σ_w^2 is the family variance.

As BLUP-based models consider the variance components are known, AIREMLF90 (Misztal et al 2002) was used to estimate variance components for the additive direct genetic effect, random family effect, and residuals. Inbreeding was considered in all analyses, and was calculated using INBUPGF90 (Misztal et al 2002) on 63,808 fish that represent five generations in the NCCWA population. Quality control (QC) of genomic data was performed using BLUBF90 (Misztal et al 2002) with the following parameters: SNP with minimum Allele Frequency (MAF) >0.05, SNP with call rate > 0.90, animals with call rate >0.90, and SNP with a difference between observed and expected allele frequency <0.15 (i.e., HWE test) were kept in the data. Out of a total of 50,006 SNPs, 35,322 SNPs passed QC.

In WssGBLUP, the weights for each SNP was assigned the same weight (e.g., 1.0, i.e., standard ssGBLUP) for the first iteration. For the second iteration, weights were calculated based on the SNP effects ($\hat{u}$) estimated in the previous iteration as $\hat{u}^2 2p(1-p)$, where p is the current allele frequency. Each iteration was performed using three steps as follows: first, weight was assigned as described above; second, BLUPF90 (Misztal et al 2002) was used to compute genomic estimated breeding values (GEBV) based on a realized relationship matrix (**H**) that combines pedigree (**A**) and genomic relationship matrix (**G**), the last considered weights for SNP; and third, postGSF90 (Misztal et al 2002) was used to calculate SNP effects and weights based on sliding variance windows of 50 adjacent SNP. A total of 2 iterations were used. A window based on physical size (i.e. specific number of nucleotides) was not used to avoid biases due to uneven distributed SNPs in the new SNP chip. A Manhattan plot based on the proportion of additive genetic variance explained by the windows was created using the qqman package in R (Turner 2014); the

genomic windows explaining significant proportion of the additive genetic variance for muscle yield could be detected.

Citrate synthase (CS) activity assay

GWA analysis (described below) showed a SNP window contained the CS gene associated with the genetic variance in muscle yield. To assess the potential effect of the SNPs in this gene, we measured the CS activity in 100 fish from the 2012 year-class as previously described (Brijs et al 2017, Seite et al 2018). Frozen muscle tissue samples were homogenized using electric homogenizer on ice followed by centrifugation at 1,000g for 15 min at 4 °C. The supernatant was used to assess the total protein concentration and CS activity. Total protein concentration was assessed using a BCA protein assay kit at 562 nm with bovine serum albumin (BSA) as the standard. CS activity was determined from the rate of appearance of reduced DTNB (5,5`-dithiobis [2-nitrobenzoic]), which was monitored with a spectrophotometer at 412 nm(Ekstrom et al 2017). For the CS assay, 10 μL of diluted tissue homogenate (1.0 mg/ ml) was incubated with 140 μL reaction medium (0.1mM DTNB, 0.2mM AcetylCoA, 0.15mM oxaloacetic acid, pH 8.0). The absorbance was read in triplicate at 412 nm (25 °C) after 4 min. CS activity was expressed as ΔOD/ mg protein.

RESULTS AND DISCUSSION

chip genotyping quality assessment

The SNP-chip was used to genotype 1,728 fish. Out of 50,006 SNPs, 32,273 SNPs (64.5%) were characterized as high quality and polymorphic and 3,458 SNPs (6.9%) were high quality monomorphic (Table 1).

Table 1. SNP chip Metric summary

Conversion Type	Count	Percentage
Poly High Resolution	32,273	64.5
Other	8,395	16.7
Mono High Resolution	3,458	6.9
No Minor Hom	2,725	5.4
Call Rate Below Threshold	2,705	5.4
Off target variant	450	0.9

The Affymetrix SNPolisher software was used to filter samples/genotypes at the default setting (Palti et al 2015). Out of 1,728 genotyped samples, 1,641 (94.9%) fish samples were retained, and 87 samples were filtered out because they failed to meet the 0.97 call rate (CR) and 0.82 Dish QC (DQC) thresholds. The average QC call rate for the passing samples was 99.6% (Table 2).

We compared the Affymetrix genotyping results of 47 SNPs that were previously genotyped by a Fluidigm PCR-based assay (Al-Tobasei et al 2017a). Using 120 samples genotyped by both methods, there was a 99.5% match in genotypes between the two assays for high-resolution polymorphic markers (data not shown). This test demonstrates the high quality of the SNP chip and reliable genotyping data for the subsequent GWA analyses. The SNP-chip showed an average minor allele frequency (MAF) of 0.25 and standard deviation of 0.134. A total of 27,280 SNPs had MAF> 0.1 and 16,101 SNP more than 0.25 (Figure 1).

Table 2. SNP chip Sample QC Summary

Number of input samples	1,728
Samples passing DQC	1,722
Samples passing DQC and QC CR	1,641
Samples passing DQC, QC CR and Plate QC	1,641 (94.9%)
Number of failing samples	87
Number of Samples Genotyped	1,641
Average QC CR for the passing samples	99.66
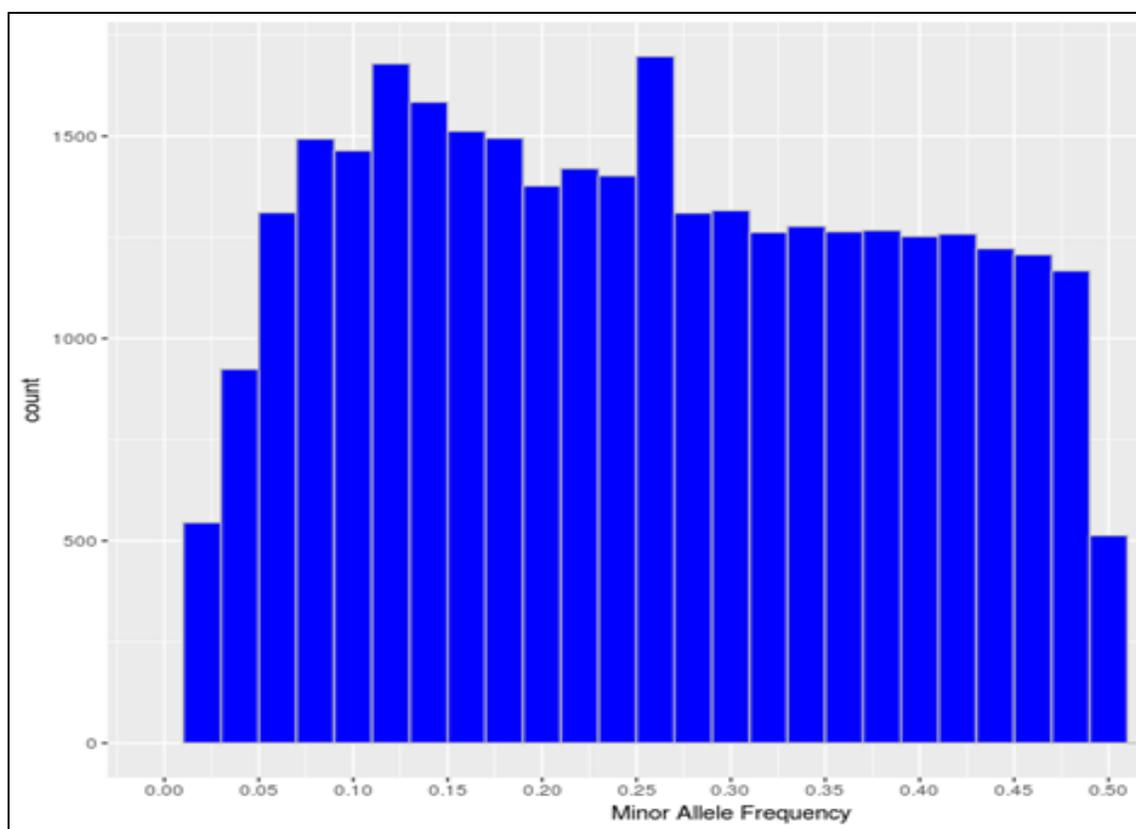
Figure 1. Minor allele frequency distribution of the polymorphic high-resolution SNPs in the SNP chip.

SNP density and genomic distribution

SNPs used in this study to build the chip were initially identified using a rainbow trout reference genome published by Berthelot et al. in 2014 (Berthelot et al 2014). However, in this reference only ~1 Gb out of a 2.1 Gb total length of the assembly is anchored to chromosomes. Recently, a newer genome assembly has been built that is currently available at NCBI (Accession GCA_002163495) (Gao et al 2018). The new assembly has a 1.94 Gb total length (89% of the genome) anchored to 29 chromosomes. A total of 45,590 SNPs out of 50,006 existing in the SNP-chip were mapped to the new genome assembly with an average of 1,572 SNPs per chromosome. The average SNP density was 1 SNP per 42.7 Kb, with a range of 1 SNP/33.5 Kb (Chromosome 16) to 1 SNP/61.6 Kb (Chromosome 23). Figure 2 shows the number of SNPs per chromosome and the SNP density distribution. A total of 21K out 50K SNPs on the chip were selected based on putative association with phenotypic traits, and hence, were expected to be clustered in specific genome loci. However, supplementing the chip with 29K SNPs (two SNPs per gene) perhaps helped in randomizing the SNP distribution in the genome. Previously, a 57K genome-wide SNP array for rainbow trout reported an average of 1,551.4 mapped SNPs per chromosome (Gonzalez-Pena et al 2016). The 57K array was designed primarily using SNPs originating from RAD-Seq sequencing of doubled-haploid clonal lines (Palti et al 2014) and whole genome re-sequencing of fish from the Aquagen (Norway) breeding program. A key point here, is that the SNPs included in the 57K chip were originated from other genetic lines. Hence, although polymorphic enough in the NCCCWA growth line used in this study for conducting GWA as we have previously shown(Gonzalez-Pena et al 2016), the SNPs used for GWA in this study were originated from the investigated

population and were expected to be more informative due to ascertainment bias (Vallejo et

al 2017b).
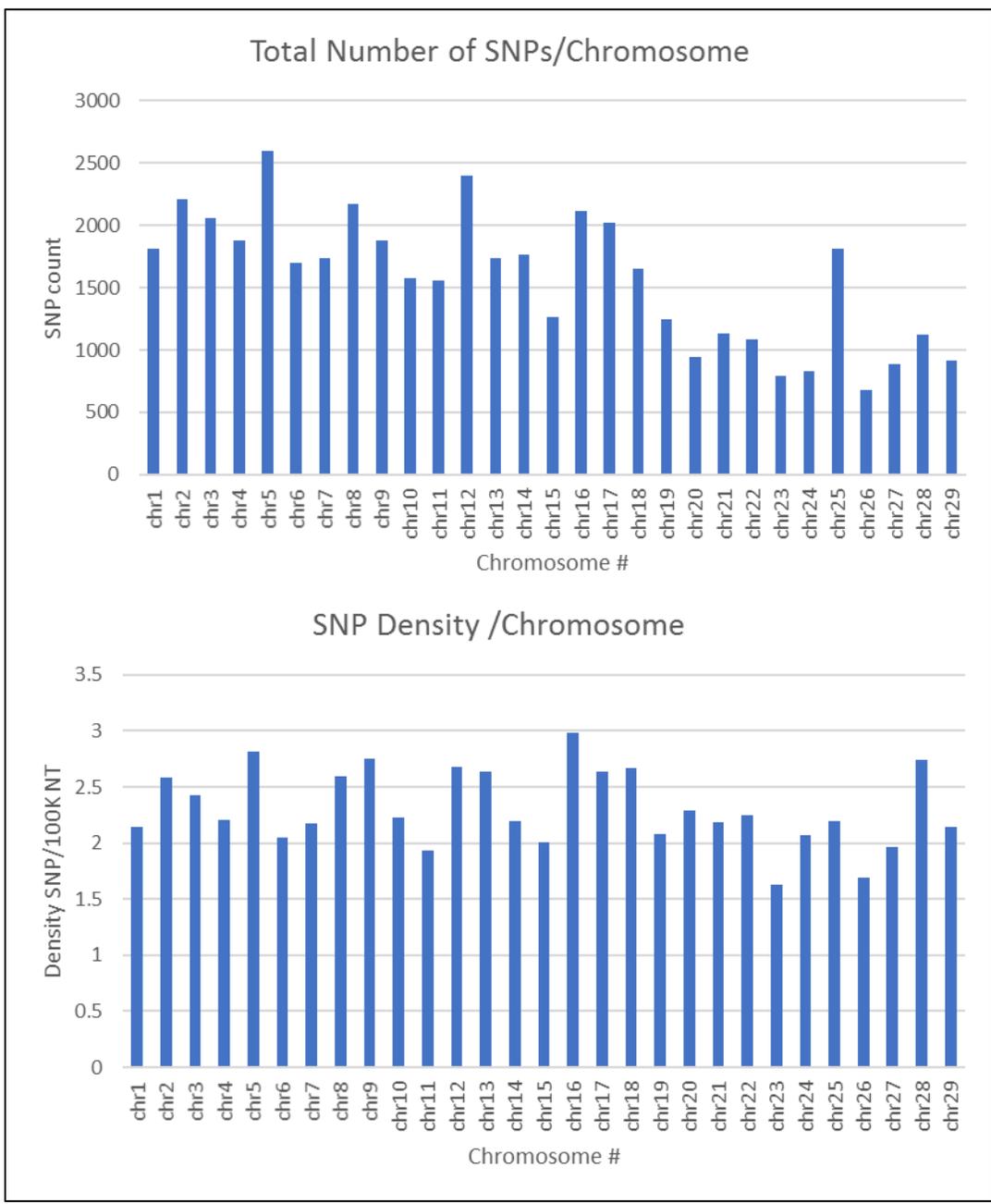
Figure 2. Number of SNPs per chromosome and SNP density distribution (SNP/100K nucleotide).

SNP annotation and classification based on functional effects

SNPeff program was used to classify and annotate functional effects of the SNPs. A total of 45,590 SNPs were included in this analysis. Classifying SNPs by impact showed 636 effects (0.23%) with high impact (stop-gain) and 20,987 effects (7.86%) with moderate impact, (missense variants). The rest (91.9%) represents low to moderate variant effect including synonymous and non-coding SNPs. Figure 3 shows percent of SNP effects by gene regions. A total of 32.8% of the effects were within transcripts with 16.5% exonic, 1.3% in the 5'-UTR and 12.8% in 3'-UTR. All SNPs on the chip were identified through transcriptome sequencing. Surprisingly, there were 14% upstream and 18.1% downstream effects (within 5 Kb of the genes). The upstream/downstream percent is consistent with our previous report that showed 17.1-20.2% SNPs within 5 Kb upstream/downstream of protein-coding genes in one of two SNP data sets used in building the SNP-chip (Al-Tobasei et al 2017a). On the other hand, there was only 1.9% of the SNP effects within intergenic regions, compared to 37.7-49.2% intergenic SNPs in the previous study (Al-Tobasei et al 2017a). In our previous study, the high percentage of intergenic and upstream/downstream SNPs was explained by the incomplete annotation of protein-coding genes and exons used in the previous version of the rainbow trout reference genome (Berthelot et al 2014). The drop in the percentage of intergenic SNP effects in this study may be due to the improved gene annotation of the current version of the genome reference.
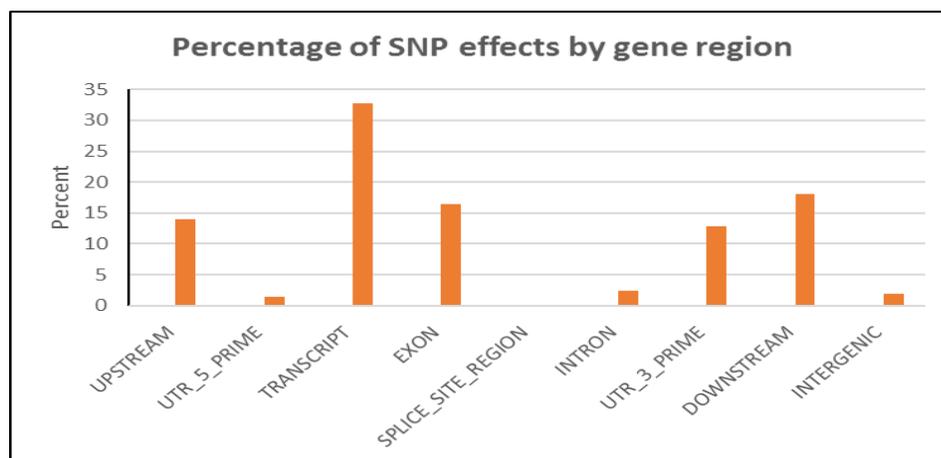
Figure 3. Percentage of SNP effects by gene region.

GWA Analysis

Genomic regions associated with muscle yield

GWA analysis using WssGBLUP identified 163 SNPs, each explaining at least 2% of the genetic variance of muscle yield (Figure 4, Tables 3 and 4; Supplementary file 1). The study identified several QTL regions explaining together up to 28.40% of the additive genetic variance for muscle yield in this rainbow trout population (Appendix A). The SNPs were clustered into 4 main chromosomes (14, 16, 9, and 17). Chromosomes 14 and 16 showed the highest peaks with genomic loci explaining up to 12.71% and 10.49% of the genetic variance, respectively. The total variance explained by these loci is 23.2%. Figure 4 shows a Manhattan plot displaying association between SNP genomic sliding window of 50 SNPs and muscle yield. Sixty-nine of the 163 SNPs (42.2%) were previously identified as SNPs with allelic imbalances associated with muscle yield in the original SNP data set used to build the SNP chip (Al-Tobasei et al 2017a). Twenty-one of the 163 SNPs caused nonsynonymous mutations. The rest of the SNPs were either silent mutations or located in UTR of the genes indicating their potential epigenetic mechanism of gene regulation.

Important SNPs with more than 5% genetic variance are discussed below and all 163 SNPs are listed in Supplementary file 1.
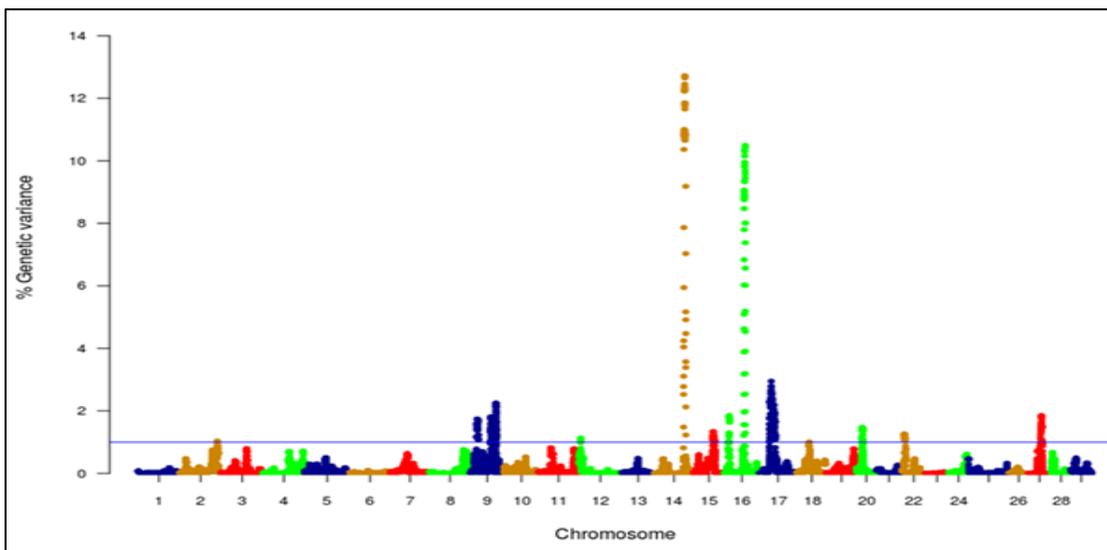
Figure 4. Manhattan plot of GWA analysis performed with WssGBLUP and showing association between SNP genomic sliding windows of 50 SNPs and muscle yield. Chromosomes 14 and 16 showed the highest peaks with genomic loci, explaining together up to 23.2% of the genetic variance. The blue line shows a threshold of 1% of additive genetic variance explained by SNPs.

With 46 SNPs clustered into 23 annotated genes, chromosome 14 had the most significant QTL windows explaining up to 12.71% of the genetic variance in muscle yield (Table 3 and Figure 4). At least four genes can be inferred to be involved in cell differentiation/proliferation and regulation of gene expression based on their RefSeq annotation. The list included fibroblast growth factor-binding protein-1(FGFBP1) which had a single nonsynonymous SNP found in a window that explained 12.24% of the additive

genetic variance. FGFBP1 plays an essential role in cell proliferation and differentiation by binding to fibroblast growth factors. The FGFBP1 expression increases during development and decreases before neuromuscular junction degeneration during aging (Taetzsch et al 2017). The list of genes on chromosome 14 also includes inositol polyphosphate 5-phosphatase (OCRL-1). OCRL is involved in terminating the PI3K signaling and thus plays an important role in modulating effects of growth factors and insulin stimulation in cell proliferation and survival (Ooms et al 2009). Prominin-1-A gene (PROM1) that encodes for a transmembrane glycoprotein had 2 SNPs. PROM1, often used as adult stem cell marker, plays a role in maintaining stem cell properties by suppressing differentiation (GeneCards-Human-Gene-Database 2018a). Another gene on chromosome 14 was farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha (FNTA) which had a SNP explaining 12.36% of the variance. FNTA may positively regulate neuromuscular junction development (UniProtKB 2018d).

In addition, chromosome 14 had three genes involved in the cell cycle regulation. The first gene is MCTS1 re-initiation and release factor that had two SNPs in a window explaining 12.65% of the additive genetic variance. MCTS1 is anti-oncogene that decreases cell doubling time by shortening the G1 and G1/S transit time (GeneCards-Human-Gene-Database 2018b). The second cell cycle control gene was cyclin-A2 which promotes transition through G1/S and G2/M and can block muscle-specific gene expression during muscle differentiation (X et al 1996). The third gene was glutathione S-transferase P (GSTP1). Although involved in numerous biological functions, GSTP1 negatively regulates CDK5 activity via p25/p35 translocation which diminishes neurodegeneration (Sun et al 2011).

Table 3. Selected SNP markers explaining the largest proportion of genetic variance (>5%) for muscle yield in chromosome 14 using 50 adjacent SNP windows. Color intensities reflect changes in additive genetic variance (green is the highest and red is the lowest).

Variance %	CHR	SNP position	Distance to next SNP	Strand	Gene	Annotation	Region/Effect
5.95	14	60291342	16113	+	etfdh	Electron Transfer Flavoprotein Dehydrogenase	CDS/nonSyn
7.86	14	60307455	366	+	etfdh	Electron Transfer Flavoprotein Dehydrogenase	CDS/syn
10.36	14	60307821	8	+	etfdh	Electron Transfer Flavoprotein Dehydrogenase	3'UTR
10.79	14	60307829	2256	+	etfdh	Electron Transfer Flavoprotein Dehydrogenase	3'UTR
10.84	14	60310085	163538	-	ppid	Peptidylprolyl Isomerase D	CDS/nonSyn
10.90	14	60473623	421210	+	rapgef2	Rap Guanine Nucleotide Exchange Factor 2	3'UTR
10.96	14	60894833	295302	NA	NA	NA	NA
11.00	14	61190135	558	+	LOC110488945	Prominin-1-A	3'UTR
11.00	14	61190693	7552	+	LOC110488945	Prominin-1-A	3'UTR
12.24	14	61198245	76178	+	LOC110488947	Fibroblast Growth Factor-Binding Protein 1	CDS/nonSyn
12.23	14	61274423	13691	-	LOC110488948	Cyclin-A2	3'UTR
12.29	14	61288114	762	-	LOC110488950	Transmembrane Protein 33	3'UTR
12.35	14	61288876	528124	-	LOC110488950	Transmembrane Protein 33	CDS/syn
12.36	14	61817000	18067	-	LOC110488956	Protein Farnesyltransferase/Geranylgeranyltransferase Type-1 Subunit Alpha	3'UTR
12.30	14	61835067	6866	-	LOC110488957	Glutathione S-Transferase P	3'UTR
12.26	14	61841933	319532	-	LOC110488957	Glutathione S-Transferase P	CDS/syn
12.24	14	62161465	1101	NA	NA	NA	NA
12.45	14	62162566	79441	NA	NA	NA	NA
12.71	14	62242007	38699	-	LOC110488962	Inositol Polyphosphate 5-Phosphatase Ocrl-1	CDS/nonSyn
12.71	14	62280706	12616	+	LOC110488963	Chloride Intracellular Channel Protein 2	3'UTR
12.70	14	62293322	4394	+	LOC110488964	C1Galt1-Specific Chaperone 1	3'UTR
12.65	14	62297716	9021	-	mcts1	Mcts1, Re-Initiation And Release Factor	3'UTR
11.85	14	62306737	36808	-	mcts1	Mcts1, Re-Initiation And Release Factor	5'UTR
11.84	14	62343545	586	+	lamp2	Lysosomal Associated Membrane Protein 2	CDS/nonSyn
11.85	14	62344131	2211	+	lamp2	Lysosomal Associated Membrane Protein 2	CDS/nonSyn
11.78	14	62346342	306	+	lamp2	Lysosomal Associated Membrane Protein 2	Intronic
11.78	14	62346648	579	+	lamp2	Lysosomal Associated Membrane Protein 2	Intronic
11.77	14	62347227	29198	+	lamp2	Lysosomal Associated Membrane Protein 2	Intronic
11.66	14	62376425	304	+	tmem255a	Transmembrane Protein 255A	CDS/syn
10.92	14	62376729	3620	+	tmem255a	Transmembrane Protein 255A	CDS/syn
10.87	14	62380349	282	+	tmem255a	Transmembrane Protein 255A	3'UTR
10.86	14	62380631	31094	+	tmem255a	Transmembrane Protein 255A	3'UTR
10.86	14	62411725	1632	+	upf3b	Upf3B, Regulator Of Nonsense Mediated Mrna Decay	CDS/nonSyn
10.86	14	62413357	1931	+	upf3b	Upf3B, Regulator Of Nonsense Mediated Mrna Decay	3'UTR
10.92	14	62415288	26359	+	LOC110488974	60S Ribosomal Protein L39	3'UTR
10.90	14	62441647	10087	+	LOC110488975	Septin-6	CDS/syn
10.88	14	62451734	10231	+	LOC110488975	Septin-6	CDS/syn
10.88	14	62461965	6983	+	LOC110488975	Septin-6	3'UTR
10.75	14	62468948	89647	NA	NA	NA	NA
10.72	14	62558595	7052	+	LOC110488979	Ets-Related Transcription Factor Elf-1	3'UTR
10.66	14	62565647	66310	-	LOC110488980	Tenomodulin	3'UTR
10.67	14	62631957	1503911	-	LOC110488980	Tenomodulin	CDS/nonSyn
10.83	14	64135868	6948	+	gla	Galactosidase Alpha	CDS/nonSyn
9.18	14	64142816	2581	-	LOC110488986	60S Ribosomal Protein L36A	CDS/syn
7.03	14	64145397	20716	-	LOC110488986	60S Ribosomal Protein L36A	CDS/nonSyn
5.17	14	64166113		+	btk	Bruton Tyrosine Kinase	CDS/nonSyn

Chromosome 14 also had SNPs in genes playing important mitochondrial functions. There were 4 SNPs in the gene encoded for the electron transfer flavoprotein dehydrogenase (ETFDH) which is an important enzyme in the mitochondrial electron transport chain. Mutations in ETFDH are associated with myopathies(Haller and DiMauro 2012). Another mitochondrial-relevant gene was peptidylprolyl isomerase D (PPID). Mutations in PPID are associated with muscular dystrophy in human (Giorgio et al 2010).

Few other genes included in the QTL region on chromosome 14 are important for maintenance of the muscle functions. Of them is the chloride intracellular channel protein 2 (CLIC2) which modulates the activity of ryanodine receptor 2 (RYR2) and inhibits calcium influx, and therefore is involved in regulating muscle contraction (Takano et al 2012). Five SNPs were in the lysosomal-associated membrane protein 2 gene (LAMP2). LAMP2 mutations were reported in patients with cardioskeletal myopathies (Sugie et al 2018). Two SNPs were located in the UPF3B gene, a regulator of nonsense-mediated mRNA decay (NMD). NMD inhibition was observed in patients with muscular dystrophy (Miller and Pearce 2014). Three SNPs were observed in the septin-6 gene. Mutations of septin-9 (another gene family member) is genetically linked to muscle atrophy (Dolat et al 2014). Two SNPs were identified in the tenomodulin gene which showed differential expression in an animal muscle atrophy model (Furlow et al 2013).

Chromosome 16 ranked second in having the most significant QTL windows with 49 SNPs clustered into 16 annotated genes (Table 4 and Figure 4). The gene within the most significant SNP window to additive genetic variance was the cysteine/serine-rich nuclear protein 2 (CSRNP2). CSRNP2 has DNA binding transcription factor/activation activity. Deletion of CSRNP1/2/3 three gene family members resulted in mice neonatal lethality

(Gingras et al 2007). Another gene within the same SNP window was solute carrier family 26 member 9 (Slc26a9). Little is known about the function of Slc26a9 in muscle, it serves as anion exchanger mediating chloride, sulfate and oxalate transport and chloride/bicarbonate exchange (UniProtKB 2018a). A single SNP was observed in the stem cell marker CD34a gene. Cd34(-/-) mice showed a defect in muscle regeneration caused by acute or chronic muscle injury (Alfaro et al 2011).

Several genes were involved in cell signaling/receptor activity. Five SNPs were predicted in 2 genes of the immune-related complement activation pathway, these are the complement receptor and C4b-binding protein alpha chain. Recent studies indicated that the complement is activated as a response of skeletal muscle injury and plays a key role during muscle regeneration(Zhang et al 2017). A single SNP was identified in the tyrosine-protein phosphatase non-receptor type 12 (PTPN12) which dephosphorylates a wide-range of proteins, and thus regulates several cellular signaling cascades such as ERBB2 and PTK2B/PYK2 (UniProtKB 2018b). This group of genes also includes the membrane-associated guanylate kinase, WW and PDZ domain-containing protein 3 (MAGI3), which is involved in the regulation of various cell signaling processes including the AKT1, TGFA, ERK and JNK signaling cascades (UniProtKB 2018c). Two SNPs were in the basement membrane-specific heparan sulfate proteoglycan core protein (HSPG2). A mouse model deficient in this gene showed muscle hypertrophy through reduced myostatin expression suggesting a role in maintaining fast muscle mass and fiber composition (Xu et al 2010). Five SNPs were in the TNF receptor superfamily member 5A gene. Recently, some proinflammatory cytokines belonging to TNF superfamily have been recognized as an important regulator of skeletal muscle mass (Tajrishi et al 2014).

Chromosome 16 also had a single SNP in the DAZ-associated protein 2 (DAZAP2). Not much is known about the DAZAP2 function in muscle, however, DAZAP2 interacts with the transforming growth factor-beta signaling molecule SARA (Smad anchor for receptor activation), eukaryotic initiation factor 4G, and an E3 ubiquitinase ubiquitinase (Sayers et al 2019). Another gene in the list was Rac GTPase-activating protein 1 (RACGAP1) that harbored 2 SNPs explaining up to 9.658% of the genetic variance. RACGAP1 regulates cytokinesis and cell differentiation (Wang et al 2018). A single SNP existed in caspase-9 which has an important non-apoptotic role in muscle differentiation (Murray et al 2008). Three SNPs were located in the kelch protein 21. Several Kelch family members play important roles in skeletal muscle development by regulating the cell proliferation and/or differentiation (Gupta and Beggs 2014).

An important gene affecting muscle function which is also located within the QTL region on chromosome 16 is the citrate synthase (CS), which is used as a marker for human mitochondrial functions. Ten SNPs explaining up to 8% of the genetic variance were located in the CS gene.

GWA studies in fish to identify QTL affecting muscle yield and quality are still in its infancy. Previous GWA analysis using a 57K genomic SNP chip on the same fish population identified two windows that explained 1.5% and 1.0% of the additive genetic variance for muscle yield and 1.2% and 1.1% for muscle weight. Interestingly, the windows are located on chromosome 9, which showed some association with muscle yield in the current study; however, none of the SNPs were annotated to the same genes. No major QTLs were identified in the previous study. This large difference in the outcomes of the two studies was somewhat unexpected. However, it may be explained by lower marker density within or near genes in the 57K chip (Gonzalez-Pena et al 2016) and by ascertainment bias, because the transcribed SNPs used in this study were discovered in the phenotyped fish and hence are expected to be more polymorphic and informative for GWA analysis in this population. Additionally, in this study, sliding windows of 50 SNP were used contrasting with 20 non-sliding windows in the previous study. Difference in window size slightly contributed to the increased proportion of variance (data not shown). By using SNP windows, it is assumed that those DNA blocks may be inherited together, which may not always be the case for all assumed windows. In common carp, genetic linkage mapping identified QTLs with large effects for muscle fiber cross-section area (21.9%) and muscle fiber density (18.9%) (Zhang et al 2011). Genome-wide significant QTL affecting growth and muscle related traits were identified in Atlantic salmon (Tsai et al 2015). The latter two studies, together with our study, indicate existence of large-effect QTLs affecting muscle yield in aquaculture species. However, the QTLs identified in this study might be population specific and thus, need to be tested in other populations.

Citrate synthase activity correlation with muscle yield

A SNP window on chromosomes 16 explaining up to 8.01% of the genetic variance in muscle yield contained ten SNPs of the CS gene. Two of the SNPs were nonsynonymous mutations. To investigate the potential effect of these SNPs, we measured the CS activity in 100 fish from the 2012 year-class. The samples included 38 fish from 5 high-ranked and 5 low-ranked families for muscle yield, 19 each, and 62 randomly selected fish. CS had 1.43-fold increase in the high-ranked fish compared to the low ranked ones (figure 5). The regression coefficient R^2 value between the muscle yield and CS activity was 0.092 (p-value 0.002). However, there was no significant association between any of the SNP genotypes and CS activity (P-value <0.001). Mitochondria are at the center of age-related sarcopenia that is characterized by decline in human muscle mass. Skeletal muscle CS decreases with aging in humans (Short et al 2005). Therefore, our results suggest an important role of mitochondrial functions to muscle growth.
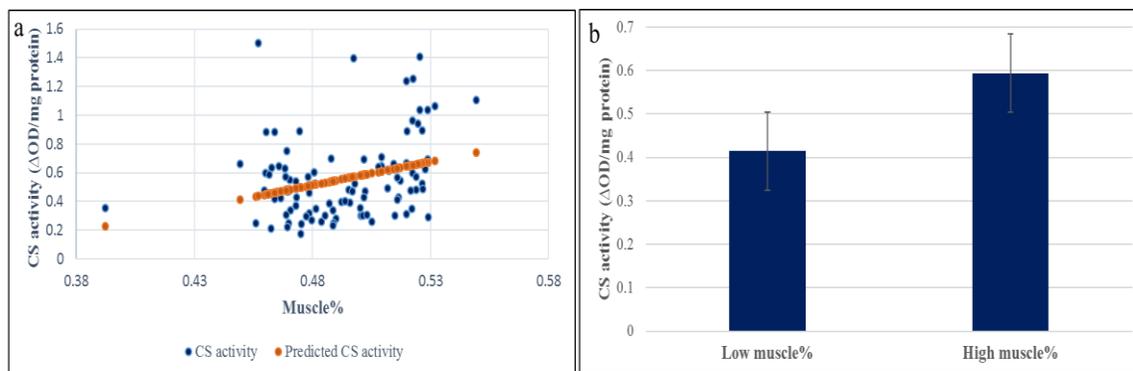
Figure 5. Correlation coefficient between muscle yield and CS activity in 96 samples. A) The regression coefficient R2 value between the muscle yield and CS activity was 0.092 (p-value 0.002). b) CS had 1.43-fold increase in the high-ranked fish compared to the low ranked ones.

CONCLUSIONS

This study provides a 50K transcribed gene SNP-chip based on RNA-Seq data from fish families showing genetic diversity for six aquaculture production traits in the USDA/NCCCWA growth- and disease-selected genetic lines. The chip was tested for GWA analysis, which led to identification of large-effect QTL for muscle yield in that population. Other muscle quality traits are currently under investigation. Collectively, these studies will allow the use of SNP markers to estimate breeding values for muscle yield and quality traits that are economically important traits for aquatic food producers, processors, and consumers. Current and future selection at the NCCCWA will select for improved fillet yield. Genetic markers are desirable for these traits because genetic improvement is limited by the inability to measure fillet yield traits directly on broodstock due to lethal sampling. Hence the accuracy and efficiency of selective breeding can be improved by taking advantage of the genomic information, even though limited phenotyping is available for this economically-important trait.

One potential limitation in this study is the use of the same population for SNP identification and GWAS. The QTLs identified in this study might be population specific and thus, need to be tested in other populations. It is worth mentioning that while the SNP chip has 50K SNPs, about 7.9K SNPs had putative allelic imbalances associated with 5 growth and muscle related traits (body weight, muscle yield, muscle fat content, shear force, and whiteness). Also, there were 13K additional SNPs with putative allelic imbalances associated with resistance BCWD. About 620 fish were used in the previous RNA-Seq analyses to identify these putative SNPs (Al-Tobasei et al 2017a, Al-Tobasei et al 2017b). In this study, only one of the 6 traits, muscle yield was considered for GWAS.

Only 40 fish were used in the previous RNA-Seq study to identify the putative SNPs that were associated with muscle yield (Al-Tobasei et al 2017a). To make sure those fish do not affect the GWAS results, we removed those 40 fish in addition to all fish involved in determining the growth/muscle putative SNPs (a total of 90 fish) and we reran GWAS. There was no significant change to the QTL identified in this study.   Also, all the 90 fish came from hatch year, 2010. In this GWAS, we used fish from 2010 hatch year (406 fish representing 98 families) and 2012 hatch year (472 fish representing 99 families).

Author Contributions

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

REFERENCES

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* **93:** 743-752.

Al-Tobasei R, Paneru B, Salem M (2016). Genome-Wide Discovery of Long Non-Coding RNAs in Rainbow Trout. *PLoS One* **11:** e0148940.

Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B *et al* (2017a). Identification of SNPs associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-Seq samples in rainbow trout. *BMC Genomics* **18:** 582.

Al-Tobasei R, Palti Y, Wiens G, Salem M: Identification of SNPs With Allelic Imbalances in Rainbow Trout Genetic Lines Showing Different Susceptibility to Infection With Flavobacterium Psychrophilum. *PAG-XXV Plant & Animal Genomes Conference*; January 14-18; San Diego, California. 2017b.

Alfaro LA, Dick SA, Siegel AL, Anonuevo AS, McNagny KM, Megeney LA *et al* (2011). CD34 promotes satellite cell motility and entry into proliferation to facilitate efficient skeletal muscle regeneration. *Stem cells (Dayton, Ohio)* **29:** 2030-2041.

Berg P, Nielsen J, Sørensen MK: EVA: Realized and predicted optimal genetic contributions. *Proc. 8th World Cong. Genet. Appl. Livest. Prod.*; Belo Horizonte, Brazil. 2006.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B *et al* (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* **5:** 3657.

Boldman KG, Kriese LA, Van Vleck LD, Van Tassell CP, Kachman SD (1995). A manual for the use of MTDFREML. A set of programs to obtain estimates of variances and covariances [Draft]: Clay Center, NE: USDA-ARS.

Brijs J, Sandblom E, Sundh H, Grans A, Hinchcliffe J, Ekstrom A *et al* (2017). Increased mitochondrial coupling and anaerobic capacity minimizes aerobic costs of trout in the sea. *Sci Rep* **7:** 45778.

Brookes AJ (2007). Single Nucleotide Polymorphism (SNP). In: John Wiley & Sons Lwen, http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0005006.pub2/pdf, accessed April 4, 2012 (ed). *ENCYCLOPEDIA OF LIFE SCIENCES (els)*.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L *et al* (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6:** 80-92.

Dolat L, Hu Q, Spiliotis ET (2014). Septin functions in organ system physiology and pathology. *Biol Chem* **395:** 123-141.

Ekstrom A, Sandblom E, Blier PU, Dupont Cyr BA, Brijs J, Pichaud N (2017). Thermal sensitivity and phenotypic plasticity of cardiac mitochondrial metabolism in European perch, Perca fluviatilis. *J Exp Biol* **220:** 386-396.

Furlow JD, Watson ML, Waddell DS, Neff ES, Baehr LM, Ross AP *et al* (2013). Altered gene expression patterns in muscle ring finger 1 null mice during denervation- and dexamethasone-induced muscle atrophy. *Physiol Genomics* **45:** 1168-1185.

Gao G, Nome T, Pearse DE, Moen T, Naish KA, Thorgaard GH *et al* (2018). A New Single Nucleotide Polymorphism Database for Rainbow Trout Generated Through Whole Genome Resequencing. *Front Genet* **9:** 147.

GeneCards-Human-Gene-Database (2018a). PROM1 Gene - GeneCards | PROM1 Protein | PROM1 Antibody.

GeneCards-Human-Gene-Database (2018b). MCTS1 Gene - GeneCards | MCTS1 Protein | MCTS1 Antibody.

Gerrard DE, Grant AL (2003). *Principles of Animal Growth and Development*. Kendall/Hunt Publishing Co.

Gingras S, Pelletier S, Boyd K, Ihle JN (2007). Characterization of a family of novel cysteine- serine-rich nuclear proteins (CSRNP). *PLoS One* **2:** e808.

Giorgio V, Soriano ME, Basso E, Bisetto E, Lippe G, Forte MA *et al* (2010). Cyclophilin D in mitochondrial pathophysiology. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1797:** 1113-1118.

Gjerde B: State of the art in selective breeding of aquaculture species. In: BARD Workshop: Aquaculture Genetics - status and prospects. February 20-23; Eilat, Israel. 2006.

Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB *et al* (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (Oncorhynchus mykiss). *Front Genet* **7:** 203.

Gupta VA, Beggs AH (2014). Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skeletal muscle* **4:** 11.

Haller RG, DiMauro S (2012). Chapter 75 - Metabolic and Mitochondrial Myopathies. In: Hill JA, Olson EN (eds). *Muscle*. Academic Press: Boston/Waltham. pp 1031-1041.

Houston RD, Gheyas A, Hamilton A, Guy DR, Tinch AE, Taggart JB *et al* (2008). Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (Salmo salar). *Dev Biol (Basel)* **132:** 199-204.

Leeds TD, Vallejo RL, Weber GM, Pena DG, Silverstein JS (2016). Response to five generations of selection for growth performance traits in rainbow trout (Oncorhynchus mykiss). *Aquaculture* **465:** 341-351.

Legarra A, Aguilar I, Misztal I (2009). A relationship matrix including full pedigree and genomic information. *J Dairy Sci* **92:** 4656-4663.

Marancik D, Gao G, Paneru B, Ma H, Hernandez AG, Salem M *et al* (2014). Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with Flavobacterium psychrophilum. *Front Genet* **5:** 453.

Miller JN, Pearce DA (2014). Nonsense-mediated decay in genetic disease: friend or foe? *Mutat Res Rev Mutat Res* **762:** 52-64.

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A *et al* (2014). Manual for BLUPF90 family of programs. University of Georgia: University of Georgia, Athens, USA.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., H. LD (2002). BLUPF90 and related programs (BGF90) [WWW Document], in Proceeding of 7th World Congress on Genetics Applied to Livestock Production (Montpellier).

Moen T, Baranski M, Sonesson AK, Kjoglum S (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. *BMC Genomics* **10:** 368.

Murray TVA, McMahon JM, Howley BA, Stanley A, Ritter T, Mohr A *et al* (2008). A non-apoptotic role for caspase-9 in muscle differentiation.

Ooms LM, Horan KA, Rahman P, Seaton G, Gurung R, Kethesparan DS *et al* (2009). The role of the inositol polyphosphate 5-phosphatases in cellular function and human disease. *Biochem J* **419:** 29-49.

Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E *et al* (2014). A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Molecular ecology resources* **14:** 588-596.

Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR *et al* (2015). The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol Ecol Resour* **15:** 662-672.

Quinlan AR, Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841-842.

Rexroad CE, Palti Y, Gahr SA, Vallejo RL (2008). A second generation genetic map for rainbow trout (Oncorhynchus mykiss). *BMC Genet* **9:** 74.

Salem M, Kenney PB, Rexroad CE, Yao J (2006). Molecular characterization of muscle atrophy and proteolysis associated with spawning in rainbow trout. *Comp Biochem Physiol Part D Genomics Proteomics* **1:** 227-237.

Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A *et al* (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS One* **7:** e36264.

Salem M, Manor ML, Aussanasuwannakul A, Kenney PB, Weber GM, Yao J (2013). Effect of sexual maturation on muscle gene expression of rainbow trout: RNA-Seq approach. *Physiol Rep* **1:** e00120.

Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K *et al* (2019). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **47:** D23-D28.

Seite S, Mourier A, Camougrand N, Salin B, Figueiredo-Silva AC, Fontagne-Dicharry S *et al* (2018). Dietary methionine deficiency affects oxidative status, mitochondrial integrity and mitophagy in the liver of rainbow trout (Oncorhynchus mykiss). *Sci Rep* **8:** 10151.

Short KR, Bigelow ML, Kahl J, Singh R, Coenen-Schimke J, Raghavakaimal S *et al* (2005). Decline in skeletal muscle mitochondrial function with aging in humans. *Proc Natl Acad Sci U S A* **102:** 5618-5623.

Sugie K, Komaki H, Eura N, Shiota T, Onoue K, Tsukaguchi H *et al* (2018). A Nationwide Survey on Danon Disease in Japan. *Int J Mol Sci* **19**.

Sun KH, Chang KH, Clawson S, Ghosh S, Mirzaei H, Regnier F *et al* (2011). Glutathione-S-transferase P1 is a critical regulator of Cdk5 kinase activity. *Journal of neurochemistry* **118:** 902-914.

Taetzsch T, Tenga MJ, Valdez G (2017). Muscle Fibers Secrete FGFBP1 to Slow Degeneration of Neuromuscular Synapses during Aging and Progression of ALS. *J Neurosci* **37:** 70-82.

Tajrishi MM, Sato S, Shin J, Zheng TS, Burkly LC, Kumar A (2014). The TWEAK-Fn14 dyad is involved in age-associated pathological changes in skeletal muscle. *Biochem Biophys Res Commun* **446:** 1219-1224.

Takano K, Liu D, Tarpey P, Gallant E, Lam A, Witham S *et al* (2012). An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity. *Hum Mol Genet* **21:** 4497-4507.

Taylor JF, Taylor KH, Decker JE (2016). Holsteins are the genomic selection poster cows. *Proc Natl Acad Sci U S A* **113:** 7690-7692.

Tsai HY, Hamilton A, Guy DR, Tinch AE, Bishop SC, Houston RD (2015). The genetic architecture of growth and fillet traits in farmed Atlantic salmon (Salmo salar). *BMC Genet* **16:** 51.

Turner SD (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*.

UniProtKB (2018a). Slc26a9 - Solute carrier family 26 member 9 - Mus musculus (Mouse) - Slc26a9 gene & protein.

UniProtKB (2018b). PTPN12 - Tyrosine-protein phosphatase non-receptor type 12 - Homo sapiens (Human) - PTPN12 gene & protein.

UniProtKB (2018c). MAGI3 - Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 3 - Homo sapiens (Human) - MAGI3 gene & protein.

UniProtKB (2018d). Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha.

Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP *et al* (2017a). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet Sel Evol* **49:** 17.

Vallejo RL, Liu S, Gao G, Fragomeni BO, Hernandez AG, Leeds TD *et al* (2017b). Similar Genetic Architecture with Shared and Unique Quantitative Trait Loci for Bacterial Cold

Water Disease Resistance in Two Rainbow Trout Breeding Populations. *Front Genet* **8:** 156.

Wang C, Wang W, Liu Y, Yong M, Yang Y, Zhou H (2018). Rac GTPase activating protein 1 promotes oncogenic progression of epithelial ovarian cancer. *Cancer Sci* **109:** 84-93.

Wang H, Misztal I, Aguilar I, Legarra A, Muir WM (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)* **94:** 73-83.

Wang H, Misztal I, Aguilar I, Legarra A, Fernando RL, Vitezica Z *et al* (2014). Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front Genet* **5:** 134.

Wang S, Sha Z, Sonstegard TS, Liu H, Xu P, Somridhivej B *et al* (2008). Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics* **9:** 450.

WorldFish Center (2009). Climate Change: Research to Meet the Challenges Facing Fisheries and Aquaculture. In: Issues Brief 1915.

X SS, J R, Kim PS, Novitch BG, Lassar AB (1996). Cyclin-mediated inhibition of muscle gene expression via a mechanism that is independent of pRB hyperphosphorylation.

Xu Z, Ichikawa N, Kosaki K, Yamada Y, Sasaki T, Sakai LY *et al* (2010). Perlecan deficiency causes muscle hypertrophy, a decrease in myostatin expression, and changes in muscle fiber composition. *Matrix Biology* **29:** 461-470.

Zhang C, Wang C, Li Y, Miwa T, Liu C, Cui W *et al* (2017). Complement C3a signaling facilitates skeletal muscle regeneration by regulating monocyte function and trafficking. *Nat Commun* **8:** 2078.

Zhang Y, Xu P, Lu C, Kuang Y, Zhang X, Cao D *et al* (2011). Genetic linkage mapping and analysis of muscle fiber-related QTLs in common carp (Cyprinus carpio L.). *Mar Biotechnol (NY)* **13:** 376-392.

APPENDICES

Variance %	RefSeq	CHR	SNP position	ProbeID	AffID	Strand	Gene	Annotation	Region	Effect
2.25	NC_035085.1	9	51756019	AX-171615940	Affx-277346584	+	LOC110532512	Pancreatic Progenitor Cell Differentiation And Proliferation Factor	CDS	Nonsynonymous
12.71	NC_035090.1	14	62242007	AX-171630283	Affx-277322554	-	LOC110488962	Inositol Polyphosphate 5-Phosphatase Ocrl-1	CDS	Nonsynonymous
10.49	NC_035092.1	16	42192751	AX-171635060	Affx-277344473	-	LOC110492098	Cysteine/Serine-Rich Nuclear Protein 2	CDS	Synonymous
2.95	NC_035093.1	17	24478297	AX-172561273	Affx-277318283	-	LOC110493874	Charged Multivesicular Body Protein 4B	CDS	Nonsynonymous

Appendix A: SNPs with the highest genetic variance for muscle yield in each chromosome.

CHAPTER III

GENOME-WIDE ASSOCIATION STUDY IDENTIFIES GENOMIC LOCI AFFECTING FILLET FIRMNESS AND PROTEIN CONTENT IN RAINBOW TROUT

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. Frontiers in Genetics 2019, 10(386)

ABSTRACT

Fillet quality traits determine consumer satisfaction and affect profitability of the aquaculture industry. Soft flesh is a criterion for fish fillet downgrades, resulting in loss of value. Fillet firmness is influenced by many factors, including rate of protein turnover. A 50K transcribed gene SNP chip was used to genotype 789 rainbow trout, from two consecutive generations, produced in the USDA/NCCCWA selective breeding program. Weighted single-step GBLUP (WssGBLUP) was used to perform genome-wide association (GWA) analyses to identify quantitative trait loci affecting fillet firmness and protein content. Applying genomic sliding windows of 50 adjacent SNPs, 212 and 225 SNPs were associated with genetic variation in fillet shear force and protein content, respectively. Four common SNPs in the ryanodine receptor 3 gene (RYR3) affected the aforementioned fillet traits; this association suggests common mechanisms underlying fillet shear force and protein content. Genes harboring SNPs were mostly involved in calcium homeostasis, proteolytic activities, transcriptional regulation, chromatin remodeling, and apoptotic processes. RYR3 harbored the highest number of SNPs (n =

32) affecting genetic variation in shear force (2.29%) and protein content (4.97%). Additionally, based on single-marker analysis, a SNP in RYR3 ranked at the top of all SNPs associated with variation in shear force. Our data suggest a role for RYR3 in muscle firmness that may be considered for genomic- and marker-assisted selection in breeding programs of rainbow trout.

INTRODUCTION

Aquaculture continues to experience rapid growth worldwide. However, for a sustainable industry, there is a need to produce fish fillets with consistent quality and high value. Consumer attitude towards fish is influenced by nutritional and sensory attributes, including fillet firmness (Bonneau and Lebret 2010). Firmness is one of the most important quality attributes that determines consumer satisfaction toward the product; and, it is affected by many intrinsic and extrinsic factors (Destefanis et al 2008). These factors include prerigor muscle processing, production and storage temperature, chilling protocols, genotype, handling stress, collagen content, extent of proteolysis, and the proximate composition of muscle (Bahuaud et al 2010, Castañeda et al 2005, Grześ et al 2017). Fillet softness shares common causes but should not be confused with gaping that results from tearing the connective tissue between muscle layers and weakening of the interface between the myotome and the myosepta causing slits in the fillet (Jacobsen et al 2017). Previous studies in farmed European whitefish showed that fillet firmness is a heritable trait (0.30 ± 0.09); whereas, gaping seems to be not heritable (Kause et al 2011). Gaping is affected by a range of perimortem harvest and handling factors and postmortem handling practices. In other words, there is a great opportunity for uncontrolled, random variation that makes elucidation of the genetic control of gaping a challenge. Loss of fillet firmness and gaping

contribute to downgrading during the secondary processing of fillet causing economic loss for the industry (Jacobsen et al 2017, Torgersen et al 2014). The increased level of stress has been reported as a major cause of gaping and fillet softness (Jacobsen et al 2017). In pigs, heat stress leads to development of pale, soft, exudative (PSE) meat (Strasburg and Chiang 2009) that is associated with defective Ca^{2+} regulation. Despite a well-developed understanding of meat tenderization that has been studied for decades in mammals, the need exists for genetic markers of the fish "gaping" and fillet softness phenotypes (Ouali et al 2013).

Connective tissue, muscle fiber density, muscle fiber type, postmortem metabolism, and postmortem autolysis are inherent factors affecting muscle texture. Proteolytic degradation of connective tissue, myofibrils, extracellular matrix, and cell membrane constituents contribute to post-mortem softening (Torgersen et al 2014). Protein content is relatively constant in fish; however, it may vary due to seasonal changes and physiological factors (Belitz et al 2009, Delbarre-Ladrat et al 2006). For instance, carbohydrate content and metabolism affect postmortem changes in protein content. Glycolysis determines the rate and extent of pH decline, which affects proteolysis and water-binding ability of the tissue. In turn, proteolysis and water-binding ability influence firmness of porcine muscle (Grześ et al 2017). However, the pH decline in fish is small due to low glycogen content in the muscle (Belitz et al 2009). There is general agreement that tenderization is enzymatic in nature and may begin with the onset of apoptosis, followed by proteolysis (Ouali et al 2013). Enzymatic degradation of key structural proteins that maintain myofibril integrity leads to postmortem tenderization. Calpains, cathepsins, proteasome, and matrix metalloproteases may act in synergy, affected by pH,

sarcoplasmic calcium, osmotic pressure, and oxidative processes, to degrade the proteins (Delbarre-Ladrat et al 2006). The increased level of stress, glycogenolysis, glycolysis, and pH decline (Thomas et al 2005) in the perimortem period, is associated with increased activity of cathepsin L, which degrades collagen and leads to fillet softening. However, protein isoforms of fish may react differently than mammalian species because fillet storage temperature are much closer to temperature optimal for proteases, glycolytic enzymes, and pyruvate dehydrogenase to name a few possibilities. Firmness of salmon muscle has been previously attributed to efficient aerobic metabolism and degradation of damaged/ misfolded proteins (Torgersen et al 2014). In addition, atrophying muscle from sexually-mature rainbow trout fish showed softer muscle that that of sterile fish (Paneru et al 2018). Transcriptomic profiling of the atrophying muscle revealed differential expression of genes related to protein ubiquitination, autophagy, extracellular matrix, myofibrillar proteins, and collagen; collectively called "the rainbow trout muscle degradome" (Paneru et al 2018). Further, profiling muscle transcriptome from fish families exhibiting divergent fillet firmness, revealed a network of protein-coding and noncoding genes related to lysosomal and proteolytic activities (Ali et al 2018b, Paneru et al 2017). Understanding the underlying mechanism of fillet firmness will help evaluate the postmortem changes affecting fillet quality, and facilitate selective breeding decisions.

Traditional genetic improvement programs to determine animals with elite genetic merit have used statistical analyses of phenotypes and pedigree information (Dang et al 2014). Genetic selection has been introduced in rainbow trout to improve fillet quality (Hu et al 2013, Kause et al 2007). Selection programs for fish, including rainbow trout, focused on growth rate and fillet quality traits; however, little attention has been paid to fillet texture

(Bahuaud et al 2010). Selection on fat content improved color and fillet texture (Florence et al 2015), feed conversion ratio (FCR), and protein-retention efficiency (Kause et al 2016). Five generations of family-based selection was established at the USDA National Center of Cool and Cold Water Aquaculture (NCCCWA) yielding a genetic gain of ~10% in body weight/ generation (Leeds et al 2016). Firmness is measured postmortem, thus the trait cannot be measured directly on breeding candidates. Only family-specific estimated breeding values (EBVs) are used for breeding candidates in traditional breeding programs. Genomic selection will allow further within-family selection for the fillet firmness traits, and thus is anticipated to increase accuracy of genetic predictions and selection response. Understanding the genetic architecture of the fillet phenotypic traits and development of genetically improved strains will improve aquaculture industry profitability and consumer satisfaction (Ali et al 2018b).

Genome-wide association (GWA) analysis compares allele frequencies at candidate loci with respect to the studied trait, and takes advantage of linkage disequilibrium (LD) between SNP marker and trait loci (Schielzeth and Husby 2014). GWA analyses have been extensively used, in mammals including human, to facilitate the investigation of variants association with complex phenotypic traits and diseases (Hindorff et al 2009). A limited number of GWA analyses have been conducted in fish including Atlantic salmon (Tsai et al 2015), catfish (Geng et al 2016), orange-spotted grouper (Yu et al 2018), and rainbow trout (Gonzalez-Pena et al 2016, Salem et al 2018). The studied traits in fish included growth (Tsai et al 2015, Yu et al 2018), disease resistance (Palti et al 2015), head size (Geng et al 2016), heat stress (Jin et al 2017), low oxygen tolerance (Zhong et al 2017), and muscle yield (Gonzalez-Pena et al 2016, Salem et al 2018). In rainbow trout, GWA

analysis revealed quantitative trait loci (QTL) associated with fillet yield and disease resistance (Gonzalez-Pena et al 2016, Liu et al 2015, Palti et al 2015). No GWA studies have been conducted in fish to identify the genetic architecture of fillet firmness. However, several GWA studies in cattle and pig revealed some genetic factors affecting meat tenderness. Calpain 1 and calpastatin are among genes that harbored genetic variants associated with meat tenderness in cattle (Ramayo-Caldas et al 2016).

A 50K transcribed gene SNP chip of average 1 SNP per 42.7 Kb, was recently developed for rainbow trout. About 21K SNPs showing potential association with important traits, including fish growth, muscle yield/quality and fillet softness, were used to build the chip. In addition, 29K SNPs were added to the chip following a strategy of 2 SNPs/ gene to randomize the SNP distribution. The recent release of rainbow trout genome (GenBank assembly Accession GCA_002163495, RefSeq assembly accession GCF_002163495) helped in assigning SNPs to chromosomes. Recently, the chip was successfully used to identify several QTL markers associated with muscle yield (Salem et al 2018). The objective of the current study was to explore the genetic architecture in one of the most important muscle quality attributes, fillet firmness in relation to protein content, and identify QTL associated with these traits in a rainbow trout population developed by the USDA/NCCCWA selective breeding program.

MATERIALS AND METHODS

Ethics statement

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 approval #056).

Fish population, tissue sampling, and phenotypic traits

Fish population and tissue sampling were previously described in detail (Al-Tobasei et al 2017). Briefly, diploid females from a growth-selected line at NCCCWA were used to carry out GWA analysis. This selective breeding program was initiated in 2004 and has gone through 5 generations of selection (Leeds et al 2016). Third- and fourth-generation fish (Year-class, YC, 2010 and YC 2012) were used for GWA analysis. Phenotypic data were collected from 789 fish representing 98 families from YC 2010 and 99 families from YC 2012. Over a 6-week period, full-sib families were produced from single-sire×single-dam matings. Eggs were reared in spring water and incubated at 7-13°C to hatch all families within a 3-week period. Each family was reared in a separate 200-L tank at ~12.5°C to retain pedigree information and were fed a commercial fishmeal-based diet (Zeigler Bros Inc., Gardners, PA). At ~5-months post-hatch, fish were tagged with a passive integrated transponder (Avid Identification Systems Inc., Norco, CA) and reared together in 800-L communal tanks supplied with partially-recirculated spring water, at ~13°C, until ~13 months post-hatch. Fish were fed a commercial fishmeal-based diet. The feeding schedule was previously described (Hinshaw 1999). Fish did not receive feed for 5 days prior to harvest to facilitate processing.

Whole body weight (WBW) was measured in fish belonging to each family and families were sorted according to their WBW. The 2nd or 3rd fish from each family was selected for muscle sampling to keep the distribution of WBW consistently adjusted around the median of each family. For each harvest year, selected fish were randomly assigned to

one of five harvest groups (~100 fish each) allowing one fish per family per harvest group. The five groups were sampled in five consecutive weeks (one group/week) each YC. Fish from the YC 2010 were harvested between 410- and 437-days post-hatch (mean body weight = 985 g; SD = 239 g), whereas those from YC 2012 were harvested between 446- and 481-days post-hatch (mean body weight = 1,803 g; SD = 305 g). Muscle shear force and protein content showed low regression coefficient (R^2) values of 0.05 and 0.04 with body weight, respectively. Fish were euthanized in a lethal dose of tricaine methane sulfonate (Tricaine-S, Western Chemical, Ferndale, WA), harvested, and eviscerated. Head-on gutted carcasses were packed in ice, transported to the West Virginia University Muscle Foods Processing Laboratory (Morgantown, WV), and stored overnight. Carcasses were manually processed into trimmed, skinless fillets (Salem et al 2013). Shear force of 4 X 8 cm section of cooked fillet was assessed using a five-blade, Allo-Kramer shear cell attached to a Texture Analyzer (Model TA-HDi®; Texture Technologies Corp., Scarsdale, NY), equipped with a 50 kg load cell; tests were performed at a crosshead speed of 127 mm/min(Aussanasuwannakul et al 2010). Texture Expert Exceed software (version 2.60; Stable Micro Systems Ltd., Surrey, U.K.) was used to record and analyze force-deformation graphs. Peak shear force (g/g sample) was recorded. All cooked texture evaluations were performed approximately 48 h post-harvest. Details of the proximate analyses, including crude protein were previously described (Manor et al 2015). Crude protein analysis was achieved using AOAC-approved methods (AOAC 2000). Percent Kjeldahl nitrogen (KjeltecTM 2300; Foss North America; Eden Prairie, MN, USA) was converted into crude protein using 6.25 as the conversion factor. The pedigree-based

heritability h^2 (h^2ped) for protein content and shear force were estimated according to Zaitlen et al. (Zaitlen et al., 2013).

SNP genotyping and quality control

Genotyping was done using a 50K transcribed gene SNP-chip that we recently described and utilized in identifying QTL affecting fillet yield (Salem et al 2018). Source of all SNPs used to build the SNP chip was described in our previous publication (Al-Tobasei et al 2017). In brief, the array has about 21K SNPs showing potential allelic imbalances with fish body weight, muscle yield, fat content, shear force, whiteness index, and susceptibility to Bacterial Cold Water Disease (BCWD) as we previously described (Al-Tobasei et al 2017, Salem et al 2018). In addition, ~5K nonsynonymous SNPs and more SNPs were added to the chip to include at least 2 SNPs per each SNP-harboring gene. The SNP chip includes a total of 50,006 SNPs.

As describe before, a total of 1,728 fish were used to assess quality of this Affymetrix SNP chip. Genotyped fish were obtained from the NCCCWA growth- and BCWD-selection lines (Salem et al 2018). The SNP chip and sample metrics were calculated. Assessment of quality control (QC) and filtration of samples/genotypes have been performed using the Affymetrix SNPolisher software at the default parameters (Liu et al 2015). A call rate of 0.97 and Dish QC (DQC) threshold of 0.82 have been applied to filter out genotyped samples. For this study, 789 fish genotyped by the SNP chip had available phenotypic data for fillet shear force and protein content. All genotypic data passed the QC. Those fish were used for the current GWA analyses.

Fifty-SNP window GWA analysis

GWA analysis was performed using the Weighted single-step GBLUP (WssGBLUP) as we previously described (Salem et al 2018). In brief, WssGBLUP allows use of genotyped and ungenotyped animals. WssGBLUP integrates phenotypic data, genotype and pedigree information in a combined analysis using the following mixed model for single trait analysis:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{w} + \mathbf{e}$$

Where **y** is the vector of the phenotypes, **b** is the vector of fixed effects including harvest group and hatch year, **a** is the vector of additive direct genetic effects (i.e., animal effect), **w** is the vector of random family effect, and **e** is the residual error. The matrices **X**, $\mathbf{Z}_1$, and $\mathbf{Z}_2$ are incidence matrices for the effects contained in **b**, **a**, and **w**, respectively. The model combines all the relationship information (based on pedigree and genotypes) into a single matrix ($\mathbf{H}^{-1}$).

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

where $\mathbf{H}^{-1}$ is the inverse of the realized relationship matrix (**H**), $\mathbf{A}^{-1}$ is the inverse of the relationship matrix based on pedigree information, $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree relationship matrix for genotyped animals only, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix. The random family effect is uncorrelated and just accounts for the fact the animals within the same family were raised in a common environment, and the covariance structure is given by $\mathbf{I}\sigma_w^2$, where **I** is an identity matrix and σ_w^2 is the family variance.

AIREMLF90 (Misztal et al 2018) was used to estimate the variance components for the additive direct genetic effect, random family effect, and residuals. The inbreeding value, was previously calculated using a pedigree data of 63,808 fish from five consecutive generations in the NCCCWA breeding program using INBUPGF90 (Misztal et al 2002, Salem et al 2018). Quality control (QC) of genomic data was performed using PREGSF90 (Misztal et al 2014) according to the following settings; MAF > 0.05, call rate > 0.90, and HWE < 0.15. In total, 35,322 SNPs (70.6%) passed QC.

In WssGBLUP analysis, two iterations were used. All SNPs were assigned the same weight during the first iteration (i.e., weight = 1.0). For the second iteration, weights were calculated according to the SNP effects ($\hat{u}$) assessed in the first iteration as $\hat{u}^2 2p(1-p)$, where p represents the current allele frequency. Three steps were performed in each iteration: 1) weight was assigned to the SNPs. 2) genomic estimated breeding values (GEBV) were computed using BLUPF90 based on $\mathbf{H}^{-1}$ (Misztal et al 2002). 3) SNP effects and weights were calculated using POSTGSF90 (Misztal et al 2002) based on sliding variance windows of 50 adjacent SNPs. Since the SNPs in the chip were not evenly distributed over the whole genome, the window size used for the current analysis was based on a specific number of adjacent SNPs (n = 50 SNPs) instead of physical size (e.g. specific number of nucleotides). A Manhattan plot showing the proportion of additive genetic variance explained by the 50 SNP windows was generated in R using the qqman package (Turner 2014).

Single marker GWA analysis

Single marker association analysis was conducted using PLINK (Purcell et al 2007). The phenotypic data were checked for normality using Kolmogorov-Smirnov and Shapiro-

Wilk test in order to make sure that the studied phenotypes are normally distributed and meet the assumption of linear model analysis in PLINK (Purcell et al 2007). For single marker association analysis, the linear model included multiple covariates and account for population structure. To control the global inflation of the test statistic, the first five Principal components (PCs) were used as covariates in the model. The Wald test, using the --assoc command, was applied to the quantitative traits in order to retrieve the R-squared values of association.

Gene annotation and enrichment analysis

To retrieve SNP annotations, SNPs bed file was intersected with the rainbow trout genome gff/gtf file using Bedtools as described before (Quinlan and Hall 2010, Salem et al 2018). SNPs located within each gene were classified as genic whereas SNPs located outside the body of the gene were classified as intergenic. Genic SNPs were subsequently classified as CDS, intronic, 5'UTR or 3'UTR SNPs. SNPs within long noncoding RNAs (lncRNAs) were determined using a gtf file of our previously published lncRNA reference assembly (Al-Tobasei et al 2016). SNP-harboring genes were uploaded to the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang et al 2009a, Huang et al 2009b) to perform gene enrichment analysis (Fisher Exact < 0.05).

RESULTS AND DISCUSSION

Soft flesh is a major criterion for downgrading fish fillets, resulting in loss of value (Michie 2001). Post-mortem muscle softness is correlated with proteolytic degradation of extracellular matrix and cell membrane components (Bahuaud et al 2010, Martinez et al 2011). The fish population used for the current GWA analysis had average shear force of 475.7 ± 83.47

(g/g) and crude protein% = 20.64 ± 0.62. For the current GWA analysis, phenotypic variations in shear force and protein are shown in Figure (1).
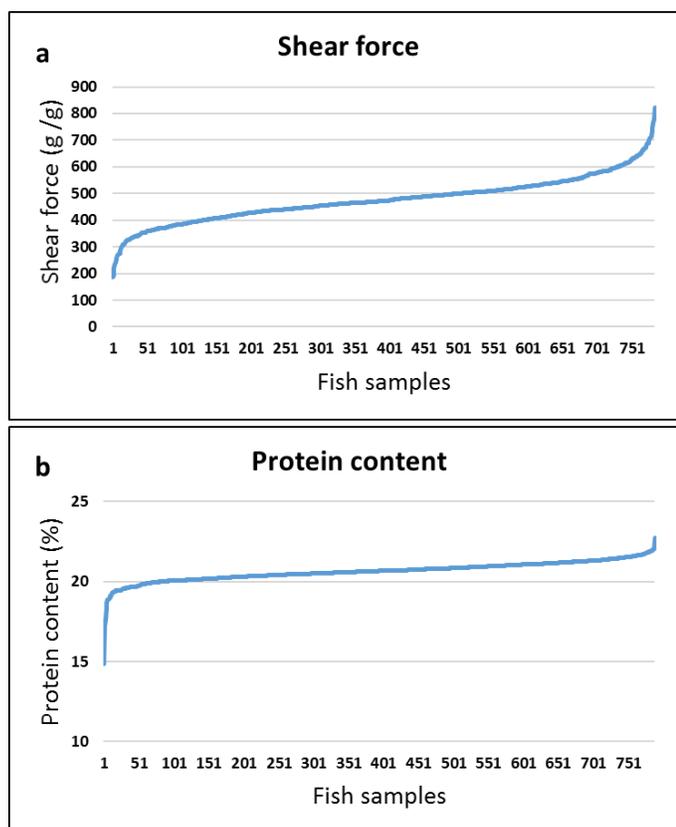
Figure 1. Phenotypic variations in shear force (a) and protein content (b) among fish samples used for GWA analysis.

The estimated heritabilities were 0.33 ± 0.07 and 0.27 ± 0.06 for shear force and protein content, respectively. Previous studies showed a significant correlation between changes in protein content and meat tenderness (Grześ et al 2017). Consistently, our data showed a significant correlation between protein content and shear force (R^2 = 0.18; p-value

<0.0001). Therefore, we used a 50K SNP chip to perform GWA analyses to identify QTL associated with both traits based on 50 SNP sliding windows using WssGBLUP and single-marker association analyses using PLINK. The chip contains SNPs potentially associated with muscle quality traits including fillet softness as we previously described (Al-Tobasei et al 2017, Salem et al 2018). However, we did not include any fish used in building the SNP-chip for GWA analysis in this study.

QTL affecting fillet shear force using WssGBLUP

The WssGBLUP-based GWA analysis identified a total of 212 SNPs affecting the additive genetic variance for shear force. These SNPs were located within 95 genes coding for proteins and 4 lncRNAs with 20 SNPs in intergenic regions. SNPs were included in windows explaining at least 2% (arbitrary value) of the additive genetic variance for shear force (Table S1). Genomic loci that harbor SNPs were clustered on 6 chromosomes (4, 7, 8, 10, 13, and 28) (Figure 2). Chromosome 13 had the most significant peaks affecting shear force (6.91%) (Table S1, Figure 2) and the highest number of SNPs (n = 83) in windows explaining additive genetic variance for shear force (Table S1). Many of the SNPs (n = 80) were located within the 3'UTR of their genes suggesting a role for these SNPs in microRNA, post-transcriptional regulation of gene expression. Among those 80 SNPs, 32 SNPs created or deleted binding sites for 56 microRNAs (Appendix A & Table S2). All QTL associated with genetic variance of shear force are listed in Table (S1). To gain insights into the biological significance of the identified QTL, we annotated the SNP-harboring genes followed by gene enrichment analysis. Functional annotation showed that SNP-harboring genes were involved in calcium binding/ metabolism, proteolytic activities, apoptotic process, and cellular adhesion and junction (Tables 1 & 2). Enriched terms

included calcium channel complex, smooth endoplasmic reticulum, ryanodine-sensitive calcium-release channel activity, calcium ion binding, and Z disc (Appendix B & Table S3).
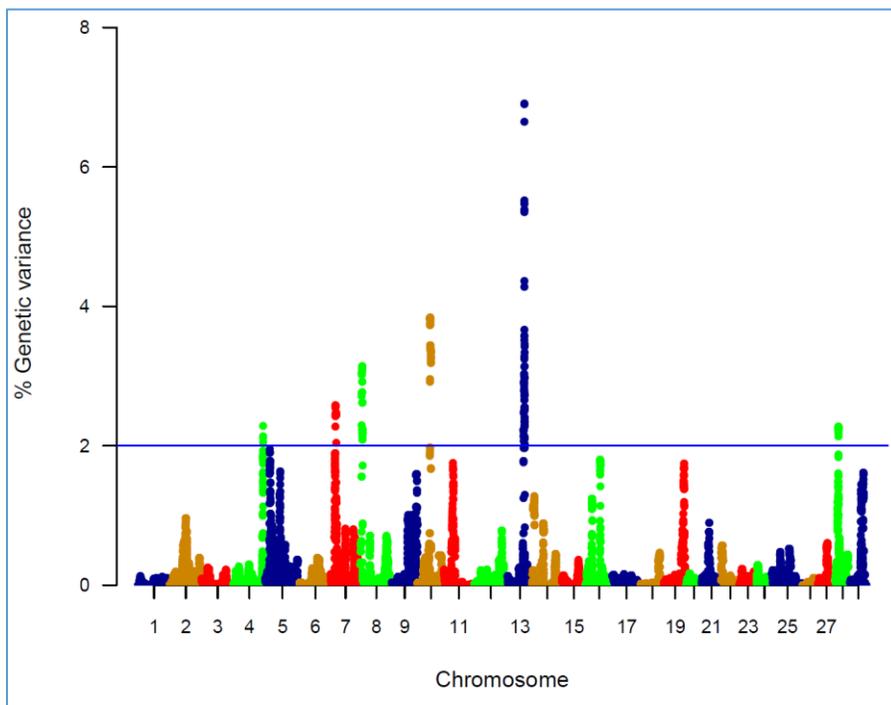
Figure 2. Manhattan plot showing association between genomic sliding windows of 50 SNPs and muscle shear force. Chromosome 13 showed the highest peaks with genomic loci explaining up to 6.91% of the additive genetic variance. The basal blue line represents 2% of additive genetic variance explained by SNPs.

SNPs in genes affecting Ca^{2+} homeostasis

Ten genes necessary for calcium metabolism harbored 47 SNPs affecting the genetic variation in shear force (Table 1). Ryanodine receptor3 (RYR3; member of the

sarcoplasmic reticulum calcium release channel) had 17 SNPs located on chromosome 4 and 8 suggesting an important role for calcium in regulating shear force.

Two SNPs were nonsynonymous, and one of these SNPs exists in the third structural repeat that is conserved in all RYR isoforms; it is located in the N-terminal part of the cytoplasmic region of the RYRs. Several studies reported a correlation between development of pale, soft and exudative (PSE) meat and abnormality in calcium release mechanism of porcine skeletal muscle as a result of a point mutation in the porcine RYR1 that led to a substitution of cysteine for arginine (Arg615Cys) (Fujii et al 1991). Poor regulation of the mutant channel led to accumulation of sarcoplasmic calcium and development of PSE meat accordingly(MacLennan and Phillips 1992). Breeding strategies were initiated to avoid this mutation from the pig populations.

Unlike in mammals where RYR1 is the main isoform expressed in skeletal muscle, fish co-express RYR3 (Murayama and Kurebayashi 2011). Absence of RYR1 in fish, causes slow swimming, weak contractions and reduced Ca^{2+} transients (Hirata et al 2007). On the other hand, RYR3 knock-down led to reduction in formation of anatomical structures called the parajunctional feet (PJF), which are located on the sides of the SR junctional cisternae (Perni et al 2015). Reduction of the PJF was coupled with reduced SR Ca^{2+} flux that causes Ca^{2+} sparks that was reported in fish muscle. However, the muscle fibers looked structurally normal and the swimming behavior was not affected (Perni et al 2015). Association of RYR1&3 mRNA expression level with fillet water holding capacity was reported in the Nile tilapia under pre-slaughter stress (Goes et al 2015). Impaired Ca^{2+} handling was reported in the muscles of the hatchery-reared salmon compared to that of wild fish (Anttila et al 2008). Levels of RYR were greatly reduced in the muscles of the

hatchery-reared salmon. Similar differences were seen in the oxidative capacity of muscles. This impairment was suggested to contribute to the lower swimming capacity of the reared fish (Anttila et al).

Chromosome 13 had 15 SNPs in Ca^{2+} homeostasis-relevant genes located in top windows affecting the genetic variability in shear force (Table 1). Nucleobindin 1, is a multidomain calcium-binding protein of unclear physiological and biochemical functions (Kapoor et al 2010) and harbored 2 SNPs within the 3'UTR representing the highest peak in this QTL. The second gene on this chromosome, myosin-binding protein C, fast (MyBP-C), encompassed 4 SNPs. MyBP-C sensitizes the actin thin filaments to Ca^{2+} (Lin et al 2018). MyBPC gene knockout in mice leads to muscle hypertrophy and impaired contractile function. The third gene, protein kinase C and casein kinase substrate in neurons protein 3 (PACSIN 3) had 4 SNPs. PACSIN 3 has been primarily identified in muscle and lung (Bai et al 2012). PACSIN 3 is known to modulate the subcellular localization of TRPV4 (Cuajungco et al 2006) which regulates Ca^{2+} homeostasis and cytoskeletal remodeling (Ryskamp et al 2016). Coronin-1 represents the fourth gene and had two SNPs. It mediates Ca^{2+} mobilization from the intracellular stores (Mueller et al 2008). The fifth gene, myosin regulatory light chain 2 (MYL2), had three SNPs within 3'UTR. MYL2 is a calcium-binding chain known to be associated with meat tenderness (Rosa et al 2018).

Table 1. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of additive genetic variance in shear force and involved in calcium homeostasis. 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
2.29	4	79275235	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.14	4	79277177	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.11	4	79279144	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/nonsyn
2.06	4	79282537	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	Intronic
2.18	8	5481101	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.18	8	5481500	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.18	8	5481770	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.23	8	5481890	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.23	8	5481962	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.23	8	5482326	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/nonsyn
2.23	8	5482409	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.23	8	5483556	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.23	8	5488334	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.24	8	5498487	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.24	8	5499229	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.19	8	5509881	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.19	8	5512331	-	LOC110529177	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.92	10	33155185	+	LOC110533811	Calcium metabolism	plastin-3	3'UTR
2.96	10	33155312	+	LOC110533811	Calcium metabolism	plastin-3	3'UTR
3.73	10	33155825	+	LOC110533811	Calcium metabolism	plastin-3	3'UTR
3.75	10	33156032	+	LOC110533811	Calcium metabolism	plastin-3	3'UTR
3.83	10	33157280	+	LOC110533811	Calcium metabolism	plastin-3	3'UTR
3.38	10	34861588	-	LOC110533854	Calcium metabolism	TBC1 domain family member 8B-like	3'UTR
3.37	10	35666172	-	LOC110533869	Calcium metabolism	galectin-9-like	CDS/nonsyn
3.36	10	35668815	-	LOC110533869	Calcium metabolism	galectin-9-like	CDS/nonsyn
5.47	13	45250062	-	LOC110486648	Calcium metabolism	nucleobindin-1-like	3'UTR
6.90	13	45250138	-	LOC110486648	Calcium metabolism	nucleobindin-1-like	3'UTR
5.52	13	45348326	+	LOC110486653	Calcium metabolism	myosin-binding protein C, fast-type-like	CDS/syn
4.29	13	45348905	+	LOC110486653	Calcium metabolism	myosin-binding protein C, fast-type-like	CDS/syn
3.59	13	45353098	+	LOC110486653	Calcium metabolism	myosin-binding protein C, fast-type-like	Intronic
3.57	13	45358893	+	LOC110486653	Calcium metabolism	myosin-binding protein C, fast-type-like	CDS/nonsyn
3.67	13	45494621	-	LOC110486657	Calcium metabolism	protein kinase C and casein kinase substrate in neurons protein 3-like	3'UTR
3.13	13	45495127	-	LOC110486657	Calcium metabolism	protein kinase C and casein kinase substrate in neurons protein 3-like	3'UTR
3.25	13	45495294	-	LOC110486657	Calcium metabolism	protein kinase C and casein kinase substrate in neurons protein 3-like	3'UTR
3.28	13	45497545	-	LOC110486657	Calcium metabolism	protein kinase C and casein kinase substrate in neurons protein 3-like	CDS/nonsyn
2.98	13	45641799	-	LOC110486661	Calcium metabolism	coronin-1A-like	3'UTR
2.74	13	45644102	-	LOC110486661	Calcium metabolism	coronin-1A-like	CDS/syn
2.35	13	45825907	+	LOC110486680	Calcium metabolism	myosin regulatory light chain 2, skeletal muscle isoform-like	Intronic
2.36	13	45826199	+	LOC110486680	Calcium metabolism	myosin regulatory light chain 2, skeletal muscle isoform-like	3'UTR
2.51	13	45826267	+	LOC110486680	Calcium metabolism	myosin regulatory light chain 2, skeletal muscle isoform-like	3'UTR
2.25	28	9763838	-	LOC110508483	Calcium metabolism	matrix metalloproteinase-14-like	3'UTR
2.26	28	9763927	-	LOC110508483	Calcium metabolism	matrix metalloproteinase-14-like	3'UTR
2.26	28	9764282	-	LOC110508483	Calcium metabolism	matrix metalloproteinase-14-like	3'UTR
2.27	28	9764561	-	LOC110508483	Calcium metabolism	matrix metalloproteinase-14-like	3'UTR
2.18	28	9767473	-	LOC110508483	Calcium metabolism	matrix metalloproteinase-14-like	CDS/syn
2.14	28	9771512	-	LOC110508483	Calcium metabolism	matrix metalloproteinase-14-like	CDS/syn
2.21	28	9784531	-	LOC110508483	Calcium metabolism	matrix metalloproteinase-14-like	5'UTR

Eight SNPs were also identified in 3 genes necessary for calcium metabolism on chromosome 10 (Table 1). Plastin-3 (PLS3) had five SNPs in windows explaining up to 3.83% of the additive genetic variance in shear force. PLS3 functions as a protective modifier of spinal muscular atrophy in Ca^{+2}-dependent manner (Lyon et al 2014). A single SNP was identified in a gene that codes for TBC1 domain family member 8B (TBC1D8) and has GO terms belong to calcium ion binding. Galectin-9 (Gal-9) harbored two SNPs within windows explaining ~ 3.37% of the additive genetic variance in shear force. Gal-9 induces apoptosis via the Ca^{2+}-calpain-caspase-1 pathway (Kashio et al 2003).

Chromosome 28 had a single gene, matrix metalloproteinase-14 (MMP14), that had 7 SNPs explaining at least 2.0% of the additive genetic variance (Table 1). MMP14 has a Ca^{2+}-dependent catalytic MP domain that degrades the extracellular matrix proteins such as collagen (Tallant et al 2010). Our recent studies showed that MMP9 was downregulated in trout families of high shear force suggesting a role for matrix metalloproteinase family in regulating fillet firmness in fish (Ali et al 2018). In addition, transcripts of stanniocalcin (STC), the main regulatory hormone of Ca^{2+} homeostasis in fish (Verma and Alim 2014), were overexpressed in trout families with high shear force (Ali et al 2018). The relationship between calcium and protein content in dystrophic muscle has been attributed to decreased functionality of the sarcoplasmic reticulum to sequester calcium ions (KAMEYAMA and ETLINGER 1979). Together, our results indicate a major role of Ca^{2+} homeostasis in determining fish fillet firmness.

SNPs in genes affecting proteolysis

Six SNP-harboring genes involved in proteolytic/ catabolic and apoptotic processes were identified on chromosomes 10, 13, and 28 (Table 2). Chromosome 10 had a gene

that codes for Gal-9 which is known to induce apoptotic process (Kashio et al 2003). Chromosome 13 had four genes harboring SNPs within top windows affecting the additive genetic variance in shear force. First, tripartite motif-containing protein 16, affecting 5.47% of the additive genetic variance, promotes apoptosis by modulating the caspase-2 activity. Second, branched-chain-amino-acid aminotransferase (cytosolic), had a single 3'UTR SNP. This enzyme catalyzes the first reaction in the catabolism of the most hydrophobic branched chain amino acids (leucine, isoleucine, and valine) that play important roles in determining the structure of globular proteins, in addition to the interaction of transmembrane domains with the phospholipid layer (Blomstrand et al 2006). Third, potassium voltage-gated channel, subfamily A member 1 harbored a single synonymous SNP. Voltage-dependent potassium channels mediates transmembrane potassium transport and are involved in the proteolytic system that causes postmortem tenderization (Mateescu et al 2017). The fourth gene in the list codes for granulins that had 2 SNPs. Granulins have possible critical lysosomal functions, and their loss is an initiating factor in lysosomal dysfunction (Holler et al 2017). In addition, chromosome 28 had four SNPs in a gene coding for apoptotic chromatin condensation inducer in the nucleus (ACIN1) (Table 2). ACIN1 belongs to the prominent canonical apoptosis signaling pathway (Schrötter et al 2012).

Two SNP-harboring genes were mapped to the autophagy pathway; immunoglobulin-binding protein 1 (IGBP1) and zinc finger FYVE domain-containing protein 1 (ZFYVE1). ZFYVE1, has been used as a marker of omegasomes (exists only during autophagosome formation) (Zientara-Rytter and Subramani 2016). Three SNPs spanning two genes coding for coronin-1A and charged multivesicular body protein 1b, were mapped to the

endosomal/phagosomal pathway. Previous studies support presence of phagocytic activities in postmortem muscle to eliminate extracellular material (Ouali et al 2013).

SNPs in genes affecting cell adhesion

Genes involved in focal adhesion and cell junction were previously reported to be associated with meat tenderness (Fonseca et al 2017). Five SNPs spanning two genes on chromosome 13 were mapped to the focal adhesion pathway (Table 2). These genes code for myosin regulatory light chain 2 and serine/threonine-protein phosphatase alpha-2. In addition, 2 SNPs were identified spanning two genes involved in tight junction pathway (Table 2). The two genes are located on chromosomes 7 and 13, and code for Na (+)/H (+) exchange regulatory cofactor NHE-RF1 and actin-related protein 3. Cerebellin-1 on chromosome 28, had a single SNP in a window explaining 2.25% of the additive genetic variance (Table 2). Functional annotation analysis showed that cerebellin-1 has GO terms belonging to heterophilic, cell-cell adhesion via plasma membrane, cell adhesion molecules. The list also includes a SNP in a gene on chromosome 13, that codes for claudin-4 (Table 2). This SNP creates a binding site for the mir-10c-5p. mir-10c-5p showed differential expression association with shear force in trout fish families of YC 2010 (Paneru et al 2017). Members of the claudins family are major integral membrane proteins existing at tight junctions, and they have Ca^{2+}-independent cell-adhesion activity (Kubota et al 1999).

Table 2. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of additive genetic variance in shear force and involved in proteolytic, apoptotic process, tight junction, and focal adhesion. 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
2.47	7	14194291	-	LOC110527456	Tight junction	actin-related protein 3	CDS/syn
3.05	8	3673516	+	LOC110529290	Autophagy	zinc finger FYVE domain-containing protein 1-like	CDS/syn
3.05	8	3673957	+	LOC110529290	Autophagy	zinc finger FYVE domain-containing protein 1-like	3'UTR
3.83	10	33231241	-	LOC110533814	Endosome	charged multivesicular body protein 1b	CDS/syn
3.44	10	34812751	+	LOC110533848	Autophagy	immunoglobulin-binding protein 1-like	CDS/syn
3.37	10	35666172	-	LOC110533869	Apoptosis	galectin-9-like	CDS/nonsyn
3.36	10	35668815	-	LOC110533869	Apoptosis	galectin-9-like	CDS/nonsyn
2.23	13	43044831	+	LOC110485193	Tight junction	Na(+)/H(+) exchange regulatory cofactor NHE-RF1-like	3'UTR
3.52	13	45087094	+	LOC110486632	Lysosome	granulins-like	CDS/syn
4.36	13	45089988	+	LOC110486632	Lysosome	granulins-like	3'UTR
5.38	13	45111046	-	LOC110486641	Amino acid catabolism	branched-chain-amino-acid aminotransferase, cytosolic-like	3'UTR
5.36	13	45142682	+	LOC110486644	Proteolysis	potassium voltage-gated channel subfamily A member 1-like	CDS/syn
5.47	13	45284153	-	LOC110486651	Apoptosis	tripartite motif-containing protein 16-like	3'UTR
3.46	13	45619953	-	LOC110486660	Focal adhesion	serine/threonine-protein phosphatase alpha-2 isoform-like	3'UTR
3.15	13	45620629	-	LOC110486660	Focal adhesion	serine/threonine-protein phosphatase alpha-2 isoform-like	3'UTR
2.98	13	45641167	+	LOC110486662	Tight junction	claudin-4-like	3'UTR
2.98	13	45641799	-	LOC110486661	Phagosome	coronin-1A-like	3'UTR
2.74	13	45644102	-	LOC110486661	Phagosome	coronin-1A-like	CDS/syn
2.35	13	45825907	+	LOC110486680	Focal adhesion	myosin regulatory light chain 2, skeletal muscle isoform-like	Intronic
2.36	13	45826199	+	LOC110486680	Focal adhesion	myosin regulatory light chain 2, skeletal muscle isoform-like	3'UTR
2.51	13	45826267	+	LOC110486680	Focal adhesion	myosin regulatory light chain 2, skeletal muscle isoform-like	3'UTR
2.23	28	9751858	+	LOC110508482	Apoptosis	apoptotic chromatin condensation inducer in the nucleus-like	CDS/nonsyn
2.23	28	9753870	+	LOC110508482	Apoptosis	apoptotic chromatin condensation inducer in the nucleus-like	CDS/syn
2.23	28	9754864	+	LOC110508482	Apoptosis	apoptotic chromatin condensation inducer in the nucleus-like	3'UTR
2.27	28	9754973	+	LOC110508482	Apoptosis	apoptotic chromatin condensation inducer in the nucleus-like	3'UTR
2.25	28	9843505	+	LOC110508486	Cell adhesion	cerebellin-1-like	3'UTR

QTL affecting protein content using WssGBLUP

In total, 225 SNPs affecting the genetic variation in muscle protein content were identified; 202 genic and 23 intergenic SNPs (Table S4). Each SNP was in a window explaining at least 2% of the additive genetic variance for the protein content. The genomic loci that harbor SNPs were clustered on five chromosomes (1, 3, 4, 7, and 11) (Figure 3). Chromosomes 4 and 1 harbored 50 SNPs located within top windows affecting the genetic

variability (variance > 4.0%) in protein content of the muscle (Table S4). Similar to shear force, 40% of the SNPs were located within 3'UTR. Thirteen SNPs created/deleted target sites for 16 microRNAs (Appendix C & Table S5). SNPs associated with genetic variation in crude protein content are listed in Table (S4). Functional annotation followed by gene enrichment analysis were performed to functionally characterize the SNP-harboring genes. Functional annotation showed that SNP-harboring genes were mainly involved in apoptotic process, proteolysis, lysosomal activities, cell proliferation, transcription, and methylation (Table 3 & 4). Enriched terms included muscle contraction, transcription, regulation of transcription, and chromatin remodeling (Appendix D & Table S6).

SNPs in genes affecting apoptosis

Thirteen SNPs were identified spanning seven genes on chromosomes 4 and 7, and engaged in apoptotic process (Table 3). Actin, alpha harbored two SNPs in windows that explained the highest genetic variability (4.62%) in this category. Alpha actin was previously suggested as a genetic marker for apoptosis (Ouali et al 2013). SNW domain-containing protein 1 (SNW1) harbored 4 SNPs in windows explaining up to 3.53% of the additive genetic variance. Depletion of SNW1 or its associating proteins induced apoptotic processes in cancer cells (Sato et al 2015). Three SNPs were identified in RNA-binding protein 25 (RBM25) and Bcl-2-like protein 1 (BCL2L1). RBM25 is involved in apoptotic cell death by regulating BCL2L1 expression (Zhou et al 2008). Two SNPs were identified in RHOB that is known to positively regulate apoptotic process (Srougi and Burridge 2011). A single 3'UTR SNP was identified in a gene coding for protein snail homolog Snai. Snai1-expressing cells resists apoptosis triggered by proapoptotic stimuli (Olmeda et al 2007). Another 3'UTR SNP was also identified in a gene coding for cell death

activator CIDE-3. This gene has a role in the execution phase of apoptosis (Liang et al 2003).
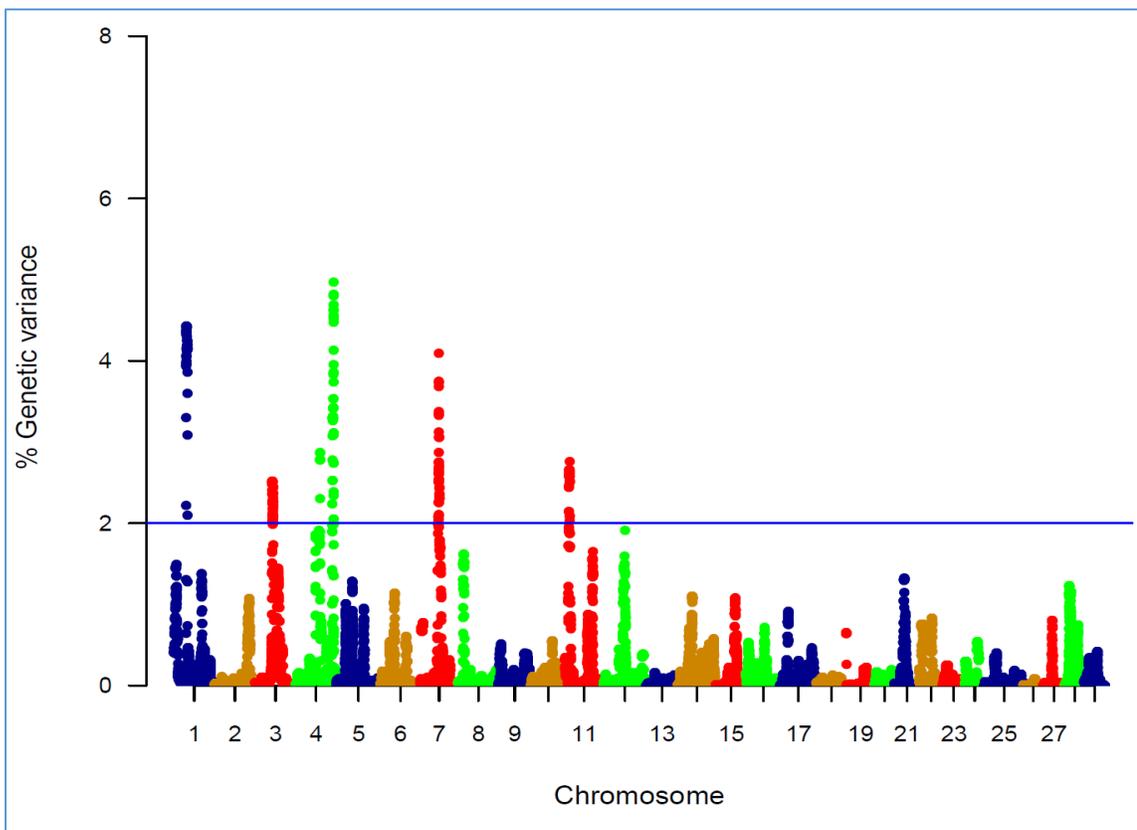
Figure 3. Manhattan plot showing association between genomic sliding windows of 50 SNPs and muscle protein content. Chromosome 4 showed the highest peaks with genomic loci explaining up to 4.97% of the additive genetic variance. The basal blue line represents 2% of additive genetic variance explained by SNPs.

SNPs in genes affecting proteolysis

Ten genes with proteolytic activities were identified that affected genetic variability in protein content (Table 3). A single SNP located in the gene coding for short transient

receptor potential channel 4-associated protein (TRPC4AP) followed by a SNP in 26S protease regulatory subunit 4 (PSMC1) came at the top of this group. TRPC4AP is involved in ubiquitination and destruction of Myc protein (Choi et al 2010) that control cell proliferation and growth (Bernard and Eilers 2006). Whereas, PSMC1 is a component of the 26S proteasome that maintains protein homeostasis through ubiquitin-mediated degradation of damaged and misfolded proteins (Kanayama et al 1992). NEDD8 ultimate buster 1 (NUB1) and inactive serine protease 35 (PRSS35) had a single SNP. NUB1 positively regulates proteasomal ubiquitin-dependent protein catabolic process (Schmidtke et al 2006) whereas, the proteolytic activities of the serine protease, PRSS35, have not been characterized yet (Diao et al 2013). Plectin had nine SNPs. In human, mutations of the plectin gene cause muscular dystrophy (Natsuga et al 2010). The list also includes two mitochondrial genes, encoding for 2-oxoisovalerate dehydrogenase subunit beta and aspartate aminotransferase, involved in amino acid catabolism (Nobukuni et al 1991, Schiele et al 1989). Of note, three genes on chromosome 4 were involved in lysosomal activities; V-type proton ATPase subunit D, Rho-related GTP-binding protein RhoB (RHOB), and lysosomal-associated transmembrane protein 4A (LAPTM4A). V-type proton ATPase subunit D had 4 SNPs in windows explaining up to 3.30% of the genetic variation in crude protein content. The vacuolar (H^+)-ATPases acidify the intracellular compartments and play an important role in protein degradation (Toei et al 2010). RHOB is involved in trafficking epidermal growth factor (EGF) receptor from late endosomes to lysosomes (Gampel et al 1999). Three SNPs were identified in LAPTM4A. The function of this gene is unclear.

Table 3. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of additive genetic variance in protein content and involved in proteolytic and apoptotic processes. 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
2.22	1	25420470	+	LOC110520559	Proteolysis	aspartate aminotransferase, mitochondrial-like	3'UTR
2.02	3	35652142	-	LOC110518458	Proteolysis	plectin-like	CDS/syn
2.41	3	35653189	-	LOC110518458	Proteolysis	plectin-like	CDS/syn
2.36	3	35653735	-	LOC110518458	Proteolysis	plectin-like	CDS/syn
2.49	3	35654687	-	LOC110518458	Proteolysis	plectin-like	CDS/nonsyn
2.49	3	35654724	-	LOC110518458	Proteolysis	plectin-like	CDS/nonsyn
2.52	3	35654840	-	LOC110518458	Proteolysis	plectin-like	CDS/nonsyn
2.51	3	35654914	-	LOC110518458	Proteolysis	plectin-like	CDS/syn
2.51	3	35657920	-	LOC110518458	Proteolysis	plectin-like	CDS/syn
2.52	3	35673163	-	LOC110518458	Proteolysis	plectin-like	CDS/syn
2.04	3	36517107	+	LOC110518871	Proteolysis	2-oxoisovalerate dehydrogenase subunit beta, mitochondrial-like	3'UTR
2.05	3	36517385	+	LOC110518871	Proteolysis	2-oxoisovalerate dehydrogenase subunit beta, mitochondrial-like	3'UTR
2.18	3	36865549	-	LOC110519010	Proteolysis	inactive serine protease 35-like	CDS/nonsyn
2.30	4	50656794	-	LOC110522013	Proteolysis & apoptosis	rho-related GTP-binding protein RhoB	3'UTR
2.87	4	50657566	-	LOC110522013	Proteolysis & apoptosis	rho-related GTP-binding protein RhoB	CDS/syn
2.79	4	50753202	+	LOC110522012	Proteolysis	lysosomal-associated transmembrane protein 4A-like	3'UTR
2.78	4	50753210	+	LOC110522012	Proteolysis	lysosomal-associated transmembrane protein 4A-like	3'UTR
2.78	4	50753431	+	LOC110522012	Proteolysis	lysosomal-associated transmembrane protein 4A-like	3'UTR
3.30	4	77336642	-	LOC110522556	Proteolysis	V-type proton ATPase subunit D-like	3'UTR
3.29	4	77336859	-	LOC110522556	Proteolysis	V-type proton ATPase subunit D-like	3'UTR
3.29	4	77338088	-	LOC110522556	Proteolysis	V-type proton ATPase subunit D-like	CDS/syn
3.29	4	77342367	-	LOC110522556	Proteolysis	V-type proton ATPase subunit D-like	CDS/syn
3.29	4	77347822	-	LOC110522557	Apoptosis	RNA-binding protein 25-like	CDS/syn
3.30	4	77352343	-	LOC110522557	Apoptosis	RNA-binding protein 25-like	CDS/nonsyn
3.31	4	77453100	-	LOC110522560	Proteolysis	26S protease regulatory subunit 4-like	CDS/syn
3.42	4	78252725	+	LOC110522566	Apoptosis	SNW domain-containing protein 1-like	CDS/syn
3.42	4	78252740	+	LOC110522566	Apoptosis	SNW domain-containing protein 1-like	CDS/nonsyn
3.42	4	78254203	+	LOC110522566	Apoptosis	SNW domain-containing protein 1-like	CDS/syn
3.53	4	78263097	+	LOC110522566	Apoptosis	SNW domain-containing protein 1-like	3'UTR
4.62	4	79069732	+	LOC110522582	Apoptosis	actin, alpha cardiac	CDS/syn
4.62	4	79071472	+	LOC110522582	Apoptosis	actin, alpha cardiac	CDS/syn
2.53	7	38918677	-	LOC110527901	Apoptosis	bcl-2-like protein 1	3'UTR
2.75	7	39093852	+	LOC110527910	Apoptosis	protein snail homolog Sna-like	3'UTR
3.38	7	39160373	-	LOC110527912	Proteolysis	short transient receptor potential channel 4-associated protein-like	CDS/syn
2.30	7	40277695	+	LOC110527935	Apoptosis	cell death activator CIDE-3-like	3'UTR
2.57	11	9183364	-	LOC110535270	Proteolysis	NEDD8 ultimate buster 1-like	CDS/syn

SNPs in genes affecting Ca^{2+} homeostasis

We identified 28 SNPs, within 5 genes on chromosomes 1, 3, and 4, that are involved in calcium homeostasis (Table 4). Interestingly, RYR3 harbored ~ 68% of those SNPs; whereas four of these SNPs affected genetic variability in shear force. This result suggests a major role for RYR3 in regulating protein content and shear force in rainbow trout. SNPs

of RYR3 were ranked first in this category and were located within windows explaining up to 4.97% of the additive genetic variance in protein content. A single SNP was identified within a gene that codes for reticulocalbin 2 (RCN2). Previous studies showed that RCN2 binds to calcium and was identified to be localized in endoplasmic reticulum. RCN2 was upregulated in hepatocellular carcinoma patients and its homozygous deletion in mice was lethal (Wang et al 2017). In addition, there were three 3'UTR SNPs within the calmodulin (CaM) gene. CaM codes for a calcium binding protein known to regulate RYR activity through direct binding to a CaM-binding domain of RYR (Oo et al 2015). In addition, two genes coding for inhibitor of Bruton tyrosine kinase (Btk) and protein FAM26E (CALHM1) harbored 5 SNPs on chromosome 3. Btk plays a role in releasing sequestered Ca^{2+} to the cytosol (Liu et al 2001). Whereas, CALHM1 detects the extracellular Ca^{2+} level and plays a role in Ca^{2+} homeostasis (Ma et al 2012). These results suggest a significant role of the genes involved in Ca^{2+} handling (release and re-sequestration). In mammals, proteolysis by calcium-dependent proteases (calpains) in the early postmortem period greatly affects muscle texture and meat tenderization (Duckett et al 2000, Koohmaraie 1992). We previously showed that calpains are elevated and calpastatin is reduced during starvation-induced muscle degradation (Salem et al 2005a, Salem et al 2007) and calpastatin expression is associated with rainbow trout muscle growth (Salem et al 2005b). Further studies are warranted to investigate postmortem autolysis caused by calpain system in regulating protein content and shear force in rainbow trout.

Table 4. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of genetic variance in protein content and involved in calcium metabolism, cell proliferation, and transcriptional/ chromatin regulations.   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
3.97	1	25480866	-	LOC110520608	Chromatin regulator	histone-lysine N-methyltransferase KMT5B-like	CDS/syn
4.43	1	25956429	+	LOC110520691	Proliferation	myocyte-specific enhancer factor 2A-like	3'UTR
4.42	1	25956638	+	LOC110520691	Proliferation	myocyte-specific enhancer factor 2A-like	3'UTR
4.20	1	28282718	+	rcn2	Calcium & proliferation	reticulocalbin 2	CDS/syn
2.51	3	36206239	+	LOC110518743	Chromatin regulator	transcription and mRNA export factor ENY2-2	3'UTR
2.51	3	36206278	+	LOC110518743	Chromatin regulator	transcription and mRNA export factor ENY2-2	3'UTR
2.10	3	36651473	-	LOC110518895	Calcium metabolism	inhibitor of Bruton tyrosine kinase-like	3'UTR
2.11	3	36662440	-	LOC110518895	Calcium metabolism	inhibitor of Bruton tyrosine kinase-like	CDS/nonsyn
2.09	3	36662456	-	LOC110518895	Calcium metabolism	inhibitor of Bruton tyrosine kinase-like	CDS/nonsyn
2.09	3	36672488	-	LOC110518895	Calcium metabolism	inhibitor of Bruton tyrosine kinase-like	CDS/syn
2.17	3	37260197	+	LOC110519152	Chromatin regulator	ubiquinone biosynthesis O-methyltransferase, mitochondrial-like	CDS/syn
2.19	3	37276563	+	LOC110519152	Chromatin regulator	ubiquinone biosynthesis O-methyltransferase, mitochondrial-like	3'UTR
2.09	3	37664570	+	LOC110519270	Calcium metabolism	protein FAM26E-like	CDS/nonsyn
2.78	4	76616397	+	LOC110522552	Chromatin regulator	ribosomal oxygenase 1-like	3'UTR
3.27	4	77424809	-	LOC110522561	Calcium metabolism	calmodulin	3'UTR
3.30	4	77425427	-	LOC110522561	Calcium metabolism	calmodulin	3'UTR
3.30	4	77425526	-	LOC110522561	Calcium metabolism	calmodulin	3'UTR
3.42	4	78252725	+	LOC110522566	Chromatin regulator	SNW domain-containing protein 1-like	CDS/syn
3.42	4	78252740	+	LOC110522566	Chromatin regulator	SNW domain-containing protein 1-like	CDS/nonsyn
3.42	4	78254203	+	LOC110522566	Chromatin regulator	SNW domain-containing protein 1-like	CDS/syn
3.53	4	78263097	+	LOC110522566	Chromatin regulator	SNW domain-containing protein 1-like	3'UTR
3.86	4	79017410	-	LOC110522579	Transcription	poly(A) polymerase beta-like	3'UTR
3.85	4	79017997	-	LOC110522579	Transcription	poly(A) polymerase beta-like	3'UTR
4.56	4	79019016	-	LOC110522579	Transcription	poly(A) polymerase beta-like	3'UTR
4.63	4	79039957	-	LOC110522579	Transcription	poly(A) polymerase beta-like	CDS/syn
4.69	4	79274809	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
4.81	4	79275235	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
4.81	4	79277177	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
4.82	4	79279144	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/nonsyn
4.97	4	79282537	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	Intronic
4.13	4	79284060	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
4.51	4	79285101	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
4.54	4	79307585	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
4.48	4	79320253	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
3.96	4	79320321	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
3.74	4	79322346	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
3.11	4	79330676	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/nonsyn
3.09	4	79330715	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/nonsyn
2.74	4	79331299	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/nonsyn
2.74	4	79335498	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/nonsyn
2.39	4	79336796	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.34	4	79347211	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.05	4	79348423	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.05	4	79349026	+	LOC110521100	Calcium metabolism	ryanodine receptor 3-like	CDS/syn
2.51	7	38897108	+	LOC110527899	Chromatin regulator	host cell factor 1-like	3'UTR
2.51	7	38897671	+	LOC110527899	Chromatin regulator	host cell factor 1-like	3'UTR

SNPs in genes affecting transcriptional process and cell proliferation

Genes involved in transcription and cell proliferation were identified (Table 4). The majority of SNP-harboring genes were involved in transcription. Sixty-six SNPs were identified in 26 genes located, mainly, on chromosomes 4 and 7. Four SNPs in a gene that code for poly(A) polymerase beta were identified in windows explaining the highest genetic variability (4.63%) in this category. Additionally, twelve SNPs located on six genes involved in cell proliferation were identified. Three SNPs on two genes that code for myocyte-specific enhancer factor 2A (MEF2) and RCN2 were ranked at the top of this group. MEF2 plays diverse roles in muscle to control myogenesis (Black and Olson 1998).

SNPs in genes affecting histone modifications

Twelve SNPs in six genes involved in epigenetic transcriptional regulation were also identified on chromosomes 1, 3, 4, and 7 (Table 4). Histone-lysine N-methyltransferase KMT5B (KMT5B) had a single SNP located in a window explaining the maximum variance in protein content in this group (3.97%). KMT5B is a histone methyltransferase that trimethylates 'Lys-20' of histone H4 (a tag for epigenetic transcriptional repression) and plays a role in myogenesis (Neguembor et al 2013). Four SNPs in a gene that codes for SNW domain-containing protein 1 were identified. This protein positively regulates histone H3-K4 methylation (Brès et al 2009). A single SNP was identified on ribosomal oxygenase 1 that functions as histone lysine demethylase, a ribosomal histidine hydroxylase, and contributes to MYC-induced transcriptional activation (Eilbracht et al 2004, Ge et al 2012, Suzuki et al 2007). Two SNPs were identified in a gene coding for host cell factor 1 (HCF-1). In human, the cell-proliferation factor HCF-1 tethers Sin3 histone deacetylase and Set1/Ash2 histone H3-K4 methyltransferase (H3K4me) complexes

that are involved in repression and activation of transcription, respectively (Wysocka et al 2003). The list includes two other genes that harbored four SNPs on chromosome 3; transcription and mRNA export factor ENY2-2 and ubiquinone biosynthesis O-methyltransferase, mitochondrial.

Taken together, our results suggest that calcium homeostasis, more likely through RYR3, and transcriptional/ chromatin regulators have major roles in regulating genetic variability in muscle protein content.

Single marker association analyses

In addition to WssGBLUP and to identify single SNP marker association with phenotypic variation in shear force and protein content, we analyzed SNPs included in the SNP chip using general linear regression model available in PLINK which allows for multiple covariates (Purcell et al 2007). In this study, PLINK identified 11 significant SNPs with potential impact on the shear force (Bonferroni-corrected $p < 2.01E-06$; Figure 4 & Table 5).

Most of the significant SNPs were located on chromosome 5 (n = 5) and chromosome 7 (n = 4). However, the most significant SNP explaining 3.4% of the phenotypic variability in shear force, was located on chromosome 8 in a gene coding for RYR3. This result was in agreement with the WssGBLUP 50 SNP-window analysis and suggests an essential role for RYR3 in regulating fillet firmness in trout.
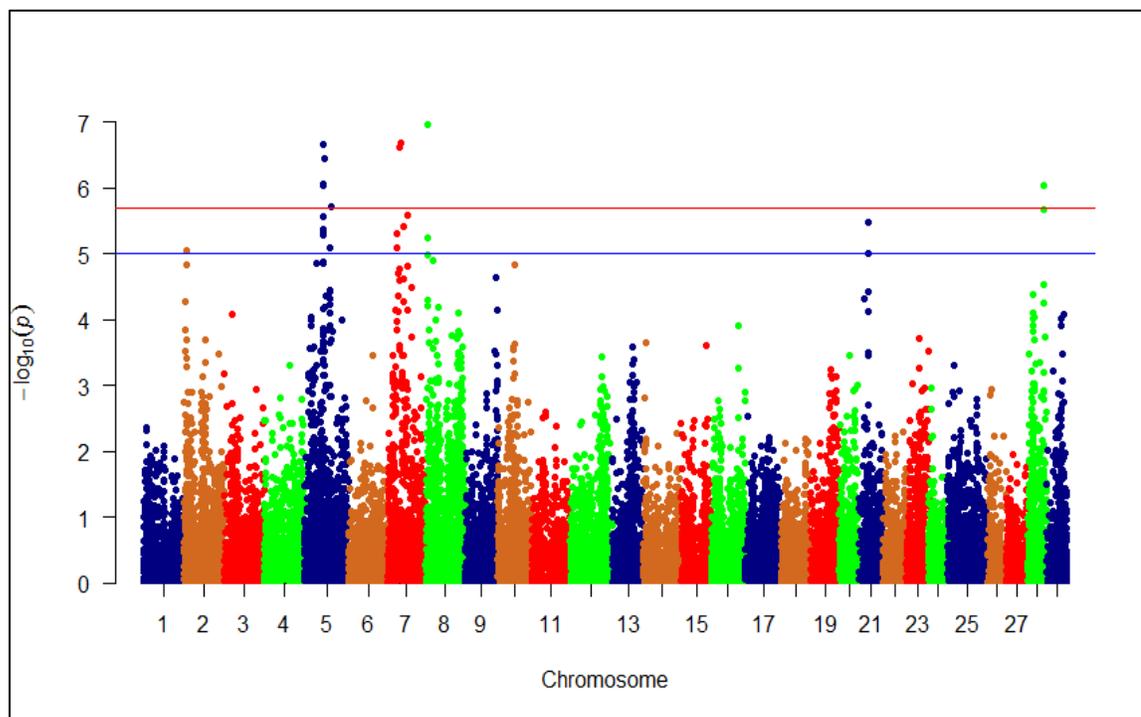
Figure 4. Manhattan plot showing single SNP markers associated with variations in shear force. Blue and red horizontal lines represent suggestive and significance threshold p-values of 1e-05 and 2.01e-06, respectively.

Cytochrome c oxidase subunit 6C-1 (COX6C1), 14-3-3B1 protein, and rho GTPase-activating protein 15 (ARHGAP15) were ranked next to RYR3 in impacting phenotypic variability in shear force. COX6 was rapidly degraded under endoplasmic reticulum stress conditions induced by Ca^{2+} depletion (Hong et al 2016) and upregulated in rainbow trout families of high shear force (Ali et al 2018b). 14-3-3B1 protein has been reported to be involved in apoptotic process (Rodrigues et al 2017). Previous studies elucidated the involvement of 14-3-3 proteins in meat tenderness (Rodrigues et al 2017). Overexpression of ARHGAP15 increases actin stress fibers and cell contraction (Seoh et al 2003).

ARHGAP15 SNP was in strong LD (D' = 1), with two SNPs located in COX6C and 14-3-3B1 protein. In addition to 14-3-3B1 protein, a gene coding for disabled homolog 2 (DAB2) was also involved in apoptotic process (Prunier and Howe 2005). Two SNP-harboring genes, phosphatidylinositol glycan anchor biosynthesis class U (PIGU) and annexin A13, were involved in lipid metabolism. PIGU has functions in lipid metabolism including membrane lipid biosynthesis. This gene exhibited differential expression in porcine muscles divergent for intramuscular fat, which correlates positively with meat tenderness (Hamill et al 2013). Annexins are Ca^{2+}-dependent phospholipid-binding proteins that have an important role in the cell cycle and apoptosis (Mirsaeidi et al 2016). The list also includes a cell adhesion receptor, nicotinamide riboside kinase 2, that modulates myogenic differentiation (Li et al 1999). Single-marker analysis did not identify SNPs in significant association with variation in protein content.

Table 5. SNP markers significantly associated with phenotypic variability in shear force using a single SNP marker analysis. SNPs are sorted according to their R^2 values.

RefSeq	CHR	SNP position	Gene ID	Strand	Gene annotation	Region	R^2	UNADJ	BONF
NC_035084.1	8	5559245	LOC110529177	-	ryanodine receptor 3	CDS	0.034	1.08E-07	2.69E-03
NC_035083.1	7	27549740	LOC110527679	-	cytochrome c oxidase subunit 6C-1	5'UTR	0.031	2.10E-07	5.22E-03
NC_035083.1	7	27591587	LOC100136192	-	14-3-3B1 protein	3'UTR	0.031	2.10E-07	5.22E-03
NC_035083.1	7	28432371	LOC110527692	+	rho GTPase-activating protein 15	3'UTR	0.031	2.10E-07	5.22E-03
NC_035081.1	5	39867207	LOC110523898	+	disabled homolog 2	3'UTR	0.026	2.24E-07	5.57E-03
NC_035083.1	7	24880637	pigu	-	phosphatidylinositol glycan anchor biosynthesis class U	3'UTR	0.031	2.37E-07	5.89E-03
NC_035081.1	5	41247573	tcp4	-	Activated RNA polymerase II transcriptional coactivator p15	CDS	0.025	3.53E-07	8.79E-03
NC_035081.1	5	39901639	LOC110523901	+	uncharacterized LOC110523901	LncRNA	0.011	8.72E-07	2.17E-02
NC_035104.1	28	34338597	LOC110509080	-	annexin A13	CDS	0.020	9.39E-07	2.34E-02
NC_035081.1	5	39866812	LOC110523898	+	disabled homolog 2	CDS	0.008	9.42E-07	2.34E-02
NC_035081.1	5	56278406	LOC110524193	-	nicotinamide riboside kinase 2	3'UTR	0.030	1.96E-06	4.88E-02

Altogether, results obtained from the single SNP analyses provided additional evidence of RYR3 role in regulating phenotypic variability in fillet firmness. Also, single-marker

analysis highlighted a role for a few more genes in fillet firmness. However, estimating the effect of each SNP individually does not allow the detection of small effects of multiple joint SNPs. This may explain the inconsistency in the significant peaks between the single-marker analysis and the WssGBLUP approach. Several studies indicated that the SNP-joint analysis is more successful than the single-SNP analysis in GWA studies of complex traits(Fridley and Biernacka 2011, Lu et al 2015). Therefore, WssGBLUP approach is assumed to be more effective in dissecting the genetic architecture of the studied traits and providing putative markers that can be used for selection purposes.

CONCLUSIONS

The current GWA analyses identified novel genomic loci with a role in regulating muscle firmness and protein content. These genomic loci code for proteins involved in calcium homeostasis, transcriptional and chromatin regulators, cell adhesion, protein synthesis/degradation, and apoptotic processes. The top windows affecting the additive genetic variance in protein content and shear force appeared on chromosome 4 and 13, respectively. RYR3 was the major gene harboring the largest number of SNPs located within windows affecting the additive genetic variance in shear force and protein content. Abnormal calcium homeostasis in muscle cells accelerates postmortem protein degradation, and meat softness (Barbut et al 2008). The current study revealed that WssGBLUP, using 50 adjacent SNP windows, provided putative markers that could be used to estimate breeding values for firmness and protein content.

Author 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 and MS wrote the manuscript. All authors reviewed and approved the publication.

REFERENCES

Al-Tobasei R, Paneru B, Salem M (2016). Genome-Wide Discovery of Long Non-Coding RNAs in Rainbow Trout. *PLoS One* **11:** e0148940.

Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B *et al* (2017). Identification of SNPs associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-Seq samples in rainbow trout. *BMC Genomics* **18:** 582.

Ali A, Al-Tobasei R, Kenney B, Leeds TD, Salem M (2018). Integrated analysis of lncRNA and mRNA expression in rainbow trout families showing variation in muscle growth and fillet quality traits. *Sci Rep* **8:** 12111.

Anttila K, Jarvilehto M, Manttari S (2008). Ca2+ handling and oxidative capacity are greatly impaired in swimming muscles of hatchery-reared versus wild Atlantic salmon (Salmo salar). *Can J Fish Aquat Sci* **65:** 10-16.

Aussanasuwannakul A, Kenney PB, Brannan RG, Slider SD, Salem M, Yao J (2010). Relating instrumental texture, determined by variable-blade and Allo-Kramer shear attachments, to sensory analysis of rainbow trout, Oncorhynchus mykiss, fillets. *J Food Sci* **75:** S365-374.

Bahuaud D, Gaarder M, Veiseth-Kent E, M T (2010). Fillet texture and protease activities in different families of farmed Atlantic salmon (Salmo salar L.). *Aquaculture* **310:** 213-220.

Bai X, Meng G, Zheng X (2012). Cloning, purification, crystallization and preliminary X-ray diffraction analysis of mouse PACSIN 3 protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68:** 159-162.

Baldwin RLt, Li RW, Li CJ, Thomson JM, Bequette BJ (2012). Characterization of the longissimus lumborum transcriptome response to adding propionate to the diet of growing Angus beef steers. *Physiol Genomics* **44:** 543-550.

Barbut S, Sosnicki AA, Lonergan SM, Knapp T, Ciobanu DC, Gatcliffe LJ *et al* (2008). Progress in reducing the pale, soft and exudative (PSE) problem in pork and poultry meat. *Meat Sci* **79:** 46-63.

Belitz HD, Grosch W, Schieberle P (2009). *Food chemistry*: Berlin: Springer.

Bernard S, Eilers M (2006). Control of cell proliferation and growth by Myc proteins. *Results Probl Cell Differ* **42:** 329-342.

Black BL, Olson EN (1998). Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu Rev Cell Dev Biol* **14:** 167-196.

Blomstrand E, Eliasson J, Karlsson HK, Köhnke R (2006). Branched-chain amino acids activate key enzymes in protein synthesis after physical exercise. *J Nutr* **136:** 269S-273S.

Bonneau M, Lebret B (2010). Production systems and influence on eating quality of pork. *Meat Sci* **84:** 293-300.

Brès V, Yoshida T, Pickle L, Jones KA (2009). SKIP interacts with c-Myc and Menin to promote HIV-1 Tat transactivation. *Mol Cell* **36:** 75-87.

Castañeda MP, Hirschler EM, Sams AR (2005). Early postmortem carcass trim effects on the tenderness of broiler breast fillets. *Poult Sci* **84:** 951-954.

Choi SH, Wright JB, Gerber SA, Cole MD (2010). Myc protein is stabilized by suppression of a novel E3 ligase complex in cancer cells. *Genes Dev* **24:** 1236-1241.

Cuajungco MP, Grimm C, Oshima K, D'hoedt D, Nilius B, Mensenkamp AR *et al* (2006). PACSINs bind to the TRPV4 cation channel. PACSIN 3 modulates the subcellular localization of TRPV4. *J Biol Chem* **281:** 18753-18762.

Dang CG, Cho SH, Sharma A, Kim HC, Jeon GJ, Yeon SH *et al* (2014). Genome-wide Association Study for Warner-Bratzler Shear Force and Sensory Traits in Hanwoo (Korean Cattle). *Asian-Australas J Anim Sci* **27:** 1328-1335.

Delbarre-Ladrat C, Chéret R, Taylor R, Verrez-Bagnis V (2006). Trends in postmortem aging in fish: understanding of proteolysis and disorganization of the myofibrillar structure. *Crit Rev Food Sci Nutr* **46:** 409-421.

Destefanis G, Brugiapaglia A, Barge MT, Dal Molin E (2008). Relationship between beef consumer tenderness perception and Warner-Bratzler shear force. *Meat Sci* **78:** 153-156.

Diao H, Xiao S, Li R, Zhao F, Ye X (2013). Distinct spatiotemporal expression of serine proteases Prss23 and Prss35 in periimplantation mouse uterus and dispensable function of Prss35 in fertility. *PLoS One* **8:** e56757.

Duckett SK, Snowder GD, Cockett NE (2000). Effect of the callipyge gene on muscle growth, calpastatin activity, and tenderness of three muscles across the growth curve. *J Anim Sci* **78:** 2836-2841.

Eilbracht J, Reichenzeller M, Hergt M, Schnölzer M, Heid H, Stöhr M *et al* (2004). NO66, a highly conserved dual location protein in the nucleolus and in a special type of synchronously replicating chromatin. *Mol Biol Cell* **15:** 1816-1832.

Florence L, Mireille C, Jérôme B, Laurent L, Françoise M, Edwige Q (2015). Selection for muscle fat content and triploidy affect flesh quality in pan-size rainbow trout, Oncorhynchus mykiss *Aquaculture* **448:** 569-577.

Fonseca LFS, Gimenez DFJ, Dos Santos Silva DB, Barthelson R, Baldi F, Ferro JA *et al* (2017). Differences in global gene expression in muscle tissue of Nellore cattle with divergent meat tenderness. *BMC Genomics* **18:** 945.

Fridley BL, Biernacka JM (2011). Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* **19:** 837-843.

Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE *et al* (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253:** 448-451.

Gampel A, Parker PJ, Mellor H (1999). Regulation of epidermal growth factor receptor traffic by the small GTPase rhoB. *Curr Biol* **9:** 955-958.

Ge W, Wolf A, Feng T, Ho CH, Sekirnik R, Zayer A *et al* (2012). Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol* **8:** 960-962.

Geng X, Liu S, Yao J, Bao L, Zhang J, Li C *et al* (2016). A Genome-Wide Association Study Identifies Multiple Regions Associated with Head Size in Catfish. *G3 (Bethesda)* **6:** 3389-3398.

Goes ES, Lara JA, Gasparino E, Del Vesco AP, Goes MD, Alexandre Filho L *et al* (2015). Pre-Slaughter Stress Affects Ryanodine Receptor Protein Gene Expression and the Water-Holding Capacity in Fillets of the Nile Tilapia. *PLoS One* **10:** e0129145.

Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB *et al* (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (Oncorhynchus mykiss). *Front Genet* **7:** 203.

Grześ B, Pospiech E, Iwańska E, Mikołajczak B, Łyczyński A, Koćwin-Podsiadła M *et al* (2017). The relationship between protein changes in porcine longissimus muscle at different courses of meat tenderisation. *European Food Research and Technology* **243:** 2025-2034.

Hamill RM, Aslan O, Mullen AM, O'Doherty JV, McBryan J, Morris DG *et al* (2013). Transcriptome analysis of porcine M. semimembranosus divergent in intramuscular fat as a consequence of dietary protein restriction. *BMC Genomics* **14:** 453.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS *et al* (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106:** 9362-9367.

Hinshaw JM (1999). Trout production: Feeds and feeding methods. *Southern Regional Aquaculture Center* **223**.

Hirata H, Watanabe T, Hatakeyama J, Sprague SM, Saint-Amant L, Nagashima A *et al* (2007). Zebrafish relatively relaxed mutants have a ryanodine receptor defect, show slow swimming and provide a model of multi-minicore disease. *Development* **134:** 2771-2781.

Holler CJ, Taylor G, Deng Q, Kukar T (2017). Intracellular Proteolysis of Progranulin Generates Stable, Lysosomal Granulins that Are Haploinsufficient in Patients with Frontotemporal Dementia Caused by. *eNeuro* **4**.

Hong SH, Chang SH, Cho KC, Kim S, Park S, Lee AY *et al* (2016). Endoplasmic reticulum-Golgi intermediate compartment protein 3 knockdown suppresses lung cancer through endoplasmic reticulum stress-induced autophagy. *Oncotarget* **7:** 65335-65347.

Hu G, Gu W, Bai QL, Wang BQ (2013). Estimation of genetic parameters for growth traits in a breeding program for rainbow trout (Oncorhynchus mykiss) in China. *Genet Mol Res* **12:** 1457-1467.

Huang dW, Sherman BT, Lempicki RA (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37:** 1-13.

Huang dW, Sherman BT, Lempicki RA (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44-57.

Jacobsen Á, Joensen H, Eysturskarð J (2017). Gaping and loss of fillet firmness in farmed salmon (Salmo salar L.) closely correlated with post-slaughter cleaning of the abdominal cavity. *Aquaculture Research* **48:** 321-331.

Jin Y, Zhou T, Geng X, Liu S, Chen A, Yao J *et al* (2017). A genome-wide association study of heat stress-associated SNPs in catfish. *Anim Genet* **48:** 233-236.

KAMEYAMA T, ETLINGER JD (1979). Calcium-dependent regulation of protein synthesis and degradation in muscle. *Nature* **279:** 344-346.

Kanayama HO, Tamura T, Ugai S, Kagawa S, Tanahashi N, Yoshimura T *et al* (1992). Demonstration that a human 26S proteolytic complex consists of a proteasome and multiple associated protein components and hydrolyzes ATP and ubiquitin-ligated proteins by closely linked mechanisms. *Eur J Biochem* **206:** 567-578.

Kapoor N, Gupta R, Menon ST, Folta-Stogniew E, Raleigh DP, Sakmar TP (2010). Nucleobindin 1 is a calcium-regulated guanine nucleotide dissociation inhibitor of G{alpha}i1. *J Biol Chem* **285:** 31647-31660.

Kashio Y, Nakamura K, Abedin MJ, Seki M, Nishi N, Yoshida N *et al* (2003). Galectin-9 induces apoptosis through the calcium-calpain-caspase-1 pathway. *J Immunol* **170:** 3631-3636.

Kause A, Paananen T, Ritola O, Koskinen H (2007). Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. *J Anim Sci* **85:** 3218-3227.

Kause A, Quinton C, Airaksinen S, Ruohonen K, Koskela J (2011). Quality and production trait genetics of farmed European whitefish, Coregonus lavaretus. *J Anim Sci* **89:** 959-971.

Kause A, Kiessling A, Martin SA, Houlihan D, Ruohonen K (2016). Genetic improvement of feed conversion ratio via indirect selection against lipid deposition in farmed rainbow trout (Oncorhynchus mykiss Walbaum). *Br J Nutr* **116:** 1656-1665.

Koohmaraie M (1992). The role of Ca(2+)-dependent proteases (calpains) in post mortem proteolysis and meat tenderness. *Biochimie* **74:** 239-245.

Kubota K, Furuse M, Sasaki H, Sonoda N, Fujita K, Nagafuchi A *et al* (1999). Ca(2+)-independent cell-adhesion activity of claudins, a family of integral membrane proteins localized at tight junctions. *Curr Biol* **9:** 1035-1038.

Leeds TD, Vallejo RL, Weber GM, Pena DG, Silverstein JS (2016). Response to five generations of selection for growth performance traits in rainbow trout (Oncorhynchus mykiss). *Aquaculture* **465:** 341-351.

Li J, Mayne R, Wu C (1999). A novel muscle-specific beta 1 integrin binding protein (MIBP) that modulates myogenic differentiation. *J Cell Biol* **147:** 1391-1398.

Liang L, Zhao M, Xu Z, Yokoyama KK, Li T (2003). Molecular cloning and characterization of CIDE-3, a novel member of the cell-death-inducing DNA-fragmentation-factor (DFF45)-like effector family. *Biochem J* **370:** 195-203.

Lin BL, Li A, Mun JY, Previs MJ, Previs SB, Campbell SG *et al* (2018). Skeletal myosin binding protein-C isoforms regulate thin filament activity in a Ca. *Sci Rep* **8:** 2604.

Liu S, Vallejo RL, Palti Y, Gao G, Marancik DP, Hernandez AG *et al* (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front Genet* **6:** 298.

Liu W, Quinto I, Chen X, Palmieri C, Rabin RL, Schwartz OM *et al* (2001). Direct inhibition of Bruton's tyrosine kinase by IBtk, a Btk-binding protein. *Nat Immunol* **2:** 939-946.

Lu ZH, Zhu H, Knickmeyer RC, Sullivan PF, Williams SN, Zou F *et al* (2015). Multiple SNP Set Analysis for Genome-Wide Association Studies Through Bayesian Latent Variable Selection. *Genet Epidemiol* **39:** 664-677.

Lyon AN, Pineda RH, Hao lT, Kudryashova E, Kudryashov DS, Beattie CE (2014). Calcium binding is essential for plastin 3 function in Smn-deficient motoneurons. *Hum Mol Genet* **23:** 1990-2004.

Ma Z, Siebert AP, Cheung KH, Lee RJ, Johnson B, Cohen AS *et al* (2012). Calcium homeostasis modulator 1 (CALHM1) is the pore-forming subunit of an ion channel that mediates extracellular Ca2+ regulation of neuronal excitability. *Proc Natl Acad Sci U S A* **109:** E1963-1971.

MacLennan DH, Phillips MS (1992). Malignant hyperthermia. *Science* **256:** 789-794.

Manor ML, Cleveland BM, Kenney PB, Yao J, Leeds T (2015). Differences in growth, fillet quality, and fatty acid metabolism-related gene expression between juvenile male and female rainbow trout. *Fish Physiology and Biochemistry* **41:** 533-547.

Martinez I, Wang PA, Slizyte R, Jorge A, Dahle SW ea (2011). Protein expression and enzymatic activities in normal and soft textured Atlantic salmon. *Food Chem* **126:** 140-148.

Mateescu RG, Garrick DJ, Reecy JM (2017). Network Analysis Reveals Putative Genes Affecting Meat Quality in Angus Cattle. *Front Genet* **8:** 171.

Michie I (2001). *Causes of downgrading in the salmon farming industry In Farmed Fish Quality*. Kestin, S.C., Warris, P. Eds.; Blackwell: Oxford.

Mirsaeidi M, Gidfar S, Vu A, Schraufnagel D (2016). Annexins family: insights into their functions and potential role in pathogenesis of sarcoidosis. *J Transl Med* **14:** 89.

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A *et al* (2014). Manual for BLUPF90 family of programs. University of Georgia: University of Georgia, Athens, USA.

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A *et al* (2018). *Manual for BLUPF90 family of programs*: Univ. Georg., Athens, USA.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., H. LD (2002). BLUPF90 and related programs (BGF90) [WWW Document], in Proceeding of 7th World Congress on Genetics Applied to Livestock Production (Montpellier).

Mueller P, Massner J, Jayachandran R, Combaluzier B, Albrecht I, Gatfield J *et al* (2008). Regulation of T cell survival through coronin-1-mediated generation of inositol-1,4,5-trisphosphate and calcium mobilization after T cell receptor triggering. *Nat Immunol* **9:** 424-431.

Murayama T, Kurebayashi N (2011). Two ryanodine receptor isoforms in nonmammalian vertebrate skeletal muscle: possible roles in excitation-contraction coupling and other processes. *Prog Biophys Mol Biol* **105:** 134-144.

Natsuga K, Nishie W, Shinkuma S, Arita K, Nakamura H, Ohyama M *et al* (2010). Plectin deficiency leads to both muscular dystrophy and pyloric atresia in epidermolysis bullosa simplex. *Hum Mutat* **31:** E1687-1698.

Neguembor MV, Xynos A, Onorati MC, Caccia R, Bortolanza S, Godio C *et al* (2013). FSHD muscular dystrophy region gene 1 binds Suv4-20h1 histone methyltransferase and impairs myogenesis. *J Mol Cell Biol* **5:** 294-307.

Nobukuni Y, Mitsubuchi H, Akaboshi I, Indo Y, Endo F, Yoshioka A *et al* (1991). Maple syrup urine disease. Complete defect of the E1 beta subunit of the branched chain alpha-ketoacid dehydrogenase complex due to a deletion of an 11-bp repeat sequence which encodes a mitochondrial targeting leader peptide in a family with the disease. *J Clin Invest* **87:** 1862-1866.

Olmeda D, Moreno-Bueno G, Flores JM, Fabra A, Portillo F, Cano A (2007). SNAI1 is required for tumor growth and lymph node metastasis of human breast carcinoma MDA-MB-231 cells. *Cancer Res* **67:** 11721-11731.

Oo YW, Gomez-Hurtado N, Walweel K, van Helden DF, Imtiaz MS, Knollmann BC *et al* (2015). Essential Role of Calmodulin in RyR Inhibition by Dantrolene. *Mol Pharmacol* **88:** 57-63.

Ouali A, Gagaoua M, Boudida Y, Becila S, Boudjellal A, Herrera-Mendez CH *et al* (2013). Biomarkers of meat tenderness: present knowledge and perspectives in regards to our current understanding of the mechanisms involved. *Meat Sci* **95:** 854-870.

Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE *et al* (2015). Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing. *PLoS One* **10:** e0138435.

Paneru B, Ali A, Al-Tobasei R, Kenney B, Salem M (2018). Crosstalk among lncRNAs, microRNAs and mRNAs in the muscle 'degradome' of rainbow trout. *Sci Rep* **8:** 8416.

Paneru BD, Al-Tobasei R, Kenney B, Leeds TD, Salem M (2017). RNA-Seq reveals MicroRNA expression signature and genetic polymorphism associated with growth and muscle quality traits in rainbow trout. *Sci Rep* **7:** 9078.

Perni S, Marsden KC, Escobar M, Hollingworth S, Baylor SM, Franzini-Armstrong C (2015). Structural and functional properties of ryanodine receptor type 3 in zebrafish tail muscle. *J Gen Physiol* **145:** 253.

Prunier C, Howe PH (2005). Disabled-2 (Dab2) is required for transforming growth factor beta-induced epithelial to mesenchymal transition (EMT). *J Biol Chem* **280:** 17540-17548.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559-575.

Quinlan AR, Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841-842.

Ramayo-Caldas Y, Renand G, Ballester M, Saintilan R, Rocha D (2016). Multi-breed and multi-trait co-association analysis of meat tenderness and other meat quality traits in three French beef cattle breeds. *Genet Sel Evol* **48:** 37.

Rodrigues RT, Chizzotti ML, Vital CE, Baracat-Pereira MC, Barros E, Busato KC *et al* (2017). Differences in Beef Quality between Angus (Bos taurus taurus) and Nellore (Bos taurus indicus) Cattle through a Proteomic and Phosphoproteomic Approach. *PLoS One* **12:** e0170294.

Rosa AF, Moncau CT, Poleti MD, Fonseca LD, Balieiro JCC, Silva SLE *et al* (2018). Proteome changes of beef in Nellore cattle with different genotypes for tenderness. *Meat Sci* **138:** 1-9.

Ryskamp DA, Frye AM, Phuong TT, Yarishkin O, Jo AO, Xu Y *et al* (2016). TRPV4 regulates calcium homeostasis, cytoskeletal remodeling, conventional outflow and intraocular pressure in the mammalian eye. *Sci Rep* **6:** 30583.

Salem M, Nath J, Rexroad CE, Killefer J, Yao J (2005a). Identification and molecular characterization of the rainbow trout calpains (Capn1 and Capn2): their expression in muscle wasting during starvation. *Comp Biochem Physiol B Biochem Mol Biol* **140:** 63-71.

Salem M, Yao J, Rexroad CE, Kenney PB, Semmens K, Killefer J *et al* (2005b). Characterization of calpastatin gene in fish: its potential role in muscle growth and fillet quality. *Comp Biochem Physiol B Biochem Mol Biol* **141:** 488-497.

Salem M, Silverstein J, Rexroad CE, Yao J (2007). Effect of starvation on global gene expression and proteolysis in rainbow trout (Oncorhynchus mykiss). *BMC genomics* **8:** 328.

Salem M, Manor ML, Aussanasuwannakul A, Kenney PB, Weber GM, Yao J (2013). Effect of sexual maturation on muscle gene expression of rainbow trout: RNA-Seq approach. *Physiol Rep* **1:** e00120.

Salem M, Al-Tobasei R, Ali A, Lourenco D, Gao G, Palti Y *et al* (2018). Genome-Wide Association Analysis With a 50K Transcribed Gene SNP-Chip Identifies QTL Affecting Muscle Yield in Rainbow Trout. *Frontiers in Genetics* **9**.

Sato N, Maeda M, Sugiyama M, Ito S, Hyodo T, Masuda A *et al* (2015). Inhibition of SNW1 association with spliceosomal proteins promotes apoptosis in breast cancer cells. *Cancer Med* **4:** 268-277.

Schiele F, Artur Y, Varasteh A, Wellman M, Siest G (1989). Serum mitochondrial aspartate aminotransferase activity: not useful as a marker of excessive alcohol consumption in an unselected population. *Clin Chem* **35:** 926-930.

Schielzeth H, Husby A (2014). Challenges and prospects in genome-wide quantitative trait loci mapping of standing genetic variation in natural populations. *Ann N Y Acad Sci* **1320:** 35-57.

Schmidtke G, Kalveram B, Weber E, Bochtler P, Lukasiak S, Hipp MS *et al* (2006). The UBA domains of NUB1L are required for binding but not for accelerated degradation of the ubiquitin-like modifier FAT10. *J Biol Chem* **281:** 20045-20054.

Schrötter A, Pfeiffer K, El Magraoui F, Platta HW, Erdmann R, Meyer HE *et al* (2012). The amyloid precursor protein (APP) family members are key players in S-adenosylmethionine formation by MAT2A and modify BACE1 and PSEN1 gene expression-relevance for Alzheimer's disease. *Mol Cell Proteomics* **11:** 1274-1288.

Seoh ML, Ng CH, Yong J, Lim L, Leung T (2003). ArhGAP15, a novel human RacGAP protein with GTPase binding property. *FEBS Lett* **539:** 131-137.

Srougi MC, Burridge K (2011). The nuclear guanine nucleotide exchange factors Ect2 and Net1 regulate RhoB-mediated cell death after DNA damage. *PLoS One* **6:** e17108.

Strasburg GM, Chiang W (2009). Pale, soft, exudative turkey--The role of ryanodine receptor variation in meat quality. *Poult Sci* **88:** 1497-1505.

Suzuki C, Takahashi K, Hayama S, Ishikawa N, Kato T, Ito T *et al* (2007). Identification of Myc-associated protein with JmjC domain as a novel therapeutic target oncogene for lung cancer. *Mol Cancer Ther* **6:** 542-551.

Tallant C, Marrero A, Gomis-Rüth FX (2010). Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim Biophys Acta* **1803:** 20-28.

Thomas PM, Pankhurst NW, Bremner A (2005). The effect of stress and exercise on post-mortem biochemistry of Atlantic salmon and rainbow trout. *Journal of Fish Biology* **54:** 1177 - 1196.

Toei M, Saum R, Forgac M (2010). Regulation and isoform function of the V-ATPases. *Biochemistry* **49:** 4715-4723.

Torgersen JS, Koppang EO, Stien LH, Kohler A, Pedersen ME, Mørkøre T (2014). Soft texture of atlantic salmon fillets is associated with glycogen accumulation. *PLoS One* **9:** e85551.

Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ *et al* (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16:** 969.

Turner SD (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*.

Verma SK, Alim A (2014). Differential activity of stanniocalcin in male and female fresh water teleost Mastacembelus armatus (Lacepede) during gonadal maturation. *PLoS One* **9:** e101439.

Wang G, Wang Q, Fan Y, He X (2017). Reticulocalbin 2 correlates with recurrence and prognosis in colorectal cancer. *Am J Cancer Res* **7:** 2169-2179.

Wysocka J, Myers MP, Laherty CD, Eisenman RN, Herr W (2003). Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* **17:** 896-911.

Yu H, You X, Li J, Zhang X, Zhang S, Jiang S *et al* (2018). A genome-wide association study on growth traits in orange-spotted grouper (Epinephelus coioides) with RAD-seq genotyping. *Science China Life sciences* **61:** 934-946.

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S *et al* (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* **9:** e1003520.

Zhong X, Wang X, Zhou T, Jin Y, Tan S, Jiang C *et al* (2017). Genome-Wide Association Study Reveals Multiple Novel QTL Associated with Low Oxygen Tolerance in Hybrid Catfish. *Mar Biotechnol (NY)* **19:** 379-390.

Zhou A, Ou AC, Cho A, Benz EJ, Huang SC (2008). Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 5' splice site selection. *Mol Cell Biol* **28:** 5924-5936.

Zientara-Rytter K, Subramani S (2016). Role of actin in shaping autophagosomes. *Autophagy* **12:** 2512-2515.

APPENDICES

Probe ID	SNP allele	miRNA	Range	No of tools	Variance %	RefSeq	CHR	SNP position	Strand	Gene ID	Gene annotation	Region
AX-171615066	AX-171615066B	mir-29b-3p	28:48	3	2.584846386	NC_035083.1	7	12213061	-	LOC110527434	ribose-phosphate pyrophosphokinase 2	3'UTR
AX-171615066	AX-171615066B	mir-29c	28:48	3	2.584846386	NC_035083.1	7	12213061	-	LOC110527434	ribose-phosphate pyrophosphokinase 2	3'UTR
AX-171615067	AX-171615067B	mir-221-3p	21:42	3	2.551697226	NC_035083.1	7	12213141	-	LOC110527434	ribose-phosphate pyrophosphokinase 2	3'UTR
AX-171615067	AX-171615067B	mir-221-5p	21:42	3	2.551697226	NC_035083.1	7	12213141	-	LOC110527434	ribose-phosphate pyrophosphokinase 2	3'UTR
AX-171641797	AX-171641797A	mir-430c-3p	35:55	4	2.4735777	NC_035083.1	7	14400358	-	LOC110527462	glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic-like	3'UTR
AX-174104128	AX-174104128A	mir-24a-4-5p	29:55	3	3.028904955	NC_035084.1	8	3617788	+	LOC110529289	rho-related GTP-binding protein RhoV-like	3'UTR
AX-171599923	AX-171599923B	mir-1388-3p	24:45	3	3.127901564	NC_035084.1	8	4319191	+	LOC110529171	EMILIN-1-like	3'UTR
AX-171626302	AX-171626302A	mir-130d	15:41	4	2.6247137	NC_035084.1	8	5389012	+	LOC110529308	KATNB1-like protein 1	3'UTR
AX-171643085	AX-171643085B	mir-200b-3p	14:35	4	3.829789342	NC_035086.1	10	33157280	+	LOC110533811	plastin-3	3'UTR
AX-171643085	AX-171643085B	mir-200c	14:35	3	3.829789342	NC_035086.1	10	33157280	+	LOC110533811	plastin-3	3'UTR
AX-171643085	AX-171643085B	mir-200c-5p	14:35	3	3.829789342	NC_035086.1	10	33157280	+	LOC110533811	plastin-3	3'UTR
AX-171643085	AX-171643085B	mir-429	14:35	3	3.829789342	NC_035086.1	10	33157280	+	LOC110533811	plastin-3	3'UTR
AX-171643085	AX-171643085B	mir-429-3p	14:35	3	3.829789342	NC_035086.1	10	33157280	+	LOC110533811	plastin-3	3'UTR
AX-171643085	AX-171643085B	mir-429-5p	14:35	3	3.829789342	NC_035086.1	10	33157280	+	LOC110533811	plastin-3	3'UTR
AX-171619048	AX-171619048B	mir-124b-5p	21:46	3	3.811492044	NC_035086.1	10	33400519	+	LOC110533822	type-2 angiotensin II receptor-like	3'UTR
AX-171619048	AX-171619048B	mir-124c-5p	21:46	3	3.811492044	NC_035086.1	10	33400519	+	LOC110533822	type-2 angiotensin II receptor-like	3'UTR
AX-171596818	AX-171596818B	mir-19a-5p	26:47	3	3.412604358	NC_035086.1	10	34183838	+	LOC110533839	protein YIPF6	3'UTR
AX-171622625	AX-171622625A	mir-205a	34:53	3	3.383331867	NC_035086.1	10	34861588	-	LOC110533854	TBC1 domain family member 8B-like	3'UTR
AX-171622625	AX-171622625A	mir-205b-1-3p	34:53	3	3.383331867	NC_035086.1	10	34861588	-	LOC110533854	TBC1 domain family member 8B-like	3'UTR
AX-171622625	AX-171622625A	mir-205b-5p	34:53	3	3.383331867	NC_035086.1	10	34861588	-	LOC110533854	TBC1 domain family member 8B-like	3'UTR
AX-171622623	AX-171622623B	mir-181b-5p	18:38	3	3.383827911	NC_035086.1	10	34913476	-	c10hxorf57	chromosome 10 CXorf57 homolog	3'UTR
AX-171634788	AX-171634788A	mir-122-5p	29:45	3	3.3652804	NC_035086.1	10	35138547	+	LOC110533863	nuclear fragile X mental retardation-interacting protein 2-like	3'UTR
AX-172548115	AX-172548115B	mir-18b-5p	13:39	3	3.351775848	NC_035086.1	10	35806564	-	LOC110533872	vascular endothelial zinc finger 1-like	3'UTR
AX-172560944	AX-172560944A	mir-9a-5p	9:35	3	3.3514595	NC_035086.1	10	35806606	-	LOC110533872	vascular endothelial zinc finger 1-like	3'UTR
AX-172560944	AX-172560944A	mir-9a-6-3p	9:35	3	3.3514595	NC_035086.1	10	35806606	-	LOC110533872	vascular endothelial zinc finger 1-like	3'UTR
AX-172560944	AX-172560944A	mir-9a-8-3p	9:35	3	3.3514595	NC_035086.1	10	35806606	-	LOC110533872	vascular endothelial zinc finger 1-like	3'UTR
AX-172560944	AX-172560944A	mir-9b-3p	9:35	3	3.3514595	NC_035086.1	10	35806606	-	LOC110533872	vascular endothelial zinc finger 1-like	3'UTR
AX-172560944	AX-172560944A	mir-9b-5p	9:35	3	3.3514595	NC_035086.1	10	35806606	-	LOC110533872	vascular endothelial zinc finger 1-like	3'UTR
AX-172560944	AX-172560944A	let-7b	23:44	3	3.3514595	NC_035086.1	10	35806606	-	LOC110533872	vascular endothelial zinc finger 1-like	3'UTR
AX-172554708	AX-172554708A	mir-2188-5p	14:37	3	2.230264884	NC_035089.1	13	43003254	+	LOC110486574	ras-related protein Rab-37-like	3'UTR
AX-171604094	AX-171604094A	mir-729-5p	21:41	3	2.232950745	NC_035089.1	13	43044831	+	LOC110485193	Na(+)/H(+) exchange regulatory cofactor NHE-RF1-like	3'UTR
AX-171609147	AX-171609147B	mir-125a-5p	18:41	4	2.153171321	NC_035089.1	13	43192101	+	LOC110486583	serine/arginine-rich splicing factor 2-like	3'UTR
AX-171609147	AX-171609147B	mir-125b-5p	18:41	4	2.153171321	NC_035089.1	13	43192101	+	LOC110486583	serine/arginine-rich splicing factor 2-like	3'UTR
AX-171609147	AX-171609147B	mir-3588	21:42	3	2.153171321	NC_035089.1	13	43192101	+	LOC110486583	serine/arginine-rich splicing factor 2-like	3'UTR
AX-171609148	AX-171609148A	mir-9341	21:43	3	2.229069684	NC_035089.1	13	43192669	-	LOC110486583	serine/arginine-rich splicing factor 2-like	3'UTR
AX-171632041	AX-171632041A	mir-33b-5p	24:41	3	2.899460901	NC_035089.1	13	44092279	+	LOC110486595	monocyte to macrophage differentiation factor	3'UTR
AX-171632041	AX-171632041A	let-7d-3p	32:53	3	2.899460901	NC_035089.1	13	44092279	+	LOC110486595	monocyte to macrophage differentiation factor	3'UTR
AX-171604509	AX-171604509A	mir-25-5p	23:51	3	3.032321208	NC_035089.1	13	44788649	+	LOC110486624	eukaryotic translation initiation factor 3 subunit D	3'UTR
AX-171629631	AX-171629631A	mir-24b	19:39	3	2.827162719	NC_035089.1	13	44806234	+	LOC110486626	GTPase IMAP family member 4-like	3'UTR
AX-171611403	AX-171611403A	mir-9404-5p	29:52	3	3.458521994	NC_035089.1	13	45619953	-	LOC110486660	serine/threonine-protein phosphatase alpha-2 isoform-like	3'UTR
AX-171611404	AX-171611404A	mir-130d	23:50	4	3.148106903	NC_035089.1	13	45620629	-	LOC110486660	serine/threonine-protein phosphatase alpha-2 isoform-like	3'UTR
AX-171611404	AX-171611404A	mir-129b	24:47	3	3.148106903	NC_035089.1	13	45620629	-	LOC110486660	serine/threonine-protein phosphatase alpha-2 isoform-like	3'UTR
AX-171611404	AX-171611404A	mir-129b-5p	24:47	3	3.148106903	NC_035089.1	13	45620629	-	LOC110486660	serine/threonine-protein phosphatase alpha-2 isoform-like	3'UTR
AX-171611405	AX-171611405B	mir-129b	22:45	3	2.977861733	NC_035089.1	13	45641167	+	LOC110486662	claudin-4-like	3'UTR
AX-171611405	AX-171611405B	mir-129b-5p	22:45	3	2.977861733	NC_035089.1	13	45641167	+	LOC110486662	claudin-4-like	3'UTR
AX-171611405	AX-171611405B	mir-10c-5p	33:54	3	2.977861733	NC_035089.1	13	45641167	+	LOC110486662	claudin-4-like	3'UTR
AX-171604417	AX-171604417B	mir-29b-3p	23:45	3	2.549795835	NC_035089.1	13	45826604	-	LOC110486679	TBC1 domain family member 10B-like	3'UTR
AX-171604417	AX-171604417A	mir-29c	23:45	3	2.549795835	NC_035089.1	13	45826604	-	LOC110486679	TBC1 domain family member 10B-like	3'UTR
AX-171604665	AX-171604665A	mir-129b	13:36	3	2.403048174	NC_035089.1	13	45913986	+	LOC110486684	DDB1- and CUL4-associated factor 7-like	3'UTR
AX-171604665	AX-171604665A	mir-129b-5p	13:36	3	2.403048174	NC_035089.1	13	45913986	+	LOC110486684	DDB1- and CUL4-associated factor 7-like	3'UTR
AX-171609135	AX-171609135B	mir-7132a-5p	36:55	3	2.100126594	NC_035089.1	13	46018191	-	LOC110486687	LIM domain-containing protein 2-like	3'UTR
AX-171609138	AX-171609138A	mir-375-3p	20:42	3	2.026688793	NC_035089.1	13	46019477	-	LOC110486687	LIM domain-containing protein 2-like	3'UTR
AX-171609138	AX-171609138A	mir-375-5p	20:42	3	2.026688793	NC_035089.1	13	46019477	-	LOC110486687	LIM domain-containing protein 2-like	3'UTR
AX-171609138	AX-171609138B	mir-19b-3p	31:38	3	2.026688793	NC_035089.1	13	46019477	-	LOC110486687	LIM domain-containing protein 2-like	3'UTR
AX-171609139	AX-171609139B	mir-124-6-5p	16:50	3	2.020014275	NC_035089.1	13	46046944	-	LOC110486688	ATP-dependent RNA helicase DDX42-like	3'UTR
AX-171609139	AX-171609139A	mir-124a	16:50	3	2.020014275	NC_035089.1	13	46046944	-	LOC110486688	ATP-dependent RNA helicase DDX42-like	3'UTR
AX-171609139	AX-171609139B	mir-124a-1-3p	16:50	3	2.020014275	NC_035089.1	13	46046944	-	LOC110486688	ATP-dependent RNA helicase DDX42-like	3'UTR
AX-171609139	AX-171609139B	mir-124b	16:50	3	2.020014275	NC_035089.1	13	46046944	-	LOC110486688	ATP-dependent RNA helicase DDX42-like	3'UTR
AX-171609139	AX-171609139A	mir-199b-3p	19:41	3	2.020014275	NC_035089.1	13	46046944	-	LOC110486688	ATP-dependent RNA helicase DDX42-like	3'UTR
AX-171609139	AX-171609139A	mir-199c-3p	19:41	3	2.020014275	NC_035089.1	13	46046944	-	LOC110486688	ATP-dependent RNA helicase DDX42-like	3'UTR
AX-171609139	AX-171609139A	mir-3604	19:41	3	2.020014275	NC_035089.1	13	46046944	-	LOC110486688	ATP-dependent RNA helicase DDX42-like	3'UTR
AX-171598937	AX-171598937A	mir-183-5p	29:51	3	2.256936632	NC_035104.1	28	9763927	-	LOC110508483	matrix metalloproteinase-14-like	3'UTR
AX-171598937	AX-171598937A	mir-3553	29:51	3	2.256936632	NC_035104.1	28	9763927	-	LOC110508483	matrix metalloproteinase-14-like	3'UTR

Appendix A: Created/deleted miRNA targets as a consequence of SNPs affecting genetic variance in shear force.

Category	Term	Count	%	P-Value	Fold Enrichment	Fisher Exact
GOTERM_BP_DIRECT	release of sequestered calcium ion into cytosol	2	2.9	4.40E-02	43.3	9.50E-04
GOTERM_BP_DIRECT	hemopoiesis	3	4.3	2.90E-02	11.1	2.50E-03
GOTERM_CC_DIRECT	calcium channel complex	2	2.9	1.10E-02	176.8	4.70E-05
GOTERM_CC_DIRECT	smooth endoplasmic reticulum	2	2.9	1.10E-02	176.8	4.70E-05
GOTERM_CC_DIRECT	sarcolemma	2	2.9	4.10E-02	47.2	8.00E-04
GOTERM_CC_DIRECT	sarcoplasmic reticulum membrane	2	2.9	3.30E-02	58.9	5.10E-04
GOTERM_CC_DIRECT	Z disc	2	2.9	8.70E-02	21.4	3.90E-03
GOTERM_MF_DIRECT	ryanodine-sensitive calcium-release channel activity	2	2.9	1.50E-02	126.4	9.80E-05
GOTERM_MF_DIRECT	calcium-induced calcium release activity	2	2.9	9.30E-03	210.6	2.90E-05
GOTERM_MF_DIRECT	calcium ion binding	6	8.7	6.90E-02	2.7	2.40E-02
INTERPRO	EF-hand domain	6	8.7	1.40E-03	7.2	1.90E-04
INTERPRO	EF-hand-like domain	6	8.7	3.00E-03	6	4.80E-04
INTERPRO	RNA recognition motif domain	5	7.2	6.90E-03	6.5	1.00E-03
INTERPRO	Ryanodine receptor Ryr	2	2.9	1.20E-02	161.3	5.70E-05
INTERPRO	Ryanodine receptor	2	2.9	1.20E-02	161.3	5.70E-05
INTERPRO	Ryanodine Receptor TM 4-6	2	2.9	1.50E-02	129	9.40E-05
INTERPRO	Nucleotide-binding, alpha-beta plait	5	7.2	1.10E-02	5.7	1.80E-03
INTERPRO	Ryanodine receptor-related	2	2.9	2.70E-02	71.7	3.40E-04
INTERPRO	RyR/IP3R Homology associated domain	2	2.9	2.70E-02	71.7	3.40E-04
INTERPRO	EF-Hand 1, calcium-binding site	4	5.8	3.10E-02	5.8	5.10E-03
INTERPRO	Inositol 1,4,5-trisphosphate/ryanodine receptor	2	2.9	2.40E-02	80.7	2.60E-04
INTERPRO	Intracellular calcium-release channel	2	2.9	2.40E-02	80.7	2.60E-04
INTERPRO	MIR motif	2	2.9	3.60E-02	53.8	6.10E-04
INTERPRO	Ion transport domain	3	4.3	6.70E-02	7	9.10E-03
SMART	RRM	5	7.2	5.10E-03	7	7.00E-04
SMART	MIR	2	2.9	3.60E-02	53.3	6.20E-04

Appendix B: DAVID functional annotation for genes harboring SNPs affecting the additive genetic variance in Shear force.

Probe ID	SNP allele	miRNA	Range	No of tools	Variance %	RefSeq	CHR	SNP position	Strand	Gene ID	Gene annotation	Region
AX-172546818	AX-172546818A	mir-124a	30:54	3	4.421235807	NC_035077.1	1	25956638	+	LOC110520691	myocyte-specific enhancer factor 2A-like	3'UTR
AX-171607505	AX-171607505A	mir-301c-3p	24:43	4	4.135149461	NC_035077.1	1	27191151	-	plin1	perilipin 1	3'UTR
AX-171607504	AX-171607504A	mir-9341	20:43	3	4.134564892	NC_035077.1	1	27191275	-	plin1	perilipin 1	3'UTR
AX-172561146	AX-172561146B	mir-200b-3p	17:41	3	4.196982555	NC_035077.1	1	28051731	+	LOC110520934	tropomyosin alpha-1 chain-like	3'UTR
AX-171625748	AX-171625748A	let-7d-3p	20:43	3	2.780320855	NC_035080.1	4	76616397	+	LOC110522552	ribosomal oxygenase 1-like	3'UTR
AX-171625748	AX-171625748A	let-7a-5p	20:43	3	2.780320855	NC_035080.1	4	76616397	+	LOC110522552	ribosomal oxygenase 1-like	3'UTR
AX-171601941	AX-171601941A	mir-211	19:41	3	3.293202605	NC_035080.1	4	77336859	-	LOC110522556	V-type proton ATPase subunit D-like	3'UTR
AX-171641551	AX-171641551A	mir-93a-5p	10:40	3	2.530240481	NC_035083.1	7	38918677	-	LOC110527901	bcl-2-like protein 1	3'UTR
AX-171618880	AX-171618880B	mir-301c-3p	36:58	3	2.101255546	NC_035083.1	7	39141647	+	LOC110527911	myosin heavy chain, cardiac muscle isoform-like	3'UTR
AX-171622258	AX-171622258A	mir-146d-3p	27:47	3	2.750151584	NC_035083.1	7	39146046	+	LOC110527911	myosin heavy chain, cardiac muscle isoform-like	3'UTR
AX-171615203	AX-171615203B	mir-7552a-5p	19:40	3	2.350722674	NC_035083.1	7	40255082	+	LOC110527932	cytochrome b-c1 complex subunit 1, mitochondrial-like	3'UTR
AX-171615202	AX-171615202A	mir-183-5p	20:45	3	2.444950785	NC_035083.1	7	40490236	-	LOC110527945	actin-related protein 2/3 complex subunit 4-like	3'UTR
AX-171615202	AX-171615202A	mir-3553	20:45	3	2.444950785	NC_035083.1	7	40490236	-	LOC110527945	actin-related protein 2/3 complex subunit 4-like	3'UTR
AX-172548169	AX-172548169B	mir-30c-5p	32:55	3	2.141673425	NC_035087.1	11	7818010	+	LOC110535249	guanine nucleotide-binding protein G(s) subunit alpha-like	3'UTR
AX-171644934	AX-171644934B	mir-153a-3p	33:53	3	2.442923695	NC_035087.1	11	8055323	+	LOC110536547	piezo-type mechanosensitive ion channel component 2-like	3'UTR
AX-171644934	AX-171644934B	mir-153a-5p	33:53	3	2.442923695	NC_035087.1	11	8055323	+	LOC110536547	piezo-type mechanosensitive ion channel component 2-like	3'UTR
AX-171644934	AX-171644934B	mir-153b-3p	33:53	3	2.442923695	NC_035087.1	11	8055323	+	LOC110536547	piezo-type mechanosensitive ion channel component 2-like	3'UTR

Appendix C: Created/deleted miRNA targets as a consequence of SNPs affecting genetic variance in protein content.

Category	Term	Count	%	P-Value	Fold Enrichment	Fisher Exact
KEGG_PATHWAY	Cardiac muscle contraction	4	6.2	8.30E-03	9.1	8.70E-04
UP_KEYWORDS	Transcription	6	9.4	4.20E-02	3.1	1.30E-02
UP_KEYWORDS	Transcription regulation	6	9.4	3.30E-02	3.3	9.40E-03
UP_KEYWORDS	Chromatin regulator	3	4.7	2.50E-02	12.1	2.00E-03
UP_KEYWORDS	Translocation	2	3.1	6.80E-02	28.1	2.30E-03
UP_KEYWORDS	Coiled coil	17	26.6	1.40E-02	1.9	7.40E-03
GOTERM_CC_DIRECT	cytoplasm	13	20.3	3.50E-02	1.8	1.80E-02
GOTERM_CC_DIRECT	nucleus	14	21.9	7.90E-02	1.6	4.70E-02
GOTERM_CC_DIRECT	cytoskeleton	4	6.2	6.30E-02	4.3	1.40E-02
GOTERM_BP_DIRECT	negative regulation of transcription, DNA-templated	3	4.7	3.50E-02	10	3.30E-03
KEGG_PATHWAY	Adrenergic signaling in cardiomyocytes	4	6.2	5.40E-02	4.4	1.20E-02
GOTERM_BP_DIRECT	covalent chromatin modification	3	4.7	2.80E-02	11.3	2.30E-03
GOTERM_MF_DIRECT	protein dimerization activity	4	6.2	1.50E-02	7.5	1.90E-03
GOTERM_CC_DIRECT	clathrin-coated vesicle	2	3.1	2.90E-02	66.1	4.00E-04
INTERPRO	Zinc finger, PHD-type	3	4.7	3.50E-02	10.1	3.30E-03
GOTERM_BP_DIRECT	transcription, DNA-templated	7	10.9	2.60E-02	3	8.40E-03
INTERPRO	Zinc finger, PHD-finger	3	4.7	2.80E-02	11.4	2.30E-03
INTERPRO	Zinc finger, PHD-type, conserved site	3	4.7	1.90E-02	13.9	1.30E-03
SMART	HLH	3	4.7	7.20E-02	6.6	1.00E-02
SMART	PHD	3	4.7	3.80E-02	9.4	3.90E-03
INTERPRO	Zinc finger, FYVE/PHD-type	3	4.7	7.80E-02	6.4	1.20E-02
GOTERM_BP_DIRECT	regulation of transcription, DNA-templated	10	15.6	1.70E-02	2.4	6.80E-03
INTERPRO	FERM conserved site	2	3.1	7.50E-02	25.4	2.80E-03
INTERPRO	Myc-type, basic helix-loop-helix (bHLH) domain	3	4.7	6.70E-02	7	9.20E-03
GOTERM_BP_DIRECT	chromatin remodeling	2	3.1	9.70E-02	19.2	4.80E-03
INTERPRO	Tropomyosin	2	3.1	1.70E-02	114.5	1.20E-04
GOTERM_MF_DIRECT	cytoskeletal protein binding	2	3.1	7.70E-02	24.4	3.00E-03

Appendix D: DAVID functional annotation for genes harboring SNPs affecting the additive genetic variance in protein content. Enriched terms involved in regulation of transcription and chromatin remodeling are highlighted in green and yellow, respectively.

CHAPTER IV

IDENTIFICATION OF GENOMIC LOCI ASSOCIATED WITH GROWTH IN RAINBOW TROUT

Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M: Identification of genomic loci associated with growth in rainbow trout. BMC Genomics

ABSTRACT

Growth is a major economic production trait in aquaculture. Improvements in growth performance will reduce time and cost to market-size. However, genes underlying growth have not been fully explored in rainbow trout. A 50K, gene-transcribed SNP chip was used to genotype 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-selective breeding program. Weighted single-step GBLUP (WssGBLUP) was used to perform 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 were identified and associated with bodyweight gain. 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, 46 SNPs were identified in association with bodyweight gain. The highest SNP associated with this trait was identified

in a gene coding for thrombospondin-1 (THBS1) ($R^2 = 0.09$). 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 genomics selection in breeding programs.

INTRODUCTION

Aquaculture is a growing agribusiness that enhances food security and increases economic opportunities, worldwide (Burbridge et al 2001). A key challenge for this industry is to sustain the increasing consumer demand for seafood (Fornshell 2002). Salmonid species have been extensively studied as cultured fish species due to their economic and nutritional value (Tsai et al 2015a). Growth performance, particularly the efficiency of converting feed to bodyweight gain, is one of the most economic important traits (Tsai et al 2015a). 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 (Gutierrez et al 2015, Wringe et al 2010). Thus, artificial selection for growth is plausible, allowing potential improvement through selective breeding programs (Gutierrez et al 2015).

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 (Gjedrem 1983, Gjedrem 1992). Selection on harvest weight can improve growth rate (Salem et al 2012) and flesh color, and reduce production cost (Dufflocqa et al 2016). Successful genetic programs depend on establishment of a base

population with natural genetic variations which help achieving 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 (Leeds et al 2016) at harvest. Further studies are required to understand the genetic basis of bodyweight gain for genetically improved strains in order to achieve fast/efficient production (Fornshell 2002).

Quantitative trait loci (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 (Tsai et al 2015a). 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 (Wringe et al 2010). However, classical QTL mapping has a number of limitations. Linkage analysis is time consuming and depends on association within a family, limiting the power to detect associations between markers and phenotypes of interest (Gutierrez et al 2015). In addition, the identified QTL encompasses several megabases which contain hundreds, if not thousands, of genes making it difficult to identify the causal gene in a QTL (Price 2006).

Genomic resources have been developed for rainbow trout, including release of the first genome assembly draft (Berthelot et al 2014) 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 construction of high

density genetic maps (Al-Tobasei et al 2017, Palti et al 2013) . About 90% of the genetic variation comes from SNPs that are highly adaptable to large-scale genotyping and therefore, most suitable for genome-wide association studies (Salem et al 2012). The rainbow trout genome was successfully used for calling variants (Al-Tobasei et al 2017), and these variants have been used to build a 50K transcribed gene SNP chip suitable for association mapping (Salem et al 2018). Genome-wide association (GWA) studies have been employed to test the association between SNP markers spread throughout the genome and complex quantitative traits of interest (Tsai et al 2015b). Owing to the drastic reduction in cost and time required for genotyping a large number of markers, GWA studies are replacing QTL linkage mapping (Schielzeth and Husby 2014). 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 (Tsai et al 2015b). Many GWA studies conducted on livestock species led to identification of genes and mutations associated with economic traits (Tsai et al 2015b). Recently, a few GWA studies have been implemented in aquaculture species (Tsai et al 2015b) including rainbow trout. These studies aimed to identify markers associated with body weight (Gonzalez-Pena et al 2016), fillet quality (Gonzalez-Pena et al 2016, Salem et al 2018), and disease resistance (Palti et al 2015). Growth traits are controlled by small-effect variants in the farmed Atlantic salmon (Yoshida et al 2017). 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 (Gonzalez-Pena et al 2016).

In this study, we used a 50K transcribed gene SNP chip, recently developed in our laboratory, to perform a GWA analyses (Salem et al 2018). 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 (Salem et al 2018), and fillet firmness and protein content (Ali et al 2019) in rainbow trout. The objective of this study was to use the 50K SNP array to identify large-effect QTL associated with growth rate that could be applied in genomic selection.

MATERIALS AND METHODS

Fish population, tissue sampling, and phenotypic traits

Fish population and sampling are described in detail in a previous publication (Al-Tobasei et al 2017). In brief, female fish from a growth-selected line developed at NCCCWA were used to perform the current GWA analyses. The breeding program was established in 2002 and has continued for 5 generations of selection. Full-sib families were produced from single-sire×single-dam matings over a 6-week period. In order to get all families hatched within a 3-week period, eggs were reared in spring-water and incubated at 7-13°C. Each family was separately reared in a 200-L tank at approximately 12.5°C to maintain pedigree information; fish were fed a standard commercial fishmeal-based diet (Zeigler Bros Inc., Gardners, PA). Fish were tagged with a passive integrated transponder (Avid Identification Systems Inc., Norco, CA) at ~5-months post-hatch. Following fish tagging, fish were reared together in 800-L communal tanks supplied with partially-recirculated spring water, at ~13°C, until ~13 months post-hatch. Fish were fed a standard

commercial fishmeal-based diet according to a previously described feeding schedule (Hinshaw 1999). Fish were starved for 5 days prior to harvest day.

Because this study was a part of a bigger study where lethal traits such as fillet yield were collected, fish were harvested from two consecutive generations produced in the NCCCWA growth-selective breeding program. For each harvest year, selected fish were randomly assigned to one of five harvest groups (~100 fish each) allowing one fish per family per harvest group. The five groups from each Year Class (YC) were sampled over five consecutive weeks (one group/week). The YC 2010 fish were harvested between 410- and 437-days post-hatch with a mean body weight of 985 g (SD = 239 g). Fish from the YC 2012 were harvested between 446- and 481-days post-hatch with a mean body weight of 1,803 g (SD = 305 g). About 100 mg/L of tricaine methane sulfonate (Tricaine-S, Western Chemical, Ferndale, WA) were used to anesthetize the fish. Phenotypic data of 789 fish from YC 2010 and YC 2012 were used for the GWA analyses. The pedigree-based heritability h^2 (h^2ped) for growth was estimated according to Zaitlen et al., (Zaitlen et al 2013).

SNP genotyping and quality control

A 50K transcribed gene SNP-chip was recently developed and used to identify QTL associated with muscle yield (Salem et al 2018), fillet firmness and protein content (Ali et al 2019). Source of all SNPs used to build up the SNP chip was previously described (Al-Tobasei et al 2017). Briefly, the chip has ~ 21K SNPs showing allelic imbalances with fish body weight, muscle yield, fat content, shear force, whiteness index, and susceptibility to Bacterial Cold Water Disease (BCWD) as we described before (Al-Tobasei et al 2017, Salem et al 2018). In addition, about 5K nonsynonymous SNPs along with ~24K SNPs

were added to the chip to make sure at least 2 SNPs per each SNP-harboring gene were included. In total, the SNP chip has 50,006 SNPs.

As described before, a total of 1,728 fish were used to asses quality of this Affymetrix SNP chip. Genotyped fish were obtained from the NCCCWA growth- and BCWD-selection lines (Salem et al 2018). The SNP chip and sample metrics were calculated. Assessment of quality control (QC) and filtration of samples/genotypes have been performed using the Affymetrix SNPolisher software at the default parameters (Liu et al 2015). A call rate of 0.97 and Dish QC (DQC) threshold of 0.82 were applied to filter out genotyped samples. Out of 1,728 fish genotyped by this chip, 789 samples 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 (Salem et al 2018). WssGBLUP allows genotyped and ungenotyped animals to be used at the same time, and integrates phenotype, genotype and pedigree information in a combined analysis using the following mixed model for single trait analysis:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{w} + \mathbf{e}$$

Where y is the vector of the phenotypes, b is the vector of fixed effects including harvest group and hatch year, $\mathbf{a}$, $\mathbf{w}$, and $\mathbf{e}$ are the vectors of direct additive genetic (i.e., animal effect), random family, and residual effects, respectively. The matrices $\mathbf{X}$, $\mathbf{Z}_1$, and $\mathbf{Z}_2$ are incidence matrices for the effects contained in $\mathbf{b}$, $\mathbf{a}$, and $\mathbf{w}$, respectively. The model

combines all the relationship information based on pedigree and genotypes into a single matrix ($\mathbf{H}^{-1}$).

$$\mathbf{H}^{-1}=\mathbf{A}^{-1}+\begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} \end{bmatrix}$$

where $\mathbf{H}^{-1}$ is the inverse of the realized relationship matrix ($\mathbf{H}$), $\mathbf{A}^{-1}$ is the inverse of the relationship matrix based on pedigree information, $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree relationship matrix for genotyped animals only, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix. The random family effect is uncorrelated and just accounts for the fact the animals within the same family were raised in a common environment, and the covariance structure is given by $\mathbf{I}\sigma_w^2$, where $\mathbf{I}$ is an identity matrix and σ_w^2 is the family variance.

AIREMLF90 (Misztal et al 2018) was used to estimate the variance components for the direct additive genetic effect, random family effect, and residual. The inbreeding value, accounted for in the construction of the inverse of inverse of the pedigree relationship matrix, was previously calculated using a pedigree data of 63,808 fish from five consecutive generations in the NCCCWA breeding program using INBUPGF90 (Misztal et al 2002, Salem et al 2018). Quality control of genomic data was performed using PREGSF90 (Misztal et al 2002) according to the following settings: MAF > 0.05, call rate for SNP and samples > 0.90, and HWE < 0.15. In total, 35,322 SNPs (70.6%) passed QC.

Two iterations were used in the current WssGBLUP analysis and all SNPs were assigned the same weight during the first iteration (i.e., weight = 1.0). SNP effects ($\hat{u}$) assessed in the first iteration were used to calculate weights during the second iteration as $\hat{u}^2 2p(1-p)$, where p represents the allele frequency in the current genotyped population.

Each iteration was performed in three steps: 1) SNPs were assigned weight 2) BLUPF90 was used to compute genomic estimated breeding values (GEBV) using $\mathbf{H}^{-1}$ (Misztal et al 2002). 3) POSTGSF90 (Misztal et al 2002) was used to compute SNP effects and weights using sliding windows of 50 adjacent SNPs. The window size was based on a specific number of adjacent SNPs rather than specific number of nucleotides (i.e. physical size) because SNPs were not evenly distributed over the whole genome. The qqman package in R was used to generate a Manhattan plot showing the proportion of additive genetic variance explained by the defined windows of SNPs (Turner 2014).

Single marker GWA analysis

PLINK software was used to perform single marker association GWA analysis (Purcell et al 2007). In order to use the linear model in PLINK for the GWA analysis, the phenotypic data were checked for normality using Kolmogorov-Smirnov and Shapiro-Wilk test. The linear regression model in PLINK, using --linear command, allows the use of multiple covariates to account for the population structure. The first two principal components, YC and harvest groups, were included as covariates in the model. The --assoc command was used to retrieve the R-squared values of association.

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 (Quinlan and Hall 2010, Salem et al 2018). SNPs were classified as genic or intergenic. Genic SNPs were classified as CDS, intronic, 5'UTR or 3'UTR SNPs. To perform gene enrichment analysis, SNP-harboring genes were uploaded to the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang et al 2009a, Huang et al 2009b).   In order to avoid counting

duplicated genes, Fisher Exact statistics was 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.

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 (Li et al 2018). Traditional selection, based on the phenotype, has been applied to select for growth traits resulting in relatively slow genetic improvement over generations (Salem et al 2012). The economic significance of growth to aquaculture encouraged several studies aimed at understanding the genetic basis/mechanisms underlying the phenotype (Li et al 2018). SNPs explain 90% of the genetic variations among individuals (Salem et al 2012); therefore, these markers are most suitable for genetic evaluation of breeding candidates in selection programs. Fish population used for the current GWA analysis had average bodyweight gain per day of 3.27 ± 0.96 (g). Variations in bodyweight gain among 789 fish individuals used for the current GWA analysis are shown in Figure (1). The estimated heritability for bodyweight gain was 0.30 ± 0.05 (Zaitlen et al 2013). In this study, we used a 50K SNP chip to perform GWA analyses and identify QTL associated with bodyweight gain, based on 50 SNP sliding windows and single-marker association analysis. The chip contains SNPs potentially associated with the body weight as we previously described (Al-Tobasei et al 2017, Salem et al 2018). For this reason, we did not include any fish used in building the SNP-chip for GWA analyses in this study.
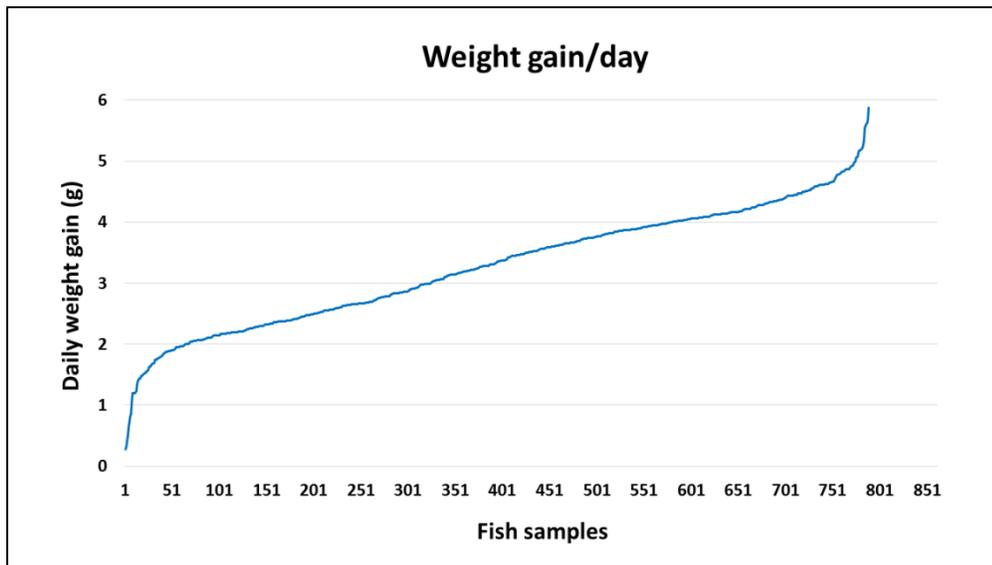
Figure 1. Variations in bodyweight gain among fish samples used in GWA analysis.

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 and 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 loci that harbor SNPs were clustered on 7 chromosomes (2, 4, 8, 9, 13, 14, and 18) (Figure 2).
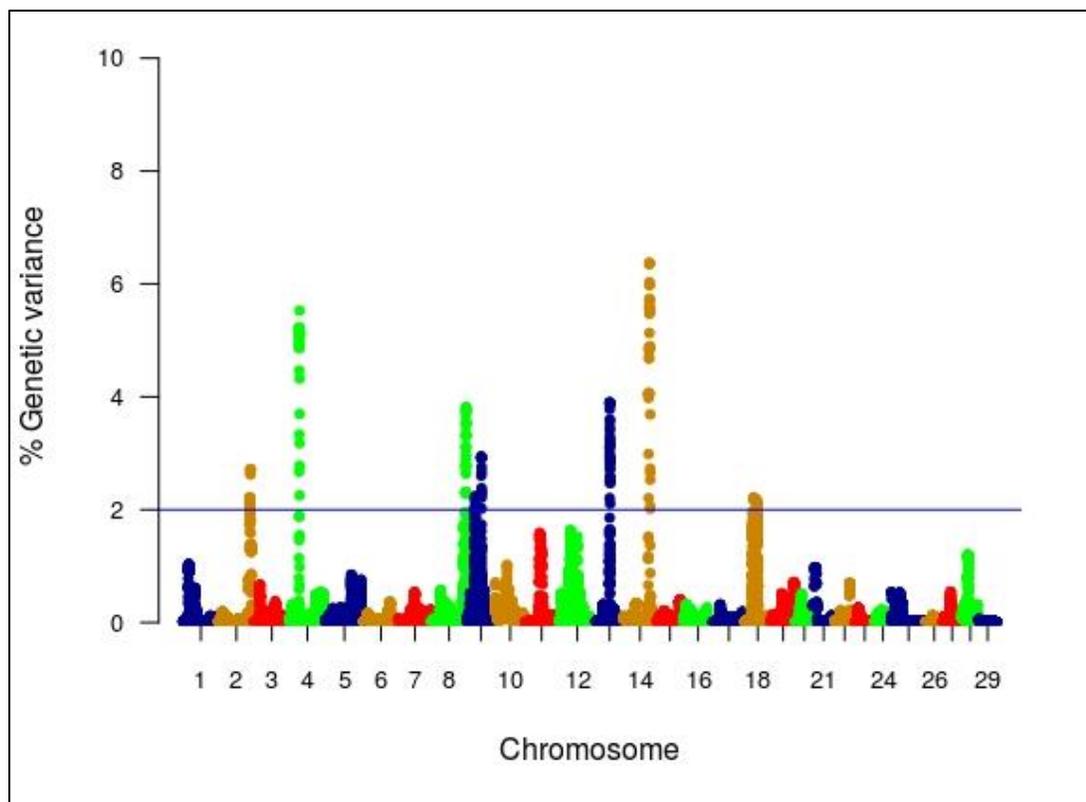
Figure 2. Manhattan plot showing association between genomic sliding windows of 50 SNPs and bodyweight gain. Chromosome 14 showed the highest peaks with genomic loci explaining up to 6.37% of the additive genetic variance. The basal blue line represents 2% of additive genetic variance explained by SNPs.

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 insights 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 (Appendix A & Table S2).

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

Coordinated hypertrophy and hyperplasia are essential for growing organisms (Goranov et al 2009). 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-β, probably, by enfolding TGF-β receptors in membrane invaginations (Le Saux et al 2008). Knockdown of CAV-1 had a tumor suppressing effect by inhibiting cell proliferation (Wang et al 2014), arresting cells in the G0/G1 phase, and inhibiting the expression of cell cycle-related proteins such as cyclin D1 (Wang et al 2014).

Table 1. SNP markers in 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(0)AT2	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	-	rbtstat3	Cell proliferation	Stat3	3'UTR
2.47	13	34488012	-	rbtstat3	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

Two SNPs were identified in each of TES and EIF4G2. TES negatively regulates cell proliferation and inhibits tumor cell growth (Tobias et al 2001, Wang et al 2004); whereas, eIF4G2 positively regulates cell growth and proliferation, prevents autophagy, and releases cells from nutrient sensing control by mTOR (Ramírez-Valle et al 2008). Each of SLC6A15 and KIF21A had a single SNP. Depletion of SLC6A15 attenuates leucine's effects to reduce the weight gain associated with a high fat diet (Drgonova et al 2013). KIF21A has been identified in association with growth in pigs (Fernández et al 2012). 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 (Yang et al 2006). Cancer cell proliferation (Okabe et al 2006) and growth of multifocal dysplastic lesions (Williams et al 2003) 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 (Shaulian and Karin 2002), which enhance expression of genes involved in growth (Shen et al 2008). Whereas, PRRC2C regulates the cell cycle and cell proliferation, and it controls the growth of lung cancer cells, *in vitro* (de Miguel et al 2014). 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 (Lynch et al 2018). We identified 2 SNPs on chromosome 9 in the gene coding for protein RCC2 homolog. RCC2 is a key regulator of cell cycle progression during the interphase (Yenjerla et al 2013). 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 (Wang and Faller 2008). 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 (Blazek et al 2011). STAT3 had two SNPs in the 3'UTR. Knockdown of STAT3 inhibits cell proliferation and leads to irreversible growth arrest (Sherry et al 2009).

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 (Mak et al 2014). FGFBP1 promotes fibroblast growth factor2 (FGF2) signaling during angiogenesis, tissue repair, and tumor growth (Tassi et al 2001). A single SNP was identified in the CCNA2 gene. This gene has a key role in cell cycle by regulating the initiation and progression of DNA synthesis (Arsic et al 2012). 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 (Levenson et al 2005). In addition, MCTS1 is involved in cell cycle progression by decreasing the length of the G1 phase without a reciprocal increase in other phases (Prosniak et al 1998). 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

arrest cell cycle (Kremer et al 2007). Over-expression of RPL36A leads to rapid cell cycling which enhances cell proliferation (Kim et al 2004). Of note, TNMD had a SNP in a window explaining 5.5% of the additive genetic variance. TNMD is essential for tenocyte proliferation and collagen fibril maturation (Docheva et al 2005). 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 (Ali et al 2018). Our results indicate a role for increased biomass and cell numbers in explaining variations in body weight.

SNPs in genes regulating lipid metabolism

Fatty acid synthesis is essential to meet the demand for phospholipids required for membrane expansion in growing cells (Lin et al 2013). We have identified 29 SNPs in 16 genes involved in lipid metabolism, explaining at least 2% of the additive genetic variance in bodyweight gain (Table 2). These SNPs spanned 5 chromosomes (4, 8, 13, 14, and 18). Chromosome 4 had 15 SNPs (56.6%) in 7 genes; peroxiredoxin 6 (PRDX6), phospholipid phosphatase 6 (PLPP6), vesicle associated membrane protein 4 (VAMP4), phosphatidylinositol Glycan, Class C (PIGC), disabled homolog 1 (DAB1), AMPK subunit alpha-2 (PRKAA2), and phospholipid phosphatase 3 (PLPP3).

Table 2. SNP markers in 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	23614046	-	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	81708446	-	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

Three SNPs were identified in the gene coding for PRDX6. The bifunctional enzyme,
PRDX6, regulates phospholipid turnover as well as protects against oxidative injury (Chen
et al 2000). A single 3'UTR SNP was identified in VAMP4 gene. This gene encodes a
protein implicated in the growth of lipid droplets in rainbow trout (Bou et al 2017). Also,
the DAB1 had a 3'UTR SNP.   DAB1 is associated with intramuscular fatty acid content

in pigs (Puig-Oliveras et al 2016). PRKAA2 harbored 3 SNPs located within windows that were among those explaining the highest genetic variability in bodyweight gain. AMPK regulates lipid metabolism by inhibiting the activity of key enzymes necessary for *de novo* biosynthesis of fatty acids and cholesterol (Aguan et al 1994) . PLPP3 had 5 SNPs in windows explaining ~5% of the additive genetic variance. This enzyme catalyzes the conversion of phosphatidic acid to diacylglycerol which is important to improve meat quality and lower body fat accumulation (Miklos et al 2011).

In total, 14 SNPs were identified on chromosomes 8, 13, 14, and 18. Chromosome 8 had three SNPs in 2 genes encoding acetyl-coenzyme A synthetase (ACSS2) and peroxisomal trans-2-enoyl-CoA reductase (PECR). ACSS2 activates acetate that can be used for lipid synthesis (Luong et al 2000). In addition, the PECR contributes to chain elongation of fatty acids (Das et al 2000). Chromosome 13 had 3 SNPs in genes coding for stAR-related lipid transfer protein 3 (STARD3) and ATP-citrate synthase (ACLY). STARD3 acts as a mediator of lipid metabolism, and is required for the growth and survival of cancer cells (Soccio and Breslow 2003). A single coding SNP was identified in a gene coding for ACLY. This enzyme has a crucial role in *de novo* biosynthesis of lipids and promoting tumor growth (Lin et al 2013). Six SNPs were identified on chromosome 14 in genes coding for electron transfer flavoprotein dehydrogenase (ETFDH), peptidylprolyl isomerase D (PPID), and galactosidase alpha (GLA). Four polymorphic sites were identified in ETFDH. Mutations in ETFDH gene leads to a disorder of fatty acid, amino acid and choline metabolism (Olsen et al 2007). An SNP was identified in PPID gene that has gene ontology (GO) terms belonging to lipid particle organization. In addition, we identified two SNPs on chromosome 18 in genes encoding AMPK subunit gamma-1

(PRKAG1) and oleoyl-ACP hydrolase. The latter enzyme contributes to release of free fatty acids from fatty acid synthase (Ritchie et al 2016). Moderate to high heritability for growth-related traits and fat content has been reported, implying existence of additive genetic variance in the fish population (Gonzalez-Pena et al 2016, Leeds et al 2012). In fish population YC 2010, fat content exhibited a moderate regression coefficient (R^2) value of 0.50 with WBW (Ali et al 2018). Many genes (n = 31) involved in lipid metabolic processes, including AMPK, were DE in fish families (YC 2010) showing contrasting WBW (Ali et al 2018). These results suggest a substantial role for fat content in explaining variations in body weight.

SNPs in genes regulating proteolytic activities

We identified 19 SNPs in 12 genes involved in proteolytic activities (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 variability (~6%) in this category. LAMP2 is essential during autophagy for the fusion of autophagosomes with lysosomes (Hubert et al 2016). ATP6V1H is a vacuolar (H+)-ATPase which is required to acidify the phagosome/ lysosome for proper processing (Trombetta et al 2003). GLA and NEU1 are lysosomal acid hydrolases (glycosidases) required to breakdown glycoproteins (Winchester 2005). NEU1 was associated with suppression of ovarian carcinoma (Chahal et al 2016).

Table 3. SNP markers in 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

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 metallopeptidase (OMIM). The OMIM is a protease essential for mitochondrial inner membrane proteostasis maintenance (Rainbolt et al 2016), and its deficiency leads to increased body weight and obesity (Quirós et al 2012). Plectin had two SNPs. Mutation in plectin results in muscular dystrophy (Natsuga et al 2010). 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) of extreme WBW phenotype (Ali et al 2018). 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 (Shamseldin et al 2015). Mutagenesis of SUCO leads to failure of osteoblast maturation, a decrease in synthesis of type I collagen, and eventually catastrophic defects in skeletal development (Sohaskey et al 2010). The gene encoding TMED5 has GO terms belonging to chromatin binding (Gaudet et al 2011). Knockdown of MYSM1, a histone H2A deubiquitinase, led to embryonic lethality and growth retardation (Huo et al 2018). 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 (Zhou et al 2016).

Table 4. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of the additive genetic variance in bodyweight gain and involved in 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	-	pigc	Development	phosphatidylinositol glycan anchor biosynthesis class C	CDS/syn
5.11	4	23200735	-	LOC110521628	Development	SUN domain-containing ossification factor	CDS/nonsyn
5.13	4	23209006	-	LOC110521628	Development	SUN domain-containing ossification factor	CDS/nonsyn
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/nonsyn
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/nonsyn
3.90	13	33571191	-	LOC110486234	Development	telethonin	3'UTR
3.89	13	33572642	-	LOC110486234	Development	telethonin	3'UTR
3.87	13	33572680	-	LOC110486234	Development	telethonin	3'UTR
3.89	13	33573086	-	LOC110486234	Development	telethonin	CDS/nonsyn
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 cis-trans isomerase FKBP10	CDS/syn
3.21	13	33863127	+	LOC110486247	Development	peptidyl-prolyl cis-trans isomerase FKBP10	CDS/syn
2.93	13	33864468	+	LOC110486247	Development	peptidyl-prolyl cis-trans isomerase FKBP10	CDS/syn
2.93	13	33866204	+	LOC110486247	Development	peptidyl-prolyl cis-trans 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/nonsyn
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/nonsyn
2.58	13	34218617	-	LOC110486259	Development|chromatin modification	histone acetyltransferase KAT2A	CDS/syn
4.06	14	60473623	+	rapgef2	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/nonsyn
5.49	14	62558595	+	LOC110488979	Development	ETS-related transcription factor Elf-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/nonsyn
2.04	18	29258567	-	LOC110496110	Chromatin modification	double-strand-break repair protein rad21 homolog	CDS/syn

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 (Gaudet et al 2011) whereas RALBP1 was involved in regulation of actin dynamics during embryogenesis (Boissel et al 2012). Knockdown of DHRS3 led to a phenotype with underdeveloped head structure and perturbed somitogenesis. 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 (Gaudet et al 2011). Four SNPs were identified in TCAP. TCAP-null mice exhibit abnormal myofiber size variation and increased levels of TCAP binding protein, myostatin (Markert et al 2010). 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 (Gruenwald et al 2014). 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 (Lietman et al 2014). CNP had the highest number of SNPs on chromosome 13. This protein regulates blood supply to the developing embryo (Cameron et al 1996). KAT2A encodes a protein that acts as a histone H3 succinyltransferase and exhibits a role in tumor cell proliferation and development (Wang et al 2017). KAT2A is involved in regulation of developmental processes by mediating acetylation of TBX5 (Ghosh et al

2018). 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 variability in bodyweight gain (~6.4%) followed by ELF1 (~5.5%). Lacking both OCRL and its paralog (Inpp5b) led to early lethality of mice embryos (Bernard and Nussbaum 2010). ELF1 has a role in maintaining cell polarity during development (Mishra et al 1998). 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 (Gaudet et al 2011). Sixty-three genes involved in development were DE in fish families (YC 2010) exhibiting divergent WBW phenotypes (Ali et al 2018). In agreement with a recent GWA study in rainbow trout (Reis Neto et al 2019), 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 general linear regression analysis available in PLINK, which allows for multiple covariates (Purcell et al 2007). In this study, we have identified 46 SNPs spread over 13 chromosomes with potential impact on the bodyweight gain (Bonferroni-corrected p < 2.01E-06; Appendix B, Table S3 & Figure 3). About one quarter of the identified SNPs (26.09%) spanned chromosome 15. Genes identified by PLINK were involved in growth,

cell proliferation, proteolysis, calcium ion binding, and lipid metabolism. Genes explaining the highest variability in bodyweight gain are coding for thrombospondin-1 (THBS1), adrenomedullin (ADM), microtubule-associated protein 4 (MAP4), malate dehydrogenase, mitochondrial-like (MDH2), poly [ADP-ribose] polymerase 6 (PARP6), EH domain-binding protein 1 (EHD1), and calsyntenin 1 (Table 5). THBS1 and ADM, ranked at the top of the list, explained ~9% and 7% of the variability in bodyweight gain, respectively. THBS1 is involved in complex biological processes including angiogenesis and tissue development (Resovi et al 2014). Mutation in THBS1 was associated with vascular permeability, accounting for embryonic lethality (Shamseldin et al 2015). ADM affects cell growth (Peláez et al 2017) and stimulates lipolysis (Dong et al 2017). Absence of ADM in mice led to 100% embryo lethality (Peláez et al 2017). MAP4 had 7 SNPs spanning ~21Kb on chromosome 15. MAP4 stabilizes microtubules (Nguyen et al 1997) and contributes to myogenesis (Mangan and Olmsted 1996). MDH2 had a synonymous SNP. This key metabolic enzyme has a role in embryonic development (Epstein et al 1969) and regulation of adipogenic differentiation (Kim et al 2012). ASNP was identified in a 3'UTR of a gene coding for PARP6. PARP6 is a negative regulator of cell proliferation and suppresses tumor growth by controlling the cell cycle (Tuncel et al 2012). EHD1 had two SNPs; one of them is nonsynonymous. This protein contributes to myoblast fusion and myogenesis (Posey et al 2011), and it regulates lipid droplet storage (Naslavsky et al 2007). Calsyntenin 1 had an SNP explaining ~3% of bodyweight gain. Two intronic SNPs in the calsyntenin 1 gene affect fillet yield and weight in rainbow trout (Gonzalez-Pena et al 2016).
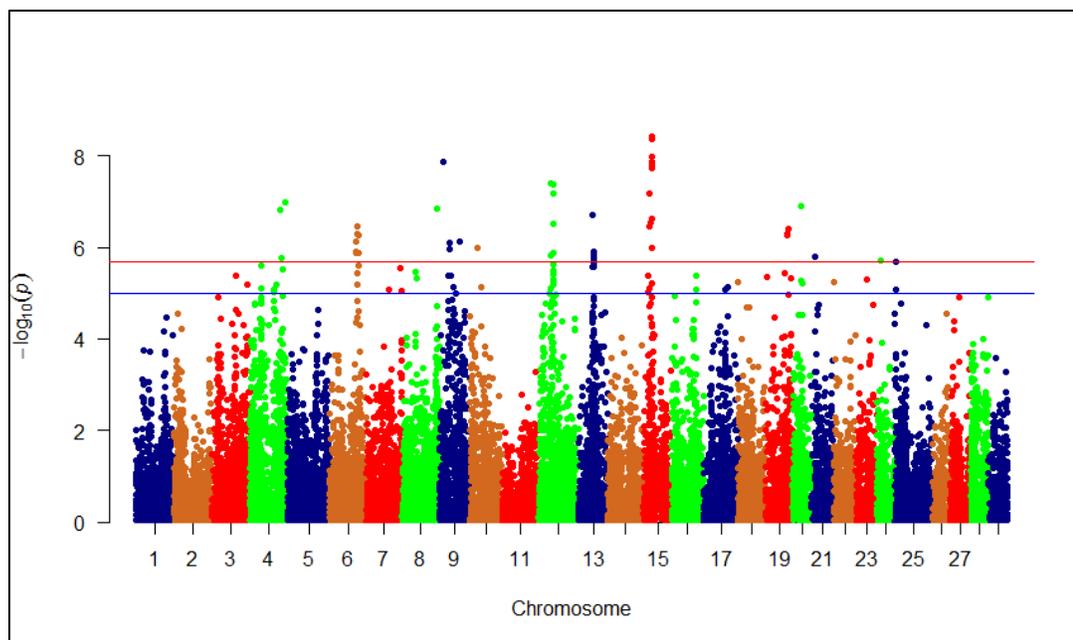
Figure 3. Manhattan plot showing single SNP markers associated with variations in bodyweight gain. Blue and red horizontal lines represent suggestive and significance threshold p-values of 1e-05 and 2.01e-06, respectively.

Three nonsynonymous SNPs were identified in genes coding for cell cycle progression protein 1 (CCPG1) and collagenase-3 (MMP13) (Table 5). CCPG1 is a positive regulator of the cell cycle and cell proliferation (Faghihi et al 2010). MMP13 plays a critical role in the skeletal system development (Inada et al 2004). A single SNP was identified in a gene encoding death-associated protein kinase 3 (Table 5). This protein is involved in regulation of autophagy (Bialik and Kimchi 2006). In addition, two SNPs were identified in a gene coding for 14-3-3 protein zeta (Table 5). This protein was mapped to different KEGG pathways including PI3K-Akt signaling pathway, in which 14-3-3 protein zeta regulates

the FOXO protein and cell cycle progression. The remaining SNPs associated with the variations in bodyweight gain are listed in Additional Table (S3).

Table 5. SNP markers significantly associated with bodyweight gain using single-SNP analysis. A color gradient on the left indicates phenotypic variability explained by a single SNP marker (green is the highest and red is the lowest). SNPs are sorted according to their chromosome positions.

R²	CHR	position	Gene ID	Strand	Gene annotation	Region|effect	UNADJ	BONF
0.09	4	79714730	LOC110522588	+	thrombospondin-1	CDS|syn	1.04E-07	2.59E-03
0.05	6	60367801	LOC110526356	+	poly [ADP-ribose] polymerase 6	3'UTR	1.27E-06	3.17E-02
0.07	6	60962934	LOC110526366	+	adrenomedullin 1	CDS|syn	1.29E-06	3.22E-02
0.03	6	63769566	LOC110526427	+	cell cycle progression protein 1	CDS|nonsyn	5.05E-07	1.26E-02
0.01	8	77245636	LOC110530779	-	death-associated protein kinase 3-like	CDS|syn	1.39E-07	0.003458
0.03	9	20731576	LOC110531861	+	EH domain-binding protein 1	CDS|nonsyn	8.10E-07	2.02E-02
0.03	9	20731964	LOC110531861	+	EH domain-binding protein 1	CDS|syn	1.07E-06	2.66E-02
0.03	9	44719009	clstn1	-	calsyntenin 1	CDS|syn	7.40E-07	1.84E-02
0.02	12	27438119	LOC110537437	-	collagenase 3	CDS|nonsyn	4.00E-08	9.96E-04
0.01	12	27439606	LOC110537437	-	collagenase 3	CDS|nonsyn	1.47E-06	3.65E-02
0.05	12	33984250	LOC110537548	+	malate dehydrogenase, mitochondrial	CDS|syn	4.13E-08	1.03E-03
0.06	15	17706212	LOC110489787	-	microtubule-associated protein 4	3'UTR	1.07E-08	2.65E-04
0.06	15	17706284	LOC110489787	-	microtubule-associated protein 4	3'UTR	1.58E-08	3.94E-04
0.06	15	17706417	LOC110489787	-	microtubule-associated protein 4	3'UTR	4.34E-09	1.08E-04
0.05	15	17706541	LOC110489787	-	microtubule-associated protein 4	3'UTR	1.80E-08	4.47E-04
0.05	15	17706838	LOC110489787	-	microtubule-associated protein 4	3'UTR	1.32E-08	3.29E-04
0.06	15	17717951	LOC110489787	-	microtubule-associated protein 4	CDS|syn	1.58E-08	3.92E-04
0.06	15	17728093	LOC110489788	-	microtubule-associated protein 4	5'UTR	3.66E-09	9.11E-05
0.01	19	48964202	LOC110498109	+	14-3-3 protein zeta-like	CDS|syn	5.08E-07	0.01263
0.02	19	48964211	LOC110498109	+	14-3-3 protein zeta-like	CDS|syn	5.32E-07	0.01323

Single SNP GWA analysis provided an additional set of SNPs, potentially, regulating variability in bodyweight gain. The two GWA approaches, adopted in the current study, revealed major roles of genes related to development and lipid metabolism in regulating weight gain. Routine use of single-SNP and multi-maker for GWA analysis has been recommended to take advantage of entire genotype information (Lorenz et al 2010).

Dividing the genome into chromosomal segments/windows, defined by 50 adjacent markers, outperformed the single-marker analysis in describing the genetic architecture of the studied trait. The recombinational history of the QTL and nearby markers determines the information content of haplotypes (Lorenz et al 2010). 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 (Reis Neto et al 2019). However, QTL associated with growth rate varied between the studies, and this discrepancy may be due to testing 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; these authors 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 (Gonzalez-Pena et al 2016). However, this window was not identified in our study because, perhaps, we considered only windows explaining at least 2% of the additive genetic variance. Several markers, each explaining less than 0.1% of the variance, were identified to be associated with bodyweight in a GWA study for Atlantic salmon (Tsai et al 2015b). 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 (Salem et al 2018). In agreement with previous GWA studies, growth is multifactorial in nature, and growth-related genes regulate development, cell proliferation, energy metabolism and growth (Peng et al 2016, Reis Neto et al 2019). 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 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-harboring genes were 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.

Author 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 and MS wrote the manuscript.

REFERENCES

Aguan K, Scott J, See CG, Sarkar NH (1994). 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* **149:** 345-350.

Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B *et al* (2017). Identification of SNPs associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-Seq samples in rainbow trout. *BMC Genomics* **18:** 582.

Ali A, Al-Tobasei R, Kenney B, Leeds TD, Salem M (2018). Integrated analysis of lncRNA and mRNA expression in rainbow trout families showing variation in muscle growth and fillet quality traits. *Sci Rep* **8:** 12111.

Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M (2019). Genome-Wide Association Study Identifies Genomic Loci Affecting Filet Firmness and Protein Content in Rainbow Trout. *Frontiers in Genetics* **10**.

Arsic N, Bendris N, Peter M, Begon-Pescia C, Rebouissou C, Gadéa G *et al* (2012). A novel function for Cyclin A2: control of cell invasion via RhoA signaling. *J Cell Biol* **196:** 147-162.

Bernard DJ, Nussbaum RL (2010). X-inactivation analysis of embryonic lethality in Ocrl wt/-; Inpp5b-/- mice. *Mamm Genome* **21:** 186-194.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B *et al* (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* **5:** 3657.

Bialik S, Kimchi A (2006). The death-associated protein kinases: structure, function, and beyond. *Annu Rev Biochem* **75:** 189-210.

Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z *et al* (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25:** 2158-2172.

Boissel L, Fillatre J, Moreau J (2012). Identification and characterization of the RLIP/RALBP1 interacting protein Xreps1 in Xenopus laevis early development. *PLoS One* **7:** e33193.

Bou M, Montfort J, Le Cam A, Rallière C, Lebret V, Gabillard JC *et al* (2017). Gene expression profile during proliferation and differentiation of rainbow trout adipocyte precursor cells. *BMC Genomics* **18:** 347.

Burbridge P., Hendrick V., Roth E., H. R (2001). Social and economic policy issues relevant to marine aquaculture. *J Appl Ichthyol* **17:** 194–206.

Cameron VA, Aitken GD, Ellmers LJ, Kennedy MA, Espiner EA (1996). The sites of gene expression of atrial, brain, and C-type natriuretic peptides in mouse fetal development: temporal changes in embryos and placenta. *Endocrinology* **137:** 817-824.

Chahal HS, Wu W, Ransohoff KJ, Yang L, Hedlin H, Desai M *et al* (2016). Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nat Commun* **7:** 12510.

Chen JW, Dodia C, Feinstein SI, Jain MK, Fisher AB (2000). 1-Cys peroxiredoxin, a bifunctional enzyme with glutathione peroxidase and phospholipase A2 activities. *J Biol Chem* **275:** 28421-28427.

Das AK, Uhler MD, Hajra AK (2000). Molecular cloning and expression of mammalian peroxisomal trans-2-enoyl-coenzyme A reductase cDNAs. *J Biol Chem* **275:** 24333-24340.

de Miguel FJ, Sharma RD, Pajares MJ, Montuenga LM, Rubio A, Pio R (2014). Identification of alternative splicing events regulated by the oncogenic factor SRSF1 in lung cancer. *Cancer Res* **74:** 1105-1115.

Docheva D, Hunziker EB, Fässler R, Brandau O (2005). Tenomodulin is necessary for tenocyte proliferation and tendon maturation. *Mol Cell Biol* **25:** 699-705.

Dong Y, Betancourt A, Belfort M, Yallampalli C (2017). Targeting Adrenomedullin to Improve Lipid Homeostasis in Diabetic Pregnancies. *J Clin Endocrinol Metab* **102:** 3425-3436.

Drgonova J, Jacobsson JA, Han JC, Yanovski JA, Fredriksson R, Marcus C *et al* (2013). Involvement of the neutral amino acid transporter SLC6A15 and leucine in obesity-related phenotypes. *PLoS One* **8:** e68245.

Dufflocqa P, Lhorenteb JP, Bangerab R, Neirac R, Newmand S, Yáñeza JM (2016). Correlated response of flesh color to selection for harvest weight in coho salmon (Oncorhynchus kisutch). *Aquaculture*.

Epstein CJ, Wegienka EA, Smith CW (1969). Biochemical development of preimplantation mouse embryos: in vivo activities of fructose 1,6-diphosphate aldolase, glucose 6-phosphate dehydrogenase, malate dehydrogenase, and lactate dehydrogenase. *Biochem Genet* **3:** 271-281.

Faghihi MA, Kocerha J, Modarresi F, Engström PG, Chalk AM, Brothers SP *et al* (2010). RNAi screen indicates widespread biological function for human natural antisense transcripts. *PLoS One* **5**.

Fernández AI, Pérez-Montarelo D, Barragán C, Ramayo-Caldas Y, Ibáñez-Escriche N, Castelló A *et al* (2012). Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip. *BMC Genet* **13:** 41.

Fornshell G (2002). Rainbow Trout — Challenges and Solutions. pp *545–557*.

Gaudet P, Livstone MS, Lewis SE, Thomas PD (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform* **12:** 449-462.

Ghosh TK, Aparicio-Sánchez JJ, Buxton S, Ketley A, Mohamed T, Rutland CS *et al* (2018). Acetylation of TBX5 by KAT2B and KAT2A regulates heart and limb development. *J Mol Cell Cardiol* **114:** 185-198.

Gjedrem T (1983). Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture* **33:** 51–71

Gjedrem T (1992). Breeding plans for rainbow trout. In: Gall GAE (ed). *The Rainbow Trout: Proceedings of the First Aquaculture-sponsored Symposium held at the Institute of Aquaculture, University of Sterling, Scotland.* pp 73–83.

Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB *et al* (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (Oncorhynchus mykiss). *Front Genet* **7:** 203.

Goranov AI, Cook M, Ricicova M, Ben-Ari G, Gonzalez C, Hansen C *et al* (2009). The rate of cell growth is governed by cell cycle stage. *Genes Dev* **23:** 1408-1422.

Gruenwald K, Castagnola P, Besio R, Dimori M, Chen Y, Akel NS *et al* (2014). Sc65 is a novel endoplasmic reticulum protein that regulates bone mass homeostasis. *J Bone Miner Res* **29:** 666-675.

Gutierrez AP, Yáñez JM, Fukui S, Swift B, Davidson WS (2015). Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (Salmo salar). *PLoS One* **10:** e0119730.

Hinshaw JM (1999). Trout production: Feeds and feeding methods. . *Southern Regional Aquaculture Center* **223**.

Huang dW, Sherman BT, Lempicki RA (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37:** 1-13.

Huang dW, Sherman BT, Lempicki RA (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44-57.

Hubert V, Peschel A, Langer B, Gröger M, Rees A, Kain R (2016). LAMP-2 is required for incorporating syntaxin-17 into autophagosomes and for their fusion with lysosomes. *Biol Open* **5:** 1516-1529.

Huo Y, Li BY, Lin ZF, Wang W, Jiang XX, Chen X *et al* (2018). MYSM1 Is Essential for Maintaining Hematopoietic Stem Cell (HSC) Quiescence and Survival. *Med Sci Monit* **24:** 2541-2549.

Inada M, Wang Y, Byrne MH, Rahman MU, Miyaura C, López-Otín C *et al* (2004). Critical roles for collagenase-3 (Mmp13) in development of growth plate cartilage and in endochondral ossification. *Proc Natl Acad Sci U S A* **101:** 17192-17197.

Kim EY, Kim WK, Kang HJ, Kim JH, Chung SJ, Seo YS *et al* (2012). Acetylation of malate dehydrogenase 1 promotes adipogenic differentiation via activating its enzymatic activity. *J Lipid Res* **53:** 1864-1876.

Kim JH, You KR, Kim IH, Cho BH, Kim CY, Kim DG (2004). Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma. *Hepatology* **39:** 129-138.

Kremer BE, Adang LA, Macara IG (2007). Septins regulate actin organization and cell-cycle arrest through nuclear accumulation of NCK mediated by SOCS7. *Cell* **130:** 837-850.

Le Saux CJ, Teeters K, Miyasato SK, Hoffmann PR, Bollt O, Douet V *et al* (2008). Down-regulation of caveolin-1, an inhibitor of transforming growth factor-beta signaling, in acute allergen-induced airway remodeling. *J Biol Chem* **283:** 5760-5768.

Leeds T, Kenney P, Manor M (2012). Genetic parameter estimates for feed intake, body composition, and fillet quality traits in a rainbow trout population selected for improved growth. *International Symposium on Genetics in Aquaculture*: Auburn, AL. p 259.

Leeds TD, Vallejo RL, Weber GM, Pena DG, Silverstein JS (2016). Response to five generations of selection for growth performance traits in rainbow trout (Oncorhynchus mykiss). *Aquaculture* **465:** 341-351.

Levenson AS, Thurn KE, Simons LA, Veliceasa D, Jarrett J, Osipo C *et al* (2005). MCT-1 oncogene contributes to increased in vivo tumorigenicity of MCF7 cells by promotion of angiogenesis and inhibition of apoptosis. *Cancer Res* **65:** 10651-10656.

Li N, Zhou T, Geng X, Jin Y, Wang X, Liu S *et al* (2018). Identification of novel genes significantly affecting growth in catfish through GWAS analysis. *Mol Genet Genomics* **293:** 587-599.

Lietman CD, Rajagopal A, Homan EP, Munivez E, Jiang MM, Bertin TK *et al* (2014). Connective tissue alterations in Fkbp10-/- mice. *Hum Mol Genet* **23:** 4822-4831.

Lin R, Tao R, Gao X, Li T, Zhou X, Guan KL *et al* (2013). Acetylation stabilizes ATP-citrate lyase to promote lipid biosynthesis and tumor growth. *Mol Cell* **51:** 506-518.

Liu S, Vallejo RL, Palti Y, Gao G, Marancik DP, Hernandez AG *et al* (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front Genet* **6:** 298.

Lorenz AJ, Hamblin MT, Jannink JL (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One* **5:** e14079.

Luong A, Hannah VC, Brown MS, Goldstein JL (2000). Molecular characterization of human acetyl-CoA synthetase, an enzyme regulated by sterol regulatory element-binding proteins. *J Biol Chem* **275:** 26458-26466.

Lynch JM, Dolman AJ, Guo C, Dolan K, Xiang C, Reda S *et al* (2018). Mutant myocilin impacts sarcomere ultrastructure in mouse gastrocnemius muscle. *PLoS One* **13:** e0206801.

Mak AB, Pehar M, Nixon AM, Williams RA, Uetrecht AC, Puglielli L *et al* (2014). Post-translational regulation of CD133 by ATase1/ATase2-mediated lysine acetylation. *J Mol Biol* **426:** 2175-2182.

Mangan ME, Olmsted JB (1996). A muscle-specific variant of microtubule-associated protein 4 (MAP4) is required in myogenesis. *Development* **122:** 771-781.

Markert CD, Meaney MP, Voelker KA, Grange RW, Dalley HW, Cann JK *et al* (2010). Functional muscle analysis of the Tcap knockout mouse. *Hum Mol Genet* **19:** 2268-2283.

Miklos R, Xu X, Lametsch R (2011). Application of pork fat diacylglycerols in meat emulsions. *Meat Sci* **87:** 202-205.

Mishra L, Cai T, Levine A, Weng D, Mezey E, Mishra B *et al* (1998). Identification of elf1, a beta-spectrin, in early mouse liver development. *Int J Dev Biol* **42:** 221-224.

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A *et al* (2018). *Manual for BLUPF90 family of programs*: Univ. Georg., Athens, USA.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., H. LD (2002). BLUPF90 and related programs (BGF90) [WWW Document], in Proceeding of 7th World Congress on Genetics Applied to Livestock Production (Montpellier).

Naslavsky N, Rahajeng J, Rapaport D, Horowitz M, Caplan S (2007). EHD1 regulates cholesterol homeostasis and lipid droplet storage. *Biochem Biophys Res Commun* **357:** 792-799.

Natsuga K, Nishie W, Shinkuma S, Arita K, Nakamura H, Ohyama M *et al* (2010). Plectin deficiency leads to both muscular dystrophy and pyloric atresia in epidermolysis bullosa simplex. *Hum Mutat* **31:** E1687-1698.

Nguyen HL, Chari S, Gruber D, Lue CM, Chapin SJ, Bulinski JC (1997). Overexpression of full- or partial-length MAP4 stabilizes microtubules and alters cell growth. *J Cell Sci* **110 (Pt 2):** 281-294.

Okabe H, Lee SH, Phuchareon J, Albertson DG, McCormick F, Tetsu O (2006). A critical role for FBXW8 and MAPK in cyclin D1 degradation and cancer cell proliferation. *PLoS One* **1:** e128.

Olsen RK, Olpin SE, Andresen BS, Miedzybrodzka ZH, Pourfarzam M, Merinero B *et al* (2007). ETFDH mutations as a major cause of riboflavin-responsive multiple acyl-CoA dehydrogenation deficiency. *Brain* **130:** 2045-2054.

Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E *et al* (2013). A resource of single-nucleotide polymorphisms for rainbow trout generated by RAD sequencing of doubled haploids. *Mol Ecol Resour*.

Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE *et al* (2015). Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing. *PLoS One* **10:** e0138435.

Peláez R, Niculcea M, Martínez A (2017). The Mammalian Peptide Adrenomedullin Acts as a Growth Factor in Tobacco Plants. *Front Physiol* **8:** 219.

Peng W, Xu J, Zhang Y, Feng J, Dong C, Jiang L *et al* (2016). Erratum: An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (Cyprinus carpio). *Sci Rep* **6:** 30101.

Posey AD, Pytel P, Gardikiotes K, Demonbreun AR, Rainey M, George M *et al* (2011). Endocytic recycling proteins EHD1 and EHD2 interact with fer-1-like-5 (Fer1L5) and mediate myoblast fusion. *J Biol Chem* **286:** 7379-7388.

Price AH (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci* **11:** 213-216.

Prosniak M, Dierov J, Okami K, Tilton B, Jameson B, Sawaya BE *et al* (1998). A novel candidate oncogene, MCT-1, is involved in cell cycle progression. *Cancer Res* **58:** 4233-4237.

Puig-Oliveras A, Revilla M, Castelló A, Fernández AI, Folch JM, Ballester M (2016). Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci Rep* **6:** 31803.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559-575.

Quinlan AR, Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841-842.

Quirós PM, Ramsay AJ, Sala D, Fernández-Vizarra E, Rodríguez F, Peinado JR *et al* (2012). Loss of mitochondrial protease OMA1 alters processing of the GTPase OPA1 and causes obesity and defective thermogenesis in mice. *EMBO J* **31:** 2117-2133.

Rainbolt TK, Lebeau J, Puchades C, Wiseman RL (2016). Reciprocal Degradation of YME1L and OMA1 Adapts Mitochondrial Proteolytic Activity during Stress. *Cell Rep* **14:** 2041-2049.

Ramírez-Valle F, Braunstein S, Zavadil J, Formenti SC, Schneider RJ (2008). eIF4GI links nutrient sensing by mTOR to cell proliferation and inhibition of autophagy. *J Cell Biol* **181:** 293-307.

Reis Neto RV, Yoshida GM, Lhorente JP, Yáñez JM (2019). Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (Oncorhynchus mykiss). *Mol Genet Genomics*.

Resovi A, Pinessi D, Chiorino G, Taraboletti G (2014). Current understanding of the thrombospondin-1 interactome. *Matrix Biol* **37:** 83-91.

Ritchie MK, Johnson LC, Clodfelter JE, Pemble CW, Fulp BE, Furdui CM *et al* (2016). Crystal Structure and Substrate Specificity of Human Thioesterase 2: INSIGHTS INTO THE MOLECULAR BASIS FOR THE MODULATION OF FATTY ACID SYNTHASE. *J Biol Chem* **291:** 3520-3530.

Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A *et al* (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS One* **7:** e36264.

Salem M, Al-Tobasei R, Ali A, Lourenco D, Gao G, Palti Y *et al* (2018). Genome-Wide Association Analysis With a 50K Transcribed Gene SNP-Chip Identifies QTL Affecting Muscle Yield in Rainbow Trout. *Frontiers in Genetics* **9**.

Schielzeth H, Husby A (2014). Challenges and prospects in genome-wide quantitative trait loci mapping of standing genetic variation in natural populations. *Ann N Y Acad Sci* **1320:** 35-57.

Shamseldin HE, Tulbah M, Kurdi W, Nemer M, Alsahan N, Al Mardawi E *et al* (2015). Identification of embryonic lethal genes in humans by autozygosity mapping and exome sequencing in consanguineous families. *Genome Biol* **16:** 116.

Shaulian E, Karin M (2002). AP-1 as a regulator of cell life and death. *Nat Cell Biol* **4:** E131-136.

Shen Q, Uray IP, Li Y, Krisko TI, Strecker TE, Kim HT *et al* (2008). The AP-1 transcription factor regulates breast cancer cell growth via cyclins and E2F factors. *Oncogene* **27:** 366-377.

Sherry MM, Reeves A, Wu JK, Cochran BH (2009). STAT3 is required for proliferation and maintenance of multipotency in glioblastoma stem cells. *Stem Cells* **27:** 2383-2392.

Soccio RE, Breslow JL (2003). StAR-related lipid transfer (START) proteins: mediators of intracellular lipid metabolism. *J Biol Chem* **278:** 22183-22186.

Sohaskey ML, Jiang Y, Zhao JJ, Mohr A, Roemer F, Harland RM (2010). Osteopotentia regulates osteoblast maturation, bone formation, and skeletal integrity in mice. *J Cell Biol* **189:** 511-525.

Tassi E, Al-Attar A, Aigner A, Swift MR, McDonnell K, Karavanov A *et al* (2001). Enhancement of fibroblast growth factor (FGF) activity by an FGF-binding protein. *J Biol Chem* **276:** 40247-40253.

Tobias ES, Hurlstone AF, MacKenzie E, McFarlane R, Black DM (2001). The TES gene at 7q31.1 is methylated in tumours and encodes a novel growth-suppressing LIM domain protein. *Oncogene* **20:** 2844-2853.

Trombetta ES, Ebersold M, Garrett W, Pypaert M, Mellman I (2003). Activation of lysosomal function during dendritic cell maturation. *Science* **299:** 1400-1403.

Tsai HY, Hamilton A, Guy DR, Tinch AE, Bishop SC, Houston RD (2015a). The genetic architecture of growth and fillet traits in farmed Atlantic salmon (Salmo salar). *BMC Genet* **16:** 51.

Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ *et al* (2015b). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16:** 969.

Tuncel H, Tanaka S, Oka S, Nakai S, Fukutomi R, Okamoto M *et al* (2012). PARP6, a mono(ADP-ribosyl) transferase and a negative regulator of cell proliferation, is involved in colorectal cancer development. *Int J Oncol* **41:** 2079-2086.

Turner SD (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*.

Wang R, Li Z, Guo H, Shi W, Xin Y, Chang W *et al* (2014). Caveolin 1 knockdown inhibits the proliferation, migration and invasion of human breast cancer BT474 cells. *Mol Med Rep* **9:** 1723-1728.

Wang S, Faller DV (2008). Roles of prohibitin in growth control and tumor suppression in human cancers. *Transl Oncogenomics* **3:** 23-37.

Wang Y, Guo YR, Liu K, Yin Z, Liu R, Xia Y *et al* (2017). KAT2A coupled with the α-KGDH complex acts as a histone H3 succinyltransferase. *Nature* **552:** 273-277.

Wang Z, Wei H, Yu Y, Sun J, Yang Y, Xing G *et al* (2004). Characterization of Ceap-11 and Ceap-16, two novel splicing-variant-proteins, associated with centrosome, microtubule aggregation and cell proliferation. *J Mol Biol* **343:** 71-82.

Williams TM, Cheung MW, Park DS, Razani B, Cohen AW, Muller WJ *et al* (2003). Loss of caveolin-1 gene expression accelerates the development of dysplastic mammary lesions in tumor-prone transgenic mice. *Mol Biol Cell* **14:** 1027-1042.

Winchester B (2005). Lysosomal metabolism of glycoproteins. *Glycobiology* **15:** 1R-15R.

Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG (2010). Growth-related quantitative trait loci in domestic and wild rainbow trout (Oncorhynchus mykiss). *BMC Genet* **11:** 63.

Yang K, Hitomi M, Stacey DW (2006). Variations in cyclin D1 levels through the cell cycle determine the proliferative fate of a cell. *Cell Div* **1:** 32.

Yenjerla M, Panopoulos A, Reynaud C, Fotedar R, Margolis RL (2013). TD-60 is required for interphase cell cycle progression. *Cell Cycle* **12:** 837-841.

Yoshida GM, Lhorente JP, Carvalheiro R, Yáñez JM (2017). Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (Salmo salar L.). *Anim Genet* **48:** 698-703.

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S *et al* (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* **9:** e1003520.

Zhou W, He Q, Zhang C, He X, Cui Z, Liu F *et al* (2016). BLOS2 negatively regulates Notch signaling during neural and hematopoietic stem and progenitor cell development. *Elife* **5**.

APPENDICES

Category	Term	Count	%	P-Value	Fold Enrichment	Fisher Exact
GOTERM_CC_DIRECT	lysosome	5	22.7	9.50E-05	19.2	4.90E-06
UP_KEYWORDS	Fatty acid biosynthesis	3	13.6	1.30E-03	54	2.30E-05
UP_KEYWORDS	Lysosome	4	18.2	2.30E-03	14.2	1.60E-04
UP_KEYWORDS	Lipid metabolism	5	22.7	8.70E-04	10.8	7.90E-05
GOTERM_MF_DIRECT	hydrolase activity	4	18.2	1.40E-03	16.8	8.20E-05
GOTERM_CC_DIRECT	lysosomal lumen	3	13.6	3.90E-03	30.6	1.20E-04
UP_KEYWORDS	Hydrolase	7	31.8	5.30E-03	3.9	1.30E-03
UP_KEYWORDS	Fatty acid metabolism	3	13.6	7.00E-03	22.6	3.00E-04
UP_KEYWORDS	Lipid biosynthesis	3	13.6	1.10E-02	18	5.90E-04
GOTERM_BP_DIRECT	fatty acid biosynthetic process	3	13.6	1.60E-03	48.4	3.10E-05
UP_KEYWORDS	Disease mutation	7	31.8	3.80E-02	2.6	1.40E-02
UP_KEYWORDS	Membrane	13	59.1	4.30E-02	1.6	2.50E-02
UP_KEYWORDS	Phosphoprotein	14	63.6	3.60E-02	1.6	2.20E-02
UP_KEYWORDS	Acetylation	8	36.4	4.70E-02	2.2	2.10E-02
UP_KEYWORDS	Mitochondrion inner membrane	3	13.6	3.10E-02	10.4	2.90E-03
UP_KEYWORDS	Ubiquinone	2	9.1	3.50E-02	53.5	6.40E-04
GOTERM_CC_DIRECT	cytoplasmic vesicle	3	13.6	2.70E-02	11.1	2.40E-03
UP_KEYWORDS	Glycosidase	2	9.1	8.00E-02	22.8	3.40E-03
GOTERM_BP_DIRECT	electron transport chain	2	9.1	7.90E-03	239.9	2.80E-05
UP_KEYWORDS	Transmembrane	10	45.5	9.40E-02	1.7	5.40E-02
UP_KEYWORDS	Transmembrane helix	10	45.5	9.20E-02	1.7	5.30E-02
GOTERM_BP_DIRECT	lipid biosynthetic process	2	9.1	1.60E-02	119.9	1.20E-04
GOTERM_CC_DIRECT	membrane	6	27.3	8.40E-02	2.4	3.40E-02
GOTERM_CC_DIRECT	mitochondrial membrane	2	9.1	9.80E-02	18.5	5.20E-03
GOTERM_CC_DIRECT	integral component of membrane	10	45.5	8.30E-02	1.7	4.70E-02
KEGG_PATHWAY	Lysosome	3	13.6	2.80E-02	10.7	2.50E-03
GOTERM_CC_DIRECT	mitochondrial inner membrane	3	13.6	8.30E-02	5.9	1.40E-02
GOTERM_BP_DIRECT	glycosphingolipid metabolic process	2	9.1	5.00E-02	37.3	1.30E-03
GOTERM_BP_DIRECT	intracellular signal transduction	3	13.6	7.50E-02	6.3	1.20E-02
KEGG_PATHWAY	Sphingolipid metabolism	2	9.1	9.80E-02	18.3	5.20E-03
GOTERM_BP_DIRECT	response to stress	2	9.1	6.70E-02	27.5	2.40E-03
UP_SEQ_FEATURE	active site:Proton acceptor	4	18.2	3.20E-02	5.4	5.60E-03
GOTERM_MF_DIRECT	receptor binding	3	13.6	6.50E-02	6.8	9.10E-03
UP_SEQ_FEATURE	mutagenesis site	6	27.3	7.10E-02	2.5	2.70E-02

Appendix A: DAVID functional annotation for genes harboring SNPs affecting the additive genetic variance for bodyweight gain. Enriched terms involved in lipid metabolic process and lysosomal activities are highlighted in yellow.

R2	CHR	SNP	UNADJ	BONF	position	Region	Strand	Gene ID	Gene annotation
0.01091	4	AX-171625486	1.56E-07	0.003879	68011538	CDS|syn	-	LOC110521066	utrophin-like
0.01617	4	AX-171607023	1.66E-06	0.04131	69989282	3'UTR	-	LOC110522434	peptidyl-prolyl cis-trans isomerase FKBP1B-like
0.08647	4	AX-171630496	1.04E-07	0.002591	79714730	CDS|syn	+	LOC110522588	thrombospondin-1-like
0.04684	6	AX-172556705	1.27E-06	0.03165	60367801	3'UTR	+	LOC110526356	poly [ADP-ribose] polymerase 6-like
0.01143	6	AX-172563666	7.48E-07	0.0186	60723828	N/A	N/A	N/A	N/A
0.06782	6	AX-171619750	1.29E-06	0.03217	60962934	CDS|syn	+	LOC110526366	ADM-like
0.0251	6	AX-171622208	5.05E-07	0.01257	63769566	CDS|nonsyn	+	LOC110526427	cell cycle progression protein 1-like
0.02126	6	AX-89924731	3.49E-07	0.008688	63769966	3'UTR	+	LOC110526427	cell cycle progression protein 1-like
0.02371	6	AX-171641164	1.33E-06	0.03316	63879610	3'UTR	-	LOC110526438	transcriptional repressor CTCF-like
0.02646	6	AX-171627451	5.35E-07	0.0133	64083708	CDS|syn	+	LOC110526452	aspartate aminotransferase, mitochondrial-like
0.01415	8	AX-171630631	1.39E-06	0.003458	77245636	CDS|syn	-	LOC110530779	death-associated protein kinase 3-like
0.02578	9	AX-171616254	1.32E-08	0.0003291	7679223	CDS|syn	-	LOC110531523	butyrophilin subfamily 1 member A1-like
0.03383	9	AX-171600786	8.10E-07	0.02015	20731576	CDS|nonsyn	+	LOC110531861	EH domain-binding protein 1-like protein 1
0.03326	9	AX-171617045	1.07E-06	0.02657	20731964	CDS|syn	+	LOC110531861	EH domain-binding protein 1-like protein 1
0.03336	9	AX-171599860	7.40E-07	0.01842	44719009	CDS|syn	-	clstn1	calsyntenin 1
0.02083	10	AX-171632649	1.05E-06	0.02623	15124739	N/A	N/A	N/A	N/A
0.02047	12	AX-89917012	4.00E-08	0.0009956	27438119	CDS|nonsyn	-	LOC110537437	collagenase 3-like
0.009093	12	AX-171608998	1.47E-06	0.03645	27439606	CDS|nonsyn	-	LOC110537437	collagenase 3-like
0.03069	12	AX-171621474	6.62E-08	0.001647	32632594	CDS|syn	-	LOC110537518	nucleolar protein 16-like
0.0001165	12	AX-171633719	3.04E-07	0.007559	33144189	CDS|syn	-	LOC110537529	ras-related protein Rab-24-like
0.01345	12	AX-171608755	1.35E-06	0.03368	33187010	3'UTR	+	LOC110537531	phosphoglycerate kinase-like
0.04754	12	AX-171608772	4.13E-08	0.001027	33984250	CDS|syn	+	LOC110537548	malate dehydrogenase, mitochondrial-like
0.006532	13	AX-174104527	1.91E-07	0.004744	33281265	3'UTR	-	LOC110486226	synaptic vesicle membrane protein VAT-1 homolog
0.006766	13	AX-171609128	1.37E-06	0.03411	34514689	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.006234	13	AX-171609127	1.71E-06	0.04263	34514726	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.006477	13	AX-171609121	1.71E-06	0.04248	34515780	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.007874	13	AX-171609117	1.26E-06	0.03138	34516041	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.006472	13	AX-171609115	1.77E-06	0.04401	34516555	3'UTR	-	LOC110486263	polymerase I and transcript release factor-like
0.007692	15	AX-171626582	3.36E-07	0.008367	11682367	CDS|syn	-	LOC110489660	PAB-dependent poly(A)-specific ribonuclease subunit PAN3-like
0.01128	15	AX-171604750	6.63E-08	0.00165	11758363	CDS|syn	+	rl21	60S ribosomal protein L21
0.01186	15	AX-171634933	2.91E-07	0.007239	16326142	3'UTR	+	LOC110489750	gamma-soluble NSF attachment protein-like
0.03656	15	AX-172547113	1.03E-06	0.02561	17317644	N/A	N/A	N/A	N/A
0.005578	15	AX-171609691	2.33E-07	0.005805	17590095	5'UTR	+	LOC110489782	myosin-7-like
0.05848	15	AX-171609715	1.07E-08	0.000265	17706212	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.05705	15	AX-171609716	1.58E-08	0.0003935	17706284	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.06168	15	AX-171609717	4.34E-09	0.0001079	17706417	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.04933	15	AX-171609719	1.80E-08	0.0004467	17706541	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.04727	15	AX-171609720	1.32E-08	0.0003293	17706838	3'UTR	-	LOC110489787	microtubule-associated protein 4-like
0.06099	15	AX-171609722	1.58E-08	0.0003919	17717951	CDS|syn	-	LOC110489787	microtubule-associated protein 4-like
0.06188	15	AX-171606782	3.66E-09	9.11E-05	17728093	5'UTR	N/A	LOC110489788	microtubule-associated protein 4-like
0.01398	19	AX-171599829	5.08E-07	0.01263	48964202	CDS|syn	+	LOC110498109	14-3-3 protein zeta-like
0.01638	19	AX-171599828	5.32E-07	0.01323	48964211	CDS|syn	+	LOC110498109	14-3-3 protein zeta-like
0.01812	19	AX-171636533	4.00E-07	0.00995	51608979	CDS|syn	+	LOC110498149	lysine-specific histone demethylase 1A-like
0.06264	20	AX-171644862	1.25E-07	0.003109	19516087	lnc_RNA	+	LOC110499159	uncharacterized LOC110499159
0.001551	21	AX-172547222	1.62E-06	0.04023	10349000	N/A	N/A	N/A	N/A
0.02373	24	AX-171626722	1.94E-06	0.0482	7400058	5'UTR	+	LOC110503649	large neutral amino acids transporter small subunit 4-like

Appendix B: All SNP markers significantly associated with bodyweight gain using single-SNP analysis. A color gradient on the left indicates phenotypic variability explained by a single SNP marker (green is the highest and red is the lowest).

CHAPTER V

GENOME-WIDE ASSOCIATION STUDY IDENTIFIES COMMON GENOMIC LOCI ASSOCIATED WITH MUSCLE FAT AND MOISTURE CONTENT IN RAINBOW TROUT

Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M

ABSTRACT

Genetic improvement of fillet quality attributes is a concern of aquaculture industry breeding programs, which must also meet safety, nutrition, and quality requirements. Muscle composition impacts quality attributes such as flavor, appearance, texture, and juiciness. Fat and moisture make up about 80% of the tissue weight. Genes underlying the fat and moisture content of muscle are still to be fully exploited in rainbow trout. A 50K gene transcribed SNP chip was used to genotype a total of 789 fish with available phenotypic data for fat and moisture content. Genotyped fish were obtained from two consecutive generations produced in the NCCWA growth-selective breeding program. Weighted single-step GBLUP (WssGBLUP) was utilized to perform genome-wide association (GWA) analysis to identify quantitative trait loci (QTL) affecting the studied traits. Using genomic sliding windows of 50 adjacent SNPs, 137 and 178 SNPs were identified as associated with the additive genetic variance for fat and moisture content, respectively. Chromosomes 19 and 29 harbored the highest number of SNPs explaining at least 2% of the genetic variation in fat and moisture content. A total of 61 common SNPs on chromosomes 19 and 29 affected the aforementioned traits; this association suggests common mechanisms underlying muscle fat and moisture content. Additionally, based on

single-marker GWA analyses, 13 and 27 SNPs were identified in association with fat and moisture content, respectively. SNP-harboring genes were mainly involved in lipid metabolism, cytoskeleton remodeling, and protein turnover. This work provided putative SNP markers that could be prioritized and used for genomic selection in breeding program

INTRODUCTION

Fillet quality traits have economic importance to the aquaculture industry (Sodeland et al 2013). For efficient aquaculture production, there is a need of fish fillet with high nutritional and organoleptic qualities. Consumer attitude towards fish is influenced by fillet quality aspects including intramuscular fat content and water holding capacity (Bonneau and Lebret 2010, Zhang et al 2013). Fish provide an important source of long chain omega 3 and other essential fatty acids for humans, which can lower risk of several chronic diseases (Zhao et al 2010). Additionally, quantity and quality of intramuscular lipid impact fillet quality attributes; particularly juiciness, flavor, color, texture, and shelf-life (Manor et al 2012).

Previous studies showed high correlation between fat, moisture, and protein contents, which together account for 99% of the muscle weight (Ang et al 1984). In mammals, the intramuscular fat content exhibited a significant negative correlation with moisture content (Li et al 2013, Watanabe et al 2018). In addition, simultaneous lower protein and moisture content along with increased fat content of the porcine meat was reported (Li et al 2013). In fish, previous studies showed that muscle mobilizes protein or fat, replenishing them with water depending on the energetic demands associated with various physiological conditions (Paneru et al 2018, Salem et al 2006, Salem et al 2013). Muscle degradation

accompanying sexual maturation in female rainbow trout caused a significant decline in muscle protein content associated with a concomitant increase in moisture content ($R^2 =$ 0.994, $p < 0.01$) (Salem et al 2006).   Meanwhile, fat content was not significantly affected by spawning in rainbow trout ($p > 0.05$), suggesting that intramuscular fat is not required to support gonadal maturation. Therefore, muscle protein catabolism was suggested as a source for energy during spawning, generating voids that were replaced by water (Salem et al 2006). However, recent studies in our laboratory revealed gravid fish, fed on a high-plane of nutrition, showed reduced intramuscular fat and increased abundance of transcripts involved in fatty acid degradation metabolic pathways compared to sterile fish. These results suggested that female trout approaching spawning more likely mobilize intramuscular fat in order to fuel gonadal maturation. Fertile fish exhibited decrease in fat content associated with concurrent increase in shear force, moisture and protein content of the muscle (Salem et al 2013). Most of the changes in muscle growth and quality attributes occur 2-3 months pre-spawning. Control of the muscle fat content in fish can enhance muscle quality attributes and aquaculture industry profitability in turn.

Fat content of the tissue is a lethally-measured trait which makes it difficult for the aquaculture industry to improve it. The aquaculture industry controls fat content of fillets by adjusting lipid content in the ration (Manor et al 2012). Rearing triploid fish has been adopted in rainbow trout to avoid loss of fillet quality due to sexual maturation (Paneru et al 2018). Selective breeding represents another approach to enhance phenotypic traits of interest. One aim in the breeding programs is to convert the feed into muscle rather than accumulating excess of visceral fats. Also, accumulating excessive lipids in the muscle makes the fillet processing difficult and reduces the fillet firmness (Sodeland et al 2013).

A two generation program of selection on fat was initiated in rainbow trout to produce lean and fat lines where the fat content increased by 15 to 31% in the fat line (Florence et al 2015). These lines were used as a model to study the effect of muscle fat content on the fillet quality (Florence et al 2015). Selection on fat content enhanced fillet color and texture (Florence et al 2015), feed conversion ratio (FCR), and protein-retention efficiency (Kause et al 2016). Also, five generations of family-based selection on body weight of rainbow trout was established at the USDA National Center of Cool and Cold Water Aquaculture (NCCCWA) (Leeds et al 2016). In fish population YC 2010, fat content showed a moderate regression coefficient (R^2) value of 0.50 with the whole body weight (WBW) (Ali et al 2018). Therefore, selection on body weight influenced the fat content in rainbow trout yielding heavier fish with more fat. Similarly, the gilthead sea bream exhibited 0.1% increase in fat content concomitant with 0.08% decline in moisture content per increment of ten grams in weight (García-Celdrán et al 2015). Muscle fat and moisture content showed moderate heritability in fish, including rainbow trout, implying existence of additive genetic variance in the fish population (Leeds et al 2012), and making genetic responses to selection possible. However, genetic architecture of fat and moisture content remains undiscovered in fish. Understanding the genetic basis of the phenotypic traits in question and development of fish strains of improved genetic gain will enhance the efficiency of breeding programs, and influence the aquaculture industry profitability and consumers' satisfaction.

Genome-wide association (GWA) studies help identify genetic variations affecting phenotypic variations and facilitate efficient implementation in breeding programs through genomic and marker-assisted selection. A few GWA studies have been conducted on

aquaculture species including salmonids. Some studies aimed to identify genomic loci responsible for the genetic variability in body weight (Gonzalez-Pena et al 2016), fillet quality (Gonzalez-Pena et al 2016, Salem et al 2018), and disease resistance (Vallejo et al 2017) traits. There were only two GWA studies performed in Atlantic salmon (Sodeland et al 2013) and common carp (Zheng et al 2016) to identify quantitative trait loci (QTL) associated with the muscle fat content. In Atlantic salmon, four significant SNPs affecting genetic variance for fat content were identified on chromosomes 9 and 10, based on 5,650 genome-wide distributed SNPs (Sodeland et al 2013). A high density 250K SNP array in common carp revealed eight genome-wide suggestive significance SNPs related to muscle fat content, whereas there were no SNPs surpassed the genome-wide significance level (Zheng et al 2016). The two studies did not identify major QTL affecting genetic variance for fat content in fish. To the best of our knowledge, no GWA studies have been implemented in rainbow trout to identify SNP markers associated with the genetic variance for fat and moisture content.

A 50K transcribed SNP-chip suitable for GWA analyses has been recently developed in our laboratory. The array has been used to identify large-effect QTL responsible for the genetic variance for fillet yield, firmness, protein content, and body weight gain (Ali et al 2019, Salem et al 2018). The chip outperformed a 57K genomic SNP panel (Gonzalez-Pena et al 2016) in identifying major QTL associated with genetic variability in fillet yield and bodyweight gain in fish populations of two consecutive generations produced from USDA growth-selection line. The current study aimed to identify major QTL affecting the additive genetic variance for fillet fat and moisture content in rainbow trout. We identified two major QTL common between the two traits on chromosomes 19 and 29. The identified

SNPs could be prioritized in estimating candidates for genomic selection in breeding programs.

MATERIALS AND METHODS

Ethics statement

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 approval #056).

Fish population, tissue sampling, and phenotypic traits

Fish population and tissue sampling were previously described (Al-Tobasei et al 2017). Briefly, diploid females from a growth-selected line at NCCWA were used in order to carry out the GWA analyses. This selective breeding program was initiated in 2004 by intercrossing seven domesticated strains of rainbow trout and has gone through five generations of selection for improved growth performance (Leeds et al 2016). Full-sib families collected from third- and fourth-generation fish (Year-class, YC, 2010 and YC 2012) were used for the current GWA analyses. Phenotypic data for muscle fat and moisture content were obtained from 789 fish representing 98 families from YC 2010 and 99 families from YC 2012. Full-sib families were produced from single-sire×single-dam matings over 6 weeks. Eggs from all families were reared in spring water, incubated at 7-13°C, and hatched within 3 weeks. A separate 200-L tank at ~12.5°C was used for each family in order to keep the pedigree information. Tagging of fish, at ~5-months post-hatch, with a passive integrated transponder (Avid Identification Systems Inc., Norco, CA) helped rearing fish together in 800-L communal tanks, at ~13°C, until ~13 months post-hatch. A

commercial fishmeal-based diet was provided to fish. The feeding schedule was previously described (Hinshaw 1999). Fish were starved for 5 days prior to harvest day to facilitate processing.

WBW was measured, and families were sorted according to the WBW. The 2nd or 3rd fish from each family was picked for sampling in order to keep the distribution of WBW around the median of each family. For each YC, fish were randomly assigned to one of five harvest groups allowing a single fish per family per harvest group (~100 fish each). The harvest groups were sampled over a five-weeks period (one group/week). Fish groups from the two year classes were harvested at different ages. Fish produced from the YC 2010 were harvested between 410- and 437-days post-hatch (mean body weight = 985 g; SD = 239 g). Whereas fish produced from YC 2012 were harvested between 446- and 481-days post-hatch (mean body weight = 1,803 g; SD = 305 g). Muscle fat and moisture content showed moderate regression coefficient (R^2) values of 0.23 and 0.38 with body weight, respectively. Fish were euthanized in a tricaine methane sulfonate (Tricaine-S, Western Chemical, Ferndale, WA), harvested, and eviscerated. Head-on gutted carcasses were shipped on ice to the West Virginia University Muscle Foods Processing Laboratory (Morgantown, WV) where carcasses were manually processed into skinless fillets (Salem et al 2013).

Proximate analyses, including crude lipid and moisture content, were previously described (Manor et al 2015). In brief, crude lipid and moisture analyses were performed using AOAC-approved methods (AOAC 2000). Crude lipid content was completed using Soxhlet extraction with petroleum ether. Moisture content was assessed by weighing the sample before and after drying at 110°C for 18 hrs. The pedigree-based heritability h^2

(h^2ped) for fat and moisture content were estimated according to Zaitlen et al., (Zaitlen et al 2013).

SNP genotyping and quality control

A 50K transcribed gene SNP-chip was recently developed and utilized in identifying genomic loci responsible for the additive genetic variance for fillet yield (Salem et al 2018). SNPs used to build the SNP chip were identified and reported in our previous publication (Al-Tobasei et al 2017). Briefly, the chip included ~21K SNPs showing potential allelic imbalances with the body weight, muscle yield, fat content, shear force, whiteness index, and susceptibility to Bacterial Cold Water Disease (BCWD) as previously described (Al-Tobasei et al 2017, Salem et al 2018). About 5K nonsynonymous SNPs were included in the chip. Additional SNPs were added to the chip in order to have a minimum of 2 SNPs per each SNP-harboring gene. The array contains a total of 50,006 SNPs.

A total of 1,728 fish were used for genotyping and quality assessment of the SNP chip. Genotyped fish were collected from the USDA/NCCCWA growth- and BCWD- selection lines (Salem et al 2018). The SNP chip and sample metrics were calculated. The Affymetrix SNPolisher software was used for QC assessment and filtration of samples/genotypes at the default settings (Liu et al 2015). Genotyped samples were filtered using a call rate of 0.97 and Dish QC (DQC) threshold of 0.82. In the current study, 789 genotyped fish having available phenotypic data for muscle fat and moisture content were used for the GWA analyses.

Fifty-SNP window GWA analysis

Weighted single-step GBLUP (WssGBLUP) was used to conduct the current GWA analysis as we previously described (Salem et al 2018). WssGBLUP allows use of

genotyped and ungenotyped animals in the model. The WssGBLUP integrates phenotypic data, genotype and pedigree information using a mixed model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{w} + \mathbf{e}$$

Where $\mathbf{y}$ is the vector of the phenotypes, $\mathbf{b}$ is the vector of fixed effects including harvest group and hatch year, $\mathbf{a}$ is the vector of additive direct genetic effects (i.e., animal effect), $\mathbf{w}$ is the vector of random family effect, and $\mathbf{e}$ is the residual error. The matrices $\mathbf{X}$, $\mathbf{Z}_1$, and $\mathbf{Z}_2$ are incidence matrices for the effects contained in $\mathbf{b}$, $\mathbf{a}$, and $\mathbf{w}$, respectively. The model combines all the relationship information (based on pedigree and genotypes) into a single matrix ($\mathbf{H}^{-1}$).

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

where $\mathbf{H}^{-1}$ is the inverse of the realized relationship matrix ($\mathbf{H}$), $\mathbf{A}^{-1}$ is the inverse of the relationship matrix based on pedigree information, $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree relationship matrix for genotyped animals only, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix. The random family effect is uncorrelated and just accounts for the fact the animals within the same family were raised in a common environment, and the covariance structure is given by $\mathbf{I}\sigma_w^2$, where $\mathbf{I}$ is an identity matrix and σ_w^2 is the family variance.

The variance components for the additive direct genetic effect, random family effect, and residuals were estimated using AIREMLF90 (Misztal et al 2018). Pedigree data of 63,808 fish from five consecutive generations produced in the NCCCWA breeding program were used to calculate the inbreeding value using INBUPGF90 (Misztal et al 2002, Salem et al 2018). PREGSF90 [21] was used for quality control (QC) of genomic

data according to the following parameters; MAF > 0.05, call rate > 0.90, and HWE < 0.15. A total of 35,322 SNPs (70.6%) passed QC and were used for the WssGBLUP analysis.

Two iterations were used in the current WssGBLUP analysis. All SNPs passing QC were assigned, by default, the same weight during the first iteration (i.e., weight = 1.0). In the second iteration, SNP weights were determined according to the SNP effects ($\hat{u}$) calculated in the first iteration as $\hat{u}^2 2p(1 - p)$, where p is the current allele frequency. In brief, three steps were conducted in each iteration: 1) weights were assigned to the SNPs. 2) genomic estimated breeding values (GEBV) were calculated, based on $\mathbf{H}^{-1}$, using BLUPF90 (Misztal et al 2002). 3) SNP effects and weights were computed by POSTGSF90 (Misztal et al 2002) using genomic sliding windows of 50 adjacent SNPs. This window size (n = 50 SNPs) was adopted instead of physical size (e.g. specific number of nucleotides) because SNPs were not evenly distributed over the whole genome. The qqman package (Turner 2014) in R was used to generate Manhattan plots showing the proportion of additive variance explained by the 50 SNP windows.

Single marker GWA analysis

Single marker association analysis was performed using PLINK as previously reported (Ali et al 2019, Purcell et al 2007). The studied phenotypes were checked for normality using Kolmogorov-Smirnov and Shapiro-Wilk test to make sure they meet the assumption of linear model analysis in PLINK (Purcell et al 2007). Multiple covariates were included in the linear model which accounted for population structure. The first two principal components (PCs) were included in the model to control the global inflation of the test statistic. The command, --assoc, was used to retrieve the R-squared values (R^2) of association between the quantitative traits and genotypes. Manhattan plots showing single

SNP markers associated with variations in muscle fat and moisture content were generated using qqman package (Turner 2014).

Gene annotation

A bed file containing SNPs associated with the quality traits of interest was intersected with the rainbow trout genome gff/gtf file using Bedtools in order to retrieve SNP annotations as mentioned before (Quinlan and Hall 2010, Salem et al 2018). SNPs were classified as genic or intergenic according to their physical position relative to the body of the gene. Genic SNPs exists in CDS, introns, 5'UTR or 3'UTR. Intergenic SNPs are defined as SNPs located in the region between genes (i.e. outside the body of the gene).

RESULTS AND DISCUSSION

Muscle fat and moisture contents are interrelated attributes that affect the organoleptic quality and nutritional value of flesh (Zheng et al 2016). In fish, high fat content my influence the fillet processing and reduce the firmness leading to fillet downgrading (Sodeland et al 2013). In mammals, increased marbling scores improve beef tenderness; accounting for ~9% of the shear force variation (Lu et al 2013). The inability to retain moisture during postmortem storage, in both fish and mammals, is associated with a high drip loss and, in turn, reduces the industry profitability by influencing processing yield and palatability (Melody et al 2004, Zhang et al 2006). In the pork industry, drip loss results in up to 10% product losses affecting profitability of processors and retailers (Melody et al 2004). Increased knowledge of the genetic basis of the muscle quality traits will help in making progress in commercial breeding of salmonids. GWA studies are powerful tools to identify variants associated with complex traits (Zheng et al 2016). However, there were no GWA studies conducted in rainbow trout to dissect the genetic architecture of fat and

moisture content. The estimated heritabilities were 0.31 ± 0.07 and 0.42 ± 0.07 for fat and moisture content, respectively, suggesting existence of adequate genetic variability in the fish population allowing increased breeding efficiency for these traits through genomic selection. Therefore, we used a 50K SNP chip to perform GWA analyses to identify QTL associated with both traits based on 50 SNP sliding windows and single-marker association analyses. The chip contains SNPs potentially associated with muscle quality traits including fat content as we previously described [28; 49]. However, we did not include any fish used in building the SNP-chip for GWA analysis in this study.

The fish population used for the current GWA analysis had an average muscle fat content of 9.2 ± 1.91 (%) and moisture content of 69.93 ± 1.75 (%). For the current GWA analysis, phenotypic variations in fat and moisture content are shown in Figure (1). Previous studies showed a significant correlation between changes in fat and moisture content (García-Celdrán et al 2015, Salem et al 2013). Consistently, our data showed a significant negative correlation between fat and moisture content ($R^2 = 0.78$; p-value = 6.3E-262).
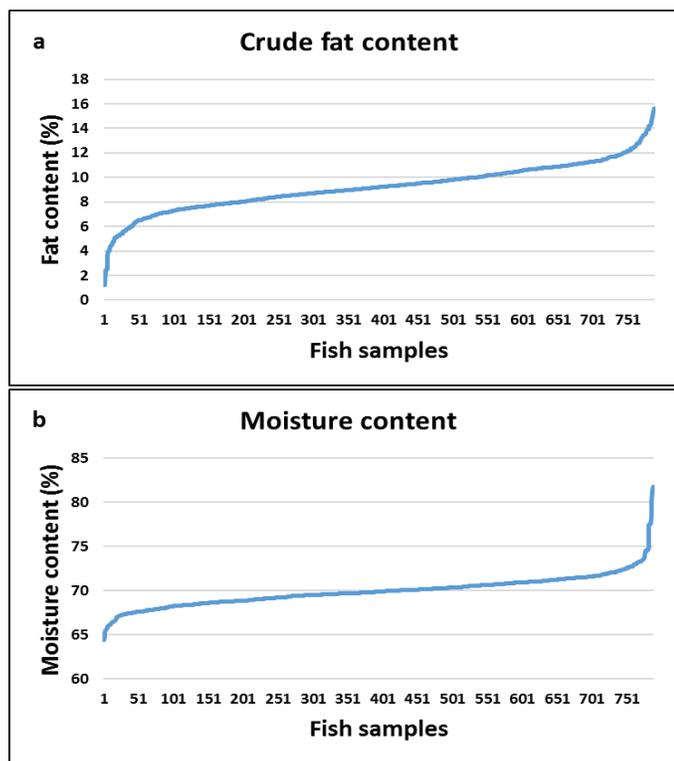
Figure 1. Phenotypic variations in crude fat content (a) and moisture content (b) among fish samples used for GWA analysis.

QTL affecting muscle fat and moisture content using WssGBLUP

All 35,322 SNPs (70.6%) which passed QC were used for the current WssGBLUP analysis. A complete list of the proportions of additive genetic variance for fat content explained by all genomic windows, was provided in Table S1. Of them, a total of 137 genomic windows explaining at least 2% (arbitrary value) of the additive genetic variance for fat content were listed in Table S2. Most of the SNPs (n = 124; ~91%) were located within 62 proteins-coding genes. Genomic loci affecting the additive variance for fat content were clustered on 5 chromosomes (1, 4, 5, 19, and 29) (Figure 2).
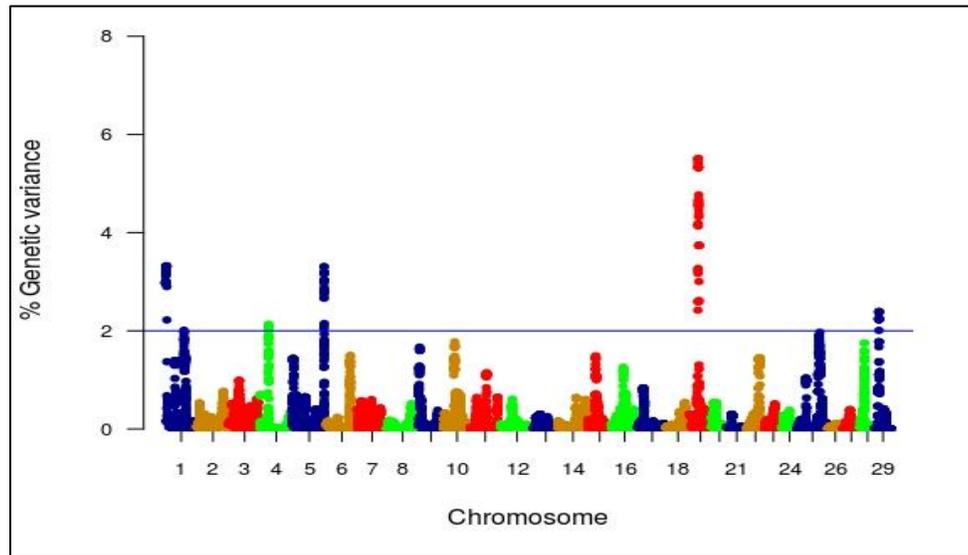
Figure 2. Manhattan plot showing association between genomic sliding windows of 50 SNPs and muscle fat content. Chromosome 19 showed the highest peaks with genomic loci explaining up to 5.51% of the additive genetic variance. The basal blue line represents 2% of additive genetic variance explained by SNPs.

Chromosome 19 harbored the highest number (n = 50) and most significant peaks affecting fat content (up to 5.51%) (Table S2, Figure 2). Many of the SNPs were located within the CDS of the SNP-harboring genes (n = 58) as well as their 3'UTR (n = 55). In order to understand the biological significance of the identified QTL, we annotated the SNP-harboring genes and searched their functions in the literature.

A complete list of the proportions of additive genetic variance for moisture content explained by all windows identified in this study, was provided in Table S3. A total of 178 genomic windows explaining at least 2% of the additive genetic variance for moisture content were listed in Table S4. Most of the SNPs (n = 165; ~93%) were located within 86 genes coding for proteins. Genomic loci affecting the additive variance for moisture

content were clustered on 5 chromosomes (5, 14, 19, 25, and 29) (Figure 3). Chromosome 29 harbored the highest number (n = 48), whereas most significant peaks affecting moisture content (up to 4.46%) were identified on chromosome 19 (Table S4, Figure 3). Many of the SNPs were located within CDS of the SNP-harboring genes (n = 68) as well as their 3'UTR (n = 72). In order to understand the biological significance of the identified QTL, we annotated the SNP-harboring genes and searched their functions in the literature. SNP-harboring genes were involved in lipid metabolism (Tables 1), cytoskeleton organization (Tables 2 & 3), and protein metabolic process (Tables 4).
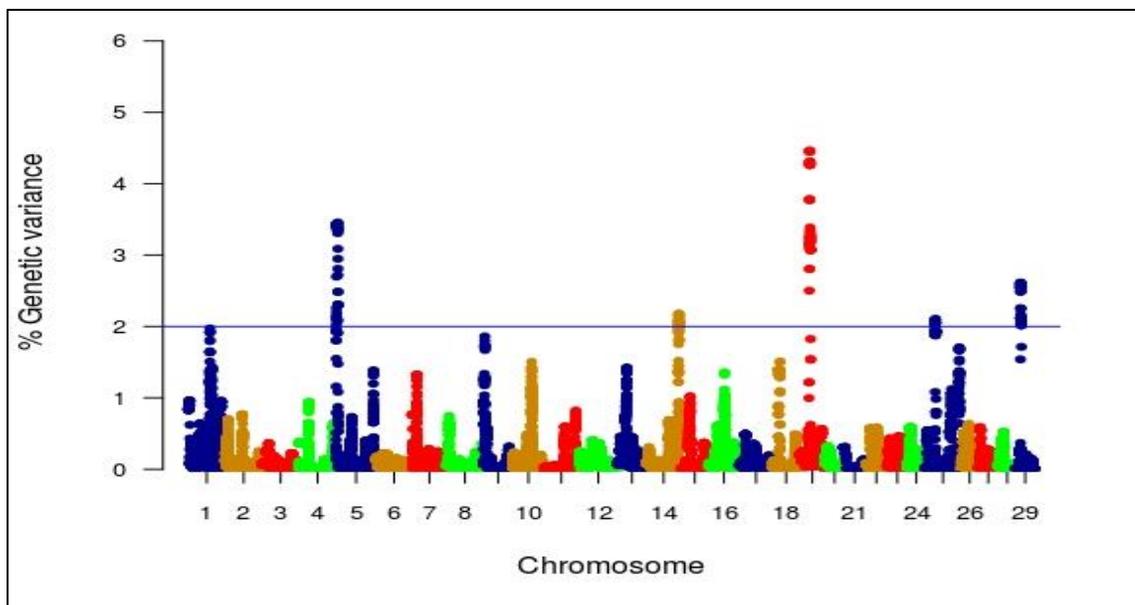
Figure 3. Manhattan plot showing association between genomic sliding windows of 50 SNPs and muscle moisture content. Chromosome 19 showed the highest peaks with genomic loci explaining up to 4.46% of the additive genetic variance. The basal blue line represents 2% of additive genetic variance explained by SNPs.

Common loci affecting muscle fat and moisture content

A substantial negative linear relationship has been established between fat and moisture content in selectively bred rainbow trout (YC 2010 and YC 2012), used in this study, suggesting a common mechanism underlying the genetic variation in the two traits. This negative correlation was consistent with other studies in fish and mammals (García-Celdrán et al 2015, Li et al 2013, Salem et al 2013, Watanabe et al 2018). However, this correlation may be disintegrated under certain physiological conditions. For example, fat content was not significantly affected in atrophying muscle during spawning in rainbow trout ($p > 0.05$) accompanied by increased moisture content suggesting that intramuscular fat is not required to support gonadal maturation at full sexual maturation and perhaps fertile female trout mobilize muscle protein in order to fuel sexual maturation. The current WssGBLUP identified common SNPs affecting the additive genetic variance for fat and moisture content on chromosomes 19 and 29 (Tables S2 & S4). The majority of the common SNPs (n = 47) were located on chromosome 19. Thirty-two SNPs, out of 47, involved in lipid metabolic process were identified in 16 protein-coding genes on chromosome 19 (Table 1). Briefly, cathepsin B had a single 3'UTR SNP. Cathepsin B regulates very low-density lipoprotein (VLDL) secretion and free fatty acids uptake in response to oleic acid exposure in mice (Thibeaux et al 2018). Thioredoxin-related transmembrane protein 1-like (TMX1) had three SNPs. Loss of TMX increases lipid peroxidation in TMX−/− mice which, in turn, enhances oxidative stress (Matsuo et al 2013). Guanine nucleotide-binding protein GI/GS/GO gamma-2 subunit (GNG2) had a single 3'UTR SNP. GNG2 expression is positively correlated with adipocyte size (Tandon et al 2018). SNPs in genes encoding beta-taxilin and Alpha-L-fucosidase 2 (FUCA2) were

spanning windows explaining the highest proportion of the additive genetic variation for fat and moisture content. Adipose tissue of obesity susceptible and resistant rats differentially expressed beta-taxilin under a high fat diet (Joo and Yun 2011). FUCA2 is a glycolipid processing enzyme (Flor et al 2017). Two SNPs in F-box only protein 30 (FBXO30) and the microtubule binding protein ensconsin were ranked next to beta-taxilin and FUCA2. A SNP in FBXO30 was located in a genomic region explaining 4.95% of the additive genetic variance for the polyunsaturated fatty acids in cattle (Lemos et al 2016). Knock-down of microtubuli-binding or -associated proteins led to changes in fat accumulation during adipogenesis (Söhle et al 2012). Dihydropyrimidinase-related protein 5-like (CRMP5) had a single synonymous SNP. CRMP5 has GO terms belong to lipid metabolic process (Gaudet et al 2011). Five SNPs were identified in a gene encoding trifunctional enzyme subunit alpha, mitochondrial (HADHA). This protein is involved in lipid metabolism; fatty acid beta-oxidation pathway (Gaudet et al 2011). mRNA decay activator protein ZFP36L1 (ZFP36L1) had three SNPs in the 3'UTR. Knockdown of mammalian ZFP36L1 led to downregulation of ERK activation and inhibition of adipogenesis (Lin et al 2012a). A single 3'UTR SNP was identified in ELM2 and SANT domain-containing protein 1 (ELMSAN1). Epigenome-wide association analysis showed DNA methylation changes in ELMSAN1 were associated with body mass index (a key measure of adiposity) (Wahl et al 2017). Prostaglandin reductase 2 (PTGR2) had two nonsynonymous SNPs. This enzyme catalyzes reduction of the conjugated α,β-unsaturated double bond of 15-keto-PGE2 in NADPH-dependent manner; a critical step in inhibition of PPARγ-mediated adipocyte differentiation (Wu et al 2008). Spectrin beta chain, erythrocytic (SPTB) harbored two SNPs. The SPTB interacts with phospholipids in natural

membranes and has a role in controlling the fluidity of the inner lipid leaflet of the cell membrane (Machnicka et al 2014).

Table 1. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of additive genetic variance for fat and moisture content, 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.

Var%_Fat	Var%_Moisture	CHR	Position	Strand	Gene ID	Function	Gene annotation	Region/effect
3.25	3.16	19	23140902	+	LOC110497560	Lipid metabolism	cathepsin B-like	3'UTR
4.19	3.77	19	23257105	+	LOC110497565	Lipid metabolism	thioredoxin-related transmembrane protein 1-like	CDS/nonsyn
4.15	3.77	19	23259435	+	LOC110497565	Lipid metabolism	thioredoxin-related transmembrane protein 1-like	3'UTR
4.15	3.78	19	23260053	+	LOC110497565	Lipid metabolism	thioredoxin-related transmembrane protein 1-like	3'UTR
4.14	3.78	19	23443626	+	gbg2	Lipid metabolism	Guanine nucleotide-binding protein GI/GS/GO gamma-2 subunit	3'UTR
5.50	4.45	19	23495709	-	LOC110497567	Lipid metabolism	beta-taxilin-like	CDS/syn
5.51	4.46	19	23496538	-	LOC110497567	Lipid metabolism	beta-taxilin-like	CDS/syn
5.51	4.46	19	23527643	-	fuca2	Lipid metabolism	alpha-L-fucosidase 2	3'UTR
5.33	4.28	19	23528291	-	fuca2	Lipid metabolism	alpha-L-fucosidase 2	3'UTR
5.42	4.31	19	23699374	-	LOC110497583	Lipid metabolism	F-box only protein 30-like	3'UTR
5.32	4.26	19	24247835	-	LOC110497599	Lipid metabolism	ensconsin-like	3'UTR
4.66	3.39	19	24329620	-	LOC110497602	Lipid metabolism	dihydropyrimidinase-related protein 5-like	CDS/syn
4.55	3.22	19	24502571	-	LOC110497612	Lipid metabolism	trifunctional enzyme subunit alpha, mitochondrial-like	3'UTR
4.56	3.23	19	24503147	-	LOC110497612	Lipid metabolism	trifunctional enzyme subunit alpha, mitochondrial-like	3'UTR
4.56	3.24	19	24509666	-	LOC110497612	Lipid metabolism	trifunctional enzyme subunit alpha, mitochondrial-like	CDS/nonsyn
4.67	3.33	19	24512704	-	LOC110497612	Lipid metabolism	trifunctional enzyme subunit alpha, mitochondrial-like	CDS/nonsyn
4.63	3.28	19	24512932	-	LOC110497612	Lipid metabolism	trifunctional enzyme subunit alpha, mitochondrial-like	CDS/syn
4.47	3.20	19	25037839	+	LOC110497620	Lipid metabolism	mRNA decay activator protein ZFP36L1-like	3'UTR
4.33	3.19	19	25039247	+	LOC110497620	Lipid metabolism	mRNA decay activator protein ZFP36L1-like	3'UTR
4.33	3.18	19	25039632	+	LOC110497620	Lipid metabolism	mRNA decay activator protein ZFP36L1-like	3'UTR
3.74	3.07	19	25366465	-	LOC110498698	Lipid metabolism	ELM2 and SANT domain-containing protein 1-like	3'UTR
3.74	3.07	19	25409419	+	ptgr2	Lipid metabolism	prostaglandin reductase 2	CDS/nonsyn
3.74	3.07	19	25414599	+	ptgr2	Lipid metabolism	prostaglandin reductase 2	CDS/nonsyn
3.74	3.07	19	25563733	+	LOC110497630	Lipid metabolism	spectrin beta chain, erythrocytic-like	CDS/nonsyn
3.74	3.07	19	25581001	+	LOC110497630	Lipid metabolism	spectrin beta chain, erythrocytic-like	CDS/syn
2.24	2.54	29	10494611	+	LOC110509620	Lipid metabolism	short-chain specific acyl-CoA dehydrogenase, mitochondrial-like	3'UTR
2.39	2.61	29	10714305	+	LOC110509628	Lipid metabolism	arrestin domain-containing protein 3-like	mRNA
2.39	2.62	29	10776686	+	LOC110509321	Lipid metabolism	myocyte-specific enhancer factor 2C-like	5'UTR
2.23	2.59	29	10798819	+	LOC110509321	Lipid metabolism	myocyte-specific enhancer factor 2C-like	mRNA
2.24	2.59	29	10800769	+	LOC110509321	Lipid metabolism	myocyte-specific enhancer factor 2C-like	CDS/syn
2.22	2.59	29	10801091	+	LOC110509321	Lipid metabolism	myocyte-specific enhancer factor 2C-like	CDS/syn
2.01	2.53	29	10801136	+	LOC110509321	Lipid metabolism	myocyte-specific enhancer factor 2C-like	CDS/syn

Chromosome 29 had 14 SNPs in genomic windows explaining at least 2% of the additive genetic variance for the muscle fat and moisture content (Tables S2 & S4). Of

them, seven SNPs were involved in lipid metabolic process (Table 1). A single SNP was identified in a gene encoding short-chain specific acyl-CoA dehydrogenase, mitochondrial (ACADS). This enzyme has a role in the fatty acid beta-oxidation (Finocchiaro et al 1987). An intronic SNP was identified in a gene coding for arrestin domain-containing protein. The latter has GO terms belong to fat pad and skin development, and regulates the body mass (Patwari et al 2011). Myocyte enhancer factor 2c (MEF2C) had the highest number of SNPs (n = 5) on chromosome 29. MEF2C is a transcription factor involved in the differentiation of skeletal muscle, however, it has been reported as a constituent of a mechanism that program the gene expression profile to develop brown adipocytes (Lin 2015). MEF2A and MEF2D isoforms showed *in vivo* DE in mammalian white adipose tissue (Mora et al 2001). To our knowledge, the role of MEF2C in white adipose tissue remains uncertain.

In addition, twelve SNPs in genes involved in transmembrane transport and cytoskeleton remodeling were identified in the common QTL affecting the additive variance for fat and moisture content (Table 2). The majority of these SNPs were identified on chromosome 19 (n = 11). Three synonymous SNPs were identified in a gene encoding intersectin-2 (ITSN2). This protein is necessary for the clathrin-mediated endocytosis and actin cytoskeleton remodeling (Novokhatska et al 2013). Six SNPs were identified in 3 genes involved in vesicle-mediated transport (i.e. exocytosis); dnaJ homolog subfamily C member 5B, visinin-like protein 1, and syntaxin-binding protein 5. The actin cytoskeleton remodeling controls each step of exocytosis (Porat-Shliom et al 2013). Three SNPs were identified in microtubule-associated protein RP/EB family member 3 (MAPRE3) and

centrin-3. MAPRE3 and centrin-3 control the dynamics of the microtubule cytoskeleton (Gaudet et al 2011, Komarova et al 2009).

Table 2. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of additive genetic variance for fat and moisture content, and involved in transmembrane transport and cytoskeleton regulation. 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.

Var%_Fat	Var%_Moisture	CHR	Position	Strand	Gene ID	Function	Gene annotation	Region/effect
2.60	2.81	19	23060380	-	LOC110497555	Endocytosis & cytoskeleton	intersectin-2	CDS/syn
2.59	2.80	19	23062680	-	LOC110497555	Endocytosis & cytoskeleton	intersectin-2	CDS/syn
3.17	3.11	19	23082360	-	LOC110497555	Endocytosis & cytoskeleton	intersectin-2	CDS/syn
5.33	4.28	19	23561806	-	LOC110497574	Exocytosis	dnaJ homolog subfamily C member 5B	3'UTR
5.33	4.29	19	23621300	-	LOC110497579	Exocytosis	visinin-like protein 1	3'UTR
5.34	4.31	19	23627722	-	LOC110497579	Exocytosis	visinin-like protein 1	5'UTR
5.42	4.31	19	23627740	-	LOC110497579	Exocytosis	visinin-like protein 1	5'UTR
5.32	4.26	19	23838038	+	LOC110497586	Exocytosis	syntaxin-binding protein 5	CDS/syn
5.33	4.26	19	23851302	+	LOC110497586	Exocytosis	syntaxin-binding protein 5	3'UTR
5.34	4.30	19	24316246	-	LOC110497601	Microtubule cytoskeleton	microtubule-associated protein RP/EB family member 3	3'UTR
4.54	3.34	19	24317314	-	LOC110497601	Microtubule cytoskeleton	microtubule-associated protein RP/EB family member 3	3'UTR
2.39	2.61	29	10746517	+	LOC110509629	Microtubule binding	centrin-3	CDS/nonsyn

Unique genomic loci affecting the additive variance for moisture

The actin cytoskeleton interacts with the plasma membrane to regulate the water transport (Wayne and Tazawa 1989). Thirty-six variants in genes (n = 14) involved in cytoskeleton remodeling were identified affecting the additive variance for moisture content in rainbow trout (Table 3).

Briefly, bone morphogenetic protein receptor type-2 (BMPR2) had a single synonymous SNP. BMPR2 is known to interact with the cytoskeleton and BMPR2 mutant mice exhibited cytoskeletal defects (Johnson et al 2012). A single SNP was identified in a

gene encoding muscle associated receptor tyrosine kinase (MUSK). The activation of MUSK in myotubes regulates the reorganization of the actin cytoskeleton (Tan-Sindhunata et al 2015). Two SNPs were identified in THAP domain containing 1 (THAP1). It has a role in regulation of mitotic cell cycle (Cayrol et al 2007). The gene encoding asparaginyl-tRNA synthetase (NARS) had 3 SNPs in windows explaining the highest additive variance (up to 3.46%) in this category. Mutations in NARS leads to cell cycle arrest in S phase (Diamond et al 1989). Actin cytoskeleton undergoes dramatic changes during the cell cycle (Nakaseko and Yanagida 2001). Ten SNPs were identified in three genes coding for cyclin-I (CCNI2), cyclin-G1 (CCNG1), and cyclin-G2 (CCNG2). Cyclins function as regulators of cell cycle and actin cytoskeleton dynamics (Bendris et al 2015). The serine/threonine-protein phosphatase 2A (PP2A) had a 3'UTR SNP. This phosphatase is associated with the microtubule stabilization where it binds and dephosphorylates the microtubule-associated proteins (Sontag et al 1996). Annexin A6 (ANXA6) had two synonymous SNPs. ANXA6 contributes to membrane and cytoskeleton organization in a Ca^{2+}-dependent manner (Qi et al 2015). Tubulin beta-4B chain (TUBB4B) had four synonymous SNPs spanning less than 1Kb on chromosome 25. TUBB4B is the key component of microtubules (Gaudet et al 2011). Five SNPs spanning ~2Kb were identified in a gene coding for mid1-interacting protein 1 (MID1IP1) (Sugiyama et al 2012). This protein enhances fatty acid biosynthesis (Sugiyama et al 2012) and stabilizes microtubule organization (Gaudet et al 2011). Two SNPs were identified in a gene encoding tubulin-specific chaperone A (TBCE). TBCE is a tubulin-folding protein required for the proper microtubule cytoskeleton organization (Sferra et al 2016). In addition, mutations in TBCE drive muscular atrophy (Sferra et al 2016). Proteinase-activated receptor 1 (PAR1) and

PAR2 had four SNPs. PAR-mediated RhoA activation is important for cytoskeletal reorganization (Greenberg et al 2003).

Table 3. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of additive genetic variance for moisture content, and involved in cell cycle and cytoskeleton regulation. 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.

Var%	CHR	Position	Strand	Gene ID	Function	Gene annotation	Region/effect
3.42	5	1581801	+	LOC110524928	Cytoskeleton	bone morphogenetic protein receptor type-2	CDS/syn
3.43	5	2416345	+	musk	Cytoskeleton	muscle associated receptor tyrosine kinase	3'UTR
3.39	5	3076608	-	thap1	Cell cycle	THAP domain containing 1	3'UTR
3.39	5	3077900	-	thap1	Cell cycle	THAP domain containing 1	CDS/syn
3.34	5	3948426	+	nars	Cell cycle	asparaginyl-tRNA synthetase	CDS/syn
3.45	5	3954845	+	nars	Cell cycle	asparaginyl-tRNA synthetase	CDS/syn
3.46	5	3956515	+	nars	Cell cycle	asparaginyl-tRNA synthetase	3'UTR
2.72	5	4499566	+	LOC110523126	Cell cycle	cyclin-I	5'UTR
2.49	5	4515345	-	LOC110523127	Cell cycle	cyclin-G2	3'UTR
2.48	5	4517902	-	LOC110523127	Cell cycle	cyclin-G2	CDS/syn
2.31	5	4523733	+	LOC110523126	Cell cycle	cyclin-I	mRNA
2.30	5	4529514	+	LOC110523126	Cell cycle	cyclin-I	CDS/nonsyn
2.30	5	4533058	+	LOC110523126	Cell cycle	cyclin-I	CDS/nonsyn
2.30	5	4539962	+	LOC110523126	Cell cycle	cyclin-I	CDS/nonsyn
2.31	5	4545000	+	LOC110523126	Cell cycle	cyclin-I	CDS/nonsyn
2.09	5	4547885	+	LOC110523126	Cell cycle	cyclin-I	3'UTR
2.08	14	74268218	+	LOC110489167	Cell cycle	cyclin-G1	3'UTR
2.08	14	74743050	-	LOC110489177	Cytoskeleton	serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	3'UTR
2.08	14	74743050	-	LOC110489177	Cytoskeleton	serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	3'UTR
2.00	14	75046230	-	LOC110489191	Cytoskeleton	annexin A6	CDS/syn
2.00	14	75047544	-	LOC110489191	Cytoskeleton	annexin A6	CDS/syn
2.05	25	23765983	-	LOC110504922	Cytoskeleton	tubulin beta-4B chain	CDS/syn
2.04	25	23766034	-	LOC110504922	Cytoskeleton	tubulin beta-4B chain	CDS/syn
2.04	25	23766334	-	LOC110504922	Cytoskeleton	tubulin beta-4B chain	CDS/syn
2.05	25	23766912	-	LOC110504922	Cytoskeleton	tubulin beta-4B chain	CDS/syn
2.09	25	23967070	+	m1ip1	Cytoskeleton	Mid1-interacting protein 1	5'UTR
2.09	25	23968673	+	m1ip1	Cytoskeleton	Mid1-interacting protein 1	3'UTR
2.11	25	23969016	+	m1ip1	Cytoskeleton	Mid1-interacting protein 1	3'UTR
2.05	25	23969082	+	m1ip1	Cytoskeleton	Mid1-interacting protein 1	3'UTR
2.05	25	23969158	+	m1ip1	Cytoskeleton	Mid1-interacting protein 1	3'UTR
2.02	29	12409282	+	LOC110509670	Cytoskeleton	tubulin-specific chaperone A	CDS/syn
2.02	29	12410990	+	LOC110509670	Cytoskeleton	tubulin-specific chaperone A	3'UTR
2.02	29	12653356	-	LOC110509674	Cytoskeleton	proteinase-activated receptor 2	3'UTR
2.02	29	12654062	-	LOC110509674	Cytoskeleton	proteinase-activated receptor 2	CDS/syn
2.07	29	12654462	-	LOC110509674	Cytoskeleton	proteinase-activated receptor 2	CDS/nonsyn
2.10	29	12665454	-	LOC110509677	Cytoskeleton	proteinase-activated receptor 1	3'UTR

Moisture content has shown dramatic/significant changes with the protein content in different species. Simultaneous decline in protein and moisture content was previously reported in mammals (Li et al 2013). Moisture content in rainbow trout exhibited a bidirectional relationship with protein content according to the physiological/metabolic status. For example, a negative correlation between moisture and protein content were previously reported under muscle catabolic conditions associated with full sexual maturation ($R^2 = 0.994$, $p < 0.01$) (Salem et al 2006), whereas positive correlation was established in female trout approaching spawning under high nutritional plane (Salem et al 2013). This was explained by selective mobilization of either protein during spawning or fat before spawning; depleted macromolecules were replenished by water. The current WssGBLUP analysis identified thirteen SNPs in genes involved in protein degradation were affecting the additive variance for moisture content (Table 4). Briefly, E3 ubiquitin-protein ligase RNF170 is an E3 ubiquitin-protein ligase which plays an essential role in ubiquitination and degradation of inositol 1,4,5-trisphosphate receptor type 1 (ITPR1). The latter controls the calcium release from the endoplasmic reticulum (Huang et al 2012) which affects the muscle protein content in rainbow trout (Ali et al 2019) and has a profound effect on the regulation of cytoskeleton (Hepler 2016). Cystatin-1, which possesses a peptidase inhibitor activity, had a single 5'UTR SNP. Thioredoxin like 1 (TXNL1) had two synonymous SNPs. Knockdown of TXNL1 moderately stabilizes the ubiquitin-protein conjugates suggesting a connection between protein reduction and proteolysis (Andersen et al 2009). Pre-mRNA-processing factor 19 (PRPF19) and ubiquitin-conjugating enzyme E2 D2 (UBE2D2) had four SNPs. These ligase enzymes catalyze polyubiquitin chain assembly and play a role in proteasomal protein degradation

(David et al 2010, Hatakeyama et al 2001). Nuclear factor NF-kappa-B p105 subunit (NFKB1) had two 3'UTR SNPs. NFKB1 is involved in negative regulation of cellular protein metabolic process (Ferreira et al 2007) and apoptotic process (Cahir-McFarland et al 2000).

Table 4. SNP markers in genomic sliding windows of 50 SNPs explaining at least 2% of additive genetic variance for moisture content, 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.

Var%	CHR	Position	Strand	Gene ID	Function	Gene annotation	Region/effect
3.42	5	1402874	-	LOC110523084	Ubiquitin-protein ligase	E3 ubiquitin-protein ligase RNF170	3'UTR
3.43	5	2249297	+	LOC110523107	peptidase inhibitor	cystatin-1	5'UTR
3.37	5	3931855	-	txnl1	Redox homeostasis	thioredoxin like 1	CDS/syn
3.34	5	3941358	-	txnl1	Redox homeostasis	thioredoxin like 1	CDS/syn
2.05	25	23666377	-	LOC110504917	Ubiquitin-protein ligase	pre-mRNA-processing factor 19	CDS/syn
2.05	25	23690277	+	LOC110504919	Ubiquitin-protein ligase	ubiquitin-conjugating enzyme E2 D2	5'UTR
2.05	25	23698043	+	LOC110504919	Ubiquitin-protein ligase	ubiquitin-conjugating enzyme E2 D2	3'UTR
2.06	25	23698573	+	LOC110504919	Ubiquitin-protein ligase	ubiquitin-conjugating enzyme E2 D2	3'UTR
2.06	25	23700241	-	LOC110504918	Transcription factor	nuclear factor NF-kappa-B p105 subunit	3'UTR
2.06	25	23700272	-	LOC110504918	Transcription factor	nuclear factor NF-kappa-B p105 subunit	3'UTR
2.12	29	11141837	+	LOC110509643	Phagosome	ras-related protein rab7	3'UTR
2.03	29	11416104	+	LOC110509654	Phagosome	V-type proton ATPase subunit B, brain isoform	3'UTR
2.03	29	12313543	+	LOC110509669	Lysosome	AP-3 complex subunit beta-1	CDS/syn

In addition to the ubiquitin-protein ligases, SNPs in three genes involved in lysosomal/phagosomal pathways were identified. Ras-related protein rab7 (RAB7A) harbored a 3'UTR SNP. RAB7A is a major regulator of endo-lysosomal maturation/ trafficking and protein targeting to lysosome inducing autophagosome formation (Lin et al 2012b). Thus, RAB7A positively regulates protein catabolic process (Caillet et al 2011). V-type proton ATPase subunit B (ATP6V1B2) had a 3'UTR SNP. V-ATPase is

responsible for acidifying the intracellular compartments including lysosomes (Trombetta et al 2003). The gene encoding the β chain of the adaptor protein-3 (AP-3) complex had a single synonymous SNP. Deletion in AP3B1 perturbs assembly of AP-3 complex and, in turn, trafficking of transmembrane lysosomal proteins (Jung et al 2006).

Taken together, most of the common genomic loci affecting the highest proportion of the additive variance were involved in lipid metabolic process suggesting a common mechanism underlying muscle fat and moisture content, and, partially, explaining the strong negative correlation between the fat and moisture contents in selectively bred rainbow trout. Unique loci affecting moisture content were mainly involved in cytoskeleton regulations and protein turnover. Inhibition of the activity of the protease systems, such as calpains, reduced degradation of proteins responsible for cell membrane-cytoskeleton attachments and postmortem drip channel formation in muscle. Presence of calcium enhances proteolysis, by μ-Calpain, of myofibrillar and other cytoskeletal proteins during postmortem storage (Zhang et al 2006).

Single marker GWA analyses

In order to identify single SNP markers associated with phenotypic variation in fat and moisture content, we analyzed SNPs which passed QC filtration, using general linear regression model in PLINK (Purcell et al 2007). The linear model allows for using multiple covariates to control for population structure, year class, and harvest group. In this study, PLINK identified 13 and 27 significant SNPs, surpassing the genome-wide significance level, had potential impact on the fat and moisture content (Bonferroni-corrected $p < 2.01E\text{-}06$; Appendices A & B, Figures 4 & 5 and Tables S5 & S6), respectively.

SNPs associated with the fat content were mainly located on chromosome 5 (n = 12), and have shown roles in lipid metabolism (Table 5). The list includes Golgin, 78 kDa glucose-regulated protein (GRP78), spindle and kinetochore associated complex subunit 1 (SKA1), apelin receptor B (APLNR-B), proliferating cell nuclear antigen (PCNA), desmoplakin, MAF bZIP transcription factor A, podocan, and calcium-binding mitochondrial carrier protein SCaMC-1 (SLC25A24). Briefly, Golgin is required for the appropriate assembly of apolipoprotein A-I as well as its transport through Golgi apparatus. Suppression of Golgin leads to improper recruitment of Rab-2, necessary for trafficking of lipid vesicles, to Golgi and less lipidation of pre-chylomicrons (Hesse et al 2013). GRP78 is essential for adipocyte differentiation and balanced secretion of adipokines. Deletion of GRP78 causes lipoatrophy accompanied by a dramatic reduction in gonadal and subcutaneous adipose tissue (Zhu et al 2013). SKA1 was downregulated in adipose tissues between samples from obese and normal control children and has been suggested as a candidate biomarker for childhood obesity (Zhu et al 2018). APLNR KO mice demonstrated excess fatty acid accumulation in skeletal muscle (Hwangbo et al 2017). Mutant mice at the phosphorylation site of PCNA failed to enter the cell cycle and develop mature adipocytes (Lo et al 2013). Abnormalities in desmoplakin have been associated with changes in lipid metabolism (Yang et al 2006). bZIP domain containing gene-deficient mice revealed increase in lipolysis and decreases in expression level of lipogenic genes (Sevane et al 2013). Podocan belongs to the small leucine-rich proteoglycans (SLRPs) which binds to low-density lipoprotein receptor-related protein (LRP-1) (Merline et al 2009). Mice fed a high fat diet exhibited increased expression level of SLC25A24,

whereas adipocyte differentiation was suppressed in Slc25a24- knockout (Writzl et al 2017).

Puromycin-sensitive aminopeptidase (NPEPPS), on chromosome 17, had a SNP explaining the highest phenotypic variability in fat content ($R^2 = 3.2\%$) (Table 5). NPEPPS has roles in different physiological processes including protein turnover and cell cycle regulation. NPEPPS was upregulated in mitten crabs fed with a linseed oil rich in linoleic acid (Wei et al 2018). However, the effect NPEPPS on lipid metabolism in fish needs further investigations.
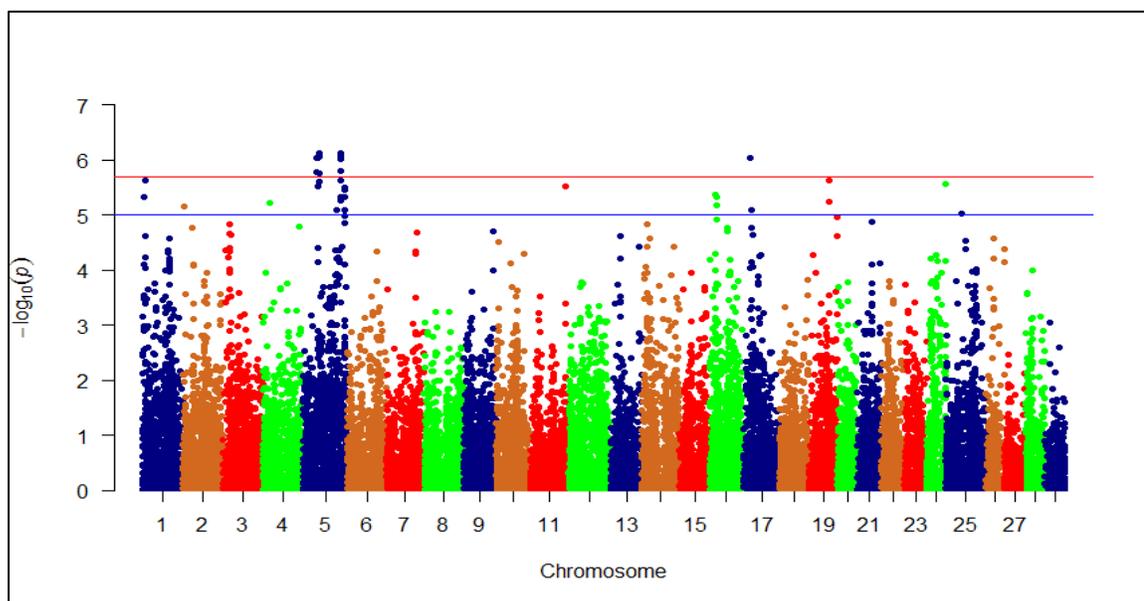

Figure 4. Manhattan plot showing single SNP markers associated with variations in muscle fat content. Blue and red horizontal lines represent suggestive and significance threshold p-values of 1e-05 and 2.01e-06, respectively.

SNPs associated with moisture content (n = 27) were involved in protein turnover, calcium metabolism, and cytoskeleton (Appendix A & Table S5). Most of these SNPs (n = 21; ~78%) were located on chromosomes 1 and 17. A SNP in a gene coding for acylphosphatase 2 ranked at the top of the list (R^2 = 7.4%) (Table 5), however, it is physiological role is not clear. Eleven SNPs associated with moisture content were identified in eight genes engaged in protein metabolic process (Table 5). These genes are WD repeat domain 11 (WDR11), adenosine kinase, ERCC excision repair 6, chromatin remodeling factor (ERCC6), NPEPPS, eukaryotic initiation factor 4A-III (EIF4A3), eukaryotic translation initiation factor 4B (EIF4B), ribosome binding protein 1 (RRPB1), and F-box only protein 46 (FBXO46). Briefly, WDR11 interacts with a protein, dysbindin, related to the ubiquitin-proteasome system (Han et al 2014). Deficiency of adenosine kinase disrupts methionine cycle (Bjursell et al 2011) which affects drip loss (Wen et al 2017). Mutations of the ERCC6 are the predominant cause of accelerated muscle wasting (Yu et al 2014). The aminopeptidase NPEPPS was associated with the variation in fat and moisture content suggesting a correlation between moisture and fat content. Five SNPs were identified in two genes encoding EIF4A3 and EIF4B suggesting a role for the translation machinery in determining the variation in moisture content. RRPB1 is ER integral membrane proteins implicated in polysome assembly; and therefore, protein synthesis (Ortega et al 2014). RRPB1 has been suggested to be essential in regulation of UPR signaling molecules and autophagy (Pan et al 2015). Finally, the F-box family SCF-E3 ubiquitin ligase, FBXO46, had a single 3'UTR SNP.
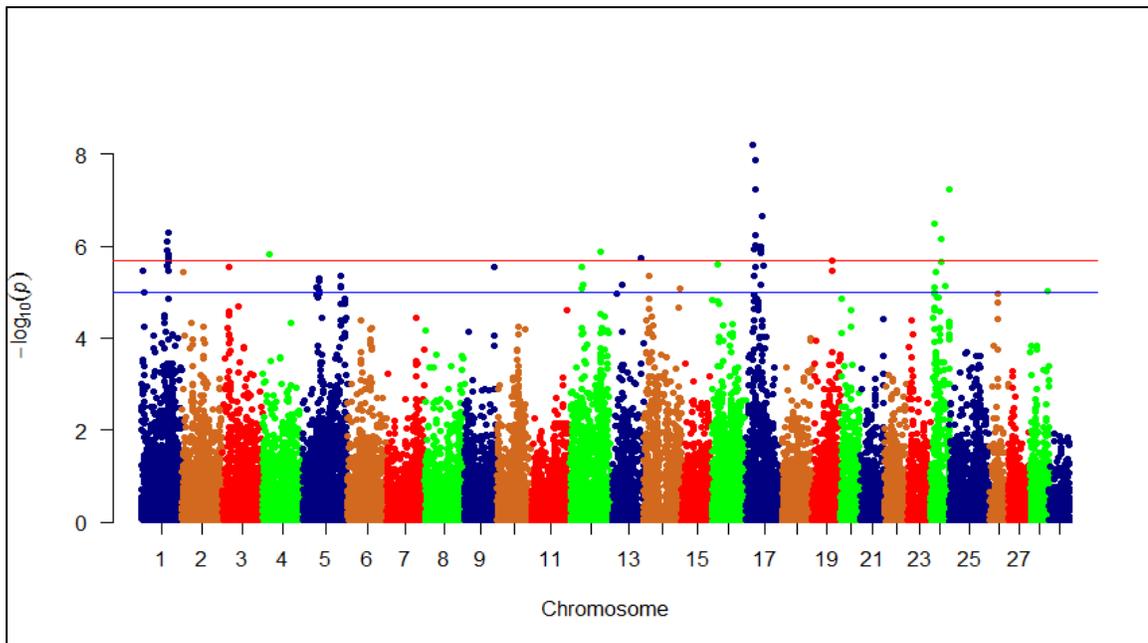
Figure 5. Manhattan plot showing single SNP markers associated with variations in moisture content. Blue and red horizontal lines represent suggestive and significance threshold p-values of 1e-05 and 2.01e-06, respectively.

A total of 10 SNPs associated with moisture content were identified in five genes engaged in cytoskeleton regulation (Table 5). Serum response factor (SRF) is a master regulator of the actin cytoskeleton (Miano et al 2007). Three SNPs have been identified in a gene encoding actin cytoskeleton-regulatory complex protein pan1 (PAN1). Genes involved in actin cytoskeleton, and cytoskeleton organization and biogenesis had positively correlated expression with drip loss (Ponsuksili et al 2008). Transducin-like enhancer protein 1 has WD40 domain which functions in cytoskeleton assembly (Sayers et al 2019). Translation initiation factors, including eIF4A and EIF4B, associate with the actin

cytoskeleton which affects protein synthesis (Gross and Kinzy 2007). Two other SNPs have been identified in genes encoding proteins involved in lipid and calcium homeostasis; 3-hydroxybutyrate dehydrogenase 2 and calcineurin subunit B type 1, respectively.

Table 5. SNP markers significantly associated with phenotypic variability in fat and moisture content using a single SNP GWA analysis. SNPs associated with each phenotype, are sorted according to their chromosome positions.

R^2	CHR	position	Gene ID	Strand	Gene annotation	Region/effect	UNADJ	BONF	Trait
0.018	5	26977667	LOC110523594	-	golgin subfamily A member 2	3'UTR	1.64E-06	4.08E-02	Fat content
0.019	5	27016969	LOC110523600	+	78 kDa glucose-regulated protein	CDS/syn	9.22E-07	2.30E-02	Fat content
0.021	5	32672947	ska1	-	spindle and kinetochore associated complex subunit 1	CDS/nonsyn	7.70E-07	1.92E-02	Fat content
0.021	5	33025927	LOC110523701	-	apelin receptor B	CDS/syn	8.51E-07	2.12E-02	Fat content
0.021	5	33296459	LOC110523713	-	proliferating cell nuclear antigen	Intronic	8.60E-07	2.14E-02	Fat content
0.019	5	77138939	LOC110524668	+	desmoplakin	CDS/syn	9.22E-07	2.30E-02	Fat content
0.019	5	78062294	mafa	-	MAF bZIP transcription factor A	3'UTR	8.95E-07	2.23E-02	Fat content
0.019	5	78880913	podn	-	podocan	CDS/syn	9.64E-07	2.40E-02	Fat content
0.019	5	79116860	LOC110524700	-	calcium-binding mitochondrial carrier protein SCaMC-1	3'UTR	1.55E-06	3.85E-02	Fat content
0.032	17	13538660	LOC110493605	+	puromycin-sensitive aminopeptidase	CDS/syn	9.26E-07	2.31E-02	Fat content
0.015	1	53970934	wdr11	+	WD repeat domain 11	CDS/syn	1.20E-06	2.98E-02	Moisture content
0.016	1	54958459	LOC110525667	-	serum response factor	CDS/syn	7.82E-07	1.95E-02	Moisture content
0.018	1	55318448	LOC110525737	-	adenosine kinase	CDS/syn	1.49E-06	3.70E-02	Moisture content
0.019	1	55875918	ercc6	+	ERCC excision repair 6, chromatin remodeling factor	3'UTR	1.84E-06	4.58E-02	Moisture content
0.019	1	55903395	LOC110525889	-	actin cytoskeleton-regulatory complex protein pan1	CDS/nonsyn	1.48E-06	3.68E-02	Moisture content
0.018	1	55903477	LOC110525889	-	actin cytoskeleton-regulatory complex protein pan1	CDS/nonsyn	1.71E-06	4.25E-02	Moisture content
0.019	1	55904783	LOC110525889	-	actin cytoskeleton-regulatory complex protein pan1	CDS/syn	1.73E-06	4.30E-02	Moisture content
0.032	4	15238170	LOC110521463	+	transducin enhancer protein 1	3'UTR	1.45E-06	3.61E-02	Moisture content
0.042	17	13538660	LOC110493605	+	puromycin-sensitive aminopeptidase	CDS/syn	6.11E-09	1.52E-04	Moisture content
0.030	17	17622851	LOC110493736	-	calcineurin subunit B type 1	3'UTR	1.14E-06	2.83E-02	Moisture content
0.056	17	20335252	LOC110493798	-	eukaryotic initiation factor 4A-III	3'UTR	9.51E-07	2.37E-02	Moisture content
0.057	17	20678734	rrbp1	+	ribosome binding protein 1	CDS/nonsyn	5.77E-07	1.44E-02	Moisture content
0.074	17	21029747	acyp2	-	acylphosphatase 2	3'UTR	1.32E-08	3.29E-04	Moisture content
0.035	17	21070906	bdh2	+	3-hydroxybutyrate dehydrogenase 2	3'UTR	5.88E-08	1.46E-03	Moisture content
0.034	17	30097354	LOC110493954	-	eukaryotic translation initiation factor 4B	3'UTR	1.37E-06	3.40E-02	Moisture content
0.055	17	30097441	LOC110493954	-	eukaryotic translation initiation factor 4B	3'UTR	1.33E-06	3.31E-02	Moisture content
0.055	17	30098943	LOC110493954	-	eukaryotic translation initiation factor 4B	CDS/syn	1.26E-06	3.14E-02	Moisture content
0.056	17	30108957	LOC110493954	-	eukaryotic translation initiation factor 4B	CDS/syn	1.01E-06	2.51E-02	Moisture content
0.019	24	9580897	LOC110503690	+	F-box only protein 46	3'UTR	3.26E-07	8.12E-03	Moisture content

In our previous work, we profiled transcriptome expression of fish families (YC 2010) showing contrasting phenotypes in fat content which revealed only 17 differentially expressed transcripts associated with fat content (Ali et al 2018). About 90% of the genetic

variation among individuals comes from SNPs (Salem et al 2012); and therefore, identifying SNP markers associated with complex traits is most suitable for genetic evaluation in selection programs. Two previous GWA studies in fish identified a few SNPs responsible for the additive variance for fat content (Sodeland et al 2013, Zheng et al 2016). The current GWA analysis identified a total of 137 SNPs in windows explaining at least 2% of the additive genetic variance for fat content suggesting a better dissection of the genetic basis underlying variation in fat content. The discrepancies among the different GWA studies might be due to; 1) usage of different algorithms in the GWA studies, 2) large variation in the population size, 3) substantial difference in the capacity of the SNP arrays, 4) different marker densities in each study, 5) polygenic nature of muscle fat content, 6) different thresholds in each study, 7) usage of different window size in each study (Gonzalez-Pena et al 2016).

Compared to our WssGBLUP analysis, the single marker GWA analysis revealed less number of SNP markers associated with the variation in the muscle fat and moisture content. In addition, the two GWA approaches revealed different significant peaks associated with the traits of interest. This result is consistent with other studied traits, such as fillet firmness, protein content (Ali et al 2019), and bodyweight gain (Data will be published somewhere else) in rainbow trout. The potential factors associated with the observed heterogeneity between the two approaches are the different algorithms, thresholds, and windows size used in each approach. The WssGBLUB was more effective in dissecting the genetic architecture of the studied traits and identifying large common QTL between the traits. The common QTL may explain the high negative correlation between the fat and moisture content. The recombinational history of the QTL and nearby

markers determines the information content of haplotypes (Lorenz et al 2010). However, SNP-harboring genes identified by the two approaches had similar biological functions and were involved in lipid metabolism, protein turnover, and cytoskeletal remodeling. Routine use of single-SNP and multi-makers for GWA analysis has been previously recommended to take the advantage of whole information content of the genotypes (Lorenz et al 2010).

CONCLUSIONS

The current GWA analyses identified novel genomic loci affecting the additive genetic variance for muscle fat and moisture content in rainbow trout. These genomic loci code for proteins involved in lipid metabolism, cytoskeleton remodeling, and protein synthesis/degradation. Compared to a previous GWA study in Atlantic salmon, this work revealed large effect QTL affecting fat content which appears to be a polygenic trait. The top common windows affecting the additive genetic variance for fat and moisture content appeared on chromosome 19. These findings provide genetic basis to understand the molecular mechanisms underlying fat and moisture content in teleost fish. In addition, this work provided putative markers that could be prioritized when estimating genomic breeding values for fat and moisture content.

Author 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 and MS wrote the manuscript.

REFERENCES

Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B *et al* (2017). Identification of SNPs associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-Seq samples in rainbow trout. *BMC Genomics* **18:** 582.

Ali A, Al-Tobasei R, Kenney B, Leeds TD, Salem M (2018). Integrated analysis of lncRNA and mRNA expression in rainbow trout families showing variation in muscle growth and fillet quality traits. *Sci Rep* **8:** 12111.

Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M (2019). Genome-Wide Association Study Identifies Genomic Loci Affecting Filet Firmness and Protein Content in Rainbow Trout. *Frontiers in Genetics* **10**.

Andersen KM, Madsen L, Prag S, Johnsen AH, Semple CA, Hendil KB *et al* (2009). Thioredoxin Txnl1/TRP32 is a redox-active cofactor of the 26 S proteasome. *J Biol Chem* **284:** 15246-15254.

Ang CYW, Young LL, Wilson RJ (1984). Interrelationships of protein, Fat and Moisture Content of broiler Meat. *Food Sci* **49:** 359

Bendris N, Lemmers B, Blanchard JM (2015). Cell cycle, cytoskeleton dynamics and beyond: the many functions of cyclins and CDK inhibitors. *Cell Cycle* **14:** 1786-1798.

Bjursell MK, Blom HJ, Cayuela JA, Engvall ML, Lesko N, Balasubramaniam S *et al* (2011). Adenosine kinase deficiency disrupts the methionine cycle and causes hypermethioninemia, encephalopathy, and abnormal liver function. *Am J Hum Genet* **89:** 507-515.

Bonneau M, Lebret B (2010). Production systems and influence on eating quality of pork. *Meat Sci* **84:** 293-300.

Cahir-McFarland ED, Davidson DM, Schauer SL, Duong J, Kieff E (2000). NF-kappa B inhibition causes spontaneous apoptosis in Epstein-Barr virus-transformed lymphoblastoid cells. *Proc Natl Acad Sci U S A* **97:** 6055-6060.

Caillet M, Janvier K, Pelchen-Matthews A, Delcroix-Genête D, Camus G, Marsh M *et al* (2011). Rab7A is required for efficient production of infectious HIV-1. *PLoS Pathog* **7:** e1002347.

Cayrol C, Lacroix C, Mathe C, Ecochard V, Ceribelli M, Loreau E *et al* (2007). The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. *Blood* **109:** 584-594.

David Y, Ziv T, Admon A, Navon A (2010). The E2 ubiquitin-conjugating enzymes direct polyubiquitination to preferred lysines. *J Biol Chem* **285:** 8595-8604.

Diamond G, Cedar H, Marcus M (1989). A temperature-sensitive mutation in asparaginyl-tRNA synthetase causes cell-cycle arrest in early S phase. *Exp Cell Res* **184:** 53-60.

Ferreira V, van Dijk KW, Groen AK, Vos RM, van der Kaa J, Gijbels MJ *et al* (2007). Macrophage-specific inhibition of NF-kappaB activation reduces foam-cell formation. *Atherosclerosis* **192:** 283-290.

Finocchiaro G, Ito M, Tanaka K (1987). Purification and properties of short chain acyl-CoA, medium chain acyl-CoA, and isovaleryl-CoA dehydrogenases from human liver. *J Biol Chem* **262:** 7982-7989.

Flor AC, Wolfgeher D, Wu D, Kron SJ (2017). A signature of enhanced lipid metabolism, lipid peroxidation and aldehyde stress in therapy-induced senescence. *Cell Death Discov* **3:** 17075.

Florence L, Mireille C, Jérôme B, Laurent L, Françoise M, Edwige Q (2015). Selection for muscle fat content and triploidy affect flesh quality in pan-size rainbow trout, Oncorhynchus mykiss *Aquaculture* **448:** 569-577.

García-Celdrán M, Ramis G, Manchado M, Estévez A, Navarro A (2015). Estimates of heritabilities and genetic correlations of raw flesh quality traits in a reared gilthead sea bream (S*parus aurata* L.) population sourced from broodstocks along the Spanish coasts. *Aquaculture* **446:** 181-186.

Gaudet P, Livstone MS, Lewis SE, Thomas PD (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform* **12:** 449-462.

Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB *et al* (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (Oncorhynchus mykiss). *Front Genet* **7:** 203.

Greenberg DL, Mize GJ, Takayama TK (2003). Protease-activated receptor mediated RhoA signaling and cytoskeletal reorganization in LNCaP cells. *Biochemistry* **42:** 702-709.

Gross SR, Kinzy TG (2007). Improper organization of the actin cytoskeleton affects protein synthesis at initiation. *Mol Cell Biol* **27:** 1974-1989.

Han MH, Hu Z, Chen CY, Chen Y, Gucek M, Li Z *et al* (2014). Dysbindin-associated proteome in the p2 synaptosome fraction of mouse brain. *J Proteome Res* **13:** 4567-4580.

Hatakeyama S, Yada M, Matsumoto M, Ishida N, Nakayama KI (2001). U box proteins as a new family of ubiquitin-protein ligases. *J Biol Chem* **276:** 33111-33120.

Hepler PK (2016). The Cytoskeleton and Its Regulation by Calcium and Protons. *Plant Physiol* **170:** 3-22.

Hesse D, Jaschke A, Chung B, Schürmann A (2013). Trans-Golgi proteins participate in the control of lipid droplet and chylomicron formation. *Biosci Rep* **33:** 1-9.

Hinshaw JM (1999). Trout production: Feeds and feeding methods. *Southern Regional Aquaculture Center* **223**.

Huang L, Chardon JW, Carter MT, Friend KL, Dudding TE, Schwartzentruber J *et al* (2012). Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia. *Orphanet J Rare Dis* **7:** 67.

Hwangbo C, Wu J, Papangeli I, Adachi T, Sharma B, Park S *et al* (2017). Endothelial APLNR regulates tissue fatty acid uptake and is essential for apelin's glucose-lowering effects. *Sci Transl Med* **9**.

Johnson JA, Hemnes AR, Perrien DS, Schuster M, Robinson LJ, Gladson S *et al* (2012). Cytoskeletal defects in Bmpr2-associated pulmonary arterial hypertension. *Am J Physiol Lung Cell Mol Physiol* **302:** L474-484.

Joo JI, Yun JW (2011). Gene expression profiling of adipose tissues in obesity susceptible and resistant rats under a high fat diet. *Cell Physiol Biochem* **27:** 327-340.

Jung J, Bohn G, Allroth A, Boztug K, Brandes G, Sandrock I *et al* (2006). Identification of a homozygous deletion in the AP3B1 gene causing Hermansky-Pudlak syndrome, type 2. *Blood* **108:** 362-369.

Kause A, Kiessling A, Martin SA, Houlihan D, Ruohonen K (2016). Genetic improvement of feed conversion ratio via indirect selection against lipid deposition in farmed rainbow trout (Oncorhynchus mykiss Walbaum). *Br J Nutr* **116:** 1656-1665.

Komarova Y, De Groot CO, Grigoriev I, Gouveia SM, Munteanu EL, Schober JM *et al* (2009). Mammalian end binding proteins control persistent microtubule growth. *J Cell Biol* **184:** 691-706.

Leeds T, Kenney P, Manor M (2012). Genetic parameter estimates for feed intake, body composition, and fillet quality traits in a rainbow trout population selected for improved growth. *International Symposium on Genetics in Aquaculture*: Auburn, AL. p 259.

Leeds TD, Vallejo RL, Weber GM, Pena DG, Silverstein JS (2016). Response to five generations of selection for growth performance traits in rainbow trout (Oncorhynchus mykiss). *Aquaculture* **465:** 341-351.

Lemos MV, Chiaia HL, Berton MP, Feitosa FL, Aboujaoud C, Camargo GM *et al* (2016). Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. *BMC Genomics* **17:** 213.

Li YX, Cabling MM, Kang HS, Kim TS, Yeom SC, Sohn YG *et al* (2013). Comparison and correlation analysis of different Swine breeds meat quality. *Asian-Australas J Anim Sci* **26:** 905-910.

Lin JC (2015). RBM4-MEF2C network constitutes a feed-forward circuit that facilitates the differentiation of brown adipocytes. *RNA Biol* **12:** 208-220.

Lin NY, Lin TY, Yang WH, Wang SC, Wang KT, Su YL *et al* (2012a). Differential expression and functional analysis of the tristetraprolin family during early differentiation of 3T3-L1 preadipocytes. *Int J Biol Sci* **8:** 761-777.

Lin WJ, Yang CY, Li LL, Yi YH, Chen KW, Lin YC *et al* (2012b). Lysosomal targeting of phafin1 mediated by Rab7 induces autophagosome formation. *Biochem Biophys Res Commun* **417:** 35-42.

Liu S, Vallejo RL, Palti Y, Gao G, Marancik DP, Hernandez AG *et al* (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front Genet* **6:** 298.

Lo YH, Ho PC, Chen MS, Hugo E, Ben-Jonathan N, Wang SC (2013). Phosphorylation at tyrosine 114 of Proliferating Cell Nuclear Antigen (PCNA) is required for adipogenesis in response to high fat diet. *Biochem Biophys Res Commun* **430:** 43-48.

Lorenz AJ, Hamblin MT, Jannink JL (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One* **5:** e14079.

Lu D, Sargolzaei M, Kelly M, Vander Voort G, Wang Z, Mandell I *et al* (2013). Genome-wide association analyses for carcass quality in crossbred beef cattle. *BMC Genet* **14:** 80.

Machnicka B, Czogalla A, Hryniewicz-Jankowska A, Bogusławska DM, Grochowalska R, Heger E *et al* (2014). Spectrins: a structural platform for stabilization and activation of membrane channels, receptors and transporters. *Biochim Biophys Acta* **1838:** 620-634.

Manor M, Weber G, Salem M, Yao J, Aussanasuwannakul A, al. e (2012). Effect of sexual maturation and triploidy on chemical composition and fatty acid content of energy stores in female rainbow trout, O*ncorhynchus mykiss*. *Aquaculture* **364:** 312–321.

Manor ML, Cleveland BM, Kenney PB, Yao J, Leeds T (2015). Differences in growth, fillet quality, and fatty acid metabolism-related gene expression between juvenile male and female rainbow trout. *Fish Physiology and Biochemistry* **41:** 533-547.

Matsuo Y, Irie K, Kiyonari H, Okuyama H, Nakamura H, Son A *et al* (2013). The protective role of the transmembrane thioredoxin-related protein TMX in inflammatory liver injury. *Antioxid Redox Signal* **18:** 1263-1272.

Melody JL, Lonergan SM, Rowe LJ, Huiatt TW, Mayes MS, Huff-Lonergan E (2004). Early postmortem biochemical factors influence tenderness and water-holding capacity of three porcine muscles. *J Anim Sci* **82:** 1195-1205.

Merline R, Schaefer RM, Schaefer L (2009). The matricellular functions of small leucine-rich proteoglycans (SLRPs). *J Cell Commun Signal* **3:** 323-335.

Miano JM, Long X, Fujiwara K (2007). Serum response factor: master regulator of the actin cytoskeleton and contractile apparatus. *Am J Physiol Cell Physiol* **292:** C70-81.

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A *et al* (2018). *Manual for BLUPF90 family of programs*: Univ. Georg., Athens, USA.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., H. LD (2002). BLUPF90 and related programs (BGF90) [WWW Document], in Proceeding of 7th World Congress on Genetics Applied to Livestock Production (Montpellier).

Mora S, Yang C, Ryder JW, Boeglin D, Pessin JE (2001). The MEF2A and MEF2D isoforms are differentially regulated in muscle and adipose tissue during states of insulin deficiency. *Endocrinology* **142:** 1999-2004.

Nakaseko Y, Yanagida M (2001). Cell biology. Cytoskeleton in the cell cycle. *Nature* **412:** 291-292.

Novokhatska O, Dergai M, Tsyba L, Skrypkina I, Filonenko V, Moreau J *et al* (2013). Adaptor proteins intersectin 1 and 2 bind similar proline-rich ligands but are differentially recognized by SH2 domain-containing proteins. *PLoS One* **8:** e70546.

Ortega A, Roselló-Lletí E, Tarazón E, Molina-Navarro MM, Martínez-Dolz L, González-Juanatey JR *et al* (2014). Endoplasmic reticulum stress induces different molecular structural alterations in human dilated and ischemic cardiomyopathy. *PLoS One* **9:** e107635.

Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE *et al* (2015). Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing. *PLoS One* **10:** e0138435.

Pan Y, Cao F, Guo A, Chang W, Chen X, Ma W *et al* (2015). Endoplasmic reticulum ribosome-binding protein 1, RRBP1, promotes progression of colorectal cancer and predicts an unfavourable prognosis. *British journal of cancer* **113:** 763-772.

Paneru B, Ali A, Al-Tobasei R, Kenney B, Salem M (2018). Crosstalk among lncRNAs, microRNAs and mRNAs in the muscle 'degradome' of rainbow trout. *Sci Rep* **8:** 8416.

Patwari P, Emilsson V, Schadt EE, Chutkow WA, Lee S, Marsili A *et al* (2011). The arrestin domain-containing 3 protein regulates body mass and energy expenditure. *Cell Metab* **14:** 671-683.

Ponsuksili S, Jonas E, Murani E, Phatsara C, Srikanchai T, Walz C *et al* (2008). Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. *BMC Genomics* **9:** 367.

Porat-Shliom N, Milberg O, Masedunskas A, Weigert R (2013). Multiple roles for the actin cytoskeleton during regulated exocytosis. *Cell Mol Life Sci* **70:** 2099-2121.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559-575.

Qi H, Liu S, Guo C, Wang J, Greenaway FT, Sun MZ (2015). Role of annexin A6 in cancer. *Oncol Lett* **10:** 1947-1952.

Quinlan AR, Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841-842.

Salem M, Kenney PB, Rexroad CE, Yao J (2006). Molecular characterization of muscle atrophy and proteolysis associated with spawning in rainbow trout. *Comp Biochem Physiol Part D Genomics Proteomics* **1:** 227-237.

Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A *et al* (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS One* **7:** e36264.

Salem M, Manor ML, Aussanasuwannakul A, Kenney PB, Weber GM, Yao J (2013). Effect of sexual maturation on muscle gene expression of rainbow trout: RNA-Seq approach. *Physiol Rep* **1:** e00120.

Salem M, Al-Tobasei R, Ali A, Lourenco D, Gao G, Palti Y *et al* (2018). Genome-Wide Association Analysis With a 50K Transcribed Gene SNP-Chip Identifies QTL Affecting Muscle Yield in Rainbow Trout. *Frontiers in Genetics* **9**.

Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K *et al* (2019). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **47:** D23-D28.

Sevane N, Armstrong E, Cortés O, Wiener P, Wong RP, Dunner S *et al* (2013). Association of bovine meat quality traits with genes included in the PPARG and PPARGC1A networks. *Meat Sci* **94:** 328-335.

Sferra A, Baillat G, Rizza T, Barresi S, Flex E, Tasca G *et al* (2016). TBCE Mutations Cause Early-Onset Progressive Encephalopathy with Distal Spinal Muscular Atrophy. *Am J Hum Genet* **99:** 974-983.

Sodeland M, Gaarder M, Moen T, Thomassen M, Kjøglum S, Kent M *et al* (2013). Genome-wide association testing reveals quantitative trait loci for fillet texture and fat content in Atlantic salmon. *Aquaculture* **408–409:** 169–174.

Sontag E, Nunbhakdi-Craig V, Lee G, Bloom GS, Mumby MC (1996). Regulation of the phosphorylation state and microtubule-binding activity of Tau by protein phosphatase 2A. *Neuron* **17:** 1201-1207.

Sugiyama M, Takenaga F, Kitani Y, Yamamoto G, Okamoto H, Masaoka T *et al* (2012). Homozygous and heterozygous GH transgenesis alters fatty acid composition and content in the liver of Amago salmon (Oncorhynchus masou ishikawae). *Biol Open* **1:** 1035-1042.

Söhle J, Machuy N, Smailbegovic E, Holtzmann U, Grönniger E, Wenck H *et al* (2012). Identification of new genes involved in human adipogenesis and fat storage. *PLoS One* **7:** e31193.

Tan-Sindhunata MB, Mathijssen IB, Smit M, Baas F, de Vries JI, van der Voorn JP *et al* (2015). Identification of a Dutch founder mutation in MUSK causing fetal akinesia deformation sequence. *Eur J Hum Genet* **23:** 1151-1157.

Tandon P, Wafer R, Minchin JEN (2018). Adipose morphology and metabolic disease. *J Exp Biol* **221**.

Thibeaux S, Siddiqi S, Zhelyabovska O, Moinuddin F, Masternak MM, Siddiqi SA (2018). Cathepsin B regulates hepatic lipid metabolism by cleaving liver fatty acid-binding protein. *J Biol Chem* **293:** 1910-1923.

Trombetta ES, Ebersold M, Garrett W, Pypaert M, Mellman I (2003). Activation of lysosomal function during dendritic cell maturation. *Science* **299:** 1400-1403.

Turner SD (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*.

Vallejo RL, Liu S, Gao G, Fragomeni BO, Hernandez AG, Leeds TD *et al* (2017). Similar Genetic Architecture with Shared and Unique Quantitative Trait Loci for Bacterial Cold Water Disease Resistance in Two Rainbow Trout Breeding Populations. *Front Genet* **8:** 156.

Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S *et al* (2017). Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541:** 81-86.

Watanabe G, Motoyama M, Nakajima I, Sasaki K (2018). Relationship between water-holding capacity and intramuscular fat content in Japanese commercial pork loin. *Asian-Australas J Anim Sci* **31:** 914-918.

Wayne R, Tazawa M (1989). The Actin Cytoskeleton and Polar Water Permeability in Characean Cells. In: Tazawa M (ed). *Cell Dynamics: Molecular Aspects of Cell Motility Cytoskeleton in Cellular Structure and Activity*. Springer Vienna: Vienna. pp 116-130.

Wei B, Yang Z, Cheng Y, Zhou J, Yang H, Zhang L *et al* (2018). **Proteomic Analysis of the Hepatopancreas of Chinese Mitten Crabs (E*riocheir sinensis)* Fed With a Linoleic Acid or α-Linolenic Acid Diet**. *Front Physiol* **9:** 1430.

Wen C, Jiang XY, Ding LR, Wang T, Zhou YM (2017). Effects of dietary methionine on growth performance, meat quality and oxidative status of breast muscle in fast- and slow-growing broilers. *Poult Sci* **96:** 1707-1714.

Writzl K, Maver A, Kovačič L, Martinez-Valero P, Contreras L, Satrustegui J *et al* (2017). De Novo Mutations in SLC25A24 Cause a Disorder Characterized by Early Aging, Bone Dysplasia, Characteristic Face, and Early Demise. *Am J Hum Genet* **101:** 844-855.

Wu YH, Ko TP, Guo RT, Hu SM, Chuang LM, Wang AH (2008). Structural basis for catalytic and inhibitory mechanisms of human prostaglandin reductase PTGR2. *Structure* **16:** 1714-1723.

Yang Z, Bowles NE, Scherer SE, Taylor MD, Kearney DL, Ge S *et al* (2006). Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circ Res* **99:** 646-655.

Yu S, Chen L, Ye L, Fei L, Tang W, Tian Y *et al* (2014). Identification of two missense mutations of ERCC6 in three Chinese sisters with Cockayne syndrome by whole exome sequencing. *PLoS One* **9:** e113914.

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S *et al* (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* **9:** e1003520.

Zhang M, Wang D, Huang W, Liu F, Zhu Y, Xu W *et al* (2013). Apoptosis during postmortem conditioning and its relationship to duck meat quality. *Food Chem* **138:** 96-100.

Zhang WG, Lonergan SM, Gardner MA, Huff-Lonergan E (2006). Contribution of postmortem changes of integrin, desmin and μ-calpain to variation in water holding capacity of pork. *Meat Sci* **74:** 578-585.

Zhao F, Zhuang P, Song C, Shi Z, Zhang L (2010). Amino acid and fatty acid compositions and nutritional quality of muscle in the pomfret, P*ampus punctatissimus*. *Food Chemistry* **118:** 224–227

Zheng X, Kuang Y, Lv W, Cao D, Sun Z, Sun X (2016). Genome-Wide Association Study for Muscle Fat Content and Abdominal Fat Traits in Common Carp (Cyprinus carpio). *PLoS One* **11:** e0169127.

Zhu G, Ye R, Jung DY, Barron E, Friedline RH, Benoit VM *et al* (2013). GRP78 plays an essential role in adipogenesis and postnatal growth in mice. *FASEB J* **27:** 955-964.

Zhu Z-L, Yang Q-M, Li C, Chen J, Xiang M, Chen M-M *et al* (2018). Identification of biomarkers for childhood obesity based on expressional correlation and functional similarity. *Molecular medicine reports* **17:** 109-116.

APPENDICES

CHR	position	SNP	UNADJ	BONF	R2	Region/effect	Strand	Gene ID	Gene annotation	Trait
5	27016969	AX-171614410	9.22E-07	0.02295	0.01944	CDS/syn	+	LOC110523600	78 kDa glucose-regulated protein-like	Fat content
5	32672947	AX-171640304	7.70E-07	0.01915	0.0207	CDS/nonsyn	-	ska1	spindle and kinetochore associated complex subunit 1	Fat content
5	33025927	AX-172550832	8.51E-07	0.02117	0.02059	CDS/syn	-	LOC110523701	apelin receptor B-like	Fat content
5	77138939	AX-172547460	9.22E-07	0.02295	0.01944	CDS/syn	+	LOC110524668	desmoplakin-like	Fat content
5	78880913	AX-172563469	9.64E-07	0.024	0.01939	CDS/syn	-	podn	podocan	Fat content
5	79115842	AX-171640473	7.48E-07	0.01861	0.01919	CDS/nonsyn	+	LOC110524703	transmembrane protein 69-like	Fat content
17	13538660	AX-171635972	9.26E-07	0.02305	0.03189	CDS/syn	+	LOC110493605	puromycin-sensitive aminopeptidase-like	Fat content
5	26977667	AX-172559241	1.64E-06	0.04083	0.01809	3'UTR	-	LOC110523594	golgin subfamily A member 2-like	Fat content
5	31174527	AX-171640705	9.22E-07	0.02295	0.01944	3'UTR	-	LOC110523653	uncharacterized LOC110523653	Fat content
5	33296459	AX-172548475	8.60E-07	0.02141	0.02053	Intronic	-	LOC110523713	proliferating cell nuclear antigen-like	Fat content
5	33380551	AX-171614340	1.78E-06	0.0442	0.01946	3'UTR	-	LOC110523715	calponin homology domain-containing protein DDB_G0272472-like	Fat content
5	78062294	AX-171614749	8.95E-07	0.02227	0.01939	3'UTR	-	mafa	MAF bZIP transcription factor A	Fat content
5	79116860	AX-171614450	1.55E-06	0.03848	0.01851	3'UTR	-	LOC110524700	calcium-binding mitochondrial carrier protein SCaMC-1	Fat content

Appendix A: All SNP markers significantly associated with fat content using single-SNP analysis.

CHR	position	SNP	UNADJ	BONF	FDR_BH	FDR_BY	R2	Region/effect	Strand	Gene ID	Gene annotation	Trait
1	53970934	AX-89976809	1.20E-06	0.0298	0.001683	0.018	0.01495	CDS/syn	+	wdr11	WD repeat domain 11	Moisture content
1	54958459	AX-171620793	7.82E-07	0.01945	0.001683	0.018	0.01568	CDS/syn	-	LOC110525667	serum response factor-like	Moisture content
1	55318448	AX-171607566	1.49E-06	0.03702	0.001683	0.018	0.01798	CDS/syn	-	LOC110525737	adenosine kinase-like	Moisture content
1	55903395	AX-171617765	1.48E-06	0.03682	0.001683	0.018	0.01928	CDS/nonsyn	-	LOC110525889	actin cytoskeleton-regulatory complex protein pan1-like	Moisture content
1	55903477	AX-171607577	1.71E-06	0.04254	0.001746	0.01868	0.01844	CDS/nonsyn	-	LOC110525889	actin cytoskeleton-regulatory complex protein pan1-like	Moisture content
1	55904783	AX-171607580	1.73E-06	0.04303	0.001746	0.01868	0.01883	CDS/nonsyn	-	LOC110525889	actin cytoskeleton-regulatory complex protein pan1-like	Moisture content
17	13538660	AX-171635972	6.11E-09	0.0001522	0.0001522	0.001628	0.04227	CDS/syn	+	LOC110493605	puromycin-sensitive aminopeptidase-like	Moisture content
17	20678734	AX-171597057	5.77E-07	0.01437	0.001683	0.018	0.05677	CDS/nonsyn	+	rrbp1	ribosome binding protein 1	Moisture content
17	30098943	AX-171610672	1.26E-06	0.03141	0.001683	0.018	0.05504	CDS/syn	-	LOC110493954	eukaryotic translation initiation factor 4B-like	Moisture content
17	30108957	AX-172561910	1.01E-06	0.02508	0.001683	0.018	0.05601	CDS/syn	-	LOC110493954	eukaryotic translation initiation factor 4B-like	Moisture content
17	33113794	AX-171597081	2.27E-07	0.005645	0.001129	0.01208	0.02864	CDS/syn	-	LOC110494012	interferon-induced protein 44-like	Moisture content
24	23448901	AX-171625236	6.82E-07	0.01698	0.001683	0.018	0.0139	CDS/nonsyn	-	LOC110504051	PHD finger protein 12-like	Moisture content
1	54948051	AX-89924960	7.92E-07	0.0197	0.001683	0.018	0.01568	3'UTR	+	fam160b1	family with sequence similarity 160 member B1	Moisture content
1	55723708	AX-172563117	1.91E-06	0.04756	0.001746	0.01868	0.01863	Intergenic	N/A	N/A	N/A	Moisture content
1	55875918	AX-171632281	1.84E-06	0.04578	0.001746	0.01868	0.01852	3'UTR	+	ercc6	ERCC excision repair 6, chromatin remodeling factor	Moisture content
1	55908056	AX-171617775	5.07E-07	0.01262	0.001683	0.018	0.02156	Intergenic	N/A	N/A	N/A	Moisture content
4	15238170	AX-172548157	1.45E-06	0.03607	0.001683	0.018	0.03164	3'UTR	+	LOC110521463	transducin-like enhancer protein 1	Moisture content
12	65648063	AX-171626534	1.34E-06	0.03329	0.001683	0.018	0.04416	3'UTR	-	LOC110538135	putative all-trans-retinol 13,14-reductase	Moisture content
13	59522477	AX-171598839	1.81E-06	0.04515	0.001746	0.01868	0.0352	3'UTR	-	LOC110487423	glycogen [starch] synthase, muscle-like	Moisture content
17	17622851	AX-172560736	1.14E-06	0.02833	0.001683	0.018	0.02993	3'UTR	-	LOC110493736	calcineurin subunit B type 1	Moisture content
17	20335252	AX-171597071	9.51E-07	0.02367	0.001683	0.018	0.05552	3'UTR	-	LOC110493798	eukaryotic initiation factor 4A-III	Moisture content
17	21029747	AX-171597053	1.32E-08	0.0003285	0.0001642	0.001757	0.074	3'UTR	-	acyp2	acylphosphatase 2	Moisture content
17	21070906	AX-171597048	5.88E-08	0.001464	0.0003728	0.003989	0.03497	3'UTR	+	bdh2	3-hydroxybutyrate dehydrogenase 2	Moisture content
17	30097354	AX-172561310	1.37E-06	0.03396	0.001683	0.018	0.0343	3'UTR	-	LOC110493954	eukaryotic translation initiation factor 4B-like	Moisture content
17	30097441	AX-171610671	1.33E-06	0.03305	0.001683	0.018	0.05473	3'UTR	-	LOC110493954	eukaryotic translation initiation factor 4B-like	Moisture content
24	9580897	AX-171613971	3.26E-07	0.008122	0.001354	0.01448	0.01881	3'UTR	+	LOC110503690	F-box only protein 46-like	Moisture content
24	38518846	AX-171628695	5.99E-08	0.001491	0.0003728	0.003989	0.019	5'UTR	+	LOC110503440	probable ATP-dependent RNA helicase DDX6	Moisture content

Appendix B: All SNP markers significantly associated with moisture content using single-SNP analysis.

GENERAL CONCLUSIONS

Depletion of natural fisheries increased the import of seafood in USA over the last two decades (USDA 2005, USDC 2008). United States has become the major seafood importer leading to a trade deficit of ~ \$14 billion in 2008 (USDA 2005). Aquaculture has the potential to improve food security and expand economic opportunities worldwide (Burbridge et al 2001). However, lack of genetically improved strains of fish represents a major challenge to compete in the international market (Ali et al 2018). The aquaculture industry could benefit from developing germplasm with enhanced production traits such as disease resistance, growth rate, and fillet quality in order to meet the customer needs/expectations. Some companies apply traditional diet-based approaches in order to improve growth rate and other desirable fillet quality traits such as fat content (Manor et al 2012); however, these approaches may decrease the profitability and are difficult to use for breeding purposes. A selection approach based on phenotypic and genotypic data is necessary in order for the U.S. aquaculture industry to compete in the world-wide market. Therefore, there is a growing interest in developing genomic resources for fish, including rainbow trout, to improve the quantitative and qualitative production traits of fish.

This project aimed to identify genes and genetic variants affecting fish growth and fillet quality traits in rainbow trout. Muscle yield and quality traits are determinants of the aquaculture industry profitability and consumer satisfaction. These traits result from multifactorial interactions between environmental and genetic factors. Because of the polygenic nature of the muscle yield and quality traits and given that the largest part of the transcriptome is noncoding with evidence of the role of lncRNA in regulating myogenesis increasing, we sought to perform an integrated analysis of mRNA and lncRNA in fish

families showing divergent phenotypes for muscle yield and quality traits. To achieve our aim, we profiled transcriptome expression of fish families showing contrasting phenotypes in WBW, muscle yield, fat content, shear force and whiteness index. We identified some candidate protein-coding genes that were differentially expressed (DE) in fish families of contrasting phenotypes. However, the lncRNA showed higher variability in terms of expression between divergent families. Given the fact that lncRNAs are poorly conserved, we identified networks/hubs between DE lncRNAs and their overlapping, neighboring, or distantly located on the genome based on expression correlation analysis. Genes identified in this study are good candidates for future validation studies such as gene knockouts to verify their roles in controlling the phenotypic variations. This study also revealed complex microRNA sponge effects for lncRNA i.e. lncRNA compete with mRNA to sequester microRNAs, and thus protect mRNA from degradation. Consequently, sponge lncRNAs may contribute to fast/efficient growth rates by controlling genes belonging to protein catabolic/anabolic pathways. Further, promoter regions of DE lncRNAs and their co-expressed protein-coding genes harbored similar transcription factor binding sites (TFBS) necessary for muscle development, suggesting common transcription initiation mechanisms. Therefore, the current study highlights possible regulatory interactions exerted by noncoding RNAs to control expression of protein-coding genes that impact muscle quality traits, and adds additional layers of complexity that may help in understanding the molecular network of muscle development.

Identification of DE lncRNAs and mRNAs help explain variations between fish families of contrasting growth/fillet quality phenotypes. However, DE cannot be directly used for selective breeding purposes. New sequencing technologies have identified SNPs

that are widely distributed throughout the genome; this SNP distribution enabled construction of high density genetic maps (Al-Tobasei et al 2017, Palti et al 2014). Most of the genetic variation comes from SNPs that are highly adaptable to large-scale genotyping and therefore, most suitable for genome-wide association studies (Salem et al 2012). Therefore, a 50K SNP-chip has been developed and used to perform four GWA studies and identify genetic markers and genomic loci associated with the desired traits, that could be used to select breeding candidates. The first study was used to test the feasibility of using the 50K SNP-chip in GWA analysis to identify loci explaining genetic variance in muscle yield. This study identified several large-effect QTL regions explaining the additive genetic variance for muscle yield. The most significant QTLs were on chromosomes 14 and 16. Many of the annotated genes in the QTL regions were previously reported in other sciences as important regulators of muscle development and cell signaling. It is worth mentioning that no major QTLs were identified in a previous GWA study using a 57K genomic SNP panel on the same fish population indicating improved detection power of the 50K transcribed gene SNP-chip in the target trait and USDA population.

Atrophying muscle from gravid rainbow trout fish showed less shear force than muscle from sterile fish indicating a correlation between muscle firmness and rate of protein turnover. Therefore, the second GWA study aimed to identify quantitative trait loci affecting fillet firmness and protein content. Ryanodine receptor 3 (RYR3) harbored the highest number of SNPs affecting genetic variation in shear force and protein content. Common SNPs in the RYR3 gene were identified affecting the aforementioned fillet traits; this association suggests common mechanisms underlying the fillet shear force and protein

content. Genes harboring SNPs were mostly involved in calcium homeostasis, proteolytic activities, transcriptional regulation, chromatin remodeling, and apoptotic processes. Our data suggest a role for RYR3 in muscle firmness that may be considered for genomic- and marker-assisted selection in breeding programs of rainbow trout. Similarly, a correlation between development of pale, soft and exudative (PSE) meat and abnormality in calcium release mechanism in porcine muscle was previously reported as a result of a point mutation in RYR1 (Fujii et al 1991). Breeding strategies were initiated to avoid this mutation from the pig populations.

Traditional selection, based on the phenotype, has been applied to select for growth traits resulting in relatively slow genetic improvement over generations. Therefore, the third GWA study aimed to identify growth-related QTL and novel genes associated with bodyweight gain in rainbow trout that could be used for development of improved germplasm for aquaculture. Compared to previous GWA studies in Atlantic salmon and rainbow trout, this work revealed relatively larger-effect QTL associated with growth which appears to be a polygenic trait in nature, controlled by many genes on multiple chromosomes. The most significant peaks explaining a reasonable proportion of the additive genetic variance for bodyweight gain appeared on chromosomes 4 and 14. The gene harboring the most significant nonsynonymous SNP encodes 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.

Muscle composition (fat and moisture content) impacts quality attributes such as flavor, appearance, texture, and juiciness. Therefore, the fourth GWA analysis aimed to identify genomic loci associated with muscle fat and moisture content. A relatively large effect QTL associated with muscle fat and moisture content, which appear to be polygenic in nature, were identified. These genomic loci encode for proteins involved in lipid metabolism, cytoskeleton remodeling, and protein synthesis/degradation. The top common windows influencing the additive genetic variance for fat and moisture content appeared on chromosome 19. These findings provide a genetic basis to help in understanding the molecular mechanisms underlying fat and moisture content in teleost fish and provide putative markers that could be prioritized for genomic and marker-assisted selection.

The current GWA analysis identified many SNPs in windows explaining a reasonable proportion of the additive genetic variance for the studied traits of interest suggesting a better dissection of the genetic basis underlying variation in the phenotypes. To the best of our knowledge, those windows were not picked up in previous GWA analyses in fish. The differences between GWA studies might be due to: 1) usage of different algorithms in the GWA studies, 2) large variation in the population size, 3) substantial difference in the capacity of the SNP arrays, 4) different marker densities in each study, 5) polygenic nature of the traits, 6) different thresholds in each study, 7) usage of different window size in each study (Gonzalez-Pena et al 2016).

Compared to WssGBLUP analysis, the single marker GWA analysis revealed less number of SNP markers associated with the variation in the phenotypes. In addition, the two GWA approaches revealed different significant peaks associated with the traits of

interest. The heterogeneity between the two adopted GWA approaches was consistent in the different studied traits, such as fillet firmness, protein content, bodyweight gain, fat and moisture content. The potential factors associated with the observed heterogeneity between the two approaches are the different algorithms, thresholds, and windows size used in each approach. The WssGBLUB was more effective in dissecting the genetic architecture of the studied traits and identifying relatively large-effect QTL associated with the phenotypes. However, routine use of single-SNP and multi-makers for GWA analysis has been previously recommended to take the advantage of whole information content of the genotypes (Lorenz et al 2010).

To conclude, the present study explored genes and genetic variants associated with important aquaculture production traits in rainbow trout and suggests candidate markers for developing germplasm with enhanced phenotypes. Aquaculture industry is interested in implementing genomic selection to increase the rates of genetic gain and sustain industry competitiveness. Therefore, the 50K SNP chip will help to improve the accuracy of predicted genomic breeding values compared to traditional pedigree-based breeding values, hence allowing genomic prediction scenarios for growth- and muscle quality-related traits in rainbow trout.

PROJECT REFERENCES

Al-Tobasei R, Paneru B, Salem M (2016). Genome-Wide Discovery of Long Non-Coding RNAs in Rainbow Trout. *PLoS One* **11:** e0148940.

Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B *et al* (2017). Identification of SNPs associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-Seq samples in rainbow trout. *BMC Genomics* **18:** 582.

Ali A, Al-Tobasei R, Kenney B, Leeds TD, Salem M (2018). Integrated analysis of lncRNA and mRNA expression in rainbow trout families showing variation in muscle growth and fillet quality traits. *Sci Rep* **8:** 12111.

Baerwald MR, Petersen JL, Hedrick RP, Schisler GJ, May B (2011). A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (Oncorhynchus mykiss). *Heredity (Edinb)* **106:** 920-926.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B *et al* (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* **5:** 3657.

Burbridge P., Hendrick V., Roth E., H. R (2001). Social and economic policy issues relevant to marine aquaculture. *J Appl Ichthyol* **17:** 194–206.

Cabianca DS, Casa V, Gabellini D (2012). A novel molecular mechanism in human genetic disease: a DNA repeat-derived lncRNA. *RNA biology* **9:** 1211-1217.

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M *et al* (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147:** 358-369.

Dang CG, Cho SH, Sharma A, Kim HC, Jeon GJ, Yeon SH *et al* (2014). Genome-wide Association Study for Warner-Bratzler Shear Force and Sensory Traits in Hanwoo (Korean Cattle). *Asian-Australas J Anim Sci* **27:** 1328-1335.

Dey BK, Pfeifer K, Dutta A (2014). The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev* **28:** 491-501.

Drew R, Schwabl H, Wheeler P, Thorgaard G (2007). Detection of QTL influencing cortisol levels in rainbow trout (O*ncorhynchus mykiss*). *Aquaculture* **272:** 183-194

Dufflocqa P, Lhorenteb JP, Bangerab R, Neirac R, Newmand S, Yáñeza JM (2016). Correlated response of flesh color to selection for harvest weight in coho salmon (Oncorhynchus kisutch). *Aquaculture*.

Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV *et al* (2004). MicroRNA-143 regulates adipocyte differentiation. *J Biol Chem* **279:** 52361-52365.

Florence L, Mireille C, Jérôme B, Laurent L, Françoise M, Edwige Q (2015). Selection for muscle fat content and triploidy affect flesh quality in pan-size rainbow trout, Oncorhynchus mykiss *Aquaculture* **448:** 569-577.

Fornshell G (2002). Rai**nbow Trout — Challenges and Solutions**. pp *545–557*.

Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE *et al* (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253:** 448-451.

Geng X, Liu S, Yao J, Bao L, Zhang J, Li C *et al* (2016). A Genome-Wide Association Study Identifies Multiple Regions Associated with Head Size in Catfish. *G3 (Bethesda)* **6:** 3389-3398.

Gjedrem T (1983). Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture* **33:** 51–71

Gjedrem T (1992). Breeding plans for rainbow trout. In: Gall GAE (ed). *The Rainbow Trout: Proceedings of the First Aquaculture-sponsored Symposium held at the Institute of Aquaculture, University of Sterling, Scotland.* pp 73–83.

Gong C, Li Z, Ramanujan K, Clay I, Zhang Y, Lemire-Brachat S *et al* (2015). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev Cell* **34:** 181-191.

Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB *et al* (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (Oncorhynchus mykiss). *Front Genet* **7:** 203.

Gutierrez AP, Yáñez JM, Fukui S, Swift B, Davidson WS (2015). Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (Salmo salar). *PLoS One* **10:** e0119730.

Haidle L, Janssen JE, Gharbi K, Moghadam HK, Ferguson MM, Danzmann RG (2008). Determination of quantitative trait loci (QTL) for early maturation in rainbow trout (Oncorhynchus mykiss). *Mar Biotechnol (NY)* **10:** 579-592.

Han X, Yang F, Cao H, Liang Z (2015). Malat1 regulates serum response factor through miR-133 as a competing endogenous RNA in myogenesis. *FASEB J* **29:** 3054-3064.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS *et al* (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106:** 9362-9367.

Houston RD, Gheyas A, Hamilton A, Guy DR, Tinch AE, Taggart JB *et al* (2008). Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (Salmo salar). *Dev Biol (Basel)* **132:** 199-204.

Hu G, Gu W, Bai QL, Wang BQ (2013). Estimation of genetic parameters for growth traits in a breeding program for rainbow trout (Oncorhynchus mykiss) in China. *Genet Mol Res* **12:** 1457-1467.

Huang CW, Li YH, Hu SY, Chi JR, Lin GH, Lin CC *et al* (2012). Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia (Oreochromis niloticus). *J Anim Sci* **90:** 4266-4279.

Jin Y, Zhou T, Geng X, Liu S, Chen A, Yao J *et al* (2017). A genome-wide association study of heat stress-associated SNPs in catfish. *Anim Genet* **48:** 233-236.

Kambara H, Niazi F, Kostadinova L, Moonka DK, Siegel CT, Post AB *et al* (2014). Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res* **42:** 10668-10680.

Kause A, Paananen T, Ritola O, Koskinen H (2007). Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. *J Anim Sci* **85:** 3218-3227.

Kause A, Kiessling A, Martin SA, Houlihan D, Ruohonen K (2016). Genetic improvement of feed conversion ratio via indirect selection against lipid deposition in farmed rainbow trout (Oncorhynchus mykiss Walbaum). *Br J Nutr* **116:** 1656-1665.

Le Bras Y, Dechamp N, Krieg F, Filangi O, Guyomard R, Boussaha M *et al* (2011). Detection of QTL with effects on osmoregulation capacities in the rainbow trout (Oncorhynchus mykiss). *BMC Genet* **12:** 46.

Leeds TD, Vallejo RL, Weber GM, Pena DG, Silverstein JS (2016). Response to five generations of selection for growth performance traits in rainbow trout (Oncorhynchus mykiss). *Aquaculture* **465:** 341-351.

Li T, Wang S, Wu R, Zhou X, Zhu D, Zhang Y (2012). Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics* **99:** 292-298.

Li Z, Rana TM (2014). Decoding the noncoding: prospective of lncRNA-mediated innate immune regulation. *RNA biology* **11:** 979-985.

Liu S, Vallejo RL, Palti Y, Gao G, Marancik DP, Hernandez AG *et al* (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front Genet* **6:** 298.

Lorenz AJ, Hamblin MT, Jannink JL (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One* **5:** e14079.

Lu L, Sun K, Chen X, Zhao Y, Wang L, Zhou L *et al* (2013). Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. *EMBO J* **32:** 2575-2588.

Manor M, Weber G, Salem M, Yao J, Aussanasuwannakul A, al. e (2012). Effect of sexual maturation and triploidy on chemical composition and fatty acid content of energy stores in female rainbow trout, O*ncorhynchus mykiss*. *Aquaculture* **364:** 312–321.

Mathieu EL, Belhocine M, Dao LT, Puthier D, Spicuglia S (2014). [Functions of lncRNA in development and diseases]. *Medecine sciences : M/S* **30:** 790-796.

Moen T, Baranski M, Sonesson AK, Kjoglum S (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. *BMC Genomics* **10:** 368.

Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR *et al* (2015). MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. *Mol Cell Biol* **35:** 498-513.

Nichols KM, Broman KW, Sundin K, Young JM, Wheeler PA, Thorgaard GH (2007). Quantitative trait loci x maternal cytoplasmic environment interaction for development rate in Oncorhynchus mykiss. *Genetics* **175:** 335-347.

O'Malley K, Sakamoto T, Danzmann R, Ferguson M (2003). Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. *J Heredity* **94:** 273-284

Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E *et al* (2014). A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Mol Ecol Resour* **14:** 588-596.

Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE *et al* (2015). Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing. *PLoS One* **10:** e0138435.

Paneru B, Al-Tobasei R, Palti Y, Wiens GD, Salem M (2016). Differential expression of long non-coding RNAs in three genetic lines of rainbow trout in response to infection with Flavobacterium psychrophilum. *Sci Rep* **6:** 36032.

Price AH (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci* **11:** 213-216.

Rasmussen RS (2001). Quality of farmed salmonids with emphasis on proximate composition, yield and sensory characteristics *Aquac Res* **32: 767–786**.

Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R *et al* (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res* **16:** 11-19.

Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A *et al* (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS One* **7:** e36264.

Salem M, Al-Tobasei R, Ali A, Lourenco D, Gao G, Palti Y *et al* (2018). Genome-Wide Association Analysis With a 50K Transcribed Gene SNP-Chip Identifies QTL Affecting Muscle Yield in Rainbow Trout. *Frontiers in Genetics* **9**.

Schielzeth H, Husby A (2014). Challenges and prospects in genome-wide quantitative trait loci mapping of standing genetic variation in natural populations. *Ann N Y Acad Sci* **1320:** 35-57.

Sodeland M, Gaarder M, Moen T, Thomassen M, Kjøglum S, Kent M *et al* (2013). Genome-wide association testing reveals quantitative trait loci for fillet texture and fat content in Atlantic salmon. *Aquaculture* **408–409:** 169–174.

Somorjai IM, Danzmann RG, Ferguson MM (2003). Distribution of temperature tolerance quantitative trait loci in Arctic charr (Salvelinus alpinus) and inferred homologies in rainbow trout (Oncorhynchus mykiss). *Genetics* **165:** 1443-1456.

Taylor JF, Taylor KH, Decker JE (2016). Holsteins are the genomic selection poster cows. *Proc Natl Acad Sci U S A* **113:** 7690-7692.

Tsai HY, Hamilton A, Guy DR, Tinch AE, Bishop SC, Houston RD (2015a). The genetic architecture of growth and fillet traits in farmed Atlantic salmon (Salmo salar). *BMC Genet* **16:** 51.

Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ *et al* (2015b). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* **16:** 969.

USDA FAS, Commodity & Marketing Programs. (2005). International Trade Report. http://www.fas.usda.gov/ffpd/Fish-Circular/Market_News/IATR_Seafood_Imports.pdf.

USDC (2008). U.S. INTERNATIONAL TRADE IN GOODS AND SERVICES-. In: Analysis USDoC-USCBUSBoE (ed).

Vallejo RL, Rexroad CE, Silverstein JT, Janss LL, Weber GM (2009). Evidence of major genes affecting stress response in rainbow trout using Bayesian methods of complex segregation analysis. *J Anim Sci* **87:** 3490-3505.

Vallejo RL, Wiens GD, Rexroad CE, Welch TJ, Evenhuis JP, Leeds TD *et al* (2010). Evidence of major genes affecting resistance to bacterial cold water disease in rainbow trout using Bayesian methods of segregation analysis. *J Anim Sci* **88:** 3814-3832.

Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP *et al* (2017). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet Sel Evol* **49:** 17.

Wang L, Zhao Y, Bao X, Zhu X, Kwok YK, Sun K *et al* (2015). LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Res* **25:** 335-350.

Watts R, Johnsen VL, Shearer J, Hittel DS (2013). Myostatin-induced inhibition of the long noncoding RNA Malat1 is associated with decreased myogenesis. *American journal of physiology Cell physiology* **304:** C995-1001.

Wei N, Pang W, Wang Y, Xiong Y, Xu R, Wu W *et al* (2014). Knockdown of PU.1 mRNA and AS lncRNA regulates expression of immune-related genes in zebrafish Danio rerio. *Developmental and comparative immunology* **44:** 315-319.

Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG (2010). Growth-related quantitative trait loci in domestic and wild rainbow trout (Oncorhynchus mykiss). *BMC Genet* **11:** 63.

Wu Z, Liu X, Liu L, Deng H, Zhang J, Xu Q *et al* (2014). Regulation of lncRNA expression. *Cellular & molecular biology letters* **19:** 561-575.

Yan B, Guo JT, Zhu CD, Zhao LH, Zhao JL (2013). miR-203b: a novel regulator of MyoD expression in tilapia skeletal muscle. *J Exp Biol* **216:** 447-451.

Yang X, Gao L, Guo X, Shi X, Wu H, Song F *et al* (2014). A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS One* **9:** e87797.

Yoshida GM, Lhorente JP, Carvalheiro R, Yáñez JM (2017). Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (Salmo salar L.). *Anim Genet* **48:** 698-703.

Yu H, You X, Li J, Zhang X, Zhang S, Jiang S *et al* (2018). A genome-wide association study on growth traits in orange-spotted grouper (Epinephelus coioides) with RAD-seq genotyping. *Science China Life sciences* **61:** 934-946.

Zhao W, Mu Y, Ma L, Wang C, Tang Z, Yang S *et al* (2015). Systematic identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development. *Sci Rep* **5:** 8957.

Zheng X, Kuang Y, Lv W, Cao D, Sun Z, Sun X (2016). Genome-Wide Association Study for Muscle Fat Content and Abdominal Fat Traits in Common Carp (Cyprinus carpio). *PLoS One* **11:** e0169127.

Zhong X, Wang X, Zhou T, Jin Y, Tan S, Jiang C *et al* (2017). Genome-Wide Association Study Reveals Multiple Novel QTL Associated with Low Oxygen Tolerance in Hybrid Catfish. *Mar Biotechnol (NY)* **19:** 379-390.

Zhou M, Wang X, Li J, Hao D, Wang Z, Shi H *et al* (2015). Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. *Molecular bioSystems* **11:** 760-769.

Zhu J, Fu H, Wu Y, Zheng X (2013). Function of lncRNAs and approaches to lncRNA-protein interactions. *Science China Life sciences* **56:** 876-885.

Zimmerman AM, Evenhuis JP, Thorgaard GH, Ristow SS (2004). A single major chromosomal region controls natural killer cell-like activity in rainbow trout. *Immunogenetics* **55:** 825-835.