bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 # Genome-wide association analysis with a 50K transcribed gene SNP-chip identifies
2 # QTL affecting muscle yield in rainbow trout

3 Mohamed Salem[1,2*] Rafet Al-Tobasei[2,3], Ali Ali[1], Daniela Lourenco[4], Guangtu Gao[5], Yniv Palti[5], Brett
4 Kenney[6], and Timothy D. Leeds[5]

5 [1]Department of Biology and Molecular Biosciences Program, Middle Tennessee State University,
6 Murfreesboro, TN 37132, U.S.

7 [2]Computational Science Program, Middle Tennessee State University, Murfreesboro, TN 37132, U.S.

8 [3]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294-0022, U.S.

9 [4]Department of Animal and Dairy Science, University of Georgia, Athens GA 30602, U.S.

10 [5]The National Center for Cool and Cold Water Aquaculture, USDA Agricultural Research Service,
11 Kearneysville, WV 25430, U.S.

12 [6]Division of Animal and Nutritional Science, West Virginia University, Morgantown, 26506-6108, WV,
13 U.S.

14

15

16 *Correspondence:

17 Mohamed Salem

18 Department of Biology and Molecular Biosciences Program,

19 Middle Tennessee State University,

20 Murfreesboro, TN, 37132, U.S.

21 Mohamed.salem@mtsu.edu

22

23

24

25

26

27

28

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

29 **Abstract**

30 Detection of coding/functional SNPs that change the biological function of a gene may lead to
31 identification of putative causative alleles within QTL regions and discovery of genetic markers
32 with large effects on phenotypes. Two bioinformatics pipelines, GATK and SAMtools, were used
33 to identify ~21K transcribed SNPs with allelic imbalances associated with important aquaculture
34 production traits including body weight, muscle yield, muscle fat content, shear force, and
35 whiteness in addition to resistance/susceptibility to bacterial cold-water disease (BCWD). SNPs
36 were identified from pooled RNA-Seq data collected from ~620 fish, representing 98 families from
37 growth- and 54 families from BCWD-selected lines with divergent phenotypes. In addition, ~29K
38 transcribed SNPs without allelic-imbalances were strategically added to build a 50K Affymetrix
39 SNP-chip. SNPs selected included two SNPs per gene from 14K genes and ~5K non-synonymous
40 SNPs. The SNP-chip was used to genotype 1728 fish. The average SNP calling-rate for samples
41 passing quality control (QC; 1,641 fish) was ≥ 98.5%. Genome-wide association (GWA) study on
42 878 fish (representing 197 families from 2 consecutive generations) with muscle yield phenotypes
43 and genotyped for 35K polymorphic markers (passing QC) identified several QTL regions
44 explaining together up to 28.40% of the additive genetic variance for muscle yield in this rainbow
45 trout population. The most significant QTLs were on chromosomes 14 and 16 with 12.71% and
46 10.49% of the genetic variance, respectively. Many of the annotated genes in the QTL regions
47 were previously reported as important regulators of muscle development and cell signaling. No
48 major QTLs were identified in a previous GWA study using a 57K genomic SNP chip on the same
49 fish population. These results indicate improved detection power of the transcribed gene SNP-chip
50 in the target trait and population, allowing identification of large-effect QTLs for important traits
51 in rainbow trout.

52

53 **Introduction**

54 Aquaculture provides sustainable production of food fish with high protein/low-saturated fat to
55 satisfy increasing U.S. and worldwide demand. To enable increased production by the aquaculture
56 industry and to meet the ever-growing demand for fish, we need fast/efficient growth and high-
57 quality fillets. However, a major constraint to increasing production efficiency is the lack of
58 genetically improved strains of fish for aquaculture [1; 2]. Development of tools that will enable
59 genomic selection for improved aquaculture production traits will greatly benefit the aquaculture
60 industry.

61

62 Fast/efficient muscle growth is a major trait affecting profitability of the aquatic muscle food
63 industry. The genetic basis of muscle growth traits is not well studied in fish. Understanding
64 molecular mechanisms of fish muscle growth can facilitate broodstock selection decisions. In
65 addition, fish fillets are nutritionally and economically valuable products. Skeletal muscle is the
66 most abundant tissue and edible portion of fish and typically constitutes about 50-60% of the fish
67 weight [3]. Growth, development and quality traits of muscle are governed by organized
68 expression of genes encoding contractile and regulatory proteins [4].

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

69

70 Genomic selection (GS) tools have been developed to increase the efficiency of genetic
71 improvement in livestock compared to conventional pedigree-based selective breeding
72 methods[5]. This concept has been recently demonstrated for bacterial cold-water disease
73 (BCWD) resistance in rainbow trout aquaculture [6]. Marker-assisted selection (MAS) can be used
74 to improve breeding for phenotypes with large-effect QTLs. This method has been recently applied
75 for the trait of infectious pancreatic necrosis virus (IPNV) resistance in Atlantic salmon [7].
76 Genetic maps, characterizing the inheritance patterns of traits, and markers have been developed
77 and used for a wide range of species, including fish. These tools target the discovery of allelic
78 variation affecting traits with an ultimate goal of identifying DNA sequences underlying
79 phenotypes [8]. Markers have been identified with a variety of molecular techniques. Single
80 nucleotide polymorphisms (SNPs) are abundant and distributed genome-wide, therefore, they are
81 most suitable for high-throughput association studies [9; 10]. SNPs located within or near coding
82 sequences, cSNPs, are especially important because they have the potential to change protein
83 function [11; 12; 13]. Therefore, cSNPs are particularly useful as genetic markers with large-
84 effect on phenotypes, allowing MAS and improved accuracy of whole-genome selection. Because
85 the muscle yield trait targeted in this study requires lethal sampling to measure the phenotype, only
86 family-specific EBVs are available for breeding candidates in traditional breeding programs. The
87 ability to use genomic selection or MAS will allow further within-family selection for the muscle
88 yield trait, and thus is anticipated to increase the accuracy of genetic predictions and selection
89 response.

90

91 Recently, we used an RNA-Seq approach to identify putative SNPs with allelic imbalances
92 associated with total body weight, muscle yield, muscle fat content, shear force, and whiteness
93 [12; 13]. Similarly, RNA-Seq data were used to identify SNPs with allelic imbalances in fish
94 families showing variations in resistance to *Flavobacterium psychrophilum,* the etiological agent
95 of BCWD in rainbow trout [14; 15]. Together about 50K and 229K transcribed SNPs were
96 identified in the two studies, respectively. Of them, ~21K SNPs had allelic-imbalances in families
97 with contrasting phenotypes. The first objective of this study was to design, develop, and validate
98 a 50K transcribed gene SNP-chip. The chip content includes the 21K transcribed SNPs with allelic-
99 imbalances associated with the aforementioned traits and ~29K SNPs without allelic-imbalances
100 that were strategically added to achieve more even genome-wide distribution. The new SNP-Chip
101 is available from Affymetrix. The second objective of this study was to test the feasibility of
102 using the new SNP-chip in GWA analysis to identify QTL explaining muscle yield variance in the
103 USDA/NCCCWA rainbow trout growth-selected line. The results were compared with a previous
104 GWA study for the same trait in the same population that we have previously conducted with a
105 genomic-based 57K SNP chip [10].

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

106 # Materials and methods

107 *ETHICS STATEMENT*

108 Institutional Animal Care and Use Committee of the United States Department of Agriculture,
109 National Center for Cool and Cold Water Aquaculture (Leetown, WV) specifically reviewed and
110 approved all husbandry practices and experimental procedures used in this study (Protocols #056
111 and 076).

112

113 *SOURCE AND SELECTION OF SNPS FOR THE CHIP*

114 Recently, we used RNA-Seq and two bioinformatic pipelines, GATK and SAMtools, for
115 discovering coding/functional SNPs from 98 rainbow trout fish families (5 fish each) showing
116 variations in whole-body weight, muscle yield, muscle fat content, shear force, and whiteness [13].
117 GATK detected 59,112 putative SNPs and SAMtools detected 87,066 putative SNPs. The two
118 datasets contained approximately 50K non-redundant common SNPs; of which, 30,529 mapped to
119 protein-coding genes (with 7.7% non-synonymous SNPs) and 4,386 mapped to lncRNAs. A total
120 of 7,930 non-redundant SNPs had allelic imbalances between the low- and high-ranked families
121 for the phenotypes. Validation of a subset of 92 SNPs revealed 1) 86.7-93.8% success rate in
122 identifying polymorphic SNPs and 2) 95.4% consistent matching between DNA and cDNA
123 genotypes, indicating a high rate of identifying SNPs using RNA-Seq. This SNP data set was
124 recently published and is available through the NCBI dbSNP database (accession numbers
125 ss#2711191806-2711287038 in addition to ss#2137497773) [13].

126

127 Similarly, we identified transcribed gene SNPs in two genetic lines, ARS-Fp-R (resistant) and
128 ARS-FP-S (susceptible), that were created by selective breeding to exhibit divergent resistance to
129 BCWD. RNA-Seq analysis of pooled RNA samples was used to identify SNPs from the resistant
130 and susceptible genetic lines. Fish belonging to resistant and susceptible genetic lines were
131 collected on day 1 and day 5 post-challenge with Fp versus PBS injection [14; 15]. Using GATK
132 bioinformatics pipelines, ~229K transcribed SNPs were identified[14]. The total number of SNPs
133 with allelic imbalance, after removing redundant SNPs, was 7,951.

134

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

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

145 *CHIP GENOTYPING QUALITY ASSESSMENT*

146 The SNP-chip was used in genotyping 1,728 fish from the USDA-ARS genetic lines. The
147 Affymetrix SNPolisher software was used to calculate the chip SNP- and sample-metrics and
148 assess QCs and filter samples/genotypes at the default setting [16]. Forty-seven SNPs previously
149 genotyped by a Fluidigm PCR-based assay [13] were used to check quality of Affymetrix chip
150 genotyping using 120 samples genotyped by both the chip and Fluidigm SNP assays. In addition,
151 we confirmed the quality of the SNPs and the order of the samples included in the genotyping
152 panel through pedigree check. Among the fish genotyped we included previously confirmed
153 parental-pairs of nine families with 470 offspring and confirmed an average of 99.4% matching
154 between offspring SNP genotypes and the genotypes of the expected parents.
155

156 *SNP GENOMIC DISTRIBUTION AND ANNOTATION*

157 SNPs used in building the chip were identified using the first draft of the rainbow trout reference
158 genome [14]. To update genomic coordinates according to the newly released genome assembly
159 (GenBank assembly Accession GCA_002163495, RefSeq assembly accession GCF_002163495)
160 [17], SNPs were mapped by BLASTing the SNP probe sequences (70 nt) to the new genome
161 sequence. Sequences with 100% identity match and no gap with single hits were assigned to the
162 new genome position. Sequences with multiple hits were re-Blasted using probe size of 150 nt by
163 adding 40 nt flanking sequence in both direction. A total of 45K SNPs out of 50K SNPs were
164 successfully assigned to the new genome and were used for the GWA analyses.
165

166 SNPeff program was used to classify and annotate functional effects of the SNPs [18]. The gff
167 file of the new rainbow trout genome reference was used to determine position of the SNPs in a
168 gene i.e. located within mRNA start and end positions (genic), within a CDS, 5'UTR or 3'UTR.
169 SNPs not within start and end positions of mRNA were considered intergenic. Upstream/
170 downstream intergenic SNPs were determined if located within 5 Kb of an mRNA. SNPs within
171 lncRNAs were determined using gtf file of our previously reported lncRNA reference [63]. SNP
172 annotation was performed by intersecting the SNPs bed file with the gff/gtf file using Bedtools
173 software [19].
174

175 *RAINBOW TROUT POPULATION AND PHENOTYPES USED FOR GWA*

176 Genome-wide association analysis was carried out using fish from a growth-selected line that
177 has been previously described [20]. Briefly, this synthetic line is a 2-yr-old winter/spring-
178 spawning population that was developed beginning in 2002, became a closed population in 2004,
179 and since then has gone through 5 generations of genetic selection for improved growth
180 performance. Fish from two consecutive generations (i.e., the third and fourth generations of
181 growth selection) were included in this study. Phenotypic data and DNA samples were collected
182 from 878 fish (representing 98 families from year-class (YC) 2010 and 99 families from YC 2012).
183 Methods used to sample fish from each nucleus family and to characterize muscle yield have been
184 described previously [10]. Eggs were hatched in spring water at 7-13°C to synchronize hatch

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

185 times. Each family was stocked separately in 200-L tanks and hand-fed a commercial fishmeal-
186 based diet beginning at swim-up. Neomales were developed from a subset of alevins from the
187 previous year class by feeding 2 mg/kg of 17α-methyltestosterone for 60 d post-swim-up, and the
188 masculinized females were used as sires for the following generation. At 5-months old, fish were
189 uniquely tagged by inserting a passive integrated transponder, and tagged fish were combined and
190 reared in 1,000-L communal tanks. Fish were fed a commercial fishmeal-based diet using
191 automatic feeders. EBV were computed based on a two-trait model, 10-mo BW and thermal
192 growth coefficient (TGC), using MTDFREML [21]. Each generation, EBV was used as selection
193 criterion and mating decisions were made to maximize genetic gain while constraining the
194 inbreeding rate to ≤1% per generation using EVA evolutionary algorithm [22]. Data from
195 masculinized fish were not used in the growth analysis.
196
197 Fish were harvested between 410 and 437 days post-hatch (mean body weight = 985 g; SD =
198 239 g), between 446 and 481 days post-hatch (mean body weight = 1803 g; SD = 305 g), for the
199 2010, and 2012 hatch years, respectively. Individual body weight data were recorded at
200 harvesting. Fish were taken off feed 5 days before harvesting. For measurement of muscle yield
201 when harvested at each of five consecutive weeks, approximately 100 fish (i.e., 1 fish per full-sib
202 family per week) were anesthetized in approximately 100 mg/L of tricaine methane sulfonate
203 (Tricaine-S, Western Chemical, Ferndale, WA) slaughtered, and eviscerated. Head-on gutted
204 carcasses were packed in ice, transported to the West Virginia University Muscle Foods Processing
205 Laboratory (Morgantown, WV), and stored overnight. The next day, carcasses were manually
206 processed into trimmed, skinless fillets by a trained faculty member and weighed; muscle yield
207 was calculated as a percent of total body weight [23].
208

209 *GWA ANALYSES*

210

211 Weighted single-step GBLUP (WssGBLUP) was used to perform GWA analysis as
212 implemented in previous studies [24; 25; 26]. In addition to phenotypic data, wssGBLUP
213 integrates genotype and pedigree information to increase estimation precision and detection power
214 [25] in a combined analysis that is executed by the BLUPF90 software [27].
215 The following mixed model was used for single trait analysis:
216

217 $$y = Xb + Z_1a + Z_2w + e$$

218

219 where y is the vector of the phenotypes, b is the vector of fixed effects including harvest group
220 and hatch year, a is the vector of additive direct genetic effects (i.e., animal effect), w is the vector
221 of random family effect, and e is the residual error. The matrices X, $Z_1$, and $Z_2$ are incidence
222 matrices for the effects contained in b, a, and w, respectively.
223 As BLUP considers the variance components are known, AIREMLF90 [27] was used to
224 estimate variance components for the additive direct genetic effect, random family effect, and

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

225 residuals. Inbreeding was considered in all analyses, and was calculated using INBUPGF90 [27]
226 on 63,808 fish that represent five generations in the NCCCWA population. Quality control (QC)
227 of genomic data was performed using BLUBF90 [27] with the following parameters: SNP with
228 minimum Allele Frequency (MAF) >0.05, SNP with call rate > 0.90, animals with call rate >0.90,
229 and SNP with a difference between observed and expected allele frequency <0.15 (i.e., HWE test)
230 were kept in the data. Out of a total of 50,006 SNPs, 35,322 SNPs passed QC. For the first iteration
231 of WssGBLUP, all SNPs were assigned the same weight (e.g., 1.0). For the next iteration, weights
232 were calculated based on the SNP effects ($\hat{u}$) estimated in the previous iteration as $\hat{u}^2 2p(1-p)$,
233 where p is the current allele frequency. Each iteration was performed using three steps as follows:
234 first, weight was assigned as described above; second, BLUPF90 [27] was used to compute
235 genomic estimated breeding values (GEBV) based on a realized relationship matrix (**H**) that
236 combines pedigree (**A**) and genomic relationship matrix (**G**), the last considered weights for SNP;
237 and third, postGSF90 [27] was used to calculate SNP effects and weights based on sliding
238 windows of 50 adjacent SNPs. A total of 2 iterations were used. A window based on physical size
239 (e.g. 1Mb) was not used to avoid biases due to uneven distributed SNPs in the new SNP chip. A
240 Manhattan plot based on the proportion of additive genetic variance explained by the windows
241 was created using the qqman package in R [28]; thus, the genomic windows explaining significant
242 proportion of the additive genetic variance for muscle yield could be detected.
243

244 *CITRATE SYNTHASE (CS) ACTIVITY ASSAY*

245 GWA analysis (described below) showed a SNP window contained the CS gene associated with
246 the genetic variance in muscle yield. To assess the potential effect of the SNPs in this gene, we
247 measured the CS activity in 100 fish from the 2012 year-class. Frozen muscle tissue samples were
248 homogenized using electric homogenizer on ice followed by centrifugation at 1,000g for 15 min
249 at 4 °C. The supernatant was used to assess the total protein concentration and CS activity. Total
250 protein concentration was assessed using a BCA protein assay kit at 562 nm with bovine serum
251 albumin (BSA) as the standard. CS activity was determined from the rate of appearance of reduced
252 DTNB (5,5`-dithiobis [2-nitrobenzoic]), which was monitored with a spectrophotometer at 412
253 nm[29]. For the CS assay, 10 µL of diluted tissue homogenate (1.0 mg/ ml) was incubated with
254 140 µL reaction medium (0.1mM DTNB, 0.2mM AcetylCoA, 0.15mM oxaloacetic acid, pH 8.0).
255 The absorbance was read in triplicate at 412 nm (25 °C) after 4 min. CS activity was expressed as
256 ΔOD/ mg protein.
257

## 258 Results and discussion

259

260 *CHIP GENOTYPING QUALITY ASSESSMENT*

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

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

264

265 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 |

266

267 The Affymetrix SNPolisher software was used to filter samples/genotypes at the default setting
268 [16]. Out of 1,728 genotyped samples, 1,641 (94.9%) fish samples were retained, and 87 samples
269 were filtered out because they failed to meet the 0.97 call rate (CR) and 0.82 Dish QC (DQC)
270 thresholds. The average QC call rate for the passing samples was 99.6% (Table 2).
271

272 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 |

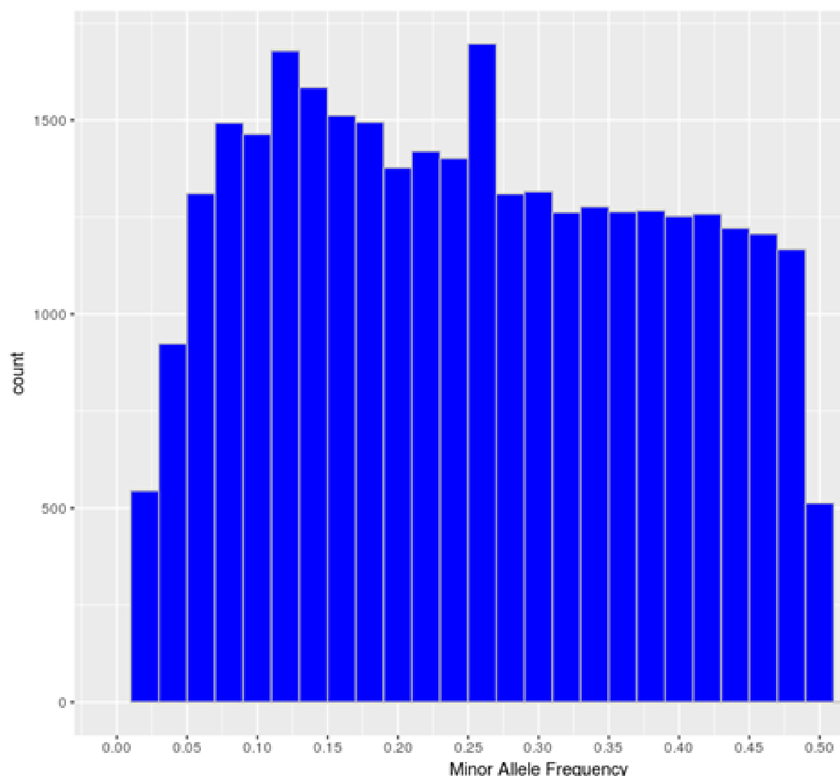
273
274
275
276
277

278

279 We compared the Affymetrix genotyping results of 47 SNPs that were previously genotyped
280 by a Fluidigm PCR-based assay [13]. Using 120 samples genotyped by both methods, there was a
281 99.5% match in genotypes between the two assays for high-resolution polymorphic markers (data
282 not shown). This test demonstrates the high quality of the SNP chip and reliable genotyping data
283 for the subsequent GWA analyses.
284

285 The SNP-chip showed an average minor allele frequency (MAF) of 0.25 and standard deviation
286 of 0.134. A total of 27,280 SNPs had MAF> 0.1 and 16,101 SNP more than 0.25 (Figure 1).

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.




287

288 Figure 1. Minor allele frequency distribution of the polymorphic high-resolution SNPs in the SNP chip.

289

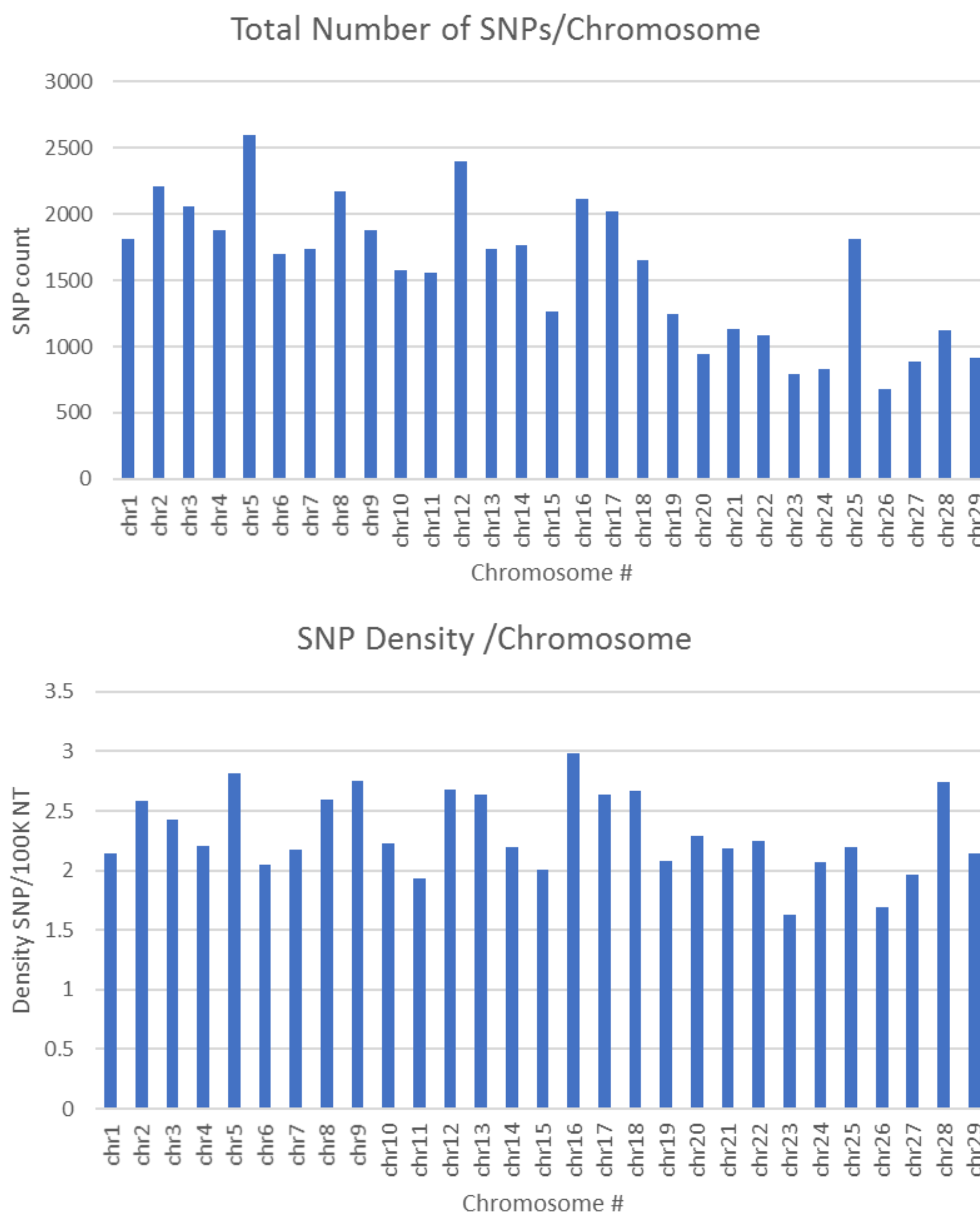
290

291 *SNP DENSITY AND GENOMIC DISTRIBUTION*

292 SNPs used in this study to build the chip were initially identified using a rainbow trout reference
293 genome published by Berthelot et al. in 2014 [30]. However, in this reference only ~1 Gb out of
294 a 2.1 Gb total length of the assembly is anchored to chromosomes. Recently, a newer genome
295 assembly has been built that is currently available at NCBI (Accession GCA_002163495) [17].
296 The new assembly has a 1.94 Gb total length (89% of the genome) anchored to 29 chromosomes.
297 A total of 45,590 SNPs out of 50,006 existing in the SNP-chip were mapped to the new genome
298 assembly with an average of 1,572 SNPs per chromosome. The average SNP density was 1 SNP
299 per 42.7 Kb, with a range of 1 SNP/33.5 Kb (Chromosome 16) to 1 SNP/61.6 Kb (Chromosome
300 23). Figure 2 shows the number of SNPs per chromosome and the SNP density distribution. A
301 total of 21K out 50K SNPs on the chip were selected based on putative association with phenotypic
302 traits, and hence, were expected to be clustered in specific genome loci. However, supplementing
303 the chip with 29K SNPs (two SNPs per gene) perhaps helped in randomizing the SNP distribution
304 in the genome. Previously, a 57K genome-wide SNP array for rainbow trout reported an average
305 of 1,551.4 mapped SNPs per chromosome [10]. The 57K array was designed primarily using SNPs
306 originating from RAD-Seq sequencing of doubled-haploid clonal lines [31] and whole genome re-
307 sequencing of fish from the Aquagen (Norway) breeding program. A key point here, is that the

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

308 SNPs included in the 57K chip were originated from other genetic lines. Hence, although
309 polymorphic enough in the NCCCWA growth line used in this study for conducting GWA as we
310 have previously shown[10], the SNPs used for GWA in this study were originated from the
311 investigated population and were expected to be more informative due to ascertainment bias [32].
312




313

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

315

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

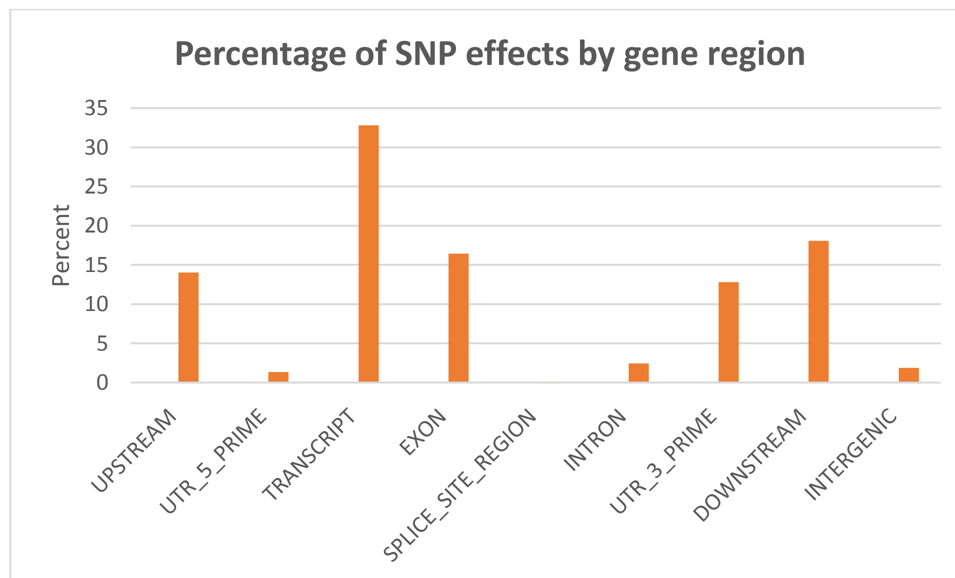
316 *SNP ANNOTATION AND CLASSIFICATION BASED ON FUNCTIONAL EFFECTS*

317

318 SNPeff program was used to classify and annotate functional effects of the SNPs. A total of
319 45,590 SNPs were included in this analysis. Classifying SNPs by impact showed 636 effects
320 (0.23%) with high impact (stop-gain) and 20,987 effects (7.86%) with moderate impact, (missense
321 variants). The rest (91.9%) represents low to moderate variant effect including synonymous and
322 non-coding SNPs. Figure 3 shows percent of SNP effects by gene regions. A total of 32.8% of the
323 effects were within transcripts with 16.5% exonic, 1.3% in the 5'-UTR and 12.8% in 3'-UTR. All
324 SNPs on the chip were identified through transcriptome sequencing. Surprisingly, there were 14%
325 upstream and 18.1% downstream effects (within 5 Kb of the genes). The upstream/downstream
326 percent is consistent with our previous report that showed 17.1-20.2% SNPs within 5 Kb
327 upstream/downstream of protein-coding genes in one of two SNP data sets used in building the
328 SNP-chip [13]. On the other hand, there was only 1.9% of the SNP effects within intergenic
329 regions, compared to 37.7-49.2% intergenic SNPs in the previous study[13]. In our previous study,
330 the high percentage of intergenic and upstream/downstream SNPs was explained by the
331 incomplete annotation of protein-coding genes and exons used in the previous version of the
332 rainbow trout reference genome [30]. The drop in the percentage of intergenic SNP effects in this
333 study may be due to the improved gene annotation of the current version of the genome reference.

334

335




336

337 Figure 3. Percentage of SNP effects by gene region.

338

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

339 *GWA ANALYSES*

340 ## Genomic regions associated with muscle yield

341 GWA analysis using WssGBLUP identified 163 SNPs, each explaining at least 2% of the
342 genetic variance of muscle yield (Tables 3 and 4; Supplementary data sheet 1). The SNPs were
343 clustered into 4 main chromosomes (14, 16, 9, and 17). Chromosomes 14 and 16 showed the
344 highest peaks with genomic loci explaining up to 12.71% and 10.49% of the genetic variance,
345 respectively. The total variance explained by these loci is 23.2%. Figure 4 shows a Manhattan plot
346 displaying association between SNP genomic sliding window of 50 SNPs and muscle yield. Sixty-
347 nine of the 163 SNPs (42.2%) were previously identified as SNPs with allelic imbalances
348 associated with muscle yield in the original SNP data set used to build the SNP chip [13]. Twenty-
349 one of the 163 SNPs caused nonsynonymous mutations. The rest of the SNPs were either silent
350 mutations or located in UTR of the genes indicating their potential epigenetic mechanism of gene
351 regulation. Important SNPs with more than 5% genetic variance are discussed below and all 163
352 SNPs are listed in Supplementary data sheet 1.

353

354 With 46 SNPs clustered into 23 annotated genes, chromosome 14 had the most significant QTL
355 windows explaining up to 12.71% of the genetic variance in muscle yield (Table 3 and Figure 4).
356 At least four genes can be inferred to be involved in cell differentiation/proliferation and regulation
357 of gene expression based on their RefSeq annotation. The list included fibroblast growth factor-
358 binding protein-1(FGFBP1) which had a single nonsynonymous SNP found in a window that
359 explained 12.24% of the additive genetic variance. FGFBP1 plays an essential role in cell
360 proliferation and differentiation by binding to fibroblast growth factors. The FGFBP1 expression
361 increases during development and decreases before neuromuscular junction degeneration during
362 aging [33]. The list of genes on chromosome 14 also includes inositol polyphosphate 5-
363 phosphatase (OCRL-1). OCRL is involved in terminating the PI3K signaling and thus plays an
364 important role in modulating effects of growth factors and insulin stimulation in cell proliferation
365 and survival [34]. Prominin-1-A gene (PROM1) that encodes for a transmembrane glycoprotein
366 had 2 SNPs. PROM1, often used as adult stem cell marker, plays a role in maintaining stem cell
367 properties by suppressing differentiation [35]. Another gene on chromosome 14 was
368 farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha (FNTA) which had a SNP
369 explaining 12.36% of the variance. FNTA may positively regulate neuromuscular junction
370 development [36].

371

372 In addition, chromosome 14 had three genes involved in the cell cycle regulation. The first gene
373 is MCTS1 re-initiation and release factor that had two SNPs in a window explaining 12.65% of
374 the additive genetic variance. MCTS1 is anti-oncogene that decreases cell doubling time by
375 shortening the G1 and G1/S transit time [37]. The second cell cycle control gene was cyclin-A2
376 which promotes transition through G1/S and G2/M and can block muscle-specific gene expression
377 during muscle differentiation [38]. The third gene was glutathione S-transferase P (GSTP1).

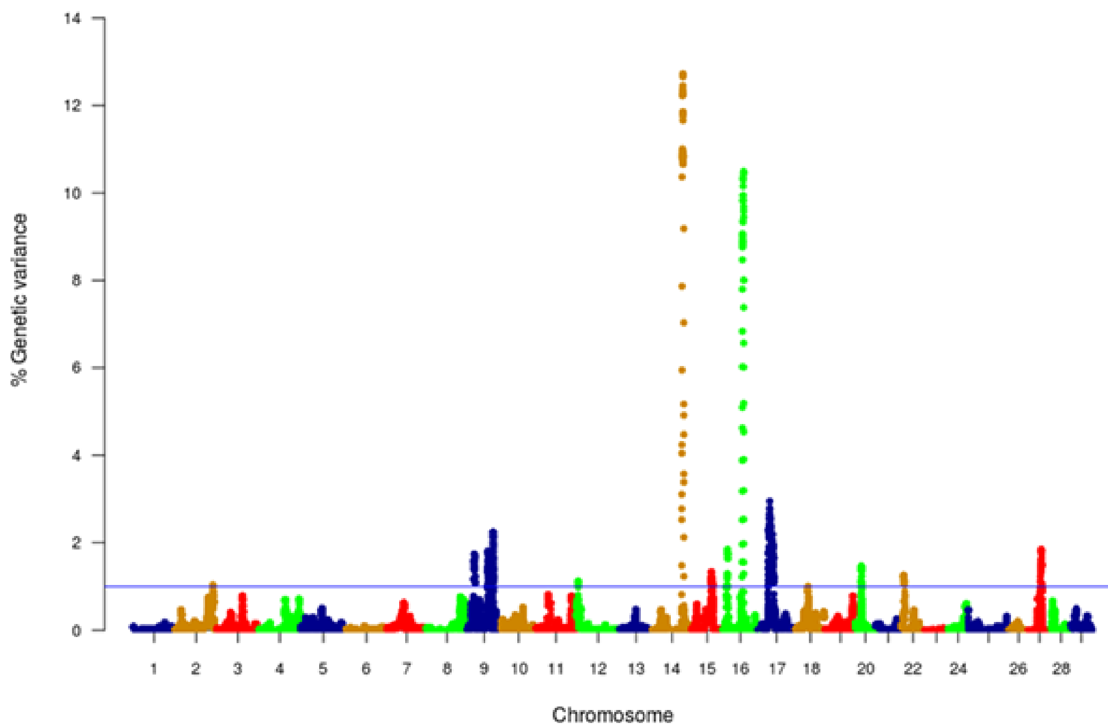
bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

378 Although involved in numerous biological functions, GSTP1 negatively regulates CDK5 activity
379 via p25/p35 translocation which diminishes neurodegeneration [39].
380
381 Chromosome 14 also had SNPs in genes playing important mitochondrial functions. There were
382 4 SNPs in the gene encoded for the electron transfer flavoprotein dehydrogenase (ETFDH) which
383 is an important enzyme in the mitochondrial electron transport chain. Mutations in ETFDH are
384 associated with myopathies[40]. Another mitochondrial-relevant gene was peptidylprolyl
385 isomerase D (PPID). Mutations in PPID are associated with muscular dystrophy in human [41].
386
387 Few other genes included in the QTL region on chromosome 14 are important for maintenance
388 of the muscle functions. Of them is the chloride intracellular channel protein 2 (CLIC2) which
389 modulates the activity of ryanodine receptor 2 (RYR2) and inhibits calcium influx, and therefore
390 is involved in regulating muscle contraction [35]. Five SNPs were in the lysosomal-associated
391 membrane protein 2 gene (LAMP2). LAMP2 mutations were reported in patients with
392 cardioskeletal myopathies [36]. Two SNPs were located in the UPF3B gene, a regulator of
393 nonsense-mediated mRNA decay (NMD). NMD inhibition was observed in patients with muscular
394 dystrophy [37]. Three SNPs were observed in the septin-6 gene. Mutations of septin-9 (another
395 gene family member) is genetically linked to muscle atrophy [38]. Two SNPs were identified in
396 the tenomodulin gene which showed downregulation in an animal muscle atrophy model [39].
397
398 Chromosome 16 ranked second in having the most significant QTL windows with 49 SNPs
399 clustered into 16 annotated genes (Table 4 and Figure 4). The gene within the most significant
400 SNP window to additive genetic variance was the cysteine/serine-rich nuclear protein 2
401 (CSRNP2). CSRNP2 has DNA binding transcription factor/activation activity. Deletion of
402 CSRNP1/2/3 three gene family members resulted in mice neonatal lethality [42]. Another gene
403 within the same SNP window was solute carrier family 26 member 9 (Slc26a9). Little is known
404 about the function of Slc26a9 in muscle, it serves as anion exchanger mediating chloride, sulfate
405 and oxalate transport and chloride/bicarbonate exchange [43]. A single SNP was observed in the
406 stem cell marker CD34a gene. Cd34(-/-) mice showed a defect in muscle regeneration caused by
407 acute or chronic muscle injury [44].
408
409 Several genes were involved in cell signaling/receptor activity. Five SNPs were predicted in 2
410 genes of the immune-related complement activation pathway, these are the complement receptor
411 and C4b-binding protein alpha chain. Recent studies indicated that the complement is activated as
412 a response of skeletal muscle injury and plays a key role during muscle regeneration[45]. A single
413 SNP was identified in the tyrosine-protein phosphatase non-receptor type 12 (PTPN12) which
414 dephosphorylates a wide-range of proteins, and thus regulates several cellular signaling cascades
415 such as ERBB2 and PTK2B/PYK2 [46]. This group of genes also includes the membrane-
416 associated guanylate kinase, WW and PDZ domain-containing protein 3 (MAGI3), which is
417 involved in the regulation of various cell signaling processes including the AKT1, TGFA, ERK

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

418 and JNK signaling cascades [47]. Two SNPs were in the basement membrane-specific heparan
419 sulfate proteoglycan core protein (HSPG2). A mouse model deficient in this gene showed muscle
420 hypertrophy through reduced myostatin expression suggesting a role in maintaining fast muscle
421 mass and fiber composition [48]. Five SNPs were in the TNF receptor superfamily member 5A
422 gene. Recently, some proinflammatory cytokines belonging to TNF superfamily have been
423 recognized as an important regulator of skeletal muscle mass [49].
424
425 Chromosome 16 also had a single SNP in the DAZ-associated protein 2 (DAZAP2). Not much
426 is known about the DAZAP2 function in muscle, however, DAZAP2 interacts with the
427 transforming growth factor-beta signaling molecule SARA (Smad anchor for receptor activation),
428 eukaryotic initiation factor 4G, and an E3 ubiquitinase [47]. Another gene in the list was Rac
429 GTPase-activating protein 1 (RACGAP1) that harbored 2 SNPs explaining up to 9.658% of the
430 genetic variance. RACGAP1 regulates cytokinesis and cell differentiation [50]. A single SNP
431 existed in caspase-9 which has an important non-apoptotic role in muscle differentiation [51].
432 Three SNPs were located in the kelch protein 21. Several Kelch family members play important
433 roles in skeletal muscle development by regulating the cell proliferation and/or differentiation [52].
434
435 An important gene affecting muscle function which is also located within the QTL region on
436 chromosome 16 is the citrate synthase (CS), which is used as a marker for human mitochondrial
437 functions. Ten SNPs explaining up to 8% of the genetic variance were located in the CS gene.
438
439
440
441

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.




442
443 Figure 4. Manhattan plot of GWA analysis performed with WssGBLUP and showing association between
444 SNP genomic sliding windows of 50 SNPs and muscle yield. Chromosomes 14 and 16 showed the highest
445 peaks with genomic loci, explaining together up to 23.2% of the genetic variance. The blue line shows a
446 threshold of 1% of additive genetic variance explained by SNPs.

447

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

448 Table 3. Selected SNP markers explaining the largest proportion of genetic variance (>5%) for muscle yield
449 in chromosome 14 using 50 adjacent SNP windows. Color intensities reflect changes in additive genetic
450 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 |

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

451 Table 4. Selected SNP markers explaining the largest proportion of genetic variance (>5%) for muscle yield
452 in chromosome 16 using 50 adjacent SNP windows. Color intensities reflect changes in additive genetic
453 variance (green is the highest and red is the lowest).

| Variance % | CHR | SNP position | Distance to next SNP | Strand | Gene | Annotation | Region/Effect |
|---|---|---|---|---|---|---|---|
| 4.62 | 16 | 39953311 | 12000 | + | tnfrsf5a | Tnf Receptor Superfamily Member 5A Precursor | 5'UTR |
| 5.09 | 16 | 39965311 | 3 | + | tnfrsf5a | Tnf Receptor Superfamily Member 5A Precursor | CDS/nonSyn |
| 6.03 | 16 | 39965314 | 689 | + | tnfrsf5a | Tnf Receptor Superfamily Member 5A Precursor | CDS/nonSyn |
| 6.83 | 16 | 39966003 | 608 | + | tnfrsf5a | Tnf Receptor Superfamily Member 5A Precursor | CDS/nonSyn |
| 7.79 | 16 | 39966611 | 666 | + | tnfrsf5a | Tnf Receptor Superfamily Member 5A Precursor | 3'UTR |
| 8.47 | 16 | 39967277 | 149527 | + | tnfrsf5a | Tnf Receptor Superfamily Member 5A Precursor | 3'UTR |
| 8.76 | 16 | 40116804 | 438 | NA | NA | NA | NA |
| 9.06 | 16 | 40117242 | 5021 | - | LOC110492067 | Kelch Protein 21 | CDS/syn |
| 9.04 | 16 | 40122263 | 471 | - | LOC110492067 | Kelch Protein 21 | CDS/syn |
| 9.04 | 16 | 40122734 | 206269 | - | LOC110492067 | Kelch Protein 21 | CDS/syn |
| 8.97 | 16 | 40329003 | 423 | + | LOC110492070 | 45 Kda Calcium-Binding Protein | 3'UTR |
| 8.88 | 16 | 40329426 | 430961 | + | LOC110492070 | 45 Kda Calcium-Binding Protein | 3'UTR |
| 8.88 | 16 | 40760387 | 133719 | + | LOC100136676 | Caspase-9 | CDS/syn |
| 8.87 | 16 | 40894106 | 16043 | + | LOC110491067 | Basement Membrane-Specific Heparan Sulfate Proteoglycan Core Protein | CDS/syn |
| 8.81 | 16 | 40910149 | 15660 | + | LOC110491067 | Basement Membrane-Specific Heparan Sulfate Proteoglycan Core Protein | CDS/nonSyn |
| 8.81 | 16 | 40925809 | 134 | NA | NA | NA | NA |
| 8.82 | 16 | 40925943 | 328 | NA | NA | NA | NA |
| 8.88 | 16 | 40926271 | 36300 | NA | NA | NA | NA |
| 8.88 | 16 | 40962571 | 1603 | + | LOC110492082 | Cdp-Diacylglycerol--Serine O-Phosphatidyltransferase | 3'UTR |
| 8.88 | 16 | 40964174 | 1011 | NA | NA | NA | NA |
| 8.89 | 16 | 40965185 | 134 | NA | NA | NA | NA |
| 8.89 | 16 | 40965319 | 214946 | NA | NA | NA | NA |
| 9.33 | 16 | 41180265 | 15995 | + | LOC110492084 | Membrane-Associated Guanylate Kinase, Ww And Pdz Domain-Containing Protein 3 | CDS/syn |
| 9.37 | 16 | 41196260 | 49825 | - | LOC110492085 | Tyrosine-Protein Phosphatase Non-Receptor Type 12 | CDS/syn |
| 9.82 | 16 | 41246085 | 3112 | + | LOC100136105 | Complement Receptor | CDS/syn |
| 9.82 | 16 | 41249197 | 474 | + | LOC100136105 | Complement Receptor | 3'UTR |
| 9.83 | 16 | 41249671 | 30475 | + | LOC100136105 | Complement Receptor | 3'UTR |
| 9.95 | 16 | 41280146 | 574 | + | c4bp | C4B-Binding Protein Alpha Chain | 3'UTR |
| 9.93 | 16 | 41280720 | 774 | + | c4bp | C4B-Binding Protein Alpha Chain | 3'UTR |
| 10.15 | 16 | 41281494 | 229 | NA | NA | NA | NA |
| 10.29 | 16 | 41281723 | 24001 | NA | NA | NA | NA |
| 10.33 | 16 | 41305724 | 20095 | - | LOC110492088 | Uncharacterized Loc110492088 | NA |
| 10.36 | 16 | 41325819 | 685099 | - | cd34a | Cd34A Molecule | 3'UTR |
| 10.44 | 16 | 42010918 | 5137 | - | slc26a9 | Solute Carrier Family 26 Member 9 | CDS/nonSyn |
| 10.45 | 16 | 42016055 | 176696 | - | slc26a9 | Solute Carrier Family 26 Member 9 | CDS/syn |
| 10.49 | 16 | 42192751 | 41683 | - | LOC110492098 | Cysteine/Serine-Rich Nuclear Protein 2 | CDS/syn |
| 9.58 | 16 | 42234434 | 23274 | + | LOC110492102 | Daz-Associated Protein 2 | 3'UTR |
| 9.68 | 16 | 42257708 | 1026 | - | LOC110492103 | Rac Gtpase-Activating Protein 1 | 3'UTR |
| 9.45 | 16 | 42258734 | 38505 | - | LOC110492103 | Rac Gtpase-Activating Protein 1 | 3'UTR |
| 8.01 | 16 | 42297239 | 2891 | + | LOC110492108 | Citrate Synthase, Mitochondrial | CDS/nonSyn |
| 8.01 | 16 | 42300130 | 5927 | + | LOC110492108 | Citrate Synthase, Mitochondrial | CDS/nonSyn |
| 7.38 | 16 | 42306057 | 101 | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |
| 6.57 | 16 | 42306158 | 92 | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |
| 6.01 | 16 | 42306250 | 1 | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |
| 5.19 | 16 | 42306251 | 60 | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |
| 4.54 | 16 | 42306311 | 303 | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |
| 3.90 | 16 | 42306614 | 605 | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |
| 3.19 | 16 | 42307219 | 57 | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |
| 2.53 | 16 | 42307276 | | + | LOC110492108 | Citrate Synthase, Mitochondrial | 3'UTR |

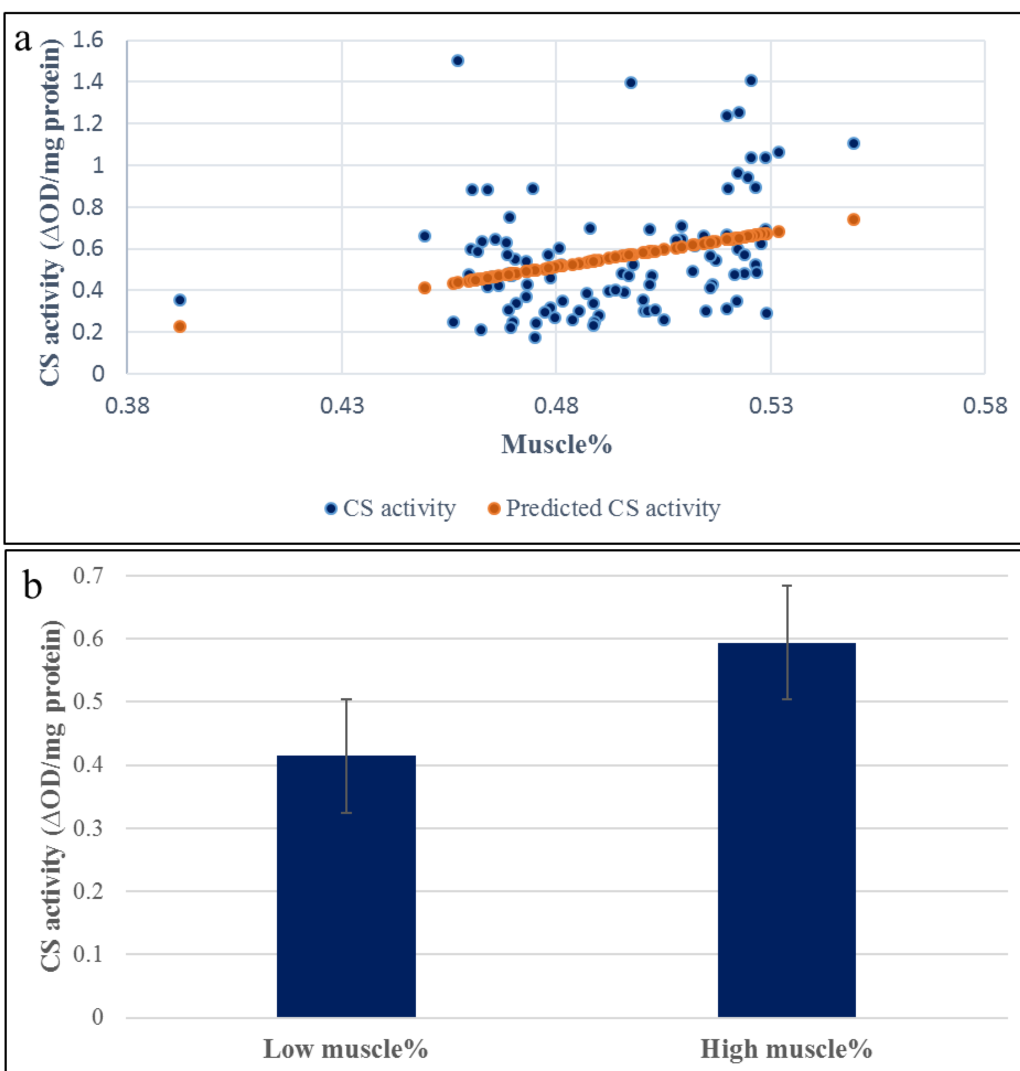
454

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

455 GWA studies in fish to identify QTL affecting muscle yield and quality are still in its infancy.
456 Previous GWA analysis using a 57K genomic SNP chip on the same fish population identified two
457 windows that explained 1.5% and 1.0% of the additive genetic variance for muscle yield and 1.2%
458 and 1.1% for muscle weight. Interestingly, the windows are located on chromosome 9, which
459 showed some association with muscle yield in the current study; however, none of the SNPs were
460 annotated to the same genes. No major QTLs were identified in the previous study. This large
461 difference in the outcomes of the two studies was somewhat unexpected. However, it may be
462 explained by lower marker density within or near genes in the 57K chip [10] and by ascertainment
463 bias, because the transcribed SNPs used in this study were discovered in the phenotyped fish and
464 hence are expected to be more polymorphic and informative for GWA analysis in this population.
465 Additionally, in this study, sliding windows of 50 SNP were used contrasting with 20 non-sliding
466 windows in the previous study. Difference in window size slightly contributed to the increased
467 proportion of variance (data not shown). By using SNP windows, it is assumed that those DNA
468 blocks may be inherited together, which may not always be the case for all assumed windows. In
469 common carp, genetic linkage mapping identified QTLs with large effects for muscle fiber cross-
470 section area (21.9%) and muscle fiber density (18.9%) [53]. Genome-wide significant QTL
471 affecting growth and muscle related traits were identified in Atlantic salmon [54]. The latter two
472 studies, together with our study, indicate existence of large-effect QTLs affecting muscle yield in
473 aquaculture species. However, the QTLs identified in this study might be population specific and
474 thus, need to be tested in other populations.
475
476 ## Citrate synthetase activity correlation with muscle yield
477 A SNP window on chromosomes 16 explaining up to 8.01% of the genetic variance in muscle
478 yield contained ten SNPs of the CS gene. Two of the SNPs were nonsynonymous mutations. To
479 investigate the potential effect of these SNPs, we measured the CS activity in 100 fish from the
480 2012 year-class. The samples included 38 fish from 5 high-ranked and 5 low-ranked families for
481 muscle yield and 62 randomly selected fish. CS had 1.43-fold increase in the high-ranked fish
482 compared to the low ranked ones (figure 5). The regression coefficient $R^2$ value between the
483 muscle yield and CS activity was 0.092 (p-value 0.002). These results indicate an important role
484 of mitochondrial functions to muscle growth. Mitochondria are at the center of age-related
485 sarcopenia that is characterized by decline in human muscle mass. Skeletal muscle CS decreases
486 with aging in humans [55].
487

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

488




489

490 Figure 5. Correlation coefficient between muscle yield and CS activity in 96 samples. A) The regression
491 coefficient R2 value between the muscle yield and CS activity was 0.092 (p-value 0.002). b) CS had 1.43-
492 fold increase in the high-ranked fish compared to the low ranked ones.

## 493 CONCLUSION

494 This study provides a 50K transcribed gene SNP-chip based on RNA-Seq data from fish
495 families showing genetic diversity for six aquaculture production traits in the USDA/NCCCWA
496 growth- and disease-selected genetic lines. The chip was tested for GWA analysis, which led to
497 identification of large-effect QTL for muscle yield in that population. Other muscle quality traits
498 are currently under investigation. Collectively, these studies will allow the use of SNP markers to
499 estimate breeding values for muscle yield and quality traits that are economically important traits
500 for aquatic food producers, processors, and consumers. Current and future selection at the
501 NCCCWA will select for improved fillet yield. Genetic markers are desirable for these traits
502 because genetic improvement is limited by the inability to measure fillet yield traits directly on

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

503 broodstock due to lethal sampling. Hence the accuracy and efficiency of selective breeding can
504 be improved by taking advantage of the genomic information, even though limited phenotyping is
505 available for this economically-important trait.
506

507 *AUTHOR CONTRIBUTIONS*

508 Conceived and designed the experiments: MS, TL, BK. Performed the experiments: RA, MS,
509 TL, BK. Analyzed the data: RA, AA, DL, GG, YP, BK, MS. Wrote the paper: MS. All authors
510 reviewed and approved the publication.
511

512 *CONFLICT OF INTEREST STATEMENT*

513 The authors declare that they have no competing interests.
514

515 *ACKNOWLEDGMENTS*

516 This study was supported by a competitive grant No. 2014-67015-21602 from the United States
517 Department of Agriculture, National Institute of Food and Agriculture (MS). R.A.T trainee's
518 projects are supported by Grant Number T32HL072757 from the National Heart, Lung, and Blood
519 Institute. The content is solely the responsibility of the authors and does not necessarily represent
520 the official views of any of the funding agents.
521

522 *REFERENCES*

523 [1] WorldFish Center, Climate Change: Research to Meet the Challenges Facing Fisheries and
524 Aquaculture. In: Issues Brief 1915, 2009.
525 [2] B. Gjerde, State of the art in selective breeding of aquaculture species. In: BARD Workshop:
526 Aquaculture Genetics - status and prospects, Eilat, Israel, 2006.
527 [3] M. Salem, P.B. Kenney, C.E. Rexroad, and J. Yao, Molecular characterization of muscle atrophy and
528 proteolysis associated with spawning in rainbow trout. Comp Biochem Physiol Part D Genomics
529 Proteomics 1 (2006) 227-37.
530 [4] D.E. Gerrard, and A.L. Grant, Principles of Animal Growth and Development, Kendall/Hunt Publishing
531 Co. , 2003.
532 [5] J.F. Taylor, K.H. Taylor, and J.E. Decker, Holsteins are the genomic selection poster cows. Proc Natl
533 Acad Sci U S A 113 (2016) 7690-2.
534 [6] R.L. Vallejo, T.D. Leeds, G. Gao, J.E. Parsons, K.E. Martin, J.P. Evenhuis, B.O. Fragomeni, G.D. Wiens,
535 and Y. Palti, Genomic selection models double the accuracy of predicted breeding values for
536 bacterial cold water disease resistance compared to a traditional pedigree-based model in
537 rainbow trout aquaculture. Genet Sel Evol 49 (2017) 17.
538 [7] R.D. Houston, J.W. Davey, S.C. Bishop, N.R. Lowe, J.C. Mota-Velasco, A. Hamilton, D.R. Guy, A.E.
539 Tinch, M.L. Thomson, M.L. Blaxter, K. Gharbi, J.E. Bron, and J.B. Taggart, Characterisation of QTL-
540 linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic
541 salmon. BMC Genomics 13 (2012) 244.
542 [8] C.E. Rexroad, Y. Palti, S.A. Gahr, and R.L. Vallejo, A second generation genetic map for rainbow trout
543 (Oncorhynchus mykiss). BMC Genet 9 (2008) 74-74.

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

544 [9] S. Wang, Z. Sha, T.S. Sonstegard, H. Liu, P. Xu, B. Somridhivej, E. Peatman, H. Kucuktas, and Z. Liu,
545 Quality assessment parameters for EST-derived SNPs from catfish. BMC Genomics 9 (2008) 450.
546 [10] D. Gonzalez-Pena, G. Gao, M. Baranski, T. Moen, B.M. Cleveland, P.B. Kenney, R.L. Vallejo, Y. Palti,
547 and T.D. Leeds, Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield,
548 Carcass, and Body Weight Traits in Rainbow Trout (Oncorhynchus mykiss). Front Genet 7 (2016)
549 203.
550 [11] A.J. Brookes, Single Nucleotide Polymorphism (SNP). in: L.w.e.n. John Wiley & Sons,
551 http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0005006.pub2/pdf, accessed April
552 4, 2012, (Ed.), ENCYCLOPEDIA OF LIFE SCIENCES (els), 2007.
553 [12] M. Salem, R.L. Vallejo, T.D. Leeds, Y. Palti, S. Liu, A. Sabbagh, C.E. Rexroad, 3rd, and J. Yao, RNA-Seq
554 identifies SNP markers for growth traits in rainbow trout. PLoS One 7 (2012) e36264.
555 [13] R. Al-Tobasei, A. Ali, T.D. Leeds, S. Liu, Y. Palti, B. Kenney, and M. Salem, Identification of SNPs
556 associated with muscle yield and quality traits using allelic-imbalance analyses of pooled RNA-
557 Seq samples in rainbow trout. BMC Genomics 18 (2017) 582.
558 [14] Al Tobasei R, Palti Y, Wiens GD, and S. M., Identification of SNPs with allelic imbalances in rainbow
559 trout genetic lines showing different susceptibility to infection with Flavobacterium
560 psychrophilum. , PAG-XXV Plant & Animal Genomes Conference, San Diego, California., 2017; .
561 [15] D. Marancik, G. Gao, B. Paneru, H. Ma, A.G. Hernandez, M. Salem, J. Yao, Y. Palti, and G.D. Wiens,
562 Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow
563 trout following experimental challenge with Flavobacterium psychrophilum. Front Genet 5
564 (2014) 453.
565 [16] Y. Palti, G. Gao, S. Liu, M.P. Kent, S. Lien, M.R. Miller, C.E. Rexroad, 3rd, and T. Moen, The
566 development and characterization of a 57K single nucleotide polymorphism array for rainbow
567 trout. Mol Ecol Resour 15 (2015) 662-72.
568 [17] G. Gao, T. Nome, D.E. Pearse, T. Moen, K.A. Naish, G.H. Thorgaard, S. Lien, and Y. Palti, A New Single
569 Nucleotide Polymorphism Database for Rainbow Trout Generated Through Whole Genome
570 Resequencing. Front Genet 9 (2018) 147.
571 [18] P. Cingolani, A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, and D.M. Ruden, A
572 program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff:
573 SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6 (2012)
574 80-92.
575 [19] A.R. Quinlan, and I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features.
576 Bioinformatics 26 (2010) 841-2.
577 [20] T.D. Leeds, R.L. Vallejo, G.M. Weber, D.G. Pena, and J.S. Silverstein, Response to five generations of
578 selection for growth performance traits in rainbow trout (Oncorhynchus mykiss). Aquaculture
579 465 (2016) 341-351.
580 [21] K.G. Boldman, L.A. Kriese, L.D. Van Vleck, C.P. Van Tassell, and S.D. Kachman, A manual for the use
581 of MTDFREML. A set of programs to obtain estimates of variances and covariances [Draft], Clay
582 Center, NE: USDA-ARS., 1995.
583 [22] P. Berg, J. Nielsen, and M.K. Sørensen, EVA: Realized and predicted optimal genetic contributions,
584 Proc. 8th World Cong. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil, 2006, pp. S. 246.
585 [23] M. Salem, M.L. Manor, A. Aussanasuwannakul, P.B. Kenney, G.M. Weber, and J. Yao, Effect of
586 sexual maturation on muscle gene expression of rainbow trout: RNA-Seq approach. Physiol Rep
587 1 (2013) e00120.
588 [24] H. Wang, I. Misztal, I. Aguilar, A. Legarra, R.L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken,
589 and W.M. Muir, Genome-wide association mapping including phenotypes from relatives without
590 genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front Genet 5
591 (2014) 134.

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

592 [25] H. Wang, I. Misztal, I. Aguilar, A. Legarra, and W.M. Muir, Genome-wide association mapping
593 including phenotypes from relatives without genotypes. Genet Res (Camb) 94 (2012) 73-83.
594 [26] I. Misztal, S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica, Manual for
595 BLUPF90 family of programs, University of Georgia, University of Georgia, Athens, USA, 2014.
596 [27] Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., and L.D. H., BLUPF90 and related programs
597 (BGF90) [WWW Document], in Proceeding of 7th World Congress on Genetics Applied to
598 Livestock Production (Montpellier), 2002.
599 [28] S.D. Turner, qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots.
600 bioRxiv (2014).
601 [29] A. Ekstrom, E. Sandblom, P.U. Blier, B.A. Dupont Cyr, J. Brijs, and N. Pichaud, Thermal sensitivity and
602 phenotypic plasticity of cardiac mitochondrial metabolism in European perch, Perca fluviatilis. J
603 Exp Biol 220 (2017) 386-396.
604 [30] C. Berthelot, F. Brunet, D. Chalopin, A. Juanchich, M. Bernard, B. Noel, P. Bento, C. Da Silva, K.
605 Labadie, A. Alberti, J.M. Aury, A. Louis, P. Dehais, P. Bardou, J. Montfort, C. Klopp, C. Cabau, C.
606 Gaspin, G.H. Thorgaard, M. Boussaha, E. Quillet, R. Guyomard, D. Galiana, J. Bobe, J.N. Volff, C.
607 Genet, P. Wincker, O. Jaillon, H. Roest Crollius, and Y. Guiguen, The rainbow trout genome
608 provides novel insights into evolution after whole-genome duplication in vertebrates. Nat
609 Commun 5 (2014) 3657.
610 [31] Y. Palti, G. Gao, M.R. Miller, R.L. Vallejo, P.A. Wheeler, E. Quillet, J. Yao, G.H. Thorgaard, M. Salem,
611 and C.E. Rexroad, A resource of single-nucleotide polymorphisms for rainbow trout generated
612 by restriction-site associated DNA sequencing of doubled haploids. Molecular ecology resources
613 14 (2014) 588-596.
614 [32] R.L. Vallejo, S. Liu, G. Gao, B.O. Fragomeni, A.G. Hernandez, T.D. Leeds, J.E. Parsons, K.E. Martin, J.P.
615 Evenhuis, T.J. Welch, G.D. Wiens, and Y. Palti, Similar Genetic Architecture with Shared and
616 Unique Quantitative Trait Loci for Bacterial Cold Water Disease Resistance in Two Rainbow Trout
617 Breeding Populations. Front Genet 8 (2017) 156.
618 [33] T. Taetzsch, M.J. Tenga, and G. Valdez, Muscle Fibers Secrete FGFBP1 to Slow Degeneration of
619 Neuromuscular Synapses during Aging and Progression of ALS. J Neurosci 37 (2017) 70-82.
620 [34] L.M. Ooms, K.A. Horan, P. Rahman, G. Seaton, R. Gurung, D.S. Kethesparan, and C.A. Mitchell, The
621 role of the inositol polyphosphate 5-phosphatases in cellular function and human disease.
622 Biochem J 419 (2009) 29-49.
623 [35] GeneCards-Human-Gene-Database, PROM1 Gene - GeneCards | PROM1 Protein | PROM1
624 Antibody, 2018.
625 [36] UniProtKB, Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha, 2018.
626 [37] GeneCards-Human-Gene-Database, MCTS1 Gene - GeneCards | MCTS1 Protein | MCTS1 Antibody.
627 (2018).
628 [38] S.S. X, R. J, P.S. Kim, B.G. Novitch, and A.B. Lassar, Cyclin-mediated inhibition of muscle gene
629 expression via a mechanism that is independent of pRB hyperphosphorylation. (1996).
630 [39] K.H. Sun, K.H. Chang, S. Clawson, S. Ghosh, H. Mirzaei, F. Regnier, and K. Shah, Glutathione-S-
631 transferase P1 is a critical regulator of Cdk5 kinase activity. Journal of neurochemistry 118
632 (2011) 902-14.
633 [40] R.G. Haller, and S. DiMauro, Chapter 75 - Metabolic and Mitochondrial Myopathies. in: J.A. Hill, and
634 E.N. Olson, (Eds.), Muscle, Academic Press, Boston/Waltham, 2012, pp. 1031-1041.
635 [41] V. Giorgio, M.E. Soriano, E. Basso, E. Bisetto, G. Lippe, M.A. Forte, and P. Bernardi, Cyclophilin D in
636 mitochondrial pathophysiology. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1797 (2010)
637 1113-1118.
638 [42] S. Gingras, S. Pelletier, K. Boyd, and J.N. Ihle, Characterization of a family of novel cysteine- serine-
639 rich nuclear proteins (CSRNP). PLoS One 2 (2007) e808.

bioRxiv preprint doi: https://doi.org/10.1101/355792; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

640 [43] UniProtKB, Slc26a9 - Solute carrier family 26 member 9 - Mus musculus (Mouse) - Slc26a9 gene &
641 protein, 2018.
642 [44] L.A. Alfaro, S.A. Dick, A.L. Siegel, A.S. Anonuevo, K.M. McNagny, L.A. Megeney, D.D. Cornelison, and
643 F.M. Rossi, CD34 promotes satellite cell motility and entry into proliferation to facilitate efficient
644 skeletal muscle regeneration. Stem cells (Dayton, Ohio) 29 (2011) 2030-41.
645 [45] C. Zhang, C. Wang, Y. Li, T. Miwa, C. Liu, W. Cui, W.C. Song, and J. Du, Complement C3a signaling
646 facilitates skeletal muscle regeneration by regulating monocyte function and trafficking. Nat
647 Commun 8 (2017) 2078.
648 [46] UniProtKB, PTPN12 - Tyrosine-protein phosphatase non-receptor type 12 - Homo sapiens (Human) -
649 PTPN12 gene & protein, 2018.
650 [47] UniProtKB, MAGI3 - Membrane-associated guanylate kinase, WW and PDZ domain-containing
651 protein 3 - Homo sapiens (Human) - MAGI3 gene & protein, 2018.
652 [48] Z. Xu, N. Ichikawa, K. Kosaki, Y. Yamada, T. Sasaki, L.Y. Sakai, H. Kurosawa, N. Hattori, and E.
653 Arikawa-Hirasawa, Perlecan deficiency causes muscle hypertrophy, a decrease in myostatin
654 expression, and changes in muscle fiber composition. Matrix Biology 29 (2010) 461-470.
655 [49] M.M. Tajrishi, S. Sato, J. Shin, T.S. Zheng, L.C. Burkly, and A. Kumar, The TWEAK-Fn14 dyad is
656 involved in age-associated pathological changes in skeletal muscle. Biochem Biophys Res
657 Commun 446 (2014) 1219-1224.
658 [50] C. Wang, W. Wang, Y. Liu, M. Yong, Y. Yang, and H. Zhou, Rac GTPase activating protein 1 promotes
659 oncogenic progression of epithelial ovarian cancer. Cancer Sci 109 (2018) 84-93.
660 [51] T.V.A. Murray, J.M. McMahon, B.A. Howley, A. Stanley, T. Ritter, A. Mohr, R. Zwacka, and H.O.
661 Fearnhead, A non-apoptotic role for caspase-9 in muscle differentiation. (2008).
662 [52] V.A. Gupta, and A.H. Beggs, Kelch proteins: emerging roles in skeletal muscle development and
663 diseases. Skeletal muscle 4 (2014) 11.
664 [53] Y. Zhang, P. Xu, C. Lu, Y. Kuang, X. Zhang, D. Cao, C. Li, Y. Chang, N. Hou, H. Li, S. Wang, and X. Sun,
665 Genetic linkage mapping and analysis of muscle fiber-related QTLs in common carp (Cyprinus
666 carpio L.). Mar Biotechnol (NY) 13 (2011) 376-92.
667 [54] H.Y. Tsai, A. Hamilton, D.R. Guy, A.E. Tinch, S.C. Bishop, and R.D. Houston, The genetic architecture
668 of growth and fillet traits in farmed Atlantic salmon (Salmo salar). BMC Genet 16 (2015) 51.
669 [55] K.R. Short, M.L. Bigelow, J. Kahl, R. Singh, J. Coenen-Schimke, S. Raghavakaimal, and K.S. Nair,
670 Decline in skeletal muscle mitochondrial function with aging in humans. Proc Natl Acad Sci U S A
671 102 (2005) 5618-23.

672