1 **Characterising the mechanisms underlying genetic resistance to amoebic gill**

2 **disease in Atlantic salmon using RNA sequencing**

3 Diego Robledo[1*], Alastair Hamilton[2,3], Alejandro P. Gutiérrez[1], James E. Bron[4], Ross

4 D. Houston[1*]

5 [1] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of

6 Edinburgh, EH25 9RG Midlothian, United Kingdom

7 [2] Landcatch Natural Selection Ltd., Roslin Innovation Centre, University of Edinburgh,

8 EH25 9RG Midlothian, United Kingdom

9 [3] Hendrix Genetics Aquaculture BV/ Netherlands, Villa 'de Körver', Spoorstraat 69,

10 5831 CK Boxmeer, NL

11 [4] Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, FK9 4LA

12 Stirling, United Kingdom

13

14 *Corresponding authors:

15 Ross D. Houston - ross.houston@roslin.ed.ac.uk

16 Diego Robledo – diego.robledo@roslin.ed.ac.uk

17 E-mail addresses:

18 Alastair Hamilton – Alastair.Hamilton@hendrix-genetics.com

19 Alejandro P. Gutiérrez – Alejandro.Gutierrez@roslin.ed.ac.uk

20 James E. Bron - bron.james@googlemail.com

21

22 **ABSTRACT**

23 **Background**: Gill health is one of the main concerns for Atlantic salmon aquaculture,

24 and Amoebic Gill Disease (AGD), attributable to infection by the amoeba

25 *Neoparamoeba perurans,* is a frequent cause of morbidity. In the absence of preventive

26 measures, increasing genetic resistance of salmon to AGD via selective breeding can

27 reduce the incidence of the disease and mitigate gill damage. Understanding the

28 mechanisms leading to AGD resistance and the underlying causative genomic features

29 can aid in this effort, while also providing critical information for the development of

30 other control strategies. AGD resistance is considered to be moderately heritable, and

31   several putative QTL have been identified. The aim of the current study was to improve

32   understanding of the mechanisms underlying AGD resistance, and to identify putative

33   causative genomic factors underlying the QTL. To achieve this, RNA was extracted

34   from the gill and head kidney of AGD resistant and susceptible animals following a

35   challenge with *N. perurans*, and sequenced.

36   **Results:** Comparison between resistant and susceptible animals pointed to differences

37   mainly in the local immune response in the gill, involving red blood cell genes and

38   genes related to immune function and cell adhesion. Differentially expressed immune

39   genes highlighted differences in the Th2 and Th17 responses, which are consistent with

40   the increased heritability observed after successive challenges with the amoeba. Five

41   QTL-region candidate genes showed differential expression, including a gene connected

42   to interferon responses (*GVINP1*), a gene involved in systemic inflammation

43   (*MAP4K4*), and a positive regulator of apoptosis (*TRIM39*). Analyses of allele-specific

44   expression highlighted a gene in the QTL region on chromosome 17, cellular repressor

45   of E1A-stimulated genes 1 (*CREG1*), showing allelic differential expression suggestive

46   of a cis-acting regulatory variant.

47   **Conclusions:** In summary, this study provides new insights into the mechanisms of

48   resistance to AGD in Atlantic salmon, and highlights candidate genes for further

49   functional studies that can further elucidate the genomic mechanisms leading to

50   resistance and contribute to enhancing salmon health via improved genomic selection.

51

52   **Keywords: AGD, genomics, amoeba, gene expression, RNA-seq, transcriptome,**

53   **Salmo salar, disease resistance, allelic specific expression**

54

55   **BACKGROUND**

56   Gill health is currently one of the major concerns for Atlantic salmon farming

57   worldwide. Fish gills are multifunctional organs fundamental for gas exchange,

58   ionoregulation, osmoregulation, acid-base balance and ammonia excretion, but also play

59   an important role in hormone production and immune defence (Rombough 2007). Gills

60   are constantly exposed to the marine environment, and are often the first line of defence

61   against pathogens. Gill damage is often observed in Atlantic salmon under farming

62   conditions, and can pose a significant welfare, management and economic burden.
63   While the aetiology of gill disorders is complex, Amoebic Gill Disease (AGD) is
64   currently regarded as a key threat to gill health. This disease adversely affects the gill,
65   and can result in respiratory distress, and ultimately mortalities if left untreated. Initially
66   limited to Tasmania, this disease is currently causing major economic and fish welfare
67   burden to Norwegian, Scottish and Australian salmon aquaculture (Shinn et al. 2015).
68   The causative agent of this disease is the amoeba *Neoparamoeba perurans*, an
69   opportunistic pathogen that can only be removed with expensive and laborious fresh
70   water or hydrogen peroxide treatments (Novak and Archibald, 2018), and there are
71   currently very limited opportunities for prevention.

72   A promising avenue to decrease the incidence of AGD in farmed Atlantic salmon is to
73   increase genetic resistance of aquaculture stocks to *N. perurans*. There is significant
74   genetic variation in resistance to AGD in commercial Atlantic salmon populations
75   (Taylor et al. 2007, 2009; Kube et al. 2012; Robledo et al. 2018; Boison et al. 2019),
76   therefore selective breeding has potential to improve gill health via a reduction in
77   amoebic load and associated gill damage. The use of genetic markers through genomic
78   selection can expedite genetic gain in aquaculture breeding programmes (e.g. Robledo
79   et al. 2018; Yoshida et al. 2018; Palaiokostas et al. 2018a, 2018b), however, the need to
80   genotype a large number of animals and to perform disease challenges in every
81   generation involves a relatively high cost. The discovery of the mechanisms leading to
82   resistance and the underlying causative genetic variants has the potential to reduce this
83   cost via incorporation of functional SNPs into the genomic prediction models.

84   Discovering the genes and pathways that lead to successful immune responses to
85   pathogens is a major goal in genetics and immunology research. Understanding disease
86   resistance can aid selective breeding via incorporation of putative causative variants
87   with greater weighting in genomic prediction models, which can improve selection
88   accuracy and reduce the need for routine trait recording (MacLeod et al. 2016; Houston
89   et al. 2017). Such information can also inform the development of improved disease
90   challenge models, and more successful prevention or treatment strategies through an
91   increased knowledge of host-pathogen interactions. Finally, with the potential role for
92   targeted genome editing (e.g. using CRISPR/Cas9) in future aquaculture breeding
93   programmes, understanding the functional mechanisms underlying disease resistance
94   traits is key to identifying target genes and variants for editing. Previous studies into

95 AGD-infected Atlantic salmon have suggested that the amoebae might elicit an
96 immunosuppressive effect on the innate response of the host, with a concurrent up-
97 regulation of the adaptive Th2-mediated response (Benedicenti et al. 2015; Marcos-
98 López et al. 2017, 2018). Th2 cytokines were also found consistently up-regulated when
99 comparing AGD infected and non-infected samples, and lesion and non-lesion areas
100 (Marcos-López et al. 2018). The heritability of resistance to AGD has been shown to
101 increase after successive cycles of disease challenge / treatment (Taylor et al. 2009),
102 which could suggest that the ability to elicit a successful adaptive immune response is
103 partly under genetic control. Finally, a higher expression of genes related to adaptive
104 immunity has been previously reported in more AGD-resistant salmon compared to
105 their more susceptible counterparts using a microarray approach to measure gene
106 expression (Wynne et al. 2008).

107 In a previous study by our group, several QTL regions with a significant contribution to
108 genetic AGD resistance were identified in Atlantic salmon derived from a commercial
109 breeding programme (Robledo et al. 2018). In the current study, the gill and head
110 kidney transcriptomes of AGD resistant and susceptible Atlantic salmon from the same
111 population were sequenced and compared. The main goals of the study were a) to assess
112 the differences in local and systemic immune responses between AGD resistant and
113 susceptible Atlantic salmon, and b) to use gene expression data to identify positional
114 and functional candidate genes underlying the previously detected resistance QTL.

115

RESULTS

Sampling and sequencing

118 Fish were classified into resistant or susceptible based on their mean gill score and their
119 gill amoebic load. A previous study by our group has shown a high positive genetic
120 correlation between these two traits (higher gill score associated with higher amoebic
121 load), and both are considered indicator traits for resistance to AGD. RNA sequencing
122 was performed on the gill and head kidney of 12 resistant and 12 susceptible fish.
123 Resistant animals had a mean gill score of 2.92 ± 0.13, mean amoebic load (qPCR ct
124 value) of 37.12 ± 3.63 and mean weight of 543 ± 116 g at the point of sampling;
125 susceptible animals had a mean gill score of 4.12 ± 0.20, mean amoebic load of $25.99 \pm$

126   1.80 and mean weight of 409 ± 96 g. Sequencing of one of the gill samples rendered an

127   extremely low number of reads and therefore was discarded. The remaining samples

128   had an average of 24M filtered paired-end reads, which were pseudoaligned to genes in

129   the Atlantic salmon genome. Exploratory analyses based on distance measures revealed

130   two head kidney samples as outliers and they were removed (Additional file 1).

131   Therefore, the final dataset comprised of 23 gill and 22 head kidney samples from 24

132   individuals. The two tissues showed clearly distinct patterns of gene expression, as

133   would be expected. The difference in global gene expression pattern between resistant

134   and susceptible samples in both tissues was much less pronounced, but still evident in

135   the gill in particular (Figure 1). Similar results were described in a Norwegian

136   commercial population (Boison et al. 2019).

137

Differential expression

139   A total of 115 and 42 differentially expressed transcripts (following multiple-testing

140   correction, FDR p-value < 0.05) were detected between resistant and susceptible

141   samples in gill and head kidney tissues respectively (Figure 2, additional file 2). The

142   clearest evidence for differential immune responses was found in gill, where several

143   differentially expressed immune-related transcripts were detected. Most differentially

144   expressed transcripts in head kidney were not obviously related to AGD or disease

145   resistance. To gain an overall view of the results we performed a gene Ontology (GO)

146   enrichment test in both gill and head kidney for sets of differentially expressed

147   transcripts according to three different significance criteria (p-value < 0.01, 0.05 and

148   0.1) (Figure 3). In the gill, various relevant GO terms were observed, such as "Response

149   to stress", "Cytoskeleton" and "Circulatory system process". We observed a larger

150   number of enriched GO terms in head kidney. While most of them are seemingly

151   unconnected to AGD resistance, terms such as "Response to stress" or "Protein

152   modification process" were observed. For instance, of 22 genes showing p-values <

153   0.01, 15 of them were assigned to "Response to stress". Similar analyses for KEGG

154   pathways did not reveal any significant enrichment.

155   Detailed inspection of the differentially expressed transcripts in the gill revealed that

156   they can be grouped into three broad categories concordant with GO enrichment results:

157   1) immune response ("Response to stress"), 2) cell adhesion or cell shape

158　("Cytoskeleton"), and 3) red blood cells and coagulation ("Circulatory system
159　process").

160　Amongst the immune-related transcripts showing differential expression in the gill was
161　interleukin-17 receptor E (*IL17RE*), which was highly expressed in resistant animals
162　(logFC = 1.1). In mice *IL17RE* is the receptor for *IL-17C*, which has an essential role in
163　host mucosal defense against infection and is critical for a successful immune response
164　against bacterial infection (Song et al. 2011). The *IL-17C – IL17RE* pair also stimulates
165　T-helper cell 17 responses, which has a proinflammatory effect (Chang et al. 2011).
166　Benedicenti et al. (2015) reported that *IL-17C* expression showed a negative correlation
167　with amoebic load in Atlantic salmon, and that the Th17 pathway in general was found
168　to be significantly down-regulated in response to AGD. This could be a mechanism of
169　immune evasion elicited by the parasite, which might be more effective in susceptible
170　fish than resistant. Another highly expressed transcript is involved in T-cell function, T-
171　cell specific surface glycoprotein CD28 (*CD28*; logFC = 1.60). *CD28* promotes T-cell
172　survival and proliferation, and enhances the production of multiple cytokines including
173　IL4 (Blotta et al. 1996) *IL4* has been found to be up-regulated in response to AGD
174　(Marcos-López et al. 2018), and this gene induces differentiation of naïve helper T cells
175　to Th2 cells. The Th2 pathway was found to be up-regulated in late stages of AGD
176　(Benedicenti et al. 2015). This pathway is linked to humoral immune responses against
177　extracellular parasites and to tissue repair (Allen and Sutherland, 2014), and therefore is
178　an expected response to AGD. A higher prevalence of this type of response in resistant
179　animals would also be consistent with the observed increase of the heritability of
180　resistance after successive cycles of disease challenge / treatment (Taylor et al. 2009),
181　reflecting genetic variability in the effectiveness of the adaptive response, and / or
182　variation in immune memory.

183　Several genes connected to red blood cells were found to be differentially expressed,
184　including five different haemoglobin subunit transcripts, which were highly expressed
185　and clearly up-regulated in resistant samples in the gill (logFC ~ 2). Reduced hematocrit
186　has been described in AGD infected Atlantic salmon, linked mainly to gill damage
187　(Hvas et al. 2017). The plasma protease C1 inhibitor gene (*SERPING1*) is also up-
188　regulated (logFC = 1.2). This gene inhibits the complement system and also has anti-
189　inflammatory functions (Davis et al. 2008). Complement proteins have been found in
190　gill mucus of AGD infected Atlantic salmon (Valdenegro-Vega et al. 2014). The lower

191   expression of *SERPING1* in susceptible samples might simply be a reflection of the

192   higher extent of gill damage in these animals, requiring activation of the complement

193   system and increase of local inflammatory responses.

194   There are also a few differentially expressed transcripts connected to cell adhesion and

195   cell shape, including a cadherin gene (cadherin-related family member 5; logFC = 4.5)

196   and an actin related gene (actin filament associated protein 1-like 1; logFC = 1.3). The

197   Cdc42 effector protein 2 (*CDC4EP2*; logFC = 0.6) was also up-regulated in resistant

198   fish, and has been associated with roles in actin filament assembly and control of cell

199   shape (Hirsch et al. 2001). A previous study identified an enrichment of cell-adhesion

200   genes in severely affected animals compared to others with healthier gills infected by

201   AGD (Boison et al. 2019). These changes are consistent with the epithelial hyperplasia

202   and other structural changes caused by the parasite in the gill of infected animals

203   (Nowak 2012).

204   Head kidney differential expression results are much harder to interpret, and most of the

205   DE transcripts are seemingly not connected to biological processes that have previously

206   been related to AGD. Tumor necrosis factor alpha-induced protein 8-like protein 1

207   (*TNFAIP8L1*; logFC = -0.9) was found to be more highly expressed in susceptible

208   samples. This gene inhibits apoptosis by suppressing the activity of caspase-8 (You et

209   al. 2001). The down-regulation of pro-apoptotic genes has been connected to AGD

210   severity (Marcos-López et al. 2018). The lack of a clear picture in head kidney might

211   reflect the relative importance of the local and systemic immune responses in response

212   to AGD. Two transcripts showed differential expression in both tissues: putative ferric-

213   chelate reductase 1 (*FRRS1*) and CG057 protein.

214   The regulation of transcripts upon infection is a strong indication of the involvement of

215   the gene product in the immune and physiological response of the host to the pathogen,

216   but a comparison between resistant and susceptible animals can offer insight into the

217   mechanisms determining the success of the immune response against the pathogen. The

218   main caveat of this approach is that it is difficult to distinguish cause and consequence,

219   i.e. is the gene differentially expressed because it confers resistance or due to

220   differential disease progression? Additional evidence, such as the co-localization of

221   differentially expressed genes with QTL or the identification of cis regulatory variants

222   in the QTL regions can further contribute to our understanding of disease resistance, and

223   help us discover underlying candidate genes.

224

225

226

227   **Integration with previous QTL**

228   The overlap between previously identified QTL regions in this population (Robledo et

229   al. 2018) and the differentially expressed genes was explored (Figure 4). A differentially

230   expressed gene, interferon-induced very large GTPase 1 (*GVINP1*), was found in one of

231   the QTL regions of chromosome 18, which explained ~20% of the genetic variance in

232   resistance to AGD (second largest QTL). Very little is known about the function of this

233   gene, but it has been shown to respond to both type I and type II interferon response in

234   mammals (Klamp et al. 2003). The genes showing FDR corrected p-values < 0.1 (a total

235   of 268 genes) were also investigated, and four additional genes were found in these

236   QTL regions. *MAP4K4*, located in a putative QTL region of chromosome 17, surpassed

237   this threshold, and is involved in systemic inflammation in mammals (Aouadi et al.

238   2009), and *TRIM39* in the second QTL region in chromosome 18, a positive regulator of

239   apoptosis (Rosenthal 2012).

240

241   **Allele specific expression**

242   To explore potential cis-acting variation underlying the resistance QTL, an allele

243   specific expression (ASE) test was performed for the SNPs in transcripts within the

244   QTL regions, finding a significant ASE event in a gene in chromosome 17; cellular

245   repressor of E1A-stimulated genes 1 (*CREG1*). In humans this protein is connected to

246   the regulation of cellular proliferation and differentiation (Di Bacco and Gill, 2003), and

247   antagonizes the proliferative effects of adenovirus E1A protein (Veal et al. 1998). This

248   gene showed a log fold change of 0.75 between resistant and susceptible samples in gill

249   (FDR p-value = 0.25). A second significant ASE event was found in an uncharacterised

250   gene in chromosome 18 (LOC106576659), however this gene showed no differences in

251   fold change between resistant and susceptible samples.

252   The polygenic nature of resistance to AGD means that different resistance mechanisms
253   might be operating in each different family. The connection between genotypes and
254   expression, through expression QTL (eQTL) or ASE, can provide strong evidence for
255   functional candidate genes underlying QTL. While eQTL studies require a relatively
256   large number of animals, the advantage of ASE is that the statistical test in performed
257   separately in each heterozygous individual. It is well known that most causative variants
258   are part of regulatory elements and affect gene expression (Keane et al. 2011; Albert
259   and Kruglyak 2015), therefore the detection of ASE in a QTL can provide strong
260   evidence linking the function of a gene to the QTL and the phenotype of interest.

261

262   **DISCUSSION**

263   The potential benefits of the identification of causative variation impacting on complex
264   traits are substantial, ranging from fundamental knowledge of the biology underlying
265   the traits of interest to their application for enhancing these traits in farmed populations.
266   However, even with the addition of various layers of information such as RNA
267   sequencing, determining the causative gene underlying a QTL is not straightforward,
268   especially because the QTL regions tend to be large and contain a large number of
269   genes, as previously described for sea lice resistance QTL (Robledo et al. 2019).

270   Eventually, functional assays are necessary to provide actual evidence of its causality.
271   The advent of CRISP-CAS9 has made this much more feasible in non-model species.
272   Likewise, this technology now provides the opportunity of using this information to
273   introduce or fix favorable alleles in farmed populations. The genetic architecture of
274   quantitative traits usually varies across populations, and indeed AGD resistance QTL
275   seem to vary across different Atlantic salmon commercial populations (Robledo et al.
276   2018, Boison et al. 2019). While the use of genome editing in farmed animals requires
277   societal and regulatory changes, the transference of causative variants across
278   populations can lead to a rapid increase in disease resistance (Jenko et al. 2015), with
279   long-lasting effects on animal welfare and food security. Nevertheless, the discovery of
280   causative variants and genes can be used to increase the weight of causative variants in
281   genomic selection, increasing its accuracy and therefore speeding up genetic gain in
282   each generation (MacLeod et al. 2016). More widely, basic knowledge about the
283   pathways leading to resistance to disease can inform drug development or preventive

284   measures such as functional feeds. To summarise, finding the underlying cause(s) of
285   resistance to disease can provide large benefits for aquaculture and society in the form
286   of healthier animals, increased food security and sustainable economic gain, directly
287   through their implementation in breeding schemes in the present and through genome
288   editing in the future.

289

290   **CONCLUSIONS**

291   The transcriptomic differences between AGD resistance and AGD susceptible Atlantic
292   salmon are limited, which might not be surprising considering the polygenic nature of
293   the trait. The differences were more evident in the gill than in head kidney, potentially
294   highlighting the importance of the local immune response. Genes involved in immune
295   response (Th2 and Th17 pathways), red blood cells and cell adhesion could be part of
296   the mechanisms leading to AGD resistance, however, it is difficult to discriminate cause
297   and consequence. The integration of previously discovered QTL and expression data
298   pointed to potential candidate genes of interest, such as *GVINP1*, *MAP4K4* or *TRIM39*.
299   An additional candidate gene, *CREG1*, showed allele specific expression in one of the
300   QTL regions. Follow-up studies to investigate the functional role of these genes in the
301   response to AGD could help us understand the molecular mechanism of resistance to
302   this parasite, and contribute to improving gill health in farmed populations through
303   incorporation of functional data to improve genomic prediction, or potentially via
304   genome editing in the longer term.

305

306   **METHODS**

307   **Experimental design**

308   The AGD challenge experiment was performed using 797 Atlantic post-smolt salmon
309   from 132 nuclear families (~18 months, mean weight after challenge ~464 g)
310   originating from a commercial breeding programme (Landcatch, UK). The challenge
311   experiment was performed as described in Robledo et al. (2018). In brief, seeder fish
312   with a uniform level of AGD infection were produced by cohabitation with fish infected
313   from an *in vivo* culture. The challenge was then performed by cohabitation of infected

314   seeder fish at a ratio of 15% seeder to naïve fish, allowing three separate cycles of

315   infection with a treatment and recovery period after the first two (Taylor *et al.* 2009)

316   using a 4 m^3 seawater tank in the experimental facilities of University of Stirling's

317   Marine Environmental Research Laboratory, Machrihanish (Scotland, UK). The fish

318   were kept under a 16-h light and 8-h dark photoperiod, starting at 05:00; the fish were

319   fed Biomar organic salmon feed, automatic every 20 minutes to approximately 1%

320   biomass; water supply was ambient flow-through filtered to approximately 90 microns,

321   for the duration of the trial, water temperature was between 13 & 14 ºC and salinity was

322   33-35ppt. For the first two challenges, fresh water treatment was performed 21 days

323   after challenge, followed by a week of recovery. The fish were checked visually four

324   times daily during this period. The disease was allowed to progress until the terminal

325   sampling point in the third challenge, when fish were terminated by an overdose of

326   anaesthetic followed by destruction of the brain. Fish were sampled and phenotypes

327   were recorded during three consecutive days. A subjective gill lesion score of the order

328   of severity ranging from 0 to 5 was recorded for both gills (Taylor *et al.* 2016). These

329   gill lesion scores were recorded by a single operator, who referred to pictures to guide

330   classification. Some fish were scored by additional operators, and the scores never

331   differed by > 0.5. Further, one of the gills was stored in ethanol for qPCR analysis of

332   amoebic load using *N. perurans* specific primers. Amoebic load has previously been

333   used as a suitable indicator trait for resistance to AGD in salmon (Taylor *et al.* 2009).

334   All fish were phenotyped for mean gill score (mean of the left gill and right gill scores)

335   and amoebic load (qPCR values using *N. perurans* specific primers, amplified from one

336   of the gills). 24 fish from 24 different families were selected for RNA sequencing

337   (Additional file 3) based on high or low levels of resistance according to the measured

338   traits (mean gill score and amoebic load as measured by qPCR). Gill and head kidney

339   samples were obtained from each animal and stored in RNAlater at 4 ºC for 24 h, and

340   then at -20ºC until RNA extraction.

341   **RNA extraction and sequencing**

342   For all the 48 samples a standard TRI Reagent RNA extraction protocol was followed.

343   Briefly, approximately 50 mg of tissue was homogenized in 1 ml of TRI Reagent

344   (Sigma, St. Louis, MO) by shaking using 1.4 mm silica beads, then 100 µl of 1-bromo-

345   3-chloropropane (BCP) was added for phase separation. This was followed by

346   precipitation with 500 µl of isopropanol and subsequent washes with 65-75 % ethanol.

347   The RNA was then resuspended in RNAse-free water and treated with Turbo DNAse

348   (Ambion). Samples were cleaned up using Qiagen RNeasy Mini kit columns and their

349   integrity was checked on Agilent 2200 Bioanalyzer (Agilent Technologies, USA).

350   Thereafter, the Illumina Truseq mRNA stranded RNA-Seq Library Prep Kit protocol

351   was followed directly. Libraries were checked for quality and quantified using the

352   Bioanalyzer 2100 (Agilent), before being sequenced on three lanes of the Illumina

353   Hiseq 4000 instrument using 75 base paired-end sequencing at Edinburgh Genomics,

354   UK. Raw reads have been deposited in NCBI's Sequence Read Archive (SRA) under

355   BioProject accession number PRJNA552604.

356   **Read mapping**

357   The quality of the sequencing output was assessed using FastQC v.0.11.5

358   (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality filtering and

359   removal of residual adaptor sequences was conducted on read pairs using Trimmomatic

360   v.0.38 (Bolger et al. 2004). Specifically, Illumina specific adaptors were clipped from

361   the reads, leading and trailing bases with a Phred score less than 20 were removed and

362   the read trimmed if the sliding window average Phred score over four bases was less

363   than 20. Only reads where both pairs were longer than 36 bp post-filtering were

364   retained. Filtered reads were mapped to the most recent Atlantic salmon genome

365   assembly (ICSASG_v2; Genbank accession GCF_000233375.1; Lien et al. 2016) using

366   STAR v.2.6.1a (Dobin et al. 2013), the maximum number of mismatches for each read

367   pair was set to 10 % of trimmed read length, and minimum and maximum intron lengths

368   were set to 20 bases and 1 Mb respectively. Transcript abundance was quantified using

369   kallisto v0.44.0 (Bray et al. 2016) and the latest Atlantic salmon genome annotation

370   (NCBI *Salmo salar* Annotation Release 100).

371   **Differential Expression**

372   Differential expression (DE) analyses were performed using R v.3.5.2 (R Core Team

373   2014). Gene count data were used to estimate differential gene expression using the

374   Bioconductor package DESeq2 v.3.4 (Love et al. 2014). Briefly, size factors were

375   calculated for each sample using the 'median of ratios' method and count data was

376   normalized to account for differences in library depth. Next, gene-wise dispersion

377   estimates were fitted to the mean intensity using a parametric model and reduced

378   towards the expected dispersion values. Finally a negative binomial model was fitted for

379 each gene and the significance of the coefficients was assessed using the Wald test. The

380 Benjamini-Hochberg false discovery rate (FDR) multiple test correction was applied,

381 and transcripts with FDR < 0.05 and absolute $\log_2$ fold change values (FC) > 0.5 were

382 considered differentially expressed genes. Hierarchical clustering and principal

383 component analyses were performed to visually identify outlier samples, which were

384 then removed from the analyses. PCA plots were created using the R package factoextra

385 (http://www.sthda.com/english/rpkgs/factoextra/).

386

387 Gene Ontology (GO) enrichment analyses were performed in R v.3.5.2 (R Core Team

388 2014) using Bioconductor packages GOstats v.2.48.0 (Falcon and Gentleman, 2007)

389 and GSEABase v.1.44.0 (Morgan et al. 2018). GO term annotation for the Atlantic

390 salmon transcriptome was obtained using the R package Ssa.RefSeq.db v1.3

391 (https://gitlab.com/cigene/R/Ssa.RefSeq.db). The over-representation of GO terms in

392 differentially expressed gene lists compared to the corresponding transcriptomes (gill or

393 head kidney) was assessed with a hypergeometric test. GO terms with ≥ 5 DE genes

394 assigned and showing a p-value < 0.05 were considered enriched. Kyoto Encyclopedia

395 of Genes and Genomes (KEGG) enrichment analyses were performed using KOBAS

396 v3.0.3 (Wu et al. 2006). Briefly, salmon genes were annotated against KEGG protein

397 database (Kanehisa and Goto 2000) to determine KEGG Orthology (KO). KEGG

398 enrichment for differentially expressed gene lists was tested by comparison to the whole

399 set of expressed genes in the corresponding tissue using Fisher's Exact Test. KEGG

400 pathways with ≥ 5 DE genes assigned and showing a Benjamini-Hochberg FDR

401 corrected p-value < 0.05 were considered enriched for differential expression. The

402 reference tissue transcriptome for both GO and KEGG enrichment comprised only those

403 genes with mean normalized counts value > 5.

Allele specific expression

404

405 Gene expression estimates and genotypes obtained from the RNA sequencing were used

406 to investigate allele specific expression. The samtools v1.6 software (Li et al. 2009) was

407 used to identify SNPs, and call genotypes for those SNPs in individual samples. PCR

408 duplicates, reads with mapping quality < 20 and bases with phred quality scores < 20

409 were excluded. SNPs within 5 bp of an indel, with quality < 20, MAF < 0.05 or less

410 than 4 reads supporting the alternative allele were discarded. The putative effect of the

411 SNPs was assessed using the official salmon genome annotation (NCBI *Salmo salar*

412   Annotation Release 100) and the SnpEff v.4.2 software (Cingolani et al. 2012). Allelic

413   specific expression was assessed using the R package AllelicImbalance (Gadin et al.

414   2015). For every SNP in the regions of interest, read counts were obtained for each

415   allele in heterozygous animals, those with less than 10 reads were filtered, and a

416   binomial test was performed to assess the significance of the allelic differences. Only

417   those genomic positions called as heterozygotes in a minimum of 4 and a maximum of

418   36 (75%) samples (75%) were considered. An allele specific expression event was

419   considered significant if the mean p-value of all heterozygotes was < 0.05. All

420   significant events were manually inspected.

421

422

423

424   **DECLARATIONS**

425   **Ethics approval and consent to participate**

426   All animals were reared in accordance with relevant national and EU legislation

427   concerning health and welfare. The challenge experiment was performed by the Marine

428   Environmental Research Laboratory (Machrihanish, UK) under approval of the ethics

429   review committee of the University of Stirling (Stirling, UK) and according to Home

430   Office license requirements. Landcatch are accredited participants in the RSPCA

431   Freedom Foods standard, the Scottish Salmon Producers Organization Code of Good

432   Practice, and the EU Code-EFABAR Code of Good Practice for Farm Animal Breeding

433   and Reproduction Organizations.

434

435   **Consent for publication**

436   Not applicable

437

438   **Availability of data and materials**

439   The RNA sequencing dataset generated during the current study is available in NCBI's

440   Sequence Read Archive (SRA) under BioProject accession number PRJNA552604. The

441   phenotypes of all sequenced samples are included in this published article (Additional

442   file 3).

443

Competing interests

445   A commerical organisation (Landcatch Natural Selection Ltd) was involved in the

446   development of this study. AH works for Landcatch Natural Selection Ltd. The

447   remaining authors declare that they have no competing interests.

448

Funding

460

Author's contributions

462   RH and DR were responsible for the concept and design of this work and drafted the

463   manuscript. AH was responsible for the disease challenge. AH and JB managed the

464   collection of the samples. AG performed the molecular biology experiments. DR

465   performed bioinformatic and statistical analyses. All authors read and approved the final

466   manuscript**.**

467

Acknowledgements

470

REFERENCES

Albert FW, Kruglyak L. The role of regulatory variation in complex traits. Nat Rev Genet. 2015; 16:197-212.

Allen JE, Sutherland TE. Host protective roles of type 2 immunity: parasite killing and tissue repair, flip sides of the same coin. Semin Immunol. 2014; 26:329-40.

Aouadi M, Tesz GJ, Nicoloro SM, Wang M, Chouinard M, Soto E, Ostroff GR, Czech MP. Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation. Nature. 2009; 458:1180-4.

Benedicenti O, Collins C, Wang T, McCarthy U, Secombes CJ. Which Th pathway is involved during late stage amoebic gill disease? Fish Shellfish Immunol. 2015; 46:417-25.

Blotta MH, Marshall JD, DeKruyff RH, Umetsu DT. Cross-linking of the CD40 ligand on human CD4+ T lymphocytes generates a costimulatory signal that up-regulates IL-4 synthesis. J Immunol. 1996; 156:3133-40.

Boison SA, Gjerde B, Hillestad B, Makvandi-Nejad S, Moghadam HK. Genomic and transcriptomic analysis of amoebic gill disease resistance in Atlantic salmon (*Salmo salar* L.). Front Genet. 2019; 10:68.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2004; 30:2114-20.

Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016; 34:525-27.

Chang SH, Reynolds JM, Pappu BP, Chen G, Martinez GJ, Dong C. Interleukin-17C promotes Th17 cell responses and autoimmune disease via interleukin-17 receptor E. Immunity. 2011; 35:611-21.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. 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). 2012; 6:80-92.

499 Di Bacco A, Gill G. The secreted glycoprotein CREG inhibits cell growth dependent on
500 the mannose-6-phosphate/insulin-like growth factor II receptor. Oncogene. 2003;
501 22:5436-45.

502 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
503 universal RNA-seq aligner. Bioinformatics. 2013; 29:15-21.

504 Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association.
505 Bioinformatics. 2007; 23:257-8.

506 Gadin JR, van't Hooft FM, Eriksson P, Folkersen L. AllelicImbalance: an
507 R/bioconductor package for detecting, managing, and visualizing allele expression
508 imbalance data from RNA sequencing. BMC Bioinformatics. 2015; 16:194.

509 Houston RD. Future directions in breeding for disease resistance in aquaculture species.
510 R Bras Zootec. 2017; 46:545-51.

511 Hvas M, Karlsbakk E, Mæhle S, Wright DW, Oppedal F. The gill parasite Paramoeba
512 perurans compromisos aerobic scope, swimming capacity and ion balance in Atlantic
513 salmon. Conserv Physiol. 2017; 5:cox066.

514 Jenko J, Gorjanc G, Cleveland MA, Varshney RK, Whitelaw BA, wooliams JA, et al.
515 Potential of promotion of alleles by genome editing to improve quantitative traits in
516 livestock breeding programs. Gen Sel Evol. 2015; 47:55.

517 Keane TM, Goodstadt L, Danecek P, White MA, Wong K, et al. Mouse genomic
518 variation and its effect on phenotypes and gene regulation. Nature. 2011; 477:289-94.

519 Klamp T, Boehm U, Schenk D, Pfeffer K, Howard J. A giant GTPase, very large
520 inducible GTPase-1, is inducible by IFNs. J Immunol. 2003; 171:1255-65.

521 Kube PD, Taylor RS, Elliott NG. Genetic variation in parasite resistance of Atlantic
522 salmon to amoebic gill disease over multiple infections. Aquaculture. 2012; 364-
523 365:165-72.

524 Morgan M, Falcon S, Gentleman R. *GSEABase:* Gene set enrichment data structures
525 and methods. R package version 1.44.0. 2018.

526 Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic salmon

527 genome provides insights into rediploidization. Nature. 2016; 533:200-5.

528 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for

529 RNA-seq data with DESeq2. Genome Biol. 2014; 15:550.

530 MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE,

531 Chamberlain AJ, et al. Exploiting biological priors and sequence variants enhances QTL

532 discovery and genomic prediction of complex traits. BMC Genom. 2016; 17:144.

533 Marcos-López M, Espinosa Ruiz C, Rodger HD, O'Connor I, MacCarthy E, Esteban

534 MA. Local and systemic humoral immune response in farmed Atlantic salmon (*Salmo

535 salar* L.) under a natural amoebic gill disease outbreak. Fish Shellfish Immunol. 2017;

536 66:207-16.

537 Marcos-López M, Calduch-Giner JA, Mirimin L, MacCarthy E, Rodger HD, O'Connor

538 I, et al. Gene expression analysis of Atlantic salmon gills reveals mucin 5 and

539 interleukin 4/13 as key molecules during amoebic gill disease. Sci Rep. 2018; 8:13689.

540 Nowak BF (2012) *Neoparamoeba perurans*. In: Woo PTK, Buchmann K, editors. Fish

541 Parasites, Pathobiology and Protection. CAB International; 2012. P. 1–18.

542 Nowak BF, Archibald JM. Opportunistic but lethal: the mystery of paramoebae. Trends

543 Parasitol. 2018; 34:404-19.

544 Palaiokostas C, Kocour M, Prchal M, Houston RD. Accuracy of genomic evaluations of

545 juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by

546 sequencing. Front Genet. 2018; 9:82.

547 Palaiokostas C, Cariou S, Bestin A, Bruant JS, Haffray P, Morin T, et al. Genome-wide

548 association and genomic prediction of resistance to viral nervous necrosis in European

549 sea bass (*Dicentrarchus labrax*) using RAD sequencing. Genet Sel Evol. 2018; 50:30.

550 Robledo D, Matika O, Hamilton A, Houston RD. Genome-wide association and

551 genomic selection for resistance to amoebic gill disease in Atlantic salmon. G3. 2018;

552 8:1195-203.

553 Robledo D, Gutiérrez AP, Barría A, Lhorente JP, Houston RD, Yáñez JM. Discovery
554 and functional annotation of quantitative trait loci affecting resistance to sea lice in
555 Atlantic salmon. Front Genet. 2019; 10:56.

556 Rombough P. The functional ontogeny of the teleost gill: which comes first, gas or ion
557 exchange? Comp Biochem Physiol A Mol Integr Physiol. 2007; 148:732-42.

558 Rosenthal CK. Trim39 ligase keeps apoptosis going. Nat Cell Biol. 2012; 14:566.

559 Shinn AP, Pratoomyot J, Bron JE, Paladini G, Brooker EE, Brooker AJ. Economic costs
560 of protistan and metazoan parasites to global mariculture. Parasitology. 2015; 142:196-
561 270.

562 Song X, Zhu S, Shi P, Liu Y, Shi Y, Levin SD, et al. IL-17RE is the functional receptor
563 for IL-17C and mediates mucosal immunity to infection with intestinal pathogens. Nat
564 Immunol. 2011; 12:1151-8.

565 Taylor RS, Wynne JW, Kube PD, Elliott NG. Genetic variation of resistance to amoebic
566 gill disease in Atlantic salmon (*Salmo salar*) assessed in a challenge system.
567 Aquaculture. 2007; 272S1: S94–S99.

568 Taylor RS, Kube PD, Muller WJ, Elliott NG. Genetic variation of gross gill pathology
569 and survival of Atlantic salmon (*Salmo salar* L.) during natural amoebic gill disease
570 challenge. Aquaculture. 2009; 294(3-4): 172–9.

571 Taylor RS, Huynh C, Cameron D, Evans B, Cook MT, et al. Gill score guide – amoebic
572 gill disease (AGD) management training document, edited by Evans B. Tassal Pty. Ltd.,
573 Hobart, Tasmania. 2016.

574 Valdenegro-Vega VA, Crosbie P, Bridle A, Leef M, Wilson R, Nowak BF.
575 Differentially expressed proteins in gill and skin mucus of Atlantic salmon (*Salmo
576 salar*) affected by amoebic gill disease. Fish Shellfish Immunol. 2014; 40:69-77.

577 Veal E, Eisenstein M, Tseng ZH, Gill G. A cellular repressor of E1A-stimulated genes
578 that inhibits activation by E2F. Mol Cell Biol. 1998; 18:5023-41.

579 Wynne JW, O'Sullivan MG, Stone G, Cook MT, Nowak BF, Lovell DR, et al.
580 Resistance to amoebic gill disease (AGD) is characterized by the transcriptional

581   dysregulation of immune and cell cycle pathways. Dev Comp Immunol. 2008; 32:1539-

582   60.

583   Yoshida GM, Bangera R, Carvalheiro R, Correa K, Figueroa R, Lhorente JP, et al.

584   Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed

585   rainbow trout. G3. 2018; 8:719-26.

586   You Z, Ouyang H, Lopatin D, Polver PJ, Wang CY. Nuclear factor-kappa B-inducible

587   death effector domain-containing protein suppresses tumor necrosis factor-mediated

588   apoptosis by inhibiting caspase-8 activity. J Biol Chem. 2001; 276:26398-404.

589

590   **FIGURES**

591   **Figure 1. Principal component analysis.** RNA-Seq samples clustered according to

592   their gene expression. The larger symbols represent group means, and ellipses represent

593   95% confidence intervals for the groups.

594

595   **Figure 2. Heatmap of differentially expressed genes between resistant and**

596   **susceptible samples.** Heatmaps of all differentially expressed genes in gill (A) and head

597   kidney (B). Samples and genes were clustered according to gene expression (mean

598   centered and scaled normalized counts).

599

600   **Figure 3. Gene Ontology enrichment for differentially expressed genes.** GO

601   enrichment is shown for all differentially expressed genes in gill and head kidney

602   according to three different significant criteria (FDR p-value < 0.1, 0.05 and 0.01). The

603   height of the bars represents fold enrichment (percentage of genes assigned to the GO

604   term in the set of differentially expressed genes compared to the percentage assigned to

605   that GO term in the transcriptome of that tissue).

606

607   **Figure 4. Differentially expressed genes located in resistance QTL.** The location of

608   the QTL regions in the chromosomes are shown in grey. Genes with significance values

609   < 0.05 are in red, those with significance values < 0.1 are in orange. Positive fold

610   changes correspond to higher expression in resistant samples.

611

612   **Figure 5. Allele specific expression CREG1.** Barplot showing the read counts for each

613   allele for those SNPs in the QTL regions showing allele specific expression. The two

614   SNPs are located in *CREG1* (Chromosome 17 – 24,545,527 bp) and the uncharacterized

615   gene LOC106576659 (Chromosome 18 – 57,163,493 bp).

616

617   **ADDITIONAL FILES**

618   **Additional file 1**

619   Portable Network Graphics (.png)

620   Principal component analysis of all RNA sequenced samples

621   RNA-Seq samples clustered according to their gene expression. Outliers were discarded

622   for further analyses.

623

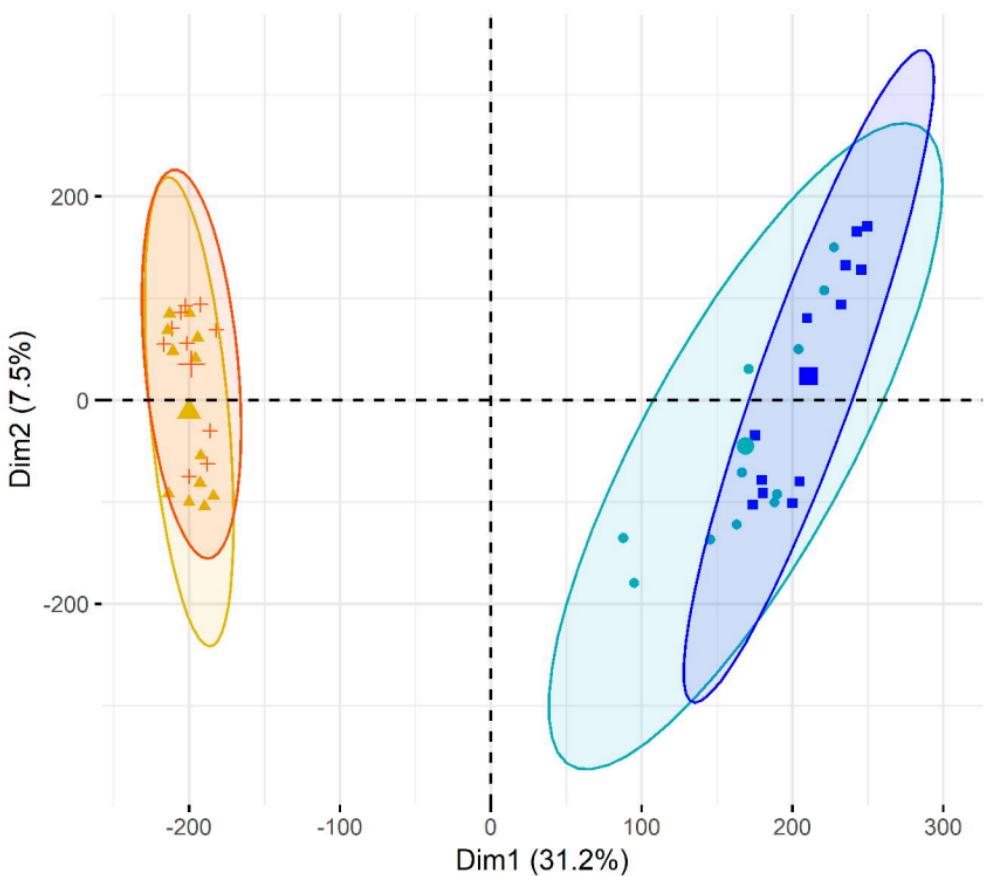624   **Additional file 2**

625   Excel file (.xlsx)

626   Differentially expressed genes between resistant and susceptible samples

627   Lists of differentially expressed genes between resistant and susceptible samples in gill

628   and head kidney. Gene ID, position in the Atlantic salmon genome (Chromosome, start

629   and end in base pairs), average expression of the gene, log 2 fold change between

630   resistant and susceptible animals (positive fold changes correspond to higher expression

631   in resistant samples), standard deviation of the fold change, FDR adjusted p-value, gene

632   annotation and gene symbol are shown.

633

634   **Additional file 3**

635   Excel file (.xlsx)

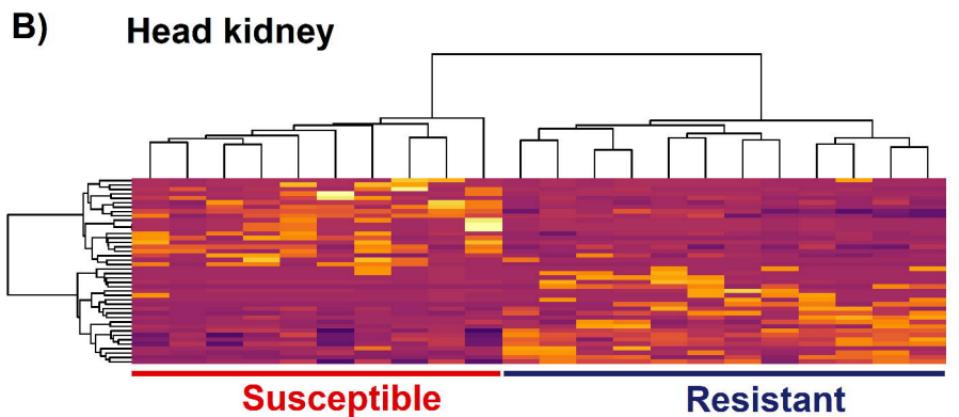636   Phenotypes of all samples used for RNA sequencing

637   All collected phenotypes for the samples used in this study. ID of the sample, tissue,

638   whether it is resistant or susceptible, finclip ID linking to the genotypes (available in

639   Robledo et al. 2018), gill scores for both gills (and mean), weight and length at the end

640   of the challenge, and amoebic load measured by qPCR are shown.

- Head kidney - resistant
- Head kidney - susceptible
- Gill - resistant
- Gill - susceptible

A) Gill

B) Head kidney

Circular bar plot showing enriched GO terms across three tissues.

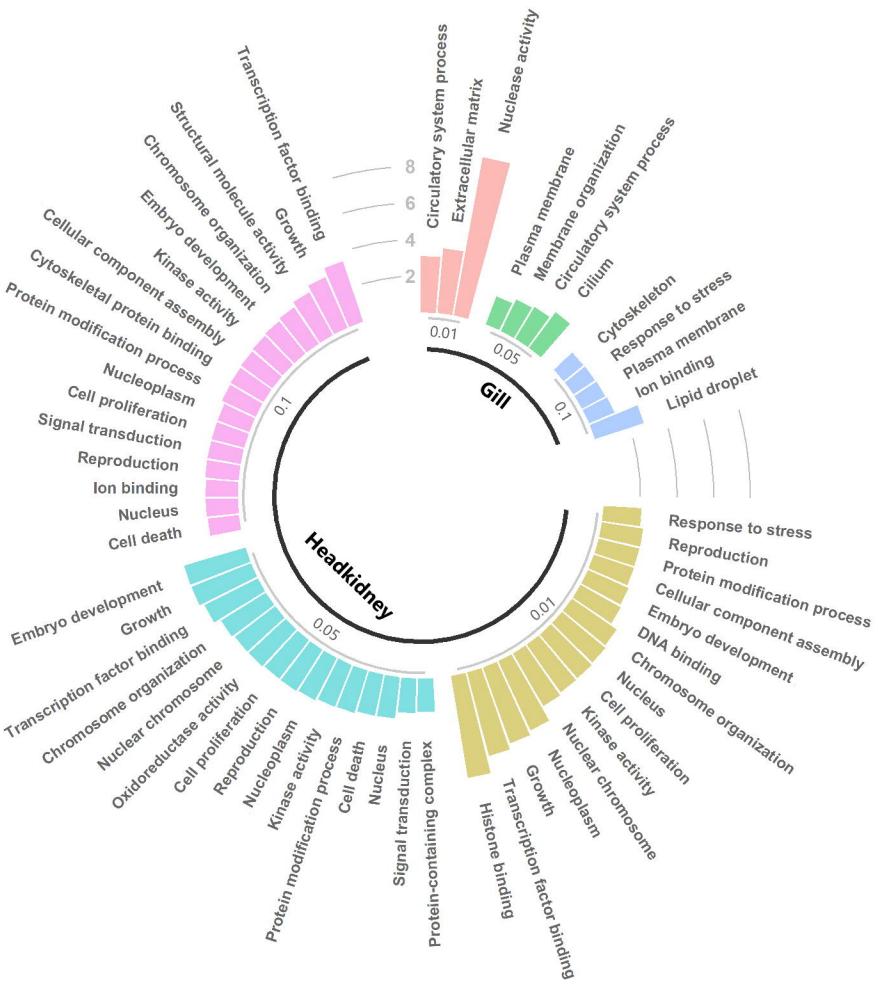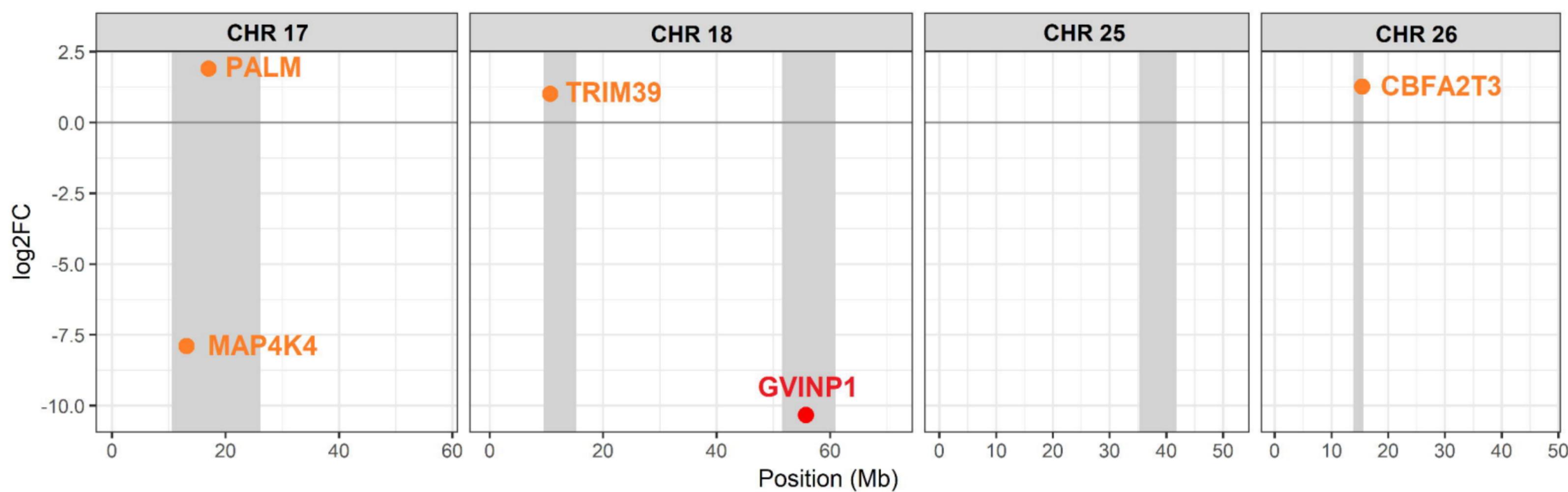**Gill** (red/salmon): Circulatory system process, Extracellular matrix, Nuclease activity, Plasma membrane, Membrane organization, Circulatory system process, Cilium

Gill (green): Cytoskeleton, Response to stress, Plasma membrane, Ion binding, Lipid droplet

Gill (blue): (Cytoskeleton, Response to stress, Plasma membrane, Ion binding, Lipid droplet)

Radial axis (Gill): 2, 4, 6, 8; 0.01, 0.05, 0.1

Headkidney (pink): Transcription factor binding, Structural molecule activity, Growth, Chromosome organization, Embryo development, Kinase activity, Cellular component assembly, Cytoskeletal protein binding, Protein modification process, Nucleoplasm, Cell proliferation, Signal transduction, Reproduction, Ion binding, Nucleus, Cell death

Headkidney (cyan/teal): Embryo development, Growth, Transcription factor binding, Chromosome organization, Nuclear chromosome, Oxidoreductase activity, Cell proliferation, Reproduction, Nucleoplasm, Kinase activity, Protein modification process, Cell death, Nucleus, Signal transduction, Protein-containing complex

Headkidney (olive/yellow): Response to stress, Reproduction, Protein modification process, Cellular component assembly, Embryo development, DNA binding, Chromosome organization, Nucleus, Cell proliferation, Kinase activity, Nuclear chromosome, Nucleoplasm, Growth, Transcription factor binding, Histone binding

Radial axis (Headkidney): 0.1, 0.05, 0.01