Title:

Cis-regulatory differences in isoform expression associate with life history strategy variation in Atlantic salmon

Authors:

Jukka-Pekka Verta [1,2], Paul Vincent Debes [1,2], Nikolai Piavchenko [1], Annukka Ruokolainen [1], Outi Ovaskainen [1], Jacqueline Emmanuel Moustakas-Verho [1,2], Seija Tillanen [1,2], Noora Parre [1], Tutku Aykanat [1, 2], Jaakko Erkinaro [3] & Craig Robert Primmer [1,2]

Affiliations:

1. Organismal and Evolutionary Biology Research Programme
   University of Helsinki
   Viikinkaari 9, 00790, Helsinki, Finland
2. Institute of Biotechnology, University of Helsinki, Finland
3. Natural Resources Institute Finland (LUKE),
   Oulu, Finland

Co-corresponding authors: craig.primmer@helsinki.fi, jukka-pekka.verta@helsinki.fi

Keywords:

life history traits, genotype-phenotype association, transcript isoforms, *cis*-regulation, puberty timing, testes differentiation

Abstract:

A major goal in biology is to understand how evolution shapes variation in individual life histories. Genome-wide association studies have been successful in uncovering genome regions linked with traits underlying life history variation in a range of species. However, lack of functional studies of the discovered genotype-phenotype associations severely restrains our understanding how alternative life history traits evolved and are mediated at the molecular level. Here, we report a *cis*-regulatory mechanism whereby expression of alternative isoforms of the transcription co-factor *vestigial-like 3* (*vgll3*) associate with variation in a key life history trait, age at maturity, in Atlantic salmon (*Salmo salar*). Using a common-garden experiment, we first show that *vgll3* genotype associates with puberty timing in one-year-old salmon males. By way of temporal sampling of *vgll3* expression in ten tissues across the first year of salmon development, we identify a pubertal transition in *vgll3* expression where maturation coincided with a 66% reduction in testicular *vgll3* expression. The *late* maturation allele was not only associated with a tendency to delay puberty, but also with expression of a rare transcript isoform of *vgll3* pre-puberty. By comparing absolute *vgll3* mRNA copies in heterozygotes we show that the expression difference between the *early* and *late* maturity alleles is largely *cis*-regulatory. We propose a model whereby expression of a rare isoform from the *late* allele shifts the liability of its carriers towards delaying puberty. These results reveal how regulatory differences can be a central mechanism for the evolution of life history traits.

Author summary:

Alternative life history strategies are an important source of diversity within populations and promote the maintenance of adaptive capacity and population resilience. However, in many cases the molecular basis of different life history strategies remains elusive. Age at maturity is a key adaptive life history trait in Atlantic salmon and has a relatively simple genetic basis. Using salmon age at maturity as a model, we report a mechanism whereby different transcript isoforms of the key age at maturity gene, *vestigial-like 3* (*vgll3*), associate with variation in the timing of male puberty. Our results show how gene regulatory differences in conjunction with variation in gene transcript structure can encode for complex alternative life histories.

Introduction:

Some of the most ecologically and evolutionarily important traits involve alternative life-history strategies and the trade-offs therein [1]. One such example is maturation age, whereby early maturing individuals avoid the increased risk of mortality associated with later reproduction, but often with the trade-off of lower fecundity [2]. Genome-wide association studies have been successful in uncovering genome regions controlling for life-history trait variation such as survival and reproductive success in Soay sheep [3], early versus late age at maturity in Atlantic salmon [4,5], early versus late run timing in Chinook salmon and steelhead trout [6–8], age and size at flowering in monkeyflowers [9] and alternative reproductive strategies in ruffs [10]. However, despite this progress mapping the genetic basis of life history traits, moving beyond genome-level associations to an understanding of the molecular basis of such variation in life history traits and trade-offs has thus far remained elusive. This lack of functional insight into the discovered genotype-phenotype associations severely restrains our understanding of the molecular basis of evolution in life history traits [11].

In Atlantic salmon, age at maturity is a key adaptive life history trait that is evolving under a trade-off between survival and fecundity. Later-maturing, and therefore larger, individuals have higher reproductive fitness at spawning than earlier maturing (smaller) individuals [12,13], but also have a higher risk of mortality prior to first reproduction [2]. Males mature, on average, earlier than females and at smaller size, which is indicative of a potential sexual conflict for age at maturity in the species [14,15]. Genetic variation in a single genome region including the transcription co-factor gene *vestigial-like 3* (*vgll3*) was recently found to explain nearly 40% of variation in sea-age at maturity in both male and female salmon across 57 Northern-European wild populations [4], making salmon age-at maturity a genetically tractable model to study the molecular bases of life-history trait variation. Further, variation in *VGLL3* also associated with age at puberty in humans [16,17], indicating that functional mechanisms of puberty timing may be shared across vertebrates. However, in all cases the functional basis of the association between genetic variation in *vgll3* and age at maturity remains unknown.

The pattern of association observed in natural salmon populations can be used to generate hypotheses pertaining the functional mechanism that translates genetic variation to different maturation phenotypes. The strongest association between age at maturity and *vgll3* genotype maps to a single nucleotide polymorphism (SNP) in a non-coding region approximately 7000 basepairs 3' of *vgll3*, followed in statistical significance by mutations in *vgll3* coding sequence [4,5]. This suggests that genetic variation in linked *cis*-regulatory sequences (e.g. enhancers) may underlie the functional

differences among salmon *vgll3* genotypes by changing their expression level or pattern. In other species, a similar pattern of linked non-coding changes is often observed in association studies for life history traits [6,7,10,18,19], suggesting that regulatory changes may be a common driver of evolution in life history strategies. However, non-coding and coding hypotheses pertaining the control of maturation age are nuanced and not mutually exclusive; enhancer mutations can have effects unlinked to their genotype (they can act in *trans*) [20,21] and non-coding mutations can alter transcript structure, stability and translation via alternative transcription start sites and splicing [22–24].

To uncover the functional genetic basis of age at maturity variation in Atlantic salmon, we first validated the *vgll3* association with male maturation timing in different genetic backgrounds and in common garden settings by using 32 controlled-crossed families with known *vgll3* genotypes. Uncertainty remains regarding the tissues and developmental stages during which *vgll3* expression potentially influences the maturation phenotype. *Vgll3* expression and function has been linked to diverse processes such as adipocyte differentiation and myogenesis in mice [25,26], immune function, adiposity and puberty timing in humans [16,17]and cell fate commitment in salmon tissues including adipose and gonads [27,28], but genotype comparisons, *vgll3* transcript structure or *cis*-regulation have not been investigated. We determined *vgll3* temporal expression dynamics during the first year of salmon development up to male maturation by way of a time-series sampling of ten tissues from a total of 273 individuals, characterized *vgll3* transcript structure and further tested the importance of *cis*-regulatory mechanisms among *vgll3* genotypes by assaying *vgll3* allele-specific expression in heterozygotes. Our results reveal a potential functional mechanism that links allele-specific isoform expression differences in *vgll3* to variation in maturation timing and thus mediates variation in life history strategies.

119 **Results:**

120

121 ***vgll3* genotype predicts male maturation frequency**

122

123 To uncover the functional genetic basis of age at maturity variation in Atlantic salmon, we first
124 validated that the genotype at the *vgll3* locus on chromosome 25 (termed "*E*" for early sea-age at
125 maturity and "*L*" for late sea-age at maturity after previous association study [4]) is associated with
126 age at maturity in controlled conditions by rearing 656 individuals with known *vgll3* genotypes from
127 32 families in common garden conditions (Fig. S1). We assessed all males for maturation status non-
128 lethally at the start of the breeding season (nine months post-hatching) and modeled maturation
129 probabilities based on observed frequencies of mature versus non-mature fish as a function of *vgll3*
130 genotype (see Methods). *Vgll3*EE* individuals showed a 10 times higher frequency of maturation
131 (20%) compared to individuals with *vgll3*LL* genotypes (2%, Fig. S2). Correspondingly, maturation
132 probability for *vgll3*EE* males (20.7%) was significantly higher compared to males with *vgll3*LL*
133 genotypes (3%) (probit 95% credible interval: 0.59 to 3.07, P=0.006, *vgll3*EE* n=93 and *vgll3*LL*
134 n=92) (Fig. 1A). Maturation probability for males heterozygous for *vgll3* (15.4%) was closer to
135 *vgll3*EE* males compared to *vgll3*LL* males, and significantly different from *vgll3*LL* males (probit
136 95% credible interval: 0.32 to 3.88, P=0.018, *vgll3*EL* n=199), but not *vgll3*EE* males.

137

138 We additionally modeled maturation probability as a function of *vgll3* genotype and fish length,
139 hypothesizing that early growth positively affects maturation probability [29]. Indeed, maturation
140 probability positively associated with fish length before the start of the breeding season (eight months
141 post-hatching) (probit 95% credible interval: 0.36 to 0.94, P=1e-4, Fig. 1B). Irrespective of size
142 effect, the *vgll3*EE* genotype was still associated with a higher maturation probability compared to
143 the *vgll3*LL* genotype (probit 95% credible interval: 0.54 to 2.84, P=0.005), as well as *vgll3*EL*
144 compared to *vgll3*LL* (probit 95% credible interval: 0.26 to 3.6, P=0.026).

145

146 These results demonstrate that the *vgll3*E* allele associated with young sea-age at maturity in natural
147 populations is associated with earlier maturation also in controlled conditions, and also supports
148 stronger dominance of *vgll3*E* allele as observed in males from natural populations [4]. Seeing that
149 concurring results were observed in another wild-derived stock of Atlantic salmon [30], the genotype-
150 phenotype association between *vgll3* and maturation timing is reproducible in controlled conditions
151 and likely common in North-European strains of Atlantic salmon.

152

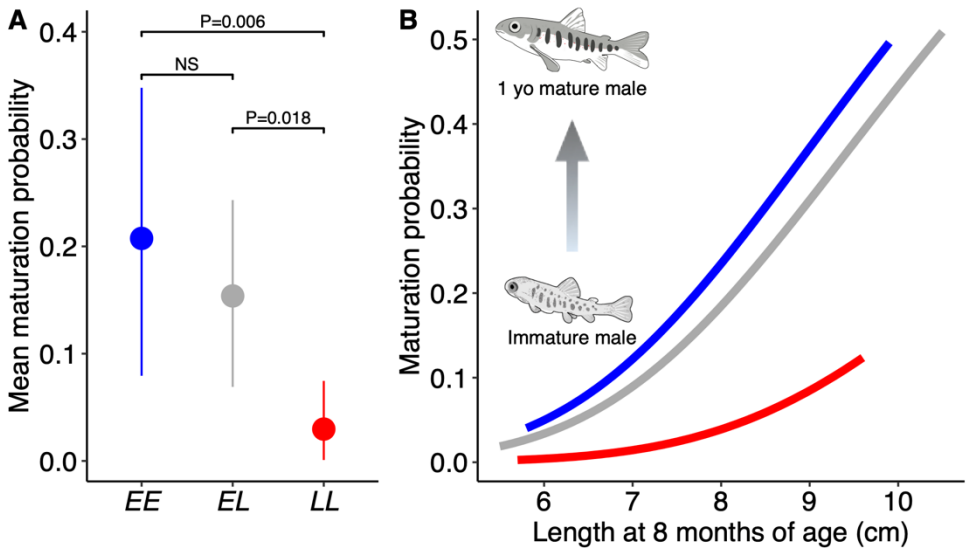


Figure 1. *Vgll3* genotype associates with male maturation probability in controlled conditions. (**A**) Probability of maturation at nine months post-hatching in controlled conditions is significantly higher in *vgll3*EE* individuals compared to *vgll3*LL* individuals. Points represent means and whiskers represent 95% credible intervals from a Bayesian Markov Chain Monte-Carlo animal model (see Methods). N_{EE}=98, N_{EL}=204, N_{LL}=82 (**B**) Maturation probability modeled as a function of fish length one month before the breeding season. Blue, *vgll3*EE*; grey, *vgll3*EL*; red, *vgll3*LL*.

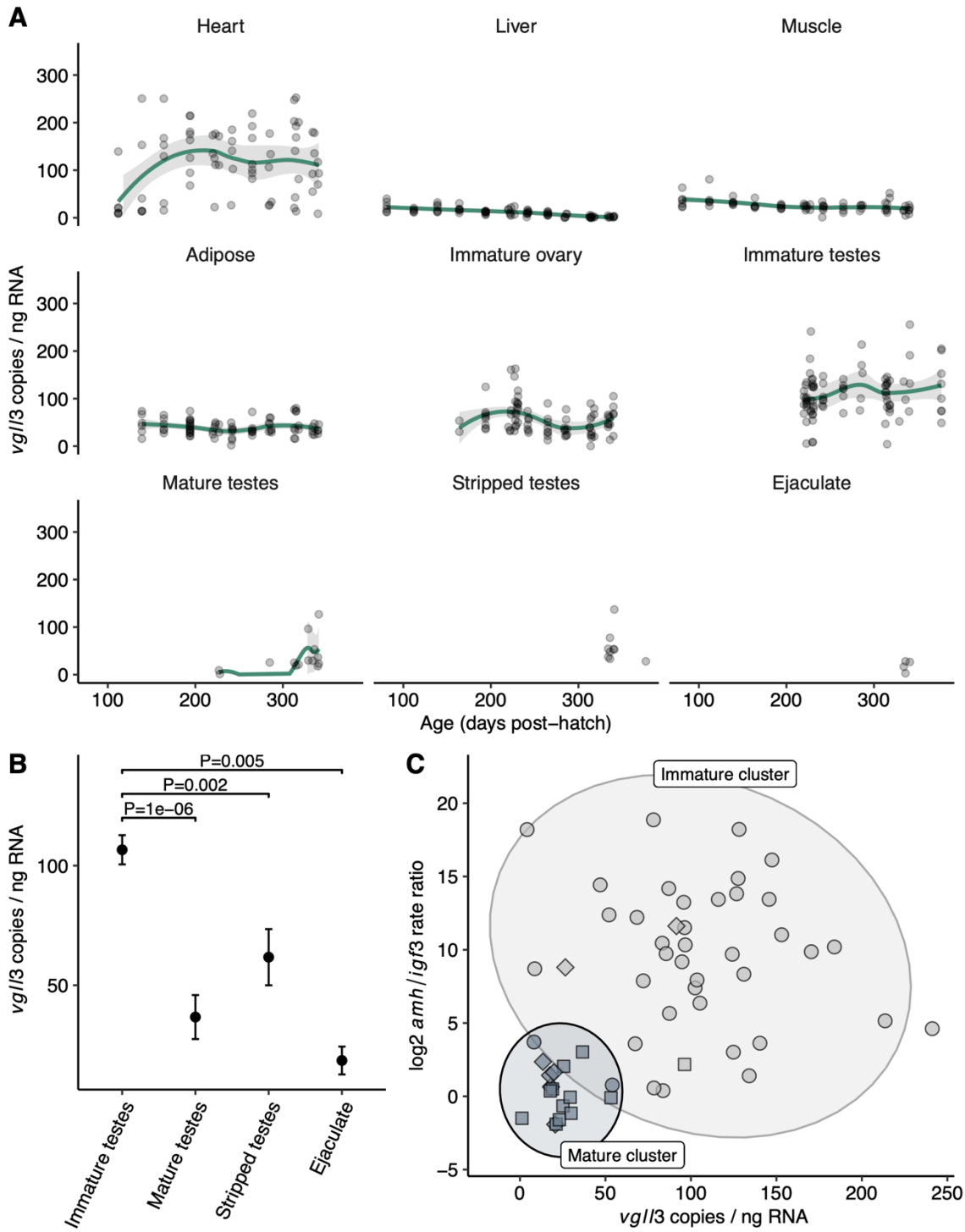
vgll3 expression correlates with maturation inhibition

Until now, it has remained unknown by which functional means, tissues, cell types and developmental time-points genetic variation in *vgll3* influences the maturation phenotype. In order to investigate the functional mechanisms resulting in earlier maturation in males carrying the *vgll3*EE* genotype and to establish the tissues and temporal time-points of relevance for *vgll3* expression, we first characterized *vgll3* transcript expression profiles of ten tissues and whole salmon juveniles in a time-series of samples including 13 time points and all *vgll3* genotypes by using reverse-transcription droplet digital PCR (RT-ddPCR) and a genotype-specific assay on exon 2 of the *vgll3* transcript (total *n*=291, Fig. S3).

Vgll3 expression in the heart was high relative to other tissues (Fig. 2A), but also very variable, similar to results from three-year old juveniles from natural populations [28]. Expression in liver showed a linear downward trend from younger to older individuals, although the overall expression level remained very low relative to other tissues (Fig. S4). *Vgll3* expression in muscle was lowest among tissues, while immature ovaries showed intermediate and relatively stable expression (Fig. 2A). Interestingly, while *vgll3* expression was high in immature testes, expression was reduced by 66% in mature testes compared to immature testes (37 copies/ng versus 107 copies/ng, i.e. 70 copies/ng fewer, 95% CI [47-93 fewer], two-tailed *t*-test, df=25.9, *P*=1e-06, n_{mature}=14, $n_{immature}$=65, Fig. 2B).

179 *Vgll3* expression was similarly lower in stripped mature testes, containing mainly mitotic cell types,
180 and in ejaculate, containing mainly mature sperm, compared to immature testes matched for sampling
181 time-point (58 copies/ng (stripped testes) versus 128 copies/ng i.e. 70 copies/ng fewer, 95% CI [30-
182 111 fewer], two-tailed *t*-test, df=21.8, P=0.002, $n_{stripped}$=8; 19 copies/ng (ejaculate) versus 134
183 copies/ng i.e. 115 copies/ng fewer, 95% CI [50-181 fewer], df=6.6, P=0.005, $n_{ejaculate}$=4, respectively,
184 Fig. 2B). Taken together, these results indicate that in salmon gonad tissues, *vgll3* expression is high
185 in immature testes and down-regulated upon the onset of male puberty.
186
187 To further test whether *vgll3* down-regulation coincides with the activation of the male puberty
188 pathway, we compared *vgll3* expression relative to the expression of *amh,* which is a repressor of
189 male puberty and Sertoli cell differentiation, and *igf3*, which promotes spermatogonial germ cell
190 differentiation [31–34]. Consistent with our visual classification of gonad morphology, immature testes
191 were characterized by typically higher *amh* expression compared to *igf3* expression, whereas
192 morphologically mature testes with differentiated (running) milt showed typically equal, or higher
193 expression of *igf3* compared to *amh* (Fig. S5). Clustering based on *amh/igf3* expression ratio and
194 *vgll3* expression level further revealed that mature testes were separated in gene expression space
195 from immature testes (bootstrap LRT [35] P<0.0001) and characterized by both low *amh/igf3* ratio
196 and low *vgll3* expression, indicating that down-regulation of both *amh* and *vgll3* expression is
197 commonly associated with completing puberty (Fig. 2C). In sum, these results point towards a
198 mechanism by which male puberty is associated with a down-regulation of *vgll3* expression,
199 corroborating previous studies that suggested *vgll3* influences Sertoli cell differentiation [27].
200
201 While the effects of *vgll3* down-regulation on the downstream transcriptional landscape of testes cell
202 types remain largely unknown, transcriptomic studies indicate that *vgll3* expression correlates with
203 changes in the Hippo signaling pathway, known to control cell differentiation and organ size [27,28].
204 To gain additional insight into the expression patterns of key players of Hippo signaling pathway and
205 *vgll3* during male puberty, we re-analyzed RNA-sequencing data from immature, prepubertal and
206 pubertal testes from an independent dataset [34]. We found that while maturation correlates with a
207 reduction in *vgll3* expression, the Hippo pathway antagonist for *vgll3*, *yap1* [26,36,37], showed the
208 opposite expression pattern with increasing expression as maturation proceeded (Fig. S6). *Vgll3* was
209 co-expressed within a group of 7073 other genes out of which 5670 were assigned function in the
210 *NCBI RefSeq* database. Gene Ontology analysis revealed an over-representation in "*transcriptional*
211 *coactivator function*" in the *vgll3* co-expressed cluster (GO:0003713, P=0.049, Fig. S6), which
212 included multiple histone acetyltransferases (LOC106607167, LOC106580491, LOC106585100),
213 nuclear receptor coactivators (LOC106607631, LOC100196047, LOC106571447, LOC106579720,

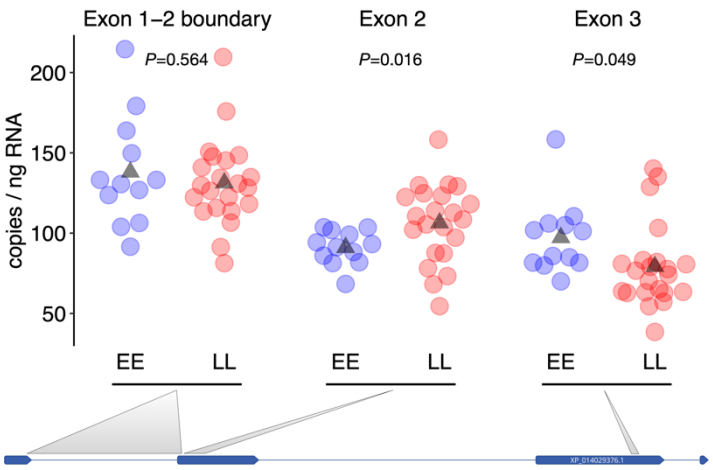
214 LOC106590658), CREB-binding proteins (LOC106590442, LOC106564464) and transcriptional
215 adapter proteins (LOC106581187, LOC106577491), suggesting that *vgll3* is embedded in a larger
216 gene regulatory network potentially influencing gene expression on a transcriptomic level.
217



218

219 **Figure 2. *Vgll3* expression across salmon tissues during the first year of development.** (**A**) *Vgll3* reaches highest
220 expression levels in the heart and immature testes. For sample size see Supplemental Fig. 3. Line, loess smoothing;
221 shading, 95% confidence interval. (**B**) Male maturation is associated with a 66% drop in *vgll3* expression compared to
222 immature testes expression levels. Reduction in *vgll3* expression is observed in all major tissues of the mature testes.
223 Stripped testes; mature testes with ejaculate removed manually. Error bars show SE. (**C**) *Vgll3* expression shows a strong

224 association with relative expression levels of *amh/igf3*. Squares; mature testes, diamonds; pubertal samples in transition
225 to mature morphology, spheres; immature testes. Ellipses delineate estimated 95% confidence intervals for each cluster.
226 Expression is measured using RT-ddPCR.
227
228 **Complex transcript expression differences between *vgll3* genotypes**
229
230 The difference in *vgll3* expression level between mature and immature testes led us to hypothesize
231 that the observed higher probability of maturation in *vgll3*EE* males may be linked to differences in
232 *vgll3* expression in immature testes of *vgll3*EE* and *vgll3*LL* males, thus resulting in differing
233 maturation probabilities for males with alternative genotypes. We tested this hypothesis by comparing
234 the testicular expression from immature individuals with different *vgll3* genotypes.
235
236 Using three different RT-ddPCR assays including two that tagged transcribed SNPs within the *vgll3*
237 coding sequence, we observed a complex pattern of expression differences where association with
238 *vgll3* genotype varied between different exons of the *vgll3* transcript. Using an assay spanning the
239 boundary of exon 1 and 2, we observed no expression difference between *vgll3* genotypes (P=0.564,
240 Fig. 3), whereas an assay spanning an SNP in exon 2 showed a 14% lower expression in *vgll3*EE*
241 genotypes compared to *vgll3*LL* genotypes (91 versus 106 copies/ng i.e. 15 copies/ng fewer, 95%
242 CI [3-28 fewer], two-sided *t*-test, df=31, P=0.014, n_{EE}=12 n_{LL}=22, Fig. 3). The opposite pattern was
243 observed using an assay spanning an SNP in exon 3, where *vgll3*EE* genotypes showed 18.5% higher
244 expression compared to *vgll3*LL* genotypes (97 versus 79 copies/ng i.e. 18 copies/ng more, 95% CI
245 [0.08-36 more], two-sided *t*-test, df=25, P=0.049, n_{EE}=13 n_{LL}=23, Fig. 3). The other two tissues with
246 highest *vgll3* expression levels, heart and adipose, did not show a clear relationship between *vgll3*
247 expression level and genotype, maturation or age as assayed on the exon 2 assay (Fig. S7 & Fig. S8).
248



249
250 **Figure 3. *Vgll3* expression differences in immature testes using three RT-ddPCR assays in different *vgll3* regions.**
251 Individuals with *vgll3*EE* (blue) and *vgll3*LL* (red) genotypes show varying levels of expression differences along the

252 *vgll3* transcript. Points represent the mean of two replicates per individual sample (triangles; genotype means, *P*-values;
253 two-sided *t*-test). Grey lines connect RT-ddPCR assays with the approximate locations of their amplification targets on
254 the NCBI annotated *vgll3* transcript model with lines representing introns, rectangles representing exons.
255
256 **Rare *vgll3* transcript isoforms in immature testes**
257
258 The results from RT-ddPCR assays spanning three exons indicated that expression differences
259 between *vgll3* genotypes are restricted to specific exons of the *vgll3* transcript. This suggests that
260 *vgll3* genotype may influence the structure of *vgll3* transcript being expressed, because conflicting
261 expression differences within a gene can be produced by confounding signal from multiple transcript
262 isoforms. We tested this by characterizing *vgll3* transcript structure in immature testes using 5' and
263 3' RACE (Rapid Amplification of cDNA Ends). Using pooled samples of five individuals for each
264 of the *vgll3*EE*, *vgll3*EL* and *vgll3*LL* genotypes, we observed that all genotypes expressed multiple
265 isoforms that differed in length 5' of exon 3 (Fig. S9). We Sanger sequenced of a total of 278 clones
266 to determine the structure of the expressed isoforms in *vgll3*EL* genotypes and observed at least six
267 distinct RACE fragments representing at least two distinct isoforms (from a total of 202 quality
268 sequences, Fig 4A). The longer isoform aligned with the NCBI annotated transcript with multiple
269 RACE fragments representing alternative lengths of the 5' UTR or exon 1. The shorter second
270 isoform was transcribed from an alternative 5' UTR within the NCBI annotated first intron with
271 translation likely initiated from start codons within exon 2. Further, Sanger sequencing revealed a
272 previously undetected C/T SNP in the alternative 5' UTR at position ssa25:28,655,795, where the
273 genotype was linked to the *vgll3*E/vgll3*L* genotype.
274
275 To further confirm the expression levels of *vgll3* isoforms, we re-analyzed publicly available RNA-
276 seq data from male gonads [34] using Cufflinks [38]. Combined RNA-seq data from three immature,
277 three prepubertal and three pubertal individuals that we identified as *vgll3*EL* heterozygotes
278 supported our discovery of *vgll3* isoforms with alternative transcription start sites. Twenty-two RNA-
279 seq reads mapped to an un-annotated region within intron 1 of the NCBI annotated transcript that
280 spliced into exon 2 with a canonical GT/AG 5'/3' (donor/acceptor) sequence at the splice site (Fig.
281 4B), and that corresponded to the alternative 5' UTR identified above. The abundance of the shorter
282 isoform accounted for 18% of the expression of the sum of *vgll3* isoforms (Cufflinks FPKM 0.56 out
283 of total of FPKM 0.56 + 2.52).
284
285 Taken together, based on varying expression across the exons of the *vgll3* transcript in conjunction
286 with the detection of multiple *vgll3* isoforms, we hypothesize that fish carrying *vgll3*EE* and
287 *vgll3*LL* genotypes may preferentially express alternative isoforms of *vgll3* that have different

transcription start sites, 5' UTRs and N-terminal protein sequences.

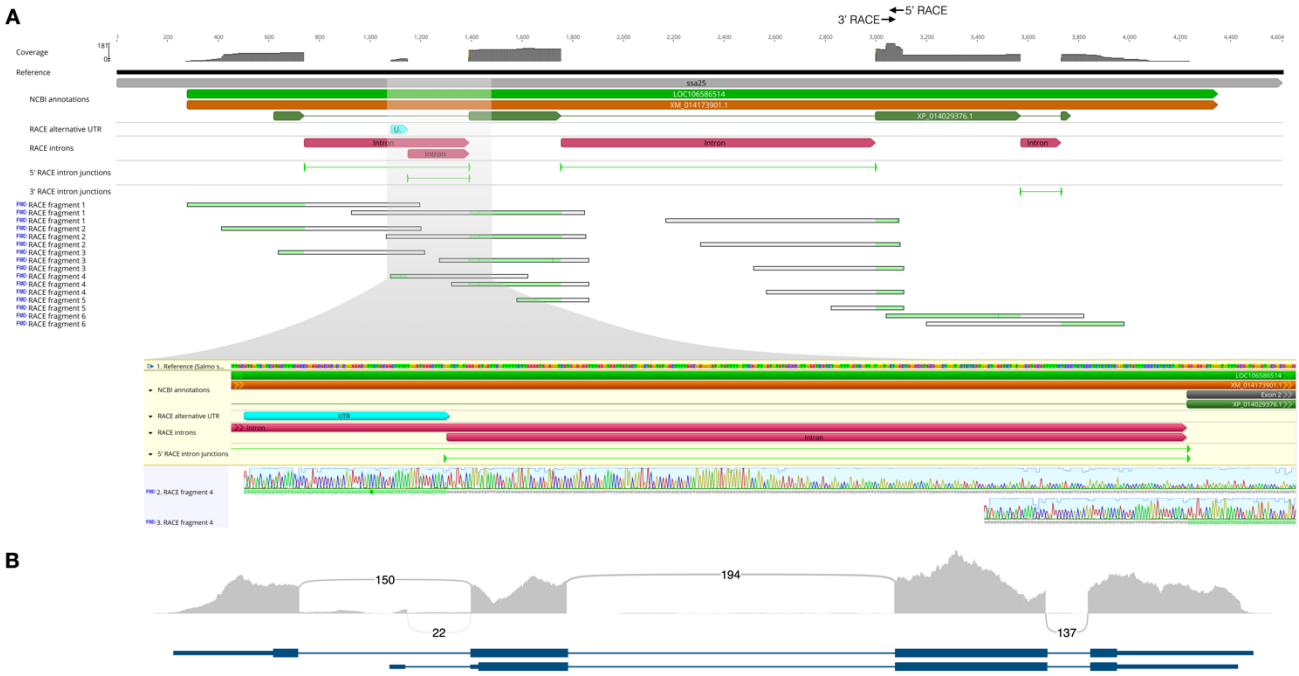


Figure 4. Characterization of a *vgll3* transcript isoform with an alternative 5' UTR. (**A**) 5' and 3' RACE sequencing of pooled total RNA from *vgll3*EL* individuals identified an isoform with an alternative 5' UTR. Browser tracks of Sanger sequencing show (from top) coverage, NCBI annotations, identified UTR, introns and intron junctions, and representative RACE fragments with their alignment to the reference gene model (in light green). Highlighted region shows gene annotations and Sanger sequencing tracks of a representative RACE fragment corresponding to the alternative 5' UTR containing isoform. (**B**) RNA-seq sequencing coverage and splice events from combined immature, pre-pubertal and pubertal testes. Numbers correspond to RNA-seq reads spanning intron splice sites. Gene models depict transcripts identified using Cufflinks [38].

Cis-regulatory differences in *vgll3* track alternative isoforms

Expression of alternative transcript isoforms can be mediated by *cis* and/or *trans*-regulatory mechanisms, where expression of isoforms is influenced by regulatory differences linked or unlinked, respectively, to the genotype at the locus. We hypothesized that if the expression differences observed between homozygote *vgll3*EE* and *vgll3*LL* individuals were due to *cis*-mediated isoform expression differences, the differing expression levels in homozygotes would be conserved as allele-specific expression within *vgll3*EL* heterozygotes. Expression level differences between the alleles in heterozygotes can be accredited to differences in *cis*-regulation because both alleles are being expressed in the same cellular *trans*-acting environment [39].

Using allele-specific assays for the newly discovered 5' UTR, exon 2 and exon 3, we observed that heterozygotes showed allele-specific expression differences consistent with a largely *cis*-regulatory

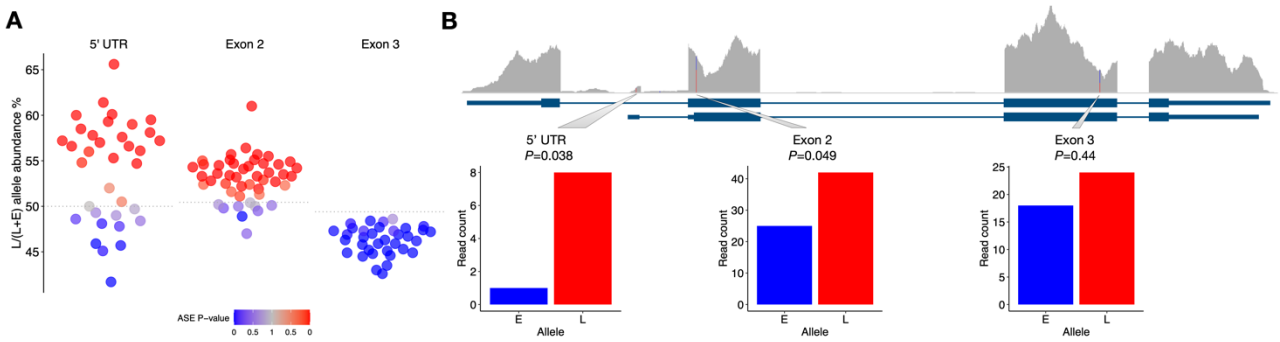
basis of isoform expression differences. For the alternative 5' UTR, heterozygotes tended to express the *vgll3*L* allele at a higher level compared to the *vgll3*E* allele, with on average 54% of mRNA molecules representing from the *vgll3*L* allele. Using Poisson modeling of expression rates, 21 of the 35 tested heterozygotes (60%) showed significantly higher *vgll3*L* expression at *P*-value threshold of 0.05, while only five (14%) showed higher *vgll3*E* allele expression at the same *P*-value threshold (Fig. 5A).

For exon 2, where we observed higher expression in *vgll3*LL* homozygotes compared to *vgll3*EE* individuals, allele-specific expression showed a tendency towards higher expression of the *vgll3*L* allele with on average 53% of the *vgll3* mRNA molecules expressed from the *vgll3*L* allele (Fig. 5A), versus 50.4% expected from heterozygous DNA used as control for differences in efficiencies of allele-specific probes. Thirty-one of the 47 heterozygotes analyzed (66%) showed significantly higher *vgll3*L* allele expression, while only one individual (2%) showed significantly higher *vgll3*E* allele expression ($P<0.05$, Poisson rate-ratio test, Fig. 5A). In terms of expressed mRNA molecules, the average magnitude of the *cis*-regulatory differences observed between *vgll3*E* and *vgll3*L* alleles (6.8 mRNA molecules per ng of RNA) was large enough to account for 89% of the average expression differences between *vgll3*EE* and *vgll3*LL* homozygotes (15.3 mRNA molecules, i.e. (2*6.8)/15.3*100=89%).

For exon 3, where *vgll3*LL* homozygotes showed lower expression compared to *vgll3*EE*, allele-specific expression differences were also consistent with *cis*-regulation with on average 46.2% of mRNA molecules expressed from the *vgll3*L* allele in *vgll3*EL* heterozygotes (compared to 49.4% in heterozygous DNA control). Thirty-one of 35 tested heterozygotes (89%) showed significantly lower expression of the *vgll3*L* allele ($P<0.05$). The average magnitude of the *cis*-regulatory difference (7.5 mRNA molecules per ng of RNA) accounted for 83% of the average homozygote expression difference (18 mRNA molecules, i.e. (2*7.5)/18*100=83%).

For further validation of the *cis*-regulatory differences in *vgll3*, we re-analyzed publicly available RNA-seq data from immature male gonads [34]. We identified three immature individuals as heterozygous for the same *vgll3* SNPs we analyzed with our allele-specific RT-ddPCR assays and quantified *vgll3* allele-specific expression based on RNA-seq reads overlapping the SNP positions. Inspection of RNA-seq reads revealed the individuals to be heterozygous all of the three linked SNPs in *vgll3* targeted by our RT-ddPCR assays (Fig. 5B). Consistent with our results from RT-ddPCR, the newly discovered alternative 5' UTR showed strong allele-specific expression, being expressed almost solely from the *vgll3*L* allele (8 versus 1 reads, $P=0.038$). The *vgll3*L* allele in exon 2 had

348 higher expression compared to the *vgll3*E* allele in all three heterozygote individuals (13 versus 8
349 reads, 10 versus 5 reads, 11 versus 10 reads), and when reads from all three males were combined
350 (42 reads versus 25 reads, binomial exact test *P*=0.049, Fig. 5B). For exon 3, the difference in *vgll3*L*
351 and *vgll3*E* alleles was non-significant (*P*=0.44).
352
353 These results show that the *vgll3* allele expression differences observed between alternative
354 homozygotes tend to be conserved in direction and in magnitude when assayed in heterozygotes. Re-
355 analysis of an independent data-set supports this conclusion. In sum, our results strongly support that
356 the differences in maturation timing between *vgll3* genotypes being driven by *cis*-regulated
357 expression differences of alternative *vgll3* isoforms.
358



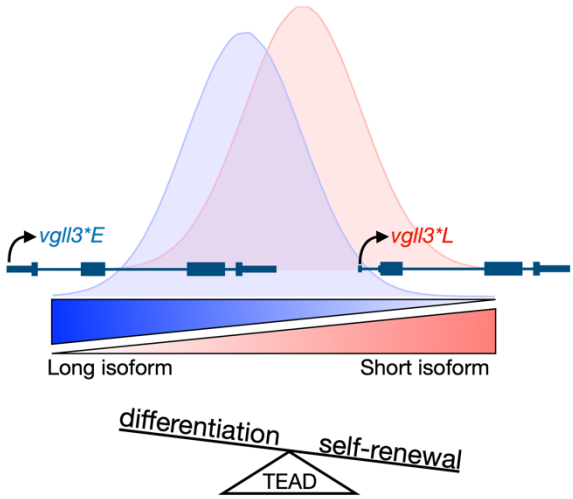
359
360 **Figure 5. Allele-specific expression of transcribed SNPs supports a *cis*-regulatory basis for differences in *vgll3***
361 **isoform expression.** (**A**) RT-ddPCR assays of *vgll3* allele-specific expression in heterozygotes show that alternative
362 homozygote expression differences are conserved within heterozygotes. For the alternative 5' UTR, exon 2 and exon 3,
363 on average 54%, 53% and 46% of *vgll3* mRNA molecules are derived from the *late* allele, respectively. For exon 2 and
364 exon 3 assay the null hypothesis for testing of 1:1 allele-specific expression ratio is calculated from heterozygote DNA
365 (dashed line, *N*=10). The color scale indicates significant (blue; *vgll3*E* allele more expressed, red; *vgll3*L* allele more
366 expressed) and non-significant (grey) allele-specific expression *P*-values (estimated from Poisson modelling of positive
367 RT-ddPCR droplets). (**B**) Re-analysis of an independent RNA-seq data set is consistent with the RT-ddPCR results and
368 shows allele-specific expression of an alternative 5' UTR and exon 2. Tracks represent (from top) RNA-seq coverage,
369 identified transcript models, and RNA-seq read coverage of SNPs. *P*-value are computed using a binomial test.
370

371 **Discussion:**

372 The molecular basis of variation in life history traits and their trade-offs have thus far remained
373 elusive because many life history traits are either polygenic, experimentally intractable, or their
374 genetic basis has only recently been discovered. Here, we show that variation in age at maturity, a
375 key life history trait, is associated with *cis*-regulatory differences in isoform expression of the
376 transcription co-factor gene *vgll3* in immature testes of Atlantic salmon males. Using targeted
377 expression assays, time series sampling in ten tissues and characterization of transcript structure, we
378 discovered that genotypes that tend to mature early versus those that tend to mature late express

379 different mixtures of *vgll3* transcript isoforms pre-puberty. The expression differences were allele-
380 specific in heterozygotes, suggesting that divergence in *cis*-regulation underlies the associations
381 between *vgll3* genotype, isoform expression and maturation probability.
382
383 We found strong evidence that *vgll3* genotypes associated with maturation age phenotypes do so
384 through a mechanism of *cis*-regulated differences in transcript isoform expression in immature testes.
385 The expressed *vgll3* isoforms differed in their transcription start sites, but are predicted to all produce
386 a functional *vgll3* protein. Alternative transcription start sites are thought to be important for
387 determining cell-types and their proliferation by influencing not only protein sequence, but also
388 through post-transcriptional regulation such as mRNA translation and stability [22,40,41]. Validating
389 the functional differences between *vgll3* isoforms and characterizing isoform expression patterns of
390 *vgll3* in additional tissues will be an important avenue for future research. Expression of different
391 transcript isoforms has been established as the primary mechanism of important developmental
392 turning points in several other species. In many insects for example, alternatively spliced isoforms of
393 the gene *doublesex* act as sex determining factors [42]. Most species-specific differences in transcript
394 isoform expression have been linked to *cis*-regulatory differences [43,44]. *Cis*-regulatory changes that
395 affect alternative splicing are thought to reside within the un-spliced pre-mRNA [44]. For expression
396 from alternative transcription start sites such conservation of *cis*-regulatory placement has not been
397 shown, as initiation of transcription can be influenced by regulatory elements not only in the core
398 promoter, but also in distal enhancer sites [45]. Overall, genetic variation influencing *cis*-regulatory
399 loci is known to be an important source of phenotypic diversity and a common source of evolutionary
400 change [46–52]. The results reported here support the notion that this is also a strong candidate for the
401 molecular mechanisms underlying variation in complex life histories as well.
402
403 The transcription co-factor function of *vgll3* suggests that isoform changes in *vgll3* expression
404 transmit downstream to expression changes in gene regulatory networks associated with cell
405 proliferation and organ development [37]. Like its vertebrate and invertebrate homologs, the salmon
406 *vgll3* has been proposed to control cell differentiation through interacting with the TEAD-family of
407 transcription factors of the Hippo signaling pathway [27,28]. Vgll3, like related co-factors in mice,
408 humans and *Drosophila*, is predicted to compete for TEAD accessibility with alternative co-factors
409 downstream of the Hippo signaling pathway controlling for organ size (Yap/Taz co-factors) [26,36].
410 Consistent with this, by analyzing RNA-sequencing data from immature, prepubertal and pubertal
411 testes, we found that *vgll3* and *yap1* showed opposite expression patterns with respect to maturity
412 status. Salmon *vgll3* expression was also found to negatively correlate with *yap1* expression in early
413 development [28]. We hypothesize that changes in *vgll3* isoform abundance associated with *vgll3*E*

414 and *vgll3*L* alleles may influence the tendency of cell differentiation in immature testes by changing
415 interaction between Vgll3 and TEADs and/or Vgll3-Yap (Fig. 6). Clarifying the molecular
416 relationships between *vgll3,* TEADs and *yap1* remain as important topics for future research.
417



418
419 **Figure 6. Model for a molecular mechanism of *vgll3* genotype-dependent liability in puberty timing.** Hypothetical
420 distributions represent relative abundances of *vgll3* isoforms in individuals with given *vgll3* genotypes (blue, *vgll3*E*;
421 red, *vgll3*L*). The association between *vgll3* genotype and maturation timing arises from *cis*-regulatory differences
422 influencing *vgll3* isoform expression. The *vgll3*L* allele tends express more of the shorter isoform compared to the
423 *vgll3*E* allele, thus changing the balance between *vgll3*-controlled differentiation versus self-renewal. TEAD interaction
424 with *vgll3* isoforms tip the balance between tendency for self-renewal (higher expression of the short isoform) and
425 differentiation (long isoform).
426
427 The relatively low expression level of *vgll3* and variation around genotype means also merit careful
428 consideration of the limits of our study. First, some of the variation around the mean expression levels
429 within *vgll3* genotypes can possibly be explained by additional genetic variation segregating in the
430 experimental families and influencing *vgll3* expression in *cis* and in *trans*. Second, as the overall
431 expression level of *vgll3* is very low, an alternative interpretation of the results could be that the small
432 but statistically significant differences in expression across different exons of the *vgll3* transcript are
433 driven by non-biological sources, such as noise in the measurement of low expression levels, or
434 differences in assay efficiency and specificity. However, concurring results from re-analysis of an
435 independent data-set, our careful control experiments, and the fact that the observed expression
436 differences were specific to tissues and not observed in a systematic manner across all samples do
437 not support this interpretation. The possibility exists also that *vgll3* is included in, but not determining
438 the output of, the molecular pathway determining maturation timing, or that variation in the timing
439 of measurement versus initiation of puberty causes variation in the genotype-phenotype association.
440 Further functional validation through manipulation of isoform expression and allele-specific
441 expression is an important future direction for validating the role of *vgll3* isoform-specific *cis*-

442 regulation in maturation timing.
443
444 It is important to note that although we observed a *cis*-regulatory differences associated with isoform
445 expression and maturation probability, it does not exclude the possibility that protein-coding changes
446 within the *vgll3* gene [4,5] could play additional functional roles in determining maturation timing.
447 Further, inferring an expression-genotype-phenotype association in male gonads does not exclude the
448 possibility that *vgll3* also has functional relevance in other tissues/processes (e.g. adipogenesis). In
449 fact, pleiotropic effects of single genes that are under the control of tissue and development-specific
450 regulatory sequences fits well with what we understand from the modularity of *cis*-regulation [46].
451 The expression pattern of *vgll3* and its paralog indeed support additional functions for the gene pair
452 in adipogenesis and female gonad development [28]. In addition, we observed relatively high *vgll3*
453 expression in heart as well as testes, which may indicate a pleiotropic action of *vgll3* in heart
454 development and function through the Hippo signaling pathway [53]. The role of changes in isoform
455 expression of *vgll3* in such pleiotropic effects warrants future research.
456
457 Although the effects of protein-coding changes on maturation phenotype were not addressed here,
458 support for the benefits of *cis*-regulatory changes compared to protein-coding changes in the
459 evolution of phenotypic diversity can be drawn from studies on morphological traits. Regulatory
460 changes often show higher modularity across tissues and development, additivity, and lower
461 pleiotropy, which are thought to inflict less evolutionary constraints compared to protein-coding
462 changes [46,54], especially in cases where proteins modulate the transcription of other genes as is the
463 case with *vgll3*. These features of evolvability of regulation are also well demonstrated here, where
464 we observed alternative expression dynamics (e.g. rising versus diminishing expression in heart and
465 liver, respectively, as individuals age), and genetic effects driving expression differences restricted
466 to one tissue and developmental stage (immature testes). Overall, these features are expected to lead
467 to *cis*-regulation being a more likely source of evolution in life history traits compared to coding
468 changes.
469
470 We have shown here that polymorphisms in the *vgll3* region associate significantly with one-year-
471 old male age at maturity in controlled conditions. The effect of *vgll3* genotype on maturation timing
472 in our experiment was arguably smaller compared to that observed in wild populations and for sea-
473 age at maturity [4]. Possible explanations for the divergent finding in the two studies can be related to
474 the differences in (*i*) traits (age at maturity versus probability of maturation), (*ii*) modeling approaches
475 (based on logit versus in our case probit distributions), (*iii*) wild versus husbandry environments, (*iv*)
476 differences in the assessment of maturity status, (*v*) variation in *vgll3* contribution to age at maturity

phenotypes across populations and (*vi*) possible differences in the genetic architecture of the distinct age at maturity phenotypes (sea-age at maturity versus early male maturity).

Taken together, here we have revealed key functional insights into the temporal expression dynamics and likely molecular mechanisms that translate genetic variation in a major effect locus to variation in life history strategies in Atlantic salmon. Many life history traits are encoded by many genes where it has remained challenging to identify the loci responsible for variation [55]. In the light of the results presented here, gene regulatory evolution and isoform variation are likely candidates for mechanisms mediating alternative life history strategies that merits further investigation across species.

487 **Materials and Methods:**

488 **Crosses used in the study**

489 We controlled-crossed eight 2x2 factorial matings of unrelated individuals that each included one
490 *vgll3*EE* and one *vgll3*LL* male and female (see ref *13* for genotyping details), thereby resulting in
491 32 families with reciprocal *vgll3* genotypes *EE*, *EL*, *LE* and *LL* (allele order denotes female and male
492 parents of origin respectively, Supplemental Fig. 1). We used parents from the "Oulujoki"
493 broodstock, which is a mixture of several salmon populations from northern Baltic rivers and is
494 maintained by the Natural Resources Institute Finland (LUKE) [56]. All parents were selected for
495 having a homogenous age-at-maturity-associated haplotype encompassing the *vgll3* (chromosome 25
496 paralog) coding sequence, the top non-coding association [4], and the coding region of the adjacent
497 *akap11* gene (collectively referred to as "*vgll3* genotype").

498

499 **Husbandry design and conditions**

500 Following fertilization in October 2017, familial egg batch replicates were raised in two egg
501 incubators until first feeding (~ 2 months post-hatching) with mesh-separators to contain families in
502 individual compartments. Families were assigned to compartments randomly. At first feeding,
503 familial batches were combined and 48 individuals per family were randomly selected and distributed
504 evenly across eight 0.25 m^3 tanks using recirculated water i.e. on average, each tank included six
505 individuals per family. Prior to tank allocation, each individual was tagged with visible elastomer
506 (VIE) at the base of the caudal fin to enable identification of the *vgll3* genotype of an individual by a
507 fluorescent color (initial n=1300). Light cycle and water temperature (min=6.3, max=17.7 °C) during
508 husbandry corresponded to the local seasonal cycle at 61°N. Fish were fed first with live *Artemia* for
509 three days, after which we fed commercial aquaculture feed *ad libitum* (Raisio Baltic Blend) with
510 pellet diameter increasing over time according to fish size. Temperature, oxygen and nitrogen-
511 component levels were monitored on a regular basis. Experimentation was conducted under animal
512 experimentation license ESAVI/2778/2018.

513

514 **Individual tagging, family assignment and morphological measurements**

515 At eight months post-hatching and before the onset of the first breeding season we tagged remaining
516 fish using passive integrated transponder (PIT) tags and collected fin-clips. We extracted DNA using
517 standard chelex or salt extraction methods, genotyped all fish for 141 informative SNPs [57], assigned
518 their genotypic sex based on genotyping coverage of the salmon sex determining locus *sdy* and
519 assigned individuals to families using the likelihood method implemented in SNPPIT [58]. We
520 recorded the length and mass of all fish monthly after PIT-tagging. We non-lethally assessed the
521 maturity status of all males by testing for the presence of running milt at nine months post-hatching.

522 Morphological measurement and non-lethal maturation assessment were performed in a random order
523 such that the assessor did not have knowledge of *vgll3* genotype.
524
525 **Statistical model of male maturation probability**
526 We modeled the probability of males reaching maturity within their first year given their *vgll3*
527 genotypes by using a GLMM (generalized linear mixed model) approach to implement a genetic
528 threshold model that takes full account of the relatedness among individuals [59,60]. We estimated
529 maturation probability using a generalized linear animal model with probit-link function under
530 Bayesian Markov Chain Monte Carlo simulations with the R-software [61] version 3.6.1 and package
531 MCMCglmm [62] for the following model: $y = vgll3 + length + a + t + e$, where y is a vector of
532 maturation binaries, *length* is mean-centered and variance scaled natural logarithm of length (cm),
533 *vgll3* is the genotype, *a* is the additive genetic relationship-matrix predicted animal effect [63], *t* is the
534 common environmental tank effect and *e* the error term (with fixed variance of 1). We used priors
535 following a χ_1^2 distribution [64] and based inferences on 10,000 samples collected every 1,000
536 iteration after burn-in of 10,000. Model convergence and adequate sampling was ensured by visual
537 trace plot inspection, autocorrelation < 0.1 for lag-2 parameter samples, and using Heidelberger and
538 Welch's convergence diagnostic [65]. We transformed maturation probabilities from probit-scale to
539 normal probability scale for ease of interpretation. Heterozygote *vgll3*LE* and *vgll3*EL* (order of
540 alleles denotes maternal origin) were considered as one group as initial analysis indicated that
541 maturation probabilities between these genotypes were not significantly different.
542
543 **Temporal sampling, RNA extraction and RT-ddPCR**
544 We sampled individuals with *vgll3*EE, vgll3*EL, vgll3*LE* and *vgll3*LL* monthly (3-32 in total per
545 genotype per time point) and assessed sex, gonad size and maturity by dissection. We sacrificed the
546 animals by MS-222 overdose and dissected ten tissues (brain (timepoints 12-13 months post hatch);
547 heart, liver, muscle (timepoints 3-13 months post hatch), adipose (timepoints 5-13 months post
548 hatch); immature ovary (timepoints 6-13 months post hatch); immature testes (timepoints 7-13
549 months post hatch); mature testes (timepoints 8 and 10-12 months post hatch); stripped mature testes
550 (timepoints 12-13 months post hatch) and ejaculate (timepoint 13 months post hatch)). As immature
551 male gonads had no recordable weight on a scale with a detection limit of 10 mg thus preventing the
552 use of gonadal-somatic index, we classified each sacrificed male fish visually as immature (not
553 enlarged, transparent color gonads), pubertal (slightly enlarged, transparent or white color gonads) or
554 mature (enlarged white color gonads). We flash froze the samples on liquid nitrogen and stored the
555 samples in -80 °C. We homogenized the tissues using an OMNI Bead Ruptor Elite machine, in 2 ml
556 or 7 ml (for mature gonads) with 2.4 mm steel beads. We used Macherey-Nagel NucleoSpin RNA

557 kits in single and in 96-well plate formats to extract total RNA and treated the samples with additional
558 DNase (Invitrogen Turbo Dnase) to remove any residual DNA. We verified total RNA quality using
559 Agilent BioAnalyzer for a random set of 24 samples and found all samples to have RIN>9. We
560 measured total RNA yield using ThermoFischer Qubit and Quant-it reagents, diluted the samples to
561 working concentrations and re-measured RNA concentrations. For samples in Figs. 2 and 3 (all gonad
562 samples) we repeated the RNA concentration measurement once more and calculated the mean of the
563 duplicate measurements, which we used in data normalization. We used ~20 ng (2-8 months post-
564 hatch) or ~50 ng (8-14 months post-hatch) of total RNA as template in one-step reverse transcription
565 droplet digital PCR (RT-ddPCR) [66] (BioRad) with TaqMan probes specific to *vgll3* alleles to
566 quantify *vgll3* absolute expression (Supplementary table 1). Measurement of *amh* and *igf3* expression
567 was performed as above, with the exception of using ~2 ng of total RNA template and TaqMan probes
568 that were labeled with alternative fluorophores and assayed in the same RT-ddPCR reactions. We
569 verified the specificity of *vgll3*-genotype specific TaqMan assays with genomic and synthetic DNA
570 templates (Fig. S10) and optimized melting temperatures for all assays using a gradient. Genomic
571 DNA could not be used as control for the *vgll3* 5' UTR assay because the assay spans an intron-exon
572 boundary. Genomic and synthetic DNA controls and null control were included in each RT-ddPCR
573 *vgll3* run, while genomic DNA and null control were included in *Amh* and *Igf3* RT-ddPCR runs.
574

575 **Expression analyses**

576 We analyzed RT-ddPCR results using BioRad QuantaSoft software. We manually set minimum
577 signal thresholds for *vgll3*L* and *vgll3*E* channels because visual inspection of RT-ddPCR signal
578 indicated that, for *vgll3* exon 2 assay, the *vgll3*L* specific allele showed low background signal on
579 the *vgll3*E* channel (Fig. S10). We excluded samples that failed *vgll3* RT-ddPCR reaction using a
580 pre-determined minimum threshold of 10,000 quality-filter passed droplets to include only samples
581 where expression could be estimated with low error. We excluded *amh/igf3* expression data from
582 samples where RT-ddPCR signal was saturated (over 98% of droplets were positive) because mRNA
583 abundances cannot be accurately estimated with saturated RT-ddPCR data. We excluded seven
584 samples where the genotypic sex was inconsistent with morphological sex of the gonads. We
585 excluded three mis-labeled individuals with inconsistent *vgll3* genotypes and allelic expression (e.g.
586 *vgll3*EE* homozygotes with expression signal only from *vgll3*LL* specific probes). We removed two
587 heterozygote samples that showed extreme monoallelic expression of the *vgll3*L* allele (fractional
588 abundance over 90%), possibly representing labeling or sampling error. We then used custom R
589 scripts to normalize the number of *vgll3* mRNA molecules to the amount of input RNA and analyze
590 the data and plot the results (available upon acceptance).
591

592 For investigation of the effect of age/season and *vgll3* genotype to *vgll3* expression in immature testes
593 and immature ovaries, we used the following regression model: $y = vgll3 + x + e$, where y is *vgll3*
594 total mRNA copies, x is days post-hatch, *vgll3* is the genotype and e the error term. We tested the
595 effect of *vgll3* and x (days post-hatch) individually using a corresponding model.
596
597 We characterized the relative expression levels of the key male maturity genes *amh* and *igf3* by way
598 of Poisson modeling of expression rate-ratio for each sample. The number of positive droplets with
599 expression signal in RT-ddPCR data follows a Poisson distribution with lambda being the rate of
600 positive droplets over all droplets ([66], https://www.bio-
601 rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf). We calculated the Poisson rates of expression from
602 *amh* and *igf3* -positive droplets over the total amount of analyzed droplets (the rate ratio of *amh/igf3*
603 expression) using the R function *poisson.test*.
604
605 We tested for presence of distinct clusters of gene expression patterns on *amh/igf3* and *vgll3*
606 expression space using Gaussian mixture modeling as implemented in the R package *Mclust* version
607 5 [35]. We chose the best fitting mixture of Gaussian distributions based on Bayesian Information
608 Criterion using the function *mclustBIC* and further tested the best fit mixture model for the presence
609 of more than one distributions using a bootstrapping likelihood-ratio test as implemented in the
610 function *mclustBootstrapLRT* with 1000 bootstrap replications.
611
612 **RACE**
613 To perform RACE we followed the manufacturer's instructions using Takara SMARTer 5'/3' RACE-
614 kit with the following modifications. We pooled equal amounts (25 ng) of immature testis total RNA
615 from six individuals per genotype with either *vgll3*EE, vgll3*EL* or *vgll3*LL* genotypes to create
616 three genotype specific pools, and performed RACE with primers aligning to exon 3 of the *vgll3*
617 transcript. After analysis on agarose gel, we purified the RACE PCR products using Macherey-Nagel
618 NucleoMag DNA binding beads, cloned the PCR products to a pRACE vector using the SMARTer
619 kit reagents, transformed the plasmids to *E. coli* and grew the transformation at 37 degrees C o/n on
620 LB-ampicillin (100 mg/ml) plates. Sanger sequencing of transformed PCR products was performed
621 as described in ref [67]. Briefly, colonies were picked into 96 deep-well plates containing 1 ml growth
622 media and incubated in a shaker o/n. An aliquot (150 ul) was transfered to a PCR plate and spin down.
623 The pellet was dissolved in 150 water and heated to 95 C for 10 min and 2 ul was used as template
624 in PCR with universal and reverse primers (UP-47 and RP-48). The obtained PCR products were
625 sequenced with UP and RP universal primers using BigDye Chemistry v 3 and analyzed on ABI
626 3130xl Sequencer (Life Technologies) at the DNA Sequencing and Genomics lab Institute of

627 Biotechnology University of Helsinki. We used *Geneious Prime 2020.1.2* to automatically detect and
628 trim the sequences of vector DNA and low-quality sequences above 1% base error rate. We aligned
629 the sequences to the *vgll3* region downloaded from NCBI using the *Geneious RNA* mapper with the
630 highest sensitivity settings and allowing for novel intron detection.
631

632 **Allele-specific expression**

633 Testing of *vgll3* allele-specific expression in heterozygotes was performed in a similar manner as
634 above for *amh/igf3* expression (test for Poisson rate ratio), with the exception that signal from positive
635 droplets now represented expression from either the *vgll3*E* or the *vgll3*L* allele. We defined the
636 expression rate of *vgll3*E* allele as the number of droplets positive for the *vgll3*E* probes over all
637 probes that passes the QuantaSoft standard quality filter (including empty droplets), and
638 correspondingly for the *vgll3*L* allele. We then tested for difference in Poisson rates for each
639 individual heterozygote using the R function *poisson.test* and *P*-value threshold of 0.05. We
640 calculated the null expectation for equal expression rates based on RT-ddPCR with 10 replicates of
641 *vgll3* heterozygous DNA as template. The assay for the alternative 5' UTR spanned an intron-exon
642 boundary, and therefore we could not use genomic DNA as control (the assay does not amplify with
643 DNA as template). We assumed 50% efficiency for this assay.
644

645 We estimated that on average the *cis*-regulatory difference in *vgll3* allele-specific expression can
646 account for 89% of the average expression difference between *vgll3*EE* and *vgll3**LL homozygotes
647 in exon 2 and 83% in exon 3. We estimated the 89% for exon 2 as follows (analogous calculation for
648 exon 3 not shown): On average *vgll3*EL* individuals expressed 113 *vgll3* mRNA molecules (per ng
649 of input RNA, Fig 3A). Of these, on average, 53.5 molecules were expressed from the *vgll3*E* allele
650 and 60.3 molecules from the *vgll3*L* allele, giving an average *vgll3*L* relative abundance of 53%
651 (60.3/(53.5+60.3)*100=53%). From above we can see that average difference between *vgll3*E* and
652 *vgll3*L* mRNA counts is 6.8 mRNA molecules (60.3-53.5=6.8). Corresponding, average difference
653 between *vgll3*EE* and *vgll3*LL* genotypes is 15.3 mRNA molecules (106.35-91.06=15.3). The
654 percentage of *vgll3*EE- vgll3*LL* difference accounted by *vgll3*E- vgll3*L cis*-expression difference
655 is thus (2*6.8)/15.3*100=89%, as *vgll3*EE* and *vgll3*LL* have two of either allele each.
656

657 **RNA-seq analyses**

658 We re-analyzed RNA-sequencing data from immature, prepubertal and pubertal Atlantic salmon
659 testis samples [34] in order to investigate the relative expression patterns of *vgll3* and *yap1* genes. We
660 downloaded raw RNA-seq data for immature (SRR8479243, SRR8479245, SRR8479246),
661 prepubertal (SRR8479244, SRR8479249, SRR8479250) and pubertal (SRR8479242, SRR8479248,

662 SRR8479247) testes from the *Sequence Read Archive* and filtered the reads for quality using *Fastp*
663 and default parameters [68].We created an alignment reference with *STAR* [69] and the parameter –
664 *runMode genomeGenerate*, using the Atlantic salmon reference genome and annotation files
665 downloaded from *NCBI* [70]. We aligned and quantified the quality-filtered reads using *STAR* in
666 manual two-pass mode and with the following parameters: *--outFilterIntronMotifs*
667 *RemoveNoncanonicalUnannotated –chimSegmentMin 10 –outFilterType BySJout –*
668 *alignSJDBoverhangMin 1 –alignIntronMin 20 –alignIntronMax 1000000 –alignMatesGapMax*
669 *1000000 –quantMode GeneCounts –alignEndsProtrude 10 ConcordantPair --limitOutSJcollapsed*
670 *5000000*. We combined the read counts from *STAR* into a single table using a custom *R* script and
671 normalized the data using *DESeq2 varianceStabilizingNormalization* [71] and independently with
672 *Clust* [72] (automatic normalization method: 101 3 4), so that expression patterns of genes could be
673 compared independently from their total expression level. Last, we used custom scripts in *R* to
674 analyze the co-expressed genes and plot the results. *NCBI RefSeq* annotations for genes were fetched
675 using *annotationhub* package and Gene Ontology over-representation in co-expression groups was
676 tested using *clusterProfiler* package [73]. Reads from isoform variants were quantified using
677 *ggsashimi* [74]
678
679 We quantified *vgll3* allele-specific expression in the RNA-seq data by comparing reads overlapping
680 SNP positions Ssa25: 28655795, Ssa25:28656101 and Ssa25:28658151 (the two latter being same
681 SNPs used in our crossing scheme and allele-specific TaqMan assays). We changed the SNP
682 nucleotide to "N" in order to remove reference-biased mapping and re-aligned the RNA-seq reads
683 using *STAR,* excluding multimapping reads and reads with more than one mis-match by defining
684 parameters *-outFilterMultimapNmax 1 -outFilterMismatchNmax 1* (in addition to the parameters
685 defined above). We then counted the number of overlapping *vgll3*L* and *vgll3*E* reads by manual
686 inspection of RNA-seq read tracks using *IGV* [75] and tested the significance of allele-specific
687 expression to the null expectation of 1:1 expression with binomial test.
688

Acknowledgements:

We thank members of Primmer lab for assistance with the common garden experiment, Lars Paulin for assistance in RACE and Sanger sequencing, and LUKE staff for assistance with the initial crosses. Financial support was provided by the University of Helsinki, Academy of Finland (grant number 314254) and the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement number 742312).

Author contributions:

- JPV and CRP designed the time-series sampling and expression experiments.
- JPV implemented and supervised the time-series sampling and expression analyses, designed and optimized the method, analyzed the data and wrote the initial draft.
- PVD and CRP designed the experiments validating *vgll3* association with male maturation probability.
- PVD implemented and supervised validation of *vgll3* association with male maturation probability and developed the associated statistical method.
- AR performed supporting lab analyses and optimized the genotyping method.
- NPi, AR, OO, JMV, ST and NPa assisted with fish rearing, tagging and sampling.
- TA designed the genotyping method.
- JE provided initial material for crosses.
- CRP conceived and supervised the project and acquired the funding.
- JPV and CRP wrote the final manuscript with input from PVD and JMV.
- All authors approved the manuscript.

Data availability:

Data will be made available on Dryad database following manuscript acceptance.

Code availability:

Code will be made available on GitHub following manuscript acceptance.

Supplementary figures.

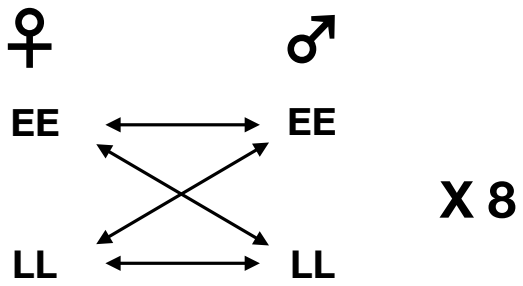


Figure S1. 16 females with 16 males were crossed in a 2x2 factorial design, producing 32 families with *vgll3* genotypes *EE, EL, LE* and *LL* (order denotes maternal and paternal origin of alleles).

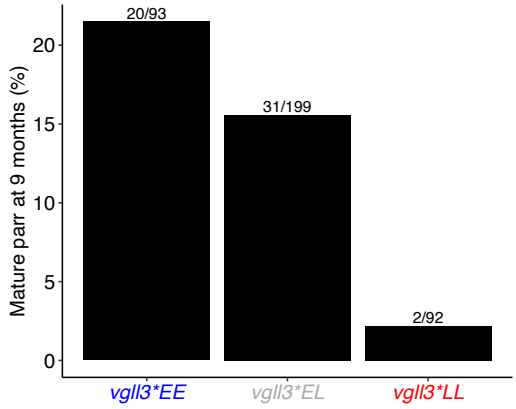


Figure S2. Maturation frequencies at nine months post-hatching. Number above bars indicates the number of mature individuals over all individuals.

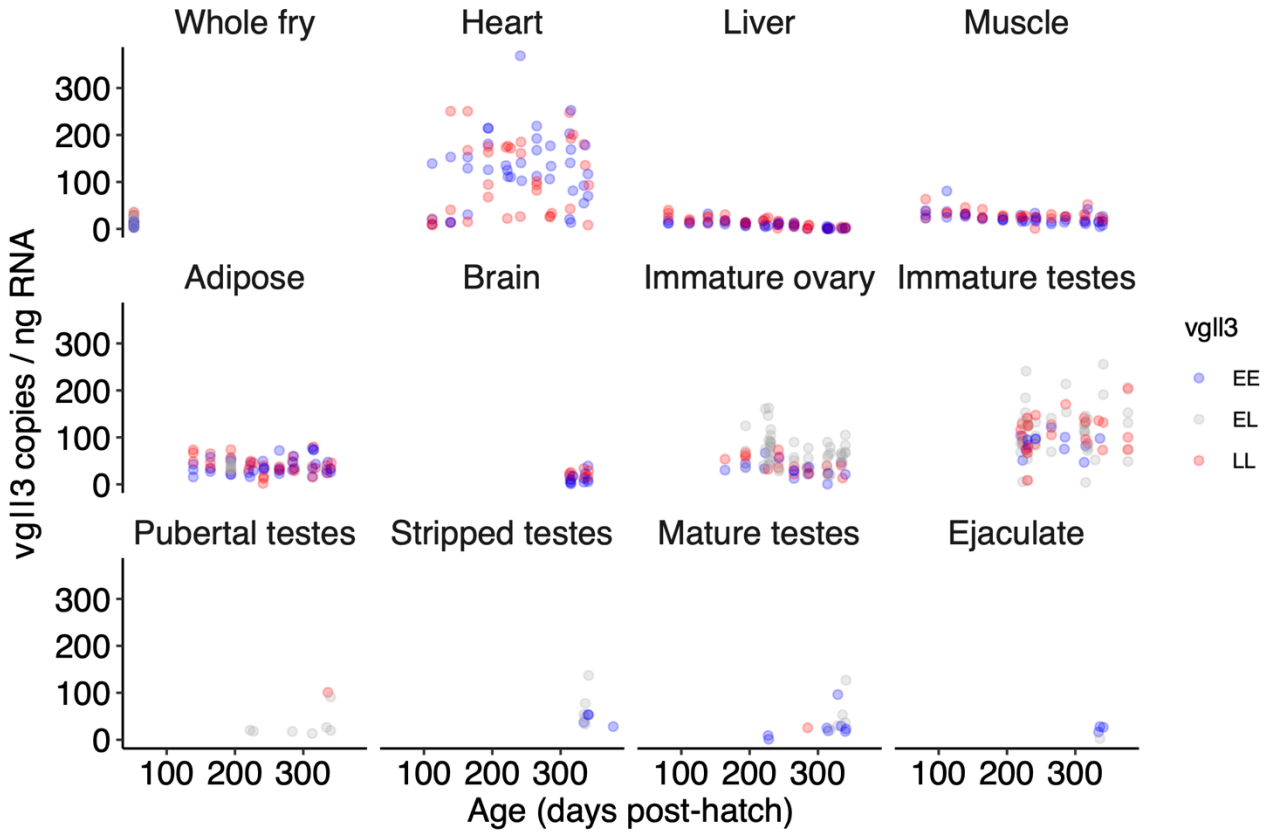


Figure S3. *Vgll3* expression in all investigated tissues and whole fry. Stripped testes; mature testes with ejaculate manually removed.

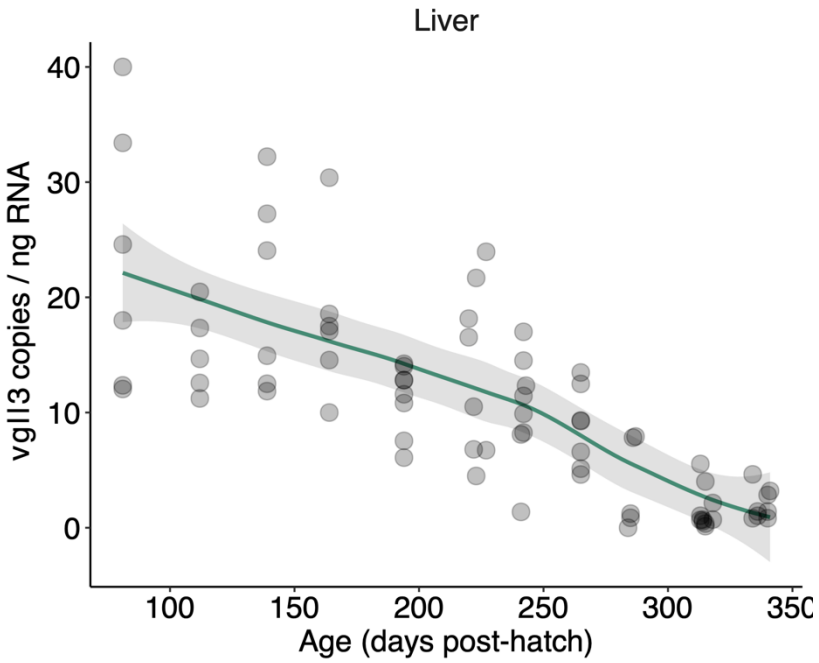


Figure S4. *Vgll3* expression in liver shows a linear downward trend, but the overall expression level is low. Green line, loess smoothing; grey shading, 95% confidence interval.

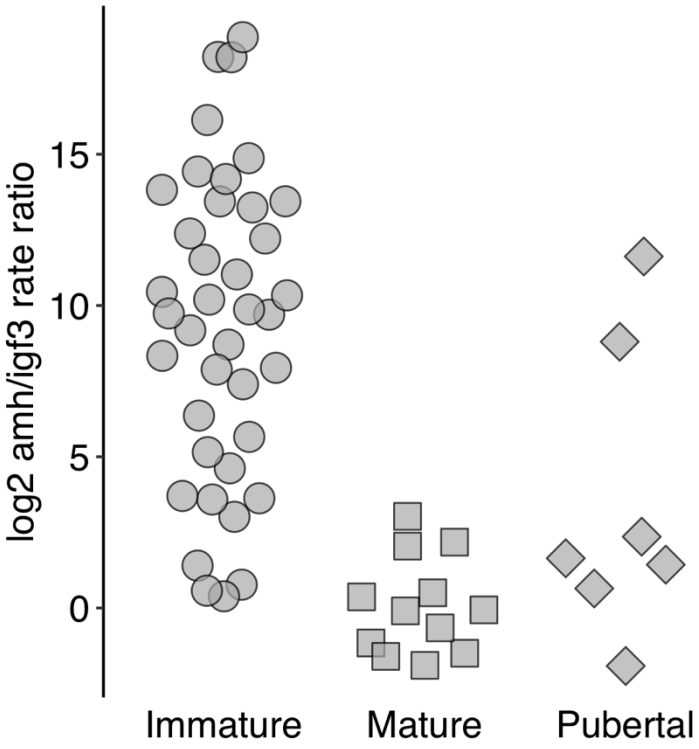


Figure S5. Expression of puberty indicator genes in male testes. Average *amh/igf3* expression ratio is higher in immature testes and maturation coincides with relative down-regulation of *amh* in relation to *igf3*. Pubertal samples show large variation in *amh/igf3* ratio. Testes are separated into categories based on morphology (see methods).

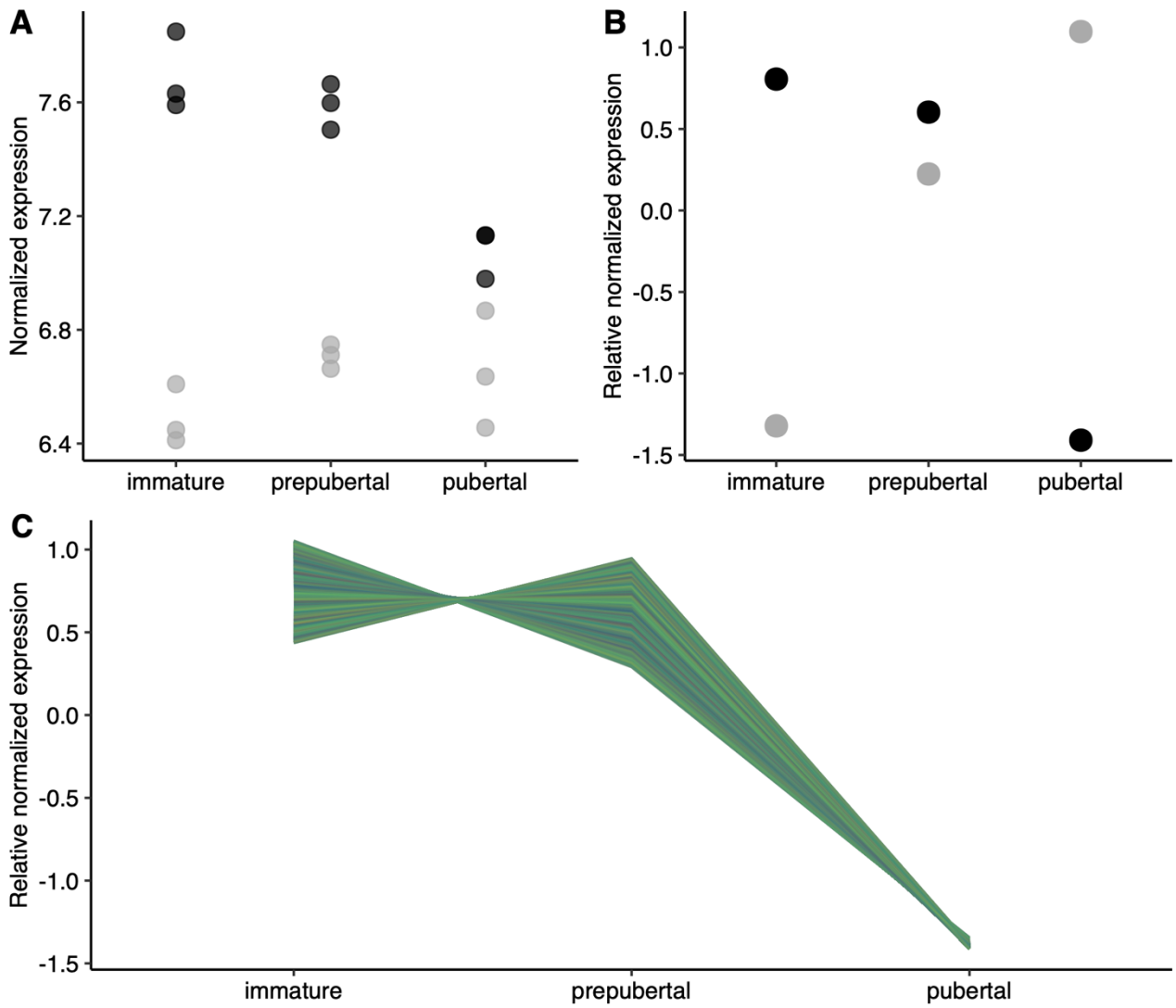
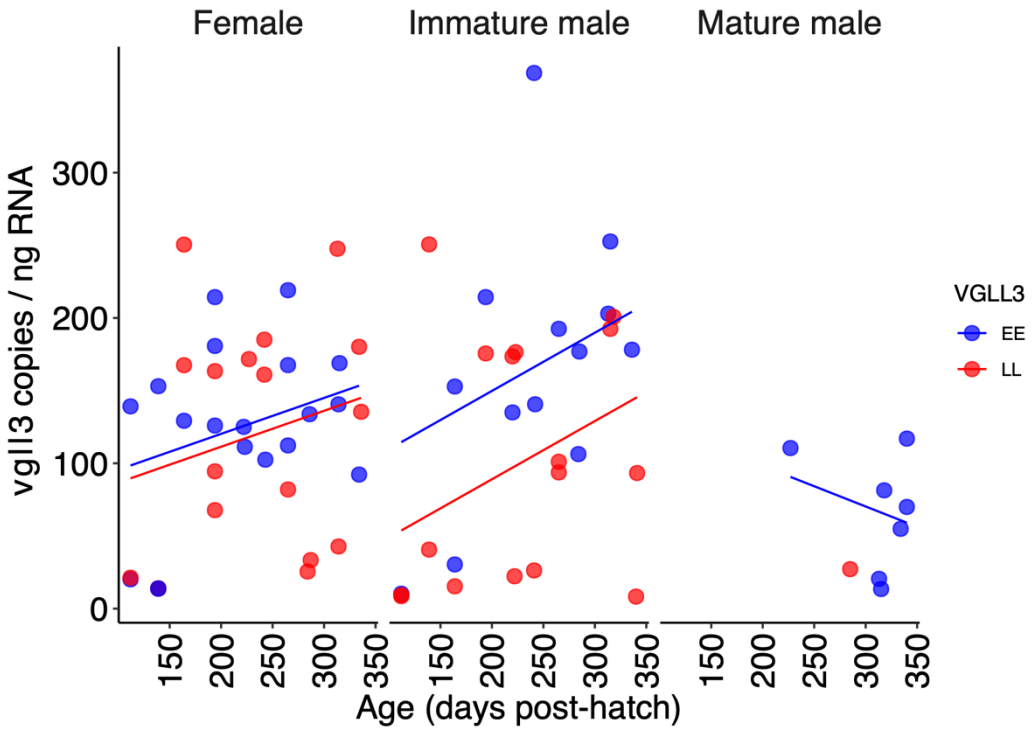


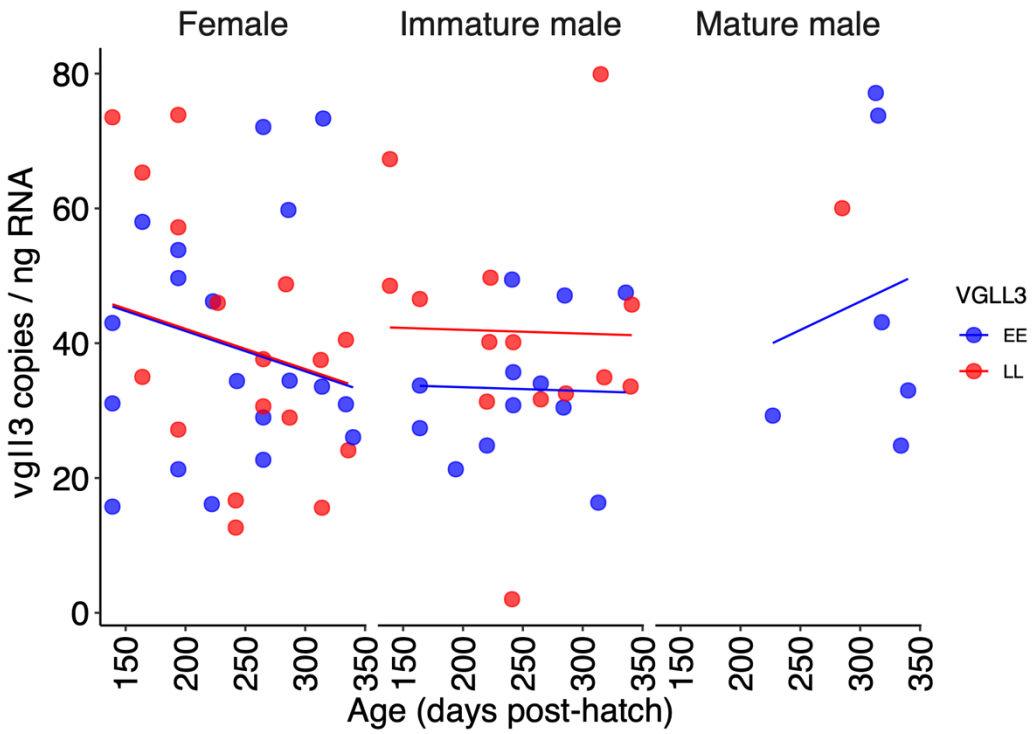
Figure S6. Re-analysis of data from ref [34] Expression patterns of *vgll3* and *yap1* in immature, prepubertal and pubertal testes. (A) Points represent normalized expression levels (variance stabilizing normalization [71]). (B) Points represent the relative normalized expression level means [72] from three biological replicates in each case. (C) Relative expression patterns of 7074 genes in *vgll3* co-expression cluster. Black points; *vgll3*, grey points; *yap1*.



751

752 **Figure S7.** Expression of *vgll3* in the heart. All trendlines non-significant at *P*-value threshold 0.05.

753



754

755 **Figure S8.** Expression of *vgll3* in adipose tissue. All trendlines non-significant at *P*-value threshold

756 0.05.

757

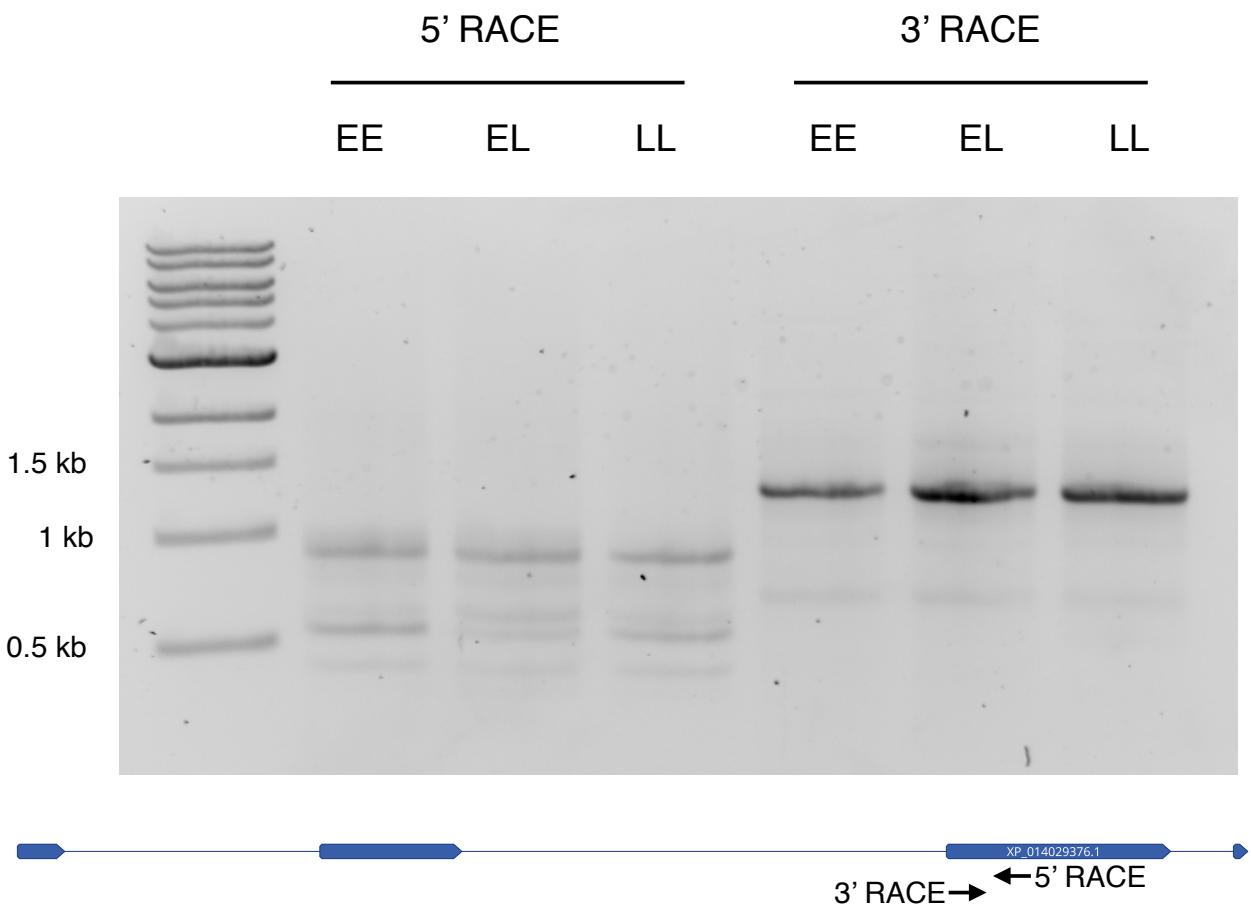


Figure S9. Agarose gel of 5' and 3' RACE PCR reactions. Both 5' and 3' RACE reactions amplify multiple RACE fragments of different lengths. RACE was performed using primers aligning to exon 3 of the *vgll3* transcript.

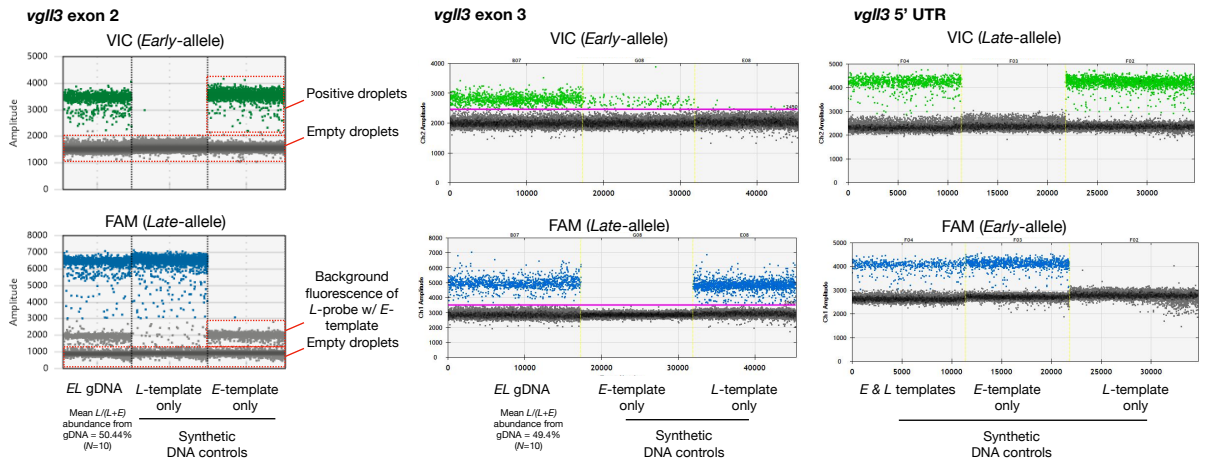


Figure S10. Genomic DNA and synthetic DNA controls of allele-specific *vgll3* assays.

768 **Supplemental tables:**

769

770 **Table S1.** TaqMan assays used in the study.

Vgll3 5' UTR Forward Primer	*Vgll3* 5' UTR Reverse Primer	*Vgll3* 5' UTR *Late* probe	*Vgll3* 5' UTR *Early* probe
GTCGTCATACTTTCACCAGAACACA	TGTAAACACCCAGCTTCTTCGAA	AAAAGTTGTAGGACGTTTC	AAAAGTTGTAGAACGTTTC

771

Vgll3 Exon 2 Forward Primer	*Vgll3* Exon 2 Reverse Primer	*Vgll3* Exon 2 *Early* probe	*Vgll3* Exon 2 *Late* probe
GAAGCTGGGTGTTTACAGTAGGAT	TGTCTCCGCCCTGGAAAC	TGCTGCTCCATGCTGT	TGCTGCTCCGTGCTGT

772

Vgll3 Exon 3 Forward Primer	*Vgll3* Exon 3 Reverse Primer	*Vgll3* Exon 3 *Early* probe	*Vgll3* Exon 3 *Late* probe
AGCCCAGGGATACACAGTGA	GTGGGCCAGGCTGAGG	CCACCTCTGTGTTCACA	CCACCTCTGTCTTCACA

773

Amh Forward Primer	*Amh* Reverse Primer	*Amh* probe
ATGAAAGAGTGACAGAAAAGCTGAGT	TGCAGCTCACACAGAAACGT	CCCTCCCCTGCTCCC

774

Igf3 Forward Primer	*Igf3* Reverse Primer	*Igf3* probe
CAACAAACTGAGATGAACAGTCTGAAC	CAGTCGCACAGCCAAGAATAAAA	TCGGCACAAAATATCT

775

776

References:

1. Stearns SC. The evolution of life histories. Oxford University Press; 1992.

2. Stearns SC. Life history evolution: successes, limitations, and prospects. Die Naturwissenschaften. 2000;87: 476–486. doi:10.1007/s001140050763

3. Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, Pemberton JM, et al. Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature. 2013;502: 93–95. doi:10.1038/nature12489

4. Barson NJ, Aykanat T, Hindar K, Baranski M, Bolstad GH, Fiske P, et al. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature. 2015;528: 405–408. doi:10.1038/nature16062

5. Ayllon F, Kjaerner-Semb E, Furmanek T, Wennevik V, Solberg MF, Dahle G, et al. The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (Salmo salar L.) males. PLoS Genetics. 2015;11. doi:10.1371/journal.pgen.1005628

6. Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, et al. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Science advances. 2017;3: e1603198. doi:10.1126/sciadv.1603198

7. Narum SR, Genova AD, Micheletti SJ, Maass A. Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. Proc Biol Sci. 2018;285. doi:10.1098/rspb.2018.0935

8. Hess JE, Zendt JS, Matala AR, Narum SR. Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. Proc Biol Sci. 2016;283: 20153064. doi:10.1098/rspb.2015.3064

9. Troth A, Puzey JR, Kim RS, Willis JH, Kelly JK. Selective trade-offs maintain alleles underpinning complex trait variation in plants. Science. 2018;361: 475–478. doi:10.1126/science.aat5760

10. Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoeppner MP, et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nat Genet. 2016;48: 84–88. doi:10.1038/ng.3430

11. Heyland A, Flatt T. Mechanisms of Life History Evolution : The Genetics and Physiology of Life History Traits and Trade-Offs. OUP Oxford; 2011.

12. Fleming IA, Einum S. Reproductive Ecology: A Tale of Two Sexes. Wiley-Blackwell; 2010. pp. 33–65. doi:10.1002/9781444327755.ch2

13. Mobley KB, Granroth-Wilding H, Ellmen M, Vaha J-P, Aykanat T, Johnston SE, et al. Home ground advantage: Local Atlantic salmon have higher reproductive fitness than dispersers in the wild. Science advances. 2019;5: eaav1112. doi:10.1126/sciadv.aav1112

813 14. Fleming IA. Pattern and variability in the breeding system of Atlantic salmon (Salmo salar),
814 with comparisons to other salmonids. Canadian Journal of Fisheries and Aquatic Sciences. 1998;55:
815 59–76.

816 15. Czorlich Y, Aykanat T, Erkinaro J, Orell P, Primmer CR. Rapid sex-specific evolution of age at
817 maturity is shaped by genetic architecture in Atlantic salmon. Nature Ecology & Evolution. 2018;2:
818 1800–1807. doi:10.1038/s41559-018-0681-5

819 16. Cousminer DL, Berry DJ, Timpson NJ, Ang W, Thiering E, Byrne EM, et al. Genome-wide
820 association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal
821 timing and childhood adiposity. Human Molecular Genetics. 2013;22: 2735–2747.
822 doi:10.1093/hmg/ddt104

823 17. Perry JRB, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-specific
824 allelic associations among 106 genomic loci for age at menarche. Nature. 2014;514: 92-+.
825 doi:10.1038/nature13545

826 18. Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, et al. Sex-
827 dependent dominance maintains migration supergene in rainbow trout. Nature Ecology &
828 Evolution. 2019;69: 1–12. doi:10.1038/s41559-019-1044-6

829 19. Lamichhaney S, Fuentes-Pardo AP, Rafati N, Ryman N, McCracken GR, Bourne C, et al.
830 Parallel adaptive evolution of geographically distant herring populations on both sides of the North
831 Atlantic Ocean. Proceedings of the National Academy of Sciences. 2017;54: 201617728-E3461.
832 doi:10.1073/pnas.1617728114

833 20. Bateman JR, Johnson JE, Locke MN. Comparing enhancer action in cis and in trans. Genetics.
834 2012;191: 1143–1155. doi:10.1534/genetics.112.140954

835 21. Tian K, Henderson RE, Parker R, Brown A, Johnson JE, Bateman JR. Two modes of
836 transvection at the eyes absent gene of Drosophila demonstrate plasticity in transcriptional
837 regulatory interactions in cis and in trans. Bosco G, editor. PLoS Genetics. 2019;15: e1008152.
838 doi:10.1371/journal.pgen.1008152

839 22. Reyes A, Huber W. Alternative start and termination sites of transcription drive most transcript
840 isoform differences across human tissues. Nucleic Acids Res. 2018;46: 582–592.
841 doi:10.1093/nar/gkx1165

842 23. Wang X, Hou J, Quedenau C, Chen W. Pervasive isoform-specific translational regulation via
843 alternative transcription start sites in mammals. Mol Syst Biol. 2016;12: 875.
844 doi:10.15252/msb.20166941

845 24. Wiestner A, Schlemper RJ, Maas A van der, Skoda RC. An activating splice donor mutation in
846 the thrombopoietin gene causes hereditary thrombocythaemia. Nat Genet. 1998;18: 49–52.
847 doi:10.1038/ng0198-49

848 25. Halperin DS, Pan C, Lusis AJ, Tontonoz P. Vestigial-like 3 is an inhibitor of adipocyte
849 differentiation. Journal of Lipid Research. 2013;54: 473–481. doi:10.1194/jlr.m032755

850 26. Figeac N, Mohamed AD, Sun C, Schönfelder M, Matallanas D, Garcia-Munoz A, et al. VGLL3
851 operates via TEAD1, TEAD3 and TEAD4 to influence myogenesis in skeletal muscle. Journal of
852 cell science. 2019;132: jcs225946. doi:10.1242/jcs.225946

853 27. Kjaerner-Semb E, Ayllon F, Kleppe L, Sørhus E, Skaftnesmo K, Furmanek T, et al. Vgll3 and
854 the Hippo pathway are regulated in Sertoli cells upon entry and during puberty in Atlantic salmon
855 testis. Scientific Reports. 2018;8: 1912. doi:10.1038/s41598-018-20308-1

856 28. Kurko J, Debes PV, House A, Aykanat T, Erkinaro J, Primmer CR. Transcription Profiles of
857 Age-at-Maturity-Associated Genes Suggest Cell Fate Commitment Regulation as a Key Factor in
858 the Atlantic Salmon Maturation Process. G3 (Bethesda). 2019; g3.400882.2019.
859 doi:10.1534/g3.119.400882

860 29. Taranger GL, Carrillo M, Schulz RW, Fontaine P, Zanuy S, Felip A, et al. Control of puberty in
861 farmed fish. General and Comparative Endocrinology. 2010;165: 483–515.
862 doi:10.1016/j.ygcen.2009.05.004

863 30. Debes PV, Piavchenko N, Ruokolainen A, Ovaskainen O, Moustakas-Verho JE, Parre N, et al.
864 Large single-locus effects for maturation timing are mediated via condition variation in Atlantic
865 salmon. bioRxiv. 2019;11: 780437. doi:10.1101/780437

866 31. Schulz RW, Franca LR, Lareyre J-J, LeGac F, Chiarini-Garcia H, Nobrega RH, et al.
867 Spermatogenesis in fish. General and Comparative Endocrinology. 2010;165: 390–411.
868 doi:10.1016/j.ygcen.2009.02.013

869 32. Pfennig F, Standke A, Gutzeit HO. The role of Amh signaling in teleost fish--Multiple functions
870 not restricted to the gonads. General and Comparative Endocrinology. 2015;223: 87–107.
871 doi:10.1016/j.ygcen.2015.09.025

872 33. Morais R, Crespo D, Nobrega RH, Lemos MS, Kant HJG van de, Franca LR, et al. Antagonistic
873 regulation of spermatogonial differentiation in zebrafish (Danio rerio) by Igf3 and Amh. Molecular
874 and Cellular Endocrinology. 2017;454: 112–124. doi:10.1016/j.mce.2017.06.017

875 34. Skaftnesmo KO, Edvardsen RB, Furmanek T, Crespo D, Andersson E, Kleppe L, et al.
876 Integrative testis transcriptome analysis reveals differentially expressed miRNAs and their mRNA
877 targets during early puberty in Atlantic salmon. BMC Genomics. 2017;18: 801.
878 doi:10.1186/s12864-017-4205-5

879 35. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, classification and density
880 estimation ssing Gaussian finite mixture models. The R journal. 2016;8: 289–317.
881 doi:10.1186/s12942-015-0017-5

882 36. Meng Z, Moroishi T, Guan K-L. Mechanisms of Hippo pathway regulation. Genes Dev.
883 2016;30: 1–17. doi:10.1101/gad.274027.115

884 37. Simon E, Faucheux C, Zider A, Theze N, Thiebaud P. From vestigial to vestigial-like: the
885 Drosophila gene that has taken wing. Development Genes and Evolution. 2016;226: 297–315.
886 doi:10.1007/s00427-016-0546-3

887 38. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and
888 transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protocols.
889 2012;7: 562–578. doi:10.1038/nprot.2012.016

890 39. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation.
891 Nature. 2004;430: 85–88.

892 40. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, et al. Translation from
893 unconventional 5′ start sites drives tumour initiation. Nature. 2017;541: 494–499.

894 41. (DGT) TFC and the RP and C. A promoter-level mammalian expression atlas. Nature.
895 2014;507: 462–470. doi:10.1038/nature13182

896 42. Burtis KC, Baker BS. Drosophila doublesex gene controls somatic sexual differentiation by
897 producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell. 1989;56:
898 997–1010. doi:10.1016/0092-8674(89)90633-8

899 43. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The
900 Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science. 2012;338: 1587–
901 1593. doi:10.1126/science.1230612

902 44. Gao Q, Sun W, Ballegeer M, Libert C, Chen W. Predominant contribution of cis-regulatory
903 divergence in the evolution of mouse alternative splicing. Mol Syst Biol. 2015;11: 816.
904 doi:10.15252/msb.20145970

905 45. Smale ST, Kadonaga JT. The RNA polymerase II core promoter. Annu Rev Biochem. 2003;72:
906 449–479.

907 46. Carroll SB. Evo-devo and an expanding evolutionary synthesis: A genetic theory of
908 morphological evolution. Cell. 2008;134: 25–36. doi:10.1016/j.cell.2008.06.030

909 47. Schwartz C, Balasubramanian S, Warthmann N, Michael TP, Lempe J, Sureshkumar S, et al.
910 Cis-regulatory Changes at FLOWERING LOCUS T Mediate Natural Variation in Flowering
911 Responses of Arabidopsis thaliana. Genetics. 2009;183: 723–732. doi:10.1534/genetics.109.104984

912 48. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution
913 of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327:
914 302–305.

915 49. Frankel N, Erezyilmaz DF, McGregor AP, Wang S, Payre F, Stern DL. Morphological
916 evolution caused by many subtle-effect substitutions in regulatory DNA. Nature. 2011;474: 598–
917 603. doi:10.1038/nature10200

918 50. O'Brown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory
919 change underlying altered expression and Wnt response of the stickleback armor plates gene EDA.
920 Elife. 2015;4. doi:10.7554/elife.05290

921 51. Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, et al. Evolving
922 new skeletal traits by cis-regulatory changes in mone morphogenetic proteins. Cell. 2016;164: 45–
923 56. doi:10.1016/j.cell.2015.12.007

924 52. Gautier M, Yamaguchi J, Foucaud J, Loiseau A, Ausset A, Facon B, et al. The genomic basis of
925 color pattern polymorphism in the harlequin ladybird. Current Biology. 2018;28: 1–7.
926 doi:10.1016/j.cub.2018.08.023

927 53. Wang J, Liu S, Heallen T, Martin JF. The Hippo pathway in the heart: pivotal roles in
928 development, disease, and regeneration. Nature Reviews Cardiology. 2018;15: 672–684.
929 doi:10.1038/s41569-018-0063-3

930 54. Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? Evolution.
931 2008;62: 2155–2177. doi:10.1111/j.1558-5646.2008.00450.x

932 55. Price T, Schluter D. ON THE LOW HERITABILITY OF LIFE-HISTORY TRAITS. Evolution.
933 1991;45: 853–861. doi:10.1111/j.1558-5646.1991.tb04354.x

934 56. Erkinaro J, Laine A, Maki-Petays A, Karjalainen TP, Laajala E, Hirvonen A, et al. Restoring
935 migratory salmonid populations in regulated rivers in the northernmost Baltic Sea area, Northern
936 Finland - biological, technical and social challenges. Journal of Applied Ichthyology. 2011;27: 45–
937 52. doi:10.1111/j.1439-0426.2011.01851.x

938 57. Aykanat T, Lindqvist M, Pritchard VL, Primmer CR. From population genomics to
939 conservation and management: a workflow for targeted analysis of markers identified using
940 genome-wide approaches in Atlantic salmon Salmo salar. Journal of fish biology. 2016;89: 2658–
941 2679. doi:10.1111/jfb.13149

942 58. Anderson EC. Computational algorithms and user-friendly software for parentage-based tagging
943 of Pacific salmonids. 2010.

944 59. Sorensen DA, Genetics SA, 1995. Bayesian inference in threshold models using Gibbs
945 sampling. Genetics Selection Evolution. 1995;27: 229–249.

946 60. Hadfield JD. Increasing the efficiency of MCMC for hierarchical phylogenetic models of
947 categorical traits using reduced mixed models. O'Hara RB, editor. Methods in Ecology and
948 Evolution. 2015;6: 706–714. doi:10.1111/2041-210x.12354

949 61. Team RC. R: A Language and Environment for Statistical Computing. 2013;

950 62. Hadfield J. MCMC methods for multi-response generalized linear mixed models: the
951 MCMCglmm R package. Journal of Statistical Software. 2010;33: 1–22.

952 63. Henderson C. Sire evaluation and genetic trends. Journal of Animal Ecology. 1973; 10–43.
953 doi:10.1093/ansci/1973.symposium.10

954 64. Villemereuil P de, Gimenez O, Doligez B. Comparing parent–offspring regression with
955 frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation
956 study for Gaussian and binary traits. Freckleton R, editor. Methods in Ecology and Evolution.
957 2013;4: 260–275. doi:10.1111/2041-210x.12011

958 65. Heidelberger P, Welch PD. A spectral method for confidence interval generation and run length
959 control in simulations. Communications of the ACM. 1981;24: 233–245.
960 doi:10.1145/358598.358630

961 66. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, et al. Absolute
962 quantification by droplet digital PCR versus analog real-time PCR. Nat Meth. 2013;10: 1003–1005.
963 doi:10.1038/nmeth.2633

964 67. Laitinen RAE, Immanen J, Auvinen P, Rudd S, Alatalo E, Paulin L, et al. Analysis of the floral
965 transcriptome uncovers new regulators of organ determination and gene families related to flower
966 organ differentiation in Gerbera hybrida (Asteraceae). Genome Res. 2005;15: 475–486.
967 doi:10.1101/gr.3043705

968 68. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
969 Bioinformatics. 2018;34: i884–i890. doi:10.1093/bioinformatics/bty560

970 69. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
971 universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635

972 70. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic salmon genome
973 provides insights into rediploidization. Nature. 2016;533: 200–205. doi:10.1038/nature17164

974 71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
975 seq data with DESeq2. Genome biology. 2014;15: 550. doi:10.1186/s13059-014-0550-8

976 72. Abu-Jamous B, Kelly S. Clust: automatic extraction of optimal co-expressed gene clusters from
977 gene expression data. Genome biology. 2018;19: 172. doi:10.1186/s13059-018-1536-8

978 73. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological
979 themes among gene clusters. Omics : a journal of integrative biology. 2012;16: 284–287.
980 doi:10.1089/omi.2011.0118

981 74. Garrido-Martín D, Palumbo E, Guigó R, Breschi A. ggsashimi: Sashimi plot revised for
982 browser- and annotation-independent splicing visualization. Plos Comput Biol. 2018;14: e1006360.
983 doi:10.1371/journal.pcbi.1006360

984 75. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative
985 genomics viewer. Nature biotechnology. 2011;29: 24–26. doi:10.1038/nbt.1754

986