Regulatory divergence in *vgll3* underlies variation in age at maturity in male Atlantic salmon

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]

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, *cis*-regulation, functional validation, puberty timing

Abstract

The optimal phenotypes to survive and reproduce in given evolutionary settings are often in conflict, which manifest as trade-offs in life history traits. Genome-wide association studies have been successful in uncovering genome regions controlling for life history trade-offs in a range of species [1-8]. However, lack of functional studies of the discovered genotype-phenotype associations restrains our understanding how alternative life history traits evolved and are mediated at the molecular level. Here we establish a molecular mechanism by which allele-specific expression differences of the transcription co-factor *vgll3* mediate variation in age at maturity in male Atlantic salmon (*Salmo salar*). By way of a common-garden experiment and temporal sampling of *vgll3* expression, we show that *vgll3* genotype and expression control male puberty timing in one-year old salmon. *Vgll3* expression was high in immature testicular tissue and males entering puberty reduced their *vgll3* expression by 66% compared to immature males. Testicular expression of the *early* maturation associated allele in immature males was constitutively lower compared to the *late* maturation allele, shifting the liability of *early* allele carriers towards earlier maturation. These results reveal how gene regulatory differences can be a central mechanism for the evolution of life history variation.

Main text

Survival and reproduction often require trade-offs because size and age at maturity are at odds with the risk of dying before reproduction. The molecular basis of variation in life history traits and such trade-offs have thus far remained elusive. In Atlantic salmon, age at maturity is a key life history trait that is evolving under a trade-off of survival and fecundity [2,9-11]. 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 across 57 Northern-European salmon populations [2]. The functional basis of this genotype-phenotype association remains unknown. Seeing that the strongest association in natural populations maps to a single nucleotide polymorphism (SNP) in a non-coding region adjacent to *vgll3* [2], we hypothesized that genetic variation in linked *cis*-regulatory sequences (e.g. enhancers) underlie the functional differences among salmon *vgll3* genotypes by changing their expression level or pattern.

To uncover the functional genetic basis of age at maturity variation in Atlantic salmon, we first validated that the genotype at the *vgll3* locus on chromosome 25 (termed "*E*" for early sea-age at maturity and "*L*" for late sea-age at maturity after previous association study [2]) is associated with age at maturity in controlled conditions by rearing 656 individuals with known *vgll3* genotypes from 32 families in common garden conditions (Extended Data Fig. 1). *Vgll3*EE* males had a significantly higher probability of maturing during their first year compared to males with *vgll3*LL* genotypes (probit 95% credible interval: 0.52 to 3.42, P=0.007), which translates to 16% vs 0% difference in mean maturation probability, respectively (Fig. 1). Males assessed for maturation status at the end of the experiment (twelve months post-hatch, *vgll3*EE* n=93, *vgll3*EL* n=199 and *vgll3*LL* n=92) revealed a 13 times higher frequency of maturity in individuals with *vgll3*EE* genotypes (39%) compared to individuals with *vgll3*LL* genotypes (3%, Extended Data Fig. 2). Maturation probability for males heterozygous for *vgll3* was intermediate between the homozygotes, indicating an additive effect of *vgll3* genotype on maturity timing in our experimental population. Results were unchanged when we modeled maturation probability with length of the fish as a covariate (Supplemental

analysis). These results demonstrate that the *vgll3*E* allele associated with young sea-age at maturity in natural populations[2] is also associated with earlier maturation also in controlled conditions. Seeing that concordant results were observed in another wild-derived stock of Atlantic salmon [12], the genotype-phenotype association between *vgll3* and maturation timing seems to be reproducible in controlled conditions and common in North-European strains of Atlantic salmon.

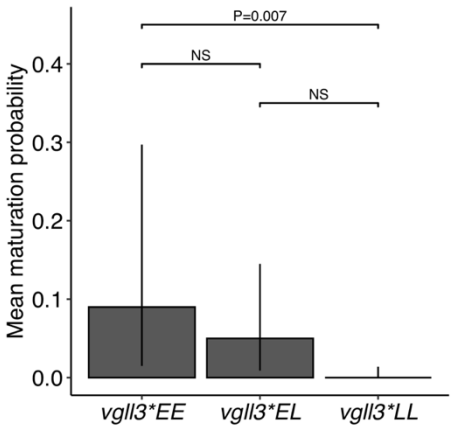


Figure 1. *Vgll3* genotype associates with male maturation probability in controlled conditions. Probability of maturation within the first year in controlled conditions is significantly higher in *vgll3*EE* individuals compared to *vgll3*LL* individuals. Whiskers represent 95% probit credible intervals. NS; non-significant.

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 (total n=291, Extended Data Fig. 3). *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 [13]. Interestingly, while *vgll3* expression was high in immature testes, expression was reduced by 65% in mature testes compared to immature testes (-70 copies/ng [-93 -47], two-tailed t-test, df=25.9, P=1e-06, n_{mature}=14, $n_{immature}$=65, Fig. 2b). *Vgll3* expression was similarly lower in stripped mature testes, containing mainly mitotic cell types, and in ejaculate, containing mainly mature sperm, compared to immature testes (-72 copies/ng [-111 -30], two-tailed t-test, df=21.8, P=0.002, $n_{stripped}$=8, -115 copies/ng [-181 -50], df=6.6, P=0.005, $n_{ejaculate}$=4, respectively, Fig. 2b). Taken together, these results indicate that among salmon tissues, *vgll3* expression is high in immature testes and down-regulated upon the onset of male puberty.

To further understand how *vgll3* expression changes in testes according to activation of the male puberty pathway, we compared *vgll3* expression relative to the expression of *amh,* which is a repressor of male puberty and Sertoli cell differentiation, and *igf3*, which promotes spermatogonial germ cell differentiation [14-17]. Clustering based on *amh/igf3* expression ratio and *vgll3* expression level revealed that mature testes were separated in gene expression space from immature testes (bootstrap LRT [18] P<0.0001) and characterized by both low *amh/igf3* ratio and low *vgll3* expression, indicating that down-regulation of both *amh* and *vgll3* expression is commonly associated with completing puberty (Fig. 2c). In sum, these results point towards a mechanism by which male puberty is associated with a down-regulation of *vgll3* expression, corroborating previous studies that suggested *vgll3* influences Sertoli cell differentiation [19].

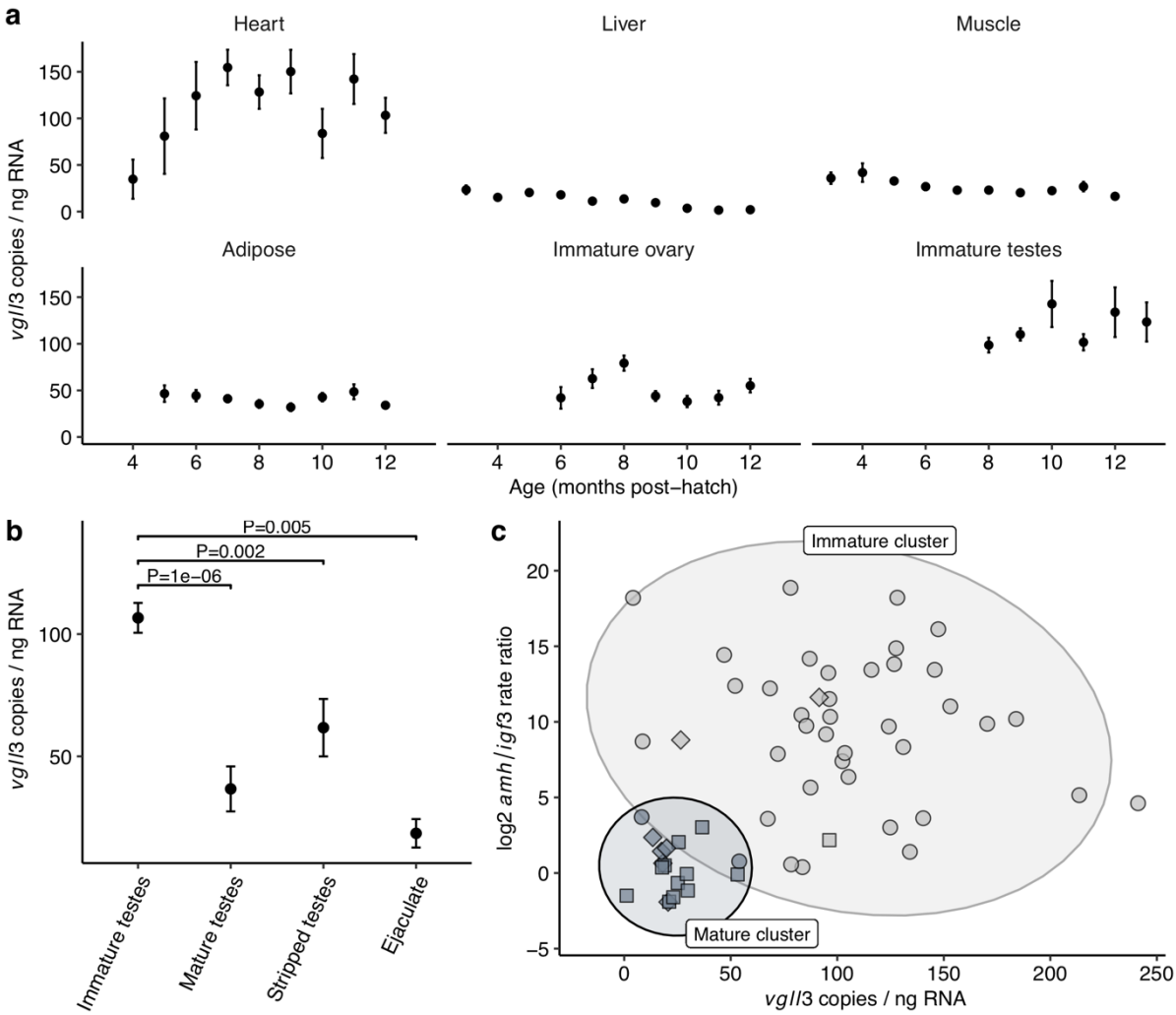


Figure 2. *Vgll3* expression across salmon tissues during the first year of development. (**a**) *Vgll3* reaches highest expression levels in the heart and immature testes. For sample size in each point see Extended Data Fig. 3. (**b**) Male maturation is associated with a 66% drop in *vgll3* expression compared to immature testes expression levels. Reduction in *vgll3* expression is observed in all major tissues of the mature testes. Stripped testes; mature testes with ejaculate removed manually. Error bars show SE. (**c**) *Vgll3* expression shows a strong association with relative expression levels of *amh/igf3*. Squares; mature testes, diamonds; pubertal samples in transition to mature morphology, spheres; immature testes. Ellipses delineate estimated 95% confidence intervals for each cluster.

These results led us to hypothesize that the observed higher incidence of maturation in *vgll3*EE* males may be linked to overall lower *vgll3* expression in immature testes that have not yet entered puberty. We tested this hypothesis by comparing the testicular expression from immature individuals carrying different *vgll3* genotypes. *Vgll3*EE* genotypes showed 20% lower expression compared to *vgll3*LL* genotypes (-22 copies/ng [-41 -3], two-sided *t*-test, df=33.8, P=0.024, n_{EE}=13 n_{LL}=23, Fig. 3). This is remarkable, since a 20% reduction in *vgll3* expression in immature *vgll3*EE* individuals accounts for roughly one third of the reduction in testicular *vgll3* expression associated with the transition to fully mature testes. Heterozygotes showed intermediate average expression levels that were statistically different from *vgll3*EE* but not *vgll3*LL* homozygotes (27 copies/ng [6 47], two-sided *t*-test, df=50.3, P=0.012 n_{EL}=42, Fig. 3). Using a regression analysis we found that, overall, *vgll3* genotype explained 8.4% of variation in *vgll3* expression in immature testes between *vgll3*EE*

and *vgll3*LL* homozygotes (P_{vgll3}=0.048, Extended Data Fig. 4), while being stable across age (P_{age}=0.38). The *vgll3*EE* genotype associated with lower average expression compared to *vgll3*LL* genotypes also in female gonads, although the difference was not statistically significant (Extended Data Fig. 5). The other two tissues with highest *vgll3* expression levels, heart and adipose, did not show a clear dependency of *vgll3* expression level on genotype (Extended Data Fig. 6 & Extended Data Fig. 7). Given the results above indicating that *vgll3* down-regulation is required for maturation, lower expression in *vgll3*EE* individuals provides a plausible mechanism for the higher maturation probability in *vgll3*EE* males.

We next tested for *cis*-regulatory differences between the *vgll3*E* and *vgll3*L* alleles by analyzing *vgll3* allele-specific expression using allele-specific probes within the same reverse transcription droplet digital PCR (RT-ddPCR) reactions. 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 [20]. In immature male testes, 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. 3b). Thirty-three of the 47 heterozygotes analyzed (70%) showed significantly higher *vgll3*L* allele expression at *P*-value threshold of 0.05, while only one individual (0.02%) showed significantly higher *vgll3*E* allele expression at the same *P*-value threshold (Poisson rate-ratio test, Fig. 3b and Extended Data Fig. 8). In terms of expressed mRNA molecules, the average magnitude of the *cis*-regulatory differences observed between *vgll3*E* and *vgll3*L* alleles was large enough to account for 62% of the average expression differences between *vgll3*EE* and *vgll3*LL* homozygotes. These results identify *cis*-regulation as a major source of expression differences between *vgll3* alleles in immature testes, which in turn associate with differences in male maturation probability.

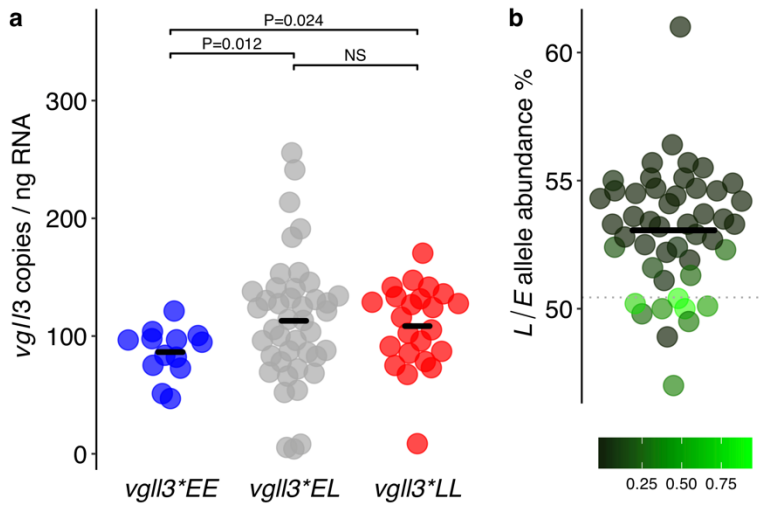


Figure 3. *Vgll3* allele-specific expression differences in immature testes. (a) *Vgll3*EE* individuals show 20% lower average expression compared to *vgll3*LL* individuals. Bars; mean. NS; non-significant. **(b)** Allele-specific expression of *vgll3* in heterozygotes shows that on average 53% of *vgll3* mRNA molecules are expressed from *late* alleles (black bar). Null hypothesis for testing of 1:1 allele-specific expression ratio is calculated from heterozygote DNA (dashed line). Color scale corresponds to low (black) and high (green) allele-specific expression *P*-value.

The functional basis of variation in life history traits and their trade-offs have thus far remained elusive. Genetic variation influencing *cis*-regulatory loci is known to be an important source of phenotypic diversity and a common source of evolutionary change [e.g. 21-27], and is a strong candidate

for molecular mechanisms underlying variation in life histories. Altogether, the data presented here support a threshold model whereby *vgll3* expression levels contribute to the liability of initiating maturation below an unknown threshold (Fig. 4). Further, our findings suggest that *cis*-regulatory differences cause individuals carrying the *vgll3*E* allele to reach maturation with higher probability by quantitatively lowering the overall level of *vgll3* co-factor being expressed, and thus shifting the liability towards early maturation. The transcription co-factor function of *vgll3* suggests that such changes in *vgll3* expression transmit downstream to expression changes in gene regulatory networks associated with cell proliferation and organ development [28]. Like its vertebrate and invertebrate homologs, the salmon *vgll3* has been proposed to control cell differentiation through interacting with the TEAD-family of transcription factors [13,29,30]. Vgll3, like related co-factors in humans and *Drosophila*, is predicted to compete for TEAD accessibility with alternative co-factors downstream of the Hippo signaling pathway controlling for organ size (Yap/Taz co-factors) [28,31]. Consistent with this, by re-analyzing RNA-sequencing data from immature, prepubertal and pubertal testes [15], we found that *vgll3* expression pattern correlates strongly and negatively with *yap1* expression pattern (*rho*=-0.825, Extended Data Fig. 9). Salmon *vgll3* expression was also found to negatively correlate with *yap1* expression in early development [13]. Quantitative changes in competing co-factor abundances is a known mechanism for controlling cell differentiation [32]. Therefore, we hypothesize that constitutive changes in *vgll3* co-factor abundance associated with the different expression levels of the *vgll3*E* and *vgll3*L* alleles influence the tendency of cell differentiation in immature testes by changing the stoichiometry of Vgll3-Yap competition for TEADs and further the balance between self-renewal (inhibition of differentiation) and differentiation.

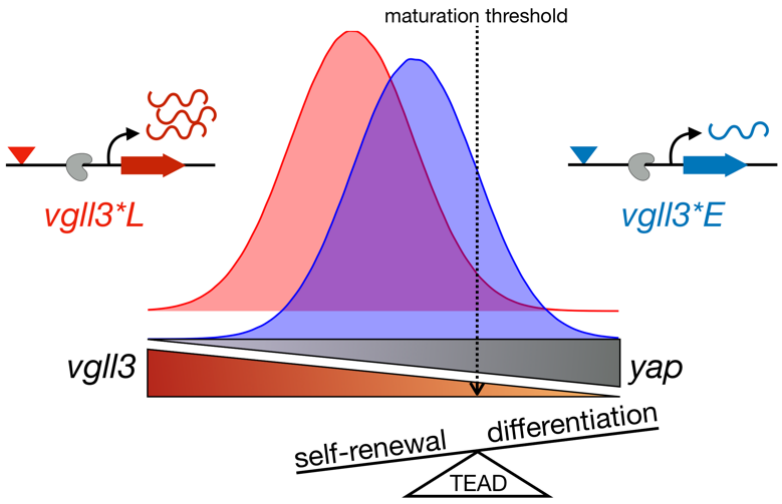


Figure 4. Model for a molecular mechanism of *vgll3* genotype-dependent liability in puberty timing. Hypothetical distributions represent the distributions of *vgll3/yap* relative abundances in individuals with given *vgll3* genotypes. The association between *vgll3* genotype and maturation timing arises from *cis*-regulatory differences influencing *vgll3* expression. The *vgll3*E* allele tends to be expressed at a lower level, thus changing the stoichiometry of competition between *vgll3* and *yap*. TEAD interaction with either *vgll3* or *yap* tips the balance between self-renewal (*vgll3*) and differentiation (*yap*).

Here we have revealed key functional insights into the temporal expression dynamics and molecular mechanisms that translate genetic variation in a major effect locus to variation in life history strategies in Atlantic salmon. Our results suggest that gene regulatory differences may be central to evolution of life history strategies.

Acknowledgements
We thank members of Primmer lab for assistance with the common garden experiment and LUKE staff for assistance with the initial crosses.

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 assisted with designing 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.

Materials and Methods

Crosses used in the study

We controlled-crossed eight 2x2 factorial matings of unrelated individuals that each included one *vgll3*EE* and one *vgll3*LL* male and female (see ref [12] for genotyping details), thereby resulting in 32 families with reciprocal *vgll3* genotypes *EE*, *EL*, *LE* and *LL* (allele order denotes female and male parents of origin respectively, Extended Data Fig. 1). We used parents from the "Oulujoki" broodstock, which is a mixture of several salmon populations from northern Baltic rivers and is maintained by the Natural Resources Institute Finland (LUKE)[33]. All parents were selected for having a homogenous age-at-maturity-associated haplotype encompassing the *vgll3* (chromosome 25 paralog) coding sequence, the top non-coding association [2], and the coding region of the adjacent *Akap11* gene (collectively referred to as "*vgll3* genotype").

Husbandry design and conditions

Following fertilization, familial egg batch replicates were distributed randomly in two egg incubators. Rearing took place in four 0.25 m^3 tanks, with each tank including, on average, six individuals per family. Light cycle and water temperature (min=6.3, max=17.7 °C) during husbandry corresponded to the local seasonal cycle at 61°N. Fish were fed first with live *Artemia* for three days, after which we fed commercial aquaculture feed *ad libitum* (fine-ground Raisio Baltic Blend). Temperature, oxygen and nitrogen-component levels were monitored on a regular basis. Experimentation was conducted under animal experimentation license ESAVI/2778/2018.

Individual tagging and morphological measurements

Before first feeding and transfer to tanks (two months post-hatching) randomly selected 24 individuals per family and inserted visible elastomer (VIE) tags at the base of the caudal fin to identify individuals according to their *vgll3* genotype by color (initial *n*=1300). At eight months post-hatching and before the onset of the first breeding season we tagged remaining fish using passive integrated transponder (PIT) tags and collected fin-clips. We extracted DNA using standard chelex or salt extraction methods, genotyped all fish for 141 informative SNPs [34], assigned their genotypic sex based on genotyping coverage of the salmon sex determining locus *sdy* and assigned individuals to families using the likelihood method implemented in SNPPIT [35]. We recorded the length and mass of all fish monthly after PIT-tagging and assessed the maturity status of males by manually testing for the presence of running milt at 12 months post-hatching.

Statistical model of male maturation probability

We modeled the probability of males reaching maturity within their first year given their *vgll3* genotypes by using a GLMM (generalized linear mixed model) approach to implement a genetic threshold model that takes full account of the relatedness among individuals [36,37]. We estimated maturation probability using a generalized linear animal model with probit-link function under Bayesian Markov Chain Monte Carlo simulations with the R-software [38] version 3.6.1 and package MCMCglmm[39] for the following model: $y = vgll3 + length + a + t + e$, where y is a vector of maturation binaries, *length* is mean-centered and variance scaled natural logarithm of length (cm), *vgll3* is the genotype, *a* is the additive genetic relationship-matrix predicted animal effect [40], *t* is the common environmental tank effect and *e* the error term (with fixed variance of 1). We used priors following a χ_1^2 distribution[41] and based inferences on 10,000 samples collected every 1,000 iteration after burn-in of 10,000. Model convergence and adequate sampling was ensured by visual trace plot

inspection, autocorrelation < 0.1 for lag-2 parameter samples, and using Heidelberger and Welch's convergence diagnostic [42].

Temporal sampling and RNA expression

We sampled individuals with *vgll3*EE, vgll3*EL, vgll3*LE* and *vgll3*LL* monthly (3-32 in total per genotype per time point) and assessed sex, gonad size and maturity by dissection. We sacrificed the animals by MS-222 overdose and dissected ten tissues (brain (timepoints 12-13 months post hatch); heart, liver, muscle (timepoints 3-13 months post hatch), adipose (timepoints 5-13 months post hatch); immature ovary (timepoints 6-13 months post hatch); immature testes (timepoints 7-13 months post hatch); mature testes (timepoints 8 and 10-12 months post hatch); stripped mature testes (timepoints 12-13 months post hatch) and ejaculate (timepoint 13 months post hatch)). We classified each male fish visually as immature (not enlarged, transparent color gonads), pubertal (slightly enlarged, transparent or white color gonads) or mature (enlarged white color gonads). We flash froze the samples on liquid nitrogen and stored the samples in -80 °C. We homogenized the tissues using an OMNI Bead Ruptor Elite machine, in 2 ml or 7 ml (for mature gonads) with 2.4 mm steel beads. We used Macherey-Nagel NucleoSpin RNA kits in single and in 96-well plate formats to extract total RNA and treated the samples with additional DNase (Invitrogen Turbo DNase) to remove any residual DNA. We verified total RNA quality using Agilent BioAnalyzer for a random set of 24 samples and found all samples to have RIN>9. We measured total RNA yield using ThermoFischer Qubit and Quant-it reagents, diluted the samples to working concentrations and re-measured RNA concentrations. For samples in Figs. 2 and 3 (all gonad samples) we repeated the RNA concentration measurement once more and calculated the mean of the duplicate measurements, which we used in data normalization (for details see Supplemental file 1). We used ~20 ng (2-8 months post-hatch) or ~50 ng (8-14 months post-hatch) of total RNA as template in one-step reverse transcription droplet digital PCR (RT-ddPCR) [43] (BioRad) with TaqMan probes specific to *vgll3* alleles to quantify *vgll3* absolute expression. Measurement of *amh* and *igf3* expression was performed as above, with the exception of using ~2 ng of total RNA template and TaqMan probes that were labeled with alternative fluorophores and assayed in the same RT-ddPCR reactions. We verified the specificity of all used TaqMan assays with genomic and synthetic DNA templates and optimized melting temperatures using a gradient.

Expression and allele-specific expression analyses

We analyzed RT-ddPCR results using BioRad QuantaSoft software and imported the results tables to R. We used custom R scripts to (i) discard samples with less than 10,000 droplets that passed QuantaSoft quality filters, (ii) normalize the number of *vgll3* mRNA molecules to the amount of input RNA and (iii) analyze the data and plot the results (for full analysis pipeline see Supplemental file 1).

For investigation of the effect of age/season and *vgll3* genotype to *vgll3* expression in immature testes and immature ovaries, we used the following regression model: $y = vgll3 + x + e$, where y is *vgll3* total mRNA copies, x is days post-hatch, *vgll3* is the genotype and e the error term. We tested the effect of *vgll3* and x (days post-hatch) individually using a corresponding model.

We characterized the relative expression levels of the key male maturity genes *amh* and *igf3* by way of Poisson modeling of expression rate-ratio for each sample. The number of positive droplets with

expression signal in RT-ddPCR data follows a Poisson distribution with lambda being the rate of positives droplets over all droplets (https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf). We calculated the Poisson rates of expression from *amh* and *igf3* -positive droplets over the total amount of analyzed droplets (the rate ratio of *amh/igf3* expression) using the R function *poisson.test*.

We tested for presence of distinct clusters of gene expression patterns on *amh/igf3* and *vgll3* expression space using Gaussian mixture modeling as implemented in the R package *Mclust* version 5 [18,44]. We chose the best fitting mixture of Gaussian distributions based on Bayesian Information Criterion using the function *mclustBIC* and further tested the best fit mixture model for the presence of more than one distributions using a bootstrapping likelihood-ratio test as implemented in the function *mclustBootstrapLRT* with 1000 bootstrap replications.

Testing of *vgll3* allele-specific expression in heterozygotes was performed in a similar manner as above for *amh/igf3* expression (test for Poisson rate ratio), with the exception that signal from positive droplets now represented expression from either the *vgll3*E* or the *vgll3*L* allele. We defined the expression rate of *vgll3*E* allele as the number of droplets positive for the *vgll3*E* probes over all probes that passes the QuantaSoft standard quality filter (including empty droplets), and correspondingly for the *vgll3*L* allele. We then tested for difference in Poisson rates for each individual heterozygote using the R function *poisson.test* and *P*-value threshold of 0.05. We calculated the null expectation for equal expression rates based on RT-ddPCR with replicates of *vgll3* heterozygous DNA as template.

We re-analyzed RNA-sequencing data from immature, prepubertal and pubertal Atlantic salmon testis samples [15] in order to investigate the relative expression patterns of *vgll3* and *yap1* genes. We downloaded raw RNA-seq data for immature (SRR8479243, SRR8479245, SRR8479246), prepubertal (SRR8479244, SRR8479249, SRR8479250) and pubertal (SRR8479242, SRR8479248, SRR8479247) testes from the *Sequence Read Archive* and filtered the reads for quality using *Fastp* and default parameters [45]. We created an alignment reference with *STAR* [46] and the parameter *--runMode genomeGenerate*, using the Atlantic salmon reference genome and annotation files downloaded from *SalmoBase* [47,48]. We aligned and quantified the quality-filtered reads using *STAR* in manual two-pass mode and with the following parameters: *--outFilterIntronMotifs RemoveNoncanonicalUnannotated --chimSegmentMin 10 --outFilterType BySJout --alignSJDBoverhangMin 1 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --quantMode GeneCounts --alignEndsProtrude 10 ConcordantPair --limitOutSJcollapsed 5000000*. We combined the read counts from *STAR* into a single table using a custom *R* script and normalized the data using *Clust* [49] (automatic normalization method: 101 3 4), so that expression patterns of genes could be compared independently from their total expression level. Last, we used custom scripts in *R* to analyze the co-expressed genes and plot the results.

Supplemental material

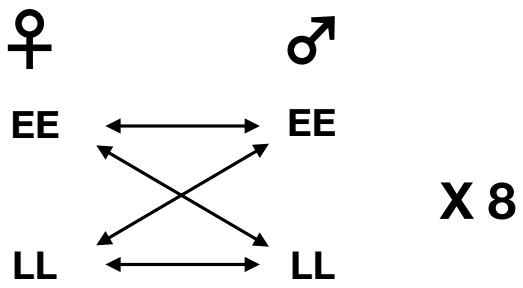
TaqMan assays

Vgll3 Forward Primer Seq.	*Vgll3* Reverse Primer Seq.	*Vgll3* Reporter 1 Sequence	*Vgll3* Reporter 2 Sequence
GAAGCTGGGTGTTTACAGTAGGAT	TGTCTCCGCCCTGGAAAC	TGCTGCTCCATGCTGT	TGCTGCTCCGTGCTGT

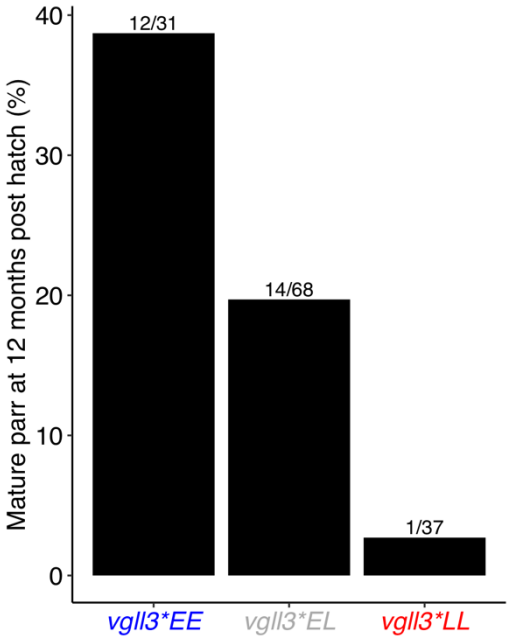
Amh Forward Primer Seq.	*Amh* Reverse Primer Seq.	Dye	*Amh* Reporter Sequence
ATGAAAGAGTGACAGAAAAGCTGAGT	TGCAGCTCACACAGAAACGT	FAM	CCCTCCCCTGCTCCC

Igf3 Forward Primer Seq.	*Igf3* Reverse Primer Seq.	Dye	*Igf3* Reporter Sequence
CAACAAACTGAGATGAACAGTCTGAAC	CAGTCGCACAGCCAAGAATAAAA	VIC	TCGGCACAAAATATCT

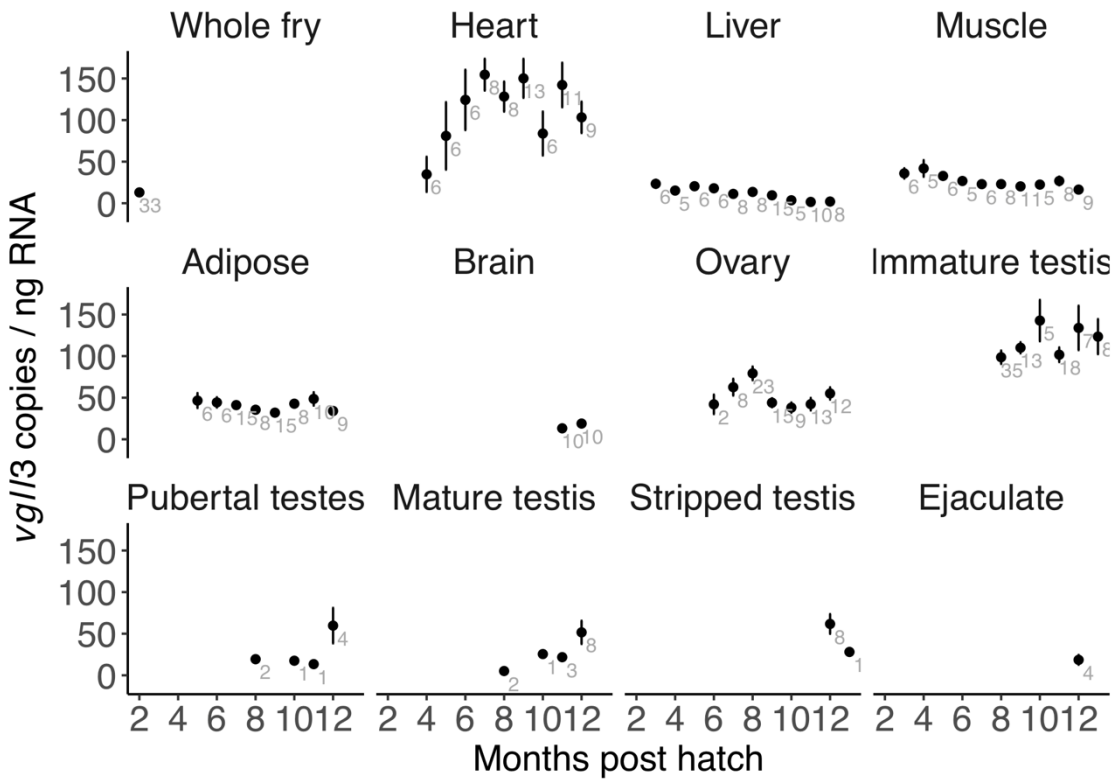
Extended Data Figures



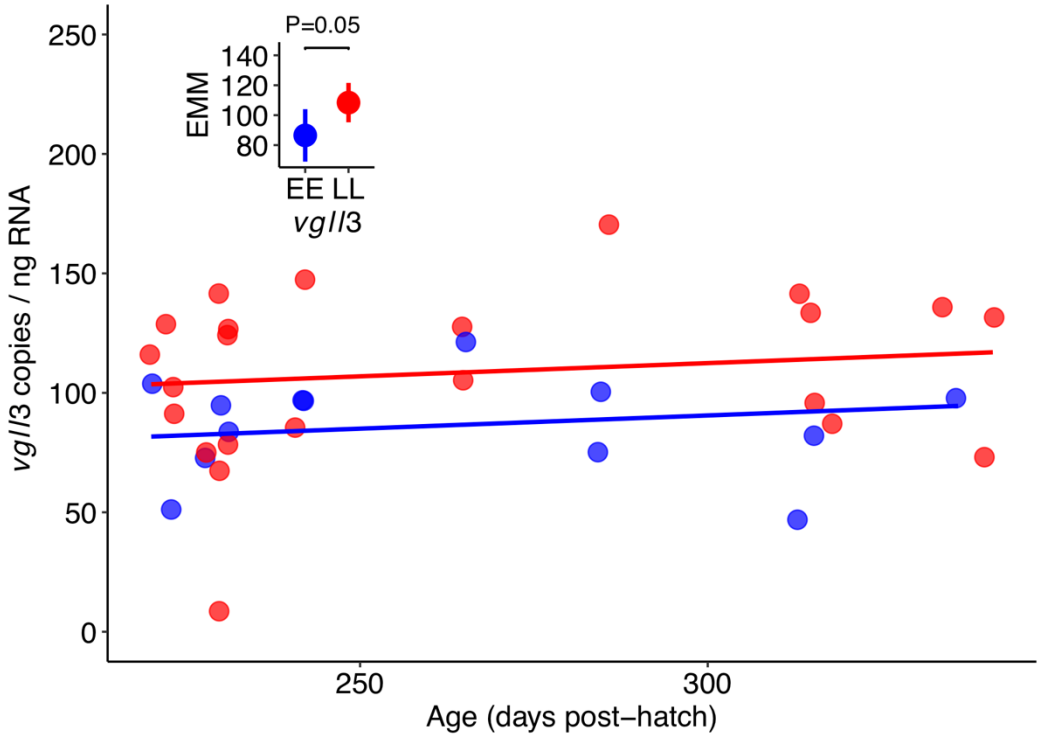
Extended Data Figure 1. 16 females with 16 males were crossed in a 2x2 factorial design, producing 32 families with *vgll3* genotypes *EE, EL, LE* and *LL* (order denotes maternal and paternal origin of alleles).



Extended Data Figure 2. Maturation frequencies at the end of the experiment (12 months post-hatch). Number above bars indicates the number of mature individuals over all individuals.

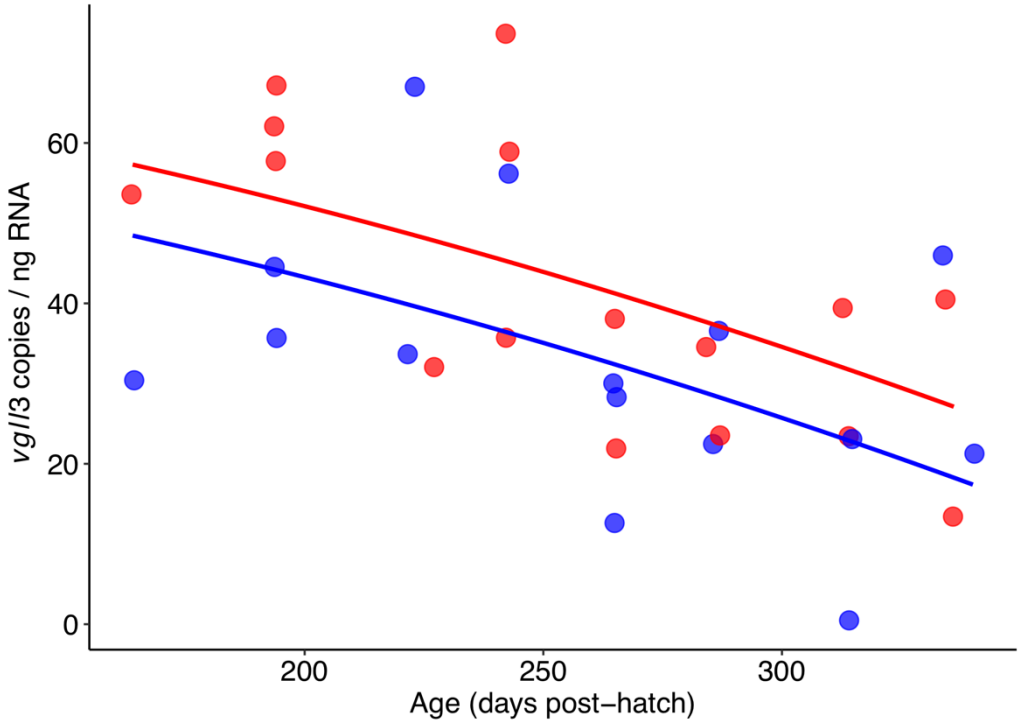


Extended Data Figure 3. *Vgll3* expression in all investigated tissues and whole fry. Stripped testes; mature testes with ejaculate manually removed. Whiskers; SE. Number in grey denotes sample size.

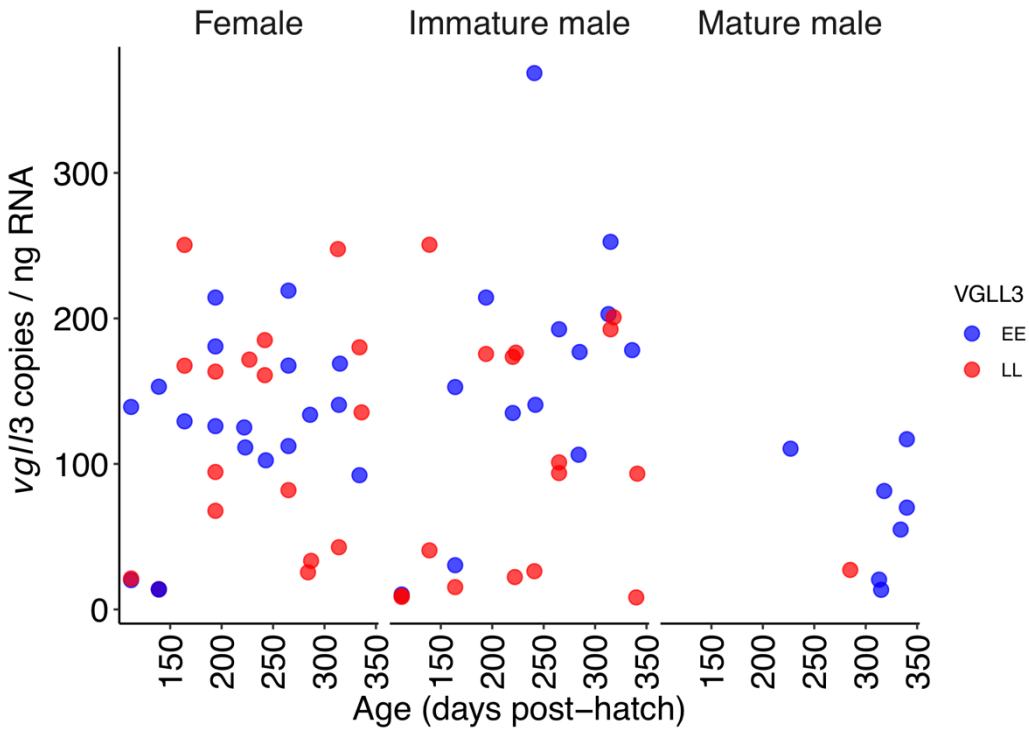


Extended Data Figure 4. Expression of *vgll3* in immature male testes across all sampled time points (blue, *vgll3**EE; red, *vgll3***LL*). Expression of *vgll3* is stable across fish age (significance of

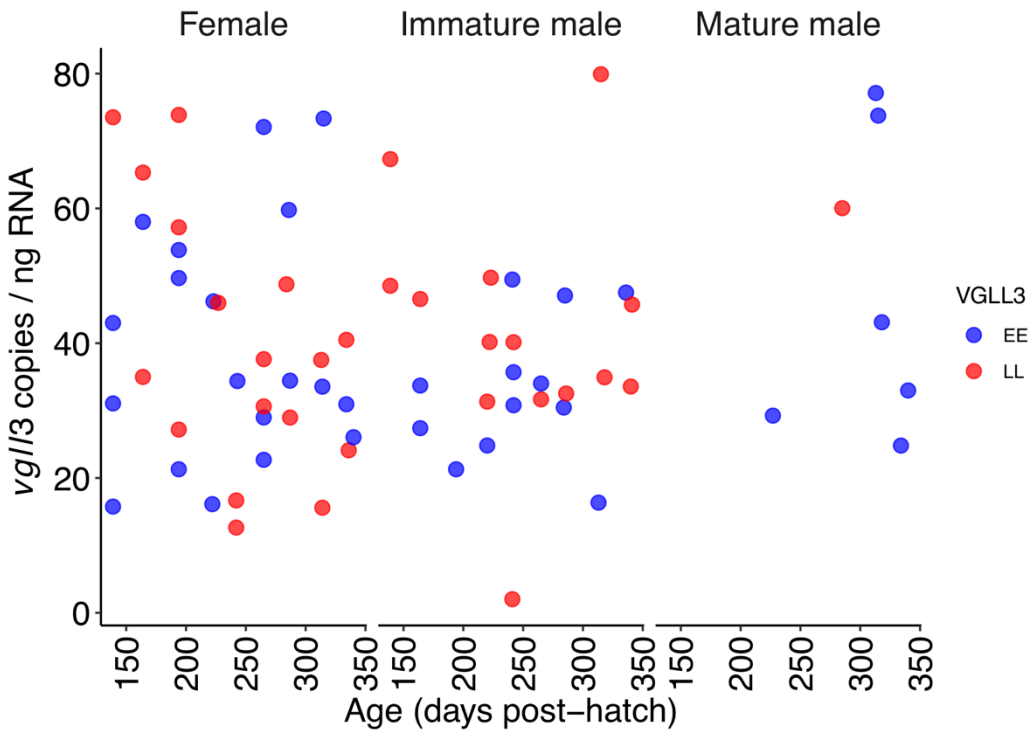
regression slope *P*=0.39). *Vgll3*LL* is more highly expressed compared to *vgll*EE* genotypes. EMM; estimated marginal mean.



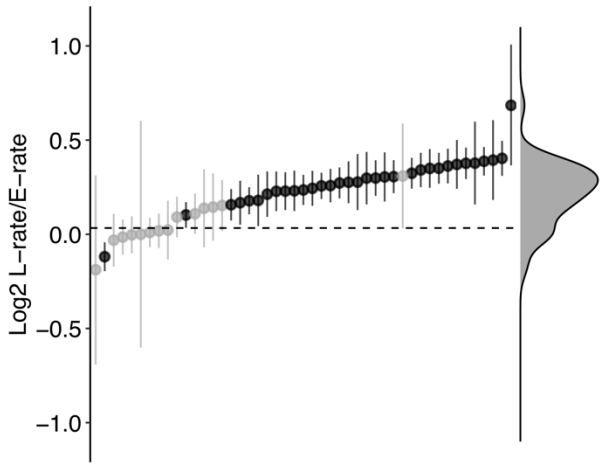
Extended Data Figure 5. Expression of *vgll3* in immature ovaries (blue, *vgll3*EE; red, *vgll3*LL*).



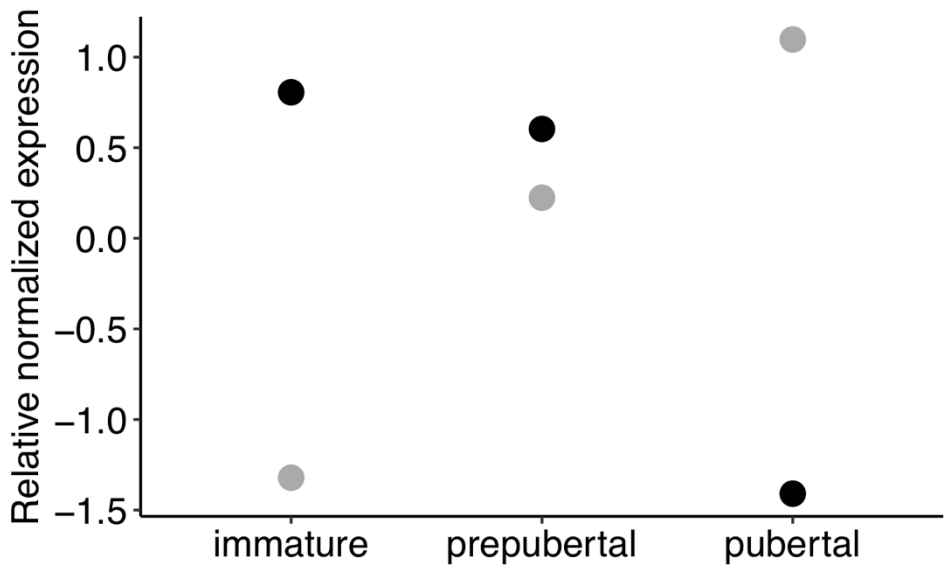
Extended Data Figure 6. Expression of *vgll3* in the heart.



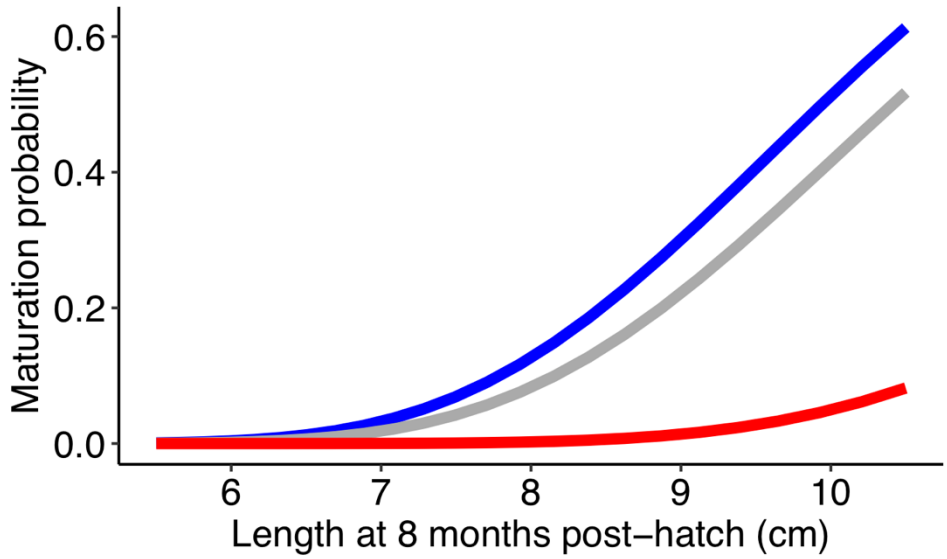
Extended Data Figure 7. Expression of *vgll3* in adipose tissue.



Extended Data Figure 8. Poisson modeling of ASE. Whiskers represent 95% confidence interval and are color coded whether or not they overlap with the null-hypothesis of 1:1 expression (dashed line). Null is calculated based on ddPCR data from replicate *vgll3*EL* genomic DNA.



Extended Data Figure 9. Expression patterns of *vgll3* and *yap1* in immature, prepubertal and pubertal testes ([re-analyzed data from 15]). Points represent the normalized relative expression level means [49] from three biological replicates in each case. Black points; *vgll3*, grey points; *yap1*.



Extended Data Figure 10. Association between *vgll3* genotype and maturation probability as a function of fish length before the breeding season (eight months of age).

Supplemental analysis

We additionally modeled maturation probability as a function of *vgll3* genotype and fish length, hypothesizing that early growth positively affects maturation probability [50]. Indeed, maturation probability positively associated with fish size before the start of the breeding season (eight months of age) (probit 95% credible interval: 0.12 to 0.95, P=0.011, Extended Data Fig. 10). Irrespective of size effect, the *vgll3*EE* genotype was still associated with a higher maturation probability compared to the *vgll3*LL* genotype (probit 95% credible interval: 0.54 to 3.35, P=0.007).

References

1. Johnston, S. E. *et al.* Life history trade-offs at a single locus maintain sexually selected genetic variation *Nature* **502,** 93-95 (2013).
2. Barson, N. J. *et al.* Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528,** 405–408 (2015).
3. Ayllon, F. *et al.* The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (Salmo salar L.) males. *PLoS Genetics* **11,** (2015).
4. Prince, D. J. *et al.* The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances* **3,** e1603198 (2017).
5. Narum, S. R., Di Genova, A., Micheletti, S. J. & Maass, A. Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proceedings of The Royal Society Biological Sciences* **285,** (2018).
6. Hess, J. E., Zendt, J. S., Matala, A. R. & Narum, S. R. Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of The Royal Society Biological Sciences* **283,** 20153064 (2016).
7. Troth, A., Puzey, J. R., Kim, R. S., Willis, J. H. & Kelly, J. K. Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science* **361,** 475–478 (2018).
8. Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature Genetics* **48,** 84–88 (2016).
9. Mobley, K. B. *et al.* Home ground advantage: Local Atlantic salmon have higher reproductive fitness than dispersers in the wild. *Science Advances* **5,** eaav1112 (2019).
10. Czorlich, Y., Aykanat, T., Erkinaro, J., Orell, P. & Primmer, C. R. Rapid sex-specific evolution of age at maturity is shaped by genetic architecture in Atlantic salmon. *Nature Ecology & Evolution* **2,** 1800–1807 (2018).
11. Fleming, I. A. Pattern and variability in the breeding system of Atlantic salmon (Salmo salar), with comparisons to other salmonids. *Canadian Journal of Fisheries and Aquatic Sciences* **55,** 59–76 (1998).
12. Debes, P. V. *et al.* Condition variation mediates maturation timing via a large effect locus in Atlantic salmon. *In preparation*
13. Kurko, J. *et al.* Transcription profiles of age-at-maturity-associated genes suggest cell fate commitment regulation as a key factor in the Atlantic salmon maturation process. *In preparation*.
14. Morais, R. D. V. S. *et al.* Antagonistic regulation of spermatogonial differentiation in zebrafish (Danio rerio) by Igf3 and Amh. *Molecular and Cellular Endocrinology* **454,** 112–124 (2017).
15. Skaftnesmo, K. O. *et al.* Integrative testis transcriptome analysis reveals differentially expressed miRNAs and their mRNA targets during early puberty in Atlantic salmon. *BMC Genomics* **18,** (2017).
16. Schulz, R. W. *et al.* Spermatogenesis in fish. *General and Comparative Endocrinology* **165,** 390–411 (2010).
17. Pfennig, F., Standke, A. & Gutzeit, H. O. The role of Amh signaling in teleost fish--Multiple functions not restricted to the gonads. *General and Comparative Endocrinology.* **223,** 87–107 (2015).
18. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, classification and density estimation ssing Gaussian finite mixture models. *R Journal* **8,** 289–317 (2016).
19. Kjaerner-Semb, E. *et al.* Vgll3 and the Hippo pathway are regulated in Sertoli cells upon entry and during puberty in Atlantic salmon testis. *Science Reports* **8,** 1912 (2018).
20. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430,** 85–88 (2004).

21. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134,** 25–36 (2008).
22. Schwartz, C. *et al.* Cis-regulatory changes at FLOWERING LOCUS T mediate natural variation in flowering responses of Arabidopsis thaliana. *Genetics* **183,** 723–732 (2009).
23. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327,** 302–305 (2010).
24. Frankel, N. *et al.* Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474,** 598–603 (2011).
25. O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D. & Kingsley, D. M. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* **4,** (2015).
26. Indjeian, V. B. *et al.* Evolving new skeletal traits by cis-regulatory changes in mone morphogenetic proteins. *Cell* **164,** 45–56 (2016).
27. Gautier, M. *et al.* The genomic basis of color pattern polymorphism in the harlequin ladybird. *Current Biology* **28,** 1–7 (2018).
28. Simon, E., Faucheux, C., Zider, A., Theze, N. & Thiebaud, P. From vestigial to vestigial-like: the Drosophila gene that has taken wing. *Development Genes and Evolution* **226,** 297–315 (2016).
29. Kjaerner-Semb, E. *et al.* Vgll3 and the Hippo pathway are regulated in Sertoli cells upon entry and during puberty in Atlantic salmon testis. *Science Reports* **8,** 1912 (2018).
30. Christensen, K. A., Gutierrez, A. P., Lubieniecki, K. P. & Davidson, W. S. TEAD3, implicated by association to grilsing in Atlantic salmon. *Aquaculture* **479,** 571–578 (2017).
31. Meng, Z., Moroishi, T., development, K. G. G.2016. Mechanisms of Hippo pathway regulation. *Genes and Development* **30,** 1–17 (2016).
32. Gualdrini, F. *et al.* SRF co-factors control the balance between cell proliferation and contractility. *Molecular Cell* **64,** 1048–1061 (2016).
33. Erkinaro, J. *et al.* Restoring migratory salmonid populations in regulated rivers in the northernmost Baltic Sea area, Northern Finland - biological, technical and social challenges. *Journal of Applied Ichthyology* **27,** 45–52 (2011).
34. Aykanat, T., Lindqvist, M., Pritchard, V. L. & Primmer, C. R. From population genomics to conservation and management: a workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon Salmo salar. *Journal of Fish Biology* **89,** 2658–2679 (2016).
35. Anderson, E. C. *Computational algorithms and user-friendly software for parentage-based tagging of Pacific salmonids.* Final report submitted to the Pacific Salmon Commission's Chinook Technical Committee. (2010).
36. Sorensen, D. A., Genetics, S. A.1995. Bayesian inference in threshold models using Gibbs sampling. *Genetics Selection Evolution* **27,** 229–249 (1995).
37. Hadfield, J. D. Increasing the efficiency of MCMC for hierarchical phylogenetic models of categorical traits using reduced mixed models. *Methods in Ecology and Evolution* **6,** 706–714 (2015).
38. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2013)
39. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* **33,** 1–22 (2010).
40. Henderson, C. Sire evaluation and genetic trends. *Journal of Animal Ecology* **1973,** 10–43 (1973).
41. de Villemereuil, P., Gimenez, O. & Doligez, B. Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution* **4,** 260–275 (2013).

42. Heidelberger, P. & Welch, P. D. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM* **24,** 233–245 (1981).
43. Hindson, C. M. *et al.* Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* **10,** 1003–1005 (2013).
44. Fraley, C. & Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97,** 611–631 (2011).
45. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34,** i884–i890 (2018).
46. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
47. Lien, S. *et al.* The Atlantic salmon genome provides insights into rediploidization. *Nature* **533,** 200–205 (2016).
48. Samy, J. K. A. *et al.* SalmoBase: an integrated molecular data resource for Salmonid species. *BMC Genomics* **18,** 482 (2017).
49. Abu-Jamous, B. & Kelly, S. Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biology* **19,** 172 (2018).
50. Taranger, G. L. *et al.* Control of puberty in farmed fish. *General and Comparative Endocrinology* **165,** 483–515 (2010).