PNAS RESEARCH ARTICLE | EVOLUTION

A complex mechanism translating variation of a simple genetic architecture into alternative life histories

Jukka-Pekka Verta[a,b,1], Jacqueline E. Moustakas-Verho[a,c], Iikki Donner[a], Morgane Frapin[a], Annukka Ruokolainen[a], Paul V. Debes[a,d], Jaakko Erkinaro[c], and Craig R. Primmer[a,e,1]

Affiliations are included on p. 10.

Edited by Leif Andersson, Uppsala Universitet, Uppsala, Sweden; received February 13, 2024; accepted October 9, 2024

Understanding the processes that link genotype to phenotype is a central challenge in biology. Despite progress in discovering genes associated with ecologically relevant traits, a poor understanding of the processes and functions via which molecules mediate evolutionary differences leaves us critically far from linking proximate and ultimate causes of evolution. This knowledge gap is particularly large in multifaceted phenotypes of ecological relevance such as life histories where multiple traits covary and influence fitness. In Atlantic salmon (*Salmo salar*), variation in a key life-history trait, maturation age, is largely linked to the transcription cofactor *vestigial-like 3* (*vgll3*). Here, we show that despite this simple genetic architecture, *vgll3* genotype influences maturation age through a complex regulatory mechanism whereby it controls the expression of diverse pubertal signaling pathways. Using a multiomic approach in salmon testes, we show that the *vgll3* genotype conferring early maturity up-regulates key genes controlling androgen production, cellular energy and adiposity, and TGF-β signaling, thereby increasing the likelihood of earlier pubertal initiation. By mapping VGLL3 regulatory elements we further show its interaction with distinct transcription factors in a genotype-dependent manner, thus coordinating differential activation of regulatory pathways. This study reveals the proximate mechanisms through which a genetically simple association leads to functionally complex molecular differences in a spectrum of cellular traits, thus explaining the molecular basis of pleiotropy in a large-effect gene. Our results indicate that evolution in correlated phenotypes, as exemplified by alternative life history strategies, can be dictated by the function of major life-history genes.

genotype–phenotype | life histories | gene regulation | pleiotropy | maturation

Significance

Understanding the molecular mechanisms through which genotype influences phenotype is crucial for solving how evolutionary differences arise at the molecular level. Excepting a few simple traits, this has remained out of reach for most ecologically relevant phenotypes. We address this in a functional study of a large-effect gene on Atlantic salmon life histories. We show that variation in the genotype of a major maturation age gene, *vestigial-like 3*, changes interactions between this regulatory protein and DNA-binding transcription factors to bring about coordinated changes in the expression of multiple genetic pathways important for initiating male puberty. This mechanism explains how single important genes can mediate evolution in complex characters such as alternative life histories.

A central challenge in biology is to understand how genetic differences alter fitness via function. Such endeavors often start by discovering the genetic basis of fitness-related phenotypes. What is largely missing, even for cases where single genes of large effect are known, is an understanding of how genotype variation influence phenotype variation via molecular function of the identified genes. Finding causal functional links between genetic variation and ecologically relevant phenotypes [the "genotype–phenotype map" (1)] is key for understanding how gene function mediates evolutionary differences. Functional studies of genetic associations can identify causal genetic changes and clarify their molecular function. They further inform on the molecular basis of pleiotropy in genetic associations, allow the prediction of evolutionary outcomes in functionally connected genes, and provide the means to identify genetic associations below the detection limit of traditional approaches. A mechanistic dissection of the genotype–phenotype map therefore allows linking proximate and ultimate drivers of evolution as well as discovering the gradient of genetic architectures and effect sizes that underlie adaptive differences.

Understanding of the genotype–phenotype map (1) for ecologically relevant traits has been largely hindered by their complex genetic basis. Evolutionary theory predicts, however, that phenotypic evolution can be also mediated by single mutations of large effect (2). Large-effect genes have been found to underlie variation in adaptive traits of ecological relevance such as coloration (3–6), morphology (7, 8), behavior (9–11), physiology (12), and life histories (13–19). In most cases, however, the mechanisms connecting genotype to the phenotype remain unknown. This knowledge gap is especially wide for functionally multifaceted phenotypes such as life history traits where differences in physiology, development, and behavior tend to covary, where consequences of variation for fitness are large, and that have therefore often been expected to be highly polygenic (20).

For understanding the genotype–phenotype map in complex life histories, few species are as powerful as Atlantic salmon (*Salmo salar*). Salmon is among the most variable vertebrates on Earth in terms of life history strategies (21), with a single gene, the

Author contributions: J.-P.V., J.E.M.-V., P.V.D., and C.R.P. designed research; J.-P.V., J.E.M.-V., I.D., M.F., and A.R. performed research; J.-P.V., J.E.M.-V., P.V.D., and J.E. contributed new reagents/analytic tools; J.-P.V. analyzed data; and J.-P.V. and C.R.P. wrote the paper.

This article is a PNAS Direct Submission.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: jukka-pekka.verta@nord.no or craig.primmer@helsinki.fi.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2402386121/-/DCSupplemental.

Published November 19, 2024.

transcription cofactor *vestigial-like 3* (*vgll3*), encoding for nearly 40% of maturation age variation in natural populations (14, 22). The phenotypic effect of *vgll3* genotype is large considering the average sea-age at maturity of 1.6 y of Atlantic salmon; individuals with the Late–Late (LL) genotype mature 0.86 y later compared to individuals with the Early–Early (EE) genotype (14). Genetic variation in *vgll3* is associated with an ensemble of puberty-related traits in Atlantic salmon including age at maturity (14, 22), early male maturation (23–25), body condition (24, 26), food acquisition preference (27), aerobic performance (28), aggressive behavior (29), and migration behavior (30). Although *vgll3* may represent the most pleiotropic large-effect life-history gene yet found, the molecular mechanisms by which *vgll3* genotypes have such pleiotropic effects on distinct phenotypes are not known.

Single "master regulators" act as hubs in gene networks and can influence multiple phenotypes in a pleiotropic manner through their effects on coregulated genes mediating development (31, 32). We therefore hypothesized that the mechanism of genotype–phenotype associations in *vgll3* is likely linked to its function as a transcriptional regulator. It was previously shown that *vgll3* genotypes differ in two nonsynonymous mutations that may influence its structure or interaction with transcription factors (14). Further haplotype analysis suggested that variation in age at maturity is best explained by a single SNP in a noncoding region adjacent to *vgll3* (33). In addition, we have shown that in the tissue with highest *vgll3* expression, the immature testes, the genotype associates with the prepubertal expression of alternative isoforms of *vgll3* (23). The mechanisms how these putative functional and regulatory differences in *vgll3* may influence downstream gene regulation are not known, however, they have been hypothesized to be linked to Sertoli cell function (23, 34, 35).

We studied the transcriptomic trajectory of the differentiating male gonad from early spring until breeding time in the autumn, and its association with *vgll3* genotype in a common-garden experiment. By using chromatin immunoprecipitation-sequencing to map VGLL3 gene regulatory elements, we further identify interacting transcription factors. We show that VGLL3 plays a role in regulating key developmental processes in the gonad that are functionally connected to other traits beyond puberty, yet also associated with the *vgll3* genotype. The results shed light on the molecular machinery behind adaptive variation in maturation age and exemplify the mechanisms by which single major-effect genes may alter multiple cellular phenotypes in a concerted manner to mediate the development of alternative life history strategies. We predict that such functional architecture of large-effect genes facilitates fast evolution in multifaceted phenotypes such as life histories.

Results

Key Pubertal Mechanisms Are Up-Regulated in the *early vgll3* Genotype. Given the pronounced impact of *vgll3* genotype on maturation age, and its high expression in immature testes (23), we hypothesized that it serves as key regulator of maturation age prior to the onset of puberty. To test this, we assayed differences in gene expression (RNA-seq), gene regulatory elements, and VGLL3 binding regions (ChIPmentation) in immature testes of individuals with alternative homozygous *vgll3* genotypes conferring either late ($vgll3_{LL}$, N = 9) or early ($vgll3_{EE}$, N = 10) maturation. We sampled individuals across the second year of growth of salmon juveniles, i.e., one growth season before expected maturation (Fig. 1*A*). We have previously shown that *vgll3* genotype association with maturation age was reproducible in controlled conditions and robust to population, ambient temperature, and husbandry conditions (23–25) (*SI Appendix*, *Supplementary Analyses* and Fig. S1).

Season, encompassing the second spring until late fall, was a major source of variation in testicular gene expression. Factor analysis (36) of testes RNA-seq data, which accounts for heterogeneity associated with e.g., cell cycle stage and sample maturation trajectories (*SI Appendix*), separated samples according to season based on the first two principal components of staged expression data (Fig. 1*B*). Using a generalized linear model analysis, we identified 957 differentially expressed genes (DEGs) affected by sampling date, while adjusting for *vgll3* genotype ($P_{ADJ} < 0.05$). These DEGs included many transforming growth factor beta (TGF-β) signaling and transcriptional regulator proteins, some of which show genetic association with pubertal age in humans (37) (*SI Appendix*, Table S1). Our analyses therefore indicated that seasonal expression dynamics involve key signaling pathways in sexual development.

Vgll3 genotype was associated with the expression of an important set of genes linked to puberty. We identified 70 DEGs between *vgll3* genotypes ($P_{ADJ} < 0.05$), while adjusting for season, with five genes of particular relevance to sexual maturation that were up-regulated in $vgll3_{EE}$ genotype which confers early maturation, compared to typically later maturing $vgll3_{LL}$ individuals (Fig. 1*C*); *nuclear receptor family 5 group A 1* (*nr5a1*, a.k.a. *SF-1*, Fig. 1*D*), encoding for a transcription factor that controls the production of sex hormones and sexual development (38); *steroid 17-alpha-hydroxylase/17,20 lyase* (*cyp17a1*), a gene regulated by *nr5a1* that encodes for an essential step in androgen synthesis and is required for male maturation; *semaphorin 3D* (*sema3d*), encoding for a secreted signaling peptide that regulates cell adhesion and cytoskeleton during cell migration; *malic enzyme 3* (*me3*), where genetic and isozyme variation has been associated with growth (39) and maturation age variation in Atlantic salmon (40); and *latent-transforming growth factor beta-binding protein 4* (*ltbp4*), which regulates TGF-β signaling. Concerted expression differences in genes associated with sex hormone production, cell migration, energy production, and TGF-β signaling pointed toward *vgll3* controlling for distinct facets of cellular puberty changes in the testes.

We additionally identified 35 DEGs with an interaction between *vgll3* genotype and sampling date ($P_{ADJ} < 0.05$), including *nuclear receptor coactivator-1* (*ncoa1* a.k.a. *steroid receptor coactivator-1* [*SRC-1*]). *Ncoa1* expression increased toward breeding time in $vgll3_{EE}$ individuals while it decreased in $vgll3_{LL}$ individuals (Fig. 1*E*). Given that increasing *ncoa1* expression promotes lipid storage usage in mammals (41), our results suggest that, toward breeding time (autumn), the contrast in maturation age between *vgll3* genotypes translates into alternative dynamics of energy expenditure; $vgll3_{EE}$ individuals invest stored energy in processes enhancing puberty initiation through upregulation of *ncoa1* expression and subsequent lipid expenditure, while the relative lower expression of *ncoa1* in $vgll3_{LL}$ individuals predicts investment in adiposity and growth, and therefore future size and reproduction.

Upregulation of Maturation Genes Associates with VGLL3 Binding. We reasoned that if functional differences in the VGLL3 protein mediate the observed expression differences, DEG loci should be associated with VGLL3 binding. To test this, we performed ChIPmentation of histone modifications H3K27ac and H3K4me3, associated with active gene transcription, as well as VGLL3, based on a custom-produced antibody for Atlantic salmon.

We identified promoter regions as genomic segments associated with both H3K27ac and H3K4me3, while we defined enhancers as genome segments with H3K27ac alone (42); both features

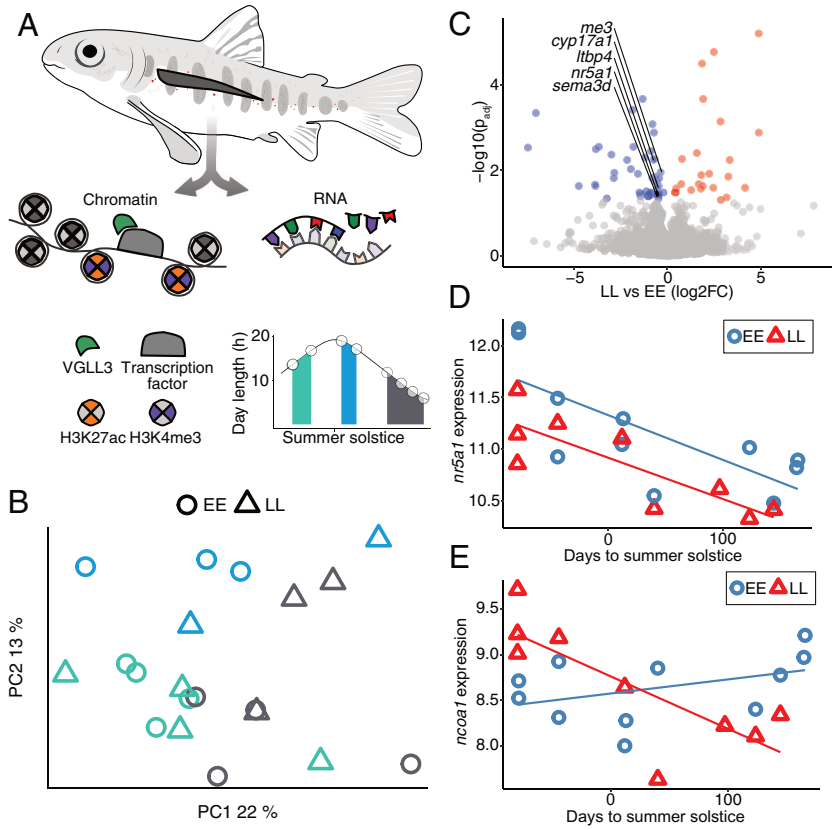

Fig. 1. Identification of differentially expressed genes (DEGs) in immature salmon testes. (*A*) Immature testes were sampled for RNA and chromatin across the second summer of growth. Sampling time points are indicated by white circles, with colored shading indicating the sampling period seasons. (*B*) Testes transcriptomic data separate samples according to season. Samples are grouped in three categories for illustration purposes. Green, blue, and gray coloring represent sampling dates from Panel *A*. (*C*) DEGs between *vgll3* genotypes. Multiple key maturity genes show consistently higher expression in $vgll3_{EE}$ genotypes. (*D*) *Nr5a1* shows higher expression in $vgll3_{EE}$ (blue) compared to $vgll3_{LL}$ (red) genotypes and both genotypes show a trend of decreasing expression toward autumn. (*E*) *Ncoa1* expression shows an interaction between *vgll3* genotype ($vgll3_{EE}$; blue, $vgll3_{LL}$; red) and season.

correlated with expression of nearest genes (Fig. 2 *A* and *B*). VGLL3 binding, when overlapping with promoters, associated with increased gene expression (Fig. 2*C*, $P < 2.2e\text{-}16$). We called VGLL3 peaks separately for *vgll3* genotypes using combined data from all biological replicates ($vgll3_{EE}$: 61,178 peaks, $vgll3_{LL}$: 73,516 peaks). Analysis of replicability across individual samples (43) showed that these peaks were nearly 100% reproducible across biological replicates. *Vgll3* genotypes showed marked differences in both promoters and enhancers that were occupied by VGLL3 (Fig. 2 *D*–*F*). VGLL3 binding was observed in a core set of promoters and enhancers in both genotypes (VGLL3 promoters: observed 2747, expected 97, $P = 0.009$, VGLL3 enhancers: observed 11,137, expected 1,172, $P = 0.009$), with nearly equal numbers of regulatory elements with VGLL3 occupancy specific to each genotype (Fig. 2*E*). To test for functional effects of this genotype difference in VGLL3 binding, we analyzed GO overrepresentation in each category relative to all VGLL3 regulatory elements (promoters and enhancers). Genes associated with VGLL3 elements unique to the $vgll3_{EE}$ genotype were overrepresented in signaling receptors, cell adhesion genes, and those involved in actin regulation, among other functions (*FDR* < 0.1, Fig. 2*F*). Genes associated with VGLL3 elements unique to $vgll3_{LL}$ were overrepresented in semaphoring receptors and regulators of cell cycle progression, among others (*FDR* < 0.1, Fig. 2*F*). Interestingly, cell differentiation process was overrepresented independently in both genotypes, indicating that VGLL3 regulatory loci influencing this process were largely unique to each genotype, or that they were acting as promoters in one genotype and enhancers in the other. Taken together, these results indicate that VGLL3 is widely associated with gene regulatory regions in immature testes and suggest that *vgll3* genotype differences have wide-scale influence on cellular physiology and development through coordinated regulation of distinct genomic loci and cellular functions.

We proceeded with identification of regulatory elements that could directly mediate the *vgll3* genotype effect on gene expression and further on maturation age. In total, 50 of the 70 DEGs between *vgll3* genotypes were associated with VGLL3 binding regions (23 with VGLL3 regions within promoters or enhancers), suggesting that expression differences may be directly mediated by functional differences in VGLL3 protein. DEGs with VGLL3 promoters or enhancers showed a notable trend with the majority having higher expression in $vgll3_{EE}$ individuals (5/5 VGLL3 promoters, 17/23 VGLL3 enhancers), suggesting that transcriptional upregulation associated with the $VGLL3_E$ protein was stronger compared to the $VGLL3_L$ protein. The up-regulated DEGs included *nr5a1, cyp17a1, me3, sema3d,* and *ltbp4,* each associated with a central process in sexual maturity, namely, androgen production, metabolism, cell motility, and TGF-β signaling (Fig. 3 *A* and *B* and *SI Appendix*, Fig. S2). Of these, VGLL3 binding in proximity of *nr5a1, cyp17a1,* and *sema3d* associated with differential H3K4me3 enrichment, with histone trimethylation observed only in $vgll3_{EE,}$ but this signal was driven by strong H3K4me3 enrichment in a single sample. Overall, our results strongly suggested that VGLL3 binding mediated the higher expression of key maturation genes in the $vgll3_{EE}$ genotype, corresponding to a concerted upregulation of signaling pathways prepuberty.

To test whether VGLL3 binding could mediate some of the most salient patterns of seasonal expression variation, namely the decrease in expression of *nr5a1* toward the autumn, we investigated the

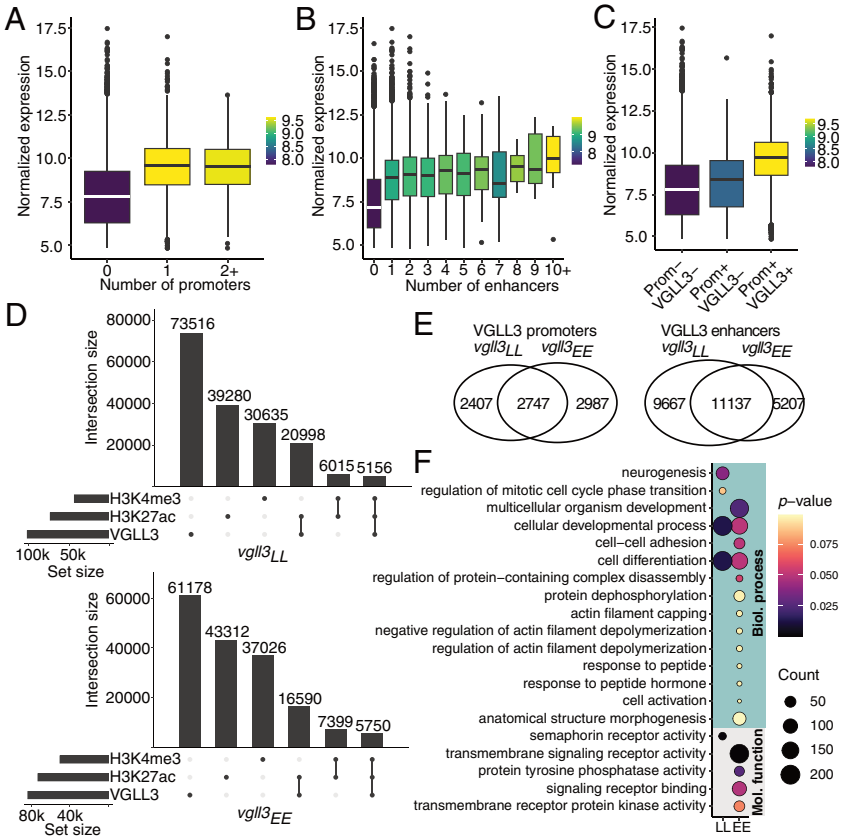

Fig. 2. Identification of gene regulatory elements by genome-wide mapping of histone modifications and VGLL3 binding regions. The number of (*A*) promoters (H3K27ac + H3K4me3) and (*B*) enhancers (H3K27ac alone), assigned to nearest expressed genes correlate positively with gene expression levels. (*C*) VGLL3 binding, when associated with promoters, is associated with increased gene expression. (*D*) Numbers of VGLL3 regions, promoters, and enhancers in *vgll3* genotypes. (*E*) Sharing of VGLL3 promoters and VGLL3 enhancers between *vgll3* genotypes ($\log_{FC} > 2$). (*F*) Overrepresented gene ontologies in genes with VGLL3 promoters or VGLL3 enhancers that are unique to each genotype ($P_{ADJ} < 0.1$).

strength of VGLL3 binding signal in the proximity of the *nr5a1* promoter. We identified two VGLL3 binding regions that showed a decrease in VGLL3 signal toward the autumn, indicating that they likely mediated *nr5a1* expression dynamics (Fig. 3*C*). We further identified proximal VGLL3 binding regions (7 VGLL3 enhancers) in 16 DEGs showing interaction between *vgll3* genotype and season. Our expression data showed that seasonal expression of *ncoa1* was dependent on *vgll3* genotype; a declining expression in $vgll3_{LL}$ genotype suggested investment on adiposity, while an increasing trend in $vgll3_{EE}$ genotypes was suggestive of utilization of fat reserves to promote pubertal initiation. VGLL3 signal in one VGLL3 enhancer in proximity of *ncoa1* showed strong differences between *vgll3* genotypes (Fig. 3*C*). An additional striking VGLL3 binding region having one of the strongest signals in the genome was observed in the first intron of *ncoa1,* and associated with differential H3K4me3 in *vgll3* genotype, albeit the H3K4me3 signal that was driven by a single sample. Overall, our results identify specific VGLL3 regulatory elements that likely mediate expression differences of key downstream genes between *vgll3* genotypes and season.

***vgll3* Self-Regulation as a Genotype–Phenotype Link in Maturation Age.** In natural populations, the strongest association between maturation age and genetic variation was observed in a noncoding region adjacent to *vgll3* (14), suggesting that divergence in gene regulatory elements acting in *cis* to *vgll3* contribute to the causal mechanism of maturation age variation. To test this, we used testis chromatin modification and VGLL3 binding data, and identified a prominent VGLL3 binding region adjacent to the analyzed SNP of strongest association signal (Fig. 3*D*) (14). Additionally, a single $vgll3_{EE}$ individual showed a strong overlapping H3K4me3 enrichment, tentatively suggesting that VGLL3 binding in the region may associate with chromatin remodeling more generally. These results suggest that the genotype–phenotype association between *vgll3* and male maturation age includes differential self-regulation of the *vgll3* gene by itself, possibly forming a feedback loop that might interplay with the nonsynonymous mutations and splicing variation also associated with maturation age (14, 23). We speculate that the different mechanism may increase regulatory complexity in *vgll3* as regulatory element activity and splicing are cell- and time-dependent, while protein-coding mutations influence all cells where *vgll3* is being expressed. Overall, our results support a mechanism where *cis* regulation in a "master regulator" influences the expression of downstream genes in *trans* (31).

VGLL3 Drives Gene Coexpression Differences in Diverse Signaling Pathways. Single master regulators may drive differences in complex phenotypes by way of influencing the expression of a coregulated set of genes (31). To test whether *vgll3* could mediate its effects on molecular phenotypes in this way, we investigated the effect of *vgll3* genotype and season on gene regulatory networks using gene coexpression network analysis. Expressed genes were assigned to 70 coexpression modules ranging in size from 31 to 10,972 genes. Using an ANOVA model to test the difference in module (eigengene) expression between $vgll3_{EE}$ and $vgll3_{LL}$ genotypes, we identified four coexpression modules with higher expression in $vgll3_{EE}$ individuals, containing 1,457 genes ($P < 0.05$, not significant after correction for multiple testing, Fig. 4*A*). Module genes were connected to DEGs by function and involved

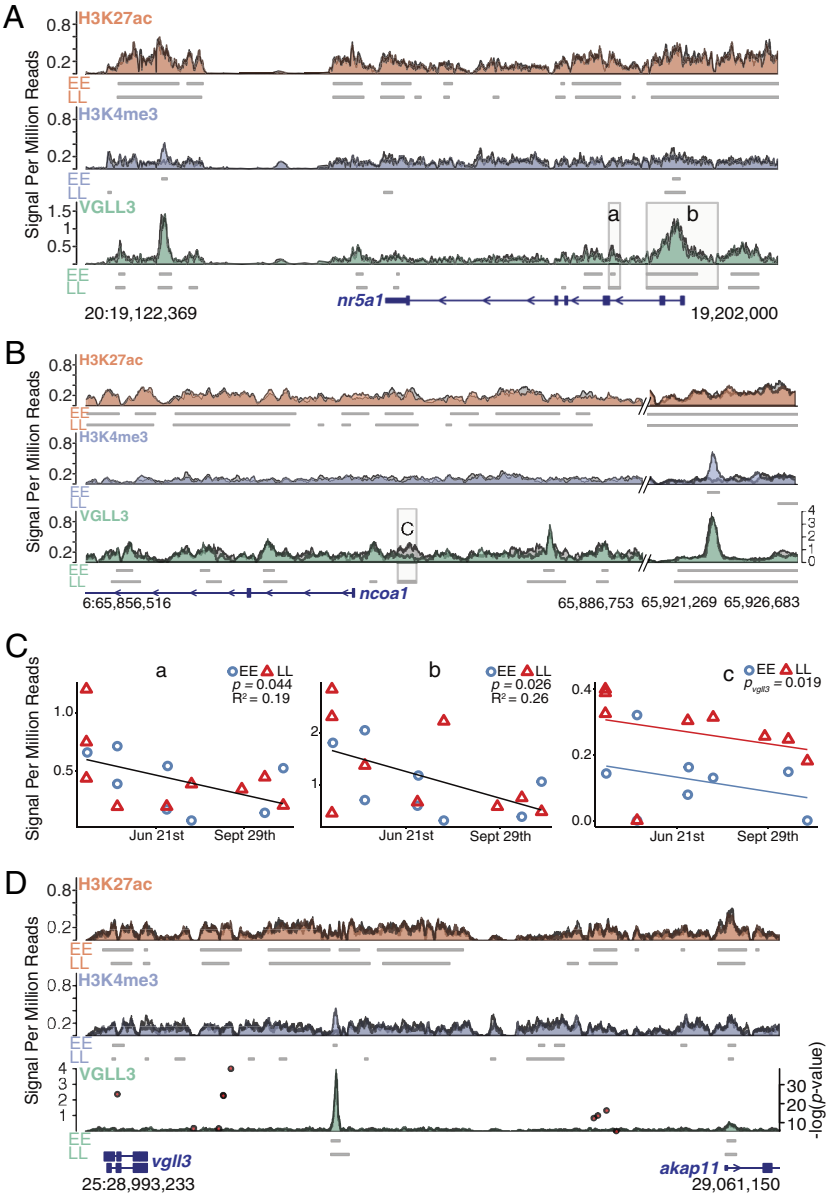

Fig. 3. VGLL3 binding is associated with *vgll3* genotype effects and seasonal expression variation in central maturation genes. ChIPmentation signals from $vgll3_{EE}$ and $vgll3_{LL}$ genotypes are overlaid for each histone modification/VGLL3 ($vgll3_{LL}$ on gray). Horizontal bars below ChIPmentation signal represent enriched regions called by MACS3 in each genotype (H3K27ac & H3K4me3 $q < 0.1$, VGLL3 $q < 0.05$). (*A*) Decreasing *nr5a1* expression toward the autumn is putatively mediated by two VGLL3 regulatory regions (shaded areas; a & b) in proximity of the *nr5a1* promoter. (*B*) *Ncoa1* expression differences between *vgll3* genotypes are putatively mediated by VGLL3 binding in a regulatory region in proximity of the *ncoa1* promoter (shaded area; c). (*C*) VGLL3 binding signal in regulatory regions of *nr5a1* and *ncoa1* associate with season (a, b) and *vgll3* genotype (c). (*D*) Intergenic region adjacent to the *vgll3* locus shows a prominent VGLL3 binding region. GWAS association *P*-value from ref. 14 is represented as red points on VGLL3 ChIP-signal background (axis on the *Left*).

in cell adhesion, immune system, chemokine and cytokine signaling, TGF-β signaling, and Hippo signaling (*SI Appendix*, Fig. S3). Additionally, one module containing 757 genes showed higher expression in $vgll3_{LL}$ individuals; it was overrepresented in genes associated with cell–cell signaling and negative regulation of the Wnt pathway (e.g., *APC regulator of WNT signaling pathway*), further supporting that *vgll3* genotype mediated concerted changes in cell adhesion and motility through gene coregulation.

The *magenta* module supported a regulatory link between Hippo and Ras signaling mediated by VGLL3 binding in association with *tead3* (Fig. 4*B*). Hippo signaling has been shown to control cell proliferation and differentiation through Ras (44), possibly acting as a switch whereby VGLL3 could influence pubertal cell phenotypes such as Sertoli cell proliferation. We examined whether *tead3* (Hippo) network neighbors in the *magenta* module were associated with VGLL3 binding, as expected if *tead3* coexpression with these genes was mediated by VGLL3. Of the 24 *tead3* network neighbors, 16 had VGLL3 binding regions in close genomic proximity (13 with VGLL3 promoters, 3 with VGLL3 enhancers), including the negative regulator of Ras signaling, *neurofibromin* (*NF*). Two additional Ras GAPs (GTPase activating proteins) not included in *magenta* showed proximal VGLL3 peaks (*RASA1*, *RASAL1*). *Tead* has been shown to directly determine Ras signaling activity in *Drosophila* through regulation of the Ras default inhibitor *cic* and activator *ets1* (a.k.a. *Pnt*), thus controlling for the balance of cell proliferation versus differentiation (44). We therefore investigated VGLL3 binding on the seven *cic* and *ets1* gene copies (not included in *magenta,* one *cica* gene included in *blue*) in Atlantic salmon and found all to be associated with strong VGLL3 peaks. Together, the results strongly suggest that Hippo-VGLL3 mediated control of cell proliferation in the testes is mediated in part by their effect on overall Ras/MAPK signaling activity, as well as its effectors. *Vgll3*-mediated regulation of Ras genes provides a putative link for mediating *vgll3*-associated oncogenic development as aberrant Ras signaling is one of the most frequent causes cancer (45, 46).

Four coexpression modules showed a significant correlation with sampling date ($P_{ADJ} < 0.05$, Fig. 4*C*). Among the three modules with decreasing trend toward spawning time, the *blue* module showed strong overrepresentation in genes involved in cell adhesion and morphogenesis, cell motility, and Wnt signaling, among others (*SI Appendix*, Fig. S4). *Blue* included *vgll3*, *vgll4*, *tead1b*, *tead2*, *nr5a1*, *cyp17a1*, *me3*, *sema3d*, *Wnt4*, among other genes (*SI Appendix*, Table S3). *Blue* module expression pattern was consistent with a

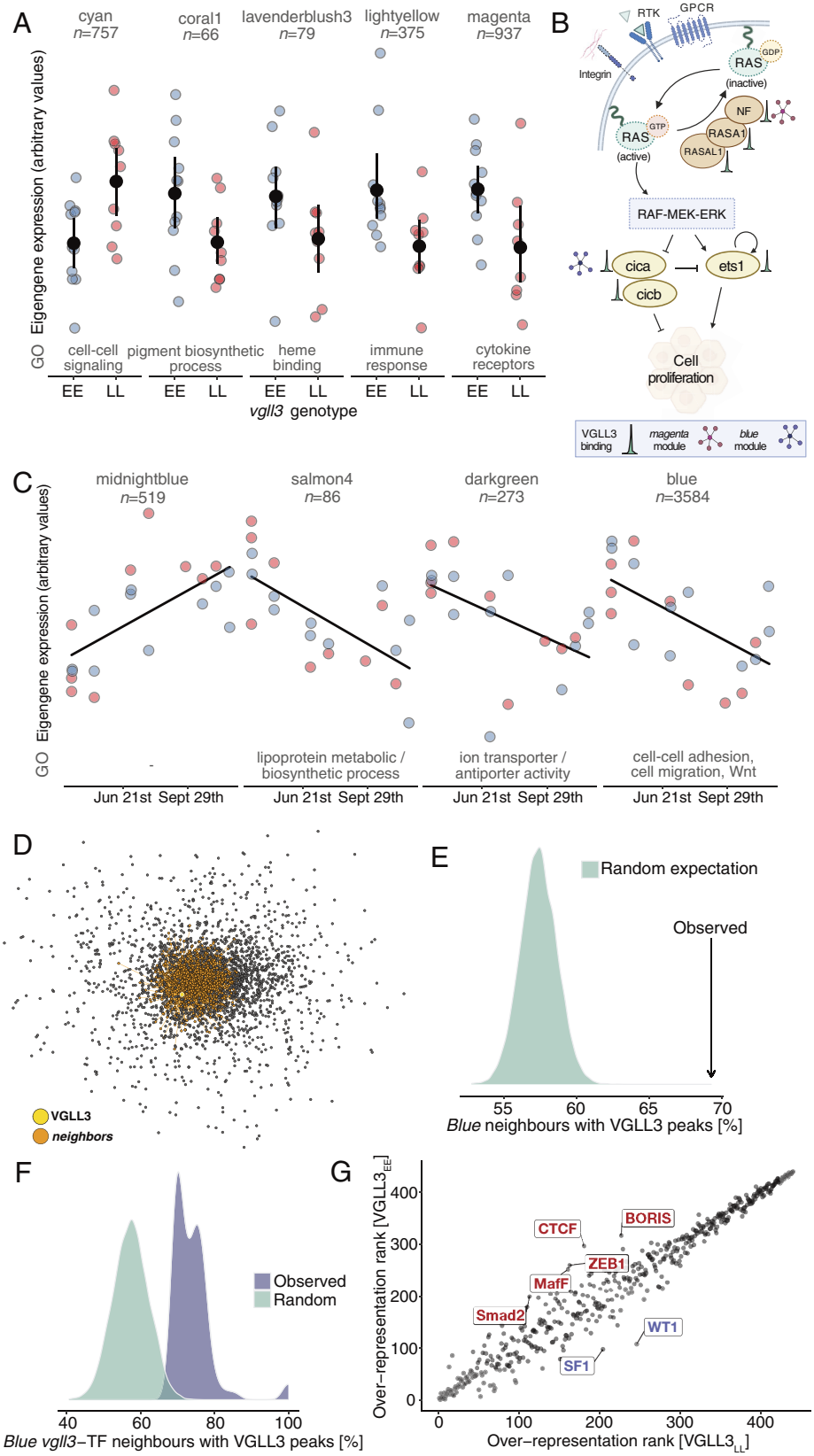

Fig. 4. Gene coexpression networks reveal signaling pathways associated with *vgll3* genotype and seasonal variation in testes transcriptomes. (*A*) Five modules show significant eigengene expression difference between *vgll3* genotypes ($P < 0.05$, not significant after correction for multiple testing). Modules are annotated according to the number of included genes and selected, overrepresented GO categories. Blue spheres, $vgll3_{EE}$ individuals; red spheres, $vgll3_{LL}$ individuals; black spheres, mean; whiskers, 95% confidence limit estimates. (*B*) Ras signaling pathway genes show functional association with *vgll3*. Genes are annotated with VGLL3 binding peaks and inclusion in coexpression modules. (*C*) Expression of four modules shows significant association with season ($P_{ADJ} < 0.05$). (*D*) The *blue* module includes *vgll3* (golden sphere). Closest *vgll3* neighbors are represented as orange spheres. Genes are ordered according to the strength of their coexpression, showing that *vgll3* expression is highly correlated with most of the genes in *blue* (*vgll3* situates toward the center of the network). (*E*) Percentage of *blue* module neighbors with proximal VGLL3 binding regions (arrow), compared to expected distribution based on random sampling of genes. (*F*) Percentage of shared neighbors between *vgll3* and transcription factors included in *blue* with proximal VGLL3 binding regions, compared to expectation from random sampling of genes. (*G*) Difference in TF motif overrepresentation in $vgll3_{LL}$ and $vgll3_{EE}$ enhancers and promoters, compared to all enhancer and promoter regions within genotypes. Overrepresented motifs are rank-transformed from most highly overrepresented to least overrepresented. Motifs with a rank difference over 80 are highlighted.

decreasing expression of genes inhibiting epithelial-to-mesenchymal transition (EMT, e.g., *cadherin*, *protocadherin*, and *Wnt4*) toward the breeding season consistent with the role of EMT in testes development and differentiation (47), and a role of *vgll3* in the regulation of EMT (48). *Vgll3* was among the most connected genes within *blue* with 1,490 neighbors of the 3,563 genes in *blue* in total (75th quantile, Fig. 4*D*). *Blue* also contained 103 transcription factor genes (GO:0003700), which allowed us to identify putative regulatory partners of *vgll3* (*SI Appendix*, Fig. S5). Top transcription factors with most network neighborhood sharing with *vgll3* included immune system, retinoic acid, Hippo, homeobox, thyroid hormone, and nuclear factor regulators, putatively connecting *vgll3* to diverse

cellular signaling pathways including membrane-bound receptors and nuclear receptors (*SI Appendix*, Fig. S5); these gene pathways have been implicated in maturation age variation in humans as well (37).

We next investigated VGLL3 binding in proximity of *vgll3* network neighbors in the *blue* module to test whether VGLL3 binding could mediate the coregulation of the genes. Of all *vgll3* neighbors, 69% were associated with VGLL3 binding regions, (significantly more than expected by chance, Fig. 4*E*), suggesting that *blue* coexpression could largely be mediated by VGLL3 regulation. Similarly, shared network neighbors between *vgll3* and transcription factors in *blue* were overrepresented in genes with VGLL3 binding regions compared to a random expectation (Fig. 4*F*), suggesting that *blue* coexpression was mediated by VGLL3 interaction with a diverse set of transcription factors.

VGLL3 Binds Regulatory Elements in Association with Distinct Transcription Factors in Different Genotypes. Our results indicated that *vgll3* genotypes were associated with large-scale differences in gene expression and regulatory element activity, and that these differences were mediated by VGLL3 binding. VGLL3 does not contain a DNA-binding domain, rather, it interacts with transcription factors to regulate gene expression (49). We used motif analysis of VGLL3 binding regions to identify transcription factors that putatively co-occupy regulatory elements with VGLL3 and may thus mediate the genotype effect on gene regulation. Although motif overrepresentation was largely conserved in VGLL3 regions between the genotypes, notable differences in motif enrichment indicated that VGLL3 co-occupied regulatory elements with distinct sets of transcription factors in different genotypes (Fig. 4*G*). Motifs with most pronounced overrepresentation in $vgll3_{EE}$, compared to $vgll3_{LL}$, were *nr5a1* (*SF-1*) and *WT1*, both positive regulators of sexual development and maturation. An alternative approach of testing the overrepresentation of motifs in $vgll3_{EE}$ peaks using $vgll3_{LL}$ peaks as a background set produced concordant results (P_{nr5a1} = 1e-199, P_{WT1} = 1e-131). In $vgll3_{LL}$, compared to $vgll3_{EE}$, most overrepresented motifs corresponded to *ZEB1, OVOL2, SMAD2, BORIS, CTCF,* and *MafF*, involved in the regulation of EMT, inhibition of androgen receptor activity, and chromatin structure. Using the $vgll3_{EE}$ peaks as a background set to test the significance of the overrepresentation concorded with these results (P_{ZEB1} = 1e-470, P_{OVOL2} = 1e-1072, P_{SMAD2} = 1e-289, P_{BORIS} = 1e-209, P_{CTCF} = 1e-296, P_{MafF} = 1e-295). These results indicate that VGLL3 interacts with distinct sets of transcription factors that vary depending on the *vgll3* genotype; either promoting sexual maturation in $vgll3_{EE}$, or putatively inhibiting it in $vgll3_{LL}$.

Discussion

A dissection of the proximal mechanisms linking genes to traits is required for a complete understanding of evolutionary change (1). Here, we dissected the molecular mechanisms that underlie the association of genetic variation in the transcription cofactor *vgll3* and male maturation age. We show that genetic variation in *vgll3* leads to differential regulation of multiple distinct signaling pathways and thousands of genes by association of VGLL3 with different transcription factors controlling pubertal cellular development such as proliferation, EMT-MET, cell adhesion, and motility. Our results exemplify a case where the mechanism translating genetic variation into phenotypic variation for a major-effect gene in life histories is, despite its simple genetic basis, functionally multifaceted.

Our results bring significant insight into the molecular basis of pleiotropy in ecologically relevant genes by showing that *vgll3* is connected to diverse signaling pathways through its central role in gene regulation. Theory predicts that the pleiotropic function of major regulatory proteins may drive covariation in traits (31), but so far, the molecular mechanisms that may mediate such dynamics for ecologically relevant phenotypes have not been empirically demonstrated and thus the validity of the theory in the context of evolution has remained untested. *Vgll3* is associated with differences in salmon maturation age, behavior, and physiology (14, 22–30). Our results show that *vgll3* regulates genes that indeed mediate these phenotypes in other species, including salmonids, providing substantial circumstantial evidence of the mechanistic role of *vgll3* in not only maturation, but also behavior and physiology. For example, we show that *vgll3* regulates *nr5a1* expression, which associates with maturation age variation in humans (50) and in rainbow trout (*Oncorhynchus mykiss*) (18); *nr5a1* also regulates aggressivity and physical activity in mice (51–53). *Cyp17a1,* another target of *vgll3*, is essential for testosterone production in the testes and *cyp17a1* knock-out zebrafish show differences in mating behaviors mediated by differences in blood testosterone levels and brain gene expression (54). We also show that *vgll3* regulates *ncoa1* expression. Ablation of *ncoa1* expression lowers cellular oxygen consumption and leads to body weight gain in mice (55), corroborating with seasonal changes in salmon body condition that depend on *vgll3* genotype (26). We anticipate that additional genes mediating the diverse cellular functions regulated by *vgll3* explain the remaining polygenic component of maturation age acting in addition to the gene itself (24, 56). These functionally connected genes are candidates for mediating maturation variation in other species, as genetic variation in humans, for example, associates with *vgll3* (37). The complexity of mechanisms identified here implies that in species where life history variation has a more polygenic genetic architecture, such as maturation age variation in humans (37), we can expect an extremely pleiotropic genotype–phenotype map approaching to what has been proposed for other complex traits (31).

We speculate that pleiotropic architectures such as that of *vgll3* may evolve in cases where fitness gains from mutation of large effect influencing one or multiple traits are more significant than fitness costs of their pleiotropic effects on others. We further hypothesize that transcriptional regulators, and especially transcription cofactors, may be particularly amenable for the evolution of such architectures as they are typically involved in the development of multiple phenotypes through multiple signaling pathways, depending on the cellular context (57, 58), thus potentiating pleiotropic control of multiple traits. As opposed to transcription factors, transcription cofactors do not bind specific DNA elements but participate in gene regulation through protein–protein interactions. This may allow cofactors to accumulate genetic variation that alter their interactions with transcription factors without compromising their function, thus leading to changes in the regulation of cellular pathways. From an ecological perspective, we predict that pleiotropic architectures may facilitate rapid evolution of phenotypically multifaceted characters such as life histories, as changes in the allele frequency of a single gene with pleiotropic effects may mediate change in multiple phenotypes through genetic correlations. Such fast evolutionary trends have indeed been observed for *vgll3* and maturation age in Atlantic salmon (59).

Our results also pertain to interpretation of the evolution of life histories and trade-offs therein. Trade-offs in life-history traits have been viewed as a consequence of resource allocation—that organism cannot allocate resources to e.g., condition and fecundity at the same time, hence there are trade-offs (60). Our results bring substantial empirical support for an alternative interpretation where trade-offs between life-history traits may have a mechanistic basis because the

traits are connected through gene regulatory networks (20). In the case of Atlantic salmon, our results indicate that the association of early maturity and higher body condition in spring is mediated through the function of *vgll3* controlling the development of both traits. The accumulation of adipose reserves necessary for producing gametes, mediated by *ncoa1*, and the development of sexually mature gonads, mediated by *nr5a1*, seem both to be under the pleiotropic control of *vgll3*. As condition is a liability trait for maturation, we speculate that evolution of maturation age differences in Atlantic salmon may have led to mutations in *vgll3* to be selected for as it seemingly controls both condition and maturation, therefore linking the underlying physiological connection between liability and threshold traits, respectively, into a genetic program. Our results therefore indicate that the evolution of trade-offs may be dictated by also the function of major life-history genes instead of solely physiological optimization of resource allocation.

Overall, our results exemplify a hidden complexity in the molecular mechanisms mediating the large effects of a single gene on alternative life histories. With this work we demonstrate a mechanism through which changes in regulatory interactions of a single gene can mediate phenotypic covariation in a spectrum of seemingly unrelated traits. Evaluating the prevalence of pleiotropic effects will require not only advances in our understanding of the molecular mechanisms of genotype–phenotype associations in other traits and species, but also the assessment of phenotypic effects of large-effect genes across phenotypes. In the light of our results, evolution in functionally multifaceted phenotypes such as alternative life histories and the "pace-of-life," typically considered to be highly polygenic characteristics of species, may indeed be mediated by mutation of a single gene with pleiotropic effects.

Material and Methods

Material. Fish crosses used are described in depth in ref. 23. Briefly, 2 × 2 factorial matings between *vgll3* LL and EE individuals were used to create 32 families of *vgll3* LL, LE, EL, and EE individuals. The broodstock is a first-generation hatchery stock composed of a mixture of northern Baltic Sea lineages and maintained by the Natural Resources Institute Finland (LUKE). Eggs were fertilized in October 2017 and raised in controlled conditions at the University of Helsinki research animal facility. Two egg incubators with families separated by mesh compartments were used for housing eggs and alevin up to first feeding. After first feeding, 48 individuals were randomly selected from each family and distributed in roughly equal densities across four 0.25 m^3 recirculating tanks.

Fish were raised using a typical annual cycle of water temperature and photoperiod corresponding to latitude of origin (min–max temp: 6.3 to 17.7 °C, latitude: 61 N) for 1.5 y with ad libitum feeding of appropriately sized commercial fish feed. At 8 mo of age, fish were individually tagged with passive integrated transponders (PIT-tags) and a fin clip was collected for genotyping. Genotyping was performed using a set of 141 SNP markers, including *vgll3*, and PCR-sequencing (61). Individuals were assigned to families using SNPPIT (62).

Individuals were preassigned randomly for sampling dates with equal representation of genotypes and sexes. At assigned time points, fish were killed using an overdose of buffered tricaine methanesulfonate (MS-222) and dissected. Their maturity phenotype was recorded and tissue samples including testes were flash-frozen on liquid nitrogen and stored in −80 °C until extraction of RNA and chromatin. Testes tissue for ChIPmentation was immersed in 1 mL Cryostor10 in cryovials and incubated on ice. Cryovials were transferred to a MrFrosty freezing container and transferred to −80 °C. Cryostored tissue was stored in −80 °C until extraction of cells for chromatin. Experimentation was conducted according to the Finnish Government Decree on the Protection of Animals Used for Scientific or Educational Purposes (564/2013), which implements EU directive 2010/63/EU. The experiments in this study were approved by the Project Authorisation Board (ELLA) on behalf of the Regional Administrative Agency for Southern Finland (ESAVI) under experimental license ESAVI/2778/2018.

RNA Extraction and RNA-seq. Gonad tissue was removed from −80 °C storage and flash-frozen in Macherey-Nagel (MN) Nucleospin 96 RNA extraction buffer. Frozen samples for 9 $vgll3_{LL}$ individuals and 10 $vgll3_{EE}$ individuals were homogenized on a OMNI Bead Ruptor Elite instrument using 2 mL tubes with 2.4 mm steel beads and a program with 6 bursts of 20 s with speed 4.5, and 10 s pauses in between bursts. Total RNA was extracted following MN kit instructions on 96-well plates. RNA elutions were treated with the Invitrogen Turbo DNase kit reagents for removal of trace DNA and quantified using Thermo Quant-iT reagents on a 96-well plate reader. RNA concentration was verified using a Qubit instrument and Quant-iT reagents, and adjusted to 50 ng/μL with pure water. 100 ng of total RNA was used for RNA-seq library construction using the Illumina stranded mRNA kit and manufacturer instructions. Libraries were sequenced at the University of Helsinki Institute of Biotechnology sequencing service using a NextSeq500 instrument and 75 bp paired-end reads (*SI Appendix*, Table S2).

Antibodies. A custom anti-VGLL3 polyclonal antibody was procured from Genscript with delivery in January 2020 as follows. Target DNA sequence of *vgll3* was optimized and synthesized. The synthesized sequence was cloned into vector pET-30a (+) with His tag for protein expression in *Escherichia coli*. *E. coli* strain BL21 star (DE3) was transformed with recombinant plasmid. A single colony was inoculated into TB medium containing related antibiotic; culture was incubated in 37 °C at 200 rpm and then induced with IPTG. SDS-PAGE was used to monitor the expression. Recombinant BL21 star (DE3) stored in glycerol was inoculated into TB medium containing related antibiotic and cultured at 37 °C. When the OD600 reached about 1.2, cell culture was induced with IPTG at 37 °C/4 h. Cells were harvested by centrifugation. Cell pellets were resuspended with lysis buffer followed by sonication. The precipitate after centrifugation was dissolved using denaturing agent. Target protein was obtained by one-step purification using Ni column. Target protein was sterilized by 0.22 μm filter before stored in aliquots. The concentration was determined by BCATM protein assay with BSA as standard. The protein purity and molecular weight were determined by standard SDS-PAGE along with western blot confirmation. Two rabbits were immunized with three injections of purified VGLL3 protein and antiserum was tested for presence of anti-VGLL3 antibodies after 3rd bleeding using western blotting of Atlantic salmon heart tissue. Rabbits were immunized for a fourth time, killed, antiserum was pooled, and anti-VGLL3 antibody was affinity-purified from the antiserum pool. Purified VGLL3 antibody was validated for reactivity using an indirect ELISA VGLL3 protein as antigen.

Additional quality control for the VGLL3 antibody was performed using immunoprecipitation followed by western blotting of Atlantic salmon heart tissue, which was shown to have high *vgll3* expression (23). Western blotting with anti-VGLL3 antibody without immunoprecipitation gave inconclusive results with background bands and/or a band of lower-than-expected size. Western blotting after immunoprecipitation identified a band of the expected size (*SI Appendix*, Fig. S6).

The VGLL3 antibody was coupled with beads from the Dynabeads Co-Immunoprecipitation Kit (Invitrogen, Massachusetts, USA) using 10 μg antibody per mg beads. Protease inhibitors were added to the Extraction Buffer and flash-frozen salmon hearts weighing approximately 0.25 g were homogenized in 2,250 μL buffer each using the Bead Ruptor Elite (Omni International, Georgia, USA). The lysates were centrifuged at 840 × *g* for 2 min in 4 °C and the supernatants combined with 150 μL antibody-coupled beads. Proteins and beads were then incubated at 4 °C for 40 min. The rest of the immunoprecipitation was performed in accordance with manufacturer protocol. The immunoprecipitated protein samples and positive controls (purified VGLL3 protein) were prepared for western blotting by adding appropriate amounts of 4X Laemmli and DTT (final concentration of DTT 100 mM) and incubating at 92 °C for 5 min. Proteins were separated on Any kD Mini-PROTEAN TGX Precast Protein Gels in a minielectrophoresis system (Bio-Rad Laboratories, California, USA). For molecular weight sizing, Precision Plus Protein WesternC Standards (Bio-Rad Laboratories, California, USA) were used. Separated proteins were transferred to a 0.2 μm nitrocellulose membrane using Trans-Blot Turbo Transfer Packs and the Trans-Blot Turbo Transfer System (Bio-Rad Laboratories, California, USA). 3% BSA in Tris-buffered saline with 0,1% Tween 20 (TBST) was used for membrane blocking and for diluting the primary antibody 3:1,000. The incubation of the membrane in primary antibody was performed O/N at 4 °C. For protein detection, we used mouse anti-rabbit IgG-HRP sc-2357 (Santa Cruz Biotechnology, Texas, USA) and Precision Protein StrepTactin-HRP Conjugate (for protein standards; Bio-Rad Laboratories, California, USA), both diluted 1:20,000

in TBST. Chemiluminescence was recorded with the Pierce ECL Western Blotting Substrates (Thermo Fisher Scientific, Massachusetts, USA) and the ChemiDoc MP imaging system (Bio-Rad Laboratories, California, USA).

Chromatin Extraction and ChIPmentation. Detailed ChIPmentation protocol is provided as a supplement. In summary, gonad tissue for individuals used in RNA-seq was removed from −80 °C storage and rapidly thawed in a 37 °C water bath. Thawed tissue was rinsed with ice-cold D-PBS and transferred to a clean tube. Tissue was homogenized in ice-cold D-PBS using a OMNI Bead Ruptor Elite instrument with 7 mL tubes and 2.4 mm ceramic beads, and a program with one burst of 5 s with speed 2.4. Cell suspension was filtered through a Flowmi Cell Strainer and inspected under a microscope with Trypan blue staining. Cell suspension was diluted with room temperature D-PBS and cross linked with Diagenode ChIP crosslinking gold in 1× concentration for 30 min, followed by fixation with 1% formaldehyde for 2 min. Formaldehyde was quenched in 0.125 M glycine for 5 min and cells collected with centrifugation at 400×*g* for 10 min. Cells were washed two times with 500 μL ice-cold PBS and centrifuged at 400×*g* for 10 min in between washes. Cells were subject to ChIPmentation with Thermo MAGnify ChIP-kit and Illumina Tn5 reagents as detailed in supplement and briefly below.

Cells were collected using centrifugation, resuspended in 50 μL lysis buffer supplemented with protease inhibitors and lysed on ice for 5 min. Chromatin was sheared in 50 μL volumes using a Bioruptor device with settings high power and 3× eight cycles of 30 s on, 30 s off. Debris was pelleted by centrifugation and sheared chromatin was diluted to four equal aliquots of 100 μL using dilution buffer supplemented with protease inhibitors. One aliquot of sheared chromatin was reserved as input control. Remaining three aliquots were immunoprecipitated in 4 °C o/n using 1 μg of Abcam ab4729, 2 μg of Abcam ab8580, and 10 μg of a custom-procured anti-VGLL3 antibody on ThermoFisher Dynabeads Protein A/G. Beads were subsequently washed following MAGnify kit protocol, with an additional final wash using 150 μL of ice-cold 10 mM Tris (pH 8). Bead-bound chromatin was then treated in 20 μL volume of tagmentation reaction containing Illumina Tn5 transposase for 5 min at 37 °C. Input controls were treated with tagmentation reaction for 5 min at 55 °C. Tagmentation was terminated by adding 7.5 volumes of RIPA buffer and incubation on ice for 5 min. Chromatin was subsequently washed twice with 150 μL of ice-cold RIPA and TE buffer. Crosslinks were reversed using a proteinase-K treatment and ChIPmentation DNA was captured using Macherey-Nagel NucleoMag magnetic beads. ChIPmentation libraries were measured using a Qubit instrument and a control-PCR was run with Nextera sequencing oligos to assess library amplification on agarose gel. Finally, libraries were indexed, pooled, and sequenced at the University of Helsinki Institute of Biotechnology sequencing and FIMM sequencing services using NextSeq500 (75 bp paired-end) and Novaseq6000 (150 bp paired-end) instruments, respectively.

RNA-seq Analysis. RNA-seq reads were trimmed using "fastp" (63) version 0.20.1 and options "p trim_front1=2 trim_front2=2 detect_adapter_for_pe". Trimmed reads were aligned using "STAR" (64) version 2.7.9a and manual two-pass mode to the Atlantic salmon genome (Salmo_salar-GCA_905237065.2) downloaded from Ensembl. Alignment options for "STAR" were defined as "out FilterIntron Motifs Remove Noncanonical Unannotated chimSegmentMin 10 outFilterType BySJout alignSJDBoverhangMin 1 alignIntronMin 10 alignIntronMax 1000000 alignMatesGapMax 1000000 alignEndsProtrude 10 ConcordantPair limitOutSJcollapsed 5000000". Reads overlapping Ensembl gene models (Salmo_salar-GCA_905237065.2) were quantified using R (version 4.2) package "Rsubread" (65) and the function "featureCounts", specifying the parameters "count ChimericFragments=FALSE, countReadPairs=TRUE, countMultiMappingReads=TRUE, fraction=TRUE," and "primaryOnly=TRUE". A principal component analysis was then applied to normalized read counts from "DESeq2 vst" (66) method to investigate major axes of variation in the expression data.

Gene-level counts were used for testing of differential expression between *vgll3* genotypes, sampling dates (as continuous variable), and their interaction using a Generalized Linear Model approach with normalization factors for unwanted variation as follows. First, genes were filtered for those expressed using "edgeR" (67) function "filterByExp" and defining treatment groups. A "naive" GLM was defined with *vgll3* genotype, sampling date (scale-normalized), and their interaction, using "DESeq2". Non-DEGs based on Wald's test and a significance threshold of 0.05 (adjusted for multiple testing) between genotypes, sampling dates, and their interaction were extracted using contrasts and the "results" function. The non-DEGs were designated as "negative control" genes to be used in estimating unwanted variation in the data. For calculating factors that describe unwanted variation, RNA-seq counts were adjusted for differential sequencing depth using the "EDASeq" package (68) and the function "between Lane Normalization". RNA-seq read counts for negative control genes were then used to calculate a sample-specific adjustment factors (weights) corresponding to the maturity trajectory of each sample using the "RUVg" method (36) and parameter "k = 2". A final GLM was defined with *vgll3* genotype, sampling date (scale-normalized), their interaction, and the two normalization factors from "RUVg". The model fit for all factors were tested against null models using the "DESeq" function with parameter "test="LRT"". All factors were retained in the final model. DEGs were extracted using Wald's test and contrasts similar to the naive analysis above.

Gene annotations for gene sets were fetched using NCBI Batch Entrez search and GO term overrepresentation was tested against expressed genes with "AnnotationHub" (69) (snapshot date 2023-04-24) function "enrichGO" and a False Discovery Rate threshold of 0.1.

Gene coexpression networks were constructed using "WGCNA" (70). Gene-level counts were normalized using RUV weights and "vst" function of "DESeq2". Gene coexpression modules were identified using the automatic network construction and module detection function "blockwiseModules" in "WGCNA R" package. A soft-thresholding power of 10 was selected based on best fit of scale-free topology while conserving a moderately large module size. Additional parameters for "blockwiseModules were defined as maxBlockSize=10000, TOMType = "signed", minModuleSize = 25, reassignThreshold = 0, and mergeCutHeight = 0.3". Module neighborhood statistics were analyzed and visualized using the "igraph" R package (71). Module genes were tested for overrepresentation of GO terms as described above.

Difference in rank-transformed module eigengene expression between *vgll3* genotype was tested using a linear model ("lm" function) and *p*-values were corrected using the FDR method of ref. 72. Correlation between module eigengenes and sampling date was calculated using "bicor" function in "WGCNA". Significance of associations was tested using "corPvalueStudent" function and corrected using the FDR method of ref. 72.

ChIPmentation Analysis. ChIPmentation reads were quality-trimmed using "fastp" and parameters "--low_complexity_filter --trim_front1=2 --trim_front2=2". Replicate sequencing runs for the same samples were trimmed individually and trimmed reads were combined into a single "fastq" file. Reads were then aligned to the Atlantic salmon genome downloaded from Ensembl (Salmo_salar-GCA_905237065.2) using "Bowtie2" (73) and parameters "--very-sensitive --maxins 1500 --end-to-end". "Samtools view" was used to filter for primary alignments with mapping quality score over 20 ("-F 256 -q 20"). "Picard MarkDuplicates" (74) was used to identify and remove duplicate reads. ChIPmentation fragment coverage was combined for all samples and for replicate samples across genotypes using "samtools merge". Combined ChIPmentation coverage was downsampled to be equal in genotypes using "Picard DownsampleSam". We used "MACS3" (https://github.com/macs3-project/MACS) with parameters "--broad --broad-cutoff 0.1" to identify genome regions associated with H3K27ac, H3K4me3, and VGLL3 (for all samples and for *vgll3* genotypes separately). We further validated the VGLL3 peak calls for genotypes by comparing to a set of peaks identified using MACS3 and a lenient *P*-value threshold of 1e-3 for individual samples, which were then summarized over biological replicates of genotypes by MSPC (43). Custom R code and "bedtools intersect" (75) were used to filter out peaks overlapping 1 kilo base windows with top 1% of sequencing coverage of control (tagmentation) libraries in order to exclude problematic genome regions. Peaks were tested for overrepresented transcription factor binding motifs using HOMER (76) and the function "findMotifsGenome.pl", specifying the parameters "-size given" and "-mset vertebrate". Peaks were filtered for those with log fold-change compared to input >2 and overlaps of peaks were analyzed in R. Peaks were assigned to closest expressed transcripts in R using the list of expressed genes from RNA-seq analysis. Significance of overlaps of peaks/chromatin features was tested based on the approach in ref. 77 and implemented by means of reservoir sampling in "poverlap.py" (https://github.com/brentp/poverlap/tree/master). Transcripts were assigned to genes using "biomaRt" function "getBM" (78) and annotations were fetched using NCBI Batch Entrez. GO overrepresentation of associated genes was tested as above against background sets defined in the main text.

Data, Materials, and Software Availability. Sequence data are publicly available in NCBI Sequence Read Archive (PRJNA1042649). Output files for gene expression and chromatin features as well as all R code used in analysis are publicly available in Dryad (https://doi.org/10.5061/dryad.vhhmgqp1g)

ACKNOWLEDGMENTS. We thank Ana Lindeza, Claudius Kratochwil, Frédéric Guillaume, the editor and reviewers for comments on the manuscript, Nikolai Piavchenko and Noora Parre for fish husbandry and Seija Tillanen for help with lab work. Funding was provided by Academy of Finland (Grant Nos. 314254, 314255, 327255, and 342851), the University of Helsinki, and the European Research Council under the European Articles Union's Horizon 2020 and Horizon Europe research and innovation programs (Grant Nos. 742312 and 101054307). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Author affiliations: [a]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki 00014, Finland; [b]Genomics Division, Faculty of Biosciences and Aquaculture, Nord University, Bodø 8026, Norway; [c]Natural Resources Institute Finland (LUKE), Helsinki 00790, Finland; [d]Department of Aquaculture and Fish Biology, Hólar University, Sauðárkrókur 551, Iceland; and [e]Institute of Biotechnology, Helsinki Institute of Life Sciences, University of Helsinki, Helsinki 00014, Finland

1. R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).
2. H. A. Orr, The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
3. C. C. Steiner, J. N. Weber, H. E. Hoekstra, Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* **5**, e219 (2007).
4. M. R. Jones *et al.*, Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**, 1355–1358 (2018).
5. P. Andrade *et al.*, Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5633–5642 (2019).
6. L. Tietgen *et al.*, Fur colour in the Arctic fox: Genetic architecture and consequences for fitness. *Proc. R. Soc. B* **288**, 20211452 (2021).
7. P. F. Colosimo *et al.*, Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
8. Y. Chan *et al.*, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302–305 (2010).
9. B. M. Horton *et al.*, Estrogen receptor α polymorphism in a species with alternative behavioral phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1443–1448 (2014).
10. S. Lamichhaney *et al.*, Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016).
11. C. K. Hu *et al.*, cis-Regulatory changes in locomotor genes are associated with the evolution of burrowing behavior. *Cell Rep.* **38**, 110360 (2022).
12. S. Xiong *et al.*, Enhanced lipogenesis through Ppary helps cavefish adapt to food scarcity. *Curr. Biol.* **32**, 2272–2280.e6 (2022).
13. S. E. Johnston *et al.*, Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93–95 (2013).
14. N. J. Barson *et al.*, Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**, 405–408 (2015).
15. J. E. Hess, J. S. Zendt, A. R. Matala, S. R. Narum, Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proc. Biol. Sci.* **283**, 20153064 (2016).
16. D. J. Prince *et al.*, The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci. Adv.* **3**, e1603198 (2017).
17. A. J. Veale, M. A. Russello, An ancient selective sweep linked to reproductive life history evolution in sockeye salmon. *Sci. Rep.* **7**, 1747 (2017).
18. D. E. Pearse *et al.*, Sex-dependent dominance maintains migration supergene in rainbow trout. *Nat. Ecol. Evol.* **69**, 1–12 (2019).
19. A. Woronik *et al.*, A transposable element insertion is associated with an alternative life history strategy. *Nat. Commun.* **10**, 5757 (2019).
20. A. Heyland, T. Flatt, *Mechanisms of Life History Evolution: The Genetics and Physiology of Life History Traits and Trade-Offs* (OUP, Oxford, UK, 2011).
21. J. Erkinaro *et al.*, Life history variation across four decades in a diverse population complex of Atlantic salmon in a large subarctic river. *Can. J. Fish Aquat. Sci.* **76**, 42–55 (2019).
22. F. Ayllon *et al.*, The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. *PLoS Genet.* **11**, e1005628 (2015).
23. J.-P. Verta *et al.*, Cis-regulatory differences in isoform expression associate with life history strategy variation in Atlantic salmon. *PLoS Genet.* **16**, e1009055 (2020).
24. P. V. Debes *et al.*, Polygenic and major-locus contributions to sexual maturation timing in Atlantic salmon. *Mol. Ecol.* **30**, 4505–4519 (2021).
25. E. R. Åsheim *et al.*, Atlantic salmon (*Salmo salar*) age at maturity is strongly affected by temperature, population and age-at-maturity genotype. *Conserv. Physiol.* **11**, coac086 (2023).
26. A. H. House, P. V. Debes, J. Kurko, J. Erkinaro, C. R. Primmer, Genotype-specific variation in seasonal body condition at a large-effect maturation locus. *Proc. R. Soc. B* **290**, 20230432 (2023).
27. T. Aykanat *et al.*, Life-history genomic regions explain differences in Atlantic salmon marine diet specialization. *J. Anim. Ecol.* **89**, 2677–2691 (2020).
28. J. M. Prokkola *et al.*, Genetic coupling of life-history and aerobic performance in Atlantic salmon. *Proc. R. Soc. B* **289**, 20212500 (2022).
29. P. B. Bangura *et al.*, Linking vgll3 genotype and aggressive behaviour in juvenile Atlantic salmon (*Salmo salar*). *J. Fish Biol.* **100**, 1264–1271 (2022).
30. P. T. Niemelä *et al.*, Life-history genotype explains variation in migration activity in Atlantic salmon (*Salmo salar*). *Proc. R. Soc. B* **289**, 20220851 (2022).
31. X. Liu, Y. I. Li, J. K. Pritchard, Trans effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034.e6 (2019).
32. A. L. Tyler, F. W. Asselbergs, S. M. Williams, J. H. Moore, Shadows of complexity: What biological networks reveal about epistasis and pleiotropy. *Bioessays* **31**, 220–227 (2009).
33. M. Sinclair-Waters *et al.*, Dissecting the loci underlying maturation timing in Atlantic salmon using haplotype and multi-SNP based association methods. *Heredity* **129**, 356–365 (2022).
34. E. Kjaerner-Semb *et al.*, Vgll3 and the Hippo pathway are regulated in Sertoli cells upon entry and during puberty in Atlantic salmon testis. *Sci. Rep.* **8**, 1912 (2018).
35. J. Kurko *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. *G3* **10**, 235–246 (2019).
36. D. Risso, J. Ngai, T. P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
37. J. R. B. Perry *et al.*, Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
38. K. L. Parker, B. P. Schimmer, Steroidogenic factor 1: A key determinant of endocrine development and function. *Endocr. Rev.* **18**, 361–377 (1997).
39. W. C. Jordan, A. F. Youngson, Genetic protein variation and natural selection in Atlantic salmon (*Salmo salav*, L.) parr. *J. Fish Biol.* **39**, 185–192 (1991).
40. W. C. Jordan, A. F. Youngson, J. H. Webb, Genetic variation at the malic enzyme-2 locus and age at maturity in sea-run Atlantic salmon (*Salmo salar*). *Can. J. Fish Aquat. Sci.* **47**, 1672–1677 (1990).
41. E. Stashi, B. York, B. W. O'Malley, Steroid receptor coactivators: Servants and masters for control of systems metabolism. *Trends Endocrinol. Metab.* **25**, 337–347 (2014).
42. M. P. Creyghton *et al.*, Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).
43. V. Jalili, M. Matteucci, M. Masseroli, M. J. Morelli, Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics* **31**, 2761–2769 (2015).
44. J. Pascual *et al.*, Hippo reprograms the transcriptional response to Ras signaling. *Dev. Cell* **42**, 667–680.e4 (2017).
45. R. C. Gimple, X. Wang, RAS: Striking at the core of the oncogenic circuitry. *Front. Oncol.* **9**, 965 (2019).
46. F. Sanchez-Vega *et al.*, Oncogenic signaling pathways in the cancer genome atlas. *Cell* **173**, 321–337.e10 (2018).
47. J. P. Thiery, H. Acloque, R. Y. J. Huang, M. A. Nieto, Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
48. N. Hori *et al.*, Vestigial-like family member 3 stimulates cell motility by inducing high-mobility group AT-hook 2 expression in cancer cells. *J. Cell Mol. Med.* **26**, 2686–2697 (2022).
49. E. Simon, C. Faucheux, A. Zider, N. Theze, P. Thiebaud, From vestigial to vestigial-like: The *Drosophila* gene that has taken wing. *Dev. Gen. Evol.* **226**, 297–315 (2016).
50. M.-C. Meinsohn, O. E. Smith, K. Bertolin, B. D. Murphy, The orphan nuclear receptors steroidogenic factor-1 and liver receptor homolog-1: Structure, regulation, and essential roles in mammalian reproduction. *Physiol. Rev.* **99**, 1249–1279 (2019).
51. L. Zhao *et al.*, Central nervous system-specific knockout of steroidogenic factor 1 results in increased anxiety-like behavior. *Mol. Endocrinol.* **22**, 1403–1415 (2008).
52. N. Grgurevic, T. Büdefeld, E. F. Rissman, S. A. Tobet, G. Majdic, Aggressive behaviors in adult SF-1 knockout mice that are not exposed to gonadal steroids during development. *Behav. Neurosci.* **122**, 876–884 (2008).
53. G. Majdic *et al.*, Knockout mice lacking steroidogenic factor 1 are a novel genetic model of hypothalamic obesity. *Endocrinology* **143**, 607–614 (2002).
54. T. Shu, G. Zhai, A. Pradhan, P.-E. Olsson, Z. Yin, Zebrafish cyp17a1 knockout reveals that androgen-mediated signaling is important for male brain sex differentiation. *Gen. Comp. Endocrinol.* **295**, 113490 (2020).
55. F. Picard *et al.*, SRC-1 and TIF2 control energy balance between white and Brown adipose tissues. *Cell* **111**, 931–941 (2002).
56. M. Sinclair-Waters *et al.*, Beyond large-effect loci: Large-scale GWAS reveals a mixed large-effect and polygenic architecture for age at maturity of Atlantic salmon. *Genet. Sel. Evol.* **52**, 9 (2020).
57. G. Stampfel *et al.*, Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**, 147–151 (2015).
58. F. Reiter, S. Wienerroither, A. Stark, Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* **43**, 73–81 (2017).
59. A. J. Jensen *et al.*, Large-effect loci mediate rapid adaptation of salmon body size after river regulation. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2207634119 (2022).
60. S. C. Stearns, *The Evolution of Life Histories* (Oxford University Press, 1992).
61. T. Aykanat, M. Lindqvist, V. L. Pritchard, C. R. Primmer, From population genomics to conservation and management: A workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon, *Salmo salar*. *J. Fish Biol.* **89**, 2658–2679 (2016).
62. E. C. Anderson, 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 (US Section) 46 (2010).
63. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
64. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
65. Y. Liao, G. K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47 (2019).
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

67. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
68. D. Risso, K. Schwartz, G. Sherlock, S. Dudoit, GC-content normalization for RNA-seq data. *BMC Bioinf.* **12**, 480 (2011).
69. M. Morgan, L. Shepherd, AnnotationHub: Client to access AnnotationHub resources (R package, 2021). https://bioconductor.org/packages/release/bioc/html/AnnotationHub.html. Accessed 26 April 2023.
70. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinf.* **9**, 559 (2008).
71. G. Csardi, T. Nepusz, The Igraph software package for complex network research. *Int. J. Complex Syst.* **1695**, 1–9 (2006).
72. J. D. Storey, A direct approach to false discovery rates. *J. R. Stat. Soc. B* **64**, 479–498 (2002).
73. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
74. Broad Institute, Picard toolkit, version 3.0.0. https://broadinstitute.github.io/picard/. Accessed 18 April 2023.
75. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
76. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
77. N. Haiminen, H. Mannila, E. Terzi, Determining significance of pairwise co-occurrences of events in bursty sequences. *BMC Bioinf.* **9**, 336 (2008).
78. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).