Integrated transcriptome, DNA methylome and chromatin state accessibility landscapes reveal regulators of Atlantic salmon maturation

Amin R. Mohamed [1, 2], Marina Naval-Sanchez[1], Moira Menzies[1], Bradley Evans[3], Harry King[4], Antonio Reverter[1], James W. Kijas [1]*

Affiliations

[1] CSIRO Agriculture and Food, Queensland Bioscience Precinct, St Lucia, Brisbane, Queensland 4067, Australia

[2] Zoology Department, Faculty of Science, Benha University, Benha 13518, Egypt

[3] Tassal Operations Pty Ltd, Hobart, Tasmania 7001, Australia

[4] CSIRO Agriculture and Food, Hobart, Tasmania 7004, Australia

* For correspondence: James.Kijas@csiro.au

Running title: Multiomics profiling during onset of salmon maturation

Abstract

Despite sexual development being ubiquitous to vertebrates, the epigenetic mechanisms controlling this fundamental transition remain largely undocumented in many organisms. Through whole-methylome, whole-transcriptome and chromatin landscape sequencing, we discovered global control mechanisms as well as specific regulators of sexual maturation in Atlantic salmon. This large integrated study was based on an experimental time course that successfully sampled the period when Atlantic salmon commence their trajectory towards sexual maturation. Co-analysis of DNA methylome and gene expression changes revealed chromatin remodelling genes *arid1b* and *smarca2* were both significantly hypermethylated and upregulated in the ovary during the onset of maturation. We also observed changes in chromatin state landscape occurred early in the transition and were strongly correlated with fundamental remodelling of gene expression. Finally, we integrated our multiomics datasets to identify *trim25* and *znf423* as key regulators in the pituitary that underwent 60 fold change in connectivity during the transition to sexual maturation. The study provides a comprehensive view of the spatiotemporal changes involved in a complex trait and opens the door to future efforts aiming to manipulate puberty in an economically important aquaculture species.

Main

Epigenetic regulation of gene expression influences a vast spectrum of complex traits, with examples spanning the onset and severity of human disease, developmental transitions during growth and the expression of ecologically and economically relevant traits across the animal kingdom. Our understanding of the epigenetic contributions to trait variation remains low in comparison to causative genes derived from approaches such as genome wide association studies. However, the relatively recent development of sophisticated sequence-based assays for the detection of chromatin state changes and methylation status have enabled the landmark development of genome wide maps of regulatory elements in human [1,2,3], mouse [4,5,6] and other model organisms such as the fruit fly (*Drosophila melanogaster*) and the nematode (*Caenorhabditis elegans*) [7,8]. These have provided the impetus for a plethora of research focussed on understanding epigenetic mechanisms and their role regulating gene expression.

Sexual maturation is a fundamental transition ubiquitous to vertebrates and provides a model for the study of epigenetic regulation. The key tissues are known given the reproductive cycle is regulated by activation across the brain, pituitary, gonadal (BPG) axis in organisms spanning mammals to teleost fish. Further, upon maturation these tissues undergo known and often profound transcriptomic remodelling providing a large dynamic range to increase the likelihood of identifying regulatory networks [9,10,11,12]. The genetic architecture of sexual maturation has been extensively studied using association studies in both wild and farmed Atlantic salmon populations [13,14,15,16]. Despite the importance of sexual maturation as a trait of

interest it can be difficult to study, as the timing of onset varies widely in response to both genetics and environmental factors and occurs prior to measurable phenotypic change. To overcome this, we chose to investigate sexual maturation in Atlantic salmon where photoperiod manipulation in an experimental system can be used to synchronise animals and access tissues across the time period when animals first commit to the onset of puberty. We also chose a multiomics approach, which has the power to identify the control mechanisms underpinning complex traits [17,18]. We describe changes in gene expression, DNA methylation and chromatin accessibility to identify epigenetic mechanisms associated with maturation in a commercially important aquaculture species.

Results

Initiation of Atlantic salmon sexual maturation and a multiomic workflow

Fish were managed in a tank based experimental system to facilitate a long-light photoperiod regime known to stimulate the onset of sexual maturation (Fig. 1a) [19,20]. Fish from a single management group were sacrificed at a timepoint immediately before initiation of the long-light regime (throughout referred to as T1) and at three timepoints afterwards (T2, T3 and T4). An increase in gonadal somatic index (GSI) of sampled fish across the time course confirmed an active response to the long photoperiod (Fig. 1b). Significant increases were observed only at T4 (t-test P-value = 0.021). Tissues from the brain – pituitary – gonad axis and liver were sampled at each timepoint to form the basis of a multiomics workflow for data generation and

integrative analysis spanning the transcriptome, DNA methylome and chromatin state datatypes (Fig. 1c; Supplementary Fig. S1).

Maturation leads to significant transcriptome changes

We sequenced messenger RNA (mRNA) from four biological replicates of each tissue before and after the onset of maturation. A total of 3.2 billion paired-end reads were mapped against the Atlantic salmon reference genome with 72% mapping efficiency to create an average depth of 50 million reads per library (Supplementary Table 1). Consistency across biological replicates within each timepoint was high for each tissue except for brain (see Supplementary Results; Supplementary Fig. S2). To begin characterization of transcriptomic remodelling, we compared gene expression levels in three post-maturation samples (T2, T3 and T4) to the control T1 (Fig. 2a). The pituitary gland showed comparatively few transcriptomic responses that involved 543 differential expressed genes (DEGs) (Supplementary Table S2). In contrast, more widespread remodelling was observed in both ovary (5,993 DEGs) and liver (9,541 DEGs, adjusted $P < 0.05$ (Fig. 2a, Supplementary Tables S3, S4). The number of DEGs increased with elapsed time following the onset of the long light photoperiod for the two BPG axis tissues (pituitary and ovary). Of these, the ovary underwent the most dramatic remodelling over time with 403, 1,709 and then 3,497 DEGs observed at timepoints T2, T3 and T4 respectively. This increasing trajectory of differential gene expression, coupled with the elevated GSI following the light stimuli (Fig. 1b), strongly suggests the experimental approach successfully initiated the onset of maturation.

Upregulation of pituitary hormones

The pituitary is expected to play a key role in the early triggers for maturation onset (21). Hierarchical clustering of the pituitary DEGs revealed two distinct groups, one of upregulated genes (n=333; 60% of pituitary DEGs) and one of downregulated genes (n=125). The upregulated group showed significant Gene Ontology (GO) enrichment for maturation-related functions including *G protein−coupled receptor signalling* and *hormone activity* (Supplementary Fig. S4 and Supplementary Table S5). Pituitary DEGs were involved in several reproduction-related functions such as steroidogenesis (*cyp17a1*), hormone receptors (*oxtr*, *pgr, dop1r1, galr1*), genes coding for pituitary hormones (*gh2*, *gthb1*, *gthb2, glha2*), retinoic acid (RA) signalling (*cyp26b1*, *rhd8-1*/*2*) and sex-related transcription factors (TFs) (several sox genes and *gata2*). The upregulation of gonadotropins subunits is highly significant (Fig. 2b). This included both *gthb1* and *gthb2* that encode the gonadotropin subunits beta-1 and 2 as well as *glha2* that encodes the glycoprotein hormone alpha chain. Together, these form the heterodimeric gonadotropins *gth-1* and *gth-2* that have previously been shown to stimulate gonadal growth in the juvenile stages of both rainbow trout and coho/chum salmon [22,23,24]. Further, physicochemical characterization of the salmon gonadotropins indicate they are functionally related to follicle stimulating hormone (fsh) and luteinizing hormone (lh) in vertebrates (reviewed by [25]). We find *glha2* (the common subunit present in gonadotropins) was consistently upregulated in the pituitary throughout the experiment (Supplementary Fig. S3). Our results directly confirm the action of these key pituitary hormones in the onset of maturation. To characterize the transcriptomic remodelling occurring in the ovary and liver, we assessed the DEG sets for GO enrichment. Upregulated genes in the ovary revealed processes related to *cell*

adhesion, *immune/inflammatory response* and *development* (Supplementary Fig. S5; Supplementary Table S6), while gene families involved in *organic acid metabolic processes* and *mitochondrial transport* were enriched among liver upregulated genes (Supplementary Fig. S7 and Supplementary Table 7). Maturation-related functions including steroidogenesis, hormonal receptors and follicular development were identified in ovary (see Supplementary Results; Supplementary Fig. S6).

DNA methylome maps of three Atlantic salmon tissues

To investigate the regulatory mechanisms controlling differential expression, the first genome-wide CpG methylation maps were developed for Atlantic salmon using whole-genome bisulfite sequencing (WGBS). Methylome data was collected from two biological replicates at the terminal time points (T1 and T4, Fig. 1c) from three tissues (pituitary, ovary and liver), generating 2.6 billion paired-end uniquely mapped reads (average coverage of 11x) (Supplementary Table S8). We found a genome-wide methylation rate of 81% per sample (Supplementary Fig. S8; Supplementary Table S8), which is similar to the rate observed in vertebrate genomes (60-90%) [26,27]. Methylome data was assessed based on coverage, read mapping and consistency between biological replicates (see Supplementary Results; Supplementary Table S8; Supplementary Fig. S8c, d). Among the different dinucleotide contexts, CpG methylation contributed the vast majority (~ 99.5% on average) compared to CHH or CHG methylation which were excluded from further analysis (Supplementary Fig. S8a; Supplementary Table S8).

By comparing the CpG methylation patterns between T4 and T1 samples (Fig. 3a), we identified 1902, 2982 and 1606 differentially methylated regions (DMRs) in the

pituitary gland, ovary and liver respectively (Fig. 3b; Supplementary Fig. 9a, b; Supplementary Tables 9, 10, 11). The average length of DMRs was short (251 bp) and their distribution was both genome wide (Fig. 3b) and highly non-random, with 52% found to overlap protein coding genes and another 18% located within 5 kb upstream or downstream (Fig. 3c; Supplementary Fig. 9c). The location of DMRs was also strongly tissue specific, with few regions shared between tissue pairs and only 8 found in all three tissues (Supplementary Fig. 9b). Next, we investigated the directionality of DMRs across tissues and found approximately equal rates of hyper-methylation (increased methylation in T4) and hypo-methylation (decreased methylation in T4) in both the pituitary and liver. Strikingly, the majority of DMRs in ovary were hyper-methylated (2175 DMRs or 73%), independent of their genomic location (genic regions, promotors, 5 kb downstream or intergenic) (Supplementary Fig. 9d). This is consistent with DNA methylation having important roles during epigenomic reprograming in embryo and stem cells development [28], compared to highly stable methylomes in somatic cells [29]. This pronounced skew towards increased methylation occurred in the tissue with both the highest total number of observed DMRs and the largest increase in upregulation of gene expression. Of the tissues investigated, the ovary also undergoes the most radical physiological change during maturation as it transforms via vitellogenesis and oocyte development in preparation for egg release during spawning.

Differentially methylated genes serve key roles in maturation

The gene catalogue present within differentially methylated regions (DMRs) was assessed for their function in relation to the trait. The majority of differentially methylated genes (DMGs) in ovary were hypermethylated (n=1165; 74%)

(Supplementary Fig. S9d) and significantly enriched for three biological process (GO-BP), one cellular component (GO-CC) and 24 molecular function (GO-MF) terms including those with maturation-related functions such as *semaphorin/glutamate receptor activity* (Supplementary Fig. S10; Supplementary Table S12). Among these genes, at least 33 genes have demonstrated roles in the biology of maturation including follicular development (*plexnb1*, *sema4f* coding for Plexin-B1 and Semaphorin-4f) [30] and the control of gonadotropin-releasing hormone excitability (*grm8* coding for the glutamate receptor 8) [31]. We plotted the normalised expression of these candidates from samples collected at T1 and T4 and found that the majority were upregulated at T4, consistent with their hyper-methylated status (Supplementary Fig. S11).

Co-analysis of DNA methylome and transcriptome reveals role for chromatin remodelling during maturation

The identification of significant transcriptional and methylation changes allowed us to explore the dynamic between these two processes by assessing the overlap of genes declared as both DEG and DMG. The overlap was low and non-significant for liver (38 / 616 or 6% of DMGs were also DEGs) and pituitary (11 / 762 or 1.4% of DMGs were DEGs) (Supplementary Fig. S9e). However, 195 or 14% of ovary DMGs (195 / 1357) were also differentially expressed, a number that exceeded random expectation in 83.8% of 1000 permutations tests (Fig. 3d). This suggests changes in methylation status may directly control gene expression in this subset of genes. If true, we would expect to see correspondence between the directionality of the expression and methylation changes. This appeared to be the case, as 82% of upregulated genes (148 / 179; Binomial P-value = 8.727E-20) were hyper-

methylated at T4 relative to T1 (Fig. 3e), matching the classical expectation of gene body methylation mediated control of gene expression [32,33]. The 148 genes were enriched for 3 GO-CC terms related to chromatin remodelling complexes (SWI/SNF and nBAF) (Fig. 3f; Supplementary Table S13), and the associated genes displaying coordinated expression and methylation status (Fig. 3g). For example, the *arid1b* gene encodes AT-rich interactive domain-containing protein 1B and *smarca2* encodes the global transcription activator snf2l2. Both proteins are involved in chromatin remodelling as they are core components of the SWI/SNF remodelling complexes[34] that carry out enzymatic change to chromatin structure by altering DNA-histone contacts [35]. This opens the possibility they act as key control points, to regulate a wide array of other genes during ovarian development. A final examination was performed to search for evidence of a generalised and genome wide association between gene expression levels and methylation status. Expression for all genes with either differential gene body or promoter methylation is shown in Supplementary Fig. S12. We found no correlation for either comparison, consistent with previous studies that reported weak correlation between DNA methylation and gene expression in humans [36,37] and more recently in fish [38]. Taken together, the results confirmed that while methylation alone does not control genome-wide patterns of gene expression, it plays a key role upregulating a defined set of genes during the maturation process.

Early and stable changes in the chromatin state

To deepen the characterization of the epigenomic features during maturation, we performed ATAC-seq (assay for transposase-accessible chromatin sequencing[39]) to produce genome-wide maps of chromatin accessibility changes. ATAC-seq was

249 performed for multiple tissues and peak enrichment around transcription start sites
250 used as the key quality control metric (TSS, Supplementary Figure S13). Following
251 data pruning, we took 12 liver libraries (3 replicates across all 4 timepoints) and a
252 total of 699 million uniquely mapped paired-end reads (Supplementary Table S14)
253 forward into joint analysis with RNA-seq and WGBS data. Principal component
254 analysis (PCA) of the 12 ATAC-seq samples revealed the T1 replicates formed a
255 tight cluster positioned separately from the T2 – T4 samples, which were less well
256 distinguished from each other (Fig. 4a). The first two principal components
257 accounted for 50% of variation, which is less than the comparative analysis of the
258 same tissue using RNA-seq (66% Fig. 4a). To characterise changes in chromatin
259 state following long light initiation, we defined differentially accessible regions
260 (DARs) where mapping counts differed significantly between T1 and other time
261 points. This revealed a strong early remodelling in the chromatin state landscape, as
262 most DARs were observed at T2 (n=1501) before decreasing in stepwise fashion at
263 T3 (n=477) and T4 (n=148; Fig.4b; Supplementary Table S15). The direction of
264 change was approximately balanced between DARs with increased and decreased
265 accessibility, broadly matching the balance between up and down regulated global
266 gene expression changes observed for liver (Fig. 2a). We next asked if the early
267 changes in chromatin state persisted throughout the time course using hierarchical
268 clustering. The majority of DARs (n=1036 or 57%) exhibit reduced accessibility at T2
269 compared with T1 and subsequently remained unchanged at later time points (Fig.
270 5a; Supplementary Figure S14). Similarly, regions that gained accessibility at T2
271 (n=696 or 38%) also remained unchanged at later timepoints. This left less than 10%
272 of DARs (n=99) that displayed an oscillating pattern following the onset of the
273 maturation. Together, this revealed the ATAC-seq signatures were predominantly

stable chromatin state changes, as opposed to pulsatile epigenomic changes that snapped back after a small number of days or weeks.

Differential chromatin accessibility strongly correlates with bidirectional regulation of global gene expression

To begin exploring the relationship between chromatin accessibility changes and gene expression, we first mapped the genomic location of DARs and found the majority were located in genes (65%) or within 5 kb upstream (6%) or 5 kb downstream (4%), affirming the quality of the ATAC-seq dataset. We also detected a quarter of DARs (25%) located on average 36 kb distal to their nearest gene, a low proportion in comparison to domesticated terrestrial livestock [40]. Starting with the subset of DARs located in genes (exons and introns) we found the proportion of variation in gene expression explained by chromatin accessibility changes was high (Fig. 4c, d). For example, chromatin state changes at T2, compared with T1, explained 56% of the variation in gene expression using linear regression. The dynamic was bidirectional, with accessibility changes associated with both up and down regulation of global gene expression, and strongest at the early timepoint T2 (Fig. 4c). We repeated the analysis for DARs located within 5 kb of transcription start sites to assess the strength of association with physically proximal putative cis-regulatory elements (CREs). These had even higher association, explaining approximately 60% of the variation in global gene expression (Supplementary Figure S15). Together, this clearly demonstrated chromatin state changes played a dominant role in directing global changes in gene expression.

298 **Cis-regulatory elements regulate metabolism genes via chromatin state**
299 **changes**

300 To examine the biological consequence of chromatin state changes, we focused on
301 CREs given their established role on transcriptional regulation via transcription
302 factors (TFs) binding (41). We focussed on the subset of CREs that underwent a
303 change in accessibility during the time course to evaluate i) the expression behaviour
304 of their closet gene; ii) the biological function of those genes, and iii) any enrichment
305 for transcription factor binding sites. We found a small subset of CREs (n = 65)
306 underwent increased accessibility early in the time course and the majority (n = 46;
307 79%' X^2 $p < 8.028^{-08}$) upregulated their nearest gene in a tightly coordinated manner
308 (Fig. 5a – c). It also appears CREs more tightly controlled the downregulation of
309 genes compared to DARs located in gene bodies, downstream regions or within
310 intergenic regions (Fig. 5b, red box; Supplementary Table S16). The gene set
311 associated with coordinated up regulation (Fig. 5c) exhibited significant GO
312 enrichment related to lipid metabolism and energy metabolism (AAcyl-CoA
313 biosynthesis) (Fig. 5d). Acyl-CoA are coenzymes involved in energy synthesis,
314 consistent with the expectation of liver function through an energetically costly
315 transition such as maturation.
316
317 To provide a fine-scale view of the coordinated response, a 15 kb region of Ssa15
318 spanning the CRE and exons of a gene that regulates hepatic lipid metabolism
319 (*hmgcr*) [42] is provided in Figure 5e. The final analysis used HOMER to search for TF
320 motifs that bind master regulators driving gene transcription [43]. Specifically, we
321 computed the enrichment of TF motifs in CREs that gained chromatin accessibility
322 against a background that remained inaccessible. This revealed significant

enrichment of 13 motifs, corresponding to the preferred binding sites of specific transcription factors present in 29 - 58% of targets following the onset of maturation (Supplementary Figure S16). Among these, the most significant motif matched the global transcriptional regulator E3 Ubiquitin-Protein Ligase CNOT4 that regulates essentially every aspect of gene expression, from mRNA synthesis to protein destruction including the degradation of RNAPII [44]. The results strongly suggest that chromatin state changes at CREs directly control gene expression in liver and upregulate energy metabolism genes via changes in TF activity.

Multiomics data integration using gene regulatory networks

The final component of our analysis sought to co-analyse all available data to infer gene regulatory networks (GRNs) responsible for the onset of maturation. GRNs provide a platform for integrating multiomic data and can be used to characterize the dynamics of perturbations during biological transitions such as puberty and other complex traits [45,46,47,48,49]. Here, we used the approach to co-analyse genes with evidence of differential behaviour using seven categories that included expression (DEGs), changed methylation at gene bodies (DMGs) or promotors (DMPs) and differential chromatin accessibility (DACs). To focus the analysis towards investigation of key regulators, we also performed regulatory impact factor (RIF) analysis. This used co-expression correlation between TFs and their target differentially expressed genes to identify 305 significant regulators (Supplementary Table S17; Supplementary Figure S17). Of the seven categories, the majority of 1,858 genes prioritised for GRN construction were DEG (n = 1,400) or DMG (n = 700). The overlap between categories, for example where genes were both DEG and DMG (n = 442), is given in Fig 6.a. The gene set showed significant GO

348 enrichment (1 GO-CC and 8 GO-MF terms) to hormone activity and steroid hormone
349 receptor activity (Fig. 6b) and their expression patterns showed clear tissue-specific
350 clustering (Supplementary Fig.S18a) suggesting biological relevance to the trait.
351 GRN construction using the 1,858 genes yielded 835,084 connections with a mean
352 of 449 connections per gene. For visualisation, we only considered genes with
353 significant correlations ≥ ±0.95 (929 gene with 17,708 connections) (Supplementary
354 Fig.S18b). Most network genes (N=777, ~42 %) belonged to pituitary compared to
355 33% and 25% in ovary and liver. These figures also were reflected in the number of
356 connections per tissue (Supplementary Fig.S18c). Genes with the highest change in
357 the number of connections are likely to be key regulators, and the top 20 included
358 five zinc finger proteins (Supplementary Table S18.) Two of these transcription
359 factors, znf664 and znf239, were expressed in pituitary suggesting their key role in
360 maturation onset. Interestingly, the most highly connected genes also included two
361 uncharacterised (dark) Atlantic salmon genes (106590493, 106612553) that
362 displayed maximum expression in ovary.
363
364 **Differential GRN connectivity identifies key regulatory factors**
365 Key regulators are likely to undergo substantial change in their number of
366 connections and identify gene networks driving the transition to maturation. This
367 prompted construction of separate networks using pre- and post-maturation stage
368 data, before identifying those genes that underwent the largest change in
369 connectivity (pipeline workflow is provided in Supplementary Figure S1b). For
370 visualisation purposes, we only included 10% of the most significant connections that
371 included 1,412 genes with 17,260 connections in the pre-maturation GRN and 1,310
372 genes with 22,059 connections in the post-maturation GRN (Fig. 6c). Next, we

computed the differences in the patterns among the tissues comprising the two networks (Fig.6c). The pituitary gland and ovary had the most abundance (~ 45% and 32%, respectively) of connections compared to a lower percentage of connections (~23%) in liver after maturation. We computed the differential connectivity for all genes and identified the most differentially connected genes (DCGs) (n=186 genes; 10%) (Supplementary Table S18) between the pre and post-maturation networks with more connections: 80,536 in post-maturation compared to pre-maturation network with 56,971. These were mainly expressed in pituitary (44%) and most connections involved DEGs (74%) and DMGs (47%). Finally, we identified regulators that gained the most connections post-maturation (Table 1, Fig. 6). The top ranked regulator was *trim25* (encodes a ubiquitin E3 ligase) and underwent a profound change in connectivity (from 10 to 599). *znf423* is a ZF-TF with multiple roles in signal transduction during development. It was predominantly expressed in pituitary and contributed to 3 categories in the network (DEG, TF, DAC) with multiple roles in signal transduction during development and a salmon dark gene highly expressed in ovary (Fig. 6d). Then, we focused on TFs contained among the top 10 regulators that were differential connected genes between the pre and post-maturation networks. This revealed *znf423* to be the most differentially connected TF (from 58 to 576) and Krueppel-like factor 11 (klf11) as the least differentially connected TF (from 513 to 9) (Fig. 6e; Table 1). Enrichment of motifs for a ubiquitin E3 ligase (*cnot4*) and ZNF TF (Zic) among ATAC-seq signals at CREs confirms their regulatory roles. Several studies have previously demonstrated roles for ZNF factors in controlling onset of female puberty in many species including humans [50,51,52,53].

Discussion

Identifying the biological mechanisms controlling complex traits is a sizable challenge. We designed our study on the assumption that the dynamic network of molecules coordinating the spatiotemporal changes driving sexual maturation would be inaccessible to investigation using only a single layer of “omics”. We anchored the study around the collection of tissue transcriptomes to visualize their changing circuitry across the period where Atlantic salmon commence their trajectory towards sexual maturation. Importantly, we also characterized the changing epigenomic landscape through interrogation of DNA methylation and chromatin state changes. Integration of the resulting multiomic dataset used rigorous quantitative approaches, and when performed inside the context of a defined biological transition, has given us an unprecedented ability to characterise the onset of maturation at the molecular level in a non-model species of worldwide aquaculture and ecological importance.

The availability of dimensional data allowed us to identify the dominant epigenomic changes controlling gene expression. We conclude that global changes in DNA methylation had little predictive power to explain changing gene expression beyond a small subset involved in chromatin remodelling. While methylation changes are a striking feature of embryonic development, they appear not to have been responsible for the rapid and numerous changes in gene expression documented here. Conversely, we observed high correlation between chromatin state changes and altered gene expression for the single tissue with ATAC-seq data (liver). The correlation was highest for differentially accessible regions immediately adjacent to coding genes, implicating *cis*-regulatory elements.

The identification of key genes relied on characterisation of differential behaviour using samples collected before, and after, fish were subjected to a photomanipulation trigger designed to stimulate maturation. Multiple data patterns confirm we successfully initiated early stage maturation. Increasing average gonadosomatic index demonstrates a physiological response occurred, and this was paralleled by significant global upregulation of gene expression in the ovary and a more modest remodelling of the pituitary transcriptome. Together, this provided confidence that the characterised DEG, DMG and DAC patterns are likely to successfully implicate genes directly involved in maturation. We showed upregulation of pituitary hormones including gonadotropins along with other pituitary genes involved in a range of reproduction related functions including steroidogenesis. Differentially methylated genes were enriched for follicular development and the control of gonadotropin-releasing hormone excitability. Integrated transcriptome and methylome analysis in ovary implicated chromatin remodelling genes in controlling maturation. Finally, differentially accessible CREs in liver were enriched for lipid metabolism and energy metabolism genes.

Despite the advantages of the multiomic approach used, limitations may be imposed by the range of tissues, timepoints and technical features of the assays used. For example, we treated tissues as homogenous entities in an approach that ignores the spectrum of constituent cell types and their differentiated roles that single-cell multiomic studies have begun to explore [17,54]. Further, suboptimal partitioning of the brain at dissection hampered our ability to assign the role of the hypothalamus separately from the brain stem, cerebellum and olfactory bulb. Consequently, the

448 variation in whole brain transcriptomes among replicates within timepoint was so
449 large as to prevent meaningful analysis. Finally, we were unable to generate high
450 quality chromatin state data from ovary samples despite repeated attempts. An
451 incomplete compendium of tissues and datatypes has resulted in an imperfect view
452 of the changing epigenomic landscape.
453
454 The promise of multiomic data will remain unfulfilled without methodological
455 approaches capable of identifying system perturbations associated with phenotypic
456 change. Here, we used a gene regulatory network approach which has proven
457 successful for investigation of puberty and other complex traits [47,48,49]. We identified
458 t*rim25*, a gene encoding a ubiquitin E3 ligase, as the network element undergoing
459 the most dramatic change in connectivity and strongly suggests it plays a key role.
460 Looking forwards, it is tempting to speculate that a number of the genes identified
461 represent high value targets for manipulation via gene editing in an attempt to delay
462 or ablate sexual maturation. Among the range of putative targets identified, the core
463 components of the SWI/SNF chromatin remodelling complex (*arid1B* and *smarca2*)
464 are appealing due to their ability to exert wide ranging change in gene expression.
465 The results described may therefore lead to better management of unwanted early
466 maturation within an aquaculture setting where the completion of maturation is
467 associated with reduced product quality and production inefficiencies.
468
469
470

References

1. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
2. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
3. Stunnenberg, H. G. & Hirst, M. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* **167**, 1145–1149 (2016).
4. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
5. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
6. Gorkin, D. et al. Systematic mapping of chromatin state landscapes during mouse development. Preprint at bioRxiv https://www.biorxiv.org/ content/10.1101/166652v1 (2017).
7. Henikoff, J. G., Belsky, J. A., Krassovsky, K., Macalpine, D. M. & Henikoff, S. Epigenome characterization at single base-pair resolution. *Proc. Natl Acad. Sci. USA* **108**, 18318–18323 (2011).
8. Rivera, J., Keränen, S. V. E., Gallo, S. M. & Halfon, M. S. REDfly: the transcriptional regulatory element database for *Drosophila*. *Nucleic Acids Res.* **47 (D1),** D828–D834 (2019)
9. Widmann, P. et al. A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. *BMC Genomics* **14**, 798 (2013).
10. Cánovas, A. et al. Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PLOS One* **9**, e102551 (2014).
11. Nguyen, L. T. et al. STAT6, PBX2, and PBRM1 emerge as predicted regulators of 452 differentially expressed genes associated with puberty in brahman heifers. *Front. Genet.* **9**, 87 (2018).
12. Schenk, S. et al. Combined transcriptome and proteome profiling reveals specific molecular brain signatures for sex, maturation and circalunar clock phase. *eLife* **8**, e41556 (2019).
13. Gutierrez, A.P. et al. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One* **10**, 3 (2015).
14. Barson, N.J. et al. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528(7582)**, 405–8 (2015).
15. Ayllon, F. et al. The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar L.)* males. *PLoS Genet*., **11(11)**,e1005628 (2015).

16. Mohamed, A.R. et al. Polygenic and sex specific architecture for two maturation traits in farmed Atlantic salmon. *BMC Genomics* **20**, 139 (2019).

17. Argelaguet, R. et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* https://doi.org/10.1038/s41586-019-1825-8 (2019).

18. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

19. Randall, C. F., Bromage, N. R., Duston, J. & Symes, J. Photoperiod induced phaseshifts of the endogenous clock controlling reproduction in the rainbow trout: a circannual phase-response curve. *Journal of Reproduction and Fertility* **112**:399–405 (1998).

20. King, H.R., Lee, P.S., Pankhurst, N.W. Photoperiod-induced precocious male sexual maturation in Atlantic salmon (*Salmo salar*). *Fish Physiol. Biochem*. **28**, 427–428 (2003).

21. Veldhuis, J. D. Changes in pituitary function with ageing and implications for patient care. *Nat. Rev. Endocrinol*. **9**, 205–215 (2013).

22. Suzuki, K., Kawauchi, H. & Nagahama, Y. Isolation and characterisation of two distinct gonadotropins from chum salmon pituitary glands. *General and Comparative Endocrinology* **71**, 292–301 (1988a).

23. Suzuki, K., Nagahama, Y., Kawauchi, H. Steroidgenic activities of two distinct salmon gonadotropins. *Gen. Comp. Endocrinol*. **71**, 452-458 (1988b).

24. Swanson, P., Bernard, M.G., Nozaki, M., Suzuki, K., Kawauchi, H., Dickhoff, W.W. Gonadotropins I and II in juvenile coho salmon. *Fish Physiology and Biochemistry* **7**, 169–176 (1989).

25. Levavi-Sivan, B., Bogerd, J., Mañanós, E. L., Gómez, A. & Lareyre, J. J. Perspectives on fish gonadotropins and their receptors. *Gen. Comp. Endocrinol*. **165**, 412–437 (2010).

26. Zhou, F. et al. Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature* **572**, 660–664 (2019).

27. Ortega-Recalde, O., Day, R. C., Gemmell, N. J., and Hore, T. A. Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. *Nat. Commun*. **10**, 3053 (2019).

28. Kim, M. & Costello, J. DNA methylation: an epigenetic mark of cellular memory. *Exp. Mol. Med*. **49**, e322 (2017).

29. Shipony, Z. et al. Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115–119 (2014).

30. Regev, A., Goldman, S., Shalev, E. Semaphorin-4D (Sema-4D), the Plexin-B1 ligand, is involved in mouse ovary follicular development. *Reprod. Biol. Endocrinol*. 5, 12 (2007).

546 31. Iremonger, K.J., Constantin, S., Liu, X. Herbison AE Glutamate regulation of GnRH
547 neuron excitability. *Brain Res.* **1364**, 35–43 (2010).

548 32. Neri, F. et al. Intragenic DNA methylation prevents spurious transcription initiation.
549 *Nature* **543**, 72–77 (2017).

550 33. Arechederra, M. et al. Hypermethylation of gene body CpG islands predicts high
551 dosage of functional oncogenes in liver cancer. *Nat. Commun*. **9**, 3164 (2018).

552 34. Wang, X. et al. SMARCB1-mediated SWI/SNF complex function is essential for
553 enhancer regulation. *Nat. Genet*. **49(2):**289-295 (2017).

554 35. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling
555 complexes and cancer: mechanistic insights gained from human genomics. *Sci. Adv.*
556 **1**, e1500447 (2015).

557 36. Park, J., Xu, K., Park, T., Yi, S.V. What are the determinants of gene expression
558 levels and breadths in the human genome? *Hum. Mol. Genet.* **21**, 46–56 (2012)

559 37. Lou, S. et al. Whole-genome bisulfite sequencing of multiple individuals reveals
560 complementary roles of promoter and gene body methylation in transcriptional
561 regulation. *Genome Biol.* **15**, 408 (2014).

562 38. Ryu, T., Veilleux, H. D., Donelson, J. M., Munday, P. L. & Ravasi, T. The epigenetic
563 landscape of transgenerational acclimation to ocean warming. *Nat Clim Change* **8**,
564 504–509 (2018).

565 39. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J.
566 Transposition of native chromatin for fast and sensitive epigenomic profiling of open
567 chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–
568 1218 (2013).

569 40. Foissac, S. et al . Multispecies annotation of transcriptome and chromatin structure in
570 domesticated animals. *BMC Biol*.**17**,108 (2019).

571 41. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in
572 development: lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).

573 42. De Jesus DF, et al. Parental metabolic syndrome epigenetically reprograms offspring
574 hepatic lipid metabolism in mice. *J. Clin. Invest*. **130(5)**, 2391-2407 (2020).

575 43. Marnetto, D. et al. Evolutionary rewiring of human regulatory networks by waves of
576 genome xpansion. *Am. J. Hum. Genet.* **102**, 207–218 (2018).

577 44. Jiang, H., Wolgast, M., Beebe, L., Reese, J. Ccr4–Not maintains genomic integrity by
578 controlling the ubiquitylation and degradation of arrested RNAPII. *Genes Dev*. **33**,
579 705–717 (2019).

580 45. Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based
581 approach to human disease. *Nature Rev. Genet*. **12**, 56–68 (2011).

582 46. Cánovas, A. et al. Multi-tissue omics analyses reveal molecular regulatory networks
583 for puberty in composite beef cattle. *PLOS One* **9**, e102551 (2014).

584 47. Alexandre, P. A. et al. Systems biology reveals NR2F6 and TGFB1 as key regulators
585 of feed efficiency in beef cattle. *Front. Genet*. **10**, 230 (2019).

586 48. Hawe, J. S., Theis, F. J. & Heinig, M. Inferring interaction networks from Multi-
587 Omics Data. *Front. Genet.* **12**, 1–13 (2019).

588 49. Lau, L. Y. et al. Dynamics of Gene Co-expression Networks in Time-Series Data: A
589 Case Study in *Drosophila melanogaster* Embryogenesis. *Front. Genet*. **11**, 517
590 (2020).

591 50. Demerath, E.W. et al. Genome-wide association study of age at menarche in African-
592 American women. *Hum. Mol. Genet*. **22**, 3329–46 (2013).

593 51. Perry, J.R. et al. Parent-of-origin-specific allelic associations among 106 genomic loci
594 for age at menarche. *Nature* **514**, 92–97 (2014).

595 52. Lomniczi, A. et al. Epigenetic control of female puberty. *Nat. Neurosci*. **16**, 281–289
596 (2013).

597 53. Fortes M.R.S. et al. Transcriptome analyses identify five transcription factors
598 differentially expressed in the hypothalamus of post-versus pre-pubertal Brahman
599 heifers. *Journal of Animal Science* **94**, 3693–702 (2016).

600 54. Chappell, L., Russell, A. J. C. & Voet, T. Single-cell (multi)omics
601 technologies. *Annu. Rev. Genom. Hum. Genet.* **19**, 15–41 (2018).

602

603

604

605 **Methods**

606

607 **Induction of maturation through photoperiod manipulation and tissue**

608 **sampling**

609 Animals were managed using photoperiod manipulation to synchronise the timing of

610 commitment into maturation. A population of female brood stock were used that were

611 ~ 36 months post fertilization in April 2017. The management of the animals and

612 associated timeline for sampling events is given in Fig.1a. In order to measure and

613 control for variation between individuals, 4 fish (biological replicates) at each of the

614 four time points (T1-T4) were used. The maturation status of animals (leading up to

the long day photoperiod initiation) was monitored by ultrasound. Control samples at T1 time point were collected on mid-June 2017 before induction of maturation occurred late-June 2017. Following the application of the long photoperiod, tissues were sampled at different three time points in 2 weeks intervals (T2-T4). At each sampling event, the gonadosomatic index GSI was calculated from the ovary mass as a proportion of the total body mass as follows: GSI = [ovary weight / total body weight] × 100.

RNA isolation, RNA-seq library preparation and sequencing

Tissue samples were preserved in RNA-Later at -80 °C and total RNA was isolated using RNeasy mini kit (QIAGEN) as previously described [55]. Tissues were lysed twice in 450 µL of lysis solution on a Precellys 24 homogenizer for 30s at 4.0 ms^{-1}. RNA was bound to a column and washed twice before elution with 40 µL at room temperature. RNA quantity and quality were assessed using a NanoDrop ND-1000 spectrometer, Qubit 2.0 fluorometer and Agilent 2100 bioanalyzer. Messenger RNA (mRNA) was isolated from 1 µg of total RNA. 64 RNA-Seq libraries (4 time points x 4 tissues x 4 biological replicates) were prepared using the TruSeq RNA Sample Preparation Kit (Illumina). Libraries were sequenced on Illumina Nova-Seq 6000 sequencing platform at the Australian Genome Research Facility (AGRF) in Melbourne, Australia. Sequencing produced a total of 4.4 billion individual 150 bp paired-end reads and ~ 70 million PE reads per library (Supplementary Table S1).

Transcriptomic data quality control (QC), genome mapping and read counting

Illumina reads were checked for quality using FastQC software. High quality reads (Q>30) were mapped to the Atlantic salmon genome ICSASG_v2 [56] using TopHat2

version 2.1.1 [57] with default parameters. Alignment files in BAM format were sorted by read name and converted into SAM format using SAMtools version 1.4 [58]. The Python package HTSeq version 0.7.2 [59] was applied to count unique reads mapped to exons using default parameters except for "*reverse*" with the strandedness.

Differential gene expression and clustering analyses

Raw counts were analysed using the edgeR package [60] in the R statistical computing environment to infer differential gene expression among tissues. The four tissues at the long photoperiod time points (T2, T3 and T4) were compared to the control samples at T1. *P*-values for differential gene expression were corrected for multiple testing using the Benjamini and Hochberg algorithm [61]. For further analyses of differential expression, only genes with a false discovery rate (FDR) of < 0.05 and have at least absolute $\log_2$(fold change) > 1 were considered significant. PCA was conducted on the lists of significant DEGs using normalised expression data ($\log_2$FPKM) using the function *--prin_comp* within trinity. Hierarchical clustering analysis was conducted using trinity's utility *analyze_diff_expr.pl* on significant DEGs in each tissue where mean-centred normalized expression ($\log_2$-transformed FPKM+1) were compared across time points [62]. Gene clusters with similar expression patterns were obtained using the Perl script *define_clusters_by_cutting_tree.pl* within trinity to cut the hierarchically clustered gene tree into clusters with similar expression using the *--Ptree* option.

Gene Ontology (GO) enrichment of the identified gene clusters

To infer the functions of the gene clusters, gene ontology (GO) enrichment was performed to identify the enriched biological themes using the R package

clusterProfiler version 3.9 using default settings [63]. The ENTREZ gene identifiers of up- and downregulated clusters per tissue were used as query gene list against the background genes in each tissue. For the purpose of the enrichment analysis, GO categories with a corrected *P*-value of < 0.05 were considered significant. Categories of candidate genes implicated in maturation were visualised as heatmaps using their normalised expression values with the R package pheatmap version 1.0.12 https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf.

Genomic DNA isolation, WGBS library preparation and sequencing

Tissue samples were snap frozen in liquid Nitrogen and stored at -80 °C until genomic DNA (gDNA) was extracted using DNeasy blood and tissue kit (QIAGEN). Tissues were lysed in 360 µL of lysis solution on a Precellys 24 homogenizer for 30s at 4.0 ms^{-1}. Samples were incubated with 40 µL of Proteinase K enzyme at 56 °C for 1 h. Following lysis, samples were treated with RNase (8 µL of RNase A incubated for 2 min at room temperature). DNA was bound to the provided columns, washed twice and eluted in 100 µL at room temperature. gDNA purity were assessed by gel electrophoresis and NanoDrop ND-1000 spectrometer. DNA concentration and integrity were assessed using Agilent 2100 bioanalyzer. gDNA was fragmented (200-400bp) by sonication using Covaris S220, followed by end repair/adenylation and adapter ligation. Bisulfite modification was performed to the DNA fragments using the EZ DNA Methylation-GoldTM Kit (Zymo Research, Inc.). Twelve libraries prepared from 3 tissues (pituitary, ovary and liver), 2 time points (T1 and T4) and 2 biological replicates. Libraries were sequenced on HiSeq 2500 sequencing platform at Novogene, Hong Kong. Sequencing produced a total of 2.2 billion individual 150

bp paired-end reads and 185 M PE reads per library. Bisulfite conversion rates (percentage of C changed to T after bisulfite treatment) were consistently >99.8%.

WGBS data QC, genome mapping and methylation calling

Raw data quality control was performed using Trim Galore v0.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to filter bases (Q scores < 30) and remove both universal and indexed adapter sequences. Processed high-quality data were mapped to into a bisulfite-converted version of the Atlantic salmon reference genome ICSASG_v2 [58] using BSseeker2 v2.1.8 [66] with default parameters for aligning paired-end libraries using Bowtie2 (Langmead & Salzberg 2012). PCR duplicates were detected and removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard/). Filtered (duplicates-free) reads (110 M PE reads) were retained for downstream methylation analysis with an average genome coverage of 11x in pituitary, ovary and liver. Methylation calling was conducted using the Python script *call-methylation.py* within BSseeker2. CGmap files were used for subsequent exploratory and differential methylation analyses. The *mstat* command within CGmap tools was used to generate global and CG context (CG, CHG, CHH) DNA methylation levels [66].

DNA methylome exploratory analyses

As CG methylation contributed to the bulk of methylated Cs, average methylation levels of genome-wide CpG positions were calculated in 50 kb bins across the genome using *mbin* command within CGmap tools and plotted as Violin plots using the R package vioplot, https://cran.r-project.org/web/packages/vioplot/index.html. To begin assessment of the quality of our libraries, common CpGs with minimum 10x

coverage among the 12 samples were used in PCA using *prcomp* implemented in R. Correlation matrices (based on Pearson coefficient) were prepared using the R package *corrplot* (https://cran.r-project.org/web/packages/corrplot/index.html). Hierarchical clustering analysis was conducted with *hclust* implemented in R using compute linkage and Euclidean distances.

Differential CpG methylation analysis

The R package DSS was used to identify differential methylation regions using common CpGs [67]. In each tissue, two replicates at T4 were compared to the control samples at T1 based on CpG methylation levels. At each CpG site, the methylation (M) level was calculated as a proportion of the total counts (coverage) as follows: M levels = [methylated counts / total counts] × 100. DSS was selected as it takes into account the biological variation among replicates (characterized by a dispersion parameter) and the sequencing depth. Differentially methylated loci (DMLs) were identified by estimating mean methylation for all CpG sites followed by estimating dispersion at each site and conducting a Wald test ($P < 0.001$). Smoothing (combining information from nearby CpG sites to improve the estimation of methylation levels) was utilised to obtain mean methylation estimates in WGBS data where the CpG sites are dense. Based on the DML results, regions with statistically significant CpG sites were identified as a differentially methylated regions (DMRs) with minimum length/distance of 50 bp and minimum CpG coverage of 3. Mean methylation between groups of greater than 10 % (delta = 0.1) and $P < 0.001$ was considered significant. A circos plot was produced to visualize multi-tissue genome-wide DMRs using Circos (http://circos.ca/software/). Individual DMRs were also visualized using the *showOneDMR* function within the DSS package to plot both the

methylation percentages (including a smoothed curve) as well as the coverage depths at each CpG site.

DMR annotation, DMGs and DEGs correspondence analysis

Differentially methylated regions were compared against the protein coding gene set annotated on reference ICSASG_v2 using custom Perl scripts. This classified DMRs as overlapping a gene body (genic), 5kb upstream of a transcription start site TSS (putative promoter), 5 kb downstream of TSS (5kb downstream), or otherwise intergenic. The distance between each DMR and nearest gene is provided in Supplementary tables 9 – 11. The overlap between significant genes from differential expression and methylation was checked using the *intersect* function within bedtools [68].

GO enrichment of DMGs

GO enrichment analyses were conducted on both the sets of hypermethylated genes (n= 1,156) and genes found to be hypermethylated and upregulated in ovary (n= 148) using the R package clusterProfiler. Genes driving GO enrichment were plotted as a heatmap using the R package pheatmap as above.

Nuclei extraction, ATAC-seq library preparation and sequencing

ATAC-seq libraries were prepared from frozen tissues using the Omni-ATAC method [69] with the following modifications. Frozen tissue (20 mg) was ground in liquid nitrogen using a mortar and pestle. The pulverized tissue was transferred to a pre-

chilled 2 ml dounce homogenizer containing 1mL cold 1x homogenisation buffer and homogenised with the pestle to form a uniform suspension (10-20 strokes). The homogenate was filtered with a 40uM nylon cell strainer (BD Falcon) before layering onto the iodixanol solution as described previously [69]. The ratio of nuclei to enzyme concentration was optimised for each sample by performing transposition reactions containing 50000, 100000 and 200000 nuclei with 2.5ul of tagment enzyme in 50ul of transposition mix [69]. The transposed DNA was amplified with custom primers as described elsewhere [70]. before libraries were purified using Agencourt AMPure XP beads (Beckman Coulter) and quality controlled using a Bioanalyser High Sensitivity DNA Analysis kit (Agilent). Twelve liver ATAC-seq libraries arising from 3 biological replicates x 4 time points (T1-T4) were sequenced at the IMB sequencing facility (University of Queensland) on an Illumina NextSeq 150 cycle (2 X 75 bp).

Chromatin accessibility data QC, genome mapping and peak calling

Sequencing produced a total of 1.2 billion individual paired-end reads (Supplementary Table 14). Raw reads were mapped to the Atlantic salmon reference genome ICSASG_v2 [58] using BOWTIE2 version 2.3.5.1 with the *--very-sensitive* parameter [71]. Duplicate reads were removed using the MarkDuplicates function in Picard (http://broadinstitute.github.io/picard/). Multi-mapped reads and mitochondrial reads were filtered out and only uniquely mapped reads (MAPQ > 10) were extracted from alignment files using SAMTOOLS for downstream analyses.

For peak calling, the model-based analysis of ChIP-seq (MACS2) (https://github.com/macs3-project/MACS) was used to identify read enrichment

regions “peaks” using default parameters. Only peaks detected in at least two replicates per condition were used for downstream analyses, and peaks across timepoints were merged to generate a unique peak list per tissue. The number of raw reads mapped to each peak was quantified using the Python package HTSeq version 0.11.1 [59].

Differential accessibility and clustering analyses

Samples from the long photoperiod time points (T2, T3 and T4) were compared to control samples (T1) for each tissue. Raw counts were analysed using the R package edgeR and P-values were corrected for multiple testing using the Benjamini and Hochberg algorithm. Peaks with FDR < 0.05 and $\log_2$FC > $\pm$ 1 were considered significantly differentially accessible regions (DARs). PCA of significant DARs used normalised accessibility data ($\log_2$CPM) prepared using the function *--prin_comp* within trinity. Hierarchical clustering analysis was conducted using *analyze_diff_expr.pl* where mean-centred normalized accessibility ($\log_2$CPM+1) were compared across time points [62]. Gene clusters with similar accessibility patterns were obtained using the Perl script *define_clusters_by_cutting_tree.pl* to cut the hierarchically clustered gene tree into clusters with similar accessibility patterns as described above.

Genomic distribution of DARs within clusters

Hierarchical clustering identified both accessible and inaccessible DAR clusters. DARs per cluster were annotated in a genomic context (genic, promoter, 5 kb downstream or intergenic) as previously done for annotation of DMRs.

ATAC-seq and RNA-seq correspondence analysis

Only DARs co-located with genes and promoters were used for co-analysis with gene expression data. The relationship between accessibility of DARs and gene expression was visualised by overlying information of significant DARs to genome-wide normalised expression estimates in liver samples and plotted as a MA-biplot. A linear regression analyses were performed to assess correlations between accessibility and expression abundance and the effect of changes in accessibility and changes in gene expression across time. Chromatin accessibility and gene expression data were visualised using Gnuplot version 5.0.7 (http://www.gnuplot.info) by overlying accessibility data of significant DARs at genes and promoters to genome-wide normalised expression estimates at each timepoint.

Multiomic heatmap analysis per time and genomic regions

All heatmaps were produced using the R package pheatmap. GO enrichment analyses have been conducted on the set of nearest genes to accessible promoters using the R package clusterProfiler as described above. The integrated genome viewer (IGV) was used to visualise the relationship between accessibility and gene expression in a 15kb region that contains *hmgrc* gene and its promoter region.

Motif enrichment analyses

The function *findMotifsGenome.pl* within Homer software version 4.11 (http://homer.ucsd.edu/homer/) was used with default parameters to find sequence motifs significantly enriched among accessible DARs vs inaccessible DARs located

839 within promoter regions.TF motifs that are highly enriched (*P* value < 1 x 10^{-10}) were
840 selected.
841
842 **Inference of master regulators**
843 Master regulator analysis was performed using regulatory impact factor (RIF) metrics
844 described by [72] to identify key regulators contributing to the differential expression in
845 the T4 vs T1 comparison in each tissue. Data for potential transcription factors (TFs)
846 in Atlantic salmon were taken from Mohamed et al., 2018. As most of the
847 transcriptional changes were detected at T4, RIF was applied to the T4-T1
848 comparisons for each tissue. Briefly, RIF exploits the differential co-expression
849 concept where regulators were contrasted against unique lists of genes that were
850 differentially expressed at T4 in each tissue. Genes with a mean expression FPKM
851 < 0.2 were excluded. Those scores deviating ± 2.57 standard deviation from the
852 mean were considered significant at *P* < 0.01. We identified a total of 305 significant
853 regulators (113, 68 and 123 in pituitary, ovary and liver, respectively at *P* > 0.01).
854 Most of these regulators (n=298; 97.7%) were unique to each tissue leaving only 7
855 that were shared among tissue pairs (Supplementary Table S17; Supplementary
856 Figure S17). The regulators identified were used as input for construction of gene
857 regulatory networks as summarized in Supplementary Figure S1b.
858
859 **Gene regulatory network (GRN) analysis**
860 Genes from different omics analyses (DEGs, DMGs, DMPs, DACs) along with key
861 transcription factors identified by RIF (TFs), as well as information for tissue-specific
862 (TS) genes and gene-harbouring GWAS SNPs (SNPs) were selected based on
863 overlap (at least once) and mean normalised expression (at least 0.2 FPKM) for

network construction. The R package UpSetR (https://cran.r-project.org/web/packages/UpSetR/vignettes/basic.usage.html) was used to investigate the cross-talk among genes from different sources.

For gene network inference, genes were used as nodes and significant connections (edges) between them were identified using the Partial Correlation and Information Theory (PCIT) algorithm [73], considering all samples. PCIT determinates the significance of the correlation between two nodes after accounting for all the other nodes in the network. Connections between gene nodes were accepted when the partial correlation was greater than two standard deviations from the mean ($P < 0.05$). The output of PCIT was visualized using Cytoscape Version 3.7.2 [74].

In order to explore differential connectivity during maturation onset, two networks were created; one using 12 samples at T1 (pre-maturation) and a second using 36 samples at T2, T3 and T4 (post-maturation). The number of connections of each gene in each network was computed, making it possible to compare the same gene in the two networks to identify differentially connected genes (DCGs). From these networks, we explored a series of subnetworks. First subnetworks based on the top trio genes and the top regulators (TFs) based on their differential connectivity between pre-and post-maturation. Pre- and post-maturation networks were constructed from the 12 control samples at T1 and 36 post-maturation (T2-T4) samples.

References

55. Mohamed, A.R., King, H., Evans, B., Reverter, A. & Kijas J.W. Multi-Tissue Transcriptome Profiling of North American Derived Atlantic Salmon. *Front. Genet.* **9**, 369 (2018).

56. Lien, S. et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200 (2016).

57. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2009).

58. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

59. Anders, S., Pyl, P. T., and Huber, W. HTseq - a Python framework to work with high-throughput sequencing data. *Bioinformatics 31*, 166–169 (2015).

60. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

61. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol*. **57**, 289–300 (1995).

62. Haas, B. J. et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc*. **8**, 1494–1512 (2013).

63. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

64. Guo, W. et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774 (2013).

911 65. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat.*
912 *Methods* **9**, 357–359 (2012).
913 66. Guo, W., Zhu, P., Pellegrini, M., Zhang, M. Q., Wang, X., and Ni, Z. (2018).
914 CGmapTools improves the precision of heterozygous SNV calls and supports allele-
915 specific methylation detection and visualization in bisulfite-sequencing
916 data. *Bioinformatics* **34**, 381–387
917 67. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect
918 differentially methylated loci from single nucleotide resolution sequencing
919 data. *Nucleic Acids Res*. **42**, e69 (2014).
920 68. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr.*
921 *Protoc. Bioinformatics* **47**, 11–34 (2014).
922 69. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and
923 enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
924 70. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of
925 regulatory variation. Nature 523, 486–490 (2015).
926 71. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat.
927 Methods 9, 357–359 (2012).
928 72. Reverter, A., Hudson, N. J., Nagaraj, S. H., Perez-Enciso, M. & Dalrymple, B. P.
929 Regulatory impact factors: unraveling the transcriptional regulation of complex traits
930 from expression data. *Bioinformatics* **26**, 896–904 (2010).
931 73. Reverter, A. & Chan, E.K.F. Combining partial correlation and an information theory
932 approach to the reversed engineering of gene co-expression
933 networks. *Bioinformatics* **24**, 2491–2497 (2008).
934 74. Shannon, P. et al. Cytoscape: a software environment for integrated models of
935 biomolecular interaction networks. *Genome Res*. **13**, 2498–2504 (2003).

936 **Data availability**

937 Sequencing reads, raw and processed data used in this multiomic study have been
938 submitted to the NCBI Gene Expression Omnibus (GEO) database under Accession
939 numbers GSE157003 (combined data types); GSE157001 (RNA-seq); GSE156998
940 (ATAC-seq) and GSE156997 (whole genome bisulphite sequencing data). Gene
941 network cystoscope file contains all networks described in the paper is supplied as
942 supplementary file.

943

944 **Code availability**

945 All codes used for gene expression (RNA-seq;
946 https://github.com/AminRM/salmon_mat_transcriptomes), DNA methylation
947 (WGBS;https://github.com/AminRM/DNAmethylomes), chromatin accessibility
948 (ATAC-seq; https://github.com/AminRM/Salmon-Chromatin-Accessibility), gene
949 networks (GRN; https://github.com/AminRM/maturation_GRN) analyses are
950 available

951

952 **Acknowledgements**

953 We acknowledge the cooperation and assistance provided by Saltas staff who
954 performed the animal husbandry component of the experiment. The CSIRO Office of
955 the Chief Executive provided financial support to Amin Mohamed. The work is
956 dedicated to the memory of Harry King.

957

958 **Author Contribution Statement**

959 AM, HK, BE and JK conceived and designed the experiment. AM generated and
960 analysed RNA-seq and WGBS data. MM generated ATAC-seq data. AM and MN-S

961 analysed ATAC-seq data. AM and AR analysed the gene regulatory networks. AM
962 prepared all the figures/tables and prepared the first draft of the manuscript. All
963 authors reviewed, commented on and approved the final manuscript.
964

965 **Ethics Statement**

966 The fish were reared and euthanized in compliance with the CSIRO Animal Ethics
967 Committee, under approval number 2017–02.
968

969 **Competing interests**

970 The authors declare no competing interests.

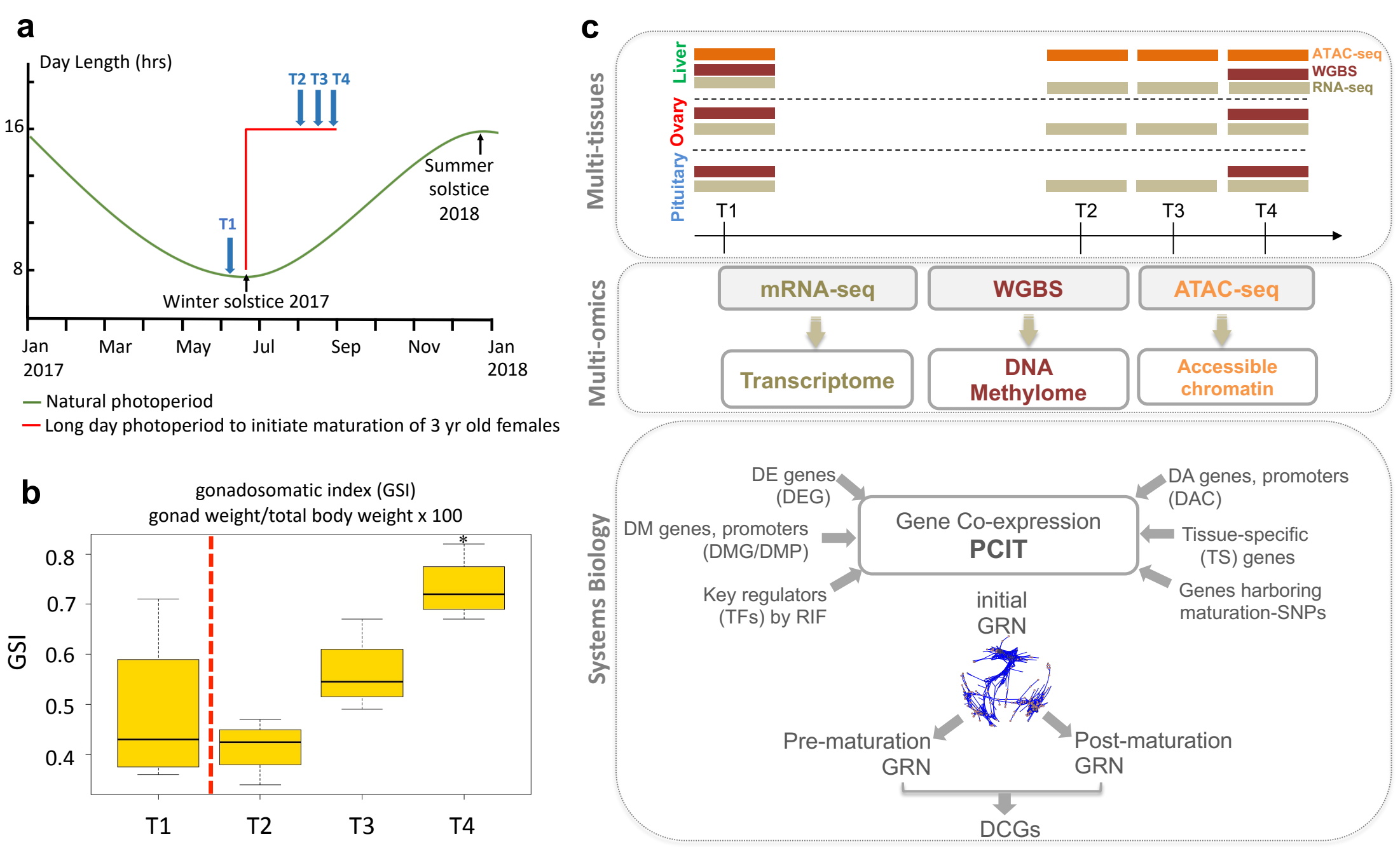

a
Day Length (hrs)
16
8
T2 T3 T4
T1
Summer solstice 2018
Winter solstice 2017
Jan 2017
Mar
May
Jul
Sep
Nov
Jan 2018
Natural photoperiod
Long day photoperiod to initiate maturation of 3 yr old females
b
gonadosomatic index (GSI)
gonad weight/total body weight x 100
GSI
0.8
0.7
0.6
0.5
0.4
T1
T2
T3
T4
c
Multi-tissues
Liver
Ovary
Pituitary
ATAC-seq
WGBS
RNA-seq
T1
T2
T3
T4
Multi-omics
mRNA-seq
WGBS
ATAC-seq
Transcriptome
DNA Methylome
Accessible chromatin
Systems Biology
DE genes (DEG)
DA genes, promoters (DAC)
DM genes, promoters (DMG/DMP)
Gene Co-expression PCIT
Tissue-specific (TS) genes
Key regulators (TFs) by RIF
Genes harboring maturation-SNPs
initial GRN
Pre-maturation GRN
Post-maturation GRN
DCGs

Fig.1 Multi-tissue transcriptomic and epigenomic changes associated with onset of salmon maturation. a, Induction of maturation through photoperiod manipulation and sampling time points. Animals were managed via photoperiod manipulation to synchronise the timing of commitment into maturation. 4 fish were sampled at each of the T1 (before long day photoperiod signal) and T2-T4 time points (during maturation) in 2 weeks intervals to control for variation between individuals. **b, Gonadosomatic index (GSI) throughout the time-course experiment.** GSI increased gradually from T2 till the last sampling event at T4 indicating active response towards maturation in these animals. The variability in GSI measures during maturation decreased compared to that of T1. **c, Tissue collection, multi-omics integrated analyses and gene regulatory networks (GRNs).** Samples from the pituitary gland, ovary and liver were collected at each sampling event. High-throughput sequencing was utilised to profile genome-wide changes in transcriptomes, DNA methylomes in the three tissues along with chromatin accessibility in liver. These genome-wide results were integrated with knowledge of key regulators (TFs) and other results from previous work that identified maturation associated-SNPs and tissue-specific genes. PCIT algorithm was then utilised to construct gene regulatory networks (GRNs). Pre- and post-maturation GRNs were constructed to identify differentially connected genes (DCGs).

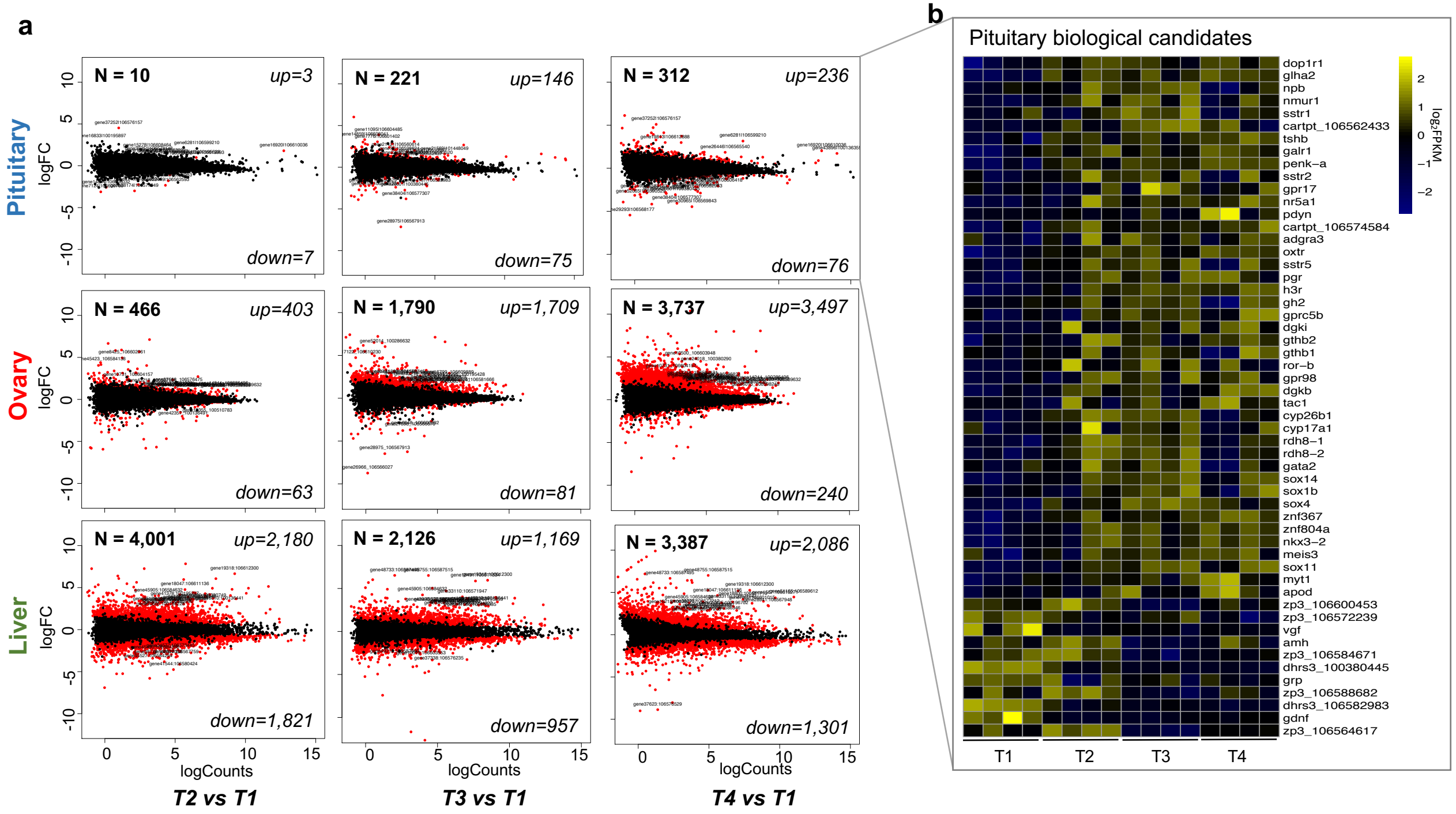

a
Pituitary
Ovary
Liver
logFC
N = 10
up=3
down=7
N = 221
up=146
down=75
N = 312
up=236
down=76
N = 466
up=403
down=63
N = 1,790
up=1,709
down=81
N = 3,737
up=3,497
down=240
N = 4,001
up=2,180
down=1,821
N = 2,126
up=1,169
down=957
N = 3,387
up=2,086
down=1,301
logCounts
T2 vs T1
T3 vs T1
T4 vs T1
b
Pituitary biological candidates
dop1r1
glha2
npb
nmur1
sstr1
cartpt_106562433
tshb
galr1
penk–a
sstr2
gpr17
nr5a1
pdyn
cartpt_106574584
adgra3
oxtr
sstr5
pgr
h3r
gh2
gprc5b
dgki
gthb2
gthb1
ror–b
gpr98
dgkb
tac1
cyp26b1
cyp17a1
rdh8–1
rdh8–2
gata2
sox14
sox1b
sox4
znf367
znf804a
nkx3–2
meis3
sox11
myt1
apod
zp3_106600453
zp3_106572239
vgf
amh
zp3_106584671
dhrs3_100380445
grp
zp3_106588682
dhrs3_106582983
gdnf
zp3_106564617
log2FPKM
T1
T2
T3
T4

Fig.2 Significant transcriptome remodelling driving onset of salmon maturation. a, 9 MA plots showing differentially expressed genes (DEGs) in pituitary, ovary and liver (FDR < 0.05 and log_2fold change >± 1) at T2, T3 and T4 during maturation compared to control samples at T1. Transcriptomic results reveal gene upregulation in both pituitary gland and ovary during maturation. **b**, Heatmap of pituitary candidate genes with significant expression. The hierarchical clustering was obtained by comparing the log_2-transformed and mean-centred FPKM values. Upregulated genes in the pituitary include several genes with reproduction-related functions such as steroidogenesis (*cyp17a1*), hormone receptors (*oxtr, pgr, dop1r1, galr1*), genes coding for pituitary hormones (*gh2, gthb1, gthb2, glha2*), retinoic acid (RA) signalling (*cyp26b1, rhd8-1/2),* sex-related TFs (several sox genes and *gata2*). The upregulation of gonadotropins subunits (*gthb1*, *gthb2* and *glha2*) is highly significant in the context of maturation onset. *gthb1* and *gthb2* encoding gonadotropin subunits beta-1 and 2 and *glha2* encoding glycoprotein hormones alpha chain that constitutes the salmon gonadotropins (*gth-1* and 2).

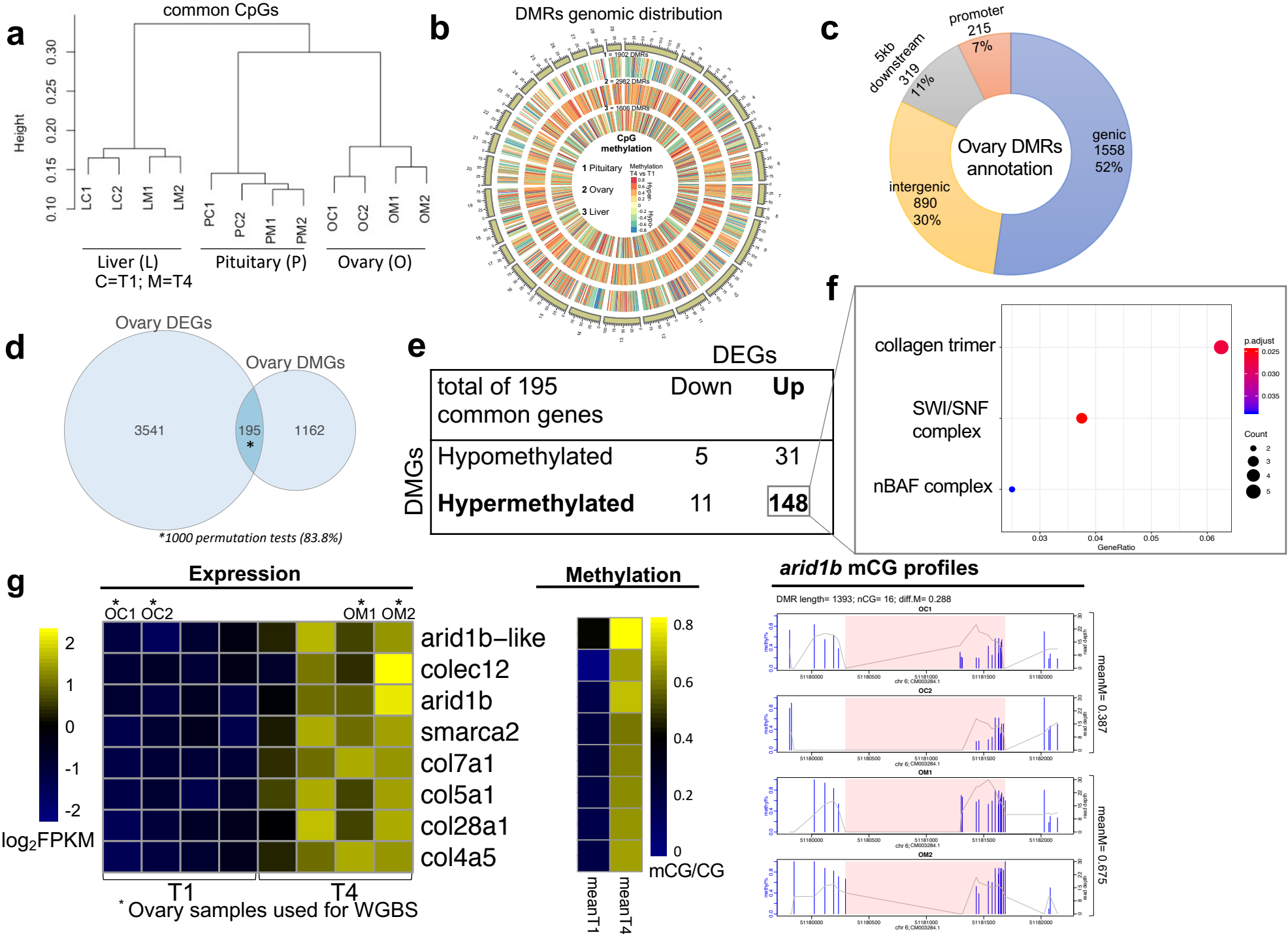
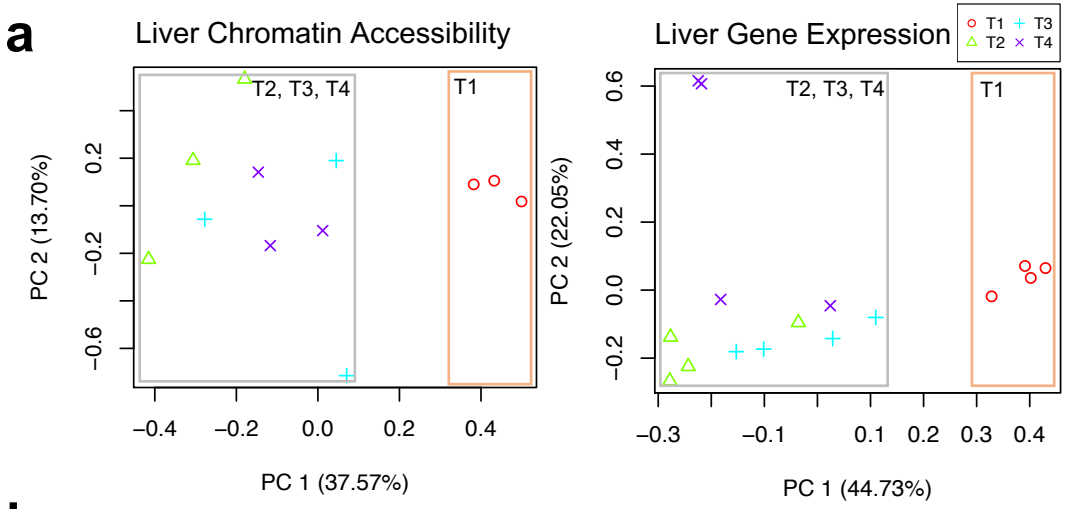
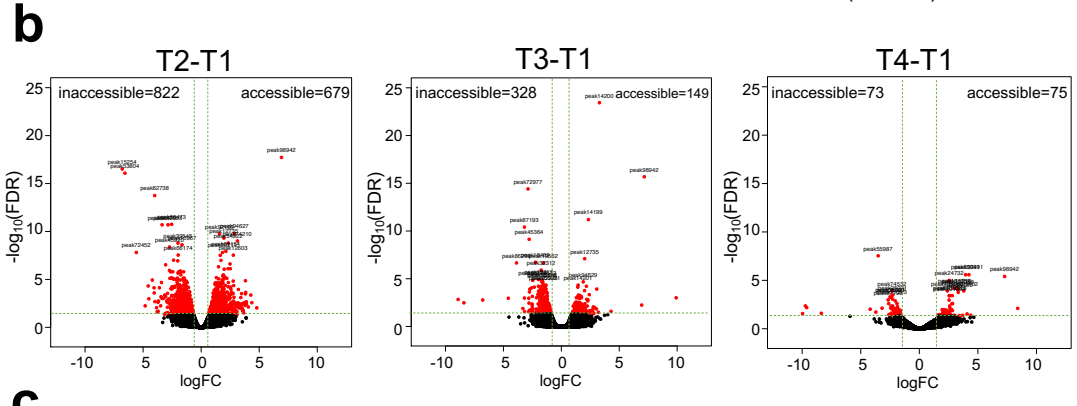

a
common CpGs
Height
0.30
0.25
0.20
0.15
0.10
LC1
LC2
LM1
LM2
PC1
PC2
PM1
PM2
OC1
OC2
OM1
OM2
Liver (L)
C=T1; M=T4
Pituitary (P)
Ovary (O)
b
DMRs genomic distribution
1 = 1902 DMRs
2 = 2982 DMRs
3 = 1606 DMRs
CpG methylation
1 Pituitary
2 Ovary
3 Liver
Methylation T4 vs T1
Hyper
Hypo
c
promoter 215 7%
5kb downstream 319 11%
genic 1558 52%
intergenic 890 30%
Ovary DMRs annotation
d
Ovary DEGs
Ovary DMRs
3541
195
1162
*
*1000 permutation tests (83.8%)
e
DEGs
DMGs
total of 195 common genes
Down
Up
Hypomethylated
5
31
Hypermethylated
11
148
f
collagen trimer
SWI/SNF complex
nBAF complex
p.adjust
0.025
0.030
0.035
Count
2
3
4
5
0.03
0.04
0.05
0.06
GeneRatio
g
Expression
OC1
OC2
OM1
OM2
2
1
0
-1
-2
log2FPKM
arid1b–like
colec12
arid1b
smarca2
col7a1
col5a1
col28a1
col4a5
T1
T4
* Ovary samples used for WGBS
Methylation
0.8
0.6
0.4
0.2
0
mCG/CG
meanT1
meanT4
arid1b mCG profiles
DMR length= 1393; nCG= 16; diff.M= 0.288
OC1
OC2
OM1
OM2
meanM= 0.387
meanM= 0.675
methyl%
read depth
51180000
51180500
51181000
51181500
51182000
chr 6; CM003284.1

Fig.3 Multi-tissue DNA methylome maps reveal methylated genes with key roles in salmon maturation. a, Hierarchical clustering of the common CpGs in the 12 methylome libraries reveals clustering by tissue type and grouping pituitary and ovary in a cluster and liver in the other. This shows variation in methylation is much lower between replicates of the same sample compared with variation among tissues as expected for high quality data. **b,** Circos plot shows genome-wide distribution of the significant differentially methylated regions (DMRs) detected in T4 samples compared to the control T1 in pituitary, ovary and liver (from the outer circle inwards). **c**, Genomic distribution of ovary DMRs across gene models in the salmon genome shows the majority co-located within genes, promoter defined as genomic regions located 5kb upstream of transcription start sites (TSSs). **d**, Significant overlap between ovary DMGs and DEGs. **e**, DNA methylation and expression directionality at gene bodies confirms that gene body methylation positively affects gene expression. **f**, Enriched gene ontology (GO) terms (hypergeometric test, Bonferroni-adjusted $P < 0.05$) among the list of 148 hypermethylated genes in ovary. **g**, Heatmaps showing significant upregulation of genes driving GO enrichment shown in part f, mean CpG methylation levels of upregulated genes before and after maturations and CpG methylation profiles for the *arid1b* gene. The methylation plot shows percentages of methylated CpGs along with coverage depth at each CpG site. Pink rectangles represent the differentially methylated regions. Genomic coordinates are indicated below the density plot and DMR details are also indicated.





c

	DAC T2			DAC T3			DAC T4		
	b	b2	R2 (%)	b	b2	R2 (%)	b	b2	R2 (%)
DEG T2	0.5123	0.291	36.8	0.6275	0.0711	20.8	0.6053	0.1044	29.5
DEG T3	0.3985	0.2131	32.5	0.562	0.0335	24.8	0.4972	0.0432	27.5
DEG T4	0.4311	0.2862	35.3	0.5548	0.0281	20.1	0.5617	0.1105	32.8

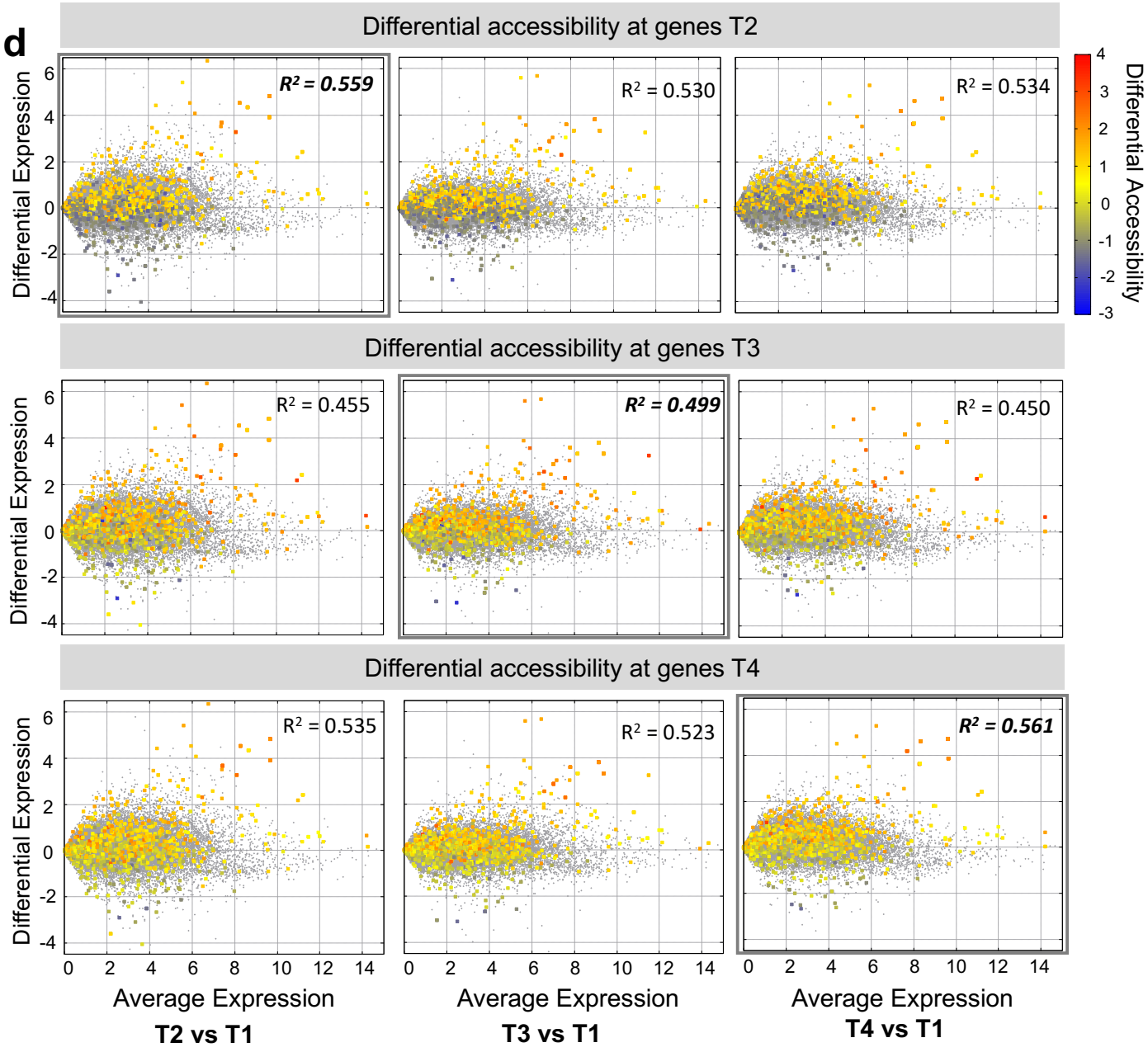


Fig.4 Genome-wide maps of chromatin accessibility and gene expression changes. a, Combined ATAC-seq and RNA-seq analysis. Principal component analysis (PCA) conducted using normalised (log_2CPM) values of the lists of significant differentially accessible regions (DARs) and significant differentially expressed genes (DEGs) at T2, T3 and T4 during maturation compared to control samples at T1, the same significance thresholds were applied (FDR < 0.05 and log_2fold change >±1). **b**, **Strong and early remodelling in the chromatin state landscape**, the volcano plots show differentially accessible regions (DARs) where mapping counts differed significantly between T1 and other time points. **c, Regression analysis conducted on significant DARs located at gene bodies and the corresponding gene expression data.** The table shows that accessibility at T2 explains the majority of the observed differential expression throughout the experiment. **d, Chromatin accessibility and gene expression are positively correlated.** 9 MA biplots showing genome-wide gene expression and overlain differentially accessible regions dynamics at gene bodies in liver at T2, T3 and T4 compared to T1. Chromatin accessibility levels are shown in red-blue spectrum reflecting open- to closed chromatin at gene bodies and the corresponding gene expression in grey colour. Note that R^2 values are the highest when using gene expression and accessibility from the same time point.

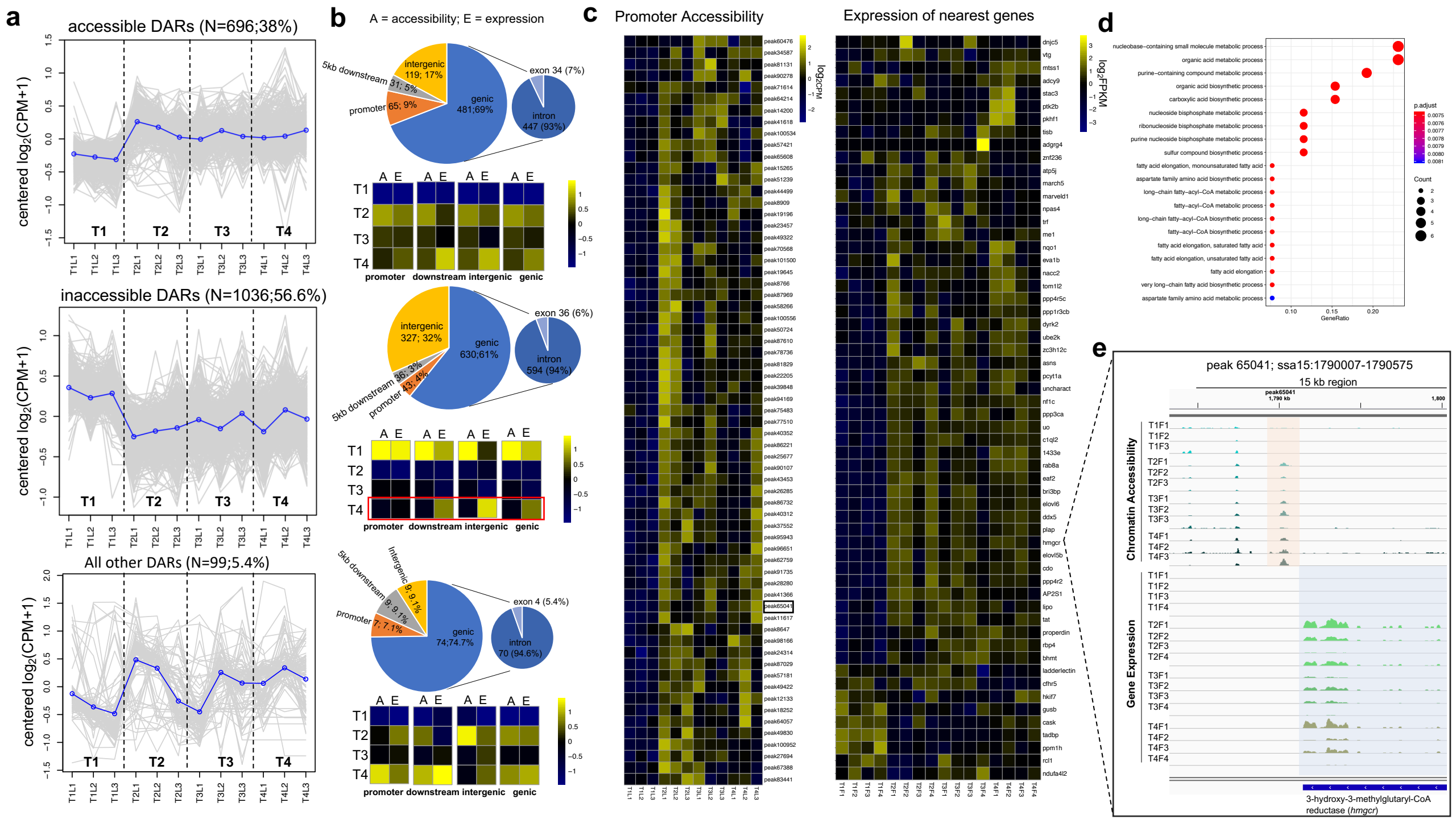
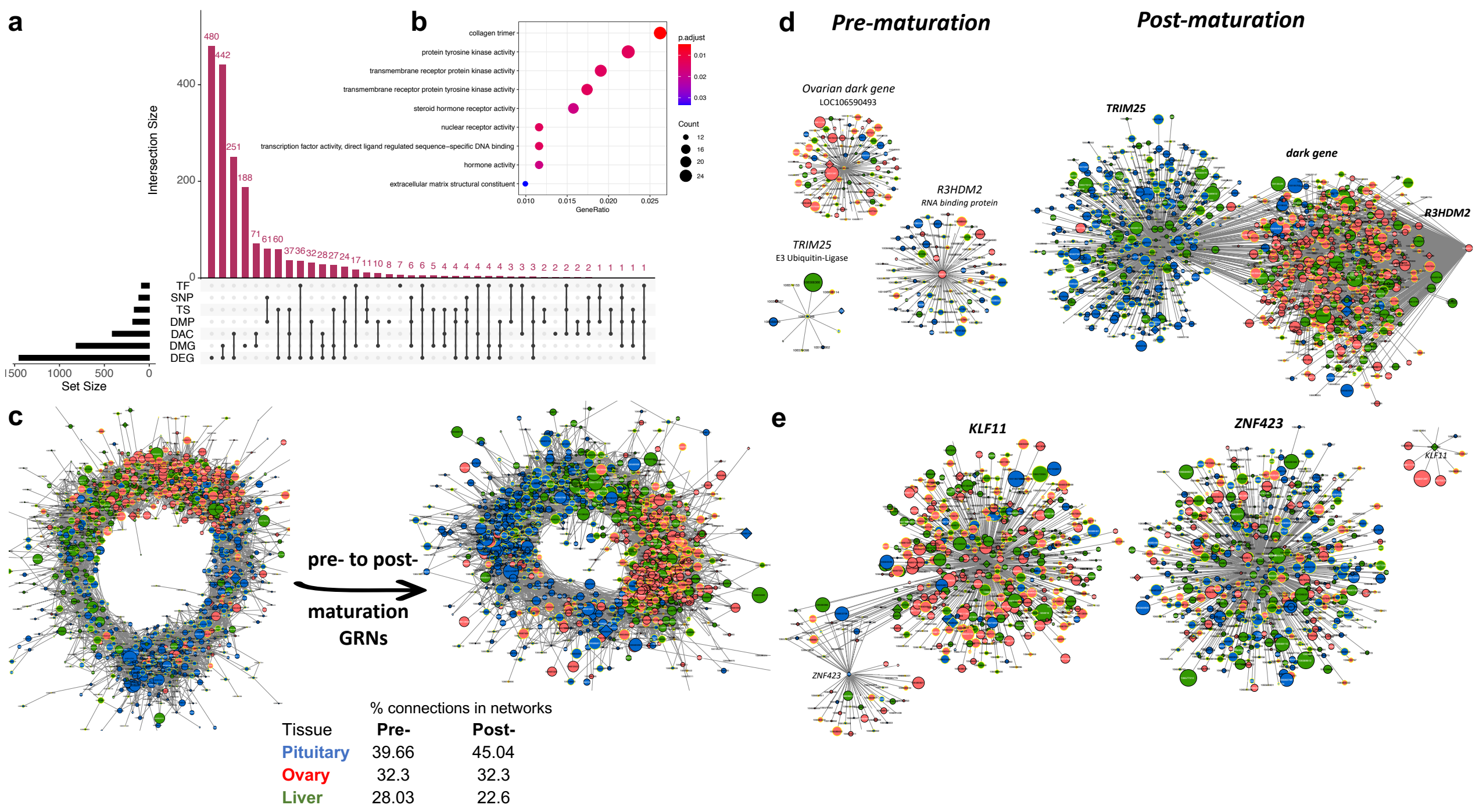

a
accessible DARs (N=696;38%)
centered log2(CPM+1)
T1
T2
T3
T4
T1L1 T1L2 T1L3 T2L1 T2L2 T2L3 T3L1 T3L2 T3L3 T4L1 T4L2 T4L3
inaccessible DARs (N=1036;56.6%)
All other DARs (N=99;5.4%)
b
A = accessibility; E = expression
intergenic 119; 17%
5kb downstream 31; 5%
promoter 65; 9%
genic 481;69%
exon 34 (7%)
intron 447 (93%)
A E
promoter downstream intergenic genic
intergenic 327; 32%
genic 630;61%
exon 36 (6%)
intron 594 (94%)
5kb downstream 36; 3%
promoter 43; 4%
Intergenic 9; 9.1%
5kb downstream 9; 9.1%
promoter 7; 7.1%
genic 74;74.7%
exon 4 (5.4%)
intron 70 (94.6%)
c
Promoter Accessibility
Expression of nearest genes
log2CPM
log2FPKM
peak60476
peak34587
peak81131
peak90278
peak71614
peak64214
peak14200
peak41618
peak100534
peak57421
peak65608
peak15265
peak51239
peak44499
peak8909
peak19196
peak23457
peak49322
peak70568
peak101500
peak19645
peak8766
peak87969
peak58266
peak100556
peak50724
peak87610
peak78736
peak81829
peak22205
peak39848
peak94169
peak75483
peak77510
peak40352
peak86221
peak25677
peak90107
peak43453
peak26285
peak86732
peak40312
peak37552
peak95943
peak96651
peak62759
peak91735
peak28280
peak41366
peak65041
peak11617
peak8647
peak98166
peak24314
peak87029
peak57181
peak49422
peak12133
peak18252
peak64057
peak49830
peak100952
peak27694
peak67388
peak83441
dnjc5
vtg
mtss1
adcy9
stac3
ptk2b
pkhf1
tisb
adgrg4
znf236
atp5j
march5
marveld1
npas4
trf
me1
nqo1
eva1b
nacc2
tom1l2
ppp4r5c
ppp1r3cb
dyrk2
ube2k
zc3h12c
asns
pcyt1a
uncharact
nf1c
ppp3ca
uo
c1ql2
1433e
rab8a
eaf2
bri3bp
elovl6
ddx5
plap
hmgcr
elovl5b
cdo
ppp4r2
AP2S1
lipo
tat
properdin
rbp4
bhmt
ladderlectin
cfhr5
hkif7
gusb
cask
tadbp
ppm1h
rcl1
ndufa4l2
d
nucleobase−containing small molecule metabolic process
organic acid metabolic process
purine−containing compound metabolic process
organic acid biosynthetic process
carboxylic acid biosynthetic process
nucleoside bisphosphate metabolic process
ribonucleoside bisphosphate metabolic process
purine nucleoside bisphosphate metabolic process
sulfur compound biosynthetic process
fatty acid elongation, monounsaturated fatty acid
aspartate family amino acid biosynthetic process
long−chain fatty−acyl−CoA metabolic process
fatty−acyl−CoA metabolic process
long−chain fatty−acyl−CoA biosynthetic process
fatty−acyl−CoA biosynthetic process
fatty acid elongation, saturated fatty acid
fatty acid elongation, unsaturated fatty acid
fatty acid elongation
very long−chain fatty acid biosynthetic process
aspartate family amino acid metabolic process
GeneRatio
p.adjust
Count
e
peak 65041; ssa15:1790007-1790575
15 kb region
peak65041
1,790 kb
1,800
Chromatin Accessibility
Gene Expression
3-hydroxy-3-methylglutaryl-CoA reductase (hmgcr)

Fig.5 Dynamics of chromatin accessibility and gene expression as a function of genomic location reveals strong association with putative cis-regulatory elements (CREs). a, co-accessibility analysis reveals stable chromatin states following the onset of the maturation. The majority of DARs exhibited either reduced (n=1036 or 57%) or increased (n=696 or 38%) accessibility at T2 and remained unchanged at later timepoints. Less than 10% of DARs (n=99) displayed an oscillating pattern. The y-axis in each graph represents the mean-centered log2(CPM+1) value across time points on the x-axis. Accessibility of single DAR is plotted in grey, while the mean accessibility of each cluster is plotted in blue. **b**, Genomic distribution of DARs across gene models in the salmon genome, promoter defined as genomic regions located 5kb upstream of TSSs and multi-omic heatmaps showing mean accessibility and expression at each of the four time points per genomic location. **c**, Heatmaps 65 DARs located 5kb upstream of TSSs representing cis-regulatory elements (CREs) that exhibit increased accessibility at T2 along with gene expression of their target (nearest) genes. This showed upregulation of the majority of the associated genes in a tightly coordinated manner (n = 46; 79%' X^2 $p < 8.028^{-08}$). **d**, Enriched gene ontology (GO) terms (hypergeometric test, Bonferroni-adjusted $P < 0.05$) among the list of the CREs- regulated genes. This list includes genes involved in hepatic lipid metabolism (*hmgcr*) and energy metabolism (*elovl5b* and *elovl6*). **e**, IGV visualisation of a 15 kb region of Ssa15 spanning the CRE and exons of the *hmgcr* gene provides fine-scale view of the coordinated gene expression response to increased accessibility.



% connections in networks

Tissue	Pre-	Post-
Pituitary	39.66	45.04
Ovary	32.3	32.3
Liver	28.03	22.6

Fig.6 Gene regulatory networks (GRNs) constructed via integrating multi-omics results. a, UpSet plot showing the intersection among the 1,858 genes selected for network analyses. Genes were selected when represented at least once and have a mean normalised expression of at least 0.2 FPKM. Network genes were divided into 7 categories: differentially expressed genes (DEG), differentially expressed genes/promoters (DMG, DMP), differentially accessible regions (genic and promoter regions as DAC), genes harbouring SNPs reported in literature as being associated with salmon maturation (SNP), tissue-specific genes (TS) and key regulators identified by RIF (TF). The nature of a given intersection is indicated by the dots below the bar plot. For example, the 442 genes in the second column are both differentially methylated and differentially expressed but not found in other categories. **b**, **Enriched gene ontology (GO) terms** (hypergeometric test, Bonferroni-adjusted $P < 0.05$) among the list of 1,858 network genes along with the gene ratio for the genes that map to each term. The majority of the enriched terms are related to hormone and receptor activities. **c**, **GRNs constructed using the PCIT algorithm for the pre- and post-maturation samples**. For visualisation purpose, only the most significant 10% of correlations their respective genes were considered. This shows increases in the connections in the pituitary that constituted 45% of the post-maturation network connections. All nodes are represented by ellipses except for genes coding key regulators (TFs) have diamond shape. Nodes with yellow borders are differentially methylated, whereas nodes with white labels are differentially accessible between pre- and post-maturation samples. Node colours are relative to the tissue of maximum expression with blue represents the pituitary, red represents ovary and green represents liver. The size of the nodes is relative to the normalized mean expression values in all samples. **d**, **Subnetworks of top differentially connected genes**, the networks created with the most trio genes differentially connected between pre- and post-maturation networks. This revealed TRIM25, a E3 Ubiquitin ligase as the key regulators with the greatest number of gained connections in the post-maturation network. TRIM25 was highly expressed in the pituitary and underwent changes in DNA methylation. **e**, **Subnetworks of top differentially connected TFs,** networks created with the most differentially connected TFs between pre- and post-maturation networks showed zinc finger protein 423 (ZNF423) as the key regulator with the greatest number of gained

connections and Kruppel-like factor 11 (KLF11) as the regulator with the least number of gained connections going from pre- to post-maturation.

geneID	Number of connections		differential connectivity	annotation	contribution	tissue of max expression	category
	Pre-	Post-	(Post vsPre)				
106574348	10	599	589	**E3 ubiquitin/ISG15 ligase TRIM25-like**	28.57	pit	DE, DMG
106590493	114	657	543	**uncharacterized protein LOC106590493, partial**	28.57	ov	DE, DMG
106566126	90	609	519	**R3H domain-containing protein 2-like**	28.57	ov	DE, DAC
106573705	58	576	518	**zinc finger protein 423**	42.86	pit	DE, TF, DAC
106589399	43	551	508	protocadherin-15-like	14.29	pit	DMG
106562515	4	508	504	RILP-like protein 1	28.57	liv	DE, DMG
106562229	30	533	503	tumor necrosis factor receptor type 1-associated DEATH domain protein-like	28.57	ov	DMG, DAC
106612386	41	535	494	zinc transporter protein DDB_G0282067	28.57	ov	DE, DMG
106568904	37	525	488	putative ferric-chelate reductase 1	14.29	ov	DE
106568722	10	489	479	cAMP-specific 3\',5\'-cyclic phosphodiesterase 4B-like	14.29	pit	DMG
106606431	487	32	-455	metalloproteinase inhibitor 2-like	14.29	pit	DE
100195217	543	87	-456	protein phosphatase 1H	28.57	liv	DE, DAC
106613686	560	104	-456	Krueppel-like factor 2	28.57	liv	DE, DAC
100195880	474	11	-463	Ependymin-1 precursor	14.29	pit	DE
100195838	488	18	-470	ependymin-2 precursor	28.57	ov	DE, DMP
106588682	594	113	-481	zona pellucida sperm-binding protein 3-like	14.29	pit	DE
106610009	546	46	-500	guanylate cyclase soluble subunit beta-1	28.57	liv	DMP, DMG
100194692	513	9	-504	**Krueppel-like factor 11**	28.57	liv	DE, TF
100196465	560	31	-529	glyoxalase domain-containing protein 5	14.29	ov	DE
106586502	559	7	-552	serine/arginine repetitive matrix protein 2-like	42.86	ov	DE, DMG, DAC