Salmonids reveal principles of regulatory evolution following autotetraploidization

Marie-Odile Baudement[1‡†], Diego Perojil Morata[2‡,3,4†], Gareth B. Gillard[1†], Pooran S. Dewari[2†], Manu Kumar Gundappa[2‡,5†], Tomasz Podgorniak[1§], Lars Grønvold[1§], Damir Baranasic[6,3,4§], Audrey Laurent[7§], François Giudicelli[8], Bojan Zunar[9], Erika Carrera-García[1], Aline Perquis[7], Aurélien Brionne[7], Tan Thi Nguyen[1], Rose Ruiz Daniels[2‡,10], Gabriela A. Merino[11], David Thybert[11], Garth R. Ilsley[11], Alexandra Louis[8], Torgeir R. Hvidsten[12], Camille Berthelot[8,13], Peter W. Harrison[11], Hugues Roest Crollius[8], Yann Guiguen[7], Boris Lenhard[3,4], Simen R. Sandve[1], Julien Bobe[7], Matthew P. Kent[1], Sigbjørn Lien[1]*, Daniel J. Macqueen[2]*

[1]Centre for Integrative Genetics (CIGENE), Faculty of BioSciences (BIOVIT), Norwegian University of Life Sciences; Ås, NO-1432 Ås, Norway.

[2]The Roslin Institute, University of Edinburgh; Easter Bush Campus, EH25 9RG, United Kingdom.

[3]MRC London Institute of Medical Sciences; London W12 0NN, United Kingdom.

[4]Institute of Clinical Sciences, Imperial College London; London W12 0NN, United Kingdom.

[5]Animal Breeding and Genomics, Wageningen University and Research; Wageningen, 6700 HB, The Netherlands.

[6]Ruđer Bošković Institute; Bijenička Cesta 54, Zagreb, 10000, Croatia.

[7]Fish Physiology and Genomics, INRAE; Rennes, UR 1037, France.

[8]Institut de Biologie de l'ENS (IBENS), Département de Biologie, Ecole normale supérieure, CNRS, Inserm, Université PSL; Paris, 75005, France.

[9] Faculty of Food Technology and Biotechnology, University of Zagreb; Zagreb, 10000 Zagreb, Croatia

[10]Institute of Aquaculture, University of Stirling, Stirling, FK9 4LA, United Kingdom.

[11]European Molecular Biology Laboratory, European Bioinformatics Institute; Cambridge, CB10 1SD, United Kingdom.

[12]Faculty of Chemistry, Biotechnology and Food Science (KBM), Norwegian University of Life Sciences; Ås, NO-1432 Ås, Norway.

[13]Institut Pasteur, Université Paris Cité, CNRS UMR 3525, INSERM UA12, Paris, F-75015, France.

[†] These authors contributed equally (as first author)

§ These authors contributed equally (as second author)

[‡] Former address

*Corresponding authors. Daniel J. Macqueen: daniel.macqueen@roslin.ed.ac.uk; Sigbjørn Lien: sigbjorn.lien@nmbu.no

Abstract

Early vertebrate autotetraploidization events may have enabled major innovations by expanding the genetic material for functional diversification, yet their ancient timing obscures how genome doubling reshaped gene regulatory evolution. Salmonids provide a unique window to these mechanisms, because they experienced a comparatively recent autotetraploidization and are earlier in the rediploidization process - which creates new genes and regulatory elements during evolution. Here, using large-scale multiomics spanning embryonic and adult tissues in two salmonids, we investigate gene regulatory evolution following genome doubling and rediploidization, which we show is governed by developmental and tissue-specific context, with a period of maximal constraint at advanced stages of embryogenesis. This work advances understanding of vertebrate genome evolution, while providing an open resource supporting salmonid aquaculture and conservation.

Main Text

Whole genome duplication (WGD, i.e. polyploidy) occurred at many stages of eukaryotic evolution (*1–3*), including three sequential rounds in the common ancestors to all vertebrates (1R), jawed vertebrates (2R) (*4–7*) and teleost fish (3R) (*8*, *9*). While many duplicated genes created by WGD (ohnologs) (*1*) are lost, retained ohnologs are enriched in functions linked to development, signaling and transcriptional regulation (*2*, *10*). Ohnologs retained from vertebrate WGDs have gained more complex regulation and specialized in expression during evolution (*11*, *12*). Events in the common ancestors to all vertebrates (~530 Mya, 1R) (*5–7*, *13*) and teleosts (~275 Mya, 3R) (*14*) are proposed autotetraploidizations (*5–7*, *15*), where a diploid genome doubled within-species, leading to four identical chromosome sets that paired non-preferentially during meiosis, resulting in tetrasomic inheritance (*1*). A separate WGD (2R) in the jawed vertebrate ancestor may have occurred by allotetraploidization (*5–7*, *13*), where genetically distinct species hybridized, leading to two subgenomes within the same nucleus (*1*).

Rediploidization involves a return to stable bivalent pairing, resulting in the disomic inheritance of unlinked ohnolog loci that can diverge in sequence and hence function (*1*, *16*, *17*). While this may be immediate in allopolyploids with distinct subgenomes, for autopolyploids, it requires the evolution of genomic reorganizations (*18*). Understanding the functional outcomes of rediploidization is necessary to unravel the genomic basis for major vertebrate radiations. However, studies of regulatory changes following the 1R-3R WGDs are challenged by the vast evolutionary time separating these events from living species.

The salmonid ancestor underwent autotetraploidization 80-100 Mya, representing a fourth round (4R) of WGD (*19–21*). Around half of all salmonid genes are retained as functional ohnolog pairs from this event (*18*, *22*), three-fold more than the 1R-3R events (*23*). Salmonid ohnologs are located in large collinear regions that have diverged in sequence to a highly variable extent (*18*, *20*). This signature results from asynchronous rediploidization (*17*, *20*), which likewise followed WGDs at the stem of teleosts (3R) (*15*), acipenseriforms (*24*) and potentially all vertebrates (1R) (*6*). Asynchronous rediploidization creates a stratification of ohnolog divergence times within affected genomes, providing an ideal study system to explore the function and regulation of ohnologs at distinct stages of evolution. A large fraction of ohnolog pairs created by the salmonid 4R have evolved divergent expression (*18*, *25*). However, the

mechanistic underpinnings of ohnolog regulatory evolution remains poorly understood in terms of promoter and enhancer activity.

To address this gap, we generated large-scale multiomics datasets for two salmonids, which were used to investigate changes in gene expression and regulatory element activity following the 4R autotetraploidization, revealing the role of ontogeny, tissue-specific context and rediploidization in shaping genome functional evolution.

Functional and comparative annotation of salmonid genomes

Following ENCODE (*26*) and within the Functional Annotation of Animal Genomes (FAANG) initiative (*27*, *28*), AQUA-FAANG aimed to deliver comprehensive genome functional annotations for farmed fishes of global economic importance (*29*). To this end, we obtained species-matched samples from Atlantic salmon and rainbow trout (hereafter: salmon and trout), among the world's most commercially-valuable fishes (*30*), with wild populations of substantial cultural and conservation importance (*31*). Sampling spanned key stages of embryogenesis ('DevMap') and a tissue panel (for both sexes) at sexually immature and mature stages ('BodyMap') (Fig. 1A). After developing robust experimental protocols (Fig. S1; Text S1-S4), RNA-seq was used to quantify gene expression, while ATAC-seq (*32*) and ChIP-seq (for histone marks: H3K27ac, H3K4me1, H3K4me3 and H3K27me3) was used to assess chromatin openness and epigenetic state (Fig. 1B). Altogether, we generated 771 high-quality sequencing datasets (Tables S1-S3; Fig. S2-S9) summing to 77.7 billion read pairs, with raw data shared on the FAANG data portal (*33*). Processed datasets are shared at salmobase.org (*34*), alongside new comparative genomic tools for data visualization and exploration (described below).

We identified 319,718 robust open chromatin regions in salmon and 194,002 in trout, expected to harbour regulatory elements bound by transcription factors (TFs) and protein complexes associated with RNA polymerase II (*32*). To enrich the functional context of these regions, we modelled chromatin epigenetic state using ChromHMM (*35*), integrating our ChIP-seq datasets. This identified 45,802 and 28,138 unique active promoters, and 78,424 and 42,164 unique active enhancers, in the salmon and trout genomes, respectively (Fig. 1C-F, Fig. S10). We performed two analyses to validate these annotations. First, Cap Analysis Gene Expression (CAGE) (*36*) using DevMap samples in both species showed that most open chromatin regions annotated as 'Active TSS' were supported by CAGE-defined promoters, with support observed for 83–91% in salmon and 53–70% in trout (Fig. S11). Second, we found that active promoters and enhancers were enriched in vertebrate TF binding sites (TFBS), matching sample type expectations (Fig. S12-13; Text S5). For example, the TFBS targeted by the Pou5f1(Oct4):Sox2 complex, essential for zebrafish zygotic genome activation (ZGA) and gastrulation initiation (*37*), was enriched in highly active salmon and trout regulatory elements at the same developmental stages (Fig. S12). We also identified enrichment of binding sites for canonical tissue TFs in both species, with extensive between-species sharing of the same enriched TFBSs (Text S5, Fig. S13-14).

For comparative analysis, we identified 10,521 high-confidence ohnolog pairs conserved in both species and 5,832 'singletons', where one ohnolog was lost in their common ancestor (Tables S4-S5; Fig. S15). To compare non-coding regions, we adapted Cactus (*38*) to align duplicated and orthologous sequences within and between the genomes of salmon and trout (Methods; Fig. S16; Table S6), alongside single orthologous regions from their sister lineage (northern pike, *Esox lucius*) that didn't undergo the 4R WGD (*19*). Our comparative dataset captured 26,157 genes and 194,528 unique promoters or enhancers showing activity in at least one sample across

both species. We provide examples of the utility of these comparative tools to infer gene regulatory dynamics among salmonid ohnolog pairs (Fig. 1G; Fig. S17).

Dynamic genome regulation across ontogeny and tissues

We classified our DevMap and BodyMap datasets using dimensionality reduction approaches (*39*). Across embryogenesis, self-organizing maps (SOM) identified respective clusters of genes (Fig. 2A; Fig. S18-19) or genomic regions (Fig. 2B; Fig. S20-21) showing closely-related changes in expression or chromatin openness. SOM captured the maternal to zygotic transition (MZT) according to clusters of maternal mRNAs showing highest levels during late cleavage or early blastulation, with declining levels thereafter (Fig. 2A; M1-4 in salmon; M1-3 in trout), alongside genes up-regulated (Fig. 2A; B1-2 and G1; both species) and chromatin regions showing maximal openness (Fig. 2B; clusters 1-7 for salmon; 1-11 for trout) at blastulation and gastrulation stages. The maternal RNA SOM clusters showed enriched Gene Ontology (GO) terms including 'cell cycle' and 'chromosome segregation (Fig. S22-23, Text S6, Table S7-S8). Transcriptomic clusters specific to segmentation stages, leading to the most developmentally advanced pharyngula stages (S1-S2 and P1-P2; both species), were enriched for terms associated with organogenesis and energy production (Fig. 2A; Fig. S22-23, Text S6, Table S7-S8). A developmental trajectory of embryogenesis was evident, with topologically-related SOM clusters co-localizing on UMAP graphs (Fig. 2A-B). Comparing the two salmonids, closely-related SOM clusters were enriched in orthologous genes and syntenic genomic regions (Fig. 2A-B, cross-species bands between clusters; Fig. S24).

These analyses were repeated using our BodyMap datasets, capturing changes in gene expression and chromatin openness across tissues, sexes, and maturation stages that were similar in both species (Fig. S25; supported by Fig. S26-32; Tables S7-S8; Text S7).

Overall, the AQUA-FAANG datasets capture dynamic changes in genome activity across life-stages and tissue types, providing a novel framework to dissect the evolution of gene regulation following the salmonid 4R autotetraploidization.

Ohnolog regulatory divergence mirrors constraints shaping gene expression across species

Ontogeny and tissue-specific context are established to influence gene expression evolution across vertebrate species (*11*, *40–44*). Do similar constraints act on rediploidization following autotetraploidization? To address this, we quantified salmon ohnolog pair expression divergence using our transcriptomic datasets, adapting a published strategy (*11*) (Fig. 3A-B). Salmonid 4R ohnolog pairs were split into two sub-classes to explicitly account for the rediploidization process (*17*, *18*). The first consists of 8,757 'Early rediploidization' pairs, where ohnolog sequence divergence started before salmon and trout separated ~25 Mya (*19*). The second consists of 1,103 'Late rediploidization' pairs, where ohnolog divergence occurred independently in salmon and trout, owing to a prolonged phase of tetrasomic inheritance (*17*, *20*).

As expected, rediploidization was the dominant factor explaining regulatory divergence (*17*), with Early rediploidization ohnolog pairs showing greater divergence for all sample types (Fig. 3A-B; non-overlapping 95% CI between rediploidization sub-classes). Ohnolog expression divergence decreased across embryogenesis, being highest at the earliest cleavage stage (pre-MZT), and lowest at the most developmentally-advanced segmentation/pharyngula stages (Fig. 3A), thought to represent the vertebrate phylotypic period (*43*). The observed pattern of decreasing regulatory divergence across embryogenesis was highly correlated between Early and Late rediploidization ohnolog pairs (Pearson's R = 0.97, $p = 1.1 \times 10^{-8}$, Fig. S33), albeit with a

170 more pronounced drop in divergence for Late rediploidization ohnolog pairs at advanced
171 segmentation stages (Fig. 3A).

172 Adult tissue types differed extensively in ohnolog expression divergence, with brain showing the
173 least divergence for Early rediploidization pairs (Fig. 3B), consistent with past work (*45*). Sexual
174 maturation had little effect on ohnolog expression divergence for most tissues, including ovaries
175 (Fig. 3B; overlapping 95% CI). However, ohnolog expression divergence was greater in sexually
176 mature compared to immature males, specifically in testis (Fig. 3B; non-overlapping 95% CI).
177 Unlike embryogenesis, expression divergence across tissues was not significantly correlated
178 between Early and Late rediploidization ohnolog pairs (Fig. S33). This was explained by
179 increased expression divergence in ovaries, alongside reduced divergence in muscle and liver,
180 for the Late rediploidization ohnolog pairs.

181 We next explored the extent to which salmonid 4R ohnolog expression divergence mirrors
182 across-species regulatory evolution. We repeated our expression divergence analysis, this time
183 using 1-to-1 ortholog pairs from salmon and trout (Fig. 3C-D). To account for impacts of WGD
184 on regulatory evolution across species, we separated these ortholog pairs into two groups. The
185 first consists of 8,661 ortholog pairs representing Early rediploidization ohnologs (with both
186 ohnolog copies retained in salmon and trout) and the second 5,099 ortholog pairs representing
187 singletons in both species.

188 Across-species expression divergence of ortholog pairs belonging to the ohnolog group was
189 highly correlated with that of salmon ohnolog pairs across embryonic development (Pearson's R
190 = 0.96, $p = 1.0 \times 10^{-7}$), and weakly correlated across adult tissue types (Pearson's R = 0.54, $p =$
191 0.037) (Fig. S33). Expression divergence of ortholog pairs differed between the ohnolog and
192 singleton groups. In both cases, expression divergence decreased across embryogenesis, with a
193 strong drop around the MZT, matching that observed for salmon ohnolog pairs (Fig. 3A, C).
194 However, after the MZT, ortholog pairs from the ohnolog group showed reduced expression
195 divergence during segmentation, compared to singletons (Fig. 3C; lack of overlap between 95%
196 CI). This result indicates that orthologs evolving under the highest constraints on expression
197 during mid-to-late embryogenesis are overrepresented as ohnolog pairs.

198 Across tissues, we observed variable and often significant differences in how ortholog pairs have
199 evolved across species, depending on whether they represent ohnologs or singletons (Fig. 3D).
200 For gill, head kidney and distal intestine, ortholog pairs from the ohnolog group showed
201 significantly less expression divergence than singletons (Fig. 3D). However, in brain, expression
202 divergence of ortholog pairs was similar for both groups, which can be explained by
203 comparatively slower evolution of singleton orthologs (Fig. 3D). Specific to ovaries, regulatory
204 divergence of ortholog pairs across species was driven by maturation stage, rather than their
205 retention status following WGD (Fig. 3D).

206 In sum, these results reveal the nature of post-WGD regulatory evolution across ontogeny and
207 tissues, implicating pleiotropic constraints during a period of rapid organ development at mid-to-
208 late embryogenesis (*40*) as one strong factor selecting for the long-term retention of ohnologs,
209 while also highlighting ohnolog retention as a driver of tissue regulatory evolution both within
210 and across species, but with highly variable effects across tissue types.

211 **Functional evolution of regulatory elements retained from autotetraploidization**

212 Our DevMap and BodyMap atlases enable robust inferences on regulatory element evolution.
213 Using our whole genome alignment, we divided all promoters and enhancers from salmon into

three categories of post-WGD evolution: “Shared” - where two duplicated sequences retained from WGD are aligned and each shows activity; “Alignable only” – where, despite being aligned, activity is restricted to one of two duplicated sequences retained from WGD; and “Exclusive” – where the active element is limited to one chromosome (Fig. 4A). To explicitly capture the rediploidization process, we further separated these elements based on their locations within collinear blocks showing Early or Late rediploidization (*17*) (Fig. 4B; Tables S9-10).

Overall, 53.5%, 21.6% and 24.9% of active promoters belonged to the Shared, Alignable only, and Exclusive categories, respectively (Table S11). The proportion of Shared promoters is consistent with the estimated 55% retention rate for salmonid 4R ohnolog pairs (*18*). For active enhancers, 33.3%, 43.8% and 22.9% belonged to the same categories (Table S11). Compared to promoters, the lower proportion of enhancers from the Shared category, alongside higher proportion from the Alignable only category (both $p < 0.0001$, Fisher’s Exact Test) shows that the faster evolutionary turnover of enhancers observed between species (*46*, *47*) also occurs following WGD.

Early rediploidization regions harbored proportionately fewer Shared promoters and enhancers, alongside more from the Exclusive category (Text S8). Thus, a longer period of divergence has facilitated the loss of duplicated regulatory elements created by rediploidization, while providing more evolutionary opportunities for the gain of novel regulatory elements post-WGD.

Ontogenetic and tissue-specific constraints on regulatory element evolution

We next explored variation in the functional retention of regulatory elements across sample types (Fig. 4C-D; supporting tests in Tables S12-S13), revealing that promoters were active in more sample types than enhancers (Fig. S34), consistent with past work (*48*). For both types of regulatory element, the breadth of activity across sample types and extent of sample type-specific activity varied across alignment categories (Fig. S35, Text S8). Mirroring ohnolog expression divergence (Fig. 3), rediploidization was the dominant factor determining the proportion of Shared promoters and enhancers across all sample types (Fig. 4C-D).

The proportion of active promoters per alignment category showed limited variation across sample types, with Shared promoters most common (Fig. 4C; Text S8). Much greater heterogeneity across sample types was observed for enhancers (Fig. 4D). During embryogenesis, the highest proportion of Shared enhancers occurred at late somitogenesis, reaching a proportion similar to brain - the highest among adult tissues (Fig. 4D). Mature testis showed the highest proportion of Exclusive enhancers, while liver and gonads showed the lowest proportion of Shared enhancers (Fig. 4D).

Thus, the retention of enhancers duplicated following the salmonid 4R WGD was strongly shaped by ontogeny and tissue-specific context, whereas promoter evolution was less dynamic.

Enhancer evolution is coupled to ohnolog expression regulation

Cis-regulatory elements govern transcriptional regulation (*47*, *49–51*), and in mammals, enhancer evolution is linked to rates of gene expression evolution (*46*). However, the association between enhancer evolution and ohnolog expression divergence following WGD remains poorly understood. We addressed this knowledge gap, using analyses that did not split genes by rediploidization history to exploit the stratification of ohnolog sequence divergence when making inferences on regulatory changes.

We first asked if enhancer alignment category is correlated with ohnolog regulatory evolution across sample types. We observed that as ohnolog expression divergence increased, the

258 proportion of Shared enhancers decreased, while the proportion of Alignable only and Exclusive
259 enhancers increased (Table S14). While the observed correlation directions are consistent with
260 enhancers contributing to ohnolog regulatory divergence, each test failed to reached significance
261 (Pearson's R = -0.41, 0.39, 0.37 for Shared, Alignable only and Exclusive enhancers,
262 respectively; each *p*-adj > 0.17). However, significant correlations were observed when
263 performing the same tests for promoters, in the same direction as enhancers (respective Pearson's
264 R = -0.72, 0.78 and 0.59; *p*-adj = 0.004, 0.002 and 0.02 for Shared, Alignable only and Exclusive
265 promoters) (Table S14). Thus, the functional retention of promoters following WGD is more
266 directly linked to ohnolog regulatory evolution than enhancers.

267 However, the above analyses indirectly test for a relationship between enhancer evolution and
268 regulatory divergence among ohnolog pairs at the sample level. As a second more direct test, we
269 defined *cis*-regulatory domains for all salmon genes then sub-divided ohnolog pairs into
270 "conserved" and "diverged" regulation classes based on expression correlation across tissues and
271 development (Fig. 5A-B, Fig. S36), before comparing the composition and conservation of
272 enhancers assigned to ohnolog pairs from each class. This distance-based approach assumes gene
273 transcription is regulated by nearby enhancers within the defined *cis*-regulatory domains. While
274 some of these gene-to-enhancer assignments are likely inaccurate, e.g. due to long-distance
275 regulation, the proximity of enhancers to promoters has been shown to be a highly accurate
276 predictor for their physical interaction (*52*), with the same approach used for cross-species
277 analyses (*46*).

278 We found that ohnolog pairs showing conserved regulation had more active enhancers ($p = 2.2 \times$
279 10^{-16}, Wilcoxon test) (Fig. S37). They also showed a higher proportion of Shared enhancers ($p =$
280 8.0×10^{-26}, odds ratio = 2.04) and reduced proportion of Exclusive enhancers ($p = 5.8 \times 10^{-37}$;
281 odds ratio = 0.62) (Fig. 5C). Breaking down enhancer characteristics by sample type, ohnologs
282 with conserved expression had higher levels of Shared enhancers active in the same samples
283 (Fig. S37, Wilcoxon test: $p = 1.2 \times 10^{-6}$). On the other hand, Alignable only enhancers were
284 equally present in ohnologs with diverged or conserved regulation (Fig. S37).

285 To further explore the link between enhancers and ohnolog pair expression divergence, we
286 calculated differences in total enhancer activity profiles across sample types, quantified by
287 Euclidean distance (termed 'enhancer-usage distance', see Methods) (Fig. 5D). Ohnolog pairs
288 showing divergent expression had higher enhancer-usage distances than pairs showing conserved
289 expression (Fig. 5D; respective medians = 1.15 vs 0.89; $p = 1.6 \times 10^{-28}$). We further asked if
290 enhancer usage distance is associated with ohnolog pair expression distance (see Methods; Fig.
291 5E, Fig. S38). While ohnolog expression distance decreased across embryogenesis, consistent
292 with our JSD-based analysis (Fig. 3), enhancer usage distance was more variable. This suggests
293 ohnolog expression divergence during embryogenesis is mainly constrained by Shared
294 enhancers, rather than total enhancer activity profiles. Conversely, we observed stronger across-
295 tissue differences in both ohnolog expression distance (Kruskal-Wallis test $\chi^2 = 166.6$, df = 4, $p <$
296 2.2×10^{-16}) and enhancer-usage distance ($\chi^2 = 77.2$, df = 4, $p = 6.7 \times 10^{-16}$). Brain showed the
297 lowest distance and dispersion in both cases, whereas liver, ovary, and testis displayed the
298 broadest range of values (Fig. S38).

299 Overall, these results reveal that the evolution of ohnolog expression is associated with the loss
300 of ancestral enhancer sequences or the gain of new enhancer sequences post-WGD.

301 **Transposable elements have shaped enhancer evolution after WGD**

302 One mechanism to evolve new enhancer sequences is through transposable element (TE)
303 activity. Indeed, the salmonid WGD was associated with a burst of TE activity, which led to
304 increased rates of TE derived *cis*-regulatory element evolution (*52*). On the other hand, TEs also
305 disable functional elements within genomes, meaning their overlap with regulatory elements
306 provides insights into constraints on gene regulation across ontogeny or tissues. Yet, how TEs
307 contribute to enhancer functional evolution following WGD remains poorly understood. To
308 explore these questions, we identified all active enhancers within *cis*-regulatory domains from
309 salmon that overlapped with a high-quality TE annotation (*52*). Overall, we found that Exclusive
310 enhancers overlapped TEs to a much larger extent (>2 fold) than Shared enhancers (Fig. 5F).
311 This could be explained by TE insertions either acting to create *de novo* enhancers (*53*, *54*), or by
312 reduced constraint on the functions of Exclusive enhancers.

313 To assess the contribution of TEs to ohnolog regulatory divergence, we contrasted the TE-
314 sequence content in enhancers of ohnolog pairs with conserved or diverged regulation (Fig. 5F-
315 G; Fig. S39). The proportion of TE sequences in Exclusive enhancers was weakly associated
316 with ohnolog regulatory divergence (Fig. 5F, $p = 0.0031$). However, this was not the case for
317 Shared enhancers (Fig. 5F). The most active TE group following WGD in salmonids are DNA
318 transposons, specifically the TcMar (DNA/TcMar) (*52*). In line with this, the overlap of DNA
319 TEs with Exclusive enhancers was increased for diverged ohnolog pairs (Fig. 5F; $p = 0.038$ for
320 all DNA elements; $p = 0.0051$ for DNA/TcMar elements), which was not the case for
321 retrotransposons. Further analysis highlighted an increased overlap between DNA or
322 DNA/TcMar TEs and Exclusive enhancers for ohnologs showing divergent regulation in early
323 somitogenesis, liver and testis (Fig. 5G, Fig. S39).

324 Past work has hypothesised that the contribution of TEs towards gene regulation following the
325 salmonid WGD was shaped by tissue-specific selective constraint (*53*). We tested this hypothesis
326 using our panel of tissues and embryonic stages. Enhancers active in brain and late
327 embryogenesis, representing sample types where ohnolog expression divergence and enhancer
328 functional conservation were most constrained following WGD (Fig. 3-4), showed the lowest
329 overall overlap with TEs for enhancers located within both the Early and Late rediploidization
330 categories (Fig. S40). Further, as the overlap between TEs and active enhancers increased across
331 sample types, so did the proportion of Alignable only and Exclusive enhancers, while the
332 proportion of Shared enhancers decreased (Fig. S40; Table S15).

333 Together, these data indicate that TE insertions broadly contribute to regulatory evolution
334 following WGD, acting to disable duplicated enhancers and create new enhancers. Moreover,
335 selection against TE insertions appears strongest for enhancers retained as duplicates, especially
336 those active in sample types under the most constrained regulatory evolution following WGD.

337 **Ontogeny governs the evolution of enhancer-CNEs following WGD**

338 The strongest retention of Shared enhancers following WGD was for late embryogenesis and
339 brain (Fig. 4C). A past mammalian study profiled different tissues across ontogeny to reveal that
340 enhancers most active in brain were also most conserved across species, especially at stages
341 showing prominent brain development (*55*). Building on this finding, we asked if the
342 conservation of enhancer sequences across species influences their evolution following WGD,
343 and if so, whether this is shaped by distinct constraints across ontogeny and tissue types.

344 Here, we exploited the fact that conserved non-coding elements (CNEs) often lie within
345 enhancers governing development and tissue identity, marking functionally important regulatory
346 sequences under purifying selection across species (*56*, *57*). We overlapped all salmon active

347 enhancers with a genome-wide dataset of northern pike CNEs stratified by phylogenetic age,
348 with the oldest category predating the split of ray- and lobe-finned fish (Fig. 6; Table S16). We
349 investigated the retention of salmon CNEs located within active enhancers (hereafter: enhancer-
350 CNEs) following WGD by their alignment to pike CNEs (Table S17). Using this approach,
351 Exclusive enhancer-CNEs represent singletons, where one in a former ohnolog pair was lost,
352 while the retained CNE is conserved with pike. Among 24,007 high-confidence enhancer-CNEs,
353 61.4%, 31.2% and 7.3% belonged to the Shared, Alignable only and Exclusive categories,
354 respectively (Fig. 6A). Thus, 92.6% of all enhancer-CNEs were retained as duplicated sequences
355 from WGD, with activity in both duplicates (Shared category) being around twice as common as
356 activity restricted to one locus (Alignable only category).

357 The most common enhancer-CNEs (35.5%) belonged to the teleost age category (Fig. 6A; Fig.
358 S41). Significant overrepresentation of enhancer-CNEs from the Shared category was observed
359 for the two oldest CNE categories, comprising 23.6% of all enhancer-CNEs (Fig. 6A). Thus,
360 selection on the retention of enhancer-CNEs following WGD was strongest for the most ancient
361 vertebrate enhancers.

362 Conversely, the small number of Exclusive enhancer-CNEs were enriched in the youngest CNEs
363 and underrepresented in CNEs of the teleost age category (Fig. 6A). Shared enhancer-CNE
364 sequence pairs showed higher sequence conservation than the Alignable only category, both
365 overall (respective median nucleotide sequence identity: 95.0 vs. 92.9%; Wilcoxon test, $p = 2.2 \times$
366 10^{-16}) or stratifying by CNE age category (Fig. S41). For enhancer-CNEs from the Alignable
367 only category, the CNE that lost enhancer activity has evolved significantly faster (reduced
368 nucleotide percentage identity with single orthologous pike CNE), with the youngest CNEs most
369 affected (Fig. S41). Thus, the phylogenetic age and functional retention of enhancer-CNEs
370 strongly shapes their ongoing sequence evolution following WGD.

371 Enhancer-CNEs were strongly overrepresented in Late rediploidization regions, overall and for
372 each phylogenetic age category (Table S18). This effect was entirely explained by Shared
373 enhancer-CNEs, which were overrepresented overall (Fig. S42), and per age category, while
374 enhancer-CNEs from the Alignable only and Exclusive categories were enriched in Early regions
375 (Fig. S42; Table S19). Remarkably, Shared enhancer-CNEs within Late regions comprised
376 39.9% of all Shared enhancer-CNE space in the genome, greater than two-fold higher than
377 expected by chance based on the length of these regions (Table S19). This overrepresentation
378 strongly suggests selection acted against the rediploidization and divergence of these elements
379 during the early stages of post-WGD evolution.

380 Finally, we partitioned enhancer-CNEs from Early and Late rediploidization regions across
381 alignment categories, CNE phylogenetic ages, and sample types (Fig. 6B). These data were
382 normalised to allow statistical comparisons in all combinations (Fisher's Exact Test results:
383 Tables S20-22). A strong enrichment of Shared enhancer-CNEs from the two oldest CNE
384 categories was observed at late somitogenesis compared to earlier embryonic stages for both
385 Early and Late regions (Fig. 6B; Table S22; $p = 0$). Specific to the oldest CNE category, in both
386 Early and Late regions, the proportion of Shared enhancer-CNEs at late somitogenesis was
387 significantly higher than brain (Fig. 6B; Table S22; $p = 0$). Comparing adult tissue types, brain
388 captured the highest proportion of enhancer-CNEs for both rediploidization regions, with a
389 significantly higher proportion of Shared-enhancer CNEs from the teleost-specific and
390 Protacanthpterygii categories compared to embryonic stages and other adult tissues (Fig. 6B). In
391 Late rediploidization regions, Shared enhancer-CNEs dominated all CNE age categories except

392 early stages of embryogenesis (Table S21). In Early regions, a smaller proportion of enhancer-
393 CNEs were active in liver, testis and ovary compared to muscle and brain (Fig. 6B).

394 Overall, these results highlight a key role for development and ontogeny in shaping the evolution
395 of enhancer-CNEs following WGD. The striking overrepresentation of the oldest CNE category
396 in Shared enhancer-CNEs active at late somitogenesis is consistent with post-WGD constraints
397 on ancient regulatory programs governing vertebrate organogenesis (*40*), and predicts that
398 similar patterns will follow other vertebrate WGD events.

399 **Discussion**

400 This study paints a novel perspective of genome evolution following autotetraploidization, where
401 ohnolog regulatory evolution is governed by the same ontogenetic and tissue-specific constraints
402 acting between species. Multiple lines of evidence capture a period of within-genome constraint
403 on regulatory evolution at late embryogenesis, which can be explained by the pleiotropic
404 regulatory networks governing organogenesis at this life stage in all vertebrates (*40*). Similarly,
405 in adult tissues, the brain represents the maxima of post-WGD regulatory constraint, while
406 gonads and liver experience weaker regulatory selection (*44*, *55*). While our study advances
407 understanding of functional evolution following the salmonid 4R WGD, additional data types
408 can penetrate additional levels of regulation, including 3D chromatin architecture, while
409 resolving changes to specific cell types (*58*).

410 Our findings illustrate the importance of better capturing rediploidization processes in future
411 studies of regulatory evolution following WGD. This comes into sharp focus in light of recent
412 discoveries that asynchronous rediploidization followed other autotetraploidization events (*6*, *15*,
413 *24*) and may represent a 'rule' of genome evolution in vertebrates, and possibly other eukaryotic
414 lineages. Thus, the stratification of regulatory divergence we describe in salmonid genomes may
415 have occurred in many genomes, including during earlier stages of vertebrate diversification. In
416 this respect, salmonids provide an ideal vertebrate system to further explore the regulatory basis
417 of functional innovations derived from autopolyploidization, including lineage-specific novelties
418 arising following delayed rediploidization (*17*).

419 Finally, our study marks the sharing of an open genomic resource in highly processed formats,
420 ready for analysis by the research community and other stakeholders. This resource can be
421 readily leveraged to support precision breeding goals in salmonid aquaculture (*28*, *29*), for
422 instance by helping to dissect traits with a regulatory basis, supporting the long-term
423 sustainability of global food production. These datasets will also enable explorations of the role
424 played by regulatory elements in the adaptation of wild salmonid populations to rapidly changing
425 environments across their natural ranges, while enriching comparative analyses of regulatory
426 evolution across vertebrate lineages.

427 **References**

428 1. S. Ohno, *Evolution by Gene Duplication* (Springer, New York, NY, 1971).

429 2. Y. Van de Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy. *Nat.*
430 *Rev. Genet.* **18**, 411–424 (2017).

431 3. J. W. Clark, P. C. J. Donoghue, Whole-genome duplication and plant macroevolution.
432 *Trends Plant Sci.* **23**, 933–945 (2018).

433 4. P. Dehal, J. L. Boore, Two rounds of whole genome duplication in the ancestral vertebrate.
434 *PLoS Biol.* **3**, e314 (2005).

435 5. O. Simakov, F. Marlétaz, J.-X. Yue, B. O'Connell, J. Jenkins, A. Brandt, R. Calef, C.-H.
436 Tung, T.-K. Huang, J. Schmutz, N. Satoh, J.-K. Yu, N. H. Putnam, R. E. Green, D. S.
437 Rokhsar, Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol.*
438 *Evol.* **4**, 820–830 (2020).

439 6. F. Marlétaz, N. Timoshevskaya, V. A. Timoshevskiy, E. Parey, O. Simakov, D.
440 Gavriouchkina, M. Suzuki, K. Kubokawa, S. Brenner, J. J. Smith, D. S. Rokhsar, The
441 hagfish genome and the evolution of vertebrates. *Nature* **627**, 811–820 (2024).

442 7. D. Yu, Y. Ren, M. Uesaka, A. J. S. Beavan, M. Muffato, J. Shen, Y. Li, I. Sato, W. Wan, J.
443 W. Clark, J. N. Keating, E. M. Carlisle, R. P. Dearden, S. Giles, E. Randle, R. S. Sansom, R.
444 Feuda, J. F. Fleming, F. Sugahara, C. Cummins, M. Patricio, W. Akanni, S. D'Aniello, C.
445 Bertolucci, N. Irie, C. Alev, G. Sheng, A. de Mendoza, I. Maeso, M. Irimia, B. Fromm, K. J.
446 Peterson, S. Das, M. Hirano, J. P. Rast, M. D. Cooper, J. Paps, D. Pisani, S. Kuratani, F. J.
447 Martin, W. Wang, P. C. J. Donoghue, Y. E. Zhang, J. Pascual-Anaya, Hagfish genome
448 elucidates vertebrate whole-genome duplication events and their evolutionary consequences.
449 *Nat. Ecol. Evol.* **8**, 519–535 (2024).

450 8. O. Jaillon, J.-M. Aury, F. Brunet, J.-L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau,
451 C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G. Lutfalla, C.
452 Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S. Castellano, V.
453 Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biémont, Z. Skalli, L.
454 Cattolico, J. Poulain, V. De Berardinis, C. Cruaud, S. Duprat, P. Brottier, J.-P. Coutanceau,
455 J. Gouzy, G. Parra, G. Lardier, C. Chapple, K. J. McKernan, P. McEwan, S. Bosak, M.
456 Kellis, J.-N. Volff, R. Guigó, M. C. Zody, J. Mesirov, K. Lindblad-Toh, B. Birren, C.
457 Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F. Quétier, W. Saurin,
458 C. Scarpelli, P. Wincker, E. S. Lander, J. Weissenbach, H. Roest Crollius, Genome
459 duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-
460 karyotype. *Nature* **431**, 946–957 (2004).

461 9. V. Ravi, B. Venkatesh, The divergent genomes of teleosts. *Annu. Rev. Anim. Biosci.* **6**, 47–
462 68 (2018).

463 10. F. G. Brunet, H. Roest Crollius, M. Paris, J.-M. Aury, P. Gibert, O. Jaillon, V. Laudet, M.
464 Robinson-Rechavi, Gene loss and evolutionary rates following whole-genome duplication in
465 teleost fishes. *Mol. Biol. Evol.* **23**, 1808–1816 (2006).

466 11. F. Marlétaz, P. N. Firbas, I. Maeso, J. J. Tena, O. Bogdanovic, M. Perry, C. D. R. Wyatt, E.
467 de la Calle-Mustienes, S. Bertrand, D. Burguera, R. D. Acemel, S. J. van Heeringen, S.
468 Naranjo, C. Herrera-Ubeda, K. Skvortsova, S. Jimenez-Gancedo, D. Aldea, Y. Marquez, L.
469 Buono, I. Kozmikova, J. Permanyer, A. Louis, B. Albuixech-Crespo, Y. Le Petillon, A.
470 Leon, L. Subirana, P. J. Balwierz, P. E. Duckett, E. Farahani, J.-M. Aury, S. Mangenot, P.
471 Wincker, R. Albalat, È. Benito-Gutiérrez, C. Cañestro, F. Castro, S. D'Aniello, D. E. K.
472 Ferrier, S. Huang, V. Laudet, G. A. B. Marais, P. Pontarotti, M. Schubert, H. Seitz, I.
473 Somorjai, T. Takahashi, O. Mirabeau, A. Xu, J.-K. Yu, P. Carninci, J. R. Martinez-Morales,

H. R. Crollius, Z. Kozmik, M. T. Weirauch, J. Garcia-Fernàndez, R. Lister, B. Lenhard, P. W. H. Holland, H. Escriva, J. L. Gómez-Skarmeta, M. Irimia, Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64–70 (2018).

12. F. Mottes, C. Villa, M. Osella, M. Caselle, The impact of whole genome duplications on the human gene regulatory networks. *PLoS Comput. Biol.* **17**, e1009638 (2021).

13. Y. Nakatani, P. Shingate, V. Ravi, N. E. Pillai, A. Prasad, A. McLysaght, B. Venkatesh, Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. *Nat. Commun.* **12**, 4489 (2021).

14. M. Qi, J. Clark, E. R. R. Moody, D. Pisani, P. C. J. Donoghue, Molecular dating of the teleost whole genome duplication (3R) is compatible with the expectations of delayed rediploidization. *Genome Biol. Evol.* **16** (2024).

15. E. Parey, A. Louis, J. Montfort, Y. Guiguen, H. R. Crollius, C. Berthelot, An atlas of fish genome evolution reveals delayed rediploidization following the teleost whole-genome duplication. *Genome Res.* **32**, 1685–1697 (2022).

16. K. J. Martin, P. W. H. Holland, Enigmatic orthology relationships between Hox clusters of the African butterfly fish and other teleosts following ancient whole-genome duplication. *Mol. Biol. Evol.* **31**, 2592–2611 (2014).

17. F. M. Robertson, M. K. Gundappa, F. Grammes, T. R. Hvidsten, A. K. Redmond, S. Lien, S. A. M. Martin, P. W. H. Holland, S. R. Sandve, D. J. Macqueen, Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* **18** (2017).

18. S. Lien, B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent, T. Nome, T. R. Hvidsten, J. S. Leong, D. R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R. A. Hermansen, K. von Schalburg, E. B. Rondeau, A. Di Genova, J. K. A. Samy, J. Olav Vik, M. D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D. I. Våge, P. de Jong, T. Moen, M. Baranski, Y. Palti, D. R. Smith, J. A. Yorke, A. J. Nederbragt, A. Tooming-Klunderud, K. S. Jakobsen, X. Jiang, D. Fan, Y. Hu, D. A. Liberles, R. Vidal, P. Iturra, S. J. M. Jones, I. Jonassen, A. Maass, S. W. Omholt, W. S. Davidson, The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–205 (2016).

19. D. J. Macqueen, I. A. Johnston, A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. Biol. Sci.* **281**, 20132881 (2014).

20. M. K. Gundappa, T.-H. To, L. Grønvold, S. A. M. Martin, S. Lien, J. Geist, D. Hazlerigg, S. R. Sandve, D. J. Macqueen, Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of Salmonid evolution. *Mol. Biol. Evol.* **39** (2022).

21. F. W. Allendorf, G. H. Thorgaard, "Tetraploidy and the evolution of Salmonid fishes" in *Evolutionary Genetics of Fishes* (Springer US, Boston, MA, 1984; http://dx.doi.org/10.1007/978-1-4684-4652-4_1), pp. 1–53.

22. C. Berthelot, F. Brunet, D. Chalopin, A. Juanchich, M. Bernard, B. Noël, P. Bento, C. Da Silva, K. Labadie, A. Alberti, J.-M. Aury, A. Louis, P. Dehais, P. Bardou, J. Montfort, C. Klopp, C. Cabau, C. Gaspin, G. H. Thorgaard, M. Boussaha, E. Quillet, R. Guyomard, D. Galiana, J. Bobe, J.-N. Volff, C. Genêt, P. Wincker, O. Jaillon, H. Roest Crollius, Y. Guiguen, The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* **5**, 3657 (2014).

23. P. P. Singh, H. Isambert, OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates. *Nucleic Acids Res.* **48**, D724–D730 (2020).

24. A. K. Redmond, D. Casey, M. K. Gundappa, D. J. Macqueen, A. McLysaght, Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nat. Commun.* **14**, 2879 (2023).

25. G. B. Gillard, L. Grønvold, L. L. Røsæg, M. M. Holen, Ø. Monsen, B. F. Koop, E. B. Rondeau, M. K. Gundappa, J. Mendoza, D. J. Macqueen, R. V. Rohlfs, S. R. Sandve, T. R. Hvidsten, Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol.* **22**, 103 (2021).

26. ENCODE Project Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Author Correction: Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **605**, E3 (2022).

27. E. Giuffra, C. K. Tuggle, FAANG Consortium, Functional Annotation of Animal Genomes (FAANG): Current achievements and roadmap. *Annu. Rev. Anim. Biosci.* **7**, 65–88 (2019).

28. E. L. Clark, A. L. Archibald, H. D. Daetwyler, M. A. M. Groenen, P. W. Harrison, R. D. Houston, C. Kühn, S. Lien, D. J. Macqueen, J. M. Reecy, D. Robledo, M. Watson, C. K. Tuggle, E. Giuffra, From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol.* **21**, 285 (2020).

29. I. A. Johnston, M. P. Kent, P. Boudinot, M. Looseley, L. Bargelloni, S. Faggion, G. A. Merino, G. R. Ilsley, J. Bobe, C. S. Tsigenopoulos, J. Robertson, P. W. Harrison, P. Martinez, D. Robledo, D. J. Macqueen, S. Lien, Advancing fish breeding in aquaculture through genome functional annotation. *Aquaculture* **583**, 740589 (2024).

30. Food and Agriculture Organization, *The State of World Fisheries and Aquaculture 2022 (SOFIA)* (Food & Agriculture Organization of the United Nations (FAO), Rome, Italy, 2022)*The State of World Fisheries and Aquaculture*.

31. D. J. Macqueen, C. R. Primmer, R. D. Houston, B. F. Nowak, L. Bernatchez, S. Bergseth, W. S. Davidson, C. Gallardo-Escárate, T. Goldammer, Y. Guiguen, P. Iturra, J. W. Kijas, B. F. Koop, S. Lien, A. Maass, S. A. M. Martin, P. McGinnity, M. Montecino, K. A. Naish, K. M. Nichols, K. Ólafsson, S. W. Omholt, Y. Palti, G. S. Plastow, C. E. Rexroad Rd, M. L. Rise, R. J. Ritchie, S. R. Sandve, P. M. Schulte, A. Tello, R. Vidal, J. O. Vik, A. Wargelius, J. M. Yáñez, FAASG Consortium, Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* **18**, 484 (2017).

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

33. P. W. Harrison, A. Sokolov, A. Nayak, J. Fan, D. Zerbino, G. Cochrane, P. Flicek, The FAANG Data Portal: Global, open-access, "FAIR", and richly validated genotype to phenotype data for high-quality Functional Annotation of ANimal Genomes. *Front. Genet.* **12**, 639238 (2021).

34. J. K. A. Samy, T. D. Mulugeta, T. Nome, S. R. Sandve, F. Grammes, M. P. Kent, S. Lien, D. I. Våge, SalmoBase: an integrated molecular data resource for Salmonid species. *BMC Genomics* **18** (2017).

35. J. Ernst, M. Kellis, ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

36. M. S. Morioka, H. Kawaji, H. Nishiyori-Sueki, M. Murata, M. Kojima-Ishiyama, P. Carninci, M. Itoh, Cap analysis of gene expression (CAGE): A quantitative and genome-wide assay of transcription start sites. *Methods Mol. Biol.* **2120**, 277–301 (2020).

37. M. T. Lee, A. R. Bonneau, C. M. Takacs, A. A. Bazzini, K. R. DiVito, E. S. Fleming, A. J. Giraldez, Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).

38. J. Armstrong, G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, D. Genereux, J. Johnson, V. D. Marinescu, J. Alföldi, R. S. Harris, K. Lindblad-Toh, D. Haussler, E. Karlsson, E. D. Jarvis, G. Zhang, B. Paten, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).

39. D. Baranasic, M. Hörtenhuber, P. J. Balwierz, T. Zehnder, A. K. Mukarram, C. Nepal, C. Várnai, Y. Hadzhiev, A. Jimenez-Gonzalez, N. Li, J. Wragg, F. M. D'Orazio, D. Relic, M. Pachkov, N. Díaz, B. Hernández-Rodríguez, Z. Chen, M. Stoiber, M. Dong, I. Stevens, S. E. Ross, A. Eagle, R. Martin, O. Obasaju, S. Rastegar, A. C. McGarvey, W. Kopp, E. Chambers, D. Wang, H. R. Kim, R. D. Acemel, S. Naranjo, M. Łapiński, V. Chong, S. Mathavan, B. Peers, T. Sauka-Spengler, M. Vingron, P. Carninci, U. Ohler, S. A. Lacadie, S. M. Burgess, C. Winata, F. van Eeden, J. M. Vaquerizas, J. L. Gómez-Skarmeta, D. Onichtchouk, B. J. Brown, O. Bogdanovic, E. van Nimwegen, M. Westerfield, F. C. Wardle, C. O. Daub, B. Lenhard, F. Müller, Multiomic atlas with functional stratification and

592 developmental dynamics of zebrafish cis-regulatory elements. *Nat. Genet.* **54**, 1037–1050
593 (2022).

594 40. H. Hu, M. Uesaka, S. Guo, K. Shimai, T.-M. Lu, F. Li, S. Fujimoto, M. Ishikawa, S. Liu, Y.
595 Sasagawa, G. Zhang, S. Kuratani, J.-K. Yu, T. G. Kusakabe, P. Khaitovich, N. Irie,
596 EXPANDE Consortium, Constrained vertebrate evolution by pleiotropic genes. *Nat. Ecol.*
597 *Evol.* **1**, 1722–1730 (2017).

598 41. B. Piasecka, P. Lichocki, S. Moretti, S. Bergmann, M. Robinson-Rechavi, The hourglass and
599 the early conservation models--co-existing patterns of developmental constraints in
600 vertebrates. *PLoS Genet.* **9**, e1003476 (2013).

601 42. J. Liu, R. R. Viales, P. Khoueiry, J. P. Reddington, C. Girardot, E. E. M. Furlong, M.
602 Robinson-Rechavi, The hourglass model of evolutionary conservation during embryogenesis
603 extends to developmental enhancers with signatures of positive selection. *Genome Res.* **31**,
604 1573–1581 (2021).

605 43. N. Irie, S. Kuratani, Comparative transcriptome analysis reveals vertebrate phylotypic period
606 during organogenesis. *Nat. Commun.* **2**, 248 (2011).

607 44. M. Cardoso-Moreira, J. Halbert, D. Valloton, B. Velten, C. Chen, Y. Shao, A. Liechti, K.
608 Ascenção, C. Rummel, S. Ovchinnikova, P. V. Mazin, I. Xenarios, K. Harshman, M. Mort,
609 D. N. Cooper, C. Sandi, M. J. Soares, P. G. Ferreira, S. Afonso, M. Carneiro, J. M. A.
610 Turner, J. L. VandeBerg, A. Fallahshahroudi, P. Jensen, R. Behr, S. Lisgo, S. Lindsay, P.
611 Khaitovich, W. Huber, J. Baker, S. Anders, Y. E. Zhang, H. Kaessmann, Gene expression
612 across mammalian organ development. *Nature* **571**, 505–509 (2019).

613 45. S. Varadharajan, S. R. Sandve, G. B. Gillard, O. K. Tørresen, T. D. Mulugeta, T. R.
614 Hvidsten, S. Lien, L. Asbjørn Vøllestad, S. Jentoft, A. J. Nederbragt, K. S. Jakobsen, The
615 Grayling genome reveals selection on gene expression regulation after whole-genome
616 duplication. *Genome Biol. Evol.* **10**, 2785–2800 (2018).

617 46. C. Berthelot, D. Villar, J. E. Horvath, D. T. Odom, P. Flicek, Complexity and conservation
618 of regulatory landscapes underlie evolutionary resilience of mammalian gene expression.
619 *Nat. Ecol. Evol.* **2**, 152–163 (2018).

620 47. D. Villar, C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R.
621 Deaville, J. T. Erichsen, A. J. Jasinska, J. M. A. Turner, M. F. Bertelsen, E. P. Murchison, P.
622 Flicek, D. T. Odom, Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566
623 (2015).

624 48. L. L. Colbran, L. Chen, J. A. Capra, Sequence characteristics distinguish transcribed
625 enhancers from promoters and predict their breadth of activity. *Genetics* **211**, 1205–1217
626 (2019).

627 49. L. Serebreni, A. Stark, Insights into gene regulation: From regulatory genomic elements to
628 DNA-protein and protein-protein interactions. *Curr. Opin. Cell Biol.* **70**, 58–66 (2021).

629 50. P. J. Wittkopp, G. Kalay, Cis-regulatory elements: molecular mechanisms and evolutionary
630 processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2011).

631 51. H. K. Long, S. L. Prescott, J. Wysocka, Ever-changing landscapes: Transcriptional
632 enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).

633 52. S. Kim, J. Wysocka, Deciphering the multi-scale, quantitative cis-regulatory code. *Mol. Cell*
634 **83**, 373–392 (2023).

635 53. Ø. Monsen, L. Grønvold, A. Datsomor, T. Harvey, J. Kijas, A. Suh, T. R. Hvidsten, S. R.
636 Sandve, The role of transposon activity in shaping cis-regulatory element evolution after
637 whole-genome duplication. *Genome Res.* **35**, 475–488 (2025).

638 54. J.-P. Verta, H. J. Barton, V. Pritchard, C. R. Primmer, Genetic drift dominates genome-wide
639 regulatory evolution following an ancient whole-genome duplication in Atlantic salmon.
640 *Genome Biol. Evol.* **13** (2021).

641 55. A. S. Nord, M. J. Blow, C. Attanasio, J. A. Akiyama, A. Holt, R. Hosseini, S.
642 Phouanenavong, I. Plajzer-Frick, M. Shoukry, V. Afzal, J. L. R. Rubenstein, E. M. Rubin, L.
643 A. Pennacchio, A. Visel, Rapid and pervasive changes in genome-wide enhancer usage
644 during mammalian development. *Cell* **155**, 1521–1531 (2013).

645 56. D. Polychronopoulos, J. W. D. King, A. J. Nash, G. Tan, B. Lenhard, Conserved non-coding
646 elements: developmental gene regulation meets genome organization. *Nucleic Acids Res.* **45**,
647 12611–12624 (2017).

648 57. J. R. Martinez-Morales, Toward understanding the evolution of vertebrate gene regulatory
649 networks: comparative genomics and epigenomic approaches. *Brief. Funct. Genomics* **15**,
650 315–321 (2016).

651 58. R. S. Taylor, R. Ruiz Daniels, D. J. Macqueen, Cell type resolved expression of duplicate
652 genes retained from whole genome duplication in Atlantic salmon. *Genome Biol. Evol.* **17**
653 (2025).

654 **Supplementary References**

655 59. M. R. Corces, A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S.
656 Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R.
657 Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A. Kundaje, P. A. Khavari,
658 T. J. Montine, W. J. Greenleaf, H. Y. Chang, An improved ATAC-seq protocol reduces
659 background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

660 60. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying
661 chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9 (2015).

662 61. Y. Ohkawa, C. Mallappa, C. S. D. Vallaster, A. N. Imbalzano, Isolation of nuclei from
663 skeletal muscle satellite cells and myofibers for use in chromatin immunoprecipitation
664 assays. *Methods Mol. Biol.* **798**, 517–530 (2012).

62. R. Ruiz Daniels, R. S. Taylor, R. Dobie, S. Salisbury, J. J. Furniss, E. Clark, D. J. Macqueen, D. Robledo, A versatile nuclei extraction protocol for single nucleus sequencing in non-model species-Optimization in various Atlantic salmon tissues. *PLoS One* **18**, e0285020 (2023).

63. C. Schmidl, A. F. Rendeiro, N. C. Sheffield, C. Bock, ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).

64. P. A. Ewels, A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M. U. Garcia, P. Di Tommaso, S. Nahnsen, The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).

65. B. E. Langer, A. Amaral, M.-O. Baudement, F. Bonath, M. Charles, P. K. Chitneedi, E. L. Clark, P. Di Tommaso, S. Djebali, P. A. Ewels, S. Eynard, J. A. Fellows Yates, D. Fischer, E. W. Floden, S. Foissac, G. Gabernet, M. U. Garcia, G. Gillard, M. K. Gundappa, C. Guyomar, C. Hakkaart, F. Hanssen, P. W. Harrison, M. Hörtenhuber, C. Kurylo, C. Kühn, S. Lagarrigue, D. Lallias, D. J. Macqueen, E. Miller, J. Mir-Pedrol, G. C. M. Moreira, S. Nahnsen, H. Patel, A. Peltzer, F. Pitel, Y. Ramayo-Caldas, M. da C. Ribeiro-Dantas, D. Rocha, M. Salavati, A. Sokolov, J. Espinosa-Carrasco, C. Notredame, T. N.-C. Community, Empowering bioinformatics communities with Nextflow and nf-core. *Genome Biol.* **26**, 228 (2025).

66. H. Patel, J. Manning, P. Ewels, M. U. Garcia, A. Peltzer, R. Hammarén, O. Botvinnik, A. Talbot, F. Hanssen, G. Sturm, N.-C. Bot, M. Zepper, D. Moreno, P. Vemuri, M. Binzer-Panchal, E. Greenberg, silviamorins, L. Pantano, R. Syme, G. Kelly, L. Sola, J. A. Fellows Yates, G. Lichtenstein, J. Espinosa-Carrasco, rfenouil, L. Zappia, C. Cheshire, E. Miller, marchoeppner, P. Zhou, *Nf-Core/Rnaseq: Nf-Core/Rnaseq v3.21.0 - Mercury Macaw* (Zenodo, 2025; http://dx.doi.org/10.5281/ZENODO.1400710).

67. F. Krueger, *TrimGalore: A Wrapper around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data* (Github; https://github.com/FelixKrueger/TrimGalore).

68. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

69. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

70. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

71. H. Patel, J. Espinosa-Carrasco, B. Langer, P. Ewels, N.-C. Bot, M. U. Garcia, R. Syme, A. Peltzer, A. Talbot, D. Behrens, G. Gabernet, M. Jin, M. Hörtenhuber, J. G. Rodriguez, K. Menden, Ö. An, *Nf-Core/Atacseq: [2.1.2] - 2022-08-07* (Zenodo, 2023; http://dx.doi.org/10.5281/ZENODO.2634132).

72. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

73. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

74. J. M. Gaspar, Improved peak-calling with MACS2, *bioRxiv* (2018). https://doi.org/10.1101/496521.

75. M. Karimzadeh, C. Ernst, A. Kundaje, M. M. Hoffman, Umap and Bismap: quantifying genome and methylome mappability. *Nucleic Acids Res.* **46**, e120 (2018).

76. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE blacklist: Identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).

77. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).

78. N. Bruse, S. J. van Heeringen, GimmeMotifs: an analysis framework for transcription factor motif analysis, *bioRxiv* (2018). https://doi.org/10.1101/474403.

79. .http://jaspar2020.uio.no/static/clustering/2020/vertebrates/CORE/interactive_trees/pairwise_comparisons.tab.

80. J. Herrero, M. Muffato, K. Beal, S. Fitzgerald, L. Gordon, M. Pignatelli, A. J. Vilella, S. M. J. Searle, R. Amode, S. Brent, W. Spooner, E. Kulesha, A. Yates, P. Flicek, Ensembl comparative genomics resources. *Database (Oxford)* **2016**, baw053 (2016).

81. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

82. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).

83. D. Earl, N. Nguyen, G. Hickey, R. S. Harris, S. Fitzgerald, K. Beal, I. Seledtsov, V. Molodtsov, B. J. Raney, H. Clawson, J. Kim, C. Kemena, J.-M. Chang, I. Erb, A. Poliakov, M. Hou, J. Herrero, W. J. Kent, V. Solovyev, A. E. Darling, J. Ma, C. Notredame, M. Brudno, I. Dubchak, D. Haussler, B. Paten, Alignathon: a competitive assessment of whole-genome alignment methods. *Genome Res.* **24**, 2077–2089 (2014).

84. I. T. Fiddes, J. Armstrong, M. Diekhans, S. Nachtweide, Z. N. Kronenberg, J. G. Underwood, D. Gordon, D. Earl, T. Keane, E. E. Eichler, D. Haussler, M. Stanke, B. Paten, Comparative Annotation Toolkit (CAT)-simultaneous clade and personal genome annotation. *Genome Res.* **28**, 1029–1038 (2018).

85. . https://github.com/bmbolstad/preprocessCore.

86. R. Wehrens, J. Kruisselbrink, Flexible self-organizing maps in kohonen 3.0. *J. Stat. Softw.* **87** (2018).

87. L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

88. J. Melville, The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction [R package uwot version 0.2.3], *Comprehensive R Archive Network (CRAN)* (2025). https://cran.r-project.org/web/packages/uwot/index.html.

89. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

90. Ggalluvial. http://corybrunson.github.io/ggalluvial/.

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

92. AnnotationForge, *Bioconductor*. https://bioconductor.org/packages/release/bioc/html/AnnotationForge.html.

93. D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, A. Kasprzyk, BioMart--biological queries made easy. *BMC Genomics* **10**, 22 (2009).

94. GO.Db, *Bioconductor*. https://www.bioconductor.org/packages/release/data/annotation/html/GO.db.html.

95. M. Murata, H. Nishiyori-Sueki, M. Kojima-Ishiyama, P. Carninci, Y. Hayashizaki, M. Itoh, Detecting expressed genes using CAGE. *Methods Mol. Biol.* **1164**, 67–85 (2014).

96. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

97. V. Haberle, A. R. R. Forrest, Y. Hayashizaki, P. Carninci, B. Lenhard, CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* **43**, e51 (2015).

98. PreprocessCore, *Bioconductor*. https://www.bioconductor.org/packages/release/bioc/html/preprocessCore.html.

99. H.-G. Drost, Philentropy: Information Theory and Distance Quantification with R. *J. Open Source Softw.* **3**, 765 (2018).

100. Authors and citation. https://rpkgs.datanovia.com/ggpubr/authors.html.

101. M. Naville, M. Ishibashi, M. Ferg, H. Bengani, S. Rinkwitz, M. Krecsmarik, T. A. Hawkins, S. W. Wilson, E. Manning, C. S. R. Chilamakuri, D. I. Wilson, A. Louis, F. Lucy Raymond, S. Rastegar, U. Strähle, B. Lenhard, L. Bally-Cuif, V. van Heyningen, D. R.

771 FitzPatrick, T. S. Becker, H. Roest Crollius, Long-range evolutionary constraints reveal cis-
772 regulatory interactions on the human X chromosome. *Nat. Commun.* **6**, 6904 (2015).

773 102. E. V. Davydov, D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, S. Batzoglou, Identifying
774 a high fraction of the human genome to be under selective constraint using GERP++. *PLoS*
775 *Comput. Biol.* **6**, e1001025 (2010).

776 103. Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density
777 Estimation [R package mclust version 6.1.1], *Comprehensive R Archive Network (CRAN)*
778 (2024). https://cran.r-project.org/web/packages/mclust/index.html.

779 104. P. Chomczynski, N. Sacchi, The single-step method of RNA isolation by acid guanidinium
780 thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat. Protoc.* **1**, 581–
781 585 (2006).

782 105. J.-L. Chew, Y.-H. Loh, W. Zhang, X. Chen, W.-L. Tam, L.-S. Yeap, P. Li, Y.-S. Ang, B.
783 Lim, P. Robson, H.-H. Ng, Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the
784 Oct4/Sox2 complex in embryonic stem cells. *Mol. Cell. Biol.* **25**, 6031–6046 (2005).

785 106. A. Rikin, T. Evans, The tbx/bHLH transcription factor mga regulates gata4 and
786 organogenesis. *Dev. Dyn.* **239**, 535–547 (2010).

787 107. A. G. Howard 4th, P. A. Baker, R. Ibarra-García-Padilla, J. A. Moore, L. J. Rivas, J. J.
788 Tallman, E. W. Singleton, J. L. Westheimer, J. A. Corteguera, R. A. Uribe, An atlas of
789 neural crest lineages along the posterior developing zebrafish at single-cell resolution. *Elife*
790 **10** (2021).

791 108. J. Kim, L. Lo, E. Dormand, D. J. Anderson, SOX10 maintains multipotency and inhibits
792 neuronal differentiation of neural crest stem cells. *Neuron* **38**, 17–31 (2003).

793 109. J. L. Sabel, C. d'Alençon, E. K. O'Brien, E. Van Otterloo, K. Lutz, T. N. Cuykendall, B. C.
794 Schutte, D. W. Houston, R. A. Cornell, Maternal Interferon Regulatory Factor 6 is required
795 for the differentiation of primary superficial epithelia in Danio and Xenopus embryos. *Dev.*
796 *Biol.* **325**, 249–262 (2009).

797 110. J. M. Santos-Pereira, L. Gallardo-Fuentes, A. Neto, R. D. Acemel, J. J. Tena, Pioneer and
798 repressive functions of p63 during zebrafish embryonic ectoderm specification. *Nat.*
799 *Commun.* **10**, 3049 (2019).

800 111. M. Carapuço, A. Nóvoa, N. Bobola, M. Mallo, Hox genes specify vertebral types in the
801 presomitic mesoderm. *Genes Dev.* **19**, 2116–2121 (2005).

802 112. J. D. Molkentin, A. B. Firulli, B. L. Black, J. F. Martin, C. M. Hustad, N. Copeland, N.
803 Jenkins, G. Lyons, E. N. Olson, MEF2B is a potent transactivator expressed in early
804 myogenic lineages. *Mol. Cell. Biol.* **16**, 3814–3824 (1996).

805 113. Y. Hinits, L. Pan, C. Walker, J. Dowd, C. B. Moens, S. M. Hughes, Zebrafish Mef2ca and
806 Mef2cb are essential for both first and second heart field cardiomyocyte differentiation. *Dev.*
807 *Biol.* **369**, 199–210 (2012).

808 114. T. Kiyama, C.-K. Chen, S. W. Wang, P. Pan, Z. Ju, J. Wang, S. Takada, W. H. Klein, C.-A.
809 Mao, Essential roles of mitochondrial biogenesis regulator Nrf1 in retinal development and
810 homeostasis. *Mol. Neurodegener.* **13**, 56 (2018).

811 115. H. H. Lau, N. H. J. Ng, L. S. W. Loo, J. B. Jasmen, A. K. K. Teo, The molecular functions
812 of hepatocyte nuclear factors - In and beyond the liver. *J. Hepatol.* **68**, 1033–1048 (2018).

813 116. E. Ktistaki, I. Talianidis, Modulation of hepatic gene expression by hepatocyte nuclear
814 factor 1. *Science* **277**, 109–112 (1997).

815 117. M. Kotulkar, D. R. Robarts, U. Apte, HNF4α in hepatocyte health and disease. *Semin. Liver*
816 *Dis.* **43**, 234–244 (2023).

817 118. J. Tong, C.-J. Han, J.-Z. Zhang, W.-Z. He, G.-J. Zhao, X. Cheng, L. Zhang, K.-Q. Deng, Y.
818 Liu, H.-F. Fan, S. Tian, J. Cai, Z. Huang, Z.-G. She, P. Zhang, H. Li, Hepatic interferon
819 regulatory factor 6 alleviates liver steatosis and metabolic disorder by transcriptionally
820 suppressing peroxisome proliferator-activated receptor γ in mice. *Hepatology* **69**, 2471–
821 2488 (2019).

822 119. M. J. Gonzalez-Rellan, E. Novoa, N. da Silva Lima, A. Rodriguez, C. Veyrat-Durebex, S.
823 Seoane, B. Porteiro, M. F. Fondevila, U. Fernandez, M. Varela-Rey, A. Senra, C. Iglesias,
824 A. Escudero, M. Fidalgo, D. Guallar, R. Perez-Fernandez, V. Prevot, M. Schwaninger, M.
825 López, C. Dieguez, R. Coppari, G. Frühbeck, R. Nogueiras, Hepatic p63 regulates glucose
826 metabolism by repressing SIRT1. *Gut* **72**, 472–483 (2023).

827 120. S. Tutukova, V. Tarabykin, L. R. Hernandez-Miranda, The role of Neurod genes in brain
828 development, function, and disease. *Front. Mol. Neurosci.* **14**, 662774 (2021).

829 121. S. Hu, J. Feng, M. Wang, R. Wufuer, K. Liu, Z. Zhang, Y. Zhang, Nrf1 is an indispensable
830 redox-determining factor for mitochondrial homeostasis by integrating multi-hierarchical
831 regulatory networks. *Redox Biol.* **57**, 102470 (2022).

832 122. K. Liu, W. Li, Y. Xiao, M. Lei, M. Zhang, J. Min, Molecular mechanism of specific DNA
833 sequence recognition by NRF1. *Nucleic Acids Res.* **52**, 953–966 (2024).

834 123. A. Coschiera, M. Yoshihara, G. Lauter, S. Ezer, M. Pucci, H. Li, A. Kavšek, C. G. Riedel,
835 J. Kere, P. Swoboda, Primary cilia promote the differentiation of human neurons through the
836 WNT signaling pathway. *BMC Biol.* **22**, 48 (2024).

837 124. C. Cai, P. Wan, H. Wang, X. Cai, J. Wang, Z. Chai, J. Wang, H. Wang, M. Zhang, N. Yang,
838 Z. Wu, J. Zhu, X. Yang, Y. Li, B. Yue, R. Dang, J. Zhong, Transcriptional and open
839 chromatin analysis of bovine skeletal muscle development by single-cell sequencing. *Cell*
840 *Prolif.* **56**, e13430 (2023).

841 125. Y. N. Gorodilov, Description of the early ontogeny of the Atlantic salmon, Salmo salar,
842 with a novel system of interval (state) identification. *Environ. Biol. Fishes* **47**, 109–127
843 (1996).

126. K. L. Hughes, E. T. Abshire, A. C. Goldstrohm, Regulatory roles of vertebrate Nocturnin: insights and remaining mysteries. *RNA Biol.* **15**, 1255–1267 (2018).

127. A. F. T. Moore, R. C. Gentry, E. Koculi, DbpA is a region-specific RNA helicase. *Biopolymers* **107**, e23001 (2017).

128. L. Bally-Cuif, C. Goutel, M. Wassef, W. Wurst, F. Rosa, Coregulation of anterior and posterior mesendodermal development by a hairy-related transcriptional repressor. *Genes Dev.* **14**, 1664–1677 (2000).

129. M. Martin, A. Akhmanova, Coming into focus: Mechanisms of microtubule minus-end organization. *Trends Cell Biol.* **28**, 574–588 (2018).

130. D. B. Dranow, K. Hu, A. M. Bird, S. T. Lawry, M. T. Adams, A. Sanchez, J. F. Amatruda, B. W. Draper, Bmp15 is an oocyte-produced signal required for maintenance of the adult female sexual phenotype in zebrafish. *PLoS Genet.* **12**, e1006323 (2016).

131. M. Escobar-Aguirre, Y. M. Elkouby, M. C. Mullins, Localization in oogenesis of maternal regulators of embryonic development. *Adv. Exp. Med. Biol.* **953**, 173–207 (2017).

132. A. Reynolds, H. Qiao, Y. Yang, J. K. Chen, N. Jackson, K. Biswas, J. K. Holloway, F. Baudat, B. de Massy, J. Wang, C. Höög, P. E. Cohen, N. Hunter, RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. *Nat. Genet.* **45**, 269–278 (2013).

133. R. B. Chalamalasetty, R. J. Garriock, W. C. Dunty Jr, M. W. Kennedy, P. Jailwala, H. Si, T. P. Yamaguchi, Mesogenin 1 is a master regulator of paraxial presomitic mesoderm differentiation. *Development* **141**, 4285–4297 (2014).

Acknowledgments: We thank the various members of the AQUA-FAANG consortium working on other teleost species for valuable discussions that shaped the direction of this study.

Funding: This study was funded by the AQUA-FAANG project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817923 (www.aqua-faang.eu). Further support was provided by the Biotechnology and Biological Sciences Research Council (BBSRC), including Institute Strategic Programme grants BBS/E/RL/230001B and BBS/E/D/10002070 and a Strategic Longer and Larger Award to DJM (BB/Z51746X/1). DB was supported by EU funds through the CroAGE project (code: NPOO.C3.2.R2-I1.06.0060 NextGenerationEU). BZ was supported by EU funds through the croESTRO project (code: NPOO.C3.2.R2-I1.06.0024 NextGrenerationEU).

Author contributions:

Conceptualization: DJM, SL, YG, SRS, JB, BL, MPK

Methodology: M-OB, DPM, GG, PSD, MKG, TP, LG, DB, AL(au), FG, BZ, EC-G, AP, AB, TTN, DT, HRC, YG, SRS, BL, JB, MPK, SL, DJM

Investigation: M-OB, DPM, GG, PSD, MKG, TP, LG, DB, AL(au), FG, BZ, EC-G, AP, AB, TTN, GAM, EP, DT, GRI, AL(ou), TRH, HRC, YG, SRS, BL, JB, MPK, DJM.

Visualization: M-OB, DPM, GG, PSD, MKG, TP, LG, DB, SRS

882 Funding acquisition: DJM, SL, YG, SRS, JB, BL, MPK

883 Project administration: DJM, SL, MKP, M-O, PWH

884 Supervision: DJM, SL, TRH, CB, PWH, DT, HRC, YG, JB, SRS, BL, MKP

885 Writing – original draft: DJM, M-OB, DPM, GG, PSD, MKG, LG, DB, SRS

886 Writing – review & editing: all authors reviewed and approved the final manuscript

887 **Competing interests**: The authors declare that they have no competing interests.

888 **Data and materials availability**: All sequencing datasets produced in this study are available
889 through the FAANG data portal (https://data.faang.org/projects/AQUA-FAANG). ENA
890 accession numbers for different studies are shared in Table S2, and as follows: PRJEB51855,
891 PRJEB51854 and PRJEB53399 (Atlantic salmon DevMap: RNA-seq, ATAC-seq and ChIP-seq,
892 respectively); PRJEB51857, PRJEB51856 and PRJEB55010 (rainbow trout DevMap: RNA-seq,
893 ATAC-seq and ChIP-seq, respectively); PRJEB47409, PRJEB47408 and PRJEB55063 (Atlantic
894 salmon BodyMap: RNA-seq, ATAC-seq and ChIP-seq, respectively); PRJEB57191,
895 PRJEB57190 and PRJEB57956 (rainbow trout BodyMap: RNA-seq, ATAC-seq and ChIP-seq,
896 respectively). All code required to reproduce the analyses and result reported is provided at:
897 https://github.com/AQUA-FAANG/AQUA-FAANG-salmonids.

898 **Figure Legends**

899 **Fig. 1. Regulatory annotation of salmonid genomes**. Data shown for salmon (equivalent data
900 for trout shown in Fig. S10). (**A**). Overview of sampling spanning embryogenesis (DevMap) and
901 adult tissues at two stages of sexual maturation (BodyMap). (**B**) Regulatory elements were
902 profiled using ATAC-seq and ChIP-seq and gene expression using RNA-seq. (C) 12-state
903 ChromHMM models for DevMap and BodyMap datasets, including signal intensity for
904 individual functional assays per chromatin state. (**D** & **E**) Coverage of the genome (**D**) and
905 number of robust open chromatin regions (**E**) per chromatin state across samples (DevMap: LB =
906 late blastulation, MG = mid-gastrulation; ES = early somitogenesis; MS = mid somitogenesis;
907 LS = late somitogenesis; BodyMap: tissue samples separated into immature (lowercase 'f' or
908 'm') and mature (uppercase 'F' and 'M') for females and males, respectively. (**F**) Number of
909 robust open chromatin regions the salmon genome, including those comprising active promoters
910 and enhancers according to ChromHMM annotation. (**G**) Integration of functional annotation
911 data for a pair of *mex3b* ohnologs (ENSSSAG0000008892 and ENSSSAG00000085168)
912 encoding a conserved RNA-binding protein. Predicted chromatin states (colored as panel **C**) and
913 corresponding gene expression are visualized across samples. This example illustrates
914 conservation of regulatory elements between ohnologs that correlate with conserved gene
915 expression. Equivalent data for rainbow trout *mex3b* ohnologs provided in Fig. S17.

916 **Fig. 2. Functional activity of salmonid genomes during embryogenesis**. (**A**) Transcriptome
917 and (**B**) chromatin openness dynamics during embryogenesis. Respective data shown are SOM
918 clusters for mean normalised transcripts per million values (TPM, RNA-Seq) and mean
919 normalised read intensity values in robust open chromatin regions (ATAC-seq). Full SOM
920 clusters shown in Figs S18-21. For panel A, light red bands illustrate the observed-to-expected
921 ratio of orthologous gene expression between the two species across all SOM cluster
922 combinations. For panel **B**, brown bands between clusters show the number of syntenic open
923 chromatin regions based on whole genome alignment. On UMAP plots shown under each
924 heatmap, dots represent mean normalised TPM values of individual genes (**A**) or mean
925 normalised ATAC-seq read intensity values for individual robust open chromatin regions (**B**).
926 Grey dots on each UMAP are genes or chromatin regions captured within constitutive SOM
927 clusters (Figs S18-S21), respectively. Schematic at top highlights embryonic stages with colours
928 matched to SOM clusters: LC = late cleavage; EB, MB, LB = early-, mid- and late blastulation;
929 EG, MG, LG= early, mid- and late gastrulation; ES, EMS, MS, MLS, LS = early, early-mid,
930 mid-, mid-to-late and late somitogenesis; EP, LP = early and late pharyngula.

931 **Fig. 3. Gene regulatory evolution following the salmonid 4R autotetraploidization**. (**A**-**B**)
932 Mean normalized Jensen–Shannon divergence (JSD) ± 1 SD (100 bootstrap samples) between
933 salmon ohnolog pairs across 14 stages of embryonic development (DevMap: A; grey area
934 between dashed lines represents the MZT period) or adult tissues of immature and mature fish
935 (triangles and squares, respectively) (BodyMap: B). Lines are colored by Early (red) and Late
936 (blue) rediploidization categories. (**C**-**D**) Mean JSD ± 1 SD for ortholog pairs from salmon and
937 trout for two groups representing ohnologs (retained as pairs in both species) or singletons (one
938 gene lost in same ohnolog pair in the salmon-trout ancestor) for the same DevMap (**C**) and
939 BodyMap (**D**) samples. JSD was calculated from quantile-normalised TPM values. Higher
940 normalized mean JSD indicates greater expression divergence between gene pairs. DevMap
941 stage abbreviations are as per the Fig. 1 legend. BodyMap tissue abbreviations: Br = brain; Te =

942 testis; Gi= gill; In = distal intestine; HK = head kidney; O =ovary; Mu = skeletal muscle; Li =
943 liver.

944 **Fig. 4. Functional retention of regulatory elements following the salmonid 4R**
945 **autotetraploidization**. (**A**) Schematic of Shared, Alignable only, and Exclusive regulatory
946 elements, categorized by whole genome alignment. (**B**) Circos plot of 29 salmon chromosomes.
947 Ribbons show collinear regions retained from the salmonid 4R WGD, with blue and red colours
948 indicating Early and Late rediploidization regions (Tables S9-10), where duplicated regions
949 share a respective 88.6% (SD: 1.51%) and 95.1% (SD: 2.67%) nucleotide sequence identity. The
950 outermost panels show percentage nucleotide sequence identity (as per the provided key)
951 between collinear blocks at 1 Mb intervals. For each alignment (Shared, Alignable only,
952 Exclusive) and Rediploidization (early, late) category, we present: The proportion of active
953 promoters (**C**) and enhancers (**D**) in each sample type.

954 **Fig. 5. Enhancer evolution is coupled to ohnolog regulatory divergence**. (**A**) Our definition of
955 *cis*-regulatory domains (*46*) of genes includes an 'exclusive' region - 5 to + 1 kb surrounding the
956 TSS, and an 'overlapping' region, extending until a neighboring exclusive region or maximum
957 distance of 1 Mb from the TSS. Active enhancers were compared within these domains for
958 ohnolog pairs from the Shared, Alignable only and Exclusive categories. (**B**) Ohnolog pairs were
959 split into "conserved" (gray) or "diverged" (red) regulation sub-classes (see Fig. S37), which
960 were compared in terms of: (**C**) proportions of Exclusive and Shared enhancers and (**D**)
961 Enhancer usage distance, representing total enhancer activity profiles across sample types
962 (expressed as Euclidian distance). (**E**) Density maps of ohnolog pair enhancer-usage Euclidian
963 distance vs. expression Euclidian distance, split by embryogenesis stage or adult tissue type. (**F**)
964 Percentage of TE basepairs within Exclusive or Shared enhancers for ohnologs from both
965 regulation sub-classes, shown for all TEs, then specifically for Exclusive enhancers by TE order
966 (DNA, LINE, LTR, SINE), and finally the DNA transposon TcMar. (**G**) TE content in Exclusive
967 enhancers across embryogenesis stage or adult tissue types. Symbols: * = $p < 0.05$, ** = $p <$
968 0.01, *** = $p < 0.001$).

969 **Fig. 6. Evolution of enhancer-CNEs following the salmonid 4R autotetraploidization**. (A)
970 The tree on the left illustrates the different phylogenetic ages of enhancer-CNEs: 'Et'
971 (Euteleostomi), 'Ac' (Actinopteri), 'Ne' (Neopterygii), 'Te' (Teleostei), 'Cl' (Clupeocephala),
972 'Eu' (Euteleostomorpha), and 'Pr' (Protacanthopterygii). Also shown is the number of Atlantic
973 salmon enhancer-CNEs belonging to the Shared (S), Alignable only (A), and Exclusive (E)
974 categories, along with their proportions per phylogenetic age category, within each alignment
975 category. On the provided barplots, different lowercase letters indicate significant differences
976 between alignment categories within each provided comparison (Fisher's Exact Test). (B)
977 Normalised proportions of Atlantic salmon enhancer-CNEs for each alignment category (S, A,
978 E), stratified by phylogenetic age and different sample-types for Early (left) and Late (right)
979 rediploidization regions. Sample abbreviations are as per the Fig. 1 and 2 legends, with
980 maturation stage split by 'I' (immature) and 'M' (mature) for tissue samples.

Figures

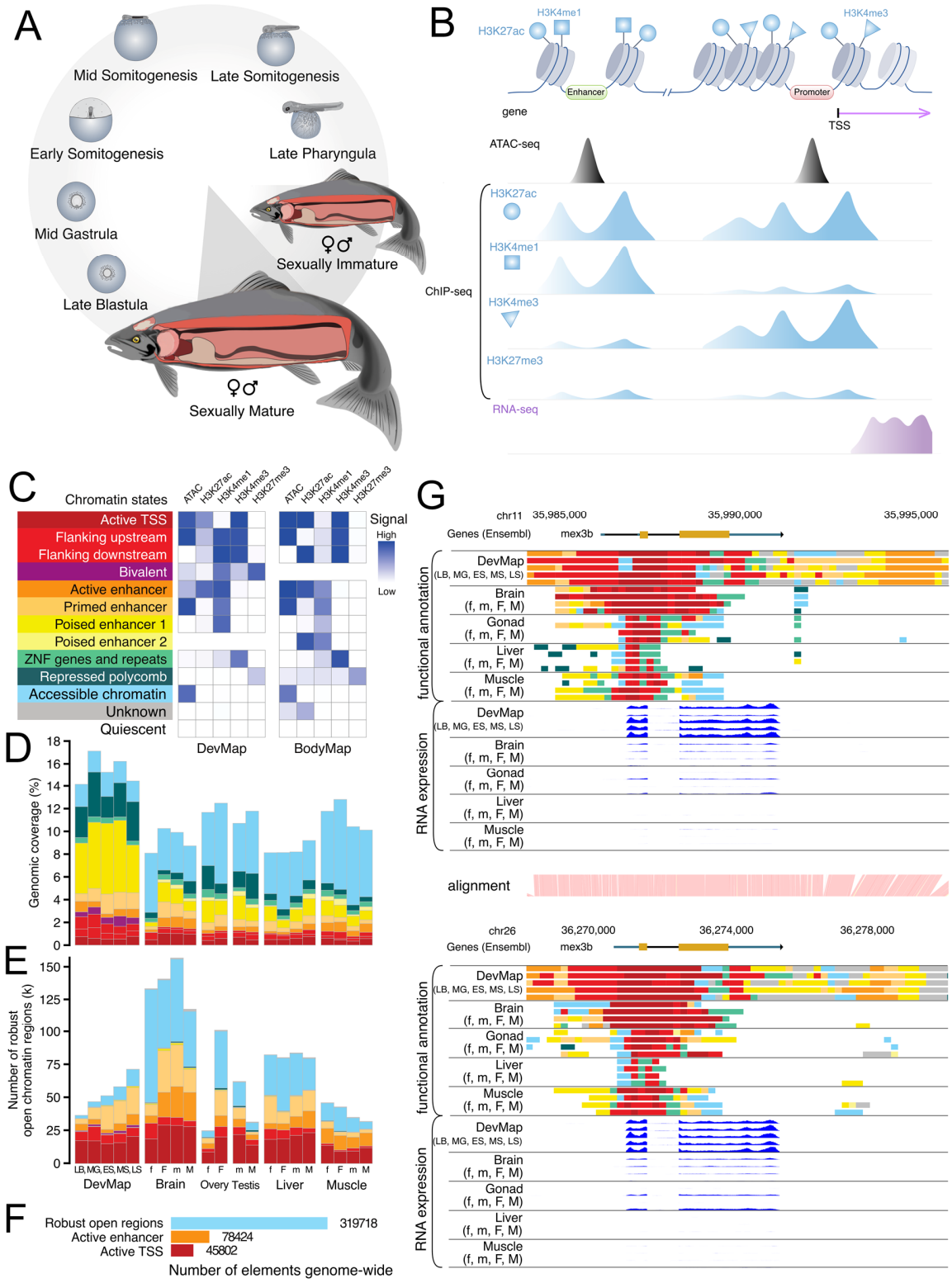

Fig. 1

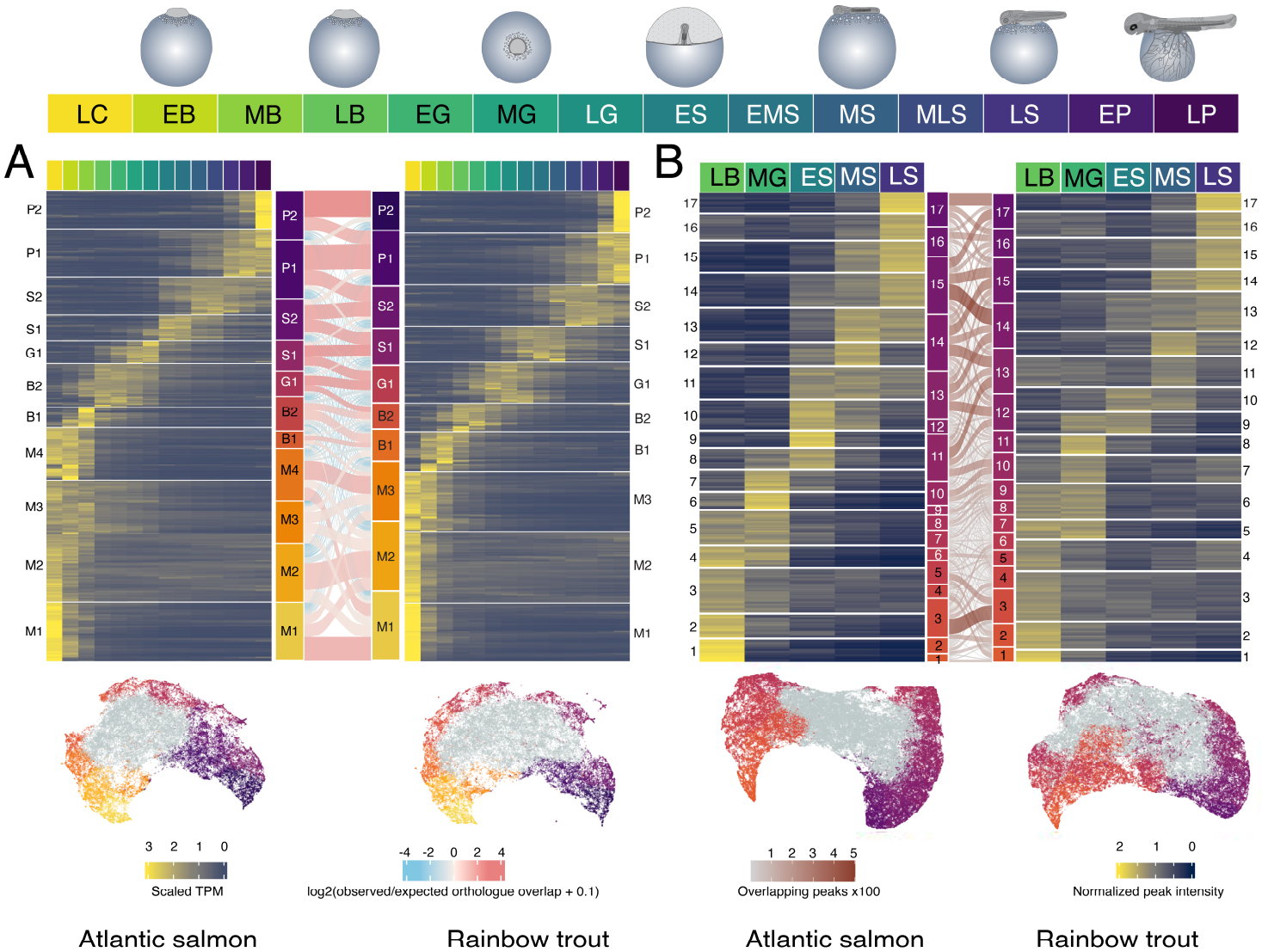
LC
EB
MB
LB
EG
MG
LG
ES
EMS
MS
MLS
LS
EP
LP
A
B
P2
P1
S2
S1
G1
B2
B1
M4
M3
M2
M1
LB MG ES MS LS
Scaled TPM
log2(observed/expected orthologue overlap + 0.1)
Overlapping peaks x100
Normalized peak intensity
Atlantic salmon
Rainbow trout
Atlantic salmon
Rainbow trout

Fig. 2

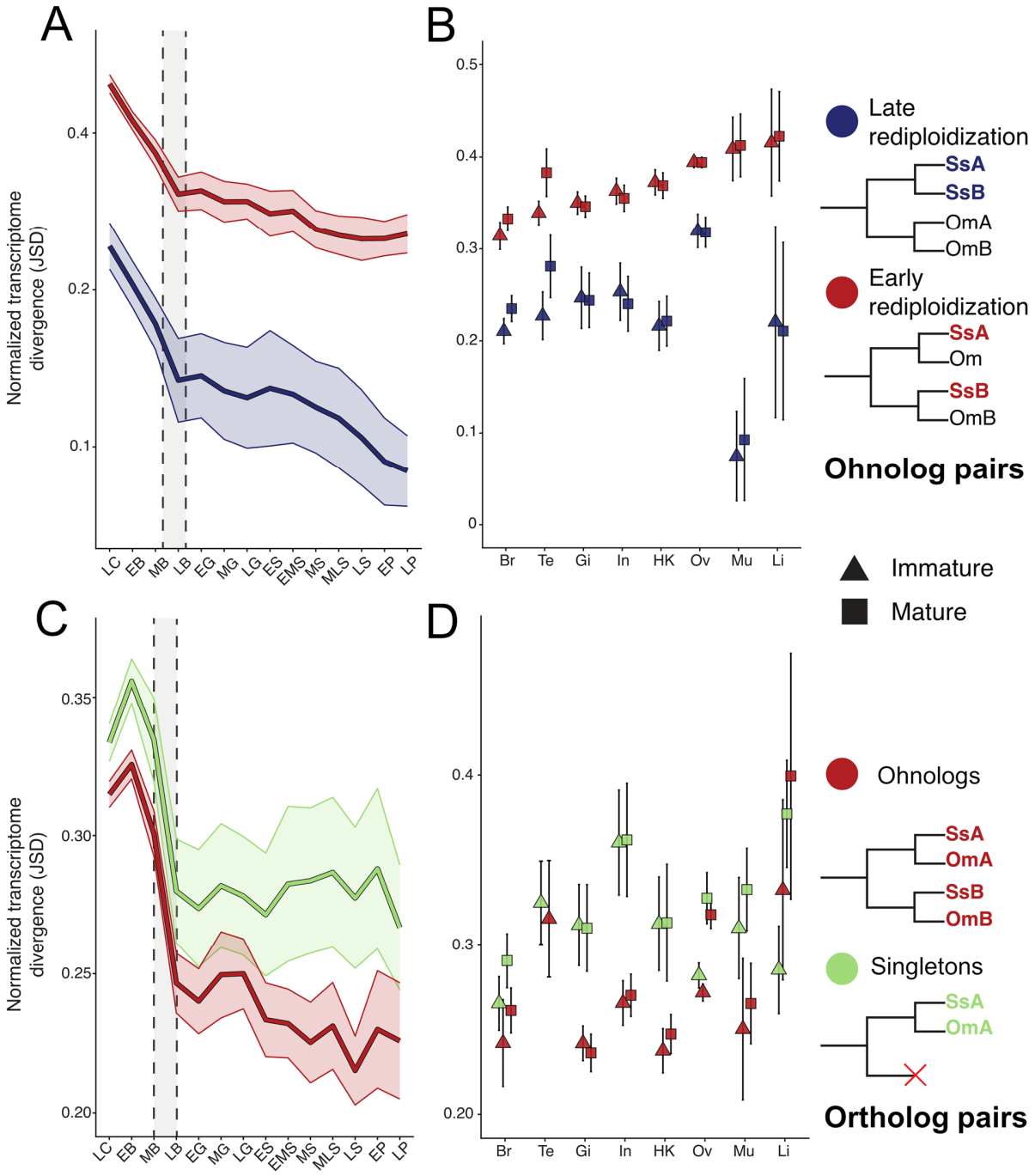
A
B
C
D
Normalized transcriptome divergence (JSD)
LC EB MB LB EG MG LG ES EMS MS MLS LS EP LP
Br Te Gi In HK Ov Mu Li
Late rediploidization
SsA
SsB
OmA
OmB
Early rediploidization
SsA
Om
SsB
OmB
Ohnolog pairs
Immature
Mature
Ohnologs
SsA
OmA
SsB
OmB
Singletons
SsA
OmA
Ortholog pairs

Fig. 3

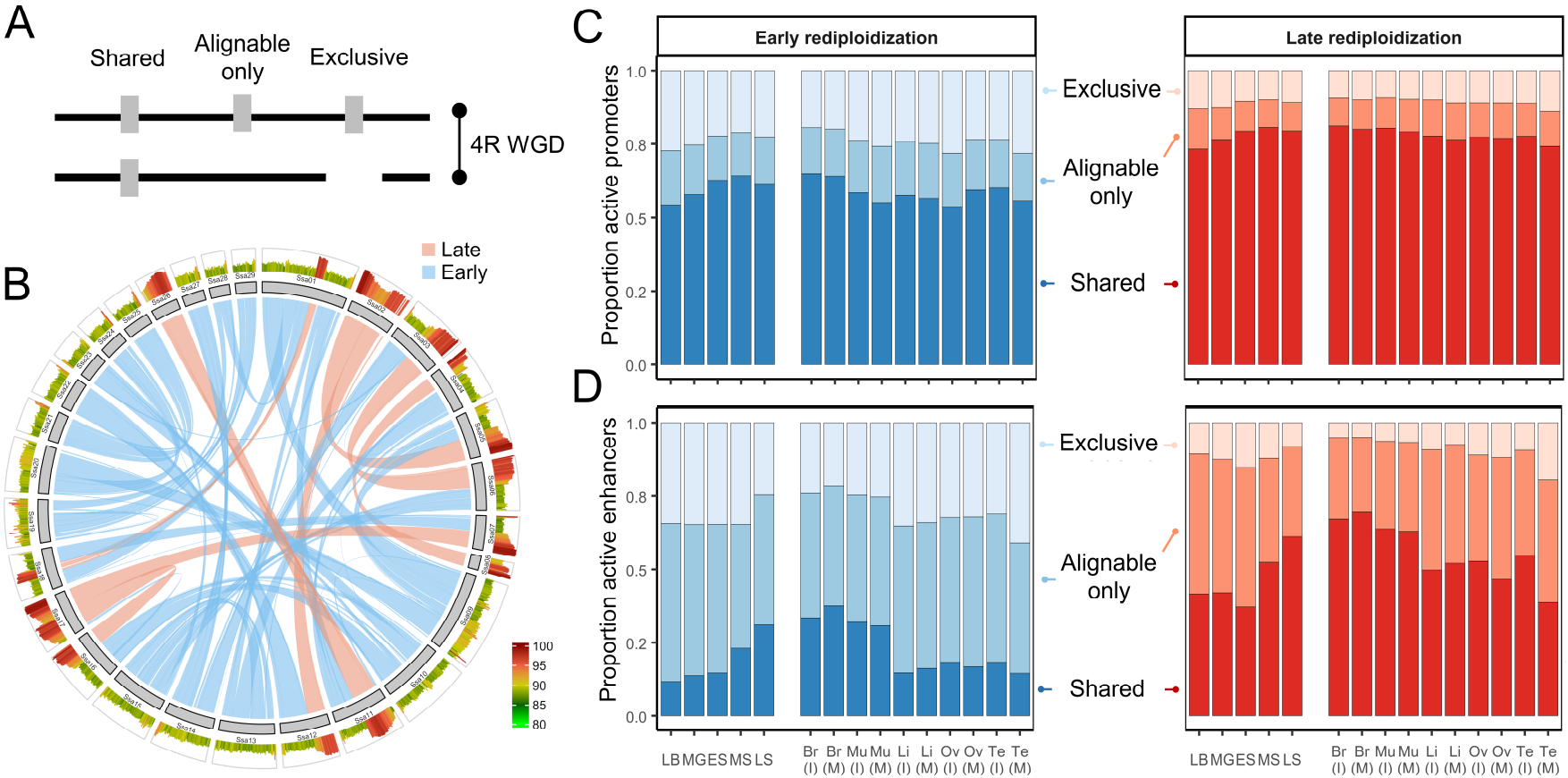
A
Shared
Alignable only
Exclusive
4R WGD
B
Late
Early
100
95
90
85
80
C
Early rediploidization
Late rediploidization
Proportion active promoters
1.0
0.8
0.5
0.2
0.0
Exclusive
Alignable only
Shared
D
Proportion active enhancers
LB MG ES MS LS
Br (I) Br (M) Mu (I) Mu (M) Li (I) Li (M) Ov (I) Ov (M) Te (I) Te (M)

Fig. 4

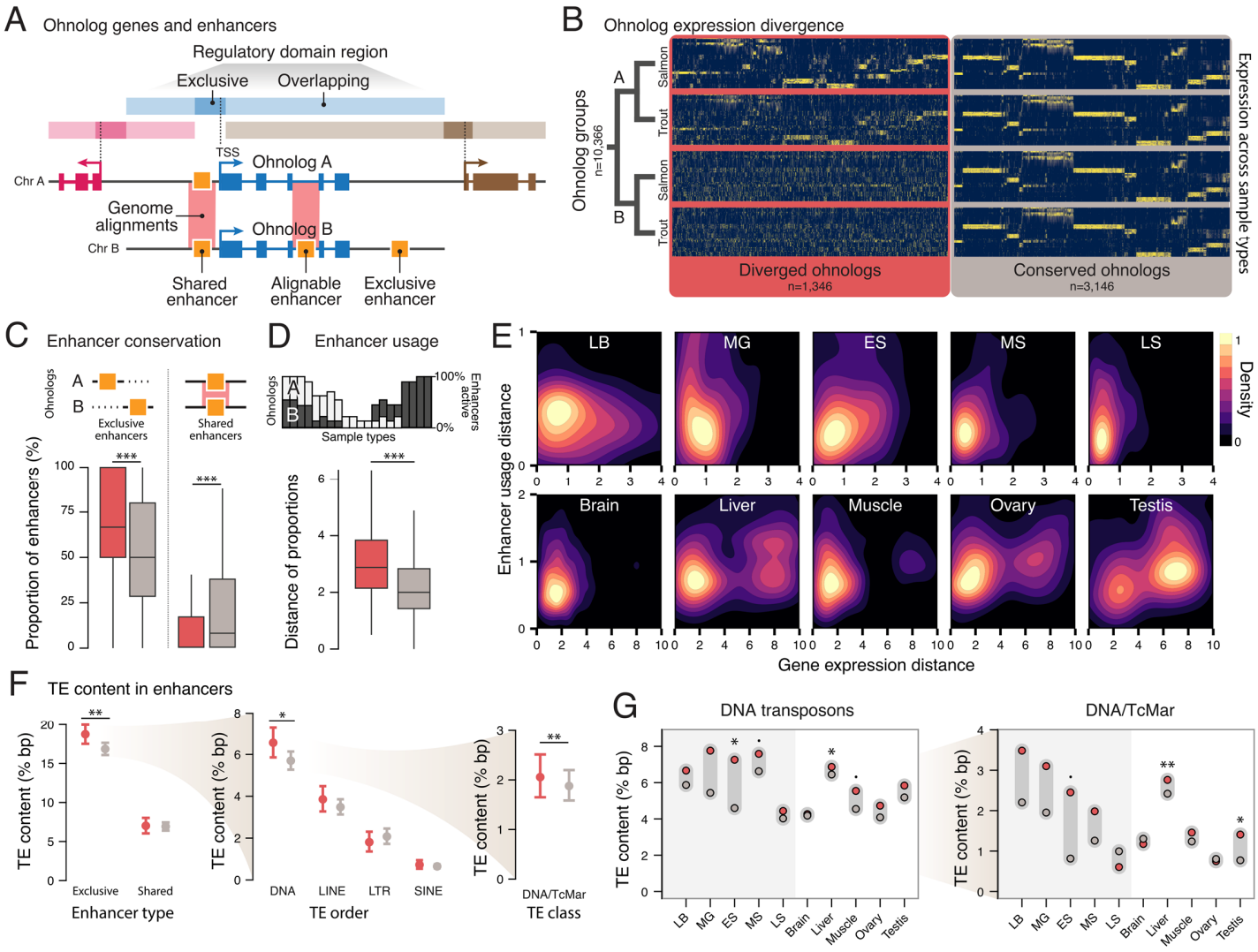
A
Ohnolog genes and enhancers
Regulatory domain region
Exclusive
Overlapping
TSS
Ohnolog A
Chr A
Genome alignments
Ohnolog B
Chr B
Shared enhancer
Alignable enhancer
Exclusive enhancer
B
Ohnolog expression divergence
Ohnolog groups
n=10,366
A
B
Salmon
Trout
Salmon
Trout
Expression across sample types
Diverged ohnologs
n=1,346
Conserved ohnologs
n=3,146
C
Enhancer conservation
Ohnologs
A
B
Exclusive enhancers
Shared enhancers
Proportion of enhancers (%)
***
***
D
Enhancer usage
Ohnologs
A
B
100%
0%
Enhancers active
Sample types
Distance of proportions
***
E
LB
MG
ES
MS
LS
Brain
Liver
Muscle
Ovary
Testis
Density
Enhancer usage distance
Gene expression distance
F
TE content in enhancers
TE content (% bp)
**
Exclusive
Shared
Enhancer type
*
DNA
LINE
LTR
SINE
TE order
**
DNA/TcMar
TE class
G
DNA transposons
DNA/TcMar
TE content (% bp)
LB
MG
ES
MS
LS
Brain
Liver
Muscle
Ovary
Testis

Fig. 5

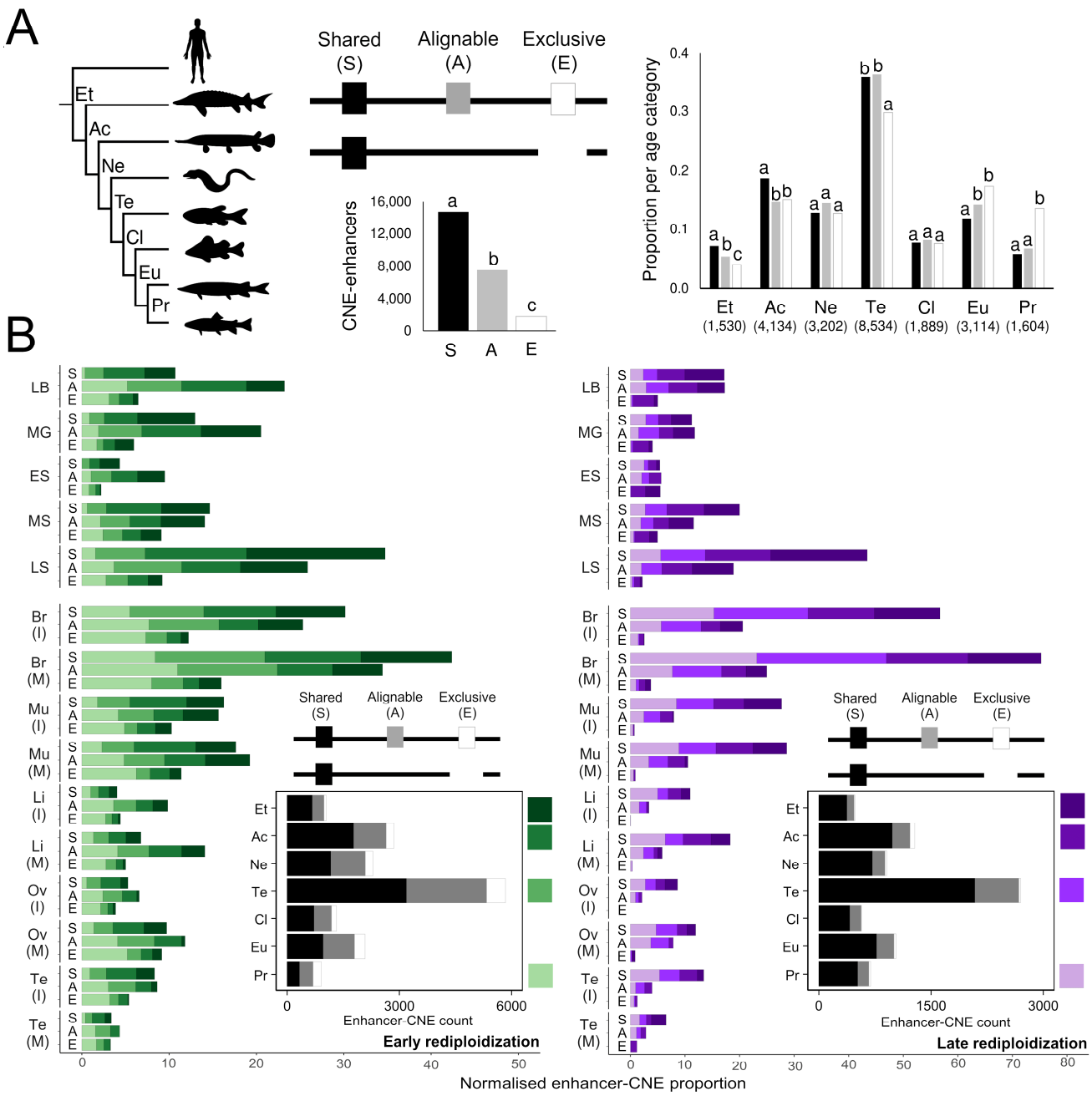
A
Et
Ac
Ne
Te
Cl
Eu
Pr
Shared (S)
Alignable (A)
Exclusive (E)
CNE-enhancers
16,000
12,000
8,000
4,000
0
S A E
Proportion per age category
0.4
0.3
0.2
0.1
0.0
Et (1,530)
Ac (4,134)
Ne (3,202)
Te (8,534)
Cl (1,889)
Eu (3,114)
Pr (1,604)
B
LB
MG
ES
MS
LS
Br (I)
Br (M)
Mu (I)
Mu (M)
Li (I)
Li (M)
Ov (I)
Ov (M)
Te (I)
Te (M)
Enhancer-CNE count
Early rediploidization
Late rediploidization
Normalised enhancer-CNE proportion

Fig. 6