


RESEARCH

Open Access



# Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication

Gareth B. Gillard<sup>1†</sup>, Lars Grønvold<sup>2†</sup>, Line L. Røsæg<sup>2</sup>, Matilde Mengkrog Holen<sup>2</sup>, Øystein Monsen<sup>2</sup>, Ben F. Koop<sup>3</sup>, Eric B. Rondeau<sup>3</sup>, Manu Kumar Gundappa<sup>4</sup>, John Mendoza<sup>5</sup>, Daniel J. Macqueen<sup>4</sup>, Rori V. Rohlf<sup>6</sup>, Simen R. Sandve<sup>2\*</sup> and Torgeir R. Hvidsten<sup>1\*</sup> 

\* Correspondence: [simen.sandve@nmbu.no](mailto:simen.sandve@nmbu.no); [torgeir.r.hvidsten@nmbu.no](mailto:torgeir.r.hvidsten@nmbu.no)

<sup>†</sup>Gareth B. Gillard and Lars Grønvold contributed equally to this work.

<sup>2</sup>Center for Integrative Genetics, Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway

<sup>1</sup>Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway  
Full list of author information is available at the end of the article

## Abstract

**Background:** Whole genome duplication (WGD) events have played a major role in eukaryotic genome evolution, but the consequence of these extreme events in adaptive genome evolution is still not well understood. To address this knowledge gap, we used a comparative phylogenetic model and transcriptomic data from seven species to infer selection on gene expression in duplicated genes (ohnologs) following the salmonid WGD 80–100 million years ago.

**Results:** We find rare cases of tissue-specific expression evolution but pervasive expression evolution affecting many tissues, reflecting strong selection on maintenance of genome stability following genome doubling. Ohnolog expression levels have evolved mostly asymmetrically, by diverting one ohnolog copy down a path towards lower expression and possible pseudogenization. Loss of expression in one ohnolog is significantly associated with transposable element insertions in promoters and likely driven by selection on gene dosage including selection on stoichiometric balance. We also find symmetric expression shifts, and these are associated with genes under strong evolutionary constraints such as ribosome subunit genes. This possibly reflects selection operating to achieve a gene dose reduction while avoiding accumulation of “toxic mutations”. Mechanistically, ohnolog regulatory divergence is dictated by the number of bound transcription factors in promoters, with transposable elements being one likely source of novel binding sites driving tissue-specific gains in expression.

**Conclusions:** Our results imply pervasive adaptive expression evolution following WGD to overcome the immediate challenges posed by genome doubling and to exploit the long-term genetic opportunities for novel phenotype evolution.



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Whole genome duplication (WGD) events have played a major role in eukaryotic evolution by increasing genomic complexity and functional redundancy [1]. This can allow gene duplicates (referred to as ohnologs) to escape selective constraints and thereby accumulate previously forbidden mutations that may become adaptive [2]. In agreement with this idea, WGD has been associated with the evolution of adaptive traits in yeast [3], plants [4, 5], and vertebrates [6–8]. At the same time, it is also evident that most polyploids go extinct shortly after formation [9] and that becoming a successful new polyploid likely requires new adaptations to overcome fitness costs stemming from having a doubled genome [10, 11]. Yet, the importance of selection in shaping polyploid genome evolution in the aftermath of WGDs is still not well understood [1, 12].

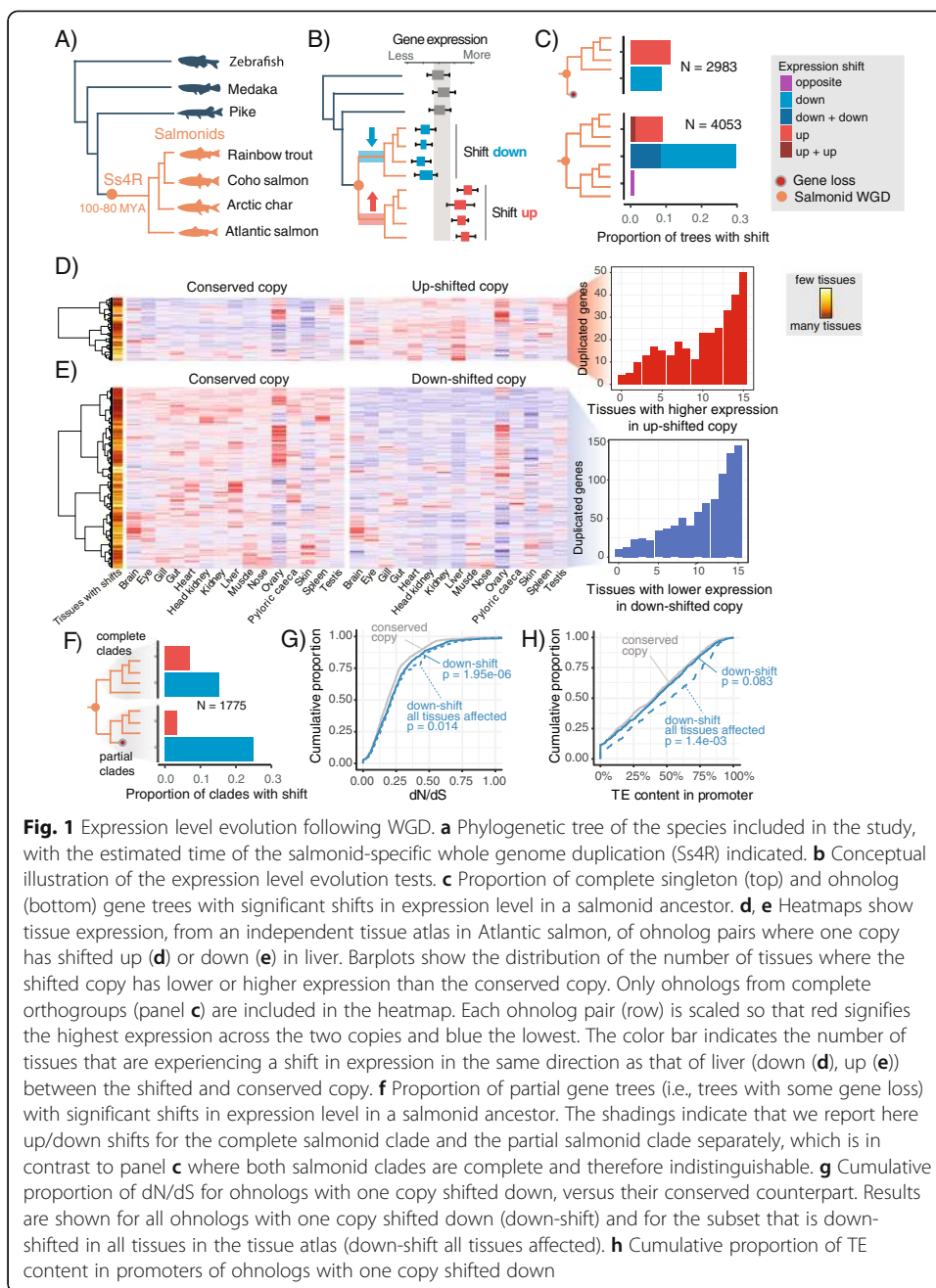
Gene expression levels are relatively easy to measure and compare, and represent a major source of complex trait variation [13] and novel adaptive phenotypes [14, 15]. Hence, there has been substantial interest in understanding consequences of WGDs on gene regulatory evolution. Comparative transcriptomics has both revealed immediate plastic responses to adjust gene dosages [16], as well as widespread regulatory divergence at evolutionary timescales (e.g., [17–20]). Ohnolog regulatory evolution is also mostly asymmetric, with one copy retaining an ancestral-like regulation, and the other copy losing and/or gaining expression in one or more tissue [12]. Although this observation can be reconciled with adaptive evolution of gene regulatory phenotypes following WGD, methodological limitations have made it difficult to distinguish between the outcomes of selection and neutral drift [12, 21].

Here we take a novel approach to improve our understanding of how selection shapes novel gene regulatory phenotypes following WGD. We first developed a flexible and user friendly version of a phylogenetic Ornstein-Uhlenbeck (OU) model of gene expression evolution [22, 23] in R (<https://gitlab.com/sandve-lab/evemodel>). The crux of this model is that it allows us to evaluate if changes in expression evolution deviate from the null hypothesis of stabilizing selection, and thereby identify putative adaptive shifts in expression regulation. We then used this model to analyze the liver transcriptome of four salmonids and three non-salmonid fish species to assess the impact of the 80–100-million-year-old salmonid-specific WGD (Ss4R) [24, 25]. We find that this WGD led to a burst of gene expression evolution, leading to rare tissue-specific gains in expression and pervasive tissue non-specific dosage selection, reflecting both adaptive possibilities afforded by genome doubling and immediate challenges that must be overcome to succeed as a polyploid lineage.

## Results

### Adaptive shifts in expression levels following WGD

To study expression level evolution following WGD, we generated RNA-seq datasets from livers (four biological replicates) of four salmonids and three non-salmonid outgroup species (Fig. 1a). We then computed gene trees to identify retained ohnologs from the salmonid WGD. In total, we included 10,154 gene trees in our analyses (Additional file 1: Figure S1), of which 65% (6689 trees) contained ohnologs derived from the salmonid WGD. For each gene tree, we then applied a phylogenetic Ornstein-Uhlenbeck (OU) process model to test for adaptive shifts in expression evolution



(referred to simply as ‘shifts’) in the ancestor of the salmonids included in this study (Fig. 1b, Additional file 1: Figures S2, S3 and S4).

Two major observations arise from this analysis. First, it is evident that the rate of adaptive gene expression evolution is increased for salmonid ohnologs. Forty percent of trees (1649) with retained ohnologs display evolution of novel expression levels in at least one ohnolog compared to only 20 % of trees with a single copy gene (Fig. 1c). Secondly, there is a clear difference in the nature of the expression evolution between ohnologs and singleton genes. Ohnologs are strongly biased towards evolving decreased expression levels following WGD (Fig. 1c), with 75% (1234/1649) of the ohnolog pairs displaying a shift down in either one or both copies. Conversely, singletons show a

small bias towards evolving increased expression (Fig. 1c). This difference could not be explained by differences in statistical power related to systematic differences in gene expression levels between singletons and ohnologs (Additional file 1: Figure S5).

To test if the identified expression level shifts following WGD were tissue-specific, we analyzed RNA-seq data from 15 Atlantic salmon tissues (Additional file 1: Figure S6A). We find that most cases of expression evolution are not liver-specific (Fig. 1d, e), and that this is true both for genes evolving increased and decreased expression following WGD. When one ohnolog copy had evolved a shift in liver expression level, this copy also displayed similar trends in the majority of the other 14 tissues compared to its conserved ohnolog partner (shift down 77% (682/885), shift up 70% (221/317)). Hence, evolution of liver-specific changes in ohnolog expression following WGD is rare, irrespective of the directionality of change.

Upon reaching a new optimal ohnolog gene dosage, the expectation is that the copy with the highest expression level contributes the most to the proteome and cell function, which will result in reduced purifying selection pressure on the more lowly expressed copy [26]. Several lines of evidence support this expectation. Firstly, species-specific gene loss events (expected for genes evolving under relaxed selection) are associated with increased probability of evolving lower liver expression in one copy (Fig. 1f) and with increased probability of the down-shifted copy to have reduced expression levels across all the other 14 tissues (Fisher's exact test,  $p = 3.1e-07$ , Additional file 1: Figure S6B). Secondly, we find that the down-shifted copy shows increased signatures of relaxed purifying selection on coding sequences in the form of elevated dN/dS rates (Fig. 1g,  $p = 2.1e-6$ ,  $N = 732$ , one-sided paired Wilcoxon test, Additional file 1: Figure S7). Lastly, we also observe that down-shifted ohnolog copies have a significantly higher load of potentially destructive transposable element (TE) insertions in promoters compared to the conserved partner (Fig. 1h, one-sided paired Wilcoxon test,  $p = 6.5e-4$ , Additional file 1: Figure S8). Importantly, the effect size of increased dN/dS and TE-load were larger when only considering ohnologs with signatures of down-shift across all tissues (Fig. 1g, h).

Pervasive differences in purifying selection pressure within individual ohnolog pairs raise the question of whether these ohnologs might belong to duplicated genome blocks experiencing large-scale differences in selective constraints. This could lead to uneven ohnolog loss rates, a process that is referred to as biased fractionation [27]. In line with previous studies on teleosts [28, 29], we found significant biases in local gene loss, albeit only in 9 of 47 syntenic duplicate blocks. However, we did not find equivalent large-scale biases in expression loss (Additional file 1: Figure S9), thus rendering regional differences in selection constraints an unlikely explanation for the large number of ohnologs experiencing loss of expression in one copy.

In conclusion, we find widespread signatures of adaptive regulatory evolution in retained ohnologs following WGD; however, most adaptive events were associated with ohnolog gene dose reduction across many tissues. Thus, ohnolog copies that evolve lower expression levels compared to their partner continue to evolve under relaxed purifying selection pressure, following a likely path towards pseudogenization.

### Strong selection on housekeeping gene dose after WGD

To test if selection on gene regulation following WGD was linked to particular cellular functions or pathways, we performed KEGG enrichment analyses for two ohnolog gene sets that had evolved either increased (up) or decreased (down) expression levels. Genes with increased expression level were enriched (Fisher's exact test,  $p < 0.05$ ) in three pathways: "fatty acid elongation," "fatty acid metabolism," and the "cell cycle" (Additional file 1: Table S1). Detailed analysis identified 29 up-shifted genes encoding proteins with essential cell division functions. These genes were highly enriched in protein-protein interactions conserved in both unicellular and multicellular eukaryotes (Additional file 1: Table S2, Additional file 1: Figure S10) and suggest compensatory regulatory adaptation to maintain a functional cell division and ensure genome stability.

Down-shifted genes had comparatively stronger functional signatures (Additional file 1: Table S1) with nine enriched pathways (Fisher's exact test,  $p < 0.05$ ). The three pathways with the strongest enrichment were "oxidative phosphorylation" ( $p = 0.003$ ) involved in mitochondrial-associated cellular energy production, "ribosome biogenesis in eukaryotes" ( $p = 0.008$ ) which consists of genes involved in assembly of the ribosome, and "ribosome" ( $p = 5.6e-9$ ) which consists of ribosomal subunit genes (Supplementary figures 11, 12 and 13). These results support strong selection on gene dosage for many housekeeping functions following WGD, which aligns well with our observation (Fig. 1d, e) that most expression level shifts occurred across most tissues.

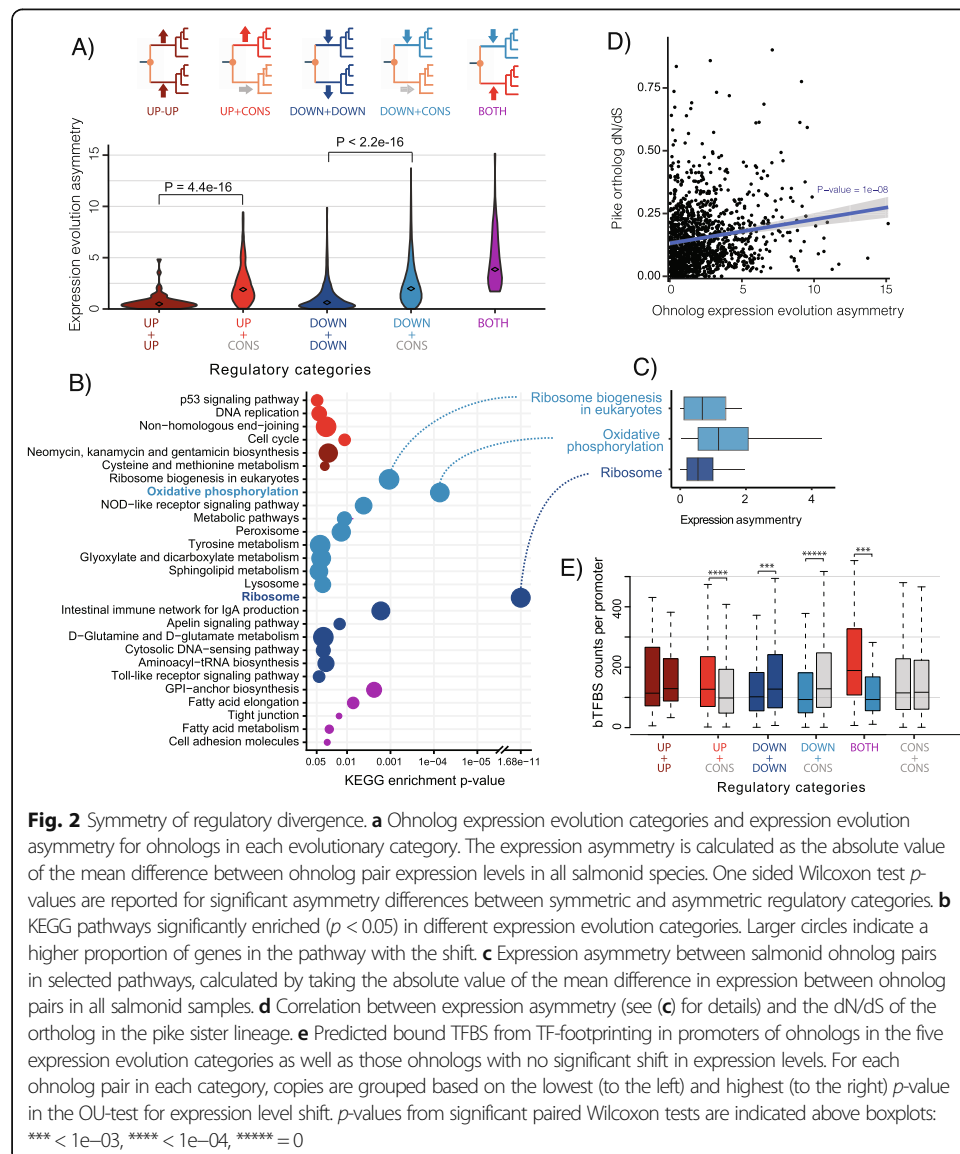
The gene balance hypothesis predicts that selection operates to maintain stoichiometry of interacting gene products [30], and this is believed to result in long-term retention of ohnologs. Using the human orthologs of salmonid genes, we queried the CORUM database of protein complexes and found that the proportion of ohnologs in protein complexes was slightly higher (22%) than the proportion of singletons (18%) (Fisher's exact test,  $p = 1.04e-5$ , Additional file 1: Figure S14A). We also found that complexes tended to contain only singletons ( $p = 0.03$ ) or only duplicates ( $p = 1E-3$ , Additional file 1: Figure S14B) more often than expected by chance. It is also plausible that stoichiometric imbalances could be rescued through evolution of novel gene dosage. Under this model, we predict that singletons in protein complexes that contain ohnologs should be enriched for shifts up in expression, while shifts down are predicted for ohnologs in complexes with singletons. These predictions are not well supported for singletons (Fisher's exact test,  $p = 0.07$ ) nor ohnologs (Fisher's exact test,  $p > 0.48$ ) (Additional file 1: Table S3).

Taken together, we find strong evidence for dosage selection following WGD on genes involved in basic cellular maintenance and cell division. In addition, we find evidence for selection to retain stoichiometric balance both at the sequence and expression level.

### Mechanism driving ohnolog regulatory divergence is associated with functional constraints

Our analysis allows us to assign ohnolog pairs to different regulatory categories (Fig. 2a) that potentially represent distinct evolutionary routes to new gene dosage optimums after WGD. Indeed our results show that ohnolog pairs with expression evolution shifts in the same direction evolve more symmetrically (down+down and up+up) while

ohnologs where expression shifts occur in only one copy or in opposite directions display stronger asymmetric divergence (e.g., up/down+conserved) (Fig. 2a). To explore the links between these modes of regulatory divergence and gene function, we performed KEGG enrichment on each expression evolution category. Twenty-seven pathways were found enriched across these categories (Fig. 2b, Additional file 1: Table S4), which is more than twice as many as when grouping ohnologs into up- or down-shifted genes (Additional file 1: Table S1). This supports that different pathways are biased towards either symmetric or asymmetric regulatory evolution. The three most enriched pathways were the same as when testing up- and down-shifted genes only, but our stratification on regulatory categories of ohnologs reveals that ribosomal subunit ohnologs (“Ribosome”) evolved lower gene dosage through highly symmetrical down-shifts, while “oxidative phosphorylation” and “ribosome biogenesis in eukaryotes” are biased towards asymmetric divergence (Fig. 2c).



**Fig. 2** Symmetry of regulatory divergence. **a** Ohnolog expression evolution categories and expression evolution asymmetry for ohnologs in each evolutionary category. The expression asymmetry is calculated as the absolute value of the mean difference between ohnolog pair expression levels in all salmonid species. One sided Wilcoxon test  $p$ -values are reported for significant asymmetry differences between symmetric and asymmetric regulatory categories. **b** KEGG pathways significantly enriched ( $p < 0.05$ ) in different expression evolution categories. Larger circles indicate a higher proportion of genes in the pathway with the shift. **c** Expression asymmetry between salmonid ohnolog pairs in selected pathways, calculated by taking the absolute value of the mean difference in expression between ohnolog pairs in all salmonid samples. **d** Correlation between expression asymmetry (see **c**) for details) and the dN/dS of the ortholog in the pike sister lineage. **e** Predicted bound TFBS from TF-footprinting in promoters of ohnologs in the five expression evolution categories as well as those ohnologs with no significant shift in expression levels. For each ohnolog pair in each category, copies are grouped based on the lowest (to the left) and highest (to the right)  $p$ -value in the OU-test for expression level shift.  $p$ -values from significant paired Wilcoxon tests are indicated above boxplots: \*\*\*  $< 1e-03$ , \*\*\*\*  $< 1e-04$ , \*\*\*\*\* = 0



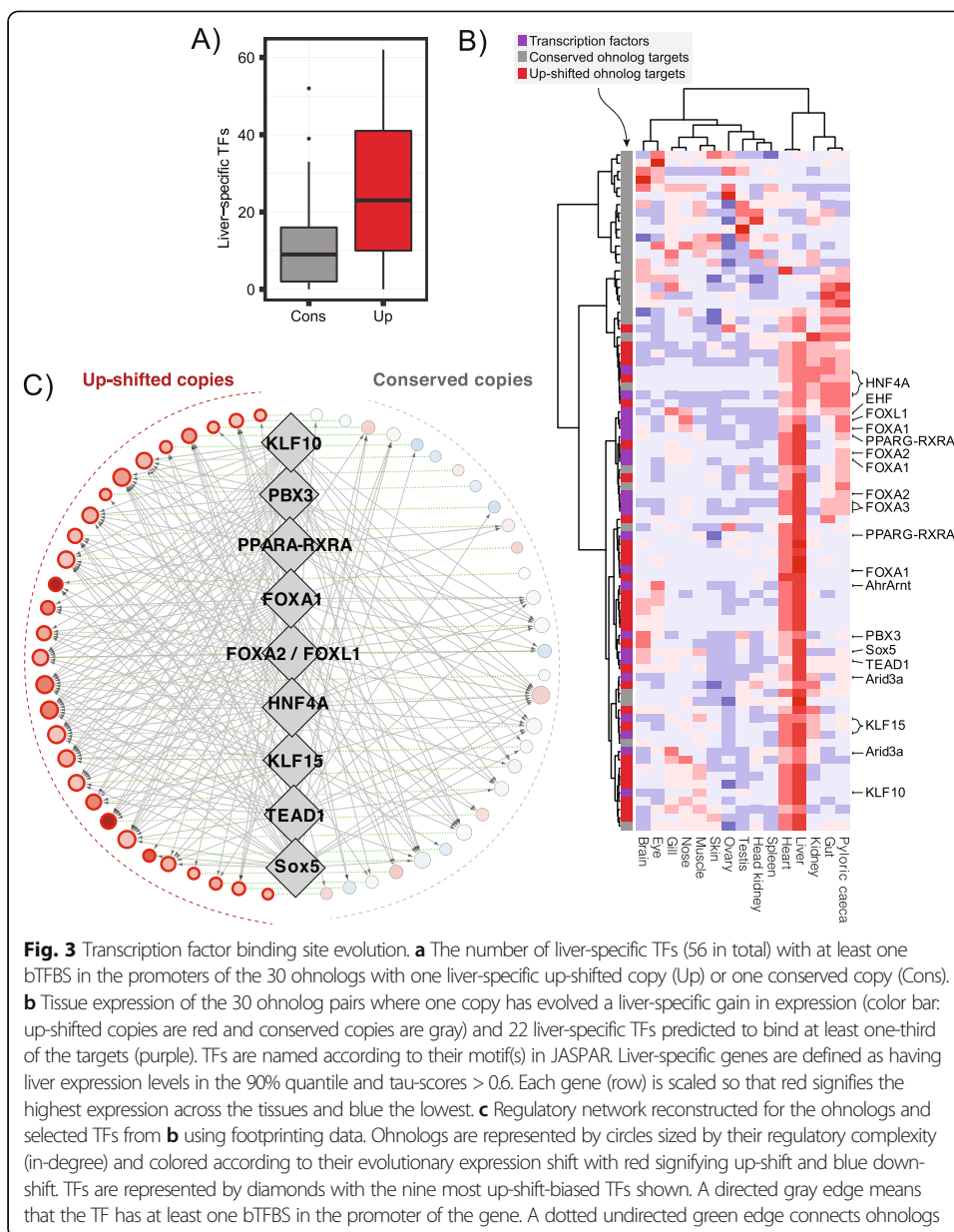
As ribosome subunit genes are known to be extremely slowly evolving genes (i.e., high sequence evolution constraints), we tested whether there is a broader correlation between sequence constraints and regulatory symmetry. Indeed, we find that ohnolog expression level symmetry is significantly correlated with the level of purifying selection on coding sequences (Spearman correlation,  $p = 1e-8$ , Fig. 2d).

To further dissect regulatory mechanisms driving ohnolog expression level evolution, we generated high coverage ATAC-seq data from the liver of Atlantic salmon and identified bound transcription factor binding sites (bTFBSs) using a footprinting approach (Additional file 1: Figure S15). We hypothesized that ohnolog regulatory evolution symmetry is shaped by the relative importance of selection on cis- versus trans-mutations. One simple prediction from this is that ohnolog pairs where one copy has evolved novel expression would have higher promoter divergence than ohnolog pairs with symmetric evolution. The divergence of bTFBSs in promoters ( $-3000/+200$  bps from transcription start site) largely matched this prediction (Fig. 2e) with ohnologs having more asymmetric expression shifts (up+cons and down+cons) differing more with respect to the number of bTFBSs in their promoters compared to symmetrically evolving ohnologs (up+up, down+down, and cons+cons) (Fig. 2e). This offers a simple explanation of expression divergence after WGD, where genes with decreased expression level have lost TFBSs, and genes with increased expression have gained TFBSs, compared to the ancestral promoter structure. Comparing the overall similarity of promoters, computed as the correlation of bTFBS between symmetrically evolving (down+down) and asymmetrically evolving (down+cons) ohnolog pairs, did not reveal a similar trend (Wilcoxon test,  $p = 0.234$ , Additional file 1: Figure S16), which is consistent with high turnover of bTFBS even for highly conserved genes [31].

Together these results support that evolutionary constraints at the coding sequence divert ohnologs down different evolutionary routes towards novel gene dosage—either in an asymmetric or symmetric fashion.

### **Adaptive gain in liver expression through acquisition of tissue-specific cis-regulatory elements**

Although the vast majority of adaptive expression evolution was associated with selection on lower gene dosage, our OU-analyses did reveal 30 ohnolog pairs where one copy had evolved liver-specific adaptive gains in expression following WGD. These genes are predicted to be involved in a variety of functions such as developmental processes, cell fate specificity, and more liver-centric functions such as endocrine signaling and lipid- and fatty-acid metabolism (Additional file 1: Table S5). To better understand the regulatory mechanisms involved in the evolution of these potential novel liver functions, we used our TF-footprinting data to test the hypothesis that adaptive gains in liver expression are linked to the acquisition of binding sites for TFs controlling liver-specific regulatory networks. Indeed, we found that promoters of up-shifted copies were occupied by many more liver-specific TFs than their non-shifted partners (Fig. 3a, Wilcoxon paired test,  $p = 7.7e-05$ ). These liver-specific TFs are thus candidates for being involved in regulatory rewiring of up-shifted ohnologs (Fig. 3b). Interestingly, many TFs with the strongest bias towards occupying the promoters of up-shifted ohnolog copies have known general liver functions (i.e., hepatocyte nuclear factors; FOX1A,



**Fig. 3** Transcription factor binding site evolution. **a** The number of liver-specific TFs (56 in total) with at least one bTFBS in the promoters of the 30 ohnologs with one liver-specific up-shifted copy (Up) or one conserved copy (Cons). **b** Tissue expression of the 30 ohnolog pairs where one copy has evolved a liver-specific gain in expression (color bar: up-shifted copies are red and conserved copies are gray) and 22 liver-specific TFs predicted to bind at least one-third of the targets (purple). TFs are named according to their motifs in JASPAR. Liver-specific genes are defined as having liver expression levels in the 90% quantile and tau-scores > 0.6. Each gene (row) is scaled so that red signifies the highest expression across the tissues and blue the lowest. **c** Regulatory network reconstructed for the ohnologs and selected TFs from **b** using footprinting data. Ohnologs are represented by circles sized by their regulatory complexity (in-degree) and colored according to their evolutionary expression shift with red signifying up-shift and blue down-shift. TFs are represented by diamonds with the nine most up-shift-biased TFs shown. A directed gray edge means that the TF has at least one bTFBS in the promoter of the gene. A dotted undirected green edge connects ohnologs

HNF4A) [32] and roles in lipid metabolism (RXR, PPARG, KLF15) [33, 34] (Fig. 3c, see the “Methods” section for details).

Next, we hypothesized that liver-specific increases in expression are driven by gains in new TFBSs. One way promoters can gain novel TFBSs is through insertions of TEs that either contain a functional TFBS or subsequently accumulate mutations that give rise to new TFBSs [35]. Indeed, we did find that TFBSs predicted to be bound by liver-specific TFs overlapped TEs more often in up-shifted copies than in conserved copies (Wilcoxon paired test,  $p = 0.037$ , Additional file 1: Figure S17A). Furthermore, at the level of TE superfamilies we found that the TIR TC1-Mariner TE superfamily were associated with gain in liver-specific bTFBS in up-shifted copies ( $p = 0.018$ , Additional file 1: Figure S17B), which included known liver and lipid metabolism transcription factors such as HNF4A, KLF15, and RXRA (Additional file 1: Table S6).



In conclusion, we find that adaptive gain in liver-specific expression is strongly associated with gain in liver-specific bound TFBSs, some of which have been facilitated by transposable element insertions.

## Discussion

The consequence of WGDs for evolution of novel adaptations, including gene expression levels, has been an actively debated topic within evolutionary biology [1]. A key challenge has been to distinguish neutral from adaptive evolution in systems where experimental evolution is not possible [12]. Here, we generated a large comparative transcriptomics dataset and for the first time applied a formal phylogenetic model to infer selection on gene expression in the aftermath of a vertebrate WGD that occurred 80–100 million years ago.

### Selection on gene dosage ameliorates immediate polyploid fitness costs

Newly formed polyploids often display augmented rates of abnormal mitosis, chromosome loss, and gross chromosomal rearrangements [36, 37]. Hence, a primary challenge for the evolutionary success of polyploids is to maintain genomic stability. In line with this, we find that adaptive evolution of gene expression was highly biased towards cellular functions not specific to the liver (Figs. 1e, f and 3b) and with a clear potential impact on genome stability. Firstly, we find genes directly involved in the cell cycle to be enriched for adaptive evolution (higher dosage). Related genes have experienced selective sweeps following WGD in plants [38, 39]. Furthermore, we find strong evidence for selection on genes involved in oxidative phosphorylation (lower dosage). Polyploidization in plants, fungi, and mammalian cells have been shown to increase levels of reactive oxygen species, which is causally linked to increased cellular stress, cell cycle failure, and increased genome instability [40–42]. Lastly, we find adaptive expression evolution (lower dosage) for genes involved in translation (ribosome subunits and ribosome assembly) after WGD. Regulation of translation also interacts with cell cycle regulation, with potential implications for genome stability [43]. However, selection for decreased expression of translation-related genes could also be linked to direct fitness costs of wasteful protein translation or harmful effects linked to the over-production of particular proteins. Overall, our study provides evidence for a scenario where a critical first step in becoming a successful polyploid lineage is pervasive adaptive evolution on gene dosage to ameliorate fitness costs linked to genome stability.

### Long-term ohnolog retention and selection on gene dosage

Following an early phase of selection on gene dosage, the long-term fates of ohnologs can be shaped by various adaptive processes [21, 44], including adaptive regulatory evolution. One potential outcome is adaptive divergence between ohnologs, resulting in two functionally non-redundant ohnologs under purifying selection. Our results demonstrate that tissue-specific shifts (up+cons) in expression are rare (Fig. 1c), and interpret this to mean that adaptive evolution of novel tissue-specific regulation likely has very little impact on genome wide ohnolog retention.

Selection on molecular stoichiometry has been proposed to play a major role in genome evolution after WGDs [30]. This narrative is supported by our finding that

molecular complexes are both enriched for retained ohnologs and biased to include only singletons or only ohnologs (Additional file 1: Figure S14). However, selection on molecular stoichiometry could also drive evolution of gene expression. Indeed, we do find some (but weak) support for selection on stoichiometric balance also operating through selection for higher expression levels of singleton genes that are in complexes with ohnologs (Additional file 1: Table S3,  $p = 0.07$ ). Moreover, it is plausible that the strong bias of “oxidative phosphorylation”-ohnologs towards highly asymmetric expression regulation also is linked to selection on stoichiometry (Fig. 2b, c). These genes are nuclear encoded genes involved in energy-related functions in mitochondria. As WGDs do not double the plastid numbers it has been proposed that in plants stoichiometric imbalances between nuclear and plastid genomes act as selection pressure to reduce the ratio of nuclear to plastid gene dosage following WGD [45]. In line with this reasoning, the driver behind the strong asymmetric down shift of “oxidative phosphorylation”-ohnologs could be the reinstatement of stoichiometric balance between the nuclear and mitochondrial genes.

At the other end of the ohnolog expression evolution symmetry spectrum, we find ohnologs belonging to the “ribosomal protein” pathway evolving lower expression in a highly symmetric fashion (Fig. 2b, c). We also demonstrate a significant correlation between constraints at the coding sequence level and symmetry of ohnolog regulatory evolution (Fig. 2d). This is in line with findings from plants that ribosomal proteins are retained over long evolutionary times and evolve slowly at both sequence and expression levels following WGD [46]. One potential explanation for this pattern could be the “toxic effects model” where long-term conservation of ohnologs is intrinsically linked to the “danger” of accumulating highly toxic coding sequence mutations [47, 48]. We therefore hypothesize that in situations where lowering the total gene dosage increases fitness, and the tolerance for accumulation of deleterious mutations is low (i.e., the toxic effect), symmetric ohnolog evolution towards lower gene dosage could be favored over pseudogenization of one copy. Eventually, mutations can arise that create completely non-functional pseudogenes without toxic-effects, and these can then be fixed in the population. This would result in an enrichment of singletons among genes that are likely to produce toxic effects, as observed in plants [45].

### Divergence of chromatin landscapes and ohnolog expression

Regulatory divergence after gene duplication is hypothesized to be linked to evolution of local chromatin landscapes [18, 49]. Using ATAC-seq data we show that signals of adaptive expression level shifts are associated with the numbers of bound TFBSs (Fig. 2e), consistent with a billboard-like model of gene regulation [50]. Furthermore, we find that both loss of expression (Fig. 1h) and tissue-specific gains in expression level (Additional file 1: Figure S17) is linked to TE activity, highlighting the dual role of TEs in regulatory evolution following WGD.

### Conclusion

Our study supports pervasive selection on gene dosage across millions of years following WGD, in particular for genes involved in basic cellular maintenance and genome stability. Interestingly, many of the homologous genes and pathways also show similar

responses in gene dosage adjustments immediately after polyploidization in plants [16]. Reconciling these immediate effects of polyploidization with our findings strongly supports the following model: Plastic genome regulatory response to polyploidization alleviate immediate fitness costs following genome doubling. Since gene loss is absent in early generations polyploids, and all genes are duplicated and in stoichiometric balance, early plastic changes in gene regulatory phenotypes is likely a result of deleterious fitness effects due to suboptimal absolute gene dosages. Over evolutionary time-scales however, selection will favor and fix regulatory mutations that can “hard code” novel transcriptional phenotypes to optimize gene dosages (as seen following the salmonid WGD). Together, this paper points to critical genome regulatory adjustments for becoming a successful polyploid lineage.

## Methods

### Ortholog inference

For ortholog inference, we used thirteen species including six salmonids (*Thymallus thymallus*, *Hucho hucho*, *Salmo salar*, *Salvelinus* sp., *Oncorhynchus mykiss*, and *Oncorhynchus kisutch*), four teleosts as outgroups to the salmonids (*Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus*, and *Esox lucius*), one non-teleost fish (*Lepisosteus oculatus*) and two mammals as outgroups to the teleosts (*Homo sapiens* and *Mus musculus*). We only report the genus name for the char (*Salvelinus* sp.) because it was recently discovered that the material used for sequencing *Salvelinus alpinus* could have been a very closely related sister species (*Salvelinus malma*) or a hybrid between the two [51]. Protein sequences were obtained from ENSEMBL (release 92) for *H. sapiens*, *M. musculus*, *L. oculatus*, *D. rerio*, *O. latipes*, and *G. aculeatus*, from NCBI RefSeq assemblies for *S. salar* (GCF\_000233375.1), *Salvelinus* sp. (GCF\_002910315.2), *O. mykiss* (GCF\_002163495.1), *O. kisutch* (GCF\_002021735.1), and *E. lucius* (GCF\_000721915.3), from the genome paper for *T. thymallus* [52] and from an in-house annotation using Transdecoder (<https://github.com/TransDecoder/TransDecoder/wiki>) for *H. hucho* (GCA\_003317085). The single longest protein per gene was assigned to gene ortholog groups (orthogroups) using OrthoFinder (v2.3.1) [53]. For each orthogroup, the corresponding CDS sequences were aligned using MACSE (v2.03) before gene trees were generated and reconciled against the species tree using TreeBest (v1.9.2). The gene trees were then split at the level of monophyletic teleost clades, defining what we refer to as trees in this article, and again at the level of the salmonid clade (excluding *T. thymallus* and *H. hucho*), defining the Ss4R duplicate clades. Trees were then selected based on their topology (Additional file 1: Figure S1). Specifically, this filtered any trees that showed more than two salmonid clades or that contained additional paralogs inside the salmonid clades or in the outgroup species. Trees with all orthologs retained in the salmonid clade(s) were designated as complete, and otherwise as partial. In addition, trees were excluded from further analysis if (1) one or both salmonid clades had no expressed genes (zero mapped reads, RNA-seq data described below), (2) the *E. lucius* ortholog was missing or not expressed, and (3) both the *D. rerio* and *O. latipes* orthologs were missing or not expressed.

### RNA-sequencing data

Liver tissue samples were collected from adult individuals of *D. rerio* (zebrafish), *O. latipes* (medaka), *E. lucius* (pike), *O. mykiss* (rainbow trout), *S. alpinus* (Arctic char), and *O. kisutch* (coho salmon) (Fig. 1a). Samples were taken in replicates of four, or three in the case of rainbow trout. All fish were raised in fresh water under standard rearing conditions in aquaculture facilities (salmonids), animal laboratory facilities (zebrafish and medaka), or restocking hatcheries (pike). Total RNA was extracted from the liver samples using the RNeasy Plus Universal Kit (QIAGEN). Quality was determined on a 2100 Bioanalyzer using the RNA 6000 Nano Kit (Agilent). Concentration was determined using a Nanodrop 8000 spectrophotometer (Thermo Scientific). cDNA libraries were prepared using the TruSeq Stranded mRNA HT Sample Prep Kit (Illumina). Library mean length was determined by running on a 2100 Bioanalyzer using the DNA 1000 Kit (Agilent) and library concentration was determined with the Qbit BR Kit (Thermo Scientific). Paired-end sequencing of sample libraries was completed on an Illumina HiSeq 2500 with 125-bp reads. Raw RNA-seq and processed count data have been deposited into ArrayExpress under the projects E-MTAB-8959 and E-MTAB-8962. For *S. salar* (Atlantic salmon), RNA-seq data was obtained from a feeding trial using four samples from individuals in freshwater fed a marine based diet [54], available in the European Nucleotide Archive (ENA) under project PRJEB24480 (samples: ERS2101563, ERS2101567, ERS2101568, ERS2101569).

To generate gene expression data, RNA-seq reads were mapped to the annotated reference genomes using the STAR aligner with default settings [55]. RSEM [56] was used to estimate read counts and Transcripts Per Million reads (TPM)-expression values that are normalized for average transcript lengths and the total number of reads from each sample.

The trimmed mean of M values (TMM), from the R package edgeR [57], was used to compute normalization factors for the gene expression data. The replicates were first normalized within each species and then between species (Additional file 1: Figure S2). Between-species normalization was accomplished by first computing species-specific normalization factors using genes from singleton orthogroups (i.e., groups containing only one gene from each species) and their mean expression values (i.e., mean of the replicates within each species), and then by normalizing the individual replicates from each species using these normalization factors. All expression values were log transformed ( $\log_2(\text{TPM}+0.01)$ ) prior to testing for expression shifts.

### Evolutionary shifts in gene expression

The EVE model [22] was used to test for shifts in gene expression levels in the salmonid clade(s) within each gene tree. For this paper, we developed and implemented a user friendly version of the EVE algorithm in R (<https://gitlab.com/sandve-lab/evemodel>). This method models an OU process, i.e., random drift in expression level that is constrained around an optimal level. The test compares a model with two optimal expression levels, one for the salmonid branch and another for the outgroup species, against the null-model which has the same optimal expression level across the entire tree (Additional file 1: Figure S3C). For ohnolog gene trees which contain two

duplicate salmonid clades, each clade was tested separately by removing the other salmonid clade.

EVE was given the expression data for each species (four samples/replicates per species) and the species tree produced by OrthoFinder. For every ortholog, a likelihood ratio test (LRT) score is calculated, representing the likelihood of the alternative hypothesis over the null hypothesis. LRT scores were compared to a chi-squared distribution with one degree of freedom and scores above the 95% quantile were considered to be significant. EVE reports estimates of the expression optimum for the salmonid branch and the rest of the tree (i.e., outgroup species), and the difference between salmonid estimates and outgroup estimates provided the direction of the expression shift.

### **Tissue atlas**

Gene expression data from an Atlantic salmon tissue atlas [17] was clustered using Pearson correlation and the R function `hclust` with `method = "ward.D"`. Heatmaps were drawn using the R function `pheatmap` with `scale = "row"`.

### **Coding sequence selection pressure**

We estimated branch-specific selection pressure on coding sequences in ohnolog gene trees by calculating dN/dS measured at the branch from the WGD node to the root of each duplicate clade using the aBSREL (adaptive Branch-Site Random Effects Likelihood) method [58] in Hyphy (Hypothesis Testing using Phylogenies) [59]. A one-sided paired Wilcoxon test was then performed to test if there is a difference in selection pressure between ohnolog pairs classified as asymmetrically shifted at the expression level.

### **Transposable elements**

Transposable element (TE) annotations were taken from [17]. For Atlantic salmon genes, we calculated the proportion of gene promoter sequence (+ 2 kb/-200b from TSS) that was overlapped with TEs using `bedtools intersect` of promoter and TE annotations. We used a one-sided paired Wilcoxon test to test the hypothesis that, for ohnologs with an asymmetric shift down in expression, the shifted copy had a higher proportion of TE overlap than the conserved copy.

### **Gene function enrichment**

We assigned KEGG pathway annotations to the orthogroups based on the Northern pike ortholog and its KEGG annotations. We then tested each set of ohnologs within an expression shift category for the enrichment of KEGG pathways using the `kegga` function from the R package `limma`, with all tested ohnologs as the background.

### **Protein complexes**

We assigned orthogroups as being in a protein complex or not based on the human ortholog and its protein complex annotations from the CORUM database [60]. We used the Fisher's exact test, for singleton and ohnolog genes, to test whether more genes within an expression shift category were in a protein complex than expected by chance. To test if complexes were biased towards only containing singletons or only

ohnologs, we randomized the singleton/ohnolog label 10,000 times and reported empirical  $p$ -values.

#### ATAC-seq generation and TF footprinting

Four Atlantic salmon (freshwater stage, 26–28 g) were euthanized using a Schedule 1 method following the Animals (Scientific Procedures) Act 1986. Around 50-mg homogenized brain and liver tissue was processed to extract nuclei using the Omni-ATAC protocol for frozen tissues [61]. Nuclei were counted on an automated cell counter (TC20 BioRad, range 4–6  $\mu\text{m}$ ) and further confirmed intact under microscope. A total of 50,000 nuclei were used in the transposition reaction including 2.5  $\mu\text{L}$  Tn5 enzyme (Illumina Nextera DNA Flex Library Prep Kit), incubated for 30 min at 37 °C in a shaker at 200 rpm. The samples were purified with the MinElute PCR purification kit (Qiagen) and eluted in 12  $\mu\text{L}$  elution buffer. qPCR was used to determine the optimal number of PCR cycles for library preparation [62] (8–10 cycles used). Sequencing libraries were prepared with short fragments and fragments > 1000 bp removed using AMPure XP beads (Beckman Coulter, Inc.). Fragment length distributions and confirmation of nucleosome banding patterns were determined on a 2100 Bioanalyzer (Agilent) and the library concentration estimated using a Qubit system (Thermo Scientific). Libraries were sent to the Norwegian Sequencing Centre, where paired-end 2 × 75 bp sequencing was done on an Illumina HiSeq 4000. The raw sequencing data for brain and liver is available through ArrayExpress (Accession: E-MTAB-9001).

Reads were mapped using BWA-MEM [63]. Duplicate reads and reads mapping to mitochondrial or unplaced scaffolds were removed. Peaks were called using MACS2 [64]. TF footprinting was performed with TOBIAS [65] based on the aligned reads, peaks, and TF motifs from JASPAR (JASPAR 2020 non-redundant vertebrate CORE PFMs) [66]. TOBIAS performs Tn5 bias correction, generates footprint scores for each base within the peaks, scans for TFBSs using the given TF motifs, and finally classifies each TFBS as bound or unbound based on the footprint scores.

For the analysis of ohnolog pairs with evolved liver-specific expression increases in one copy, we identified 30 up+cons pairs (60 target genes) where the liver expression of the up-copy was at least 90% of the maximum expression in the tissue atlas and the up-copy had a tissue specificity score ( $\tau$ ) > 0.6 [17]. To identify regulators of these genes, we BLASTed UniProt TF sequences with a motif in JASPAR to the Atlantic salmon proteome, and retained the top four hits with E-value < 1E–10 and alignment length > 100. We then filtered these TFs for having bTFBS in the promoter of at least 20 of the target genes and for having liver-specific expression (same criteria as for up-targets). This resulted in 22 liver-specific TFs predicted to bind 17 different JASPAR motifs in 52 target promoters (Fig. 3b, c). Finally, to draw the network in Fig. 3c, we (1) selected, for each JASPAR motif, the single TF with the strongest evolutionary shift in expression; (2) removed JASPAR motifs with highly similar binding profiles (> 80% overlap in target genes, retaining the TF with the strongest evolutionary shift); and (3) merged TFs associated with more than one JASPAR motif into one node and selected the nine TFs with the strongest bias towards up-shifted targets.



### Reproducibility

The scripts developed to implement analyses described in this study are available here: <https://gitlab.com/sandve-lab/gillard-groenvold> [67] and <https://doi.org/10.5281/zenodo.4478402> [68].

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02323-0>.

**Additional file 1: Supplementary figures S1-S17 and Supplementary Tables S1-S6.**

**Additional file 2.** Review history.

### Acknowledgements

We thank Zuzana Storchová, Galal Metwalli, Marc Robinson-Rechavi, Gavin Conant, Camille Berthelot, and Jeremy Coate for valuable discussions.

### Review history

The review history is available as Additional file 2.

### Peer review information

Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Authors' contributions

SRS, TRH, and RR conceived the study. GG, LG, LR, MMH, ØM, MKG, DJM, JM, BK, and EBR were involved in formal generation, analysis, and/or curation of data. LG, GG, SRS, and TRH wrote the paper. All authors contributed intellectually to data analyses and interpretation. The authors read and approved the final manuscript.

### Funding

The research was conducted as part of the NRC funded project Rewired (NRC project number 274669) and the NRC and NMBU funded project Transpose (NRC project number 275310). GG was funded by NMBU.

### Availability of data and materials

Raw RNA-seq and processed count data have been deposited into ArrayExpress under the projects E-MTAB-8959 [69] and E-MTAB-8962 [70]. Raw ATAC-seq data is also available through ArrayExpress under project E-MTAB-9001 [71]. The scripts developed to implement analyses described in this study are available on GitLab [67] and Zenodo [68].

### Declarations

#### Ethics approval and consent to participate

The fish used in this study were treated according to the Norwegian Animal Research Authority (NARA) in accordance with the Norwegian Animal Welfare Act of 19th of June 2009.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway. <sup>2</sup>Center for Integrative Genetics, Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway. <sup>3</sup>Department of Biology, University of Victoria, Victoria, Canada. <sup>4</sup>The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, UK. <sup>5</sup>Department of Computer Science, San Francisco State University, San Francisco, USA. <sup>6</sup>Department of Biology, San Francisco State University, San Francisco, USA.

Received: 27 August 2020 Accepted: 23 March 2021

Published online: 13 April 2021

### References

1. Van de Peer Y, Mizrahi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet.* 2017;18(7):411–24. <https://doi.org/10.1038/nrg.2017.26>.
2. Ohno S. *Evolution by gene duplication.* Berlin, Heidelberg: Springer Berlin Heidelberg; 1970. <https://doi.org/10.1007/978-3-642-86659-3>.
3. Merico A, Sulo P, Piskur J, Compagno C. Fermentative lifestyle in yeasts belonging to the Saccharomyces complex. *FEBS J.* 2007;274(4):976–89. <https://doi.org/10.1111/j.1742-4658.2007.05645.x>.

4. Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol.* 2016;30:159–65. <https://doi.org/10.1016/j.pbi.2016.03.015>.
5. Lohaus R, Van de Peer Y. Of dups and dinos: evolution at the K/Pg boundary. *Curr Opin Plant Biol.* 2016;30:62–9. <https://doi.org/10.1016/j.pbi.2016.01.006>.
6. Holland PW, Garcia-Fernández J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. *Development.* 1994;Supplement:125–33.
7. Meyer A, Van De Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays.* 2005;27(9):937–45. <https://doi.org/10.1002/bies.20293>.
8. Voff JN. Genome evolution and biodiversity in teleost fish. *Heredity.* 2005;94(3):280–94. <https://doi.org/10.1038/sj.hdy.6800635>.
9. Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytol.* 2014;202(4):1105–17. <https://doi.org/10.1111/nph.12756>.
10. Andalis AA, Storchova Z, Styles C, Galitski T, Pellman D, Fink GR. Defects arising from whole-genome duplications in *Saccharomyces cerevisiae*. *Genetics.* 2004;167(3):1109–21. <https://doi.org/10.1534/genetics.104.029256>.
11. Kuznetsova AY, Seget K, Moeller GK, de Pagter MS, de Roos JADM, Dürrbaum M, Kuffer C, Müller S, Zaman GJR, Kloosterman WP, Storchová Z. Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. *Cell Cycle.* 2015;14(17):2810–20. <https://doi.org/10.1080/15384101.2015.1068482>.
12. Sandve SR, Rohlfs RV, Hvidsten TR. Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nat Genet.* 2018;50(7):908–9. <https://doi.org/10.1038/s41588-018-0162-4>.
13. Boyle EA, Li Yi, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell.* 2017;169(7):1177–86. <https://doi.org/10.1016/j.cell.2017.05.038>.
14. Verta J-P, Jones FC. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *elife.* 2019;8. <https://doi.org/10.7554/eLife.43785>.
15. Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, Tezuka A, Nagano AJ, Yamasaki YY, Suzuki Y, Kokita T, Takahashi H, Lucek K, Marques D, Takehana Y, Naruse K, Mori S, Monroig O, Ladd N, Schubert CJ, Matthews B, Peichel CL, Seehausen O, Yoshizaki G, Kitano J. A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science.* 2019;364(6443):886–9. <https://doi.org/10.1126/science.aau5656>.
16. Song MJ, Potter B, Doyle JJ, Coate JE. Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana*. *Plant Cell.* 2020;32(5):1434–48. <https://doi.org/10.1105/tpc.19.00832>.
17. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, di Genova A, Samy JKA, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Inge Våge D, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones SJM, Jonassen I, Maass A, Omholt SW, Davidson WS. The Atlantic salmon genome provides insights into rediploidization. *Nature.* 2016;533(7602):200–5. <https://doi.org/10.1038/nature17164>.
18. Marlétaz F, Firas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D, Acemel RD, van Heeringen SJ, Naranjo S, Herrera-Ubeda C, Skvortsova K, Jimenez-Gancedo S, Aldea D, Marquez Y, Buono L, Kozmikova I, Permanyer J, Louis A, Albuixech-Crespo B, le Petillon Y, Leon A, Subirana L, Balwier P, Duckett PE, Farahani E, Aury JM, Mangenot S, Wincker P, Albalat R, Benito-Gutiérrez É, Cañestro C, Castro F, D'Aniello S, Ferrier DEK, Huang S, Laudet V, Marais GAB, Pontarotti P, Schubert M, Seitz H, Somorjai I, Takahashi T, Mirabeau O, Xu A, Yu JK, Carninci P, Martínez-Morales JR, Crollius HR, Kozmik Z, Weirauch MT, Garcia-Fernández J, Lister R, Lenhard B, Holland PWH, Escriva H, Gómez-Skarmeta JL, Irimia M. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature.* 2018;564(7734):64–70. <https://doi.org/10.1038/s41586-018-0734-6>.
19. De Smet R, Sabaghian E, Li Z, Saeys Y, Van de Peer Y. Coordinated functional divergence of genes after genome duplication in *Arabidopsis thaliana*. *Plant Cell.* 2017;29(11):2786–800. <https://doi.org/10.1105/tpc.17.00531>.
20. Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, DiFazio SP. Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. *Genome Res.* 2012;22(1):95–105. <https://doi.org/10.1101/gr.125146.111>.
21. Hallin J, Landry CR. Regulation plays a multifaceted role in the retention of gene duplicates. *PLoS Biol.* 2019;17(11):e3000519. <https://doi.org/10.1371/journal.pbio.3000519>.
22. Rohlfs RV, Nielsen R. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst Biol.* 2015;64(5):695–708. <https://doi.org/10.1093/sysbio/syv042>.
23. Rohlfs RV, Harrigan P, Nielsen R. Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Mol Biol Evol.* 2014;31(1):201–11. <https://doi.org/10.1093/molbev/mst190>.
24. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, da Silva C, Labadie K, Alberti A, Aury JM, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Voff JN, Genêt C, Wincker P, Jaillon O, Crollius HR, Guiguen Y. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun.* 2014;5(1):3657. <https://doi.org/10.1038/ncomms4657>.
25. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc Biol Sci.* 2014;281(1778):20132881. <https://doi.org/10.1098/rspb.2013.2881>.
26. Gout J-F, Lynch M. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol Biol Evol.* 2015;32(8):2141–8. <https://doi.org/10.1093/molbev/msv095>.
27. Langham RJ, Walsh J, Dunn M, Ko C, Goff SA, Freeling M. Genomic duplication, fractionation and the origin of regulatory novelty. *Genetics.* 2004;166(2):935–45. <https://doi.org/10.1534/genetics.166.2.935>.
28. Conant GC. The continuing impact of an ancient polyploidy on the genomes of teleosts. *BioRxiv.* 2019. <https://doi.org/10.1101/619205>.

29. Xu P, Xu J, Liu G, Chen L, Zhou Z, Peng W, Jiang Y, Zhao Z, Jia Z, Sun Y, Wu Y, Chen B, Pu F, Feng J, Luo J, Chai J, Zhang H, Wang H, Dong C, Jiang W, Sun X. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nat Commun*. 2019;10(1):4625. <https://doi.org/10.1038/s41467-019-12644-1>.
30. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A*. 2012;109(37):14746–53. <https://doi.org/10.1073/pnas.1207726109>.
31. Otto W, Stadler PF, López-Giraldéz F, Townsend JP, Lynch VJ, Wagner GP. Measuring transcription factor-binding site turnover: a maximum likelihood approach using phylogenies. *Genome Biol Evol*. 2009;1:85–98. <https://doi.org/10.1093/gbe/evp010>.
32. Lau HH, Ng NHJ, Loo LSW, Jasmen JB, Teo AKK. The molecular functions of hepatocyte nuclear factors - in and beyond the liver. *J Hepatol*. 2018;68(5):1033–48. <https://doi.org/10.1016/j.jhep.2017.11.026>.
33. Prosdócimo DA, Anand P, Liao X, Zhu H, Shelkay S, Artero-Calderon P, Zhang L, Kirsh J, Moore D'V, Rosca MG, Vazquez E, Kerner J, Akat KM, Williams Z, Zhao J, Fujioka H, Tuschl T, Bai X, Schulze PC, Hoppel CL, Jain MK, Haldar SM. Kruppel-like factor 15 is a critical regulator of cardiac lipid metabolism. *J Biol Chem*. 2014;289(9):5914–24. <https://doi.org/10.1074/jbc.M113.531384>.
34. Carmona-Antoñanzas G, Tocher DR, Martínez-Rubio L, Leaver MJ. Conservation of lipid metabolic gene transcriptional regulatory networks in fish and mammals. *Gene*. 2014;534(1):1–9. <https://doi.org/10.1016/j.gene.2013.10.040>.
35. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 2008;9(5):397–405. <https://doi.org/10.1038/nrg2337>.
36. Storchova Z, Pellman D. From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol*. 2004;5:45–54. <https://doi.org/10.1038/nrm1276>.
37. Storchová Z, Breneman A, Cande J, Dunn J, Burbank K, O'Toole E, Pellman D. Genome-wide genetic analysis of polyploidy in yeast. *Nature*. 2006;443:541–7. <https://doi.org/10.1038/nature05178>.
38. Marburger S, Monahan P, Seear PJ, Martin SH, Koch J, Paajanen P, Bohutinská M, Higgins JD, Schmickl R, Yant L. Interspecific introgression mediates adaptation to whole genome duplication. *Nat Commun*. 2019;10(1):5218. <https://doi.org/10.1038/s41467-019-13159-5>.
39. Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet*. 2012;8(12):e1003093. <https://doi.org/10.1371/journal.pgen.1003093>.
40. Roh M, van der Meer R, Abdulkadir SA. Tumorigenic polyploid cells contain elevated ROS and ARE selectively targeted by antioxidant treatment. *J Cell Physiol*. 2012;227(2):801–12. <https://doi.org/10.1002/jcp.22793>.
41. Thomson GJ, Hemon C, Austriaco N, Shapiro RS, Belenky P, Bennett RJ. Metabolism-induced oxidative stress and DNA damage selectively trigger genome instability in polyploid fungal cells. *EMBO J*. 2019;38:e101597. <https://doi.org/10.15252/emboj.2019101597>.
42. del Pozo JC, Ramirez-Parra E. Deciphering the molecular bases for drought tolerance in *Arabidopsis* autotetraploids. *Plant Cell Environ*. 2014;37(12):2722–37. <https://doi.org/10.1111/pce.12344>.
43. Zhou X, Liao W-J, Liao J-M, Liao P, Lu H. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol*. 2015;7(2):92–104. <https://doi.org/10.1093/jmcb/mjv014>.
44. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*. 2008;9(12):938–50. <https://doi.org/10.1038/nrg2482>.
45. De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc Natl Acad Sci U S A*. 2013;110(8):2898–903. <https://doi.org/10.1073/pnas.1300127110>.
46. Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell*. 2004;16(7):1679–91. <https://doi.org/10.1105/tpc.021410>.
47. Roux J, Liu J, Robinson-Rechavi M. Selective constraints on coding sequences of nervous system genes are a major determinant of duplicate gene retention in vertebrates. *Mol Biol Evol*. 2017;34(11):2773–91. <https://doi.org/10.1093/molbev/msx199>.
48. Singh PP, Affeldt S, Cascone I, Selimoglu R, Camonis J, Isambert H. On the expansion of "dangerous" gene repertoires by whole-genome duplications in early vertebrates. *Cell Rep*. 2012;2(5):1387–98. <https://doi.org/10.1016/j.celrep.2012.09.034>.
49. Lan X, Pritchard JK. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science*. 2016;352(6288):1009–13. <https://doi.org/10.1126/science.aad8411>.
50. Kulkarni MM, Arnosti DN. Information display by transcriptional enhancers. *Development*. 2003;130(26):6569–75. <https://doi.org/10.1242/dev.00890>.
51. Christensen KA, Rondeau EB, Minkley DR, Leong JS, Nugent CM, Danzmann RG, Ferguson MM, Stadnik A, Devlin RH, Muzzerall R, Edwards M, Davidson WS, Koop BF. Retraction: the Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLoS One*. 2021;16(2):e0247083. <https://doi.org/10.1371/journal.pone.0247083>.
52. Varadharajan S, Sandve SR, Gillard GB, Tørresen OK, Mulugeta TD, Hvidsten TR, Lien S, Asbjørn Vøllestad L, Jentoft S, Nederbragt AJ, Jakobsen KS. The grayling genome reveals selection on gene expression regulation after whole-genome duplication. *Genome Biol Evol*. 2018;10(10):2785–800. <https://doi.org/10.1093/gbe/evy201>.
53. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
54. Gillard G, Harvey TN, Gjuvsland A, Jin Y, Thomassen M, Lien S, Leaver M, Torgersen JS, Hvidsten TR, Vik JO, Sandve SR. Life-stage-associated remodelling of lipid metabolism regulation in Atlantic salmon. *Mol Ecol*. 2018;27(5):1200–13. <https://doi.org/10.1111/mec.14533>.
55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
56. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):323. <https://doi.org/10.1186/1471-2105-12-323>.
57. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.

58. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol.* 2015;32(5):1342–53. <https://doi.org/10.1093/molbev/msv022>.
59. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, Wisotsky S, Spielman SJ, Frost SDW, Muse SV. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol.* 2020;37(1):295–9. <https://doi.org/10.1093/molbev/msz197>.
60. Giorgi M, Reinhard J, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Ruepp A. CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* 2019;47(D1):D559–63. <https://doi.org/10.1093/nar/gky973>.
61. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods.* 2017;14(10):959–62. <https://doi.org/10.1038/nmeth4396>.
62. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol.* 2015;109(1):21.29.1–9. <https://doi.org/10.1002/0471142727.mb2129s109>.
63. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* 2013;arXiv:1303.3997.
64. Gaspar JM. Improved peak-calling with MACS2. *BioRxiv.* 2018. <https://doi.org/10.1101/496521>.
65. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, Kim J, Looso M. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun.* 2020;11(1):4267. <https://doi.org/10.1038/s41467-020-18035-1>.
66. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2020;48:D87–92. <https://doi.org/10.1093/nar/gkz1001>.
67. Gillard GB, Grønvold L, Sandve SR, Hvidsten TR. Gillard Groenvold GitLab repository. *GitLab.* 2020; <https://gitlab.com/sandve-lab/gillard-groenvold>.
68. Gillard GB, Grønvold L, Hvidsten TR, Sandve SR. Gillard Groenvold source code. *Zenodo.* 2021. <https://doi.org/10.5281/zenodo.4478402>.
69. Gillard GB, Sandve SR. RNA-seq of tissue panel samples from zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), and rainbow trout (*Oncorhynchus mykiss*). E-MTAB-8959. *ArrayExpress.* 2020; <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8959/>.
70. Gillard GB, Sandve SR, Rondeau EB, Koop BF. RNA-Seq of liver tissue samples from northern pike (*Esox lucius*), coho salmon (*Oncorhynchus kisutch*) and Arctic charr (*Salvelinus alpinus*). E-MTAB-8962. *ArrayExpress.* 2020; <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8962/>.
71. Røsæg LL, Holen MM, Sandve SR, Kent MP, Lien S. Fresh versus slow-frozen ATACseq samples for salmon tissues. E-MTAB-9001. *ArrayExpress.* 2020; <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9001/>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

