Inter-individual gene expression variability implies stable regulation of brain-biased genes across organs

Christabel F. Bucao[1,2], Consolée Aletti[1], Alexandre Laverré[1,2], Sébastien Moretti[1,2], Alexandra Trouvé[1], Andrew W. Thompson[3,4,5], Brett L. Racicot[4], Catherine A. Wilson[6], Julien Bobe[7], Ingo Braasch[4,5], Yann Guiguen[7], John H. Postlethwait[6], Marc Robinson-Rechavi[1,2]

[1] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[2] SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3] Department of Biological Sciences, Western Michigan University, Kalamazoo, Michigan, USA
[4] Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA
[5] Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, Michigan, USA
[6] Institute of Neuroscience, University of Oregon, Eugene, Oregon, USA
[7] INRAE, LPGP, Rennes 35000, France

Abstract

Phenotypic variation among individuals plays a key role in evolution, since variation provides the material on which natural selection can act. One important link between genetic and phenotypic variation is gene expression. As for other phenotypes, the range of accessible expression variation is limited and biased by different evolutionary and developmental constraints. Gene expression variability broadly refers to the tendency of a gene to vary in expression (i.e., between individuals or cells) due to stochastic fluctuations or differences in genetic, epigenetic, or environmental factors, separately from the differences between conditions (e.g. organs), and is often estimated based on observed expression variance. The variance due to biomolecular stochasticity (transcriptional 'noise') and cell-to-cell heterogeneity has been well-studied in isogenic populations of unicellular organisms such as bacteria and yeasts. However, for more complex organisms with multiple cells, tissues, and organs sharing the same genetic background, the interplay between inter-individual expression variability, gene and organ function, and gene regulation remains an open question. In this study, we used highly multiplexed 3'-end Bulk RNA Barcoding and sequencing (BRB-seq) to generate transcriptome profiles spanning at least nine organs in outbred individuals of three ray-finned fish species: zebrafish, Northern pike, and spotted gar. For each condition, we measured the expression variance per gene independent of mean expression level as an estimate of its variability. We observed that lowly variable genes are enriched in cellular housekeeping functions whereas highly variable genes are enriched in stimulus-response functions. Furthermore, genes with highly variable expression between individuals evolve under weaker purifying selection at the coding sequence level, indicating that intra-species gene expression variability predicts inter-species protein sequence divergence. Genes that are broadly expressed across organs tend to be both highly expressed and lowly variable between individuals, whereas organ-biased genes are typically highly variable within their top organ of expression. For genes with organ-biased expression profiles, we inferred differences in selective pressure on gene regulation depending on their top organ. We found that

genes with peak expression in the brain have low inter-individual expression variability across non-nervous organs, suggesting stabilizing selection on regulatory evolution of brain-biased genes. Conversely, liver-biased genes have highly variable expression across organs, implying weaker regulatory constraints. These patterns show that gene regulatory mechanisms evolved differently based on constraints on the primary organ.

Introduction

Phenotypic variation among individuals is crucial for evolutionary processes, as it serves as the substrate for natural selection. Gene expression is a key mediator of this phenotypic diversity. Similar to how phenotypic variation is shaped and constrained by evolutionary and developmental factors, variation in gene expression among individuals is also subject to comparable limitations and biases, with some genes having a wider expression distribution than others. This gene expression variance is influenced by various 'intrinsic' (gene-specific) and 'extrinsic' factors spanning biological scales, including transcriptional or translational stochasticity [1], cellular heterogeneity [2], genetic variation [3], epigenetic modifications [2,4], as well as micro- and macro-environmental cues [5]. The observed expression variance (or dispersion) is often used as a proxy measure for expression *variability*, broadly defined as the *propensity* of a gene to vary its expression level in response to stochastic, genetic, or environmental perturbations [6].

Focusing on transcriptional or translational stochasticity ('expression noise') in genetically identical cells, many studies on isogenic populations of unicellular organisms (e.g., *Escherichia coli* [7–11] and *Saccharomyces* [12–18]) have probed the relationship between expression noise and gene promoter features [8–11], as well as properties of the gene regulatory network [19–21]. Evidence suggests that natural selection can shape expression noise according to gene function [8,12,13,15,22] and that constraints on expression noise are mediated by environmental context [23]. Furthermore, a correlation between gene expression noise in isogenic yeast cells and broader expression variation in non-isogenic populations has been reported [24], suggesting that some shared regulatory mechanisms underlie both intra-genotype expression noise and inter-genotype expression variability [25]. The extent to which these findings in unicellular organisms can be extrapolated to complex multicellular organisms, which have functionally specialized organs with distinct expression profiles despite a common genetic background, is not well understood. While many studies have advanced our understanding of genetic variants underlying variation in gene expression levels within a population (e.g., eQTL analysis [26,27]), much remains to be known about the relationship among expression variability, gene function, and gene regulation in animals. Although organ-specific enhancers exist to facilitate spatiotemporal control of gene expression [28], we expect some level of pleiotropy of regulation of gene expression in different organs. Thus selection on gene expression in different organs is not totally independent, and the expression variability of a gene within any given organ may potentially be constrained by antagonistic organ-specific expression optima.

Studies on gene expression variability in animals have primarily focused on select cell types or organs in model organisms [29–31], with particular interests in the role of stochasticity in cell differentiation [32]

(including shifts in inter-cell [33,34] or inter-embryo expression variability over development [35]), as well as expression variability in human organs and its association with disease [36–43]. (Of note, although most studies compare expression in different organs, the terminology of 'tissues' and 'tissue-specificity' is frequently used in the literature instead.) Although the functional relevance of expression variability is increasingly recognized [44–46], how such variability is modulated by organ context remains largely unexplored, even in studies that sample multiple organs [25,40,42,43]. Few studies have been published on variability in outbred, non-model animal populations (but see [47–49]). This is partly due to heavy resource requirements such as the need for many replicates for statistical significance, as well as the challenges associated with disentangling different sources of expression variation [50], whether in bulk or at the single-cell level.

Recently, the advent of low-cost high-throughput sequencing methods has made it increasingly feasible to generate sufficient transcriptome data to measure within-organ gene expression variance across several individuals and estimate within-organ expression *variability*. In this study, we present the first analysis of inter-individual gene expression variability across multiple organs, using newly generated 3'-end Bulk RNA Barcoding and sequencing (BRB-seq) [51] data in three ray-finned fish species. We confirmed that gene expression variability between individuals is a functionally relevant trait. Genes that are broadly expressed across organs have lower within-organ variability in expression between individuals compared to organ-biased genes, reflecting pleiotropic constraints on expression. We further found that genes with organ-biased expression profiles tend to be variably, and thus dynamically, expressed in their top ('focal') organ, and that moreover their variability across other organs depends on their focal organ. Investigating gene expression variability within and between animal organs provides a new framework for understanding gene regulation and the evolution of gene function in complex multicellular organisms.

Results

Organ transcriptome profiles of ray-finned fishes

To investigate inter-individual gene expression variability within and between organs, we generated transcriptomic data across multiple organs using 3'-end Bulk RNA Barcoding and sequencing (BRB-seq) [51] for three species of ray-finned fishes (Actinopterygii): two teleost species (*Danio rerio* (zebrafish) and *Esox lucius* (Northern pike)) and one non-teleost outgroup (*Lepisosteus oculatus* (spotted gar)). Zebrafish and Northern pike represent two of the major teleost clades (Otomorpha and Euteleosteomorpha, respectively) and neither species have undergone additional lineage-specific genome duplications after the teleost whole-genome duplication [52], while the spotted gar is a slowly evolving outgroup to the teleosts [53]. For each species, we sampled multiple outbred individuals spanning several organs in both sexes (when known) (**Figure 1A**) and applied a two-step filter to remove low quality samples based on sequencing quality (**Figures S1-S3**) and within-organ expression correlation (Materials and Methods). We observed that this stringent two-step filtering protocol is sufficient to remove any strong batch effects in our dataset (**Figures S4**-

S6). Hierarchical clustering of zebrafish samples based on pairwise Pearson correlation shows that transcriptome profiles cluster into five major groups by related organ systems (**Figure 1B**; see **Figures S7-S9** for PCA plots). Reproductive organs (cluster 1) appear to have the most divergent expression profiles, while the following clusters can be observed for non-reproductive organs: gastrointestinal and urinary systems (kidney, intestine, and liver; cluster 2); nervous system (brain and eye; cluster 3); external organs (pectoral fin, skin, and gills; cluster 4); and lastly, muscular system (heart and muscle; cluster 5). Apart from the gonads, we do not observe strong clustering by sex except for the liver and pectoral fin. Although some differences can be observed in Northern pike and spotted gar, clustering of organs of the nervous system, muscular system, and external organs are generally consistent among species (**Figure S10**).

A

	D. rerio		*E. lucius*		*L. oculatus*
	♀	♂	♀	♂	♀ / ♂
blood			2	3	6
brain	13	11	10	9	11
eye	11	11	6	5	10
gills	7	9	9	9	6
gonads	11	9	7	5	5
heart	1	2	10	9	9
intestine	11	12	8	8	15
kidney	4	6	9	8	
liver	10	9	4	10	8
muscle	1	1	7	8	7
pectoral fin	10	12	10	10	8
skin	10	11			12
swim bladder			3	5	7
Σ	89	93	85	89	104

B

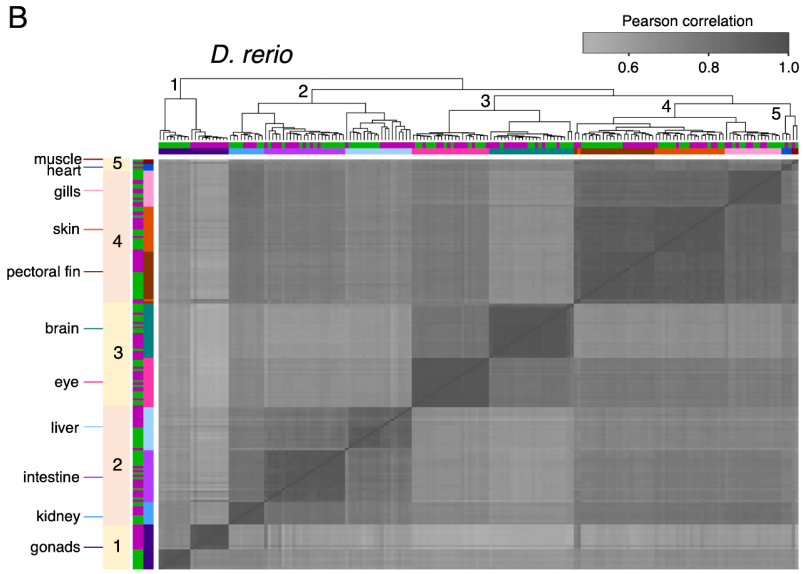


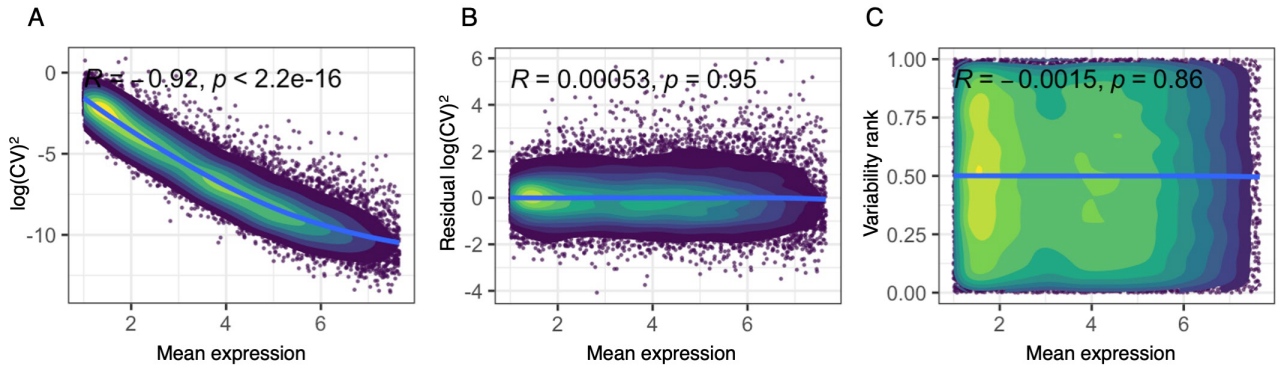
Figure 1. Study design for multi-organ analysis of inter-individual gene expression variability in fishes. (A) Summary of BRB-seq dataset. Table shows the number of individuals per organ (rows) and sex (columns) within each species, after removal of low-quality samples. There is no information on sex for spotted gar (*L. oculatus*). Organ-sex conditions with fewer than 4 replicates (light gray) were excluded from expression variability analyses. **(B)** Hierarchical clustering heatmap of organ transcriptomes (n = 182) in zebrafish (*D. rerio*). Samples cluster into 5 groups by related organ systems: reproductive system (1), gastrointestinal and urinary system (2), nervous system (3), external organs (4), and muscular system (5). Pairwise distances between samples are computed based on Pearson correlation of log-transformed, TMM-normalized counts per million ($\log_2$ TMM-CPM) of protein-coding genes (n = 19,466). The color keys for organ (inner color bars) and sex (outer color bars) are shown in (A).

Gene expression variability is estimated independently of expression level

Here, we use 'variation' to refer to a measurable statistical feature of an ensemble (i.e., statistical dispersion), and 'variability' to refer to a dispositional property inferred from such measurements [6]. Measuring expression variation as a proxy for quantifying variability has not been standardized in the literature, as it is not well-studied beyond unicellular organisms or model cell lines. Different measures of statistical dispersion (e.g., standard deviation, variance, median absolute deviation, or coefficient of variation (CV)) are known to be strongly anti-correlated with mRNA abundance (**Figure 2A, Figure S11**) [54,55]. To correct for this mean-dispersion relationship, many studies use regression to model variation as a function of expression level [55] and use either the ratio [30,35,42] or difference [25,41] (i.e., model residual) of observed to expected variation as an adjusted estimate of expression variability. Although this removes the global correlation between expression mean and variation, this does not correct for the non-homogenous distribution of variation across the range of expression levels [29,40]. To address this, other studies apply percentile rank normalization (e.g., 'local coefficient of variation' (LCV) [40,47] or 'relative noise rank' [29]), in which the effect of mean expression is removed by ranking the expression variation of each gene relative to a window of genes with similar expression levels. Here, we used the local (residual) coefficient of variation as our estimate of expression variability because it is globally uncorrelated with mean expression level, as in regression-based approaches, and it also imposes a uniform expression variability distribution across expression levels. Although the latter is a very strong correction, by forcing a uniform distribution we maximally remove mean expression as an explanatory factor for downstream analyses.

Since actual expression variation may be skewed by samples with more divergent expression profiles, we computed the LCV for each gene within an organ (for unsexed spotted gar) or organ-sex condition (for sexed Northern pike and zebrafish) using a jackknife resampling or 'leave-one-out' procedure to mitigate the effect of outliers on expression variability estimates. For each subset of n - 1 replicates, we first computed the log-transformed, squared coefficient of variation ($\log(CV)^2$, where CV is the ratio of standard deviation to the mean) and fitted a local polynomial regression (LOESS) to model $\log(CV)^2$ as a function of mean expression (**Figure 2B**). Second, we ranked the residual $\log(CV)^2$ ('residual CV') of each gene within sliding windows of 100 neighboring genes sorted by expression level, such that genes with rank 0.00 have the lowest dispersion whereas genes with rank 1.00 have the highest dispersion within a local gene neighborhood (**Figure 2C**). Lastly, we calculated the mean LCV over all n - 1 subsets within a condition and used this metric, which we call 'variability rank', as our estimate of gene expression variability (**Figures S12-S14**).

159 Because variability ranks are computed relative to the set of genes expressed within a specific
160 condition, it is not an absolute measure of variability that is directly comparable between conditions. A gene
161 with a high variability rank in the brain is considered highly variable only with respect to other genes expressed
162 in the brain (e.g., a gene with a variability rank of 0.80 in the brain is not necessarily more variable, in the
163 absolute sense, than a gene with a variability rank of 0.60 in the liver). By its construction, variability ranks are
164 thus generally organ-specific, with weak to moderate correlation across organs (**Figures S15-S17**).
165



166
167 **Figure 2. Measures of expression variation as a function of expression level.** Data from female zebrafish brain (*n* = 13 samples). All
168 values are means estimated from jackknife (*n* - 1) resampling. Mean expression refers to log-transformed, TMM-normalized counts
169 per million ($\log_2$ TMM-CPM). **(A)** Log-transformed (squared) coefficient of variation (CV, defined as the ratio of standard deviation to
170 the mean). **(B)** Residuals of $\log_2$ (squared) CV. **(C)** Local percentile rank of residual $\log(CV)^2$ ('variability rank') within sliding windows of
171 100 neighboring genes sorted by expression level. Blue lines are LOESS regression curves and *R* is Pearson's correlation coefficient.
172

173 One limitation of the variability rank metric is that it does not describe the modality of the underlying
174 expression distribution. Overdispersion may be due to either a broader range of expression centered at a
175 single mean (i.e., a unimodal distribution) or the presence of two or more expression peaks (a multimodal
176 distribution), the latter of which may be indicative of hidden substructures within the population [46]. To
177 investigate this possibility, we assessed whether more bimodal expression profiles can be observed for sets of
178 increasingly more variable genes (**Figure S18**) (Supplementary Results; Supplementary Methods). Overall, we
179 find weak evidence for bimodality in our dataset, only in some conditions, and only at very high variability
180 ranks. Thus, we infer that higher variability is likely primarily driven by a wider range of expression centered
181 around a single peak, rather than by bimodality of expression.
182

183 **Gene expression variability is associated with gene function**

184 After applying a strong correction to remove the effect of mean expression, we verified that our estimate of
185 gene expression variability retains a functionally relevant signal. For this, we performed a Gene Ontology (GO)
186 enrichment analysis of GO biological process terms repeatedly represented (*n* ≥ 3 conditions) in genes with
187 lowly or highly variable expression. We observed that sets of lowly (respectively highly) variable genes show a
188 substantial overlap in enriched GO biological process categories across several conditions. Underdispersed
189 genes (variability rank ≤ 0.20) are consistently enriched in core cellular (housekeeping) functions such as

protein localization and transport, protein modification, RNA processing, and organelle organization (**Figure 3A**). Conversely, overdispersed genes (variability rank ≥ 0.80) are enriched in functions such as signaling, chemotaxis, response to stimuli, and immune response (**Figure 3B**). Functional categories that are under-represented in lowly variable genes also tend to be over-represented in highly variable genes, and vice versa (**Figure S25**). We also performed a GO enrichment analysis for moderately dispersed genes (0.40 ≤ variability rank ≤ 0.60) but observed virtually no over-represented terms, and under-representation of terms related to defense response, immune response, and muscle movement (**Figure S26**). These findings are consistent with previous studies in humans [42,43] and *Drosophila* [25] which have shown that lowly and highly variable genes are enriched in distinct functional categories.

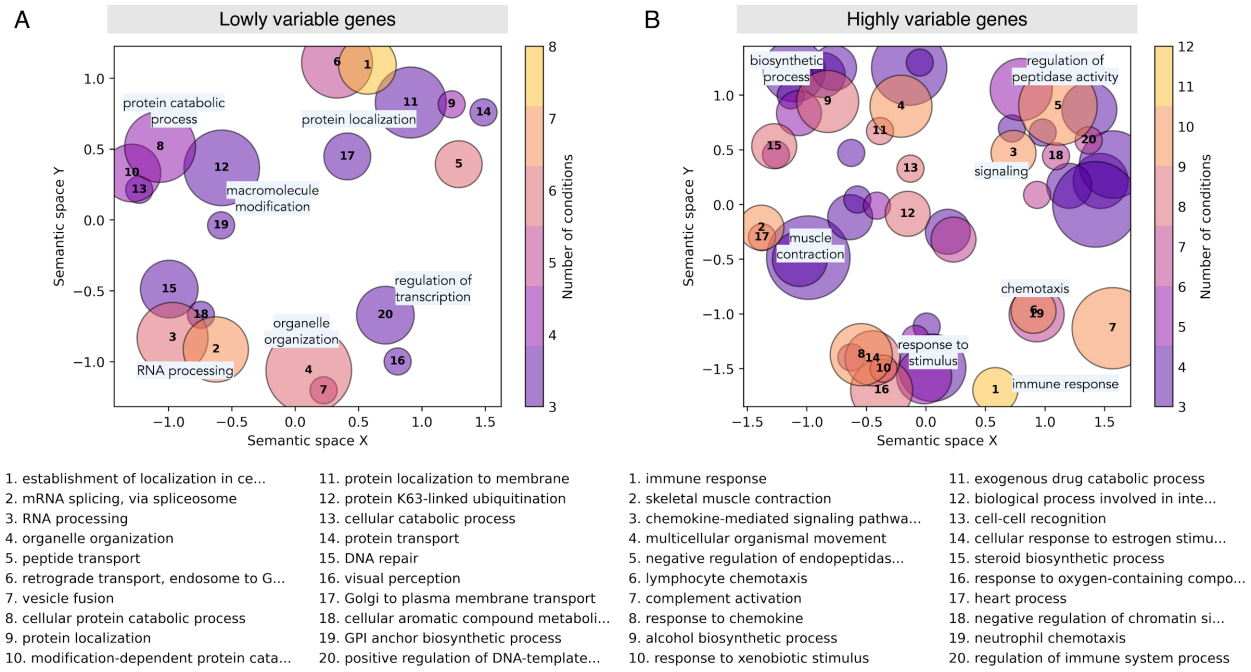


Figure 3. Semantic similarity bubble plots of enriched Gene Ontology (GO) biological process terms for (A) underdispersed (variability rank ≤ 0.20) and (B) overdispersed (variability rank ≥ 0.80) genes in zebrafish organs. Over-representation analysis of GO terms was performed separately for the set of under- and overdispersed genes in each organ-sex condition using the conditional hypergeometric test in *GOstats* **[56]**. Terms that are significant (unadjusted *p*-value < 0.01) in at least three organ-sex conditions were retained. Redundancy reduction and visualization of terms was performed using *GO-Figure!* **[57]**. Each circle represents a group of similar GO terms collapsed into a representative term. Representative terms are plotted in a semantic similarity space such that more similar terms are positioned closer together. The top 20 representative terms are listed and sorted by the number of organ-sex conditions in which the representative is significant. Circle color indicates the number of conditions in which a representative term is significant while circle size is scaled by the number of related terms summarized in a cluster.

214 **Variably expressed genes evolve under weaker purifying selection**

215 By testing for an association between inter-individual gene expression variability and dN/dS, we investigated
216 whether variability evolves under selection pressures that run in parallel to those acting on the protein-coding
217 sequence. For this, precomputed selection statistics were retrieved from Selectome [58,59], a database of
218 conservative branch-site likelihood tests for selection conducted on vertebrate gene trees (Materials and
219 Methods). For each species, the most underdispersed genes (median variability rank ≤ 0.2 averaged across
220 conditions) have evolved under stronger negative selection (dN/dS) compared to the most overdispersed
221 genes (median variability rank > 0.8 across conditions). As there is little variance in dN/dS on the bulk of genes,
222 the global inverse correlation is significant but relatively weak (**Figure 4**; **Figure S20**).
223 Next, we compared evidence for positive selection for genes with consistently low, moderate, or high
224 expression dispersion across multiple conditions within a species. We observed a statistically significant
225 association between estimated variability and the proportion of branches with evidence for positive selection
226 in zebrafish (Pearson's chi-squared test, χ^2= 21.866, p = 1.786 x 10^{-5}) and Northern pike (χ^2= 48.122, p = 3.552
227 x 10^{-11}), but not in spotted gar (χ^2= 1.4127, p = 0.4935). Specifically, we found depletion of branches under
228 positive selection for frequently underdispersed genes (variability rank ≤ 0.20 in at least half of available
229 conditions) in both zebrafish and pike, as well as enrichment of positive selection for frequently overdispersed
230 genes (variability rank ≥ 0.80 at least half of conditions) in pike (**Table S4**). Thus, higher gene expression
231 variability between individuals is associated with faster protein-coding sequence evolution between species,
232 both due to relaxed negative selection and to increased positive selection.
233

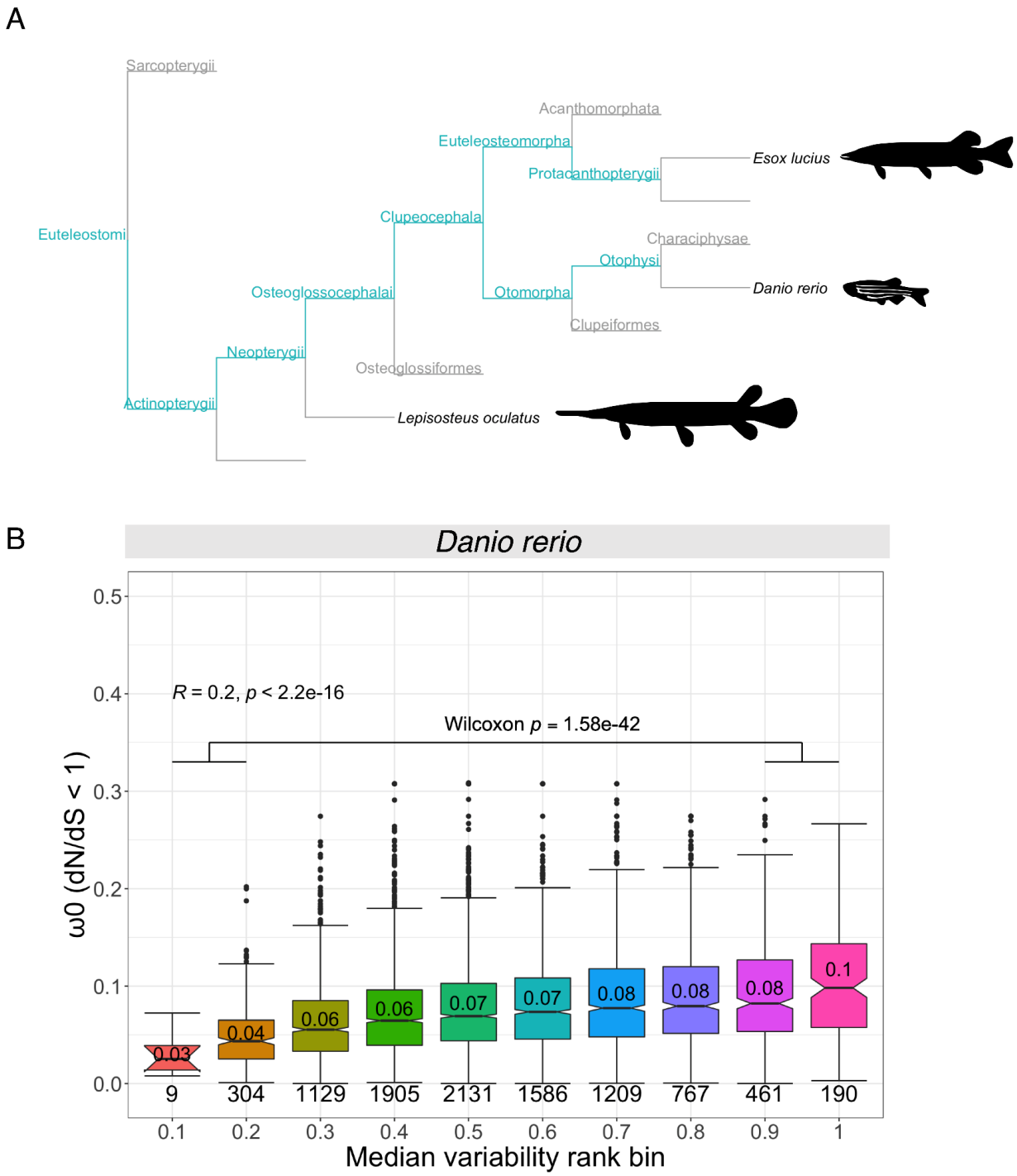


Figure 4. Purifying selection of zebrafish genes grouped by median expression variability across conditions. Selection on coding sequences computed from conservative branch-site likelihood tests were retrieved from the Selectome database **[58,59]**. **(A)** Species phylogeny with tested internal branches highlighted (turquoise). Tested branches included in all three species are: Euteleostomi, Actinopterygii, Neopterygii; for teleosts, i.e. zebrafish and pike, only: Osteoglossocephalai, Clupeocephala; for zebrafish only: Otomorpha, Otophysi; and for pike only: Euteleosteomorpha, Protacanthopterygii. Icons retrieved from PhyloPic: *Danio rerio* by Jake Warner (CC0 1.0), *Esox lucius* by Timothy Knepp and T. Michael Keesey (PDM 1.0), *Lepisosteus oculatus* by Ingo Braasch (CC0 1.0). Edge lengths are not scaled by divergence time. **(B)** Zebrafish protein-coding genes were grouped into 10 bins of size 0.1 ([0,0.1), (0.1,0.2], …, (0.8,0.9], (0.9,1.0]) based on their median variability rank across organ-sex conditions. Boxplots show the distribution of ratios of nonsynonymous substitutions (dN) to synonymous substitutions (dS) of sites that have evolved under purifying selection (ω_0 ratio, for dN/dS < 1). The median ω_0 is indicated within each boxplot and the number of genes included are annotated below each. *R* is the Spearman correlation coefficient between ω_0 and the unbinned median variability across conditions, and its corresponding *p*-value. The Wilcoxon *p*-value corresponds to a Wilcoxon rank-sum test comparing the least variable genes (variability rank ≤ 0.2) against the most variable genes (variability rank > 0.8). See **Figure S27** for equivalent analyses in Northern pike and spotted gar.

248 **Brain-biased genes have low expression variability across non-nervous organs**

249 We have shown that genes which are lowly variable across individuals in multiple conditions are enriched in
250 housekeeping functions, while highly variable genes are enriched in stimulus-sensitive functions (**Figure 3**).
251 Since housekeeping genes are, by definition, ubiquitously expressed across conditions (but see [60] for a
252 discussion), this suggests that the relationship works in both directions: genes expressed broadly across organs
253 should also show low within-organ variability across individuals, consistent with stronger pleiotropic
254 constraints. To test this hypothesis, we quantified expression breadth per gene using the expression specificity
255 metric tau (τ), with range [0,1], in which genes with τ close to 0 are uniformly expressed across organs whereas
256 genes with τ close to 1 are highly organ-specific [61,62] (Materials and Methods). Importantly, τ only considers
257 mean expression differences between conditions and does not take within-condition expression dispersion
258 into account, thus it is independent of our estimate of variability. Across conditions in each species, results
259 consistently show a positive association between variability rank and organ specificity (**Figures S28**-**S30**).
260 Based on the lower quartile of the τ distribution averaged across species, we designated genes with $\tau \leq 0.3$ as
261 'broadly expressed' and genes with $\tau > 0.3$ as 'organ-biased' (**Figure S31**). Note that we explicitly use the term
262 'organ-biased' rather than 'organ-specific' because our relaxed τ cutoff necessarily includes genes that are
263 preferentially expressed in one organ but are still expressed in other organs at lower levels. Per organ, broadly
264 expressed genes are generally underdispersed relative to organ-biased genes (**Figure 5**, **Figures S32-S34**,
265 **Figures S35-S37**, **Figures S38-S40**). Similar results were obtained with a more stringent τ cutoff ($\tau > 0.5$ as
266 organ-biased) (**Figures S41-S43**). These results are consistent with previous studies suggesting that expression
267 variability is associated with both tissue and cell-type-specific expression [49,63].

268 Since organ-biased genes have likely evolved under selection for an organ-specific function, one
269 prediction is that these genes would have a narrower expression range within their top (focal) organ of
270 expression, indicating more fine-tuned regulation for its putative primary function. On the contrary, we
271 observed significant overdispersion of organ-biased genes within their focal organ (**Figure 5A-C**, **5E-G**, **5I-K;**
272 **Figures S32-S34**, **Figures S35-S37**, **Figures S38-S40**). For example, genes that are most highly expressed in the
273 pectoral fin are overdispersed within the pectoral fin, relative both to broadly expressed genes as well as to
274 other organ-biased genes expressed in this organ (**Figure 5A, 5E, 5I**). This is consistent with a previous claim
275 that highly variable genes within a cell type population are typically indicative of cell type function [64]. This
276 pattern is lost when the top organ is assigned randomly, such that focal-biased genes have dispersion levels
277 that are indistinguishable from the background of other organ-biased genes (**Figures S45-S47; Figure S48**). We
278 infer that expression of organ-biased genes within their focal organ tends to be highly plastic in response to
279 genetic or environmental cues, and that context-dependent upregulation of expression of organ-biased genes
280 may be mediated by *trans*-acting factors that amplify variation.

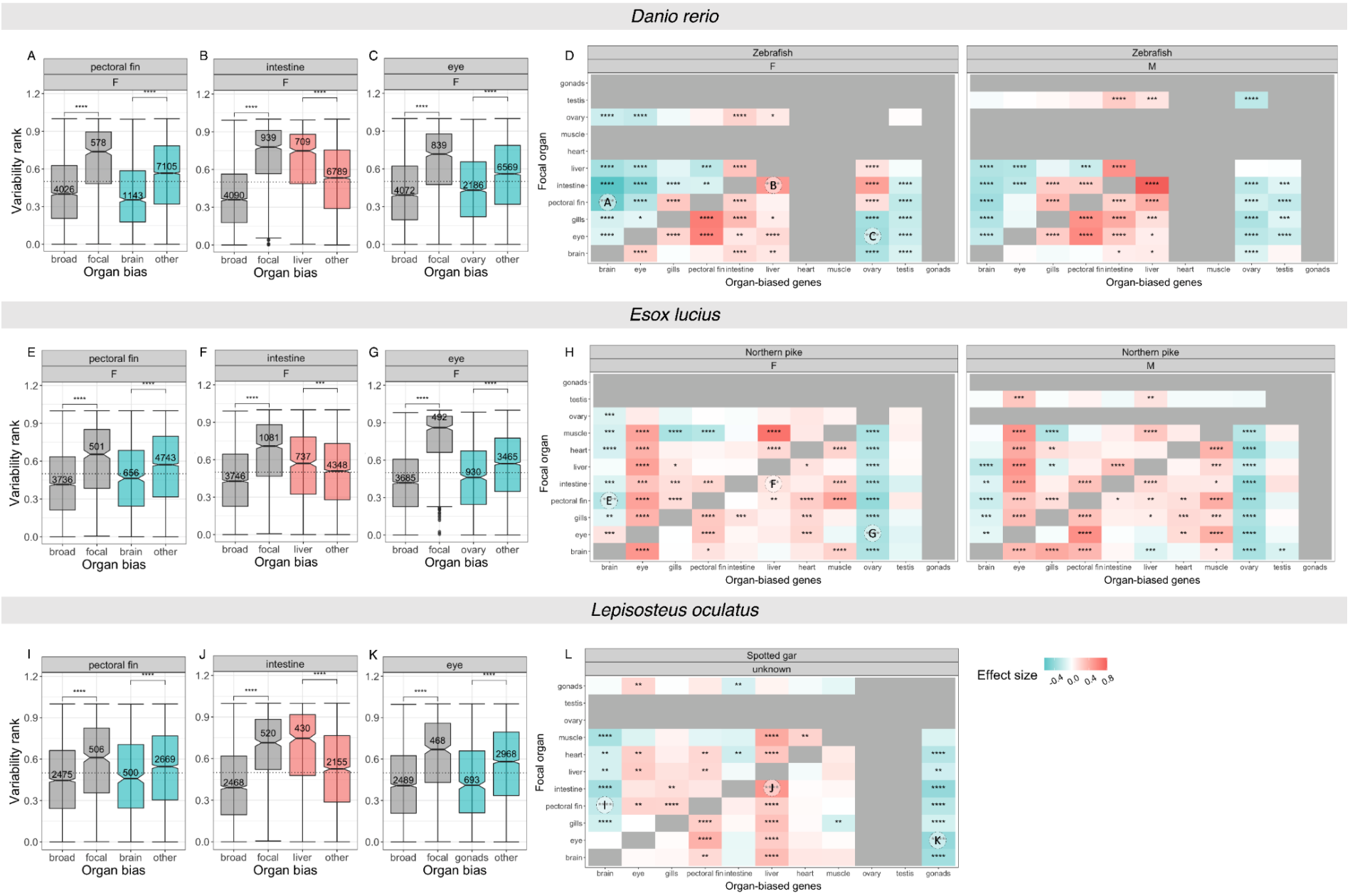
281 Patterns of expression dispersion across organs can reveal the extent to which the inter-individual
282 expression variability of a gene is independently regulated per condition [43]. On the one hand, if control of
283 variability is organ-specific, dispersion should differ across organ contexts. On the other hand, if variability is

284 controlled by a shared underlying mechanism, dispersion should appear consistent across organs. The second
285 scenario entails differences in the evolution of regulatory architecture depending on the top organ of a gene.
286 Interestingly, we observe a consistent pattern across organ contexts that is dependent on the focal organ. For
287 example, we find that brain-biased genes have consistently underdispersed expression across non-nervous
288 system organs relative to other organ-biased genes, regardless of differences in cell type composition per
289 organ (**Figure 5D, 5H, 5L; Figures S32-S34**). (See **Figures S41-S44** for the same analysis given a cutoff of $\tau >$
290 0.5.) By contrast, liver-biased genes, which evolve under weaker purifying selection at the protein-coding
291 sequence level (**Figure S49**) [65], tend to be overdispersed across multiple organs (**Figures S35-S37**), implying
292 weaker global expression constraints. Furthermore, we found that genes preferentially expressed in the
293 gonads, especially the ovary, tend to be underdispersed across organs (**Figure 5D, 5H, 5L**) **(Figures S38-S40).**
294 Taken together, these results are consistent with the expected pattern if gene-specific expression variability
295 is regulated by a shared mechanism across organ contexts [43] that is dependent on the top organ of a gene.
296 We observed qualitatively similar organ-dependent patterns of expression dispersion after controlling for
297 differences in transcriptome complexity (i.e., number of genes expressed) across organs [66–68] (**Figure S50**),
298 with particularly robust results in zebrafish (**Figure S51D**), although the effect is weaker in Northern pike
299 (**Figure S51H**) and spotted gar (**Figure S51L**). No pattern is recovered in any of the species when the top organ
300 is assigned randomly (**Figures S52**-**S54**), confirming that these results are biological and not due to biases in
301 data generation or processing. These results suggest that organ-biased genes have evolved particular *cis*-
302 regulatory mechanisms to facilitate either highly stable expression downregulation across other organs (for
303 brain- and gonads-biased genes) or more variable expression (for liver-biased genes). Indeed, we observed
304 that zebrafish promoter regions for each set of organ-biased genes can be distinguished from both matched
305 random promoter sequences (**Figure S55**) and from each other (**Figure S56**) (Methods). Promoter regions from
306 zebrafish brain-, ovary-, and testis-biased genes yield better model performance, consistent with stronger
307 promoter sequence conservation for lowly variable genes. By contrast, promoters regions from liver-, pectoral
308 fin-, gill-, and intestine-biased genes have somewhat lower model performance than sets of random
309 promoters, suggesting weaker sequence conservation (**Figure S57**).

310 Suboptimal expression can be more or less deleterious in different organs, due to differences in
311 physiological function, cellular composition and turnover, and sensitivity to perturbations [69]. In particular,
312 it has been suggested that neural tissues are highly sensitive to protein misfolding, and that this sensitivity
313 contributes to strong evolutionary constraints of genes expressed in the brain [70]. We confirm that protein-
314 coding sequences of brain-biased genes have evolved under stronger purifying selection not only compared
315 to other organ-biased genes, but also to broadly expressed genes **(Figure S42)**, which is consistent with reports
316 of strong negative selection on the mammalian brain relative to other organs [71–73,65,74]. We observed
317 moderate rates of purifying selection on both testis- and ovary-biased genes (**Figure S42**), though highly testis-
318 specific genes have been shown to be faster-evolving compared to other organ-specific genes in mammals
319 [65]. Overall, our findings show a generally positive association between inter-species protein sequence

divergence and intra-species expression variability (**Figure 4**), suggesting that evolutionary constraints at the protein sequence level co-vary with organ-dependent patterns of expression variability. One interesting exception involves eye-biased genes, which evolve under strong purifying selection across the three fish species (**Figure S49**) but which tend to have overdispersed expression in pike and spotted gar (**Figure 5H, 5L**). We suggest that the differential variability of eye-biased genes may be an interesting subject for future research, particularly given the remarkable diversity of visual opsin gene repertoires in fishes and its relation to differences in light environment and ecology [75].

From these results, we infer that (1) organ-biased genes are highly variable within their focal organ, likely due to context-sensitive expression, but that (2) their global variability reflects organ-dependent evolutionary constraints on both protein-coding and *cis*-regulatory sequences. We found that these organ-dependent patterns of expression variability are largely consistent in all three sampled species, thus constituting an evolutionarily conserved signal across at least 320 million years of evolution.


Danio rerio
Esox lucius
Lepisosteus oculatus
Variability rank
Organ bias
Focal organ
Organ-biased genes
Zebrafish
Northern pike
Spotted gar
Effect size

332

Figure 5. Relationship between expression variability and organ-biased expression for (A-D) zebrafish, (E-H) Northern pike, and (I-L) spotted gar. (A) Boxplot of expression variability ranks of genes expressed in female ('F') zebrafish pectoral fin, for sets of genes classified by organ bias and with an emphasis on brain-biased genes. Broadly expressed genes ('broad') have a measured organ expression specificity index (tau) of $\tau \leq 0.3$; genes with $\tau > 0.3$ are considered organ-biased. Among the set of organ-biased genes, genes are further classified into three categories: focal-biased genes ('focal') are preferentially expressed in the focal organ (i.e., pectoral fin), brain-biased genes ('brain') are preferentially expressed in the brain, and the remaining ('other') are preferentially expressed in other organs. The number of genes included are annotated for each category. Comparisons between broadly expressed and focal-biased genes, and between brain-biased and other organ-biased genes were performed using Wilcoxon rank-sum tests (*p*-value significance levels: '****': $p \leq 0.0001$, '***': $p \leq 0.001$, '**': $p \leq 0.01$, '*': $p \leq 0.05$, 'ns': $p > 0.05$). The dashed line indicates a variability rank of 0.5. **(B)** Boxplot of variability ranks of genes expressed in female zebrafish intestine, with genes classified by organ bias and with an emphasis on liver-biased genes. **(C)** Boxplot of variability ranks of genes expressed in female zebrafish eye, with genes classified by organ bias and with an emphasis on ovary-biased genes. **(D)** Heatmap summarizing pairwise comparisons of expression variability ranks between subsets of organ-biased genes for female ('F', left) and male ('M', right) zebrafish. Rows correspond to focal organ whereas columns correspond to subsets of organ-biased genes**.** Pairwise comparisons were performed using Wilcoxon rank-sum tests with Benjamini-Hochberg ('BH') correction for multiple testing (adjusted *p*-value < 0.05). Effect size was computed using Glass's delta (δ), with color scale shown in (L). Cells in gray are not included in the dataset. Three cells are annotated, corresponding to the following comparisons in boxplots (A-C): (A) brain-biased vs. other organ-biased genes expressed in female zebrafish pectoral fin, (B) liver-biased vs. other organ-biased biased genes expressed in female zebrafish intestine, and (C) ovary-biased vs. other organ-biased genes expressed in female zebrafish eye. A negative effect size entails lower expression variability relative to other organ-biased genes (A,C) (turquoise), whereas a positive effect size entails higher variability (B) (salmon). **(E-H)** Corresponding plots for Northern pike, with cells annotated corresponding to comparisons in boxplots (E-G). **(I-L)** Corresponding plots for unsexed spotted gar, with cells annotated corresponding to comparisons in boxplots (I-K).

Discussion

Until recently, most gene expression studies have focused on testing for differences in average expression levels between groups of interest, largely to the exclusion of expression variance and variability. Yet a more complex picture is now emerging: gene expression variability is increasingly implicated in disease [36–39] and drug efficacy [40], aging [41], circadian rhythm [76], cellular differentiation [32–34], embryogenesis [35,77], developmental canalization [47], and other biological processes. Here, we apply an evolutionary perspective based on a comparative analysis of inter-individual gene expression variability across multiple adult organs in three species of ray-finned fishes. We confirm that genes which are lowly variable among individuals are typically enriched in housekeeping functions, whereas highly variable genes are enriched in stimulus-sensitive functions, as reported previously in other species [25,42,43,63]. This biologically intuitive result reaffirms that expression variability is a functionally relevant trait independent of average expression. There is a growing body of literature that recognizes the functional relevance of expression variability [44–46] and how this variability is associated with promoter characteristics such as sequence, core promoter motifs, and transcriptional start site architecture [25,63,78]. However, there have been few attempts to characterize how gene expression variability in animals is systematically modulated by the interplay between gene function and organ context, though some studies have sampled diverse organs [40,43]. To our knowledge, this is one of the first studies to investigate how gene expression variability is influenced both by expression specificity and organ context.

Our cross-organ approach helps clarify the connection between two previously reported findings. First, although lowly and highly variable genes show distinct functional signatures, some studies report that gene expression variability tends to be tissue-specific (i.e., organ-specific) [43], with little overlap in the set of

377 most overdispersed genes between organs [41,42] or cell types [64] (but see [25]). Second, organ- or cell-type-
378 specific genes are found to have highly variable expression [49,63,64]. We verify that organ-biased genes are
379 more variably expressed relative to broadly expressed genes, but in addition, we have also found that organ-
380 biased genes are typically highly variable within their primary (focal) organ. Since the most variable genes
381 within an organ are typically genes which are biased or specific to it, we think it is unsurprising to find few
382 highly variable genes shared between multiple conditions. We infer that plastic expression within a focal organ
383 is a functional feature of more specialized genes, and that expression is dynamically induced by context-
384 specific regulatory cues, such as organ-specific transcription factor binding to enhancer elements or enhancer-
385 promoter interactions. Promoters of specialized genes (e.g., developmental genes) have been shown to
386 interact with different classes of enhancers compared to those of housekeeping genes [79]. In addition, such
387 specialized genes tend to have more enhancer-promoter interactions than housekeeping genes [28,80], and
388 the number of interacting enhancers has been found to have an additive effect on gene expression levels
389 [28,81,82]. Thus, to augment transcription of organ-biased genes in the relevant context, we expect an
390 increase in regulatory inputs, which introduces more sources of expression variation [11]. Since gene-specific
391 expression variability is influenced by a gene's regulatory inputs [11] and its position in the gene regulatory
392 network [83], we expect subsets of genes to have co-regulated patterns of variability, analogous to co-
393 regulation of gene expression levels. An analysis of the co-regulation of gene expression variability would
394 require even larger sample sizes in order to estimate variability across clusters of individuals (e.g., from
395 different populations).

396 Outside of their focal organ, organ-biased genes have basal expression levels that are relatively more
397 stable among individuals. We find striking evidence across species that the degree to which biased genes are
398 stably downregulated across non-focal organs reflects organ-dependent evolutionary constraints. Notably, we
399 observe consistently underdispersed expression of brain-biased genes across non-nervous organs, despite
400 these organs having very different transcriptional repertoires, cell-type compositions, and local micro-
401 environments. Low variability within a condition suggests stabilizing selection on expression, which is
402 consistent with strong purifying selection on *cis*-regulatory sequences, and/or with buffering of expression
403 variation by compensatory *cis-trans* effects [84]. Given that we observe consistent underdispersion across
404 multiple organs having very different local micro-environments, we reason that stable basal expression of
405 brain-biased genes is likely to be predominantly driven by strong purifying selection at the *cis*-regulatory level
406 rather than by context-dependent *trans* effects. Our finding that organ-biased genes have distinguishable
407 promoter sequences (**Figure S55**) further supports the view that inter-individual variability across organ
408 contexts is at least partly under *cis*-regulatory control. Moreover, the results support stronger selection on
409 promoters that regulate lower variability in non-focal organs.

410 We recognize that comparing gene expression variability between organs is challenging because any
411 measured changes in dispersion may simply be due to technical variation (e.g., sampling design, sequencing
412 depth or RNA quality [40]) or 'extrinsic' sources of biological variation (e.g., cell type composition [49,50],

413 transcriptome complexity [66,67]). Indeed, a limitation of bulk RNA-seq is that it does not allow us to
414 distinguish between multiple sources of variation within an organ, whether it is expression variation within a
415 cell type population, expression variation between cell type populations, or heterogeneity in cell type
416 composition between individuals. Observing a consistent signal of underdispersion across conditions with
417 heterogeneous backgrounds is thus particularly informative and suggests that our method captures a signal
418 of stable *cis*-mediated downregulation that is independent of organ-context. We infer that brain-biased genes
419 have evolved under strong purifying selection, not only at the protein-coding sequence level, but also at the
420 *cis*-regulatory level. From this, we hypothesize that genes selected for function in the brain may have co-
421 evolved with more stringent 'enhancer grammar' (e.g., number, type, order, affinity, spacing, or orientation
422 of transcription factor binding sites) [85,86] relative to other organ-biased genes. Our results highlight how
423 analysis of gene expression variability across organs can yield additional insights into gene regulatory
424 strategies in animals. Extending on these findings, we propose that intra-specific expression variability can
425 potentially be informative of broader macro-evolutionary trends, such as patterns of duplicate gene retention
426 and functional divergence after whole-genome duplication.
427

428 ## Materials and Methods

429 **Sample preparation**

430 *Sample collection*

431 Zebrafish

432 Thirteen female and twelve male zebrafish (*Danio rerio*), strain AB, were raised at the University of Oregon
433 Aquatic Animal Care Services. Animals were raised in E5 medium [87] and fed rotifers from 4-10dpf [88]. After
434 10dpf, fish were moved to a circulating fish water system and fed Zeigler Adult Zebrafish Diet. Three-month
435 old adults were euthanized by cold shock [87], and organs were dissected and immediately stored in RNAlater.
436 Procedures were approved by the University of Oregon IACUC protocol #18-13.
437

438 Northern pike

439 Ten female and ten male one-year-old Northern pike (*Esox lucius*) were sourced from a commercial hatchery
440 (Aqua 2B, Pouancé, France). The fish were euthanized using a lethal dose of tricaine methanesulfonate (MS
441 222, 400 mg/L), supplemented with 150 mg/L of sodium bicarbonate. Following euthanasia, organs were
442 immediately dissected and preserved in RNAlater. All animal procedures were conducted in full compliance
443 with French and European legislation (French decree 2013-118 and European Directive 2010-63-UE),
444 governing the ethical use and care of laboratory animals in scientific research.
445

446 Spotted gar

447 Spotted gars (*Lepisosteus oculatus*) were obtained as embryos from hormone-induced spawns of wild-caught
448 broodstock from bayous near Thibodaux, Louisiana and raised at Michigan State University for 23 or 31 months

449 (April 2018 - March/November 2020) in 150 to 300 gallon tanks on a 12h light : 12h dark cycle and a diet
450 starting from *Artemia* brine shrimp to feeder fish (mostly fathead minnows). Twenty-two (22) unsexed, pre-
451 mature individuals of 20.1 - 23.3 cm standard length (tip of snout to caudal peduncle) and 24.8 - 29.0cm total
452 length (tip of snout to tip of caudal fin) were euthanized in 300 mg/L MS-222 (Sigma). Tissues were dissected
453 and stored in RNAlater (Invitrogen) at -80°C until shipped frozen to the University of Lausanne. All spotted gar
454 experiments were approved by the Institutional Animal Care & Use Committee (IACUC) of Michigan State
455 University (animal use protocol PROTO201900309).
456
457 *RNA extraction*
458 Tissue samples were thawed on ice, removed from RNA*later* RNA Stabilization Solution (Invitrogen, catalog
459 no. AM7024) and transferred to new Eppendorf tubes containing 1 mL TRIzol Reagent (Invitrogen, ref.
460 15596018) and about 10 - 15 SiLi beads (Type ZS, 1.4 - 1.6 mm) (Sigmund Lindner, ref. 9315-33). Excluding
461 blood, samples were homogenized at 4°C using PreCellys Evolution (Bertin Technologies, ref. P000062 _PEV0)
462 for at least 2 cycles at 6600 rpm for 30 s. Phase separation of homogenized samples was performed following
463 the TRIzol Reagent user guide (MAN0001271). For Northern pike and spotted gar, RNA extraction with TRIzol
464 Reagent was carried out following the same manual. After washing and resuspending RNA, samples were
465 treated with RNase-free DNase I (Thermo Scientific, catalog no. EN0525) to remove DNA in accordance with
466 the user guide (MAN0012000). For zebrafish, column-based RNA extraction with TRIzol was performed using
467 the PureLink RNA Mini Kit (Invitrogen, ref. 12183018A), following manufacturer's instructions (MAN0000406).
468 DNA digestion during on-column RNA purification was done using the PureLink DNase Set (Invitrogen, ref.
469 12185010). All samples were eluted in 50 - 60 µl RNase-free water. Following extraction, RNA concentration
470 was quantified with the QuantiFluor RNA System (Promega, ref. E3310). A subset of samples was also assessed
471 for RNA quality using the 5200 Fragment Analyzer System (Agilent Technologies, M5310AA) with PROSize
472 3.0.1.6 data analysis software (Agilent Technologies). Samples with low RNA concentration or signs of
473 degradation were removed from subsequent steps. RNA preparations were diluted to a minimum
474 concentration of 50 ng/µl, transferred to low-binding 96-well plates, and stored in -80°C prior to sequencing.
475
476 **Data generation**
477 *Bulk RNA barcoding and sequencing (BRB-seq)*
478 Samples were sent to Alithea Genomics (Lausanne, CH) for library preparation and sequencing using highly
479 multiplexed 3'-end bulk RNA barcoding and sequencing (BRB-seq) [51] over three phases (corresponding to
480 projects AG0012, AMP0027, and AMP0020) (**Table S1**) (Supplementary Data). BRB-seq libraries were prepared
481 using the MERCURIUS BRB-seq library preparation kit for Illumina and following the manufacturer's manual
482 (Alithea Genomics, #18013). For project AG0012 (Northern pike, spotted gar), libraries were sequenced using
483 Illumina NovaSeq 6000, generating multiplexed, asymmetric paired-end reads, where read R1 (25 bp) contains
484 the sample barcode and UMI and read R2 (80 bp) contains the cDNA fragment sequence to be aligned to the

485 reference genome. For projects AMP0027 (spotted gar) and AMP0020 (zebrafish), libraries were paired-end
486 sequenced using Illumina NextSeq 550 (R1 = 21 bp, R2 = 55 bp).
487
488 *Read alignment and quantification*
489 Read trimming, alignment, and quantification steps were performed by Alithea Genomics using STARsolo
490 (preprint) [89] from STAR 2.7.9a [90]. R2 reads were trimmed ("--clipAdapterType CellRanger4") and aligned
491 to the following reference genomes: for spotted gar, LepOcu1 (Ensembl 103 [91]); zebrafish, GRCz11 (Ensembl
492 101 [91]); Northern pike, Eluc_v4 (Ensembl 101). To generate both raw and UMI-deduplicated counts, the
493 following parameter was used: "--soloUMIdedup NoDedup 1MM_Directional". UMI-deduplicated count
494 matrices were used for downstream analyses.
495
496 **Data analysis**
497 Excluding the promoter sequence analysis, data analyses were performed using R 4.1.0 [92], Bioconductor
498 3.13 [93], and RStudio 2023.03.0+386 [94] unless specified otherwise.
499
500 *Sample filtering*
501 Sample libraries were filtered in two main steps independently for each species. First, samples with low
502 sequencing quality were removed based on a minimum number of uniquely mapped reads or number of genes
503 with nonzero counts ("detected genes"), following Liu et al. [35]. For example, in zebrafish, samples with less
504 than 500,000 uniquely mapped reads or less than 10,000 detected genes were excluded from downstream
505 analysis. To account for differences in reference genome annotation quality, we selected species-specific
506 minimum uniquely mapped reads and detected genes. (See **Table S2** for filtering criteria.) For each species,
507 we visually inspected the plot of detected genes against uniquely mapped reads to estimate the point in which
508 the relationship between the two variables becomes nonlinear and set this point as the minimum threshold
509 (**Figure S1A**, **Figure S2A**, **Figure S3A**). We verified that samples which do not meet specified minimum criteria
510 are visible outliers on a principal component analysis (PCA) (**Figure S1B-C**, **Figure S2B-C**, **Figure S3B-C**).
511 Second, samples that were poorly correlated with other samples within its annotated organ were
512 removed. We performed iterative removal of lowly correlated samples based on a protocol described by
513 Fukushima & Pollock [95]. For each iteration, we first generated the all-against-all pairwise Pearson correlation
514 matrix of retained samples based on gene-filtered and normalized expression data in the form of log-
515 transformed, TMM-normalized counts per million ($\log_2$ TMM-CPM) (Gene filtering and normalization). Then,
516 for each sample, we computed its mean Pearson correlation with each organ. We removed any sample which
517 is more highly correlated with another organ than the one in which it is annotated to. This process was
518 repeated until no more samples fit removal criteria. Considering the set of protein-coding genes in each
519 species, we verified that all retained samples cluster by organ based on hierarchical clustering of the pairwise

520 Pearson correlation matrix (**Figure 1B, Figure S10**) and PCA of gene-filtered and normalized expression
521 (**Figures S7-S9**). Gene biotype information was retrieved for each species from Ensembl 105 [96].
522 For computing expression specificity (Organ expression specificity), we required a minimum of 1
523 replicate per organ-sex for Northern pike and zebrafish, and 2 replicates per organ for spotted gar (**Table S2**).
524 For estimating expression variability (Gene expression variability) and testing its association with gene function
525 (Gene Ontology (GO) statistical over- and under-representation analysis) and coding-sequence selection
526 (Selection analysis), we required a minimum of 4 replicates per organ (for spotted gar) or organ-sex condition.
527 Any conditions which did not meet the minimum number of replicates after quality filtering were excluded
528 from downstream analysis (**Table S3**). See **Tables S2-S3** for samples retained after each step of the sample
529 filtering pipeline and **Figure 1A** for a complete breakdown of the number of replicates retained per condition.
530
531 *Gene filtering and normalization*
532 At the start of each step of the sample filtering pipeline (Sample Filtering), we removed non-expressed genes
533 from the full raw count matrix by retaining only genes with both mean and median counts per million (CPM)
534 greater than 1.0 in at least one organ. From the gene-filtered raw count matrix, we performed library size
535 normalization over all conditions using the trimmed mean of M-values (TMM) method in *edgeR* (3.34.1) [97],
536 then re-converted counts to CPM and transformed CPM to $\log_2$-scale ($\log_2$ TMM-CPM). To remove negative
537 values, $\log_2$ TMM-CPM was vertically offset by the absolute value of the global expression minimum such that
538 the minimum value in the normalized expression matrix is always 0.
539
540 *Gene expression variability estimation*
541 To estimate gene expression variability within an organ (for spotted gar) or organ-sex condition (for Northern
542 pike and zebrafish), we considered only the set of conditions with at least 4 replicates after sample filtering.
543 Gene filtering and normalization steps were performed as described in the previous section, but with each
544 condition treated separately. As an additional filter to remove lowly expressed genes specific to a condition,
545 we removed genes with mean or median $\log_2$ TMM-CPM < 1.0. Since actual expression variation may be
546 sensitive to outliers, we used a jackknife resampling ('leave-one-out') procedure to compute expression
547 dispersion statistics and generate a less biased estimate of gene expression variability. That is, for a given
548 condition with n replicates, we formed all n possible subsets of n - 1 replicates and computed expression
549 dispersion metrics for each subset, then computed the mean for each statistic over all subsets. We initially
550 considered different measures of expression dispersion (Supplementary Methods), with the aim to correct for
551 the strong covariation between expression variation and average expression (**Figure S11**). From the metrics
552 we considered, we selected the local (residual) coefficient of variation (LCV) [29,40] as our estimate of
553 expression variability, as it maximally removes the association between mean expression level and expression
554 dispersion (**Figure 2C; Figures S12-S14**). The LCV was computed in two steps then averaged over all n - 1
555 subsets, as described below.

556

557 Residual coefficient of variation

558 The coefficient of variation (CV) is the ratio of the standard deviation to the mean. First, we computed the
559 squared CV and applied a $\log_2$ transformation ($\log_2(CV)^2$) for each gene, following Faure et al. [29]. Next, using
560 the 'loess' function from the *stats* package in R, we fit a LOESS regression curve (with span = 0.6 and default
561 parameters) to model $\log_2(CV)^2$ as a function of mean expression level. The difference between observed and
562 predicted $\log_2(CV)^2$ is the residual $\log_2(CV)^2$ ('residual CV', for brevity) after adjusting for expression level.
563 Genes with residual CV > 0 overdispersed, while genes genes with residual CV < 0 are underdispersed.

564

565 Local (residual) coefficient of variation (LCV)

566 Given residual CVs for each gene, we performed a percentile rank normalization step following the 'local
567 coefficient of variation' [40,47] or 'relative noise rank' [29] metric recommended in previous publications.
568 First, genes were sorted by mean expression level. Next, we defined a sliding window of 100 neighboring genes
569 and a step size of 1, with a focal gene at the center of each window. At each iteration of the sliding window,
570 the residual CV of the focal gene was ranked against the residual CVs of genes within the window, on a
571 percentile rank scale of 0.00 (lowest dispersion) to 1.00 (highest dispersion). We computed the LCV statistic
572 only for complete windows; as such, for each n - 1 subset, we do not have variability estimates for the 50 most
573 lowly and 50 most highly expressed genes.

574

575 Summarizing jackknife results

576 Since transcriptome-level expression distributions are positively skewed, with the most highly expressed genes
577 spanning a wide expression range, expression variability tends to be poorly estimated for outlier genes with
578 very high expression. As such, for each condition, we removed the top 5% most highly expressed genes based
579 on jackknifed mean expression level. Then, we computed the mean LCV over all n - 1 subsets and used this
580 metric, which we call 'variability rank', as our estimate of expression variability.

581

582 *Gene Ontology (GO) statistical over- and under-representation analysis*

583 For each organ-sex condition in zebrafish, we retrieved three sets of genes categorized by expression
584 variability: genes with variability rank ≥ 0.80 (highly variable), genes with rank ≤ 0.20 (lowly variable), and
585 genes with 0.40 ≤ rank ≤ 0.60 (moderately variable). For each variability category per condition, we used
586 *GOstats* (2.58.0) [56] to conduct conditional hypergeometric tests for statistical over- and under-
587 representation of GO biological process terms, based on the org.Dr.eg.db (3.13.0) zebrafish genome
588 annotation [98]. Per test, the background gene set was defined as the set of genes which have expression
589 variability estimates within the organ-sex condition. For each variability category, we retained GO terms that
590 were significantly over-represented (or under-represented) (unadjusted p-value < 0.01) in at least three

591 conditions. Redundancy reduction and visualization of GO terms was done using the Python package *GO-*
592 *Figure!* (1.0.0) [57] (**Figure 3, Figures S25-S26**).
593
594 *Coding-sequence selection analysis*
595 Given a list of protein-coding genes (Ensembl 105 [96]) and taxonomic branches for each species, we retrieved
596 precomputed selection statistics from Selectome (https://selectome.org/), a database of evidence for positive
597 selection based on a conservative branch-site likelihood test applied to *Euteleostomi* gene trees (Ensembl 98)
598 [58,59]. For each gene - internal branch (i.e., taxonomic clade), the Selectome dataset provides a measure of
599 selective pressure for sites under purifying selection (ω_0 ratio for sites with dN/dS < 1), as well as a likelihood-
600 ratio test statistic for positive selection. The tested internal branches for each species are as follows. For
601 spotted gar: Euteleostomi, Actinopterygii, Neopterygii. For zebrafish: Euteleostomi, Actinopterygii,
602 Neopterygii, Osteoglossocephalai, Clupeocephala, Otomorpha, Otophysi. For Northern pike: Euteleostomi,
603 Actinopterygii, Neopterygii, Osteoglossocephalai, Clupeocephala, Euteleosteomorpha, Protacanthopterygii
604 (**Figure 4A**). We included only distinct gene - internal branch entries in our downstream analysis, as duplicate
605 gene - internal branch entries may be a result of uncertainties in the gene tree. For each species, the
606 background gene set consisted of protein-coding genes that are included in the Selectome dataset and that
607 have expression variability data in at least half of the available conditions.
608 First, to test for an association between expression variability and purifying selection, we grouped
609 genes into 10 bins of size 0.1 based on their median variability rank computed across available conditions (per
610 species) and compared the ω_0 distribution across bins (**Figure 4**; **Figure S27**). Second, to check for an
611 association between variability and positive selection, we computed the proportion of gene - internal
612 branches with evidence for positive selection based on a *q*-value cutoff of $q < 0.05$ on the branch-site likelihood
613 test (**Table S4**). For this, genes that were frequently lowly, moderately, or highly variable (i.e., the same
614 classification in at least half of available conditions) were grouped into three large categories based on the
615 following thresholds: rank ≤ 0.20 for lowly variable genes, rank ≥ 0.80 for highly variable genes, and 0.20 <
616 rank < 0.80 for moderately variable genes. To test for significance, we conducted a randomization test by
617 permuting the *q*-values on the background gene list (n = 2000 permutations), in addition to a Pearson's chi-
618 squared test.
619
620 *Organ expression specificity*
621 Data preprocessing
622 To compute organ expression specificity within a species, we required at least 2 individuals per organ for
623 spotted gar, and at least 1 individual per organ-sex condition for Northern pike and zebrafish. Furthermore, to
624 make expression specificity measurements comparable between species, we considered only the set of organs
625 with sufficient samples in every species, i.e., brain, eye, intestine, liver, heart, muscle, gills, pectoral fin, and

gonads. After excluding organs with missing data (blood, kidney, skin, swim bladder), we repeated our gene filtering and normalization step starting from raw read counts (Gene filtering and normalization).

Expression specificity index

For each retained gene per species, organ expression specificity was measured using the tau (τ) index, which is defined as:

$$\tau = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(1-\frac{\bar{x}_i}{\bar{x}_{max}}\right)$$

where n is the number of organs, $\bar{x}_i$ is the mean expression of gene x in organ i, and $\bar{x}_{max}$ is the maximum mean expression of gene x within an organ $1 \leq i \leq n$ [61,62]. The tau index gives a single measure of organ specificity per gene, ranging from broadly expressed across all organs ($\tau \rightarrow 0$) to specifically expressed in a single organ ($\tau \rightarrow 1$). For species with differentiated sexes (Northern pike and zebrafish), the ovary and testis were considered separate anatomical entities. However, for non-reproductive organs in the same species, we did not consider each sex separately.

Organ-biased expression and gene expression variability

Per species, genes with tau (τ) ≤ 0.3 were classified as broadly expressed, while genes with tau above this threshold were classified as organ-biased. (See **Figures S41**-**S44** for an alternative cutoff of $\tau \leq 0.5$.) For genes which are organ-biased, the 'focal organ' is defined as the organ in which the gene is maximally expressed.

Multiple pairwise comparisons

For each species (spotted gar) or species-sex (Northern pike and zebrafish), let S_1 be the set of m organs which (a) have been used in expression specificity computations, and (b) have gene expression variability estimates (**Figure 1**):

$$S_1 = \{1, 2, \ldots, m\}.$$

Let S_2 be the set of $n \geq m$ organs which satisfy (a) but not (b):

$$S_2 = \{1, 2, \ldots, m, \ldots, n\}.$$

For each organ $k \in S_1$, we compared the variability rank distribution within k for broadly expressed genes ($\tau \leq 0.3$) versus genes which are organ-biased ($\tau > 0.3$) and with top expression in focal organ k. For example, given the pectoral fin as the focal organ, we compared the variability ranks of broadly expressed genes against pectoral-fin-biased genes (**Figure 5A, 5E, 5I**).

Next, still given k, we iterated through the set of other (non-focal) organs in S_1 which are not k:

$$\{i \in S_1 \mid i \neq k\}.$$

For each iteration of $i \neq k$, we defined a set of i-biased genes, which are organ-biased and with top expression in non-focal organ i, but still expressed in focal organ k. Then, we compared the variability rank distribution

660 within k for genes which are i-biased versus other organ-biased genes with top expression in the set J of (non-
661 focal) organs in S_2 which are neither k nor i:

662 $$J = \{j \in S_2 \mid j \neq k, j \neq i\}.$$

663 As an example, for expression variability in the pectoral fin, we compared the variability ranks of brain-biased
664 (or liver-biased, ovary-biased, etc.) genes expressed in the pectoral fin against the variability ranks of the set
665 of other organ-biased genes that are neither pectoral-fin-biased nor brain-biased (or liver-biased, ovary-
666 biased, etc.), and are also expressed in the pectoral fin (**Figure 5A, 5E, 5I**).

667 For each pairwise comparison of variability ranks of i-biased versus (other-biased) J-biased genes,
668 iterated for each organ $k \in S_1$, we tested for a difference in group means using Wilcoxon-rank sum tests, and
669 applied a Benjamini-Hochberg ('BH') correction for multiple testing. In addition, we calculated the effect size
670 using Glass's δ, using the set of J-biased genes as the reference group:

671 $$\delta_{k,i} = \frac{\bar{v}_{k,i} - \bar{v}_{k,J}}{s_{k,J}}$$

672 where $\bar{v}_{k,i}$ is the mean variability rank of i-biased genes in focal organ k, $\bar{v}_{k,J}$ is the mean variability rank of
673 J-biased genes, and $s_{k,J}$ is the sample standard deviation of the variability ranks of J-biased genes. We
674 visualized the distribution of effect sizes using a heatmap, where rows correspond to focal organ k and the
675 columns correspond to sets of i-biased genes (**Figure 5D, 5H, 5L**). We showed pairwise comparisons that are
676 considered statistically significant after BH multiple testing correction (adjusted *p*-value < 0.05).
677
678 <u>Analysis on randomized data</u>
679 Given the gene x organ mean expression matrix of a species, we permuted the values of the mean expression
680 vector per gene so that the focal organ of each gene is assigned randomly without changing tau. Since tau is
681 unchanged, the same set of genes are classified as broadly expressed ($\tau \leq 0.3$) and organ-biased, but the
682 subclassifications of organ-biased genes (e.g., brain-biased, liver-biased, etc.) differ. We pre-set two seeds
683 (12345, 67890) for our sample randomizations. Multiple pairwise comparisons of expression variability ranks
684 for subsets of organ-biased genes were performed as previously described (**Figure S48**).
685
686 <u>Analysis on nonzero expression matrix</u>
687 To control for differences in number of genes expressed among organs (**Figure S50**), we repeated the same
688 analysis (i.e., gene expression normalization, expression variability estimation, comparisons of expression
689 variability between subsets of genes) as previously described, considering only genes that have nonzero
690 counts across all samples within a species (**Figure S51**).
691
692

Organ-biased expression and coding-sequence selection

Per species, we retrieved expression specificity measurements for each gene (Organ expression specificity) and filtered selection statistics for each gene - internal branch (Coding-sequence selection analysis), as described in previous sections. We annotated each gene from the Selectome dataset with its top (focal) organ of expression (i.e., organ bias). If a gene has $\tau \leq 0.3$, its organ bias is labeled 'broad'. We compared the ω_0 distribution of genes across organ bias categories to check for any association between focal organ and strength of purifying selection at the coding-sequence level (**Figure S49**).

Promoter sequence analysis

To test whether organ-biased promoters can be distinguished from random sequences and from each other based on variability patterns, we trained gapped k-mer SVM models (gkmSVM) on promoter sequences. Analyses were first performed on each organ-biased promoter set, using subsamples of up to 1,000 genes to avoid sample-size bias **(Figure S55)**. Organ-biased promoter sets were also tested against promoters from the other organ-biased sets to evaluate whether they could be distinguished from one another **(Figure S49)**. We then grouped zebrafish organ-biased genes into two categories: those typically lowly variable in other organs (brain, ovary, testis, eye) and those highly variable in the others (liver, pectoral fin, gills, intestine), with up to 5,000 genes per set. In addition, three random gene sets were generated **(Figure S57)**. Promoter regions were defined as the 2-kb sequence upstream of the canonical TSS (Ensembl Biomart release 114) [99], and exonic regions were removed with bedtools subtract [100].

For each promoter set, a matched negative set of genomic sequences was drawn from the DanRer11 assembly, matched for length, repeat content, and GC content. This was done by building a BSgenome object with UCSC annotations [101] and sampling with the genNullSeqs function from the gkmSVM R package [102]. Models were trained with gkmtrain (LS-GKM) [103] using default parameters, and performance was assessed by 5-fold cross-validation using the -x option and receiver operating characteristic (ROC) curves.

Author Contributions

M.R.-R., I.B., Y.G., and J.H.P. conceived the experimental design. C.F.B. and M.R.-R. conceived the design of data analysis. A.W.T., B.L.R., and I.B. sourced and dissected the spotted gar samples; C.A.W. and J.H.P., the zebrafish samples; J.B. and Y.G., the Northern pike samples. C.A. performed RNA extraction. S.M. provided Selectome data for positive selection analysis. A.L. and A.T. performed the promoter sequence analysis. C.F.B. performed the data analysis and generated the figures with input from M.R.-R. C.F.B. wrote the original draft of the manuscript with input from M.R.-R; C.F.B., A.L., A.W.T., J.B., I.B., Y.G., J.H.P., and M.R.-R wrote the final version of the manuscript. C.F.B., A.L., A.W.T., J.B., I.B., Y.G., J.H.P., and M.R.-R. contributed to the interpretation and discussion of results.

Acknowledgements

We would like to thank Allyse Ferrara and Quenton Fontenot (Nicholls State University, USA) for providing gar embryos and members of the Braasch Lab for help with gar husbandry. We acknowledge having used ChatGPT (GPT-5) (OpenAI, 2025, https://chat.openai.com) solely to improve the clarity of the manuscript. Any errors or omissions are our own.

Funding

This work was supported by the Swiss National Science Foundation (SNSF, grants 173048 and 207853) to M.R.-R.; the Agence Nationale de la Recherche, France (ANR) on the GenoFish project, 2016-2021 (grant No. ANR-16-CE12-003) to J.B., J.H.P., M.R.-R., and Y.G.; the National Institute of Health (NIH) (award #R01OD011116) to J.H.P.; and the National Science Foundation (NSF) (award #202916) to I.B..

Competing Interests

The authors declare no competing interests.

Data Availability

Demultiplexed FASTQ files generated via BRB-seq for zebrafish (192 samples), Northern pike (252 samples), and spotted gar (177 samples) are available on the NCBI Sequence Read Archive (SRA) under accession number PRJNA1195359.

Data analysis scripts are available on GitHub: https://github.com/christabel-bucao/fish-variability-across-organs

Additionally, scripts and supplementary data files are available on Zenodo: https://doi.org/10.5281/zenodo.14063787

753 **References**

754 1. Kærn M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: From theories to phenotypes.
755 Nat Rev Genet. 2005;6: 451–464. doi:10.1038/nrg1615

756 2. Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. Nat Rev Genet. 2021;22: 235–250.
757 doi:10.1038/s41576-020-00300-0

758 3. Hill MS, Vande Zande P, Wittkopp PJ. Molecular and evolutionary processes generating variation in gene
759 expression. Nat Rev Genet. 2020. doi:10.1038/s41576-020-00304-w

760 4. Choi JK, Kim YJ. Epigenetic regulation and the variability of gene expression. Nat Genet. 2008;40: 141–
761 147. doi:10.1038/ng.2007.58

762 5. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid
763 responses to evolutionary adaptation. Nat Rev Genet. 2008;9: 583–593. doi:10.1038/nrg2398

764 6. Wagner GP, Altenberg L. Complex adaptations and the evolution of evolvability. Evolution. John Wiley &
765 Sons, Ltd; 1996. pp. 967–976. doi:10.1111/j.1558-5646.1996.tb02339.x

766 7. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297:
767 1183–1186. doi:10.1126/science.1070919

768 8. Silander OK, Nikolic N, Zaslaver A, Bren A, Kikoin I, Alon U, et al. A genome-wide analysis of promoter-
769 mediated phenotypic noise in Escherichia coli. PLoS Genet. 2012;8: 1002443.
770 doi:10.1371/journal.pgen.1002443

771 9. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene
772 expression. Science. 2014;346: 1533–1536. doi:10.1126/science.1255301

773 10. Wolf L, Silander OK, van Nimwegen E. Expression noise facilitates the evolution of gene regulation. eLife.
774 2015;4. doi:10.7554/eLife.05856

775 11. Urchueguía A, Galbusera L, Chauvin D, Bellement G, Julou T, van Nimwegen E. Genome-wide gene
776 expression noise in Escherichia coli is condition-dependent and determined by propagation of noise
777 through the regulatory network. Balaban N, editor. PLOS Biol. 2021;19: e3001491.
778 doi:10.1371/journal.pbio.3001491

779 12. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise minimization in eukaryotic gene expression.
780 PLoS Biol. 2004;2: e137. doi:10.1371/journal.pbio.0020137

781 13. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic
782 analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006;441: 840–846.
783 doi:10.1038/nature04785

784 14. Zhang Z, Qian W, Zhang J. Positive selection for elevated gene expression noise in yeast. Mol Syst Biol.
785 2009;5: 1–12. doi:10.1038/msb.2009.58

786 15. Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene
787 expression. Mol Syst Biol. 2008;4: 170. doi:10.1038/msb.2008.11

788 16. Lehner B. Conflict between noise and plasticity in yeast. PLoS Genet. 2010;6: e1001185.
789 doi:10.1371/journal.pgen.1001185

790 17. Bódi Z, Farkas Z, Nevozhay D, Kalapis D, Lázár V, Csörgő B, et al. Phenotypic heterogeneity promotes
791 adaptive evolution. PLoS Biol. 2017;15: e2000644. doi:10.1371/journal.pbio.2000644

792 18. Duveau F, Hodgins-Davis A, Metzger BPH, Yang B, Tryban S, Walker EA, et al. Fitness effects of altering
793 gene expression noise in saccharomyces cerevisiae. eLife. 2018;7. doi:10.7554/eLife.37272

794 19. Kittisopikul M, Süel GM. Biological role of noise encoded in a genetic network motif. Proc Natl Acad Sci.
795 2010;107: 13300–13305. doi:10.1073/pnas.1003975107

796 20. Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, et al. Interplay between gene
797 expression noise and regulatory network architecture. Trends Genet. 2012;28: 221–232.
798 doi:10.1016/j.tig.2012.01.006

799 21. Munsky B, Neuert G, van Oudenaarden A. Using Gene Expression Noise to Understand Gene Regulation.
800 Science. 2012;336: 183–187. doi:10.1126/science.1216379

801 22. Richard M, Yvert G. How does evolution tune biological noise? Front Genet. 2014;5: 374.
802 doi:10.3389/fgene.2014.00374

803 23. LaBoone PA, Assis R. Stress-Induced Constraint on Expression Noise of Essential Genes in E. coli. J Mol
804 Evol. 2024 [cited 21 Oct 2024]. doi:10.1007/s00239-024-10211-x

805 24. Dong D, Shao X, Deng N, Zhang Z. Gene expression variations are predictive for stochastic noise. Nucleic
806 Acids Res. 2011;39: 403–413. doi:10.1093/nar/gkq844

807 25. Sigalova OM, Shaeiri A, Forneris M, Furlong EE, Zaugg JB. Predictive features of gene expression
808 variation reveal mechanistic link with differential expression. Mol Syst Biol. 2020;16.
809 doi:10.15252/msb.20209539

810 26. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, et al. Common Regulatory
811 Variation Impacts Gene Expression in a Cell Type–Dependent Manner. Science. 2009;325: 1246–1250.
812 doi:10.1126/science.1174148

813 27. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philos Trans R Soc B Biol
814 Sci. 2013;368: 20120362. doi:10.1098/rstb.2012.0362

815 28. Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. Nat Rev
816 Genet. 2019;20: 437–455. doi:10.1038/s41576-019-0128-0

817 29. Faure AJ, Schmiedel JM, Lehner B. Systematic Analysis of the Determinants of Gene Expression Noise in
818 Embryonic Stem Cells. Cell Syst. 2017;5: 471-484.e4. doi:10.1016/j.cels.2017.10.003

819 30. Barroso GV, Puzovic N, Dutheil JY. The evolution of gene-specific transcriptional noise is driven by
820 selection at the pathway level. Genetics. 2018;208: 173–189. doi:10.1534/genetics.117.300467

821 31. Foreman R, Wollman R. Mammalian gene expression variability is explained by underlying cell state. Mol
822 Syst Biol. 2020;16. doi:10.15252/msb.20199146

823 32. Simon CS, Hadjantonakis AK, Schröter C. Making lineage decisions with biological noise: Lessons from
824 the early mouse embryo. Wiley Interdiscip Rev Dev Biol. 2018;7: e319. doi:10.1002/WDEV.319

825 33. Ohnishi Y, Huber W, Tsumura A, Kang M, Xenopoulos P, Kurimoto K, et al. Cell-to-cell expression
826 variability followed by signal reinforcement progressively segregates early mouse lineages. Nat Cell Biol.
827 2014;16: 27–37. doi:10.1038/ncb2881

828 34. Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, et al. Single-Cell-Based Analysis
829 Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a
830 Differentiation Process. PLOS Biol. 2016;14: e1002585. doi:10.1371/journal.pbio.1002585

831 35. Liu J, Frochaux M, Gardeux V, Deplancke B, Robinson-Rechavi M. Inter-embryo gene expression
832 variability recapitulates the hourglass pattern of evo-devo. BMC Biol. 2020;18: 1–12.
833 doi:10.1186/s12915-020-00842-z

834 36. Ho JWK, Stefani M, Dos Remedios CG, Charleston MA. Differential variability analysis of gene expression
835 and its application to human diseases. Bioinformatics. 2008;24: i390–i398.
836 doi:10.1093/bioinformatics/btn142

837 37. Li J, Liu Y, Kim TH, Min R, Zhang Z. Gene expression variability within and between human populations
838 and implications toward disease susceptibility. PLoS Comput Biol. 2010;6: e1000910.
839 doi:10.1371/journal.pcbi.1000910

840 38. Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, et al. Variance of Gene Expression
841 Identifies Altered Network Constraints in Neurological Disease. PLOS Genet. 2011;7: e1002207.
842 doi:10.1371/journal.pgen.1002207

843 39. Dinalankara W, Bravo HC. Gene Expression Signatures Based on Variability can Robustly Predict Tumor
844 Progression and Prognosis. Cancer Inform. 2015;14: CIN.S23862. doi:10.4137/CIN.S23862

845 40. Simonovsky E, Schuster R, Yeger-Lotem E. Large-scale analysis of human gene expression variability
846 associates highly variable drug targets with lower drug effectiveness and safety. Bioinformatics.
847 2019;35: 3028–3037. doi:10.1093/bioinformatics/btz023

848 41. Bashkeel N, Perkins TJ, Kærn M, Lee JM. Human gene expression variability and its dependence on
849 methylation and aging. BMC Genomics. 2019;20: 941. doi:10.1186/s12864-019-6308-7

850 42. Alemu EY, Carl JW, Bravo HC, Hannenhalli S. Determinants of expression variability. Nucleic Acids Res.
851 2014;42: 3503–3514. doi:10.1093/nar/gkt1364

852 43. Wolf S, Melo D, Garske KM, Pallares LF, Lea AJ, Ayroles JF. Characterizing the landscape of gene
853 expression variance in humans. PLOS Genet. 2023;19: e1010833. doi:10.1371/journal.pgen.1010833

854 44. Geiler-Samerotte K, Bauer C, Li S, Ziv N, Gresham D, Siegal M. The details in the distributions: why and
855 how to study phenotypic variability. Curr Opin Biotechnol. 2013;24: 752–759.
856 doi:10.1016/j.copbio.2013.03.010

857 45. De Jong TV, Moshkin YM, Guryev V. Gene expression variability: The other dimension in transcriptome
858 analysis. Physiol Genomics. 2019;51: 145–158. doi:10.1152/physiolgenomics.00128.2018

859 46. Mar JC. The rise of the distributions: why non-normality is important for understanding the
860 transcriptome and beyond. Biophys Rev. 2019;11: 89–94. doi:10.1007/s12551-018-0494-4

861 47. Horta-Lacueva QJ-B, Jónsson ZO, Thorholludottir DAV, Hallgrímsson B, Kapralova KH. Rapid and biased
862 evolution of canalization during adaptive divergence revealed by dominance in gene expression
863 variability during Arctic charr early development. Commun Biol. 2023;6: 1–12. doi:10.1038/s42003-023-
864 05264-5

865 48. Hagai T, Chen X, Miragaia RJ, Rostom R, Gomes T, Kunowska N, et al. Gene expression variability across
866 cells and species shapes innate immunity. Nature. 2018;563: 197–202. doi:10.1038/s41586-018-0657-2

867 49. Fair BJ, Blake LE, Sarkar A, Pavlovic BJ, Cuevas C, Gilad Y. Gene expression variability in human and
868 chimpanzee populations share common determinants. eLife. 2020;9: 1–26. doi:10.7554/eLife.59929

869 50. McCall MN, Illei PB, Halushka MK. Complex Sources of Variation in Tissue Expression Data: Analysis of
870 the GTEx Lung Transcriptome. Am J Hum Genet. 2016;99: 624–635. doi:10.1016/j.ajhg.2016.07.007

871 51. Alpern D, Gardeux V, Russeil J, Mangeat B, Meireles-Filho ACA, Breysse R, et al. BRB-seq: Ultra-
872 affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome
873 Biol. 2019;20: 1–15. doi:10.1186/s13059-019-1671-x

874 52. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Maucell E, et al. Genome duplication in the
875 teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 2004;431: 946–
876 957. doi:10.1038/nature03025

877 53. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. The spotted gar genome
878 illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48: 427–
879 437. doi:10.1038/ng.3526

880 54. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, et al. Noise in protein expression scales
881 with natural protein abundance. Nat Genet 2006 386. 2006;38: 636–643. doi:10.1038/ng1807

882 55. Eling N, Morgan MD, Marioni JC. Challenges in measuring and understanding biological noise. Nat Rev
883 Genet. 2019;20: 536–548. doi:10.1038/s41576-019-0130-6

884 56. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007;23:
885 257–258. doi:10.1093/bioinformatics/btl567

886 57. Reijnders MJMF, Waterhouse RM. Summary Visualizations of Gene Ontology Terms With GO-Figure!
887 Front Bioinforma. 2021;1: 6. doi:10.3389/fbinf.2021.638255

888 58. Proux E, Studer RA, Moretti S, Robinson-Rechavi M. Selectome: a database of positive selection. Nucleic
889 Acids Res. 2009;37: D404–D407. doi:10.1093/nar/gkn768

890 59. Moretti S, Laurenczy B, Gharib WH, Castella B, Kuzniar A, Schabauer H, et al. Selectome update: quality
891 control and computational improvements to a database of positive selection. Nucleic Acids Res.
892 2014;42. doi:10.1093/nar/gkt1065

893 60. Joshi CJ, Ke W, Drangowska-Way A, O'Rourke EJ, Lewis NE. What are housekeeping genes? PLOS Comput
894 Biol. 2022;18: e1010295. doi:10.1371/journal.pcbi.1010295

895 61. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange
896 transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics.
897 2005;21: 650–659. doi:10.1093/bioinformatics/bti042

898 62. Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics.
899 Brief Bioinform. 2017;18: 205–214. doi:10.1093/bib/bbw008

900 63. Einarsson H, Salvatore M, Vaagensø C, Alcaraz N, Bornholdt J, Rennie S, et al. Promoter sequence and
901 architecture determine expression variability and confer robustness to genetic variants. Eyras E, Perry
902 GH, editors. eLife. 2022;11: e80943. doi:10.7554/eLife.80943

903 64. Osorio D, Yu X, Zhong Y, Li G, Serpedin E, Huang JZ, et al. Single-Cell Expression Variability Implies Cell
904 Function. Cells. 2020;9: 14. doi:10.3390/cells9010014

905 65. Kryuchkova-Mostacci N, Robinson-Rechavi M. Tissue-specific evolution of protein coding genes in
906 human and mouse. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0131673

907 66. Ramsköld D, Wang ET, Burge CB, Sandberg R. An Abundance of Ubiquitously Expressed Genes Revealed
908 by Tissue Transcriptome Sequence Data. PLOS Comput Biol. 2009;5: e1000598.
909 doi:10.1371/journal.pcbi.1000598

910 67. Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, et al. Cellular Source and Mechanisms of

911 High Transcriptome Complexity in the Mammalian Testis. Cell Rep. 2013;3: 2179–2190.
912 doi:10.1016/j.celrep.2013.05.031

913 68. Darbellay F, Necsulea A. Comparative Transcriptomics Analyses across Species, Organs, and
914 Developmental Stages Reveal Functionally Constrained lncRNAs. Mol Biol Evol. 2020;37: 240–259.
915 doi:10.1093/molbev/msz212

916 69. Gu X, Su Z. Tissue-driven hypothesis of genomic evolution and sequence-expression correlations. Proc
917 Natl Acad Sci. 2007;104: 2779–2784. doi:10.1073/pnas.0610797104

918 70. Drummond DA, Wilke CO. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on
919 Coding-Sequence Evolution. Cell. 2008;134: 341–352. doi:10.1016/j.cell.2008.05.042

920 71. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, et al. Parallel Patterns of Evolution
921 in the Genomes and Transcriptomes of Humans and Chimpanzees. Science. 2005;309: 1850–1854.
922 doi:10.1126/science.1108296

923 72. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene
924 expression levels in mammalian organs. Nature. 2011;478: 343–348. doi:10.1038/nature10532

925 73. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. Nat Rev
926 Genet. 2014;15: 734–748. doi:10.1038/nrg3802

927 74. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, et al. Gene expression across
928 mammalian organ development. Nature. 2019;571: 505–509. doi:10.1038/s41586-019-1338-5

929 75. Musilova Z, Salzburger W, Cortesi F. The Visual Opsin Gene Repertoires of Teleost Fishes: Evolution,
930 Ecology, and Function. Annu Rev Cell Dev Biol. 2021;37: 441–468. doi:10.1146/annurev-cellbio-120219-
931 024915

932 76. Cortijo S, Aydin Z, Ahnert S, Locke JC. Widespread inter-individual gene expression variability in
933 Arabidopsis thaliana. Mol Syst Biol. 2019;15: e8591. doi:10.15252/msb.20188591

934 77. Hasegawa Y, Taylor D, Ovchinnikov DA, Wolvetang EJ, de Torrenté L, Mar JC. Variability of Gene
935 Expression Identifies Transcriptional Regulators of Early Human Embryonic Development. PLoS Genet.
936 2015;11: e1005428. doi:10.1371/journal.pgen.1005428

937 78. Schor IE, Degner JF, Harnett D, Cannavò E, Casale FP, Shim H, et al. Promoter shape varies across
938 populations and affects promoter evolution and expression noise. Nat Genet. 2017;49: 550–558.
939 doi:10.1038/ng.3791

940 79. Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, et al. Enhancer–core-promoter
941 specificity separates developmental and housekeeping gene regulation. Nature. 2015;518: 556–559.
942 doi:10.1038/nature13994

943 80. Furlong EEM, Levine M. Developmental enhancers and chromosome topology. Science. 2018;361: 1341–
944 1345. doi:10.1126/science.aau0320

945 81. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome
946 Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell.
947 2016;167: 1369-1384.e19. doi:10.1016/j.cell.2016.09.037

948 82. Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, et al. The
949 pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome
950 Res. 2015;25: 582–597. doi:10.1101/gr.185272.114

951 83. Puzović N, Madaan T, Dutheil JY. Being noisy in a crowd: Differential selective pressure on gene
952 expression noise in model gene regulatory networks. PLOS Comput Biol. 2023;19: e1010982.
953 doi:10.1371/journal.pcbi.1010982

954 84. Signor SA, Nuzhdin SV. The Evolution of Gene Expression in cis and trans. Trends Genet. 2018;34: 532–
955 544. doi:10.1016/j.tig.2018.03.007

956 85. Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development
957 and Evolution. Cell. 2016;167: 1170–1187. doi:10.1016/j.cell.2016.09.018

958 86. Jindal GA, Farley EK. Enhancer grammar in development, evolution, and disease: dependencies and
959 interplay. Dev Cell. 2021;56: 575–587. doi:10.1016/j.devcel.2021.02.016

960 87. Westerfield M. The zebrafish book; a guide for the laboratory use of zebrafish (Brachydanio rerio). 5th
961 Edition. Eugene: University of Oregon Press; 2007.

962 88. Best J, Adatto I, Cockington J, James A, Lawrence C. A Novel Method for Rearing First-Feeding Larval
963 Zebrafish: Polyculture with Type L Saltwater Rotifers (Brachionus plicatilis). Zebrafish. 2010;7: 289–295.
964 doi:10.1089/zeb.2010.0667

965 89. Kaminow B, Yunusov D, Dobin A. STARsolo: accurate, fast and versatile mapping/quantification of single-
966 cell and single-nucleus RNA-seq data. bioRxiv. 2021; 2021.05.05.442755.
967 doi:10.1101/2021.05.05.442755

968 90. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq
969 aligner. Bioinforma Oxf Engl. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635

970 91. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids
971 Res. 2020;48: D682–D688. doi:10.1093/nar/gkz966

972 92. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation
973 for Statistical Computing; 2021. p. https://www.R-project.org. Available: http://www.r-project.org

974 93. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput
975 genomic analysis with Bioconductor. Nat Methods. 2015;12: 115–121. doi:10.1038/nmeth.3252

976 94. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, PBC; 2021. p.
977 http://www.rstudio.com/. Available: http://www.rstudio.com/

978 95. Fukushima K, Pollock DD. Amalgamated cross-species transcriptomes reveal organ-specific propensity in
979 gene expression evolution. Nat Commun. 2020;11: 1–14. doi:10.1038/s41467-020-18090-8

980 96. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Ridwan Amode M, et al. Ensembl 2021. Nucleic
981 Acids Res. 2021;49: D884–D891. doi:10.1093/nar/gkaa942

982 97. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression
983 analysis of digital gene expression data. Bioinformatics. 2009;26: 139–140.
984 doi:10.1093/bioinformatics/btp616

985 98. Carlson M. org.Dr.eg.db: Genome wide annotation for Zebrafish. R package version 3.13.0. 2021.

986 99. Dyer SC, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Barrera-Enriquez VP, et al. Ensembl 2025.
987 Nucleic Acids Res. 2025;53: D948–D957. doi:10.1093/nar/gkae1071

988 100. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
989 Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

990 101. Pagès H. BSgenome: Software infrastructure for efficient representation of full genomes and their
991 SNPs. 2025. Available: doi:10.18129/B9.bioc.BSgenome

992 102. Ghandi M, Mohammad-Noori M, Ghareghani N, Lee D, Garraway L, Beer MA. gkmSVM: an R package
993 for gapped-kmer SVM. Bioinformatics. 2016;32: 2205–2207. doi:10.1093/bioinformatics/btw203

994 103. Lee D. LS-GKM: a new gkm-SVM for large-scale datasets. Bioinformatics. 2016;32: 2196–2198.
995 doi:10.1093/bioinformatics/btw142

996 104. Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR. The Bimodality Index: A Criterion for
997 Discovering and Ranking Bimodal Signatures from Cancer Gene Expression Profiling Data. Cancer Inform.
998 2009;7: 199–216.

999 105. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data.
1000 NAR Genomics Bioinforma. 2020;2. doi:10.1093/nargab/lqaa078