Detecting Signatures Of Selection On Gene Expression

Peter D Price[1*], Daniela H Palmer Droguett[1,2], Jessica A Taylor[1,3], Dong W Kim[4], Elsie S Place[5], Thea F Rogers[1], Judith E Mank[6,7,8], Christopher R Cooney[1+] & Alison E Wright[1+*]

[1]*Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, United Kingdom*

[2]*Ecology, Evolution, and Behavior Program, Michigan State University, USA*

[3]*Department of Biosciences, Durham University, United Kingdom*

[4]*Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, USA*

[5]*Development, Regeneration and Neurophysiology, School of Biosciences, University of Sheffield, United Kingdom*

[6]*Department of Zoology, University of British Columbia*

[7]*Beaty Biodiversity Research Centre, University of British Columbia*

[8]*Centre for Ecology and Conservation, University of Exeter, Penryn, UK*

*Corresponding authors: pprice3@sheffield.ac.uk, a.e.wright@sheffield.ac.uk

[+] These authors contributed equally

Keywords: gene expression, comparative phylogenetic methods, adaptive evolution, transcriptomics, complex phenotypes, stabilizing selection

ABSTRACT

A substantial amount of phenotypic diversity results from changes in gene expression levels and patterns. Understanding how the transcriptome evolves is therefore a key priority in identifying mechanisms of adaptive change. However, in contrast to powerful models of sequence evolution, we lack a consensus model of gene expression evolution. Furthermore, recent work has shown that many of the comparative approaches used to study gene expression are subject to biases that can lead to false signatures of selection. Here, we first outline the main approaches for describing expression evolution and their inherent biases. Next, we bridge the gap between the fields of phylogenetic comparative methods and transcriptomics to reinforce the main pitfalls of inferring selection on expression patterns and use simulation studies to show that shifts in tissue composition can heavily bias inferences of selection. We close by highlighting the multi-dimensional nature of transcriptional variation and identifying major, unanswered questions in disentangling how selection acts on the transcriptome.

INTRODUCTION

A growing body of evidence indicates that changes in patterns of gene expression play a key role in phenotypic divergence. Within species, a single genome can encode multiple distinct traits by varying expression levels of the underlying loci[1,2]. Similarly, across species, divergence in gene expression is implicated in major phenotypic differences that underlie adaptive change[3–7]. Given the importance of gene regulation in shaping phenotypic diversity, transcriptome analyses are widely used as a genomic tool to identify the genes that underlie phenotypic variation and the selective regimes acting on them[1,7]. However, the dominant mode of evolution acting on gene expression remains controversial. Current evidence supports the notion that global patterns of gene expression evolve predominantly under stabilizing selection but the extent of neutral evolution is heavily debated[8–11].

Much of this debate is driven by the lack of a consensus neutral model of transcriptome evolution. In contrast to established models of sequence evolution that allow us to scan coding sequence data for regions of adaptive evolution, gene expression can be complex and non-additive in its phenotypic effects. This complexity has resulted in a wide range of approaches to study the evolution of gene expression[7,12,13]. Importantly, these approaches make direct assumptions about how expression evolves across species, many of which have yet to be robustly validated, and these assumptions vary extensively across models. Over the last decade, statistical frameworks developed in the field of phylogenetic comparative methods have been applied to transcriptome data to infer selection[12,14], and these have provided important insights into patterns of expression divergence. However, in recent years it has become clear that several of these phylogenetic comparative approaches suffer from biases that often lead to false inferences of stabilizing selection when applied to real phenotypic data[15,16]. Many of the root causes of these biases are even more pronounced in transcriptomic data, but the issues uncovered in the phylogenetic comparative literature[15–17] are only rarely discussed in the genomics field[18,19].

Finally, most studies make the explicit assumption that when differential gene expression is observed, it is the direct result of regulatory change. In reality, this fundamental assumption may often be flawed as most studies of expression evolution measure transcript abundance in bulk across heterogeneous tissue samples and so cannot distinguish changes in gene expression from differences in tissue composition[20–22]. This problem undermines our current understanding of the nature and abundance of variation in gene expression across species, and how it contributes to phenotypic divergence. Although the implications of varying tissue composition across species for measuring differential expression have been discussed[20–22], the consequences of how it affects the inference of expression evolution have received less attention.

Here, we examine our current understanding of the evolutionary processes generating variation in gene expression. First, we outline the main approaches for describing gene expression evolution, examine their inherent biases, and synthesize findings to provide new perspectives to the debate over how selection acts on the transcriptome. Second, we attempt to bridge the gap between the fields of comparative phylogenetic methods and transcriptomics to reinforce the main pitfalls of inferring selection on expression levels. Importantly, we discuss the consequences of changes in tissue composition across taxa for the study of expression evolution, and use simulation studies to show that this issue can heavily bias inferences of selection. We close by highlighting the multi-dimensional nature of transcriptional variation and identifying major, unanswered questions in disentangling how selection acts on the transcriptome.

INFERRING THE MODE OF GENE EXPRESSION EVOLUTION

Currently, a number of different approaches for analysing expression evolution have been proposed in the absence of a single consensus model. These can be divided into three broad categories; (i) contrasts between divergence and variation in expression (Fig 1A), (ii) phylogenetic comparative methods (Fig 1B) and (iii) fitness-based approaches (Fig 1C). Importantly, each makes different assumptions regarding the mode of expression divergence and are subject to distinct biases. With a

few exceptions[18,19,23,24], studies rarely interrogate multiple approaches and so it remains unclear whether discrepancies between studies are biologically meaningful or caused by inherent methodological differences. Below we synthesise results from different analytical frameworks to provide an overview on the debate concerning the importance of selection versus genetic drift in shaping divergence in gene expression levels.

Contrasting divergence and variation in expression

Many early analyses of expression evolution tested for selection by contrasting expression divergence between species against diversity within species[5,25–28]. This method relies on the assumption that neutral changes are based solely on the underlying mutation rate[29,30] and so divergence between species relative to polymorphism within species will be equal at neutral loci[30]. When applied to expression data, mutation leads to polymorphism, which can be inferred through variation in expression level amongst individuals. Therefore, a neutral model of evolution can be rejected when there are deviations from a balanced ratio of within to between species expression variation (Fig. 1A). Studies employing this approach are dominated by two competing viewpoints. One posits that gene expression is predominantly neutrally evolving[13,25,26,31] and the other suggests widespread conservation and purifying selection of expression levels[27,28,32,33] with evidence of positive selection acting on certain loci[34–39].

Analogous approaches using alternative neutral models of expression divergence have also found broad support for stabilizing selection[7,10]. One such approach uses mutation accumulation studies to estimate neutral expectations of expression divergence and infer selection through contrasts with natural populations[40–42]. Most recently, the distribution of expression levels of F2 offspring from a genetic cross has been used to estimate expected levels of neutral change[43]. Here, under neutrality, expression variance of the two parental populations should be equal to the F2 progeny as F2 expression levels result from random combinations of segregating alleles. Following this logic, directional selection can be inferred when parental divergence is significantly greater than the neutral

125 expectation and stabilizing selection can be inferred when expression of parental populations is
126 significantly less diverged than expected. This study found widespread stabilizing selection on
127 expression level across a range of species, the magnitude of which was dependent on the species'
128 effective population size, consistent with population genetics theory that selection is more effective
129 in species with larger effective population sizes. Selection has also been inferred through comparisons
130 of additive genetic variance of expression (Q_{ST}) with sequence divergence in neutral molecular markers
131 (F_{ST}) across populations[44]. However, while Q_{ST}:F_{ST} approaches have been successfully applied to gene
132 expression variation in a few instances[45–49] accurately estimating the additive genetic basis of gene
133 expression level can be challenging[50] and there is a tendency for dominance variance to bias Q_{ST}
134 estimates, potentially leading to incorrect inferences of neutrality[44].

135 Nonetheless, the broad approach of contrasting inter- and intra-specific expression variation offers a
136 tractable method to investigate selective forces shaping expression levels. However, one drawback is
137 that these tests assume species or populations are phylogenetically independent and do not account
138 for shared and often complex evolutionary histories. Therefore, in cases where more than one pair of
139 species are compared, these methods can produce evolutionary patterns that are generated by the
140 structure of the underlying phylogeny[51,52]. Furthermore, the neutral expectation that expression
141 divergence equals diversity tends to break down over longer evolutionary time periods. This is because
142 gene expression divergence cannot accumulate indefinitely due to upper limits on the rate of
143 transcription. With increasing genetic distance, changes in expression among taxa may become
144 nonlinear, leading to instances of genetic drift being mistaken as directional selection[13,19]. To test for
145 selection across multiple species and evolutionary distances, approaches that take a phylogenetic
146 perspective are required.

147 ***Phylogenetic comparative methods***

148 Phylogenetic comparative methods have been widely adopted to infer selection acting on phenotypic
149 traits for a number of decades[14,52–54]. By incorporating phylogenetic information, these methods

account for shared ancestry and therefore can overcome issues of statistical non-independence. Recently, these approaches have been widely applied to transcriptome data to infer selection acting on gene expression by fitting a number of evolutionary models to expression data for a given gene[12,55,56] (Fig. 1B). A commonly used model, Brownian Motion (BM), assumes that expression divergence between species will be a function of divergence time and evolutionary rate ($\sigma 2$), and, as such, is often seen as analogous to genetic drift. A second model, the Ornstein-Uhlenbeck (OU) model, adds an 'elastic band' element (α) drawing expression values towards an optimum (θ) across the phylogeny, akin to stabilizing selection[12,57]. The OU model can be extended to allow for branch-specific events, such as shifts in optimum trait values[12,58], analogous to directional selection in particular lineages.

To date, comparative transcriptomic analyses have found overwhelming support for stabilizing selection on expression levels across a wide range of species, including *Drosophila*[12,59], African cichlids[60] and mammals[61]. While this appears consistent with past work[27,28,32,33], using OU models to infer selection has received repeated criticism within the phylogenetic comparative literature (BOX 1). In essence, any factor that leads to a reduction of phylogenetic signal in species' trait values will favour the inference of an OU process over BM, regardless of the underlying evolutionary process. Importantly, failing to account for biological intraspecific variance or methodological measurement error by running these models on a mean species expression value has been shown to erode phylogenetic signal and lead to false inferences of stabilizing selection[15,16,18] or branch-specific selection[19]. These issues are particularly relevant to expression data, which can be noisy (i.e. subject to a high degree of measurement error), particularly when environmental and developmental variance is not strictly controlled for. The OU framework has been adapted to specifically include within-species expression variability as an error term[18,58,62], and whilst it has been shown to reduce false inferences of stabilizing selection, this approach has only been employed by a handful of studies[24,63].

Recently, Rohlfs *et al*.[19] built on this approach with the Expression Variance and Evolution (EVE) model for testing expression evolution. This approach is grounded in the OU framework but incorporates

contrasts of expression variance within versus between species, analogous to divergence-diversity ratio comparisons (Fig. 1A). This is a major advance as it accounts for evolutionary relationships between species as well as incorporating a neutral expectation for expression divergence that is dataset-specific. Interestingly, the few studies that have employed this approach have typically revealed a higher proportion of genes evolving under directional than under stabilizing selection[19,24], and evidence for elevated rates of expression evolution consistent with adaptive evolution[63–66]. This contrasts with past evidence for stabilizing selection, outlined above, and may reveal the inherent biases of simpler OU models. However, it should be noted that the studies that employed EVE were primarily focused on contrasts of stabilizing versus directional selection not stabilizing versus neutral evolution and so do not explicitly rule out neutral processes. Finally, EVE also relies on accurately estimating parameters of the OU process, so it is still likely subject to similar pitfalls identified by the phylogenetic comparative literature (BOX 1).

Fitness-based approaches

Most recently, fitness-based approaches have been applied to study contemporary patterns of selection acting on gene expression[67,68]. One classical approach, which has been used to study a wide range of morphological traits, uses regression-based methods to estimate the strength of selection[69]. In this approach, the covariance between fitness and gene expression is calculated to infer selection differentials at each locus, which signify the mode of selection[67,68] (Fig. 1C). To reduce noise and computation time, as well as increase robustness of model prediction, expression data can be transformed to reduce dimensionality (i.e., by PCA) and selection gradients can then be obtained to estimate direct selection on suites of correlated transcripts. Recent studies have used these principles to measure selection on gene expression in experimental contexts (e.g. by quantifying flowering success and fecundity of rice grown in wet versus drought conditions[67]) and in natural settings (e.g. by measuring parasite load and survivorship of wild trout using mark-recapture[68]). In contrast to comparative approaches, neither of these studies found strong support for stabilizing selection, and in one case, the dominant mode of selection was disruptive[68]. Causes of this discrepancy require

202 further investigation, particularly whether or not this reflects methodological biases or difficulties in
203 accurately estimating fitness. However, it is possible that selection pressures vary over short- versus
204 long-term evolutionary time frames, and these approaches are capturing different snapshots of the
205 evolutionary process. Furthermore, unlike these fitness-based approaches, comparative phylogenetic
206 studies primarily rely on contrasting expression across highly-conserved orthologous genes, often
207 between very distantly related species, which likely biases our understanding of how gene expression
208 evolves. Gene duplicates are likely key to the evolution of tissue-specific expression patterns[70] and so
209 further work in this area might shed new light on how selection on gene expression varies across
210 genes.

211 **DECOMPOSING TRANSCRIPTIONAL VARIATION**

212 Approaches designed to test for selection on gene expression all make the explicit assumption that
213 differential expression is the direct result of regulatory change. However, in most cases, it is unclear
214 whether this assumption is valid as processes other than regulatory evolution can generate apparent
215 gene expression differences among taxa. For example, to date, studies have primarily used bulk
216 sequencing approaches to measure expression across aggregate tissues or even entire body regions,
217 which are often composed of many different cell types with variable expression profiles. In doing so,
218 these 'bulk' expression values represent an average of expression across entire populations of distinct
219 cell types. Here, we use existing single-cell expression data (scRNA-seq) for the developing chicken
220 hypothalamus[71] to illustrate this (Fig. 2A). The developing hypothalamus at Hamburger-Hamilton
221 stage 10 is composed of three major cell types, where the FOXA1 cell type represents the greatest
222 proportion of cells. Each cell type exhibits a distinct gene expression profile but average expression
223 estimated across all cells, analogous to a bulk RNA-seq approach for the whole hypothalamus, is not
224 reflective of genuine variation in gene expression. The magnitude of this effect varies across genes,
225 consistent with recent work in the mouse gonad[21] and primate heart tissue[22].

Within species, dramatic changes in tissue composition are well documented throughout development[71–73] and between the sexes[72]. This is exemplified by gonadal tissue, which exhibits sex-specific cell types[72] as well as a mix of haploid and diploid cells at various stages of differentiation[73–76]. Similarly, changes in cell type abundance between homologous tissues are common across species, particularly in the brain[77–79] and testes[21,80,81], the latter likely as a result of varying levels of sperm competition and sexual selection. For instance, New World Blackbird species under more intense sperm competition exhibit a greater proportion of sperm-producing tissue in the testes[80]. Importantly, this means that samples that vary in tissue composition can produce patterns of differential expression that are often mistaken as evidence of changes in gene regulation. Conversely, this approach can also dampen or mask genuine differences in expression within or between populations and species[20–22]. Of course, changes in tissue composition, which encompass both changes in cell type abundance within tissues and allometric scaling across them, are likely due to changes in gene expression across development. However, these changes in expression will not be detected if transcriptomes are measured after development is completed. Instead, the resulting differences in gene abundance will be mistaken as causative adaptive changes (Fig 2B).

To our knowledge, only a handful of studies have directly accounted for the consequences of varying tissue allometry when studying modes of expression evolution[20–22,82,83]. Addressing this is a major priority for the field. Recent advances in single-cell transcriptomics permit direct comparisons of expression across homologous cell types in a comparative framework and so overcome issues of tissue composition variation. However, scRNA-seq currently presents its own set of challenges both in terms of expense and robustly identifying homologous cell types across species[84] (BOX 2). Importantly, several fields, including cancer and developmental biology, have developed methods to deconvolve expression data from complex tissues, and these are likely to be extremely valuable to evolutionary genomics studies. We urge future studies to carefully consider these points in project design (BOX 2).

CHALLENGES OF INFERRING SELECTION

While the implications of varying tissue allometry for measuring gene expression change across species have been discussed[20–22] (Fig. 2), the consequences of tissue composition on inferences of expression evolution have received less attention. Most studies that test for selection on the transcriptome use expression data generated from heterogeneous tissue, with the exception of recent work that used cell sorting to isolate distinct cell types in the mouse testes[66]. As discussed, there is a tendency for phylogenetic comparative methods to falsely infer stabilizing selection or more complex adaptive processes if non-evolutionary processes (such as measurement error) reduce phylogenetic signal. Perceived changes in expression that are driven by variation in tissue composition across species represent a prominent source of non-evolutionary expression variance and could therefore bias inferences of selection. This possibility has yet to be formally examined and so, using a series of simulated scenarios, we directly explore how shifts in tissue composition can bias the inference of evolutionary processes in a phylogenetic framework.

We simulated three distinct scenarios to explore how asymmetry in tissue composition across a phylogeny can drive false model inferences of expression evolution when applying comparative methods (Fig. 3). We imagine a simple situation where a tissue is composed of two distinct cell types. We estimate bulk expression values as a function of expression level in each cell type and their relative abundances in the tissue, and fit a set of evolutionary models to this bulk expression.

First, we describe a scenario of extreme stabilizing selection on gene expression of a single locus. This locus is highly expressed in one cell type and lowly expressed in the other, but importantly, expression values are identical (i.e. not evolving) across species. However, the relative abundance of each cell type is evolving under genetic drift and so varies across species (Fig. 3A, scenario i). As predicted, the composite expression value is not reflective of single-cell expression levels nor consistent with extreme stabilizing selection (Fig. 3B, scenario i). A phylogenetic comparative approach consistently rejects a ‘static’ model of expression evolution and finds the greatest support for genetic drift as the

dominant mode of evolution (Fig. 3C, scenario i). In this instance, the false positive rate is around ~86% relative to when these models are run on single-cell expression levels. This suggests that shifts in tissue composition can lead to false inferences of evolutionary processes acting on gene expression in the complete absence of any change in expression level within each cell type.

Second, we assume that gene expression is evolving under genetic drift. The two cell types are of equal abundance in all species with one exception in which a lineage-specific change in cellular composition occurs so that one cell type dominates (Fig. 3A, scenario ii). After model fitting, we find that this type of composition shift in one lineage leads to false inferences of a shift in gene expression, consistent with adaptive evolution (Fig. 3B & C, scenario ii). The scale of this bias is highly dependent on the size of the allometric shift (Fig. 4A). Where the shift leads to a single cell type dominating, the actual mode of evolution (i.e. genetic drift), will be rejected in ~35% of instances. While this extreme situation is arguably biologically unrealistic, our simulations show that even marginal shifts in relative cell type proportion result in elevated type 1 error rates. For example, across New World Blackbirds, the proportion of seminiferous tissue in the testes ranges from 87% to 96%[80]. This equates to a shift in the proportion of ~9%. Even though our simulations use different starting conditions, it is clear that shifts of a similar magnitude can result in increased type 1 errors.

Finally, we simulated a scenario where gene expression and cell type abundance both evolve under genetic drift (Fig. 3, scenario iii). Here, we recover the true signal of genetic drift more reliably (Fig. 3C, scenario iii). However, in all instances so far, we have assumed that gene expression at a single locus is evolving independently in each cell type. While this is likely a reasonable assumption for some loci that have evolved tissue- or cell-specific regulatory machinery[85,86], expression changes are probably correlated in many instances. Interestingly, we find that this has implications for which evolutionary processes are inferred (Fig. 4B). When tissue composition evolves across the phylogeny, the type 1 error rate is highly dependent on the level of expression covariance between the cellular components of that tissue. In particular, if expression across cell types negatively covaries, where an increase in expression in one cell type is associated with a decrease in expression in another cell type at a single

locus, the type 1 error rate can exceed 40% (Fig. 4B). The extent to which gene regulation is decoupled across cell types is, in and of itself, an interesting question. But here we have shown that gene expression covariation across cell types can also have profound implications for how we infer which selective processes are operating.

These scenarios demonstrate the potential challenges of inferring selection on expression level using data from heterogeneous tissues. It is also worth noting that our simulations are conservative as we do not model other non-evolutionary sources of variation (such as measurement error and tree topology error) that are likely to be common in transcriptome studies. We believe this highlights an urgent need to reappraise our current understanding of expression evolution in light of these underlying methodological issues. In particular, establishing (i) how often and by what magnitude changes in tissue composition occur and (ii) the extent to which transcriptional variation is correlated across cell types are important factors to consider when studying expression evolution using phylogenetic comparative approaches with bulk RNA-seq. Unfortunately, we are not aware of a simple solution for correcting the biases we have uncovered, beyond recommending the use of single-cell data to study expression evolution where possible. However, while single-cell approaches are increasingly available, the technical demands of this approach mean that they remain challenging for many species. In the meantime, we urge caution when using phylogenetic comparative approaches with bulk RNA-seq and recommend some steps to minimise other sources of error (BOX 2).

FUTURE DIRECTIONS

Given the importance of changes in gene regulation to phenotypic divergence, studying transcriptome evolution is key to understanding adaptive change. As we discussed, we currently lack a consensus neutral model of transcriptome evolution and it is debatable whether we expect this to be universal across all loci due to the complex transcriptional architecture of many phenotypes. Here, we argue that our understanding of the evolution of gene expression will permit critical advances as we continue to link insights across layers of the genotype-to-phenotype map, developmental contexts, and

evolutionary timescales, with organismal ecology as our foundation. Below we identify major, unanswered questions in disentangling how selection acts on the entire transcriptome. We note that a complete understanding of how the transcriptome evolves also requires detailed knowledge of how regulatory elements combine to facilitate expression change and how selection acts on these non-coding regions[87,88], recently discussed elsewhere[7].

Transcriptional diversity and layers of gene regulation

Variation in splicing, whereby the same gene can express different RNA variants that produce distinct proteins or isoforms, are a common source of transcriptional variation across species[33,89–91] with important phenotypic effects (recently reviewed[92,93]). For genes with constraints on expression levels (e.g. because of pleiotropic effects) alternative splicing may act as another adaptive mechanism of gene regulation[94]. Long-read sequencing methods have the advantage of producing full-length transcript sequences[95], which can be a more reliable way to identify alternatively spliced variants in transcriptomic datasets. Understanding the evolution of gene regulation will ultimately require an integrated understanding of how and when differences in expression level and splicing contribute to phenotypes under selection.

For transcriptional variation – whether in terms of expression level or alternative splicing – to be selected upon, it must contribute to variation at the protein layer of the genotype-to-phenotype map. Due to difficulties in assaying proteins in comparison to RNA, the links between transcription and translation are underexplored, particularly in non-model organisms. Recent methodological advances that measure rates of protein synthesis to assay the translatome show that protein expression levels evolve under stronger evolutionary constraint than transcript levels[96], and report a higher correlation between the translatome and proteome than between the transcriptome and proteome[97]. However, this effect tends to decrease for functionally relevant loci, such as differentially expressed genes[98]. This indicates that in many cases, mRNA abundance does not fully capture transcriptional variation,

and more work is needed to understand the complex relationship between transcription and translation (e.g. mechanisms of buffering, feedback, degradation)[9,99].

Regulatory and co-expression networks

The intrinsically correlated nature of gene expression means that identifying selection at a single locus is hard to disentangle from the expression patterns at loci with shared architectures. To account for this, we must either take on network-based approaches and try to account for connectivity or covariance between loci, or we must reduce the dimensionality of our data. Furthermore, recent work identifying key nodes in gene regulatory networks of health and disease phenotypes between sexes also established that genes that appear architecturally central to a phenotype may also not appear differentially expressed[100]. Similarly, GWAS studies have revealed that complex phenotypes are often the product of many different loci where regulatory networks are likely highly interconnected and heritability is distributed across the entire genome[101–104]. Together, this means that studying expression on a locus by locus basis and not through inter-locus interactions may limit our ability to understand the transcriptional architectures underlying adaptive phenotypes, and how this impacts the mode and strength of selection on gene expression[101].

Developmental context

Phenotypic variation is produced by dynamic developmental changes through space and time. While gene regulation is highly context-dependent in terms of tissue identity and developmental stage[105–107], studies primarily test for expression evolution in a single snapshot, most often in adult tissues. Single-cell transcriptomic methods offer a promising path to better understanding how these sources of variation interface with gene expression through development and inform models of gene expression evolution.

Genotype to phenotype to adaptation

If our goal is to uncover how gene regulation underlies adaptation, we must link transcriptional variation with organismal ecology and natural history. This effort is twofold, as it requires

understanding when and how selection acts on organisms, and how transcriptional variation contributes to phenotypic responses to selection. Methods of surveying variation in gene expression offer increasing precision and resolution. However, our ability to identify the evolutionary processes causing this variation ultimately depends on our understanding of the organisms in question. Model systems like yeast continue to enable high-throughput analyses that have yielded pivotal insights into the evolution of the transcriptome[7,108–111], but non-model systems also hold promise for studying how gene expression evolves under natural settings which may yield novel and more ecologically relevant findings[68,112]. Furthermore, it remains to be seen how results from microevolutionary studies within or across a single generation integrate with those from macroevolutionary studies comparing diverged lineages, and the relative roles of stabilizing versus directional selection across these scales.

METHODS

Single-cell transcriptomics

We analysed existing single-cell expression data (scRNA-seq) for the developing chicken hypothalamus[71]. Cell types expressing 'PAX6', 'FOXA1' or 'SIX6' at Hamburger-Hamilton (HH)10 were used in this study. Methods to identify cell types and estimate expression levels are published previously[71]. Pseudo-bulk datasets were generated at HH10 by calculating the average expression across cells in the 3 cell types.

Simulations

For the first scenario (Fig 3Ai), expression values were set at one and two in two cell types (A,B) respectively. The relative proportion of each cell type (pr) was simulated under Brownian Motion (BM) for 1000 unique trees of either 25 or 100 tips, using fastBM from phytools[113] in R v4.1.1. The resultant values were normalised between 0 and 1. Composite expression values for each tip (i) were calculated as follows: $(eq.1)\ exp_i = (pr_i \times 1) + \big((1 - pr_i) \times 2\big)$.

For the second scenario (Fig 3Aii), expression values were evolved under BM over 1000 unique phylogenies of 25 and 100 tips. The relative proportion of each cell type (pr) was set to 0.5 across the phylogeny except for one randomly-chosen tip (t). For this tip, the relative abundance of one cell type was shifted to an alternate value within the range 0 to 1 in 0.05 increments (Fig 4A). The composite expression value of the shifted tip (t) was calculated as follows: $(eq.2)\ exp_t = (pr_t \times exp_{At}) + \big((1 - pr_t) \times exp_{Bt}\big)$. Expression for the other tips (i) was calculated as above using equation (1).

For the third scenario (Fig 3Aiii), expression values were evolved under BM with varying covariances between cell types A and B, with covariance values varying from -1 to 1 in increments of 0.05. In all cases, σ^2 was set to 1.0001 and we examined trees of 25 and 100 tips. Simulations for each covariance value were run 1000 times on unique trees. This scenario was run with both a fixed cell type proportion (pr), where pr = 0.5 in A and B at all tips, and with proportion values evolving under BM (Fig 4B). The composite expression value at each tip (i) was calculated as follows: $(eq.3)\ exp_i = (pr_i \times exp_{Ai}) + \big((1 - pr_i) \times exp_{Bi}\big)$.

Fitting evolutionary models to composite expression levels

We fit evolutionary models in R using phylolm[114] for scenario one (Fig 3Ai) and OUwie[115] for scenario two (Fig 3Aii) and three (Fig 3Aiii). For the first scenario (Fig 3Ai), a static evolutionary model was rejected if the 95% bootstrapped confidence interval for σ^2 crossed 0. If rejected, a BM, an OU (Ornstein-Uhlenbeck), and a WN (White Noise) model were fit and their Akaike weights calculated. The WN model was fitted by suppressing phylogenetic signal by fixing Pagel's λ to 0. We calculated the type 1 error rate for scenario one (Fig 3Ai) as the rate at which a non-static model was accepted in favour of the static model, relative to when the same set of models were applied to a single-cell type simulation. For scenarios two (Fig 3Aii) and three (Fig 3Aiii), we fit a BM, an OU and an OU-shift model, where in the latter the optimum value of the trait is allowed to vary on a single tip. For scenario two (Fig 3Aii), the OU-shift model was fit so that the tip with the proportion shift was allowed the alternate optima, whereas for scenario three (Fig 3Aiii), a random tip was allocated. For scenario two

(Fig 3Aii) and three (Fig 3Aiii), the type 1 error rate was calculated as the rate at which a non-BM model was favoured (i.e. where ΔAICc > 2) relative to BM, relative to when the same set of models were applied to a single-cell type simulation.

ACKNOWLEDGEMENTS

This work was funded by a NERC Independent Research Fellowship to CRC (NE/T01105X/1), a NERC Independent Research Fellowship to AEW (NE/N013948/1), a grant from the European Research Council (grant agreement 680951) and a Canada 150 Research Chair to JEM, a NERC ACCE DTP to PDP, and an NSF Postdoctoral Research Fellowship and MSU Presidential Postdoctoral Fellowship to DHPD. We thank Marysia Placzek, Paula Escuer Pifarré, Emily Josephs, Adrian Platts, Miles Roberts, Rebecca Panko, Maya Wilson Brown, Sophie Buysse and four anonymous reviewers for helpful comments and suggestions on the manuscript.

AUTHOR CONTRIBUTIONS

AEW, CRC, DHPD, PDP and JEM designed the review. DWK, ESP, AEW, CRC and PDP analyzed the data. All authors wrote and edited the manuscript.

DATA AVAILABILITY

All data has been published previously[71].

CODE AVAILABILITY

All code is publicly available at (https://github.com/Wright-lab-2021-Transcriptome-Evo/Inferring_expression_evolution_review).

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

BOX 1: Common pitfalls of inferring selection using Ornstein-Uhlenbeck models

Recent work from the phylogenetic comparative methods field has revealed inherent biases in estimating OU processes, often leading to false inferences of stabilizing selection. As these have already been discussed elsewhere[15–17], we summarise the main pitfalls in relation to transcriptome studies.

Small phylogenetic samples

Recent work has shown that the ability to accurately estimate parameters of the OU model is strongly influenced by the number of species. Cooper *et al*[15] simulated a range of phylogenies of varying size under Brownian Motion and compared the fit of BM and OU models to test how often stabilizing selection was falsely inferred. They found a high type 1 error rate, especially when the number of sampled taxa was limited. For example, with a phylogeny of 25 species, stabilizing selection was falsely inferred ~10% of the time. This is especially concerning for transcriptomic studies, which are frequently comprised of far fewer species due to sampling and computational costs and employ thousands of model comparisons in order to infer selection at each orthologous locus separately. We do anticipate this concern will lessen as expression data becomes available for more species. However, even with phylogenies of 100 species, Cooper *et al*[15] still estimate a type 1 error rate > 0.05.

Measurement error

Here, we use the term measurement error to broadly refer to any factor that adds noise to heritable expression values. This includes (i) data quality problems, such as RNA degradation, sequencing and assembly issues, (ii) low sample sizes and (iii) unwanted biological variance arising from the failure to control for environmental variation across samples. Measurement error across lineages can erode phylogenetic signal in the data, falsely biasing model selection away from BM models and towards OU processes and the inference of stabilizing selection[15,16]. Recent work has shown that even small amounts of measurement error can be problematic, particularly when the number of taxa sampled is small. For instance, Cooper *et al*[15] estimate that with a phylogeny of 25 species and a 10% trait

measurement error, stabilizing selection will be falsely concluded ~50% of the time. This is a particular concern for gene expression studies, as the environment can strongly influence gene regulation. Studies should endeavour to control environmental conditions so that differences in expression across samples reflects the heritable, genetic component of expression, as has been discussed previously[8,116]. Second, it is clear that using a single mean expression value for each species can lead to spurious inferences of selection[18], making multiple replicates essential. Importantly, the OU framework has been extended to parameterise within-species variance as an error term[18,58,62] and appears to be a promising approach. Finally, there are methods to control for technical problems that can introduce noise into measurements of expression, such as controlling for batch effects[117–119].

Complex patterns of trait evolution

Many phenotypic traits exhibit complex patterns of evolution and evolve at different rates across lineages[120]. While few studies have directly tested the tempo of expression change across species[60], it seems likely that gene expression does not evolve at a constant rate but instead shifts as mutation rate, selective pressures and pleiotropic constraints vary[47,121,122]. However, many evolutionary models, including BM and OU, assume a homogeneous process of trait change across lineages and/or through time. This is analogous to fitting a fixed d_N/d_S across all branches when estimating selection on coding sequences. Recent work has shown that fitting single-process models masks complexity and leads to inaccurate inferences about the underlying evolutionary process[120]. Comparative methods that account for rate heterogeneity are available (discussed in[120]), analogous to allowing d_N/d_S to vary across branches, but to our knowledge have not been widely applied in the context of gene expression evolution.

BOX 2: Best practises for inferring selection in a comparative framework

Best practises for inferring selection on traits using comparative approaches have been discussed in length in the phylogenetic literature[15–17]. Briefly, to avoid false inferences of stabilizing selection (BOX 1), studies should (i) strive to minimise measurement error, (ii) maximise the number of species sampled and (iii) use comparative approaches that parameterise within-species variance as an error term. Below, we discuss additional recommendations.

Validation of model fit

As discussed, many factors can bias model inference to conclude stabilizing selection over genetic drift. The best fitting model is often chosen by comparing the relative fit of different models. However, studies rarely examine the absolute model fit[120]. This simple step, performed using existing methods such as ARBUTUS[123] or in RevBayes[124], can be used to assess confidence in model selection. This approach relies on the process of posterior predictive simulations, in which datasets are simulated on the estimated parameters, and then a series of test-statistics are run on the simulated data. Similarly, parametric bootstrapping approaches can be applied, resampling the data to generate a bootstrapped sampling distribution from which test statistics are calculated. These results can then be compared to the empirical data to assess the adequacy of the model. Using such approaches for model estimation has been shown to outperform maximum likelihood approaches in specific cases[125].

Multiple testing and False Discovery Rate

Comparative transcriptomics studies perform multiple statistical tests across thousands of genes making them susceptible to the effects of multiple testing. Procedures including False Discovery Rate (FDR) and Bonferroni corrections can easily manage this phenomenon[126,127] yet are frequently not included as standard in phylogenetic comparative transcriptomic approaches. Neutral simulations under predicted parameters permits the estimation of the false discovery rate to account for the inflation of false positives and can be implemented in EVE[65].

Single-cell transcriptomics

By directly comparing gene expression level across equivalent cell types, comparative single-cell transcriptomics (scRNA-seq) can circumvent problems arising when expression is measured from heterogeneous tissue (Fig. 2). However, scRNA-seq is more challenging for many non-model organisms, especially those sampled from the wild, as scRNA-seq performs optimally when single cells are isolated and processed immediately after harvesting tissue. Although tissue dissociation and storage techniques are being developed, such as methanol fixation and cryopreservation, there are concerns that these can either bias expression profiles[128] or lead to cell death. However, we anticipate that these challenges will be overcome as the field progresses and the costs of scRNA-seq will decrease.

Consider tissue composition

We suggest that where possible, studies should quantify cellular composition of the tissue in question and how this varies across species. For instance, if a single cell type dominates or expression level is dominated by one cell type, then our simulations suggest that the potential for bias is reduced. Importantly, if scRNA-seq data is available for the tissue, it is possible to use this to directly test for biases in cellular composition in bulk RNA-seq data[83,129]. Deconvolution methods, such as Decon2[130], BayesPrism[131] or ABIS[132], can be used to estimate cell type abundances and subsequently resolve expression profiles closer to those observed from purified cell subpopulations or scRNA-seq. Such methods have been widely implemented[22,133,134], and may prove valuable if they are able to be co-opted into evolutionary genomic studies. Finally, we urge the use of sampling techniques to directly isolate specific regions or cells of interest using microdissection or cell sorting to greatly reduce cell composition complications, as discussed by Hunnicutt *et al*[21].

FIGURES

Figure 1. Approaches to detect selection on gene expression.

Panel (a) Gene expression evolution has been inferred by contrasting levels of variation within a focal species to divergence across species in a pairwise framework. This principle is analogous to the Hudson Kreitman Aguadé (HKA) test used to detect selection at the DNA level. The neutral expectation is that divergence covaries linearly with intra-specific variance, at least over shorter evolutionary distances. Loci with the highest or lowest levels of intra-specific expression variation relative to neutrality are the best candidates for balancing or directional selection respectively. Loci under stabilizing selection should exhibit limited biological variance and divergence. **Panel (b)** Phylogenetic comparative analyses enable comparisons across species to distinguish between evolutionary processes. Brownian Motion models neutral trait evolution via an unconstrained random walk. It assumes that divergence time between species will describe the diversity across the phylogeny with only one parameter $\boldsymbol{\sigma}^2$, the drift rate, and that variance at the tips of the phylogeny will equal $T\boldsymbol{\sigma}^2$. The Ornstein-Uhlenbeck (OU) model assumes that gene regulation follows a stochastic process that is attracted towards a single optimum value, consistent with stabilizing selection. The additional parameters are therefore α, the strength of pull, and θ, the evolutionary optima. This framework has been extended to test for branch specific processes by incorporating multiple optima to test for expression divergence in specific lineages (red line). **Panel (c)** Phenotypic selection analyses have been applied to gene expression data to infer the mode and strength of selection. These employ multiple regression of relative fitness on multiple traits to calculate selection differentials that estimate total selection (direct and indirect) on gene expression. The covariance between fitness and expression is calculated to infer linear (*S*) and quadratic (*C*) selection differentials at each locus, which signify directional, stabilizing, or disruptive selection. The linear selection differential estimates positive versus negative directional selection, while the quadratic selection differential estimates disruptive versus stabilizing selection. This panel is adapted from Groen et al (2020)[67], which used this approach to measure selection on gene expression in rice. Rice was grown under wet (blue) and dry (yellow) environmental conditions, and organism traits and fitness were measured.

Figure 2. Variation in tissue composition can lead to the perception of differential expression.

Schematic illustrating how variation in tissue composition can bias perception of expression measured from bulk RNA-Seq within (panel a) and across species (panel b). **Panel (a)** The chicken hypothalamus is comprised of 3 major cell types at developmental stage HH10. Pie chart (top) shows the proportion of cells in each cell type. Heatmap (bottom) shows gene expression measured across cells in each cell type and average 'bulk' expression estimated across all cells, equivalent to generating RNA-seq data from the whole tissue. Each cell type exhibits a distinct gene expression profile and bulk expression does not accurately reflect this. Data from[71]. **Panel (b)** shows how differences in tissue composition between species can lead to the false perception of differential expression. Here we illustrate a single tissue comprised of two cell types, type 1 (blue) and type 2 (yellow), in two species. During development in Species 1 (left panel), cell type 1 and 2 have the same rate of cell proliferation. The resulting tissue is evenly comprised of each cell type. Cell type 1 only expresses gene 1 and cell type 2 only expresses gene 2. Bulk RNA-Seq expression reflects single cell expression. In Species II (right panel), an increase in the rate of cellular proliferation for cell type 1 results in a greater proportion of cells of type 1 in the resulting adult tissue. Even though there has been no change in per cell expression of either gene 1 or 2, the relative expression from bulk RNA-Seq of the entire tissue results in the perception of higher expression of gene 1 and lower expression of gene 2 compared to Species 1.

Figure 3. Inferring selection when expression level is measured from a heterogeneous tissue.

Three scenarios illustrating potential pitfalls of inferring selection on gene expression level at a single locus using phylogenetic approaches when expression is measured from bulk sequencing. **Panel (a)** The first column shows the expression level of a single gene in two different cell types across a phylogeny. High levels of expression are in dark pink and low expression in light pink. The relative proportion of each cell type is indicated by the size of the rectangle where cell type A is on the left and cell type B is on the right. **Panel (b)** This column shows the composite expression level of the gene as a function of cell type proportion and gene expression in each species. This would be analogous to measuring expression in bulk from a heterogeneous tissue. **Panel (c)** Results of simulated phylogenetic comparative analyses for each scenario with a phylogeny of 25 (blue) or 100 (yellow) tips on 1000 unique trees. Abbreviations of phylogenetic models are BM (Brownian Motion), WN (White noise), OU (Ornstein-Uhlenbeck model) and BS (OU model with a branch shift). These models were fitted on

the simulated bulk expression values and the relative support for each model is calculated using Akaike weights. Error bars show standard deviation around the mean across simulations. Type 1 error rates for each scenario relative to when these models are fit to expression at the single-cell level are shown. In scenario (i), expression values are static across the phylogeny for each cell type but cell type abundance is evolving under BM. However, phylogenetic approaches falsely infer that expression is evolving under BM. For (ii), expression in both cell types is evolving under BM, whereas tissue composition is stable across the phylogeny with the exception of one tip which has undergone an allometric shift. Here, phylogenetic approaches falsely infer an adaptive shift in expression on a single branch. For (iii), gene expression in both cell types, as well as cell-type abundance, is evolving under BM. However, phylogenetic approaches increasingly falsely infer stabilizing selection on expression level.

Figure 4. The magnitude of allometric shift and covariance of expression level biases the inference of selection.

Panel (a) The probability that regulatory selection is incorrectly inferred increases substantially with the magnitude of an allometric shift. This plot is a more detailed representation of Fig. 3ii, where one species undergoes a shift in tissue composition, ranging from when a tissue is composed of two cell types at equal proportion to when only a single cell type is present. All other species have a tissue composition of 50:50 and expression is evolving under Brownian Motion (BM) in each cell type. **Panel (b)** Covariance of expression between cell types biases inferences of selection. This plot is an extension of Fig. 3iii. Expression is evolving under BM but cell type composition is either static (dotted lines) or also evolving under BM (solid lines). We varied the extent to which gene expression is correlated between cell types, ranging from negative covariance, where expression levels increase in one cell type at the same time as decreasing in the other cell type, to positive covariance, where expression levels decrease or increase in both cell types in a correlated manner. The relative type 1 error rate was calculated as the rate at which a BM model was not best fit to the composite expression value relative to the equivalent error rate when models are fit to single cell simulations.

REFERENCES:

1. Mank, J. E. The transcriptional architecture of phenotypic dimorphism. *Nat Ecol Evol* **1**, 6 (2017).
2. Parsch, J. & Ellegren, H. The evolutionary causes and consequences of sex-biased gene expression. *Nat. Rev. Genet.* **14**, 83–87 (2013).
3. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
4. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
5. Gilad, Y., Oshlack, A. & Rifkin, S. A. Natural selection on gene expression. *Trends Genet.* **22** 456–461 (2006).
6. Necsulea, A. & Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat. Rev. Genet.* **15**, 734–748 (2014).
7. Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **22**, 203–215 (2021).
8. Romero, I. G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* **13**, 505–516 (2012).
9. Signor, S. A. & Nuzhdin, S. V. The evolution of gene expression in cis and trans. *Trends Genet.* **34**, 532–544 (2018).
10. Fay, J. C. & Wittkopp, P. J. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**, 191–199 (2008).
11. Khaitovich, P., Enard, W., Lachmann, M. & Pääbo, S. Evolution of primate gene expression. *Nat. Rev. Genet*. **7**, 693–702 (2006).
12. Bedford, T. & Hartl, D. L. Optimization of gene expression by natural selection. *Proc. Natl. Acad. Sci. USA* **106**, 1133–1138 (2009).
13. Whitehead, A. & Crawford, D. L. Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. USA* **103**, 5425–5430 (2006).

14. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351 (1997).
15. Cooper, N., Thomas, G. H., Venditti, C., Meade, A. & Freckleton, R. P. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc. Lond.* **118**, 64–77 (2016).
16. Silvestro, D., Kostikova, A., Litsios, G., Pearman, P. B. & Salamin, N. Measurement errors should always be incorporated in phylogenetic comparative analysis. *Methods Ecol. Evol.* **6**, 340–346 (2015).
17. Ho, L. S. T. & Ané, C. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* **5**, 1133–1146 (2014).
18. Rohlfs, R. V., Harrigan, P. & Nielsen, R. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. *Mol. Biol.* **31**, 201-211 (2014).
19. Rohlfs, R. V. & Nielsen, R. Phylogenetic ANOVA: The expression variance and evolution model for quantitative trait evolution. *Syst. Biol.* **64**, 695–708 (2015).
20. Montgomery, S. H. & Mank, J. E. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol.* **25**, 5114–5128 (2016).
21. Hunnicutt, K. E., Good, J. M. & Larson, E. L. Unraveling patterns of disrupted gene expression across a complex tissue. *Evolution* (2021).
22. Fair, B. J. *et al.* Gene expression variability in human and chimpanzee populations share common determinants. *Elife* **9**, e59929 (2020).
23. Nourmohammad, A. *et al.* Adaptive evolution of gene expression in *Drosophila*. *Cell Rep.* **20**, 1385–1395 (2017).
24. Catalán, A., Briscoe, A. D. & Höhna, S. Drift and directional selection are the evolutionary forces driving gene expression divergence in eye and brain tissue of *Heliconius* butterflies. *Genetics* **213**, 581–594 (2019).

25. Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261–266 (2002).
26. Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol.* **2**, e132 (2004).
27. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**, 138–144 (2003).
28. Lemos, B., Meiklejohn, C. D., Cáceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**, 126–137 (2005).
29. Hudson, R. R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
30. Kimura, M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res* **11** 247–270 (1968).
31. Staubach, F., Teschke, M., Voolstra, C. R., Wolf, J. B. W. & Tautz, D. A test of the neutral model of expression change in natural populations of house mouse subspecies. *Evolution* **64**, 549–560 (2010).
32. Somel, M. *et al.* Transcriptional neoteny in the human brain. *Proc. Natl. Acad. Sci. USA* **106**, 5743–5748 (2009).
33. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
34. Moghadam, H. K., Pointer, M. A., Wright, A. E., Berlin, S. & Mank, J. E. W chromosome expression responds to female-specific selection. *Proc. Natl. Acad. Sci. USA* **109**, 8207–8211 (2012).
35. Gilad, Y., Oshlack, A., Smyth, G. K., Speed, T. P. & White, K. P. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440**, 242–245 (2006).
36. Enard, W. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).

725 37. Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Gene regulation in primates
726 evolves under tissue-specific selection pressures. *PLoS Genet.* **4**, e1000271 (2008).
727 38. Warnefors, M. & Eyre-Walker, A. A selection index for gene expression evolution and its
728 application to the divergence between humans and chimpanzees. *PLoS One* **7**, e34935 (2012).
729 39. Ometto, L., Shoemaker, D., Ross, K. G. & Keller, L. Evolution of gene expression in fire ants: the
730 effects of developmental stage, caste, and species. *Mol. Biol. Evol.* **28**, 1381–1392 (2011).
731 40. Rifkin, S. A., Houle, D., Kim, J. & White, K. P. A mutation accumulation assay reveals a broad
732 capacity for rapid evolution of gene expression. *Nature* **438**, 220–223 (2005).
733 41. Denver, D. R. *et al.* The transcriptional consequences of mutation and natural selection in
734 *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544–548 (2005).
735 42. Huang, W. *et al.* Spontaneous mutations and the origin and maintenance of quantitative
736 genetic variation. *Elife* **5**, e14625 (2016).
737 43. Fraser, H. B. Detecting selection with a genetic cross. *Proc. Natl. Acad. Sci. USA* **117**, 22323–
738 22330 (2020).
739 44. Leinonen, T., McCairns, R. J. S., O'Hara, R. B. & Merilä, J. Q(ST)-F(ST) comparisons: evolutionary
740 and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.* **14**, 179–190 (2013).
741 45. Mähler, N. *et al.* Gene co-expression network connectivity is an important determinant of
742 selective constraint. *PLoS Genet.* **13**, e1006402 (2017).
743 46. Kohn, M. H., Shapiro, J. & Wu, C.-I. Decoupled differentiation of gene expression and coding
744 sequence among *Drosophila* populations. *Genes Genet. Systems* **83**, 265–273 (2008).
745 47. Papakostas, S. *et al.* Gene pleiotropy constrains gene expression changes in fish adapted to
746 different thermal conditions. *Nat. Commun.* **5**, 4071 (2014).
747 48. Leder, E. H. *et al.* The evolution and adaptive potential of transcriptional variation in
748 sticklebacks--signatures of selection and widespread heritability. *Mol. Biol. Evol.* **32**, 674–689
749 (2015).
750 49. Blanc, J., Kremling, K. A. G., Buckler, E. & Josephs, E. B. Local adaptation contributes to gene

751 expression divergence in maize. *G3* **11**, (2021).
752 50. Pujol, B., Wilson, A. J., Ross, R. I. C. & Pannell, J. R. Are QST-FST comparisons for natural
753 populations meaningful? *Mol. Ecol.* **17**, 4782–4785 (2008).
754 51. Dunn, C. W., Zapata, F., Munro, C., Siebert, S. & Hejnol, A. Pairwise comparisons across species
755 are problematic when analyzing functional genomic data. *Proc. Natl. Acad. Sci. USA* **115**, e409–
756 e417 (2018).
757 52. Felsenstein, J. Phylogenies and the Comparative Method. *Am. Nat.* **125**, 1–15 (1985).
758 53. Pennell, M. W. & Harmon, L. J. An integrative view of phylogenetic comparative methods:
759 connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.*
760 **1289**, 90–105 (2013).
761 54. Felsenstein, J. Maximum-likelihood estimation of evolutionary trees from continuous
762 characters. *Am. J. Hum. Genet.* **25**, 471–492 (1973).
763 55. Oakley, T. H., Gu, Z., Abouheif, E., Patel, N. H. & Li, W.-H. Comparative methods for the analysis
764 of gene-expression evolution: an example using yeast functional genomic data. *Mol. Biol. Evol.*
765 **22**, 40–50 (2005).
766 56. Gu, X. Statistical framework for phylogenomic analysis of gene family expression profiles.
767 *Genetics* **167**, 531–542 (2004).
768 57. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: A modeling approach for adaptive
769 evolution. *Am. Nat.* **164**, 683–695 (2004).
770 58. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**,
771 343–348 (2011).
772 59. Kalinka, A. T. *et al.* Gene expression divergence recapitulates the developmental hourglass
773 model. *Nature* **468**, 811–814 (2010).
774 60. El Taher, A. *et al.* Gene expression dynamics during rapid organismal diversification in African
775 cichlid fishes. *Nat Ecol Evol* **5**, 243–250 (2021).
776 61. Chen, J. *et al.* A quantitative framework for characterizing the evolutionary history of

mammalian gene expression. *Genome Res.* **29**, 53–63 (2019).

62. Pal, S., Oliver, B. & Przytycka, T. M. Modeling gene expression evolution with EvoGeneX uncovers differences in evolution of species, organs and sexes. Preprint at *bioRxiv* https://doi.org/10.1101/2020.01.06.895615 (2021).
63. Greenway, R. *et al.* Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. *Proc. Natl. Acad. Sci. USA* **117**, 16424–16430 (2020).
64. Vegesna, R. *et al.* Ampliconic genes on the Great Ape Y chromosomes: Rapid evolution of copy number but conservation of expression levels. *Genome Biol. Evol.* **12**, 842–859 (2020).
65. Gillard, G. B. *et al.* Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol.* **22**, 103 (2021).
66. Kopania, E. E. K., Larson, E. L., Callahan, C. & Keeble, S. Molecular evolution across mouse spermatogenesis. Preprint at *bioRxiv* https://doi.org/10.1101/2021.08.04.455131 (2021).
67. Groen, S. C. *et al.* The strength and pattern of natural selection on gene expression in rice. *Nature* **578**, 572–576 (2020).
68. Ahmad, F. *et al.* The strength and form of natural selection on transcript abundance in the wild. *Mol. Ecol.* **30**, 2724–2737 (2021).
69. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
70. Guschanski, K., Warnefors, M. & Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **27**, 1461–1474 (2017).
71. Kim, D. W. *et al.* Single-cell analysis of early chick hypothalamic development reveals that hypothalamic cells are induced from prethalamic-like progenitors. *Cell Rep.* **38**, 3 (2022).
72. Estermann, M. A. *et al.* Insights into gonadal sex differentiation provided by single-cell transcriptomics in the chicken embryo. *Cell Rep.* **31**, 107491 (2020).
73. Niu, W. & Spradling, A. C. Two distinct pathways of pregranulosa cell differentiation support

follicle formation in the mouse ovary. *Proc. Natl. Acad. Sci. USA* **117**, 20015–20026 (2020).

74. Witt, E., Benjamin, S., Svetec, N. & Zhao, L. Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*. eLife **8**, e47138 (2019).

75. Hermann, B. P. *et al.* The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. *Cell Rep.* **25**, 1650–1667.e8 (2018).

76. Green, C. D. *et al.* A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-Seq. *Dev. Cell* **46**, 651–667.e10 (2018).

77. La Manno, G. *et al.* Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19 (2016).

78. Tosches, M. A. *et al.* Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).

79. Bakken, T. E. *et al.* Evolution of cellular diversity in primary motor cortex of human, marmoset monkey, and mouse. Preprint at *bioRxiv* https://doi.org/10.1101/2020.03.31.016972 (2020).

80. Lüpold, S., Linz, G. M., Rivers, J. W., Westneat, D. F. & Birkhead, T. R. Sperm competition selects beyond relative testes size in birds. *Evolution* **63**, 391–402 (2009).

81. Shami, A. N. *et al.* Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis. *Dev. Cell* **54**, 529–547.e12 (2020).

82. Harrison, P. W. *et al.* Sexual selection drives evolution and rapid turnover of male gene expression. *Proc. Natl. Acad. Sci. USA* **112**, 4393–4398 (2015).

83. Bauernfeind, A. L. *et al.* Tempo and mode of gene expression evolution in the brain across Primates. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.21.440670 (2021).

84. Shafer, M. E. R. Cross-species analysis of single-cell transcriptomic data. *Front .Cell Dev. Biol.* **7**, 175 (2019).

85. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**,

481–487 (2005).

86. Prud'homme, B. *et al.* Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006).

87. Liu, J., Mosti, F. & Silver, D. L. Human brain evolution: Emerging roles for regulatory DNA and RNA. *Curr. Opin. Neurobiol.* **71**, 170–177 (2021).

88. Sarropoulos, I. *et al.* Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science* **373**, (2021).

89. Brown, J. B. *et al.* Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399 (2014).

90. Gibilisco, L., Zhou, Q., Mahajan, S. & Bachtrog, D. Alternative splicing within and between *Drosophila* species, sexes, tissues, and developmental stages. *PLoS Genet.* **12**, e1006464 (2016).

91. Mazin, P. V., Khaitovich, P., Cardoso-Moreira, M. & Kaessmann, H. Alternative splicing during mammalian organ development. *Nat. Genet.* **53**, 925–934 (2021).

92. Gómez-Redondo, I., Planells, B., Navarrete, P. & Gutiérrez-Adán, A. Role of alternative splicing in sex determination in vertebrates. *Sex Dev.* **15**, 381-391 (2021).

93. Singh, P. & Ahi, E. P. The importance of alternative splicing in adaptive evolution. *Mol. Ecol.* (2022).

94. Rogers, T. F., Palmer, D. H. & Wright, A. E. Sex-specific selection drives the evolution of alternative splicing in birds. *Mol. Biol. Evol.* **38**, 519–530 (2021).

95. Naftaly, A. S., Pau, S. & White, M. A. Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. *Genome Res.* **31**, 1486–1497 (2021).

96. Khan, Z. *et al.* Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**, 1100–1104 (2013).

97. Wang, Z.-Y. *et al.* Transcriptome and translatome co-evolution in mammals. *Nature* **588**, 642–647 (2020).

98. Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J. & Anne Smith, V. Relationship

between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci. Rep.* **5**, 10775 (2015).

99. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).

100. Lopes-Ramos, C. M. *et al.* Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell Rep.* **31**, 107795 (2020).

101. Liu, X., Li, Y. I. & Pritchard, J. K. Trans effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034.e6 (2019).

102. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).

103. Mathieson, I. The omnigenic model and polygenic prediction of complex traits. *Am. J. Hum. Genet.* **108**, 1558–1563 (2021).

104. O'Connor, L. J. *et al.* Extreme polygenicity of complex traits is explained by negative selection. *Am. J. Hum. Genet.* **105**, 456–476 (2019).

105. Pantalacci, S. & Sémon, M. Transcriptomics of developing embryos and organs: A raising tool for evo-devo. *J. Exp. Zool. B Mol. Dev. Evol.* **324**, 363–371 (2015).

106. Liu, J. & Robinson-Rechavi, M. Developmental constraints on genome evolution in four bilaterian model species. *Genome Biol. Evol.* **10**, 2266–2277 (2018).

107. Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).

108. Metzger, B. P. H., Yuan, D. C., Gruber, J. D., Duveau, F. & Wittkopp, P. J. Selection on noise constrains variation in a eukaryotic promoter. *Nature* **521**, 344–347 (2015).

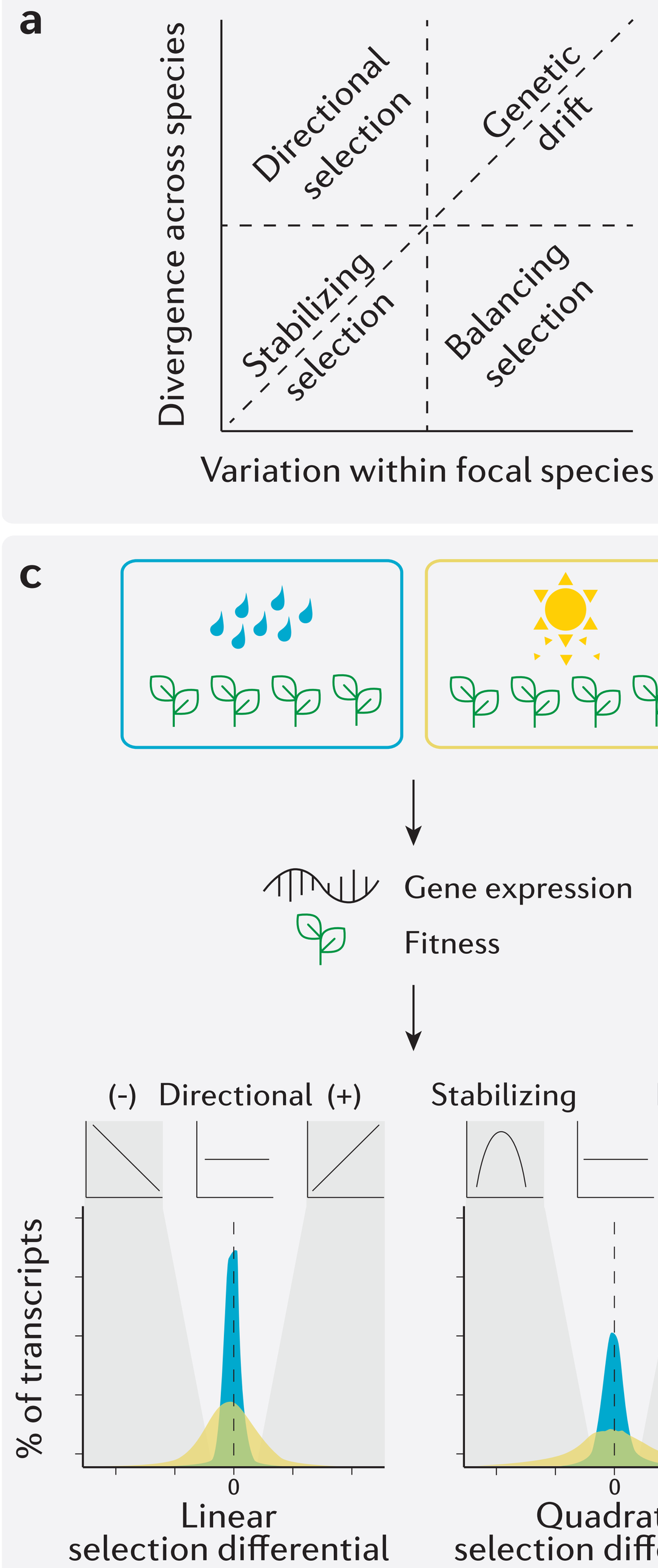
109. Metzger, B. P. H. *et al.* Contrasting frequencies and effects of cis- and trans-regulatory mutations affecting gene expression. *Mol. Biol. Evol.* **33**, 1131–1146 (2016).

110. Hodgins-Davis, A., Duveau, F., Walker, E. & Wittkopp, P. J. Empirical measures of mutational effects define neutral models of regulatory evolution in *Saccharomyces cerevisiae*. *Proc. Nat.*

881 *Acad. Sci. USA* **116**, 21085-21093 (2019).
882 111. Vaishnav, E. D. *et al.* A comprehensive fitness landscape model reveals the evolutionary history
883 and future evolvability of eukaryotic cis-regulatory DNA sequences. Preprint at *bioRxiv*
884 https://doi.org/10.1101/2021.02.17.430503 (2021).
885 112. Josephson, M. P. & Bull, J. K. Innovative mark–recapture experiment shows patterns of
886 selection on transcript abundance in the wild. *Mol. Ecol.* **30**, 2707–2709 (2021).
887 113. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things).
888 *Methods Ecol. Evol.* **3**, 217–223 (2011).
889 114. Ho, L. si T. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution
890 models. *Syst. Biol.* **63**, 397–408 (2014).
891 115. Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C. & O'Meara, B. C. Modeling stabilizing selection:
892 expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383
893 (2012).
894 116. Harrison, P. W., Wright, A. E. & Mank, J. E. The evolution of gene expression and the
895 transcriptome–phenotype relationship. *Semin. Cell Dev. Biol.* **23**, 222–229 (2012).
896 117. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing
897 batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**,
898 882–883 (2012).
899 118. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq
900 count data. *NAR Genom. Bioinform.* **2**(3), lqaa078 (2020).
901 119. Jacob, L., Gagnon-Bartsch, J. A. & Speed, T. P. Correcting gene expression data when neither the
902 unwanted variation nor the factor of interest are observed. *Biostatistics* **17**, 16–28 (2016).
903 120. Chira, A. M. & Thomas, G. H. The impact of rate heterogeneity on inference of phylogenetic
904 models of trait evolution. *J. Evol. Biol.* **29**, 2502–2518 (2016).
905 121. Allen, S. L., Bonduriansky, R. & Chenoweth, S. F. Genetic constraints on microevolutionary
906 divergence of sex-biased gene expression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, (2018).

907 122. Dean, R. & Mank, J. E. Tissue specificity and sex-specific regulatory variation permit the
908 evolution of sex-biased gene expression. *Am. Nat.* **188**, e74–84 (2016).
909 123. Pennell, M. W., FitzJohn, R. G., Cornwell, W. K. & Harmon, L. J. Model adequacy and the
910 macroevolution of angiosperm functional traits. *Am. Nat.* **186**, e33–50 (2015).
911 124. Höhna, S. *et al.* Probabilistic graphical model representation in phylogenetics. *Syst. Biol.* **63**,
912 753–771 (2014).
913 125. Slater, G. J. & Pennell, M. W. Robust regression and posterior predictive simulation increase
914 power to detect early bursts of trait evolution. *Syst. Biol.* **63**, 293–308 (2014).
915 126. Barr, W. A. & Scott, R. S. Phylogenetic comparative methods complement discriminant function
916 analysis in ecomorphology. *Am. J. Phys. Anthropol.* **153**, 663–674 (2014).
917 127. Brzyski, D. *et al.* Controlling the Rate of GWAS False Discoveries. *Genetics* **205**, 61–75 (2017).
918 128. Wang, X., Yu, L. & Wu, A. R. The effect of methanol fixation on single-cell RNA sequencing data.
919 *BMC Genom.* **22**, 420 (2021).
920 129. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust
921 integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
922 130. Aguirre-Gamboa, R. *et al.* Deconvolution of bulk blood eQTL effects into immune cell
923 subpopulations. *BMC Bioinform.* **21**, 243 (2020).
924 131. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Bayesian cell-type deconvolution and gene
925 expression inference reveals tumor-microenvironment interactions. Preprint at *bioRxiv*
926 https://doi.org/10.1101/2020.01.07.897900 (2021)
927 132. Monaco, G. *et al.* RNA-Seq signatures normalized by mRNA abundance allow absolute
928 deconvolution of human immune cell types. *Cell Rep.* **26**, 1627–1640.e7 (2019).
929 133. Chakravarthy, A. *et al.* Pan-cancer deconvolution of tumour composition using DNA
930 methylation. *Nat. Commun.* **9**, 3220 (2018).
931 134. Shen-Orr, S. S. *et al.* Cell type-specific gene expression differences in complex tissues. *Nat.*
932 *Methods* **7**, 287–289 (2010).

933


a
Divergence across species
Directional selection
Genetic drift
Stabilizing selection
Balancing selection
Variation within focal species
c
Gene expression
Fitness
(-) Directional (+)
Stabilizing
Disruptive
% of transcripts
0
0
Linear selection differential
Quadratic selection differential

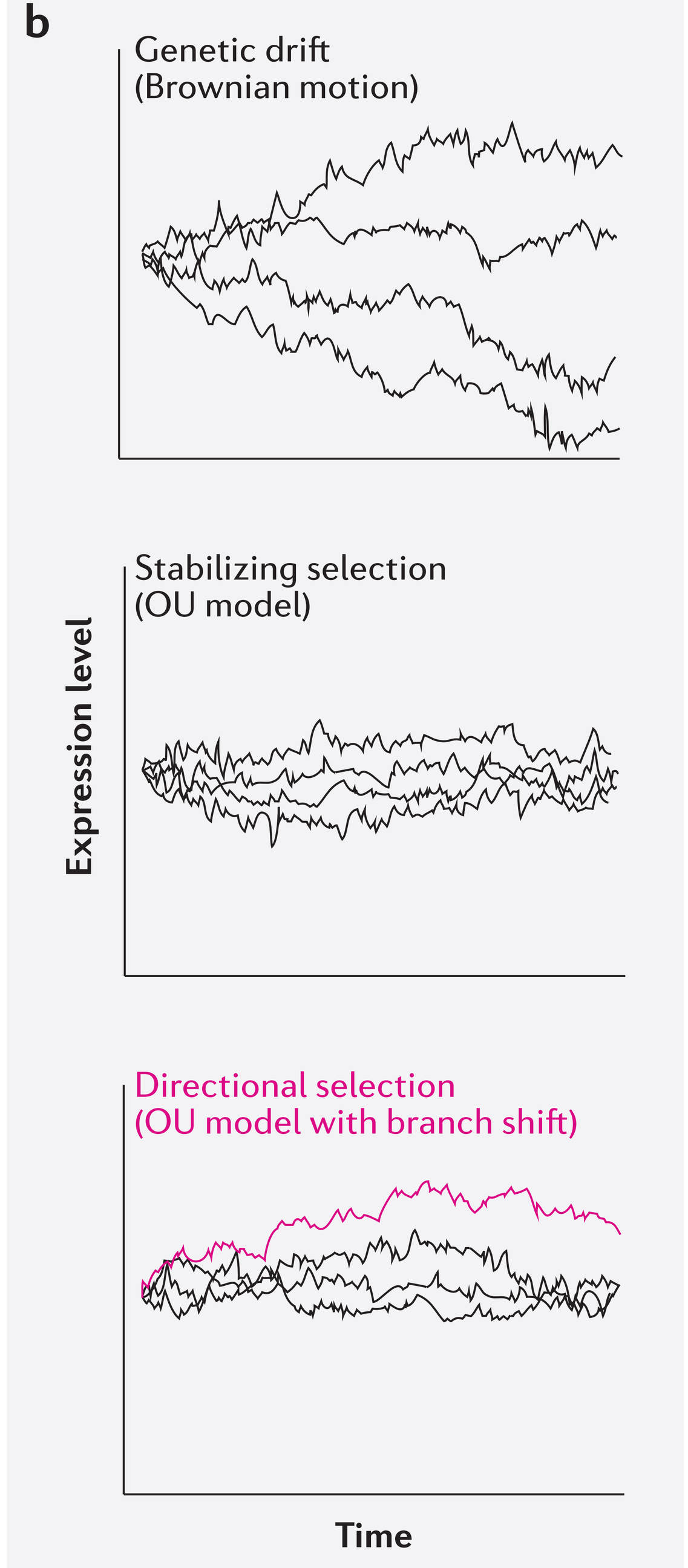

b
Genetic drift
(Brownian motion)
Stabilizing selection
(OU model)
Expression level
Directional selection
(OU model with branch shift)
Time

Figure 1

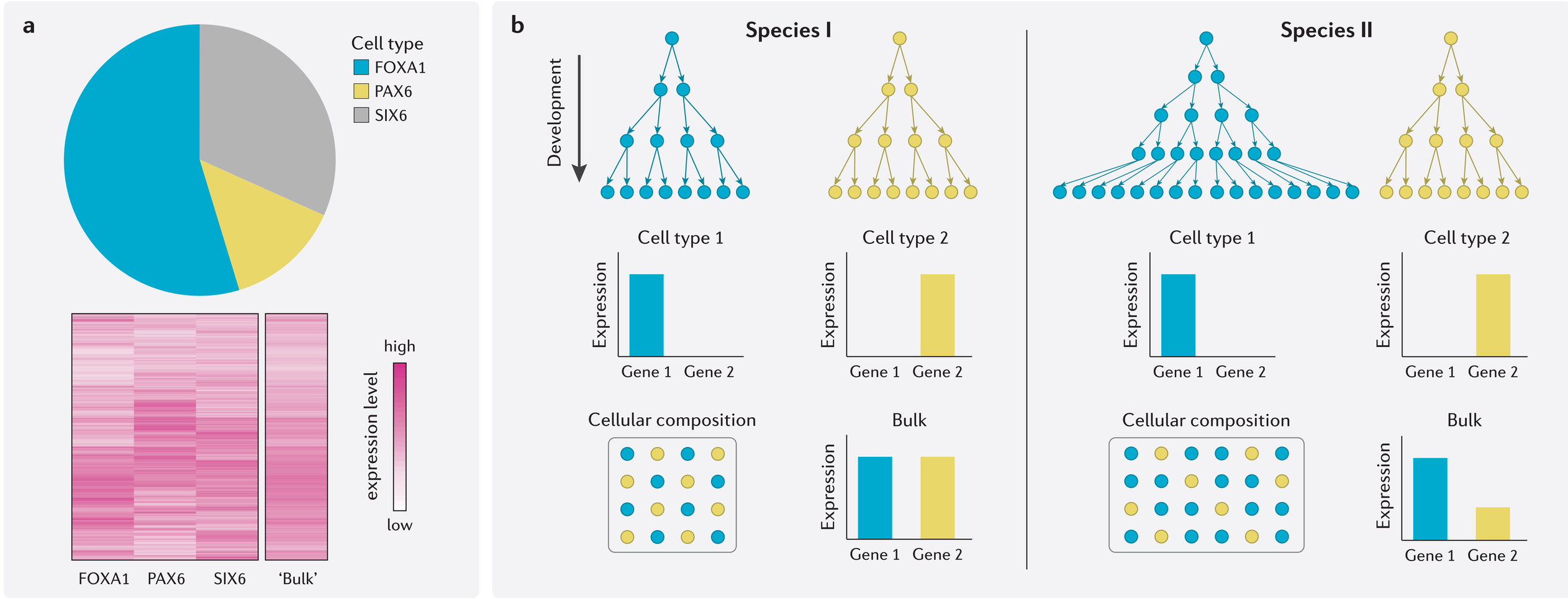

a
Cell type
FOXA1
PAX6
SIX6
high
expression level
low
FOXA1
PAX6
SIX6
'Bulk'
b
Species I
Species II
Development
Cell type 1
Cell type 2
Expression
Gene 1
Gene 2
Cellular composition
Bulk

Figure 2

a Expression level & relative abundance of two cell types

b Bulk expression level

c Model fit to bulk expression level

i.

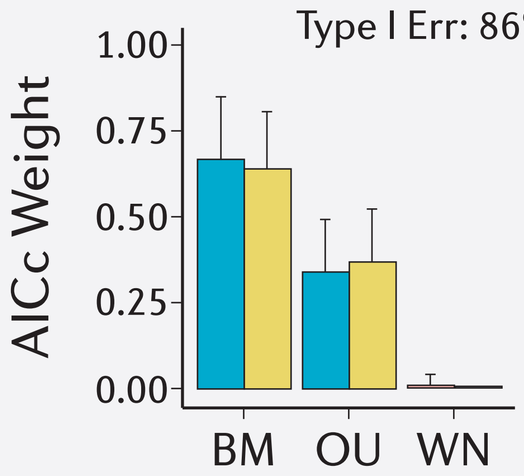
ii.

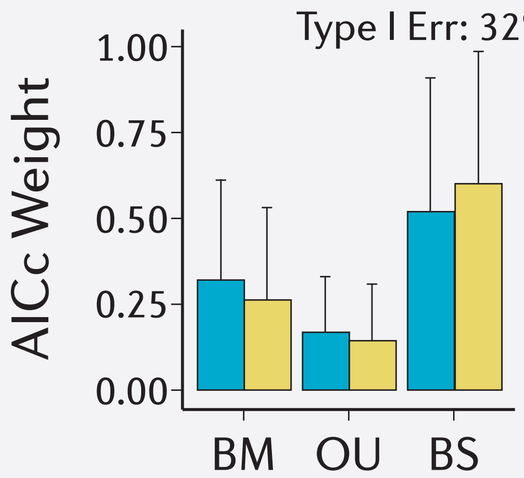
iii.

high

expression level

low





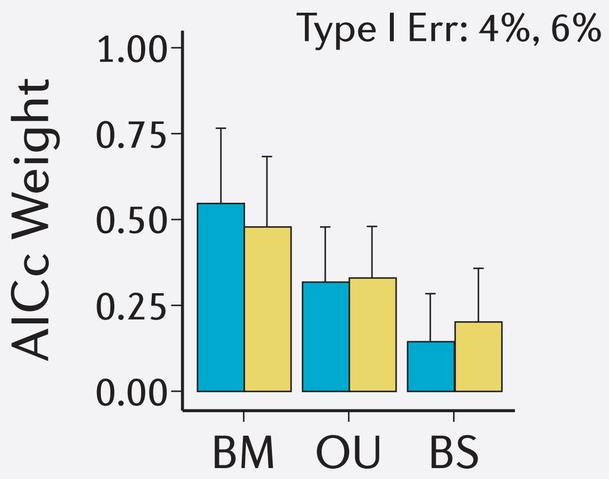


Figure 3

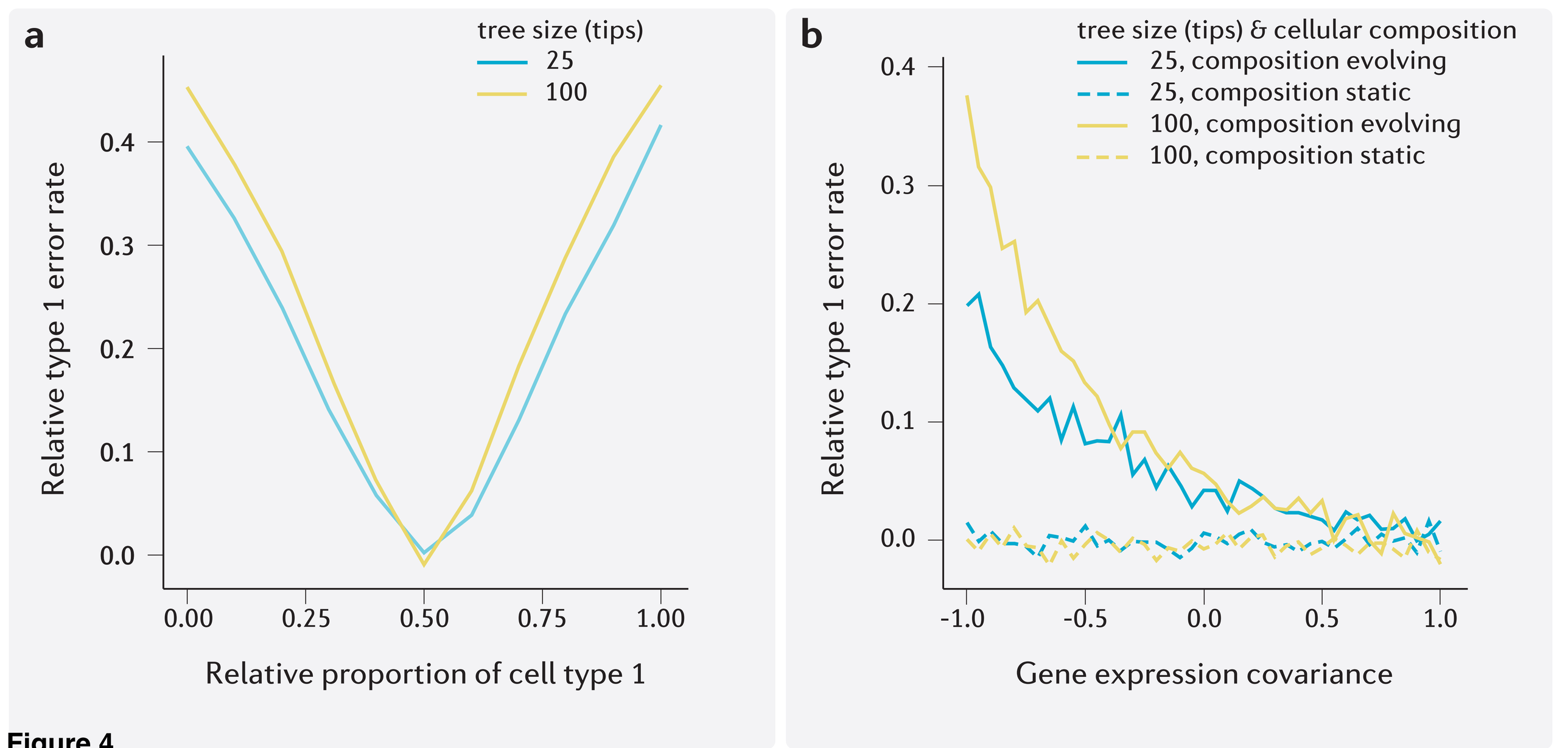

a
tree size (tips)
25
100
Relative type 1 error rate
0.0
0.1
0.2
0.3
0.4
0.00
0.25
0.50
0.75
1.00
Relative proportion of cell type 1
b
tree size (tips) & cellular composition
25, composition evolving
25, composition static
100, composition evolving
100, composition static
Relative type 1 error rate
0.0
0.1
0.2
0.3
0.4
-1.0
-0.5
0.0
0.5
1.0
Gene expression covariance

Figure 4