nature ecology & evolution

ANALYSIS

https://doi.org/10.1038/s41559-022-01761-8

Check for updates

Detecting signatures of selection on gene expression

Peter D. Price[1]✉, Daniela H. Palmer Droguett[1,2], Jessica A. Taylor[1,3], Dong Won Kim[4], Elsie S. Place[5], Thea F. Rogers[1], Judith E. Mank[6,7,8], Christopher R. Cooney[1,9] and Alison E. Wright[1,9]✉

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.

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 selection 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 hence 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 phylogenetic comparative 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.

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK. [2]Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA. [3]Department of Biosciences, Durham University, Durham, UK. [4]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [5]Development, Regeneration and Neurophysiology, School of Biosciences, University of Sheffield, Sheffield, UK. [6]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada. [7]Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada. [8]Centre for Ecology and Conservation, University of Exeter, Penryn, UK. [9]These authors contributed equally: Christopher R. Cooney, Alison E. Wright. ✉e-mail: pprice3@sheffield.ac.uk; a.e.wright@sheffield.ac.uk

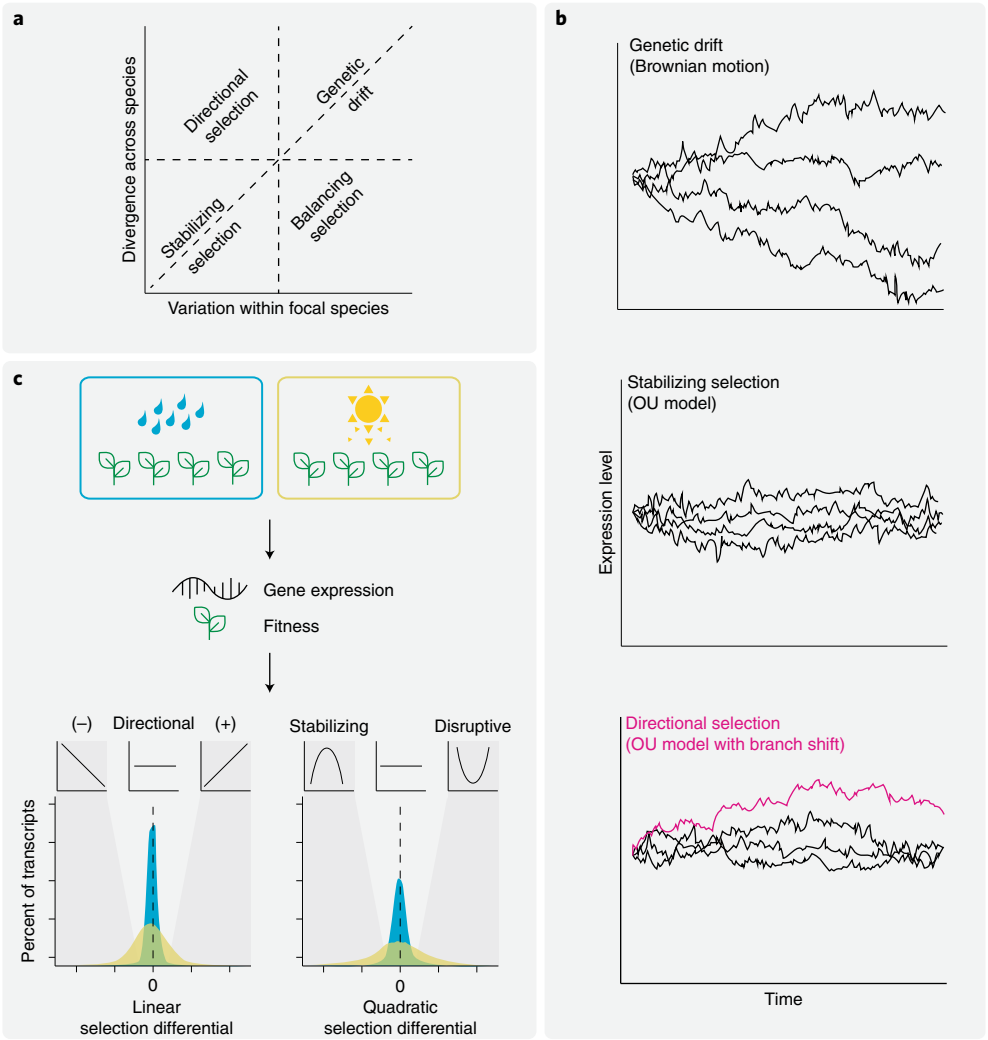

Fig. 1 | Approaches to detect selection on gene expression. a, Gene expression evolution has been inferred by contrasting levels of variation within a focal species to divergence across species in a pairwise framework. This principle is analogous to the Hudson Kreitman Aguadé (HKA) test used to detect selection at the DNA level. The neutral expectation is that divergence covaries linearly with intraspecific variance, at least over shorter evolutionary distances. Loci with the highest or lowest levels of intraspecific 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. **b**, Phylogenetic comparative analyses enable comparisons across species to distinguish between evolutionary processes. Brownian motion (top) models neutral trait evolution via an unconstrained random walk. It assumes that divergence time between species (T) will describe the diversity across the phylogeny with only one parameter σ^2, the drift rate, and that variance at the tips of the phylogeny will equal $T\sigma^2$. The Ornstein–Uhlenbeck (OU) model (middle) 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 a (the strength of pull) and θ (the evolutionary optimum). This framework has been extended to test for branch-specific processes (bottom) by incorporating multiple optima to test for expression divergence in specific lineages (pink line). **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. Groen et al.[67] 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. Panel **c** adapted with permission from ref. [67], Springer Nature Ltd.

Inferring the mode of gene expression evolution

Currently, several 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: (1) contrasts between divergence and variation in expression (Fig. 1a), (2) phylogenetic comparative methods (Fig. 1b) and (3) 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, hence it remains unclear whether discrepancies between studies are biologically meaningful or are caused by inherent methodological differences. Below we synthesize 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], hence 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 among 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 that of 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 markedly greater than the neutral expectation, and stabilizing selection can be inferred when expression of parental populations is significantly less diverged than expected. This study found widespread stabilizing selection on expression level across a range of species, the magnitude of which was dependent on the species' effective population size, consistent with population genetics theory that selection is more effective in species with larger effective population sizes. Selection has also been inferred through comparisons of additive genetic variance of expression (Q_{ST}) with sequence divergence in neutral molecular markers (F_{ST}) across populations[44]. However, while Q_{ST}:F_{ST} approaches have been successfully applied to gene expression variation in a few instances[45–49], accurately estimating the additive genetic basis of gene expression level can be challenging[50], and there is a tendency for dominance variance to bias Q_{ST} estimates, potentially leading to incorrect inferences of neutrality[44].

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

Phylogenetic comparative methods. Phylogenetic comparative methods have been widely adopted for several decades to infer selection acting on phenotypic traits[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 several 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 (σ^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.

Until now, 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 in phylogenetic signal of 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 (that is, 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 while 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 and incorporates 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 between stabilizing versus directional selection, not stabilizing versus neutral evolution, hence do not explicitly rule out neutral processes. Finally, EVE also relies on accurately estimating parameters of the OU process, so it is still probably 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 (that is, by principal component analysis) and selection gradients can then be obtained to estimate direct selection on suites of correlated

Box 1 | Common pitfalls of inferring selection using OU 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 summarize 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 composed of far fewer species due to sampling and computational costs and employ thousands of model comparisons to infer selection at each orthologous locus separately. We anticipate this concern will diminish as expression data become 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 (1) data quality problems, such as RNA degradation, sequencing and assembly issues, (2) low sample sizes and (3) 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 reflect 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 parameterize within-species variance as an error term[18,58,62] and this 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 probable 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 (the ratio of substitutions at non-synonymous and synonymous sites) 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 ref. [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.

transcripts. Recent studies have used these principles to measure selection on gene expression in experimental contexts (for example, by quantifying flowering success and fecundity of rice grown in wet versus drought conditions[67]) and in natural settings (for example, 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 further investigation, particularly whether this reflects methodological biases or difficulties in accurately estimating fitness. However, it is possible that selection pressures vary over short- versus long-term evolutionary time frames, and these approaches are capturing different snapshots of the evolutionary process. Furthermore, unlike these fitness-based approaches, phylogenetic comparative studies primarily rely on contrasting expression across highly conserved orthologous genes, often between very distantly related species, which probably biases our understanding of how gene expression evolves. Gene duplicates are probably key to the evolution of tissue-specific expression patterns[70], hence further work in this area might shed new light on how selection on gene expression varies across genes.

Decomposing transcriptional variation

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

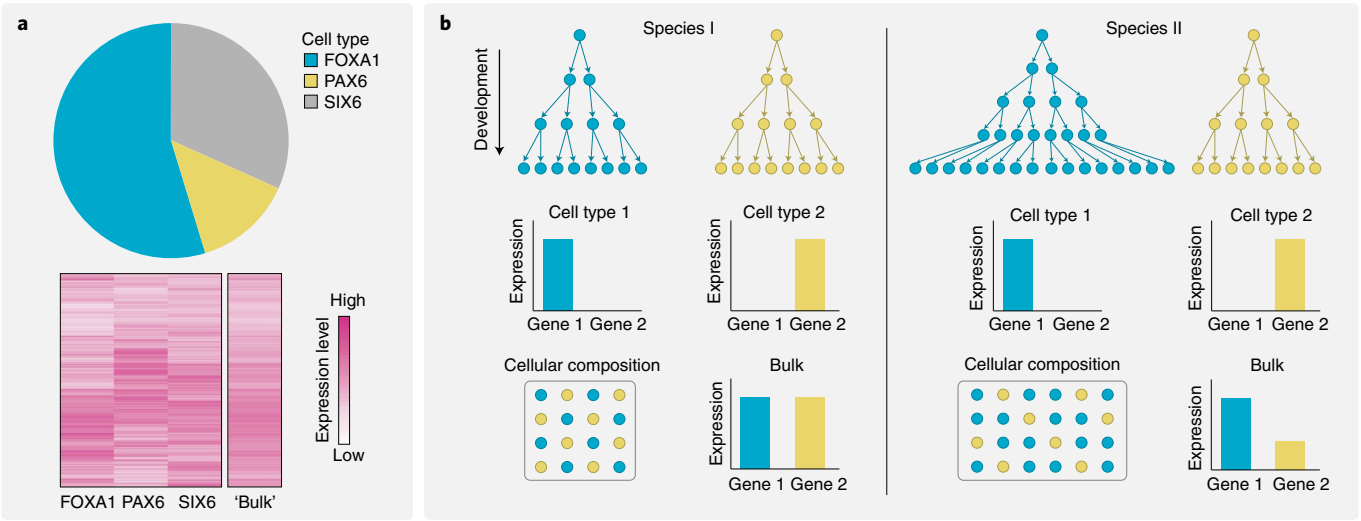

Fig. 2 | Variation in tissue composition can lead to the perception of differential expression. a,**b**, Schematic illustrating how variation in tissue composition can bias perception of expression measured from bulk RNA-seq within (**a**) and across species (**b**). **a**, The chicken hypothalamus is composed 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 ref. [71]. **b**, Illustration of how differences in tissue composition between species can lead to the false perception of differential expression. Here we illustrate a single tissue composed of two cell types, type 1 (blue) and type 2 (yellow), in two species. During development in species I (left), cell types 1 and 2 have the same rate of cell proliferation. The resulting tissue is evenly composed 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), 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. Although 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 with species I.

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 probably 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, overcoming 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 will probably 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, except for recent work that used cell sorting to isolate distinct cell types in 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, hence 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 fitted 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 (that is, not evolving) across species. However, the relative abundance of each cell type is evolving under genetic drift and thus varies across species (Fig. 3a(i)). As predicted, the bulk expression value is neither reflective of single-cell expression levels nor consistent with extreme stabilizing selection (Fig. 3b(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(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

Box 2 | Best practices for inferring selection in a comparative framework

Best practices for inferring selection on traits using comparative approaches have been discussed at length in the phylogenetic literature[15–17]. Briefly, to avoid false inferences of stabilizing selection (Box 1), studies should (1) strive to minimize measurement error, (2) maximize the number of species sampled and (3) use comparative approaches that parameterize 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 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 (FDR). Comparative transcriptomics studies perform multiple statistical tests across thousands of genes, making them susceptible to the effects of multiple testing. Procedures including FDR and Bonferroni correction can easily manage this phenomenon[126,127], yet are frequently not included as standard in phylogenetic comparative transcriptomic approaches. Neutral simulations under predicted parameters permit the estimation of the FDR to account for the inflation of false positives and can be implemented in EVE[65].

Single-cell transcriptomics. By directly comparing gene expression levels 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 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 are available for the tissue, it is possible to use these 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 can 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].

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(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(ii),c(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 (that is, 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%. Although 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(iii)). Here we recover the true signal of genetic drift more reliably (Fig. 3c(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 probably 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 probably 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 (1) how often and by what magnitude changes in tissue composition occur and (2) 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

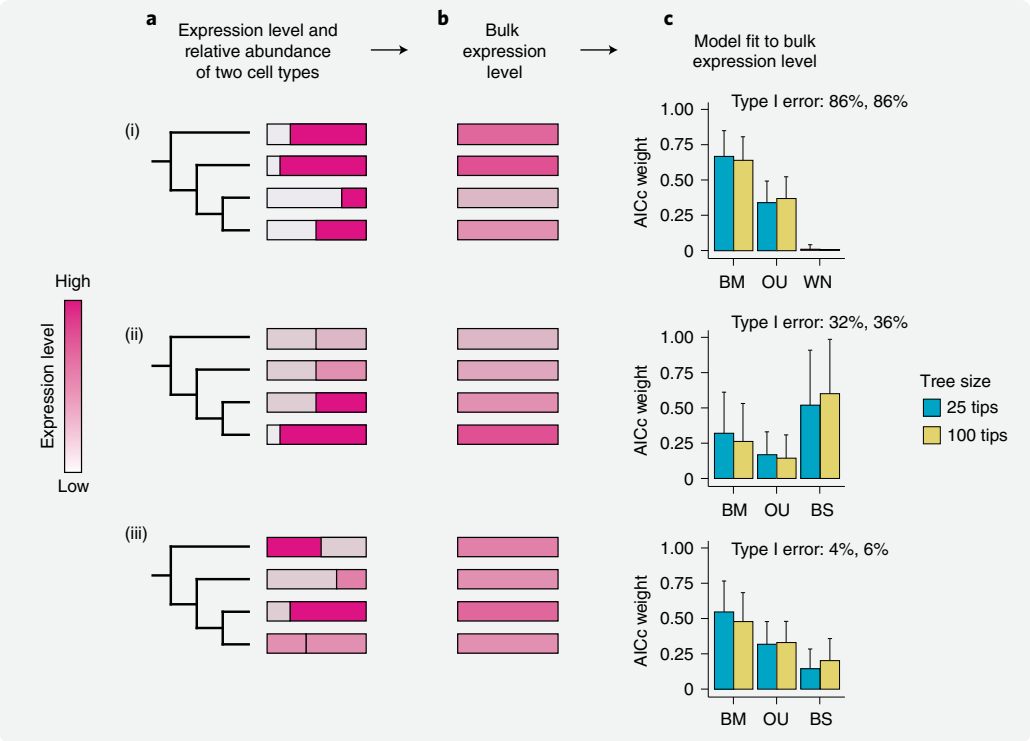

Fig. 3 | Inferring selection when expression level is measured from a heterogeneous tissue. a–c, Three scenarios illustrating potential pitfalls of inferring selection on gene expression level at a single locus using phylogenetic approaches when expression is measured from bulk sequencing. **a**, Expression 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. **b**, 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. **c**, Results of simulated phylogenetic comparative analyses for each scenario with a phylogeny of 25 (blue) or 100 (yellow) tips on 1,000 unique trees. BM, Brownian motion; WN, white noise; OU, Ornstein–Uhlenbeck model; BS, OU model with a branch shift. These models were fitted on the simulated bulk expression values and the relative support for each model was calculated using Akaike information criterion (AIC) weights while accounting for varying sample size (AICc). Error bars show standard deviation around the mean across simulations. Type 1 error rates (for trees of 25 and 100 tips, respectively) for each scenario relative to when these models were fitted 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 except for one tip that 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.

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 minimize other sources of error (Box 2).

Future directions

Given the importance of changes in gene expression 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 biology 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, is 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 (for example, 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

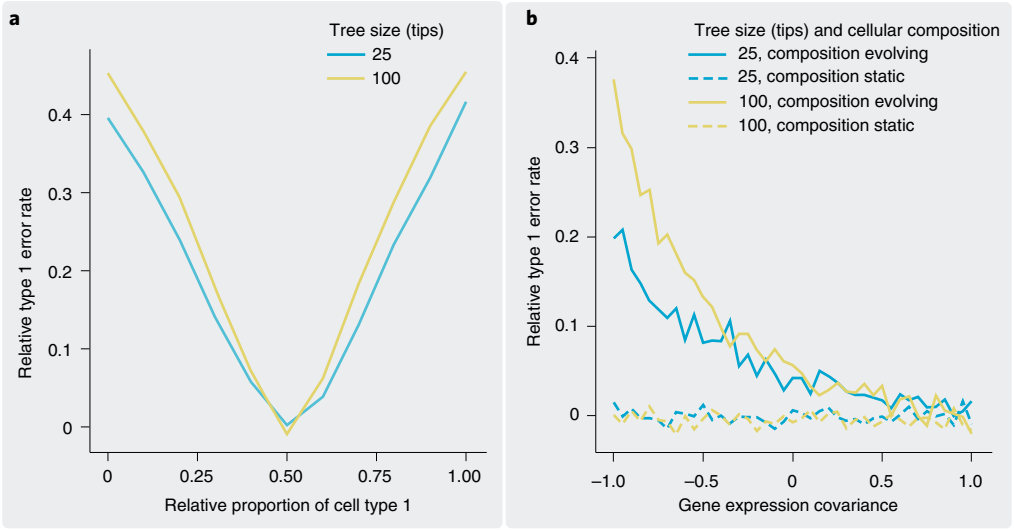

Fig. 4 | The magnitude of allometric shift and covariance of expression level biases the inference of selection. a, The probability that selection on gene expression is incorrectly inferred increases substantially with the magnitude of an allometric shift. This plot is a more detailed representation of Fig. 3(ii), 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 BM in each cell type. **b**, Covariance of expression between cell types biases inferences of selection. This plot is an extension of Fig. 3(iii). Expression is evolving under BM but cell type composition is either static (dashed 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 while at the same time 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 fitted to the composite expression value relative to the equivalent error rate when models were fitted to single-cell simulations.

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 (for example, 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, genome-wide association studies have revealed that complex phenotypes are often the product of many different loci where regulatory networks are probably 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 understand 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 such as 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 scRNA-seq data 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 have been previously published[71]. Pseudo-bulk datasets were generated at HH10 by calculating the average expression across cells in the three cell types.

Simulations. For the first scenario (Fig. 3a(i)), expression values were set at 1 and 2 in two cell types (A and B), respectively. The relative proportion of each cell type (pr) was simulated under BM for 1,000 unique trees of either 25 or 100 tips, using fastBM from phytools[113] in R v4.1.1. The resultant values were normalized between 0 and 1. Composite expression values for each tip (*i*) were calculated as follows:

$$\exp_i = \left(\mathrm{pr}_i \times 1\right) + \left(\left(1 - \mathrm{pr}_i\right) \times 2\right). \quad (1)$$

For the second scenario (Fig. 3a(ii)), expression values were evolved under BM over 1,000 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:

$$\exp_t = \left(\mathrm{pr}_t \times \exp_{\mathrm{A}t}\right) + \left(\left(1 - \mathrm{pr}_t\right) \times \exp_{\mathrm{B}t}\right). \quad (2)$$

Expression for the other tips (*i*) was calculated as above using equation (1).

For the third scenario (Fig. 3a(iii)), 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 trees of 25 and 100 tips were examined. Simulations for each covariance value were run 1,000 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:

$$\exp_i = \left(\mathrm{pr}_i \times \exp_{\mathrm{A}i}\right) + \left(\left(1 - \mathrm{pr}_i\right) \times \exp_{\mathrm{B}i}\right). \quad (3)$$

Fitting evolutionary models to composite expression levels. We fitted evolutionary models in R using phylolm[114] for scenario 1 (Fig. 3a(i)) and OUwie[115] for scenarios 2 (Fig. 3a(ii)) and 3 (Fig. 3a(iii)). For the first scenario (Fig. 3a(i)), 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 fitted and their Akaike weights calculated. The WN model was fitted by suppressing the phylogenetic signal by fixing Pagel's λ to 0. We calculated the type 1 error rate for scenario 1 (Fig. 3a(i)) 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 was applied to a single-cell-type simulation. For scenarios 2 (Fig. 3a(ii)) and 3 (Fig. 3a(iii)), we fitted a BM, an OU and an OU-shift model, where in the latter the optimum value of the trait was allowed to vary on a single tip. For scenario 2 (Fig. 3a(ii)), the OU-shift model was fitted so that the tip with the proportion shift was allowed the alternate optima, whereas for scenario 3 (Fig. 3a(iii)), a random tip was allocated. For scenarios 2 (Fig. 3a(ii)) and 3 (Fig. 3a(iii)), the type 1 error rate was calculated as the rate at which a non-BM model was favoured (that is, where $\Delta\mathrm{AICc} > 2$) relative to BM, relative to when the same set of models was applied to a single-cell-type simulation.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data have been previously published[71].

Code availability

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

Received: 2 November 2021; Accepted: 1 April 2022;
Published online: 12 May 2022

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* **76**, 275–291 (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).

37. Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet.* **4**, e1000271 (2008).
38. Warnefors, M. & Eyre-Walker, A. A selection index for gene expression evolution and its application to the divergence between humans and chimpanzees. *PLoS ONE* **7**, e34935 (2012).
39. Ometto, L., Shoemaker, D., Ross, K. G. & Keller, L. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol. Biol. Evol.* **28**, 1381–1392 (2011).
40. Rifkin, S. A., Houle, D., Kim, J. & White, K. P. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* **438**, 220–223 (2005).
41. Denver, D. R. et al. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544–548 (2005).
42. Huang, W. et al. Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *eLife* **5**, e14625 (2016).
43. Fraser, H. B. Detecting selection with a genetic cross. *Proc. Natl Acad. Sci. USA* **117**, 22323–22330 (2020).
44. Leinonen, T., McCairns, R. J. S., O'Hara, R. B. & Merilä, J. Q(ST)-F(ST) comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.* **14**, 179–190 (2013).
45. Mähler, N. et al. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.* **13**, e1006402 (2017).
46. Kohn, M. H., Shapiro, J. & Wu, C.-I. Decoupled differentiation of gene expression and coding sequence among *Drosophila* populations. *Genes Genet. Syst.* **83**, 265–273 (2008).
47. Papakostas, S. et al. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat. Commun.* **5**, 4071 (2014).
48. Leder, E. H. et al. The evolution and adaptive potential of transcriptional variation in sticklebacks—signatures of selection and widespread heritability. *Mol. Biol. Evol.* **32**, 674–689 (2015).
49. Blanc, J., Kremling, K. A. G., Buckler, E. & Josephs, E. B. Local adaptation contributes to gene expression divergence in maize. *G3* **11**, jkab004 (2021).
50. Pujol, B., Wilson, A. J., Ross, R. I. C. & Pannell, J. R. Are QST-FST comparisons for natural populations meaningful? *Mol. Ecol.* **17**, 4782–4785 (2008).
51. Dunn, C. W., Zapata, F., Munro, C., Siebert, S. & Hejnol, A. Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proc. Natl Acad. Sci. USA* **115**, e409–e417 (2018).
52. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
53. Pennell, M. W. & Harmon, L. J. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.* **1289**, 90–105 (2013).
54. Felsenstein, J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471–492 (1973).
55. Oakley, T. H., Gu, Z., Abouheif, E., Patel, N. H. & Li, W.-H. Comparative methods for the analysis of gene-expression evolution: an example using yeast functional genomic data. *Mol. Biol. Evol.* **22**, 40–50 (2005).
56. Gu, X. Statistical framework for phylogenomic analysis of gene family expression profiles. *Genetics* **167**, 531–542 (2004).
57. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* **164**, 683–695 (2004).
58. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
59. Kalinka, A. T. et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).
60. El Taher, A. et al. Gene expression dynamics during rapid organismal diversification in African cichlid fishes. *Nat. Ecol. Evol.* **5**, 243–250 (2021).
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. *Mol. Biol. Evol.* **39**, msac023 (2022).
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**, 110251 (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. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
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**, eabg4696 (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.* **31**, 1928–1938 (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. & Smith, V. A. 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* **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. Acad. Sci. USA* **116**, 21085–21093 (2019).
111. Vaishnav, E. D. et al. A comprehensive fitness landscape model reveals the evolutionary history and future evolvability of eukaryotic *cis*-regulatory DNA sequences. Preprint at *bioRxiv* https://doi.org/10.1101/2021.02.17.430503 (2021).
112. Josephson, M. P. & Bull, J. K. Innovative mark–recapture experiment shows patterns of selection on transcript abundance in the wild. *Mol. Ecol.* **30**, 2707–2709 (2021).
113. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2011).
114. Ho, L., Si, T. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).
115. Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C. & O'Meara, B. C. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383 (2012).
116. Harrison, P. W., Wright, A. E. & Mank, J. E. The evolution of gene expression and the transcriptome–phenotype relationship. *Semin. Cell Dev. Biol.* **23**, 222–229 (2012).
117. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
118. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**, lqaa078 (2020).
119. Jacob, L., Gagnon-Bartsch, J. A. & Speed, T. P. Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. *Biostatistics* **17**, 16–28 (2016).
120. Chira, A. M. & Thomas, G. H. The impact of rate heterogeneity on inference of phylogenetic models of trait evolution. *J. Evol. Biol.* **29**, 2502–2518 (2016).
121. Allen, S. L., Bonduriansky, R. & Chenoweth, S. F. Genetic constraints on microevolutionary divergence of sex-biased gene expression. *Phil. Trans. R. Soc. B* **373**, 20170427 (2018).
122. Dean, R. & Mank, J. E. Tissue specificity and sex-specific regulatory variation permit the evolution of sex-biased gene expression. *Am. Nat.* **188**, e74–e84 (2016).
123. Pennell, M. W., FitzJohn, R. G., Cornwell, W. K. & Harmon, L. J. Model adequacy and the macroevolution of angiosperm functional traits. *Am. Nat.* **186**, e33–e50 (2015).
124. Höhna, S. et al. Probabilistic graphical model representation in phylogenetics. *Syst. Biol.* **63**, 753–771 (2014).
125. Slater, G. J. & Pennell, M. W. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. *Syst. Biol.* **63**, 293–308 (2014).
126. Barr, W. A. & Scott, R. S. Phylogenetic comparative methods complement discriminant function analysis in ecomorphology. *Am. J. Phys. Anthropol.* **153**, 663–674 (2014).
127. Brzyski, D. et al. Controlling the rate of GWAS false discoveries. *Genetics* **205**, 61–75 (2017).
128. Wang, X., Yu, L. & Wu, A. R. The effect of methanol fixation on single-cell RNA sequencing data. *BMC Genomics* **22**, 420 (2021).
129. Jew, B. et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
130. Aguirre-Gamboa, R. et al. Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics* **21**, 243 (2020).
131. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Bayesian cell-type deconvolution and gene expression inference reveals tumor-microenvironment interactions. Preprint at *bioRxiv* https://doi.org/10.1101/2020.01.07.897900 (2021)
132. Monaco, G. et al. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.* **26**, 1627–1640.e7 (2019).
133. Chakravarthy, A. et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat. Commun.* **9**, 3220 (2018).
134. Shen-Orr, S. S. et al. Cell type-specific gene expression differences in complex tissues. *Nat. Methods* **7**, 287–289 (2010).

Acknowledgements

This work was funded by an NERC Independent Research Fellowship to C.R.C. (NE/T01105X/1); an NERC Independent Research Fellowship to A.E.W. (NE/N013948/1); a grant from the European Research Council (grant agreement 680951) and a Canada 150 Research Chair to J.E.M.; an NERC ACCE DTP to P.D.P.; and an NSF Postdoctoral Research Fellowship and an MSU Presidential Postdoctoral Fellowship to D.H.P.D. We thank M. Placzek, P. E. Pifarré, E. Josephs, A. Platts, M. Roberts, R. Panko, M. Wilson Brown and S. Buysse for helpful comments and suggestions on the manuscript.

Author contributions

A.E.W., C.R.C., D.H.P.D., P.D.P. and J.E.M. designed the study. D.W.K., E.S.P., A.E.W., C.R.C. and P.D.P. analysed the data. All authors wrote and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01761-8.

Correspondence and requests for materials should be addressed to Peter D. Price or Alison E. Wright.

Peer review information *Nature Ecology & Evolution* thanks Camille Berthelot and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

nature portfolio

Corresponding author(s): Peter D Price

Last updated by author(s): Mar 31, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NA
Data analysis	R v4.1.1 with the following packages - GEIGER 2.0.7 - Phylolm 2.6.2 - dplyr 1.0.8 - MuMIn 1.46.0 - phytools 1.0-1 - OUwie 2.6 - ggplot2 3.3.5 - ggpubr 0.4.0 - tidyr 1.2.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data has been published previously71

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	For a hypothetical tissue of two cell types, gene expression was simulated under brownian motion for each cell type. The relative proportion of each cell type was then varied, and a final composite expression value was calculated. A variety of quantitative trait models of evolution were then fit to these values under a variety of scenarios where tissue composition shifted either symmetrically or asymmetrically over a phylogeny. Model fit was compared using AICc weights with the best fitting model being > 2 deltaAICc than the next highest.
Research sample	Simulated data
Sampling strategy	NA
Data collection	NA
Timing and spatial scale	NA
Data exclusions	NA
Reproducibility	NA
Randomization	NA
Blinding	NA

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging