Detecting Signatures Of Selection In Regulatory Variation

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

+Joint senior author

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

ABSTRACT

A substantial amount of phenotypic diversity results from changes in gene regulation. Understanding how regulatory diversity 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 regulatory evolution. Furthermore, recent work has shown that many of the comparative approaches used to study gene regulation are subject to biases that can lead to false signatures of selection. In this review, we first outline the main approaches for describing regulatory evolution and their inherent biases. Next, we bridge the gap between the fields of comparative phylogenetic methods and transcriptomics to reinforce the main pitfalls of inferring regulatory selection 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 regulatory variation and identifying major, unanswered questions in disentangling how selection acts on the transcriptome.

INTRODUCTION

A growing body of evidence indicates that changes in gene regulation play a key role in phenotypic divergence[1]. Within species, a single genome can encode multiple distinct phenotypes by varying expression levels of the underlying loci[2]. Similarly, across species, regulatory variation is implicated in major phenotypic differences that underlie adaptive change[1]. 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[3]. 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[4–7].

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 predict the phenotypic effects of different types of coding mutations and scan coding sequence data for regions of adaptive evolution, gene regulation can be complex and non-additive in its phenotypic effects. This complexity has resulted in a wide range of approaches to study regulatory evolution[3,8,9]. 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[8,10], and these have provided important insights into patterns of regulatory variation. 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[11,12]. Many of the root causes of these biases are even more pronounced in transcriptomic data, but the issues uncovered in the phylogenetic comparative literature[11,12] are only rarely discussed in the genomics field[13,14].

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 measure expression in bulk across heterogeneous tissue samples and so cannot distinguish changes in gene regulation from differences in tissue composition[15,16]. 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 regulatory changes in development. However, these developmental regulatory differences will not be detected if transcriptomes are measured after development is completed and instead the resulting differences in gene abundance will be mistaken as causative adaptive changes (Fig. 1). This problem undermines our current understanding of the nature and abundance of regulatory variation, and how it contributes to phenotypic divergence. Although the implications of varying tissue composition across species for measuring regulatory change have been discussed[15,16], the consequences of how it affects the inference of expression evolution have received little attention and so are not widely appreciated.

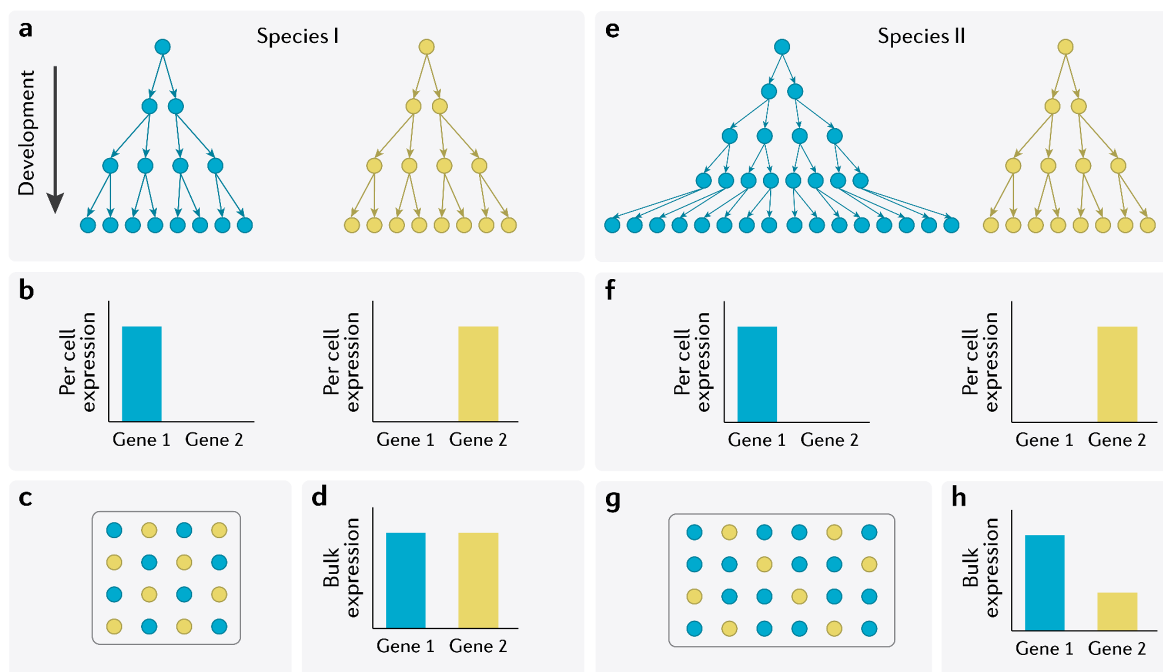

Figure 1. Variation in tissue composition can lead to the perception of differential expression. Schematic illustrating how variation in tissue composition can bias perception of expression measured from bulk RNA-Seq. Here, a single tissue is comprised of two cell types, type 1 (blue) which only expresses gene 1, and type 2 (yellow) which only expresses gene 2. During development in Species I **(a-d)**, cell type 1 and 2 have the same rate of cell proliferation **(a)** and per cell expression is the same for both genes within each cell type **(b)**. The resulting tissue is evenly comprised of each cell type **(c)** and bulk RNA-Seq expression for both gene 1 and 2 is equal **(d)**. In Species II **(e-h)**, a slight increase in the rate of cellular proliferation for cell type 1 **(e)** results in a greater proportion of cells of type 1 in the resulting adult tissue **(g)**. Even though there has been no change in per cell expression of either gene 1 or 2 **(f)**, 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 expression in Species 1 **(h)**.

Here, we examine our current understanding of the evolutionary processes generating variation in gene regulation. First, we outline the main approaches for describing regulatory 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 regulatory variation. Importantly, we identify a previously overlooked challenge to the study of expression evolution concerning shifts in tissue composition across taxa, and use simulation studies to show that this issue can heavily bias inferences of selection. We close by highlighting the multi-dimensional nature of regulatory 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 describing regulatory 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 2A), (ii) phylogenetic comparative methods (Fig 2B) and (iii) fitness-based approaches (Fig 2C). Importantly, each makes different assumptions regarding the mode of expression divergence and are subject to distinct biases. With a few exceptions[13,14,17,18], 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 regulatory variation.

Contrasting divergence and variation in expression

Many early analyses of regulatory evolution tested for selection by contrasting expression divergence between species and expression diversity within species[19–23]. This method relies on the assumption that neutral changes are based solely on the underlying mutation rate[24,25] and so divergence between species relative to polymorphism within species will be equal at neutral loci[25]. When applied to expression data, mutation leads to polymorphism, which can be inferred through variation in expression level amongst individuals. Thus, a neutral model of evolution can be rejected when there are deviations

from an equal ratio of within to between species regulatory variation (Fig. 2A). Studies employing this approach are dominated by two competing viewpoints. One posits that gene regulation is predominantly neutrally evolving[9,19,20,26] and the other suggests widespread conservation and purifying selection of gene expression levels[21,23,27,28] with evidence of positive selection acting on certain loci[29–34].

Analogous approaches using alternative neutral models of expression divergence have also found broad support for stabilizing selection[3,6]. One such approach uses mutation accumulation studies to estimate neutral expectations of expression divergence and infer selection through contrasts with natural populations[35–37]. Most recently, the distribution of expression levels of F2 offspring from a genetic cross has been used to estimate expected levels of neutral divergence[38]. Here, under neutrality, expression variance of the two parental populations should be equal to the F2 progeny as the F2 phenotypes result from random combinations of segregating alleles. Following this logic, directional selection can be inferred when parental divergence is significantly greater than the neutral expectation and stabilizing selection can be inferred when parents are significantly less diverged than expected. This study found widespread stabilizing selection across a range of species, the magnitude of which was dependent on the species' effective population size, consistent with population genetic theory. 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[39]. However, while Q_{ST}:F_{ST} approaches have been successfully applied to gene expression variation in a few instances[40–44], accurately estimating the additive genetic basis of gene regulation can be challenging[45]. There is a tendency for dominance variance to bias Q_{ST} estimates, potentially leading to incorrect inferences of neutrality[39].

Nonetheless, the broad approach of contrasting inter- and intra-specific regulatory 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[46,47]. 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, expression divergence among taxa may become nonlinear, leading to instances of genetic drift being mistaken as directional selection[9,14]. 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 to infer selection acting on phenotypic traits for a number of decades[10,47–49]. 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 discrete evolutionary processes to expression data for a given gene[8,50,51] (Fig. 2B). A commonly used model, Brownian Motion (BM), assumes that expression divergence between species will be a function of divergence time and evolutionary rate, and, as such, is often seen as analogous to genetic drift. A second model, the Ornstein-Uhlenbeck (OU) model, adds an 'elastic band' element drawing expression values towards an optimum across the phylogeny, akin to stabilizing selection[8,52]. The OU model can be extended to allow for branch-specific events, such as shifts in optimum trait values[8,53], analogous to directional selection in particular lineages.

Comparative transcriptomic analyses based on the OU model have found overwhelming support for stabilizing selection on expression levels across a wide range of species, including *Drosophila*[8,54], African cichlids[55] and mammals[56]. While this appears consistent with past work[21,23,27,28], 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[11–13] or branch-specific selection[14]. 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[13,53,57], and whilst it has been shown to reduce false inferences of stabilizing selection, this approach has only been employed by a handful of studies[18,58].

Recently, Rohlfs *et al.*[14] 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. 2A). 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[14,18], and evidence for elevated rates of expression evolution consistent with adaptive evolution[58–61], contrasting with previous findings. However, this method 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 regulation[62,63]. 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[64]. 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[62,63] (Fig. 2C). 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 regulatory selection in experimental contexts (e.g. by quantifying flowering success and fecundity of rice grown in wet versus drought conditions[62]) and in natural settings (e.g. by measuring parasite load and survivorship of wild trout using mark-recapture[63]). 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[63]. Causes of this discrepancy require further investigation, particularly whether or not this reflects methodological biases. 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.

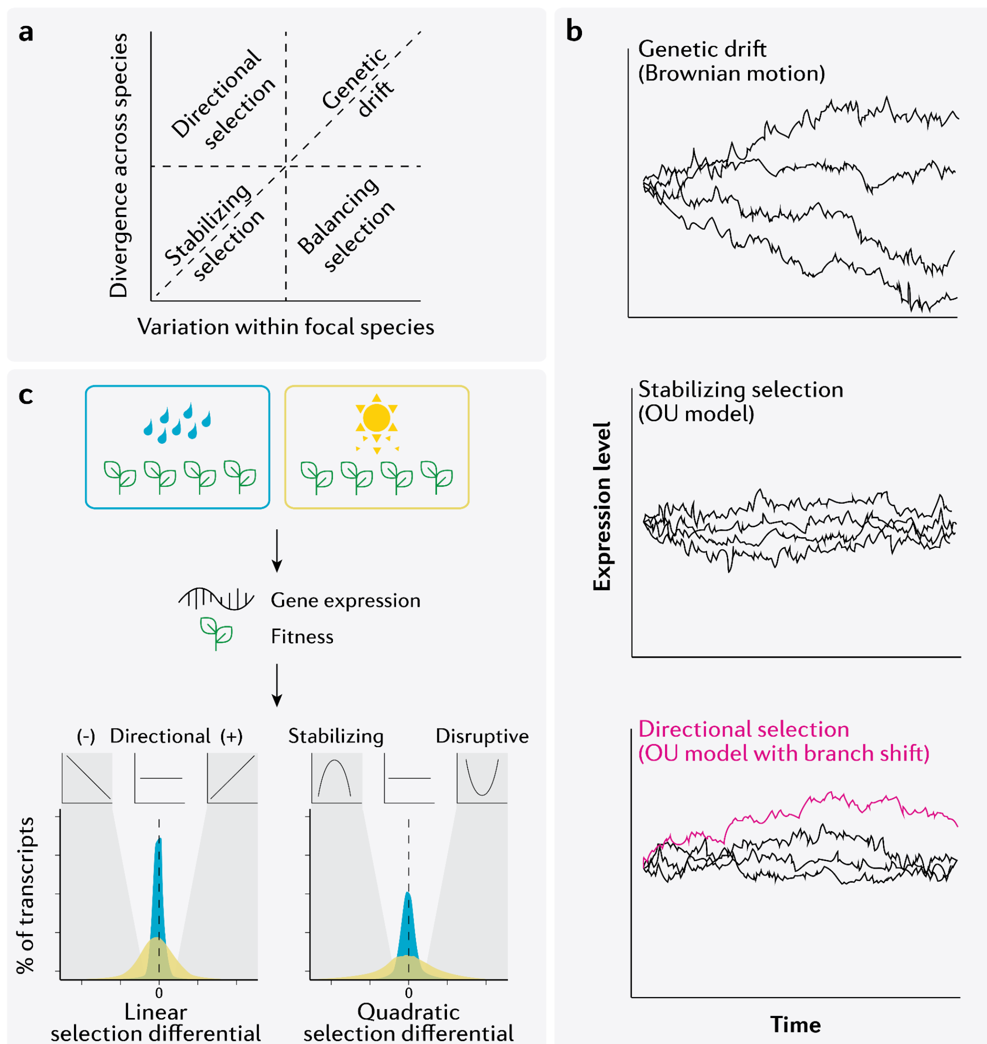

Figure 2. Approaches to detect selection on regulatory variation.

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 regulatory 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 σ^2, the drift rate, and that variance at the tips of the phylogeny will equal $T\sigma^2$. 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 trait

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)[62], which used this approach to measure selection on gene expression in rice. Rice was grown under wet (blue) and dry (yellow) environmental conditions, and phenotypes and fitness were measured.

DECOMPOSING TRANSCRIPTIONAL VARIATION

Importantly, approaches designed to test for selection on regulatory variation all make the explicit assumption that when differential gene expression is observed, it 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, to date, 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[1]. In doing so, these 'bulk' expression values represent an average of expression across entire populations of distinct cell types. Importantly, this means that samples that vary in tissue composition can produce patterns of differential expression that are often mistaken as evidence of regulatory change (Fig. 1). Conversely, this approach can also dampen and/or mask genuine regulatory differences[15,16].

Recent advances in single-cell transcriptomics are providing new insights into tissue composition and how this can vary both within and across species. Within species, dramatic changes in tissue composition are well documented throughout development[65–67] and between the sexes[65]. This is exemplified by gonadal tissue, which exhibits sex-specific cell types[65] as well as a mix of haploid and diploid cells at various stages of differentiation[67–70]. Similarly, changes in cell type abundance between homologous tissues are common across species, particularly in the testes[16,71,72], likely as a result of varying levels of sperm competition and sexual selection. For instance, species of New World Blackbirds under more intense sperm competition exhibit a greater proportion of sperm-producing tissue within their testes[71] than species subject to weaker sperm competition. In addition to differences in cell type abundance within a tissue, inter-specific single-cell analyses are starting to show that allometric

shifts might be common in many other tissues, including the brain[73–75]. Therefore, in many instances, differentially expressed genes that are identified from bulk transcriptomic approaches might simply be a product of variation in cellular heterogeneity rather than true regulatory change.

To our knowledge, only a handful of studies have directly addressed the consequences of varying tissue allometry for inferring regulatory variation across species[15,16,76,77]. To provide further insight into this issue, here we use existing single-cell expression data (scRNA-seq) for the developing chicken hypothalamus[66] to investigate this further (Fig. 3). At this stage in development (HH10), the hypothalamus is composed of three major cell types, where the FOXA1 cell type represents the greatest proportion of cells (Fig. 3A). Importantly, each cell type exhibits a distinct gene expression profile (Fig. 3B). We condensed the expression of single cells from these three cell types into a composite expression value for each loci, analogous to a bulk RNA-seq approach for the whole hypothalamus. We find that broad expression patterns across the entire tissue are not reflective of true regulatory variation, although the magnitude of this effect varies across genes (Fig. 3B), consistent with recent work in the mouse gonad[16]. This illustrates that changes in tissue composition can have profound implications for quantifying gene regulation and we urge future studies to carefully consider the composition of samples.

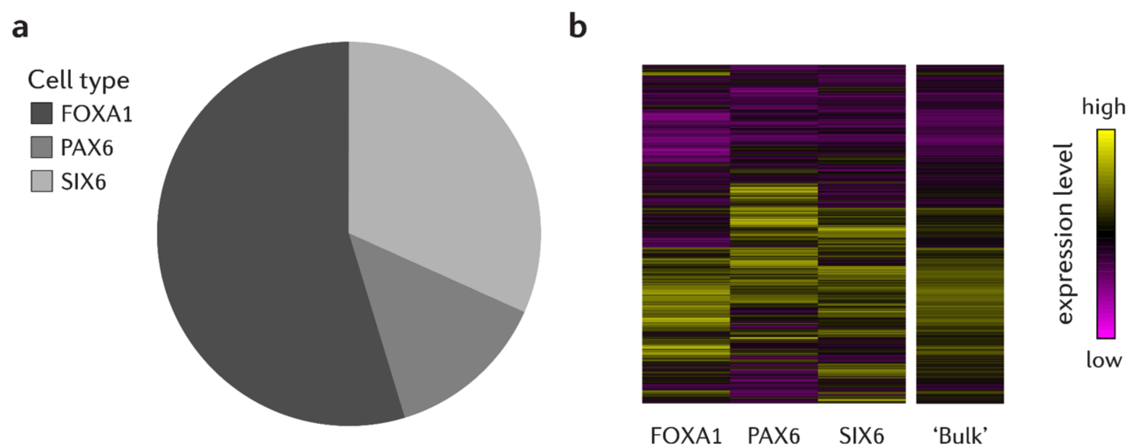

Figure 3. Deconstructing gene expression measured in bulk from heterogeneous tissue. Tissue composition and gene regulation of the chicken hypothalamus at a single developmental stage (HH10). **Panel (a)** Pie chart shows the proportion of cells in each major cell type in the hypothalamus. **Panel (b)** Heatmap shows gene expression measured across individual cells. Highly expressed genes are shown in yellow and lowly expressed genes are in pink. The first three columns of the heatmap show average expression for each gene across cells in each of the three major cell types. The final column shows average expression estimated across the entire tissue. In this case, we ignore cell identity and convert scRNA-seq data to 'bulk' data, equivalent to generating RNA-seq data from the whole tissue. Data from[66].

CHALLENGES OF INFERRING SELECTION

While the implications of varying tissue allometry for measuring regulatory change across species have been discussed[15,16] (Fig. 1 & 3), the consequences of tissue composition on inferences of expression evolution have received very little attention. Nearly all studies that test for regulatory selection use transcriptomic data generated from heterogeneous tissue, with the exception of recent work that used cell sorting to isolate distinct cell types in the mouse testes[61]. Given that changes in tissue composition across species are likely to be common, this could pose an underappreciated challenge to comparative studies of regulatory evolution. As discussed in BOX 1, 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. 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 examined and so, using a series of simulated scenarios, we directly explore how compositional shifts on a phylogeny can bias the inference of evolutionary processes.

We simulated three distinct scenarios to explore how asymmetry in tissue composition across a phylogeny can drive false model inferences of regulatory selection when applying comparative methods (Fig. 4). Specifically, 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 in each cell type and their relative abundances in the tissue, and fit a set of discrete evolutionary models to this bulk expression.

First, we describe a scenario of extreme stabilizing selection on gene regulation 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. 4A, scenario i). As predicted, the composite expression value is not reflective of single-cell expression levels nor consistent with extreme stabilizing selection (Fig. 4B, scenario i). Intuitively, 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 regulatory change (Fig. 4C, scenario i). In this instance, the false positive (i.e. Type 1 error rate) rate is around ~85% relative to when these models are run on single-cell expression levels.

This striking result suggests that shifts in tissue composition can lead to false inferences of evolutionary processes acting on gene expression in the complete absence of any regulatory change within each cell type.

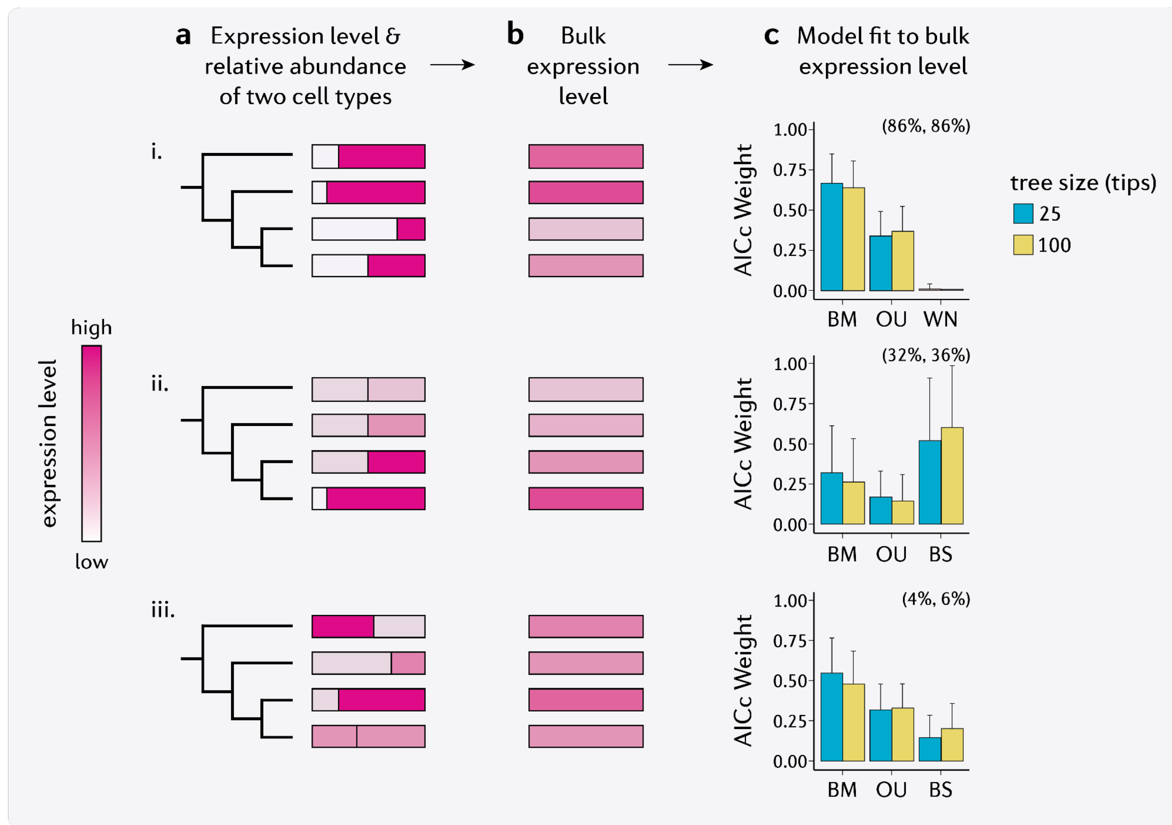

Figure 4. Inferring selection when expression level is measured from a heterogeneous tissue.
Three scenarios illustrating potential pitfalls of inferring regulatory selection at a single locus using phylogenetic approaches when expression is measured from bulk sequencing. **Panel (a)** The first column 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 1 is on the left and cell type 2 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. Shown in parenthesis (25 tips, 100 tips) are type 1 error rates for each scenario relative to when these models are fit to expression at the single-cell level. Full details: (https://github.com/Wright-lab-2021-Transcriptome-Evo/Inferring_expression_evolution_review). In scenario (i), expression values are static across the phylogeny for each cell type but cell type abundance is evolving under Brownian Motion (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 evolution.

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. 4A, 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 regulation, consistent with adaptive evolution (Fig. 4B & C, scenario ii). The scale of this bias is highly dependent on the size of the allometric shift (Fig. 5A). Where the shift leads to a single cell type dominating, the actual mode of regulatory 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 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%[71]. 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 (e.g. 0.50 to 0.60 in Fig. 5A) can result in increased type 1 errors.

Finally, we simulated a scenario where gene expression and cell type abundance are both evolving under genetic drift (Fig. 4, scenario iii). Here, we are able to recover the true signal of genetic drift more reliably (Fig. 4C, 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[78,79], expression changes are probably correlated in many instances. Interestingly, we find that this has implications for how regulatory evolution is inferred (Fig. 5B). 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. 5B). 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 regulatory selection using expression 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 regulatory evolution in the 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 regulatory variation is correlated across cell types are important prerequisites for 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 regulatory evolution where possible. However, while single-cell approaches are increasingly available, the technical demands of this approach means that they currently remain unfeasible for many species. In the meantime, we urge caution when using phylogenetic comparative approaches and recommend some steps to minimise other sources of error (BOX 2).

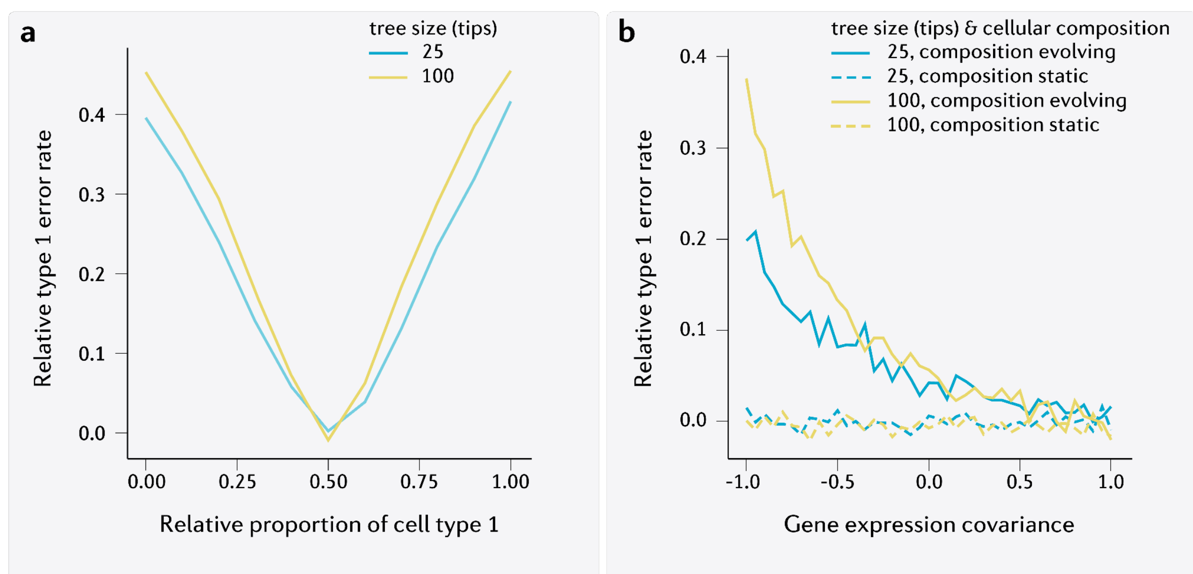

Figure 5. The magnitude of allometric shift and covariance of expression level biases the inference of regulatory evolution. Panel **(a)** The probability that regulatory selection is incorrectly inferred increases substantially with the magnitude of an allometric shift. This plot is a more detailed representation of Fig. 4 (scenario ii), where one species undergoes a shift in tissue composition, ranging from a scenario where a tissue is composed of two cell types at equal proportion to a scenario where only a single cell type is present. All other species have a tissue composition of 50:50 and expression is evolving under Brownian motion in each cell type. Panel **(b)** Covariance of expression between cell types biases inferences of selection. This plot is an extension of Fig. 4 (scenario iii). Expression is evolving under Brownian motion 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.
Full details: (https://github.com/Wright-lab-2021-Transcriptome-Evo/Inferring_expression_evolution_review).

FUTURE DIRECTIONS

To date, studies of regulatory evolution have primarily focused on differences in gene expression level. However, regulatory variation is highly complex and multi-dimensional, and below we identify major, unanswered questions in disentangling how selection acts on the entire transcriptome. As we discuss, our understanding of gene regulatory evolution will make 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.

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 regulatory diversity across species[28,80–82] with important phenotypic effects (recently reviewed[83,84]). For genes with constraints on expression levels (e.g. because of pleiotropic effects) alternative splicing may act as another adaptive mechanism of gene regulation[85]. Long-read sequencing methods have the advantage of producing full-length transcript sequences[86], 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 regulatory 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, report a higher correlation between the translatome and proteome than between the transcriptome and proteome [87]. However, this effect tends to decrease in instances surrounding functionally relevant loci, such as differentially expressed genes[88]. This indicates that in many cases, mRNA abundance does not fully capture regulatory variation, and more work is needed to understand the complex relationship between transcription and translation (e.g. mechanisms of buffering, feedback, degradation)[5,89].

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[90]. If this is common for evolutionary relevant loci, studying expression on a locus by locus basis and not through inter-locus interactions may limit our ability to understand the architectures underlying adaptive phenotypes.

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, studies primarily test for regulatory selection in a single snapshot, most often in adult tissues[1]. 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 regulatory variation with organismal ecology and natural history. This effort is twofold, as it requires understanding when and how selection acts on organisms, and how regulatory variation contributes to phenotypic responses to selection. Methods of surveying regulatory variation offer increasing precision and resolution. However, our ability to identify the evolutionary processes causing this regulatory 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 regulatory evolution[3,91–94], but non-model systems also hold promise for studying regulatory evolution under natural settings which may yield novel and more ecologically relevant findings[63,95]. 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.

ACKNOWLEDGEMENTS

This work was funded by a NERC Independent Research Fellowship to AEW (NE/N013948/1), a NERC Independent Research Fellowship to CRC (NE/T01105X/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, and Sophie Buysse 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. AEW, CRC, DHPD, JAT, JEM and PDP wrote the manuscript with input from all authors.

DATA AND CODE AVAILABILITY STATEMENT

All data has been published previously[66] and all scripts are available at (https://github.com/Wright-lab-2021-Transcriptome-Evo/Inferring_expression_evolution_review).

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[11,12,96], 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*[11] 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*[11] still estimate a type 1 error rate > 0.05.

Measurement error

Error in measuring traits 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[11,12]. 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*[11] 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 regulatory variation. Studies should endeavour to control environmental conditions so that regulatory variation across samples reflects the heritable, genetic component of expression, as has been discussed previously[4,97]. Second, it is clear that using a single mean expression value for each species can lead to spurious inferences of selection[13], making multiple replicates essential. Importantly, the OU framework has been extended to parameterise within-species variance as an error term[13,53,57] and appears to be a promising approach.

Complex patterns of trait evolution

Many phenotypic traits exhibit complex patterns of evolution and evolve at different rates across lineages[98]. While few studies have directly tested the tempo of expression change across species[55], it seems likely that gene regulation does not evolve at a constant rate but instead shifts as mutation rate, selective pressures and pleiotropic constraints[42,99,100] vary. 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[98]. Comparative methods that account for rate heterogeneity are available (discussed in[98]), 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. 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[98]. This simple step, performed using existing methods such as ARBUTUS[101] or in the probabilistic language RevBayes[102], 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[103].

Consider tissue composition

By directly comparing regulatory variation across equivalent cell types, comparative single-cell transcriptomics (scRNA-seq) can circumvent problems arising when expression is measured from heterogeneous tissue (Fig. 1 & 3). However, scRNA-seq is not yet feasible for many non-model organisms as it is necessary to isolate and process single cells immediately after harvesting tissue. Although tissue dissociation and storage techniques are being developed, bulk transcriptomic approaches are currently the only feasible option for many species, particularly those sampled from the wild. Accepting these difficulties, 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[77,104]. 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*[16].

GLOSSARY

Alternative splicing: a post-transcriptional modification involving the differential removal of introns, resulting in the production of multiple transcripts from a single gene.

Brownian motion (BM) model: a model of neutral evolution via unconstrained, random fluctuations in trait values.

Gene regulation: all pre and post-transcriptional mechanisms involved in controlling the level of gene expression.

Genome: the complete set of genetic elements encoded by the entire DNA sequence of an organism.

Ornstein-Uhlenbeck (OU) model: a modified Brownian Motion model of evolution with random fluctuations in trait values constrained towards a single optimum value.

Proteome: the set of proteins produced in a specific tissue or cell type at a particular time.

Ribo-Seq: a translatome profiling technique that involves sequencing transcripts bound and being actively translated by ribosomes.

Transcriptome: the set of RNA molecules produced by the genome in a specific tissue or cell type at a particular time.

Translatome: the set of mRNA molecules being actively translated in a specific tissue or cell type at a particular time. The regulation of the translatome determines the formation of the proteome.

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. 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).
4. 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).
5. Signor, S. A. & Nuzhdin, S. V. The evolution of gene expression in cis and trans. *Trends Genet.* **34**, 532–544 (2018).
6. Fay, J. C. & Wittkopp, P. J. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**, 191–199 (2008).
7. Khaitovich, P., Enard, W., Lachmann, M. & Pääbo, S. Evolution of primate gene expression. *Nat. Rev. Genet.* **7** 693–702 (2006).
8. Bedford, T. & Hartl, D. L. Optimization of gene expression by natural selection. *Proc. Natl. Acad. Sci. USA* **106**, 1133–1138 (2009).
9. Whitehead, A. & Crawford, D. L. Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. USA* **103**, 5425–5430 (2006).
10. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351 (1997).
11. 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).
12. 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).
13. 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).

14. Rohlfs, R. V. & Nielsen, R. Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Syst. Biol.* **64**, 695–708 (2015).
15. Montgomery, S. H. & Mank, J. E. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol.* **25**, 5114–5128 (2016).
16. Hunnicutt, K. E., Good, J. M. & Larson, E. L. Unraveling patterns of disrupted gene expression across a complex tissue. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.08.451646 (2021).
17. Nourmohammad, A. *et al.* Adaptive Evolution of Gene Expression in Drosophila. *Cell Rep.* **20**, 1385–1395 (2017).
18. 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).
19. Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261–266 (2002).
20. Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol.* **2**, E132 (2004).
21. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**, 138–144 (2003).
22. Gilad, Y., Oshlack, A. & Rifkin, S. A. Natural selection on gene expression. *Trends Genet.* **22**, 456–461 (2006).
23. 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).
24. Hudson, R. R., Kreitman, M. & Aguadé, M. A Test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
25. 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).
26. 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).
27. Somel, M. *et al.* Transcriptional neoteny in the human brain. *Proc. Natl. Acad. Sci. USA* **106**,

5743–5748 (2009).

28. 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).

29. 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. U. S. A.* **109**, 8207–8211 (2012).

30. 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).

31. Enard, W. Intra- and Interspecific Variation in Primate Gene Expression Patterns. *Science* **296**, 340–343 (2002).

32. 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).

33. 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).

34. 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).

35. 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).

36. Denver, D. R. *et al.* The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. *Nat. Genet.* **37**, 544–548 (2005).

37. Huang, W. *et al.* Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *Elife* **5**, (2016).

38. Fraser, H. B. Detecting selection with a genetic cross. *Proc. Natl. Acad. Sci. USA* **117**, 22323–22330 (2020).

39. 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).

40. Mähler, N. *et al.* Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.* **13**, e1006402 (2017).

41. Kohn, M. H., Shapiro, J. & Wu, C.-I. Decoupled differentiation of gene expression and coding sequence among *Drosophila* populations. *Genes Genet. Systems* **83**, 265–273 (2008).
42. Papakostas, S. *et al.* Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat. Commun.* **5**, 4071 (2014).
43. 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).
44. Blanc, J., Kremling, K. A. G., Buckler, E. & Josephs, E. B. Local adaptation contributes to gene expression divergence in maize. *G3* **11**, (2021).
45. 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).
46. 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).
47. Felsenstein, J. Phylogenies and the Comparative Method. *Am. Nat.* **125**, 1–15 (1985).
48. 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).
49. Felsenstein, J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471–492 (1973).
50. 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).
51. Gu, X. Statistical framework for phylogenomic analysis of gene family expression profiles. *Genetics* **167**, 531–542 (2004).
52. Butler, M. A. & King, A. A. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *Am. Nat.* **164**, 683–695 (2004).
53. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
54. Kalinka, A. T. *et al.* Gene expression divergence recapitulates the developmental hourglass

model. *Nature* **468**, 811–814 (2010).

55. El Taher, A. *et al.* Gene expression dynamics during rapid organismal diversification in African cichlid fishes. *Nat. Ecol. Evol.* **5**, 243–250 (2021).
56. Chen, J. *et al.* A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res.* **29**, 53–63 (2019).
57. 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).
58. 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).
59. 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).
60. Gillard, G. B. *et al.* Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol.* **22**, 103 (2021).
61. 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).
62. Groen, S. C. *et al.* The strength and pattern of natural selection on gene expression in rice. *Nature* **578**, 572–576 (2020).
63. Ahmad, F. *et al.* The strength and form of natural selection on transcript abundance in the wild. *Mol. Ecol.* **30**, 2724–2737 (2021).
64. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
65. Estermann, M. A. *et al.* Insights into Gonadal Sex Differentiation Provided by Single-Cell Transcriptomics in the Chicken Embryo. *Cell Rep.* **31**, 107491 (2020).
66. Kim, D. W. *et al.* Single-cell analysis of early chick hypothalamic development reveals that hypothalamic cells are induced from prethalamic-like progenitors. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.09.438683 (2021).
67. Niu, W. & Spradling, A. C. Two distinct pathways of pregranulosa cell differentiation support

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

68. 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)
69. Hermann, B. P. *et al.* The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Rep.* **25**, 1650–1667.e8 (2018).
70. Green, C. D. *et al.* A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev. Cell* **46**, 651–667.e10 (2018).
71. 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).
72. 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).
73. La Manno, G. *et al.* Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19 (2016).
74. 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).
75. 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).
76. 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).
77. 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).
78. 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).
79. Prud'homme, B. *et al.* Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006).
80. Brown, J. B. *et al.* Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399 (2014).

81. 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).

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

83. Gómez-Redondo, I., Planells, B., Navarrete, P. & Gutiérrez-Adán, A. Role of Alternative Splicing in Sex Determination in Vertebrates. *Sex Dev.* 1–11 (2021).

84. Singh, P. & Ahi, E. P. The importance of alternative splicing in adaptive evolution. Preprint at EcoEvoRxiv https://doi.org/10.32942/osf.io/wak9g (2021).

85. 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).

86. 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).

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

88. 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).

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

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

91. 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).

92. 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).

93. 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).

94. 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).
95. 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).
96. Ho, L. S. T. & Ané, C. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* **5**, 1133–1146 (2014).
97. 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).
98. 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).
99. Allen, S. L., Bonduriansky, R. & Chenoweth, S. F. Genetic constraints on microevolutionary divergence of sex-biased gene expression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, (2018).
100. Dean, R. & Mank, J. E. Tissue Specificity and Sex-Specific Regulatory Variation Permit the Evolution of Sex-Biased Gene Expression. *Am. Nat.* **188**, E74–84 (2016).
101. Pennell, M. W., FitzJohn, R. G., Cornwell, W. K. & Harmon, L. J. Model Adequacy and the Macroevolution of Angiosperm Functional Traits. *Am. Nat.* **186**, E33–50 (2015).
102. Hohna, S. et al. Probabilistic graphical model representation in phylogenetics. *Syst. Biol.* **63**, 753–771 (2014).
103. 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).
104. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).