This approach illustrates the utility of accounting for correlated error in a statistical inference framework. The ability to make inferences on individual genes is an important advantage in carefully measuring the extent of compensatory evolution. Indeed, any time estimates of *cis* and *trans* are considered jointly to make biological conclusions, correlated error should be considered, not just in cases of compensatory evolution. Modern data sets should be even better suited to addressing such questions, as lower sequencing costs allow us to achieve higher and higher replication, not only eliminating the correlated error problem but also improving statistical power. Indeed, it would be irresponsible not to replicate parental and hybrid treatments in future ASE studies.

Acknowledgments

J.J.E. thanks Daniel Rokhsar for thoughtful comments. This work was supported by National Institutes of Health grant R01GM123303-1 (J.J.E.) and University of California, Irvine, setup funds (J.J.E).

Supplemental Information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tig.2018.11.003.

[1]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA
[2]Center for Complex Biological Systems, University of California, Irvine, CA, USA
@Twitter: @JJ_Emerson

*Correspondence: jje@uci.edu (J.J. Emerson).

https://doi.org/10.1016/j.tig.2018.11.003

References

1. Lee, T.I. and Young, R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell* 152, 1237–1251
2. Lappalainen, T. *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506–511
3. Carroll, S.B. (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36
4. Emerson, J.J. and Li, W.-H. (2010) The genetic basis of evolutionary change in gene expression levels. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2581–2590
5. Wittkopp, P.J. *et al.* (2004) Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430, 85–88
6. Wittkopp, P.J. *et al.* (2008) Regulatory changes underlying expression differences within and between Drosophila species. *Nat. Genet.* 40, 346–350
7. McManus, C.J. *et al.* (2010) Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res.* 20, 816–825
8. Emerson, J.J. *et al.* (2010) Natural selection on *cis* and *trans* regulation in yeasts. *Genome Res.* 20, 826–836
9. Fraser, H.B. (2018) Improving estimates of compensatory *cis-trans* regulatory divergence. *Trends Genet.* Published online September 27, 2018. http://dx.doi.org/10.1016/j.tig.2018.09.003
10. Gierliński, M. *et al.* (2015) Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. *Bioinformatics* 31, 3625–3630

Forum

Improving Estimates of Compensatory *cis–trans* Regulatory Divergence

Hunter B. Fraser[1,2,*,@]

Interspecific hybrids have played a key role in research on gene expression regulation. A growing number of studies have measured genome-wide allele-specific expression in hybrids and observed that *cis*-regulatory changes often oppose *trans*-acting changes affecting the same genes, suggesting stabilizing selection for compensatory changes. However, the most common method for estimating these effects is biased, producing artifactual patterns of compensatory evolution. Here I introduce a simple modification leveraging biological replicates that ameliorates the bias.

First-generation hybrids between divergent lineages have been an invaluable model to study allele-specific expression (ASE), where one allele of a gene is more highly expressed than the other. Hybrid ASE specifically reflects *cis*-acting differences between alleles, since the two alleles of each gene are exposed to the same *trans*-acting regulatory environment. Quantifying *cis*-regulatory divergence can help pinpoint genes and pathways underlying adaptive traits, as well as reveal large-scale patterns of regulatory evolution [1,2].

Using high-throughput RNA-sequencing (RNA-seq), genome-wide ASE has been measured in a diverse menagerie of hybrids [1]. A popular analysis of these data compares hybrid ASE to expression differences between the two parental species; since the parental difference reflects both *cis*- and *trans*-acting divergence, the *trans* effects impacting each gene can be estimated as the parental difference minus the *cis* effect (Figure 1A). A surprisingly consistent result of these comparisons has been that compensatory changes, where *cis* and *trans* effects on a specific gene differ in sign, are far more common than reinforcing changes [3–11]. Indeed, a recent review highlighted this as a major unsolved puzzle, suggesting mechanisms such as stabilizing selection, feedback, or transvection to explain its ubiquity [1].

My colleagues and I previously noted that this approach is intrinsically biased: any error in estimating *cis* effects will introduce an artifactual negative correlation with *trans* effects [12] (Figure 1B). In a hybrid between parental species A and B, any error that overestimates the A/B ASE ratio will lead to underestimation of the A/B *trans* ratio. This leads to the undesirable situation where greater error in ASE estimates will lead to stronger *cis–trans* correlations. For example, if ASE ratios are estimated with 50% error, then even with no true correlation between *cis* and *trans* divergence, the observed (artifactual) *cis–trans* correlation will be $r \approx -0.5$ (Box 1). Although this concern has been reiterated by others [13], no solution has yet been proposed.

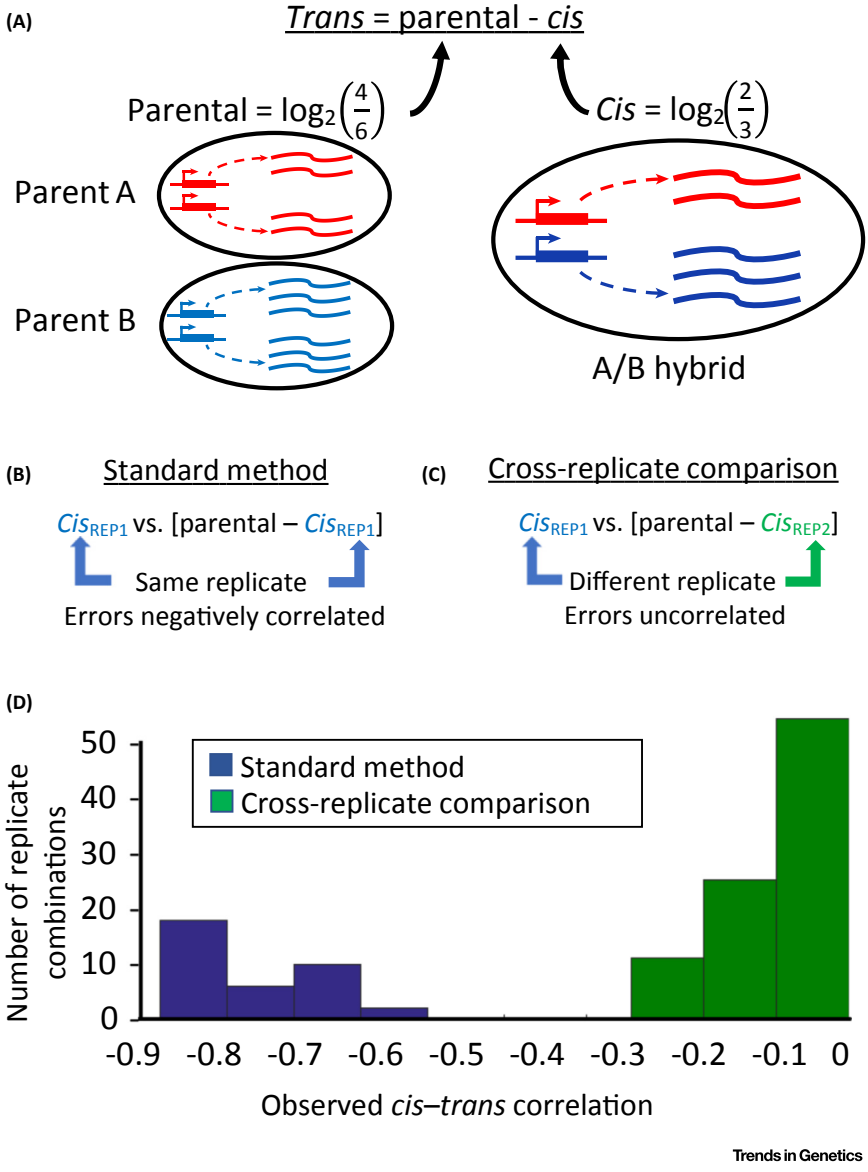

Figure 1. Improving Estimates of *cis–trans* Divergence. (A) Estimating *trans*-acting divergence from the difference between parental expression and hybrid ASE. Parental divergence is estimated from the ratio of parental expression levels, and is assumed to be the product of *cis* and *trans* effects (additive in log-space). Figure adapted from [14]. (B) The standard method of *cis–trans* comparison leads to artifactual negative correlation when *cis* estimates have any error. Note that 'REP1' could represent either a single replicate or an average of multiple replicates. (C) The proposed method of cross-replicate comparison is not inflated by random measurement error. (D) Histogram showing the difference between the two methods applied to the same data set [4]. Each *cis–trans* correlation is based on one pair of hybrid/parental replicates (36 pairs for the standard method and 90 for cross-replicate comparison, each with over 4000 informative genes).

I propose a simple solution to this bias: cross-replicate comparison. Instead of using the same ASE measurements for both the *trans* estimation and the subsequent *cis–trans* comparison, performing biological replicate ASE measurements allows one to use one replicate for *trans* estimation, and another for *cis–trans* comparison (Figure 1C). This is not subject to the same bias as the standard method, since any random error (e.g., due to low read counts) from one replicate will generally not be shared by another. Overestimation of the A/B ASE ratio in one replicate will still lead to underestimation of the A/B *trans* ratio, but this should not be correlated with the ASE ratio of an independent replicate, thus eliminating the bias.

To test this approach, I applied it to an extensively replicated study of two inbred mouse strains, with RNA-seq in six replicates of each parental line and six of each reciprocal hybrid (24 total samples; Box 1) [4]. Performing a standard *cis–trans* comparison (Figure 1B), all replicates showed strong negative correlation ranging from Pearson's $r = -0.57$ to -0.82 (Figure 1D, blue; mean $r = -0.71$). However, performing the cross-replicate approach (Figure 1C), these correlations were far weaker (Figure 1D, green; Pearson's $r = -0.21$ to -0.004, mean $r = -0.079$). This suggests that the cross-replicate design eliminates much of the negative bias when comparing *cis* versus *trans* divergence.

The cross-replicate approach can be applied with as few as two replicate hybrid samples and one sample from each parent. For example, this was the number of replicates in a study of two *Saccharomyces cerevisiae* yeast strains and their hybrids [13]. With the standard *cis–trans* approach, the two ASE replicates yielded *cis–trans* correlations of $r = -0.40$ and -0.37. However, the cross-replicate design yielded correlations of $r = -0.02$ and -0.002. Notably, these insignificant estimates are more consistent with results of an earlier study of the same two strains that found a slight excess of reinforcing *cis–trans* effects, using expression quantitative trait locus (eQTL) mapping, which is not subject to the negative bias discussed here [12].

Box 1. Methods

When no true correlation exists between *cis* and *trans* changes, no *cis* × *trans* interactions exist, and *cis* and *trans* changes have equal variance, the expected *cis*–*trans* correlation can be estimated as follows, where *cis* is the true $\log_2$ ASE ratio, *trans* is the true $\log_2$ *trans* ratio, parental is the true $\log_2$ parental ratio, and ε is an error term:

$trans = \text{parental} - cis$

$\text{observed } cis = cis + \varepsilon$

$\text{observed } trans = \text{parental} - (cis + \varepsilon) = trans - \varepsilon$

$\text{var}(cis + \varepsilon) = \text{var}(cis) + \text{var}(\varepsilon)$

$\text{var}(trans - \varepsilon) = \text{var}(trans) + \text{var}(\varepsilon) = \text{var}(cis) + \text{var}(\varepsilon)$

$\text{cov}(cis, trans) = 0$

$\text{cov}(cis + \varepsilon, trans - \varepsilon) = \text{cov}(cis, trans) + \text{cov}(\varepsilon, -\varepsilon) = 0 - \text{var}(\varepsilon)$

$\text{corr}(cis + \varepsilon, trans - \varepsilon) = \text{cov}(cis + \varepsilon, trans - \varepsilon)/\text{sqrt}[\text{var}(cis + \varepsilon) \times \text{var}(trans - \varepsilon)] = -\text{var}(\varepsilon)/[\text{var}(cis) + \text{var}(\varepsilon)]$

The numerator of the final equation leads to the artifactual negative correlation; if instead the errors are uncorrelated (as in Figure 1C in main text), the numerator becomes zero. Although it may appear that cross-replicate comparison could increase error, this is not the case (see Supplemental Information online).

Applying the equations above, if *cis* effects are estimated with 50% error (i.e., $r \approx 0.7$ between true ASE and estimated ASE), then $\text{var}(\varepsilon) \approx \text{var}(cis)$, and the expected Pearson's correlation ≈ -0.5. This level of error is not unrealistic; for example, the average ASE correlation between replicate hybrids is $r = 0.32$ and $r = 0.47$ in the mouse and yeast data, respectively, suggesting greater than 50% error per replicate. This model is meant only as an approximation to reality; in practice, many other factors (e.g., error in parental estimates) will affect the correlation as well.

Data analysis was performed on raw read counts, requiring at least five reads per allele in hybrid samples or per gene in the parental samples to include that gene in the analysis. Results were similar at other cutoffs (e.g., requiring 10 reads in the mouse data, mean $r = -0.67$ for the standard method and $r = -0.071$ for cross-replicate comparison). Hybrids from only one direction of the mouse cross were included, to avoid confounding effects of imprinted genes in the reciprocal crosses. I analyzed each replicate separately to maximize the number of comparisons (see Figure 1D in main text), but to achieve a single estimate of *cis*–*trans* concordance it is also possible to combine replicates, as long as no replicates are used for both *cis* and *trans* effect estimation.

Although many authors have used discrete cutoffs to classify genes into distinct categories of compensatory or reinforcing changes, I chose to focus on correlation due to its generality. These analyses are not meant to match the details of the previously published analyses, and therefore should not be interpreted as refutation of their specific results; rather my goal was to illustrate a more general issue with the method itself.

In sum, the bias inherent in a widely used method has led to overestimation of the ubiquity of compensatory *cis*–*trans* evolution. Although cross-replicate comparison controls for the effects of random error in ASE estimates, any bias that is shared between replicates (e.g., allelic mapping bias) could still lead to an artifactual negative correlation; therefore, methods that independently estimate *cis* and *trans* effects, such as eQTL mapping, may still be preferable. Whether previous reports of compensatory evolution can be entirely explained by this bias will be a key question for future work.

Acknowledgments

The author thanks M. Feldman, D. Petrov, J. Pritchard, and the Fraser Lab for feedback. This work was funded by NIH grant 2R01GM097171-05A1.

Supplemental Information

Supplemental information associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tig.2018.09.003.

[1]Department of Biology, Stanford University, Stanford, CA, USA
[2]Lab website: https://web.stanford.edu/group/fraserlab/
@Twitter: @FraserLab

*Correspondence: hbfraser@stanford.edu (H.B. Fraser).
https://doi.org/10.1016/j.tig.2018.09.003

References

1. Signor, S.A. and Nuzhdin, S.V. (2018) The evolution of gene expression in *cis* and *trans*. *Trends Genet.* 34, 532–544
2. Fraser, H.B. (2011) Genome-wide approaches to the study of adaptive gene expression evolution. *BioEssays* 33, 469–477
3. Schaefke, B. *et al.* (2013) Inheritance of gene expression level and selective constraints on *trans*-and *cis*-regulatory changes in yeast. *Mol. Biol. Evol.* 30, 2121–2133
4. Goncalves, A. *et al.* (2012) Extensive compensatory *cis*-*trans* regulation in the evolution of mouse gene expression. *Genome Res.* 22, 2376–2384
5. Metzger, B.P.H. *et al.* (2017) Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol. Evol.* 9, 843–854
6. Tirosh, I. *et al.* (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324, 659–662
7. Fear, J.M. *et al.* (2016) Buffering of genetic regulatory networks in *Drosophila melanogaster*. *Genetics* 203, 1177–1190
8. Carlson, C.H. *et al.* (2017) Dominance and sexual dimorphism pervade the *Salix purpurea* L. transcriptome. *Genome Biol. Evol.* 9, 2377–2394
9. Mack, K.L. *et al.* (2016) Gene regulation and speciation in house mice. *Genome Res.* 26, 451–461
10. Coolon, J.D. *et al.* (2014) Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* 24, 797–808
11. Shi, X. *et al.* (2012) *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat. Commun.* 3, 950
12. Fraser, H.B. *et al.* (2010) Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2977–2982
13. Albert, F.W. *et al.* (2014) Genetic influences on translation in yeast. *PLoS Genet.* 10, e1004692
14. Artieri, C.G. and Fraser, H.B. (2014) Evolution at two levels of gene expression in yeast. *Genome Res.* 24, 411–421