OXFORD GENETICS

GENETICS, 2026, iyag233
https://doi.org/10.1093/genetics/iyag233
Advance Access Publication Date: 31 August 2026
Investigation

Gene expression is stable despite widespread *cis* and *trans* regulatory divergence in *Saccharomyces* yeasts

Danithza S. Rojas [1], Gregory I. Lang [1,*]

[1]Department of Biological Sciences, Lehigh University, Bethlehem PA 18015, United States

*Corresponding author: Department of Biological Sciences, Lehigh University, Iacocca Hall, 111 Research Drive, Bethlehem PA 18015, United States. Email: glang@lehigh.edu

Regulatory evolution can alter phenotypes, but *cis*- and *trans*-regulatory mechanisms may also diverge extensively while total transcript abundance remains stable. Comparisons of parental expression with allele-specific expression in F1 hybrids provide a framework for separating *cis*- and *trans*-regulatory effects because both parental alleles are measured in a shared *trans*-regulatory environment. Here, we analyzed RNA sequencing data from *Saccharomyces cerevisiae*, *S. paradoxus*, and their F1 hybrid. Among the 4,164 genes with sufficient allele-specific support for strict classification, 2,134 (51.2%) showed detectable *cis*- and/or *trans*-regulatory divergence. However, hybrid expression remained largely conserved, with 81.5% of genes not significantly different from either parent. Compensatory *cis*–*trans* divergence predominated over reinforcing divergence; cross-replicate estimation reduced the apparent magnitude of this excess, but opposite-sign effects remained predominant in all 20 nonoverlapping replicate comparisons. To connect gene expression to genome sequence, we analyzed the strongly *cis*-diverged locus *LYS2* and found species differences in promoter architecture, including an *S. cerevisiae*-specific AT-rich insertion, altered spacing among candidate regulatory features, and a promoter-proximal TATA-like element unique to *S. cerevisiae*. Sequence-based nucleosome prediction suggests that these differences create a broader promoter-proximal nucleosome-depleted region in *S. cerevisiae* than in *S. paradoxus*. We also quantified allele-resolved intron retention and found that it was broadly conserved, with only rare locus-specific hybrid-associated shifts. Together, these results show that regulatory divergence is widespread but often buffered in the hybrid, whereas intron-retention divergence is comparatively limited.

Keywords: gene expression; *cis* regulation; *trans* regulation; *Saccharomyces*; splicing; Yeast2026

Introduction

Gene expression levels can remain stable even when the regulatory mechanisms that produce them have diverged. This distinction is central to understanding regulatory evolution. Genetic variation can affect gene expression in one of two general ways. *Cis*-regulatory elements influence expression locally and are physically linked to the gene they regulate. *Trans*-acting factors are diffusible regulators that can be encoded anywhere in the genome and can regulate any number of genes. Variation in *cis*-regulatory or *trans*-acting factors may alter the architecture of gene regulation without necessarily producing large changes in total transcript abundance. In some cases, opposing *cis* and *trans* effects can produce compensatory regulatory divergence, allowing transcript output to remain stable even as the underlying regulatory mechanisms diverge. Such hidden regulatory divergence is important because it can reveal how transcript output is maintained over evolutionary time while the underlying *cis* and *trans* components continue to change. Therefore, distinguishing between local *cis*-regulatory effects and diffusible *trans*-regulatory effects is essential for understanding not only how transcriptomes diverge, but also how gene expression stability is preserved despite regulatory evolution (King and Wilson 1975; Carroll 2005; Wray 2007; Wittkopp and Kalay 2012; Signor and Nuzhdin 2018; Hill et al. 2021).

The use of hybrids between closely related species provides a framework for separating the contributions of *cis*- and *trans*-regulatory elements because the two parental alleles coexist within the same nucleus and are exposed to the same pool of diffusible *trans*-acting factors. In this shared *trans*-regulatory environment, allele-specific expression (ASE) differences between the two parental alleles are interpreted as evidence of *cis*-regulatory divergence (Cowles et al. 2002). In contrast, *trans* effects are inferred by comparing total-expression differences between the parental species with the ASE pattern observed in the hybrid (Wittkopp et al. 2004; Tirosh et al. 2009; Schaefke et al. 2013). Previous studies in *Saccharomyces* have documented genome-wide *cis* and *trans* differences, dominance of many *trans* effects, complex *trans*-regulatory architectures, and divergence in expression dynamics across environments (Tirosh et al. 2009; Schaefke et al. 2013; Metzger et al. 2017; Albert et al. 2018; Krieger et al. 2020; Shih and Fay 2021). Research across diverse systems, including *Saccharomyces* yeast, Drosophila, and mammals, has shown that both *cis* and *trans* changes are common and often accumulate in opposing directions, a pattern consistent with compensatory evolution under stabilizing selection on transcript abundance

Received on 06 June 2026; accepted on 20 August 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of The Genetics Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

(Landry et al. 2005; McManus et al. 2010; Goncalves et al. 2012; Metzger et al. 2017; Hill et al. 2021). Recent work, however, has challenged this interpretation, showing that measurement error in the hybrid can lead to spurious negative correlations between *cis* and *trans* effects when the same sample is used to calculate both (Fraser 2019; Zhang and Emerson 2019). In highly replicated mouse and yeast datasets, using independent replicate estimates substantially weakened the negative *cis* and *trans* correlation, in some cases toward zero (Fraser 2019; Zhang and Emerson 2019). To our knowledge, whether opposing effects remain predominant after reducing this shared-estimation error has not been directly tested in an interspecific *S. cerevisiae*/*S. paradoxus* hybrid. Beyond total expression, allele-resolved intron retention provides a complementary test of whether diverged parental intronic and splice-site sequences remain similarly processed when each allele is placed in the shared F1 *trans* environment.

Here, we combine *cis*–*trans*-regulatory decomposition with analysis of total expression in the F1 hybrid between *S cerevisiae* and *S. paradoxus*. These closely related yeasts remain cross-compatible while harboring substantial sequence divergence. For the 5,366 one-to-one orthologs analyzed here, the coding sequences contain on average 99.2 single-nucleotide polymorphisms (SNPs) per kb, whereas the 600-bp upstream promoter regions contain on average 117.3 SNPs per kb. Rather than asking only whether genes differ in expression between *S. cerevisiae* and *S. paradoxus*, we ask how those differences are encoded, including as *cis*, *trans*, compensatory, or reinforcing effects, and whether expression in the hybrid remains within the parental range or reveals misregulation. To address these questions, we integrate total RNA sequencing, ASE, and allele-resolved intron-retention analysis in *S. cerevisiae*, *S. paradoxus*, and their F1 hybrid. This approach allows us to ask how regulatory divergence is partitioned into *cis* and *trans* components, how these architectures shape hybrid expression patterns, and whether promoter architecture or intron processing reveals additional layers of divergence. We find widespread compensatory *cis*–*trans* evolution together with generally stable hybrid expression, rare and locus-specific hybrid-associated shifts in intron retention, and species-specific promoter architecture at the strongly *cis*-diverged *LYS2* locus. Finally, using all possible three-vs-two F1 replicate partitions, we show that separating the ASE estimates used for *cis* and *trans* reduces the apparent compensatory excess but does not eliminate the predominance of opposing effects.

Materials and methods

Yeast strains and media

The yeast strains derive from prototrophic and heterothallic versions of *S. cerevisiae* S288C and *S. paradoxus* CBS432 and contain the pCLB2::3xHA::KanMX cassette upstream of *MSH2* and *SGS1* as described previously (Martínez and Lang 2026). The F1 hybrid was generated by mating an *S. cerevisiae* *MAT*a haploid with an *S. paradoxus* *MAT*α haploid. Full strain genotypes are described in Supplementary Table 1. In preparation for RNA extraction, strains were grown in YPD (1% yeast extract, 2% peptone, and 2% glucose) at 30 °C with shaking and were harvested in mid-log phase at approximately OD600 of ∼0.6. Cells were collected by rapid centrifugation, flash-frozen in liquid nitrogen, and stored at −80 °C until RNA extraction.

RNA extraction, library preparation, and sequencing

Total RNA was extracted using hot acid phenol–chloroform followed by silica-column cleanup with the Qiagen RNeasy kit for two, three, and five biological replicates of *S. cerevisiae*, *S. paradoxus*, and the F1 hybrid, respectively. RNA quantity and integrity were assessed using Qubit fluorometry and Bioanalyzer analysis, and only samples with RNA integrity numbers of at least 7 were used for library preparation. Poly(A) + RNA was enriched from 1 to 2 µg of total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490). Strand-specific RNA-seq libraries were prepared with the NEBNext Ultra II Directional RNA Library Prep Kit (NEB E7760) according to the manufacturer's deoxyuridine triphosphate (dUTP)-based protocol. Indexed libraries were amplified using ten PCR cycles, size-selected with AMPure beads for an approximate 200-bp insert distribution and quantified by Qubit fluorometry and Bioanalyzer analysis. Libraries were pooled and sequenced on an Illumina NovaSeq platform in paired-end mode. Across the ten libraries included in the parental and F1 analyses, sequencing depth ranged from 7.49 to 64.59 million input read pairs per sample.

Read processing and alignment

Adapters and low-quality bases were removed with Trim Galore/Cutadapt, and read quality was summarized across samples with MultiQC (Martin 2011; Ewels et al. 2016). We constructed a combined reference genome by concatenating the *S. cerevisiae* S288C and *S. paradoxus* CBS432 assemblies obtained from the Yeast Population Reference Panel/Yeast PacBio resource (Yue et al. 2017), adding species-specific prefixes to chromosome names while preserving chromosome numbering. The corresponding gene-annotation files were merged after harmonizing identifiers for one-to-one orthologs. Trimmed paired-end reads were aligned to the combined reference genome with STAR in two-pass, splice-aware mode (Dobin et al. 2013). Across the ten libraries included in the final analysis, 89.11% to 91.27% of input read pairs mapped uniquely, whereas 7.46% to 9.60% mapped to multiple loci. Alignment files were coordinate-sorted, and PCR duplicates were marked. The resulting Binary Alignment/Map (BAM) files were used for gene-level quantification, ASE analysis, and intron-retention profiling.

Gene quantification and expression filtering

Gene-level counts were obtained with featureCounts from the Subread package using exon features grouped by gene ID in the merged annotation (Liao et al. 2014). Counting was performed in paired-end mode, required both mates to be aligned, and excluded chimeric read pairs. Multimapping fragments were not counted because the featureCounts-M option was not enabled. Gene-level counting was performed without using strand information. PCR duplicates had been marked in the input BAM files.

One-to-one orthologs were assigned shared gene identifiers across species. The gene-level count matrix contained 5,386 orthologs and was used for total-expression analyses, including parental differential expression and hybrid expression inheritance. For the *cis*/*trans* analysis, genes were additionally required to have at least 1 count per million in at least two samples, leaving 5,243 expression-tested orthologs. ASE analysis imposed further requirements for informative interspecies SNPs, allele-informative read depth, and replicate-level ASE support. Of the 5,243 expression-tested genes, 4,164 met the strict allele-resolution criteria and received a *cis*-/*trans*-regulatory classification; genes that did not meet these requirements were left unclassified rather than assigned to the conserved class.

Because total-expression quantification and ASE analysis have different information requirements, the analyzable gene sets were not identical. Hybrid inheritance was first classified across the 5,386 genes represented in the joint parental–hybrid DESeq2 count matrix. Analyses relating hybrid inheritance to *cis*-/*trans*-regulatory

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

architecture were restricted to the 5,243 genes included in the *cis/trans*-tested set, and regulatory-class-specific summaries included only genes that received a strict *cis/trans* classification. All downstream statistical analyses were performed in R.

Differential expression between parents

Differential expression between the *S.c* and *S.p* diploid parents was tested with DESeq2 (Love et al. 2014). Size factors were estimated using the median-of-ratios method, and gene-wise dispersions were fit and shrunk with the package's empirical Bayes framework. Wald tests were used for inference, and *P* values were adjusted for multiple testing by the Benjamini–Hochberg false discovery rate (FDR) procedure. Unless otherwise noted, genes were considered significantly divergent at $P_{adj} < 0.005$ and absolute log2 fold-change ($|LFC| \geq 0.58$, corresponding to an approximately 1.5-fold difference. In figures where low-count genes required more stable effect-size estimates, we used apeglm shrinkage for LFC visualization without changing significance calls (Zhu et al. 2019). For interpretation, we recorded the DESeq2 baseMean for each gene and calculated group-specific mean expression values for *S.c*, *S.p*, and the hybrid.

ASE and *cis/trans* decomposition

ASE in the F1 hybrid was used to estimate *cis*-regulatory divergence. To reduce reference-mapping bias, reads overlapping known *S. cerevisiae/S. paradoxus* SNPs were processed using WASP, a tool for unbiased allele-specific read mapping and molecular quantitative trait locus discovery (Castel et al. 2015; van de Geijn et al. 2015). For each allele-informative read, the overlapping allele was computationally swapped and the read was remapped. Reads that failed to map consistently after allele swapping were discarded.

Allele-specific counts were obtained from the WASP-filtered alignments using GATK ASEReadCounter at known biallelic *S. cerevisiae/S. paradoxus* SNPs (McKenna et al. 2010). SNP sites were retained when supported by at least 20 reads. Counts were aggregated by gene and biological replicate. An F1 replicate was considered informative for a gene when at least 20 retained SNP sites were available for that gene, and strict regulatory classification required informative support in at least three F1 biological replicates.

For each gene in each informative replicate, deviation of the *S. cerevisiae* and *S. paradoxus*-allele counts from a 50:50 expectation was evaluated using a two-sided binomial test. Replicate-level results were converted to signed Z scores according to the direction of allelic imbalance, defined as log2(*S. cerevisiae* allele/*S. paradoxus* allele), and combined across replicates using Stouffer's method. Meta-analysis *P* values were adjusted for multiple testing.

The *cis* effect was defined as

$$LFC_{ASE} = \log_2\left(\frac{S.\ cerevisiae\ \text{allele}}{S.\ paradoxus\ \text{allele}}\right).$$

The parental-expression difference was defined as

$$LFC_{parent} = \log_2\left(\frac{S.\ cerevisiae\ \text{parent}}{S.\ paradoxus\ \text{parent}}\right),$$

and the inferred *trans* effect was calculated as

$$LFC_{trans} = LFC_{parent} - LFC_{ASE}.$$

Standard errors for the inferred *trans* effect were calculated by the propagation of uncertainty from the parental and ASE estimates. *Trans*-regulatory *P* values were obtained using Z tests and corrected using the Benjamini–Hochberg procedure.

A *cis* effect was considered significant when $P_{adjASE} \leq 0.005$ and $|LFC_{ASE}| \geq 0.58$. A *trans* effect was considered significant when $P_{adjtrans} \leq 0.005$ and $|LFC_{trans}| \geq 0.58$. Genes with a significant *cis* effect but no significant *trans* effect were classified as *cis*-only. Genes with a significant *trans* effect but no significant *cis* effect were classified as *trans*-only. Genes with significant *cis* and *trans* effects in the same direction were classified as reinforcing, whereas genes with significant effects in opposite directions were classified as compensatory. Genes for which neither component met both the significance and effect-size thresholds were classified as conserved.

Final strict regulatory classes were assigned only to genes that passed expression filtering and had ASE support in at least three F1 biological replicates. Expression-filtered genes that did not meet the allele-resolution requirement were left unclassified rather than assigned to the conserved class. The parental-divergence-vs-hybrid-ASE plot represented this same decomposition and was not used as a separate classification procedure. Because RNA-seq measures steady-state transcript abundance, the inferred *cis* and *trans* effects may reflect differences in RNA synthesis, RNA degradation, or both. We therefore interpret these classes as regulatory effects on steady-state mRNA abundance rather than exclusively as transcriptional effects.

Cross-replicate sensitivity analysis for shared-ASE measurement error

We generated all ten possible divisions of the five F1 hybrid replicates into groups of three and two. Both reciprocal orientations of each division were analyzed, producing 20 cross-replicate comparisons. Within each replicate subset, genes were required to have allele-specific support in at least two biological replicates and at least 20 total allele-informative reads.

Within each subset, the deviation of the *S. cerevisiae*- and *S. paradoxus*-allele counts from a 50:50 expectation was evaluated separately in each informative biological replicate using a two-sided binomial test. Replicate-level results were converted to signed *Z* scores according to the direction of allelic imbalance and combined within the subset using Stouffer's method. The split-specific ASE effect was calculated from the pooled allele counts within the subset as the log2 ratio of the *S. cerevisiae* allele count to the *S. paradoxus*-allele count, with a pseudocount of 0.5 added to each allele count.

For each reciprocal comparison, the *cis* effect was estimated using one subset of hybrid replicates, whereas the *trans* effect was calculated as the parental-expression log2 fold change minus the ASE estimate from the complementary, nonoverlapping subset. Thus, the hybrid ASE estimate used to define *cis* was not reused in the *trans* component against which it was compared. The parental-expression estimate remained shared, and the resulting *cis* and *trans* estimates were therefore not fully statistically independent.

The standard error of the cross-replicate *trans* estimate was obtained by propagating uncertainty from the parental-expression estimate and the ASE estimate from the complementary replicate subset. Cross-replicate *trans* Z statistics were calculated by dividing the *trans* estimate by its propagated standard error, and two-sided *P* values were obtained from the standard normal distribution. *Cis* and *trans* *P* values were adjusted separately within each replicate subset or reciprocal comparison using the Benjamini–Hochberg procedure. *Cis* and *trans* effects were considered significant at an adjusted *P* value of at most 0.005 and an absolute log2 fold change of at least 0.58.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

The primary cross-replicate analysis was restricted to the same 5,243 expression-tested genes used in the main *cis/trans* analysis. Unlike the full five-replicate analysis, the cross-replicate analysis did not begin by selecting genes already classified as having both *cis* and *trans* effects. Instead, *cis* significance and *trans* significance were evaluated separately within each of the 20 reciprocal comparisons. Only genes meeting the allele-specific support requirements in both complementary replicate subsets were eligible for a given comparison.

For each reciprocal comparison, genes significant for both cross-replicate-estimated components were classified as opposite-sign when the split-specific *cis* estimate and cross-replicate *trans* estimate had opposite signs and as same direction when the two estimates had the same sign. To determine whether the observed number of opposite-sign effects exceeded the expectation generated by the overall frequencies of positive and negative *cis* and *trans* estimates, we used an exact fixed-margin sign-randomization test (hypergeometric test). *Trans*-effect signs were reassigned among genes without replacement while preserving the total numbers of positive and negative *cis* and *trans* estimates within each comparison. The expected number of opposite-sign pairs was calculated from the observed marginal sign distributions, and the one-sided probability of obtaining at least the observed number of opposite-sign pairs was calculated from the corresponding hypergeometric distribution.

We also compared conventional shared-ASE *cis/trans* correlations with cross-replicate correlations. First, we calculated the conventional Pearson correlation using the ASE estimate from all five F1 biological replicates in both the *cis* and inferred *trans* components. For each reciprocal partition, we then calculated a matched-subset conventional correlation in which *cis* and *trans* used the same subset-specific ASE estimate and a cross-replicate correlation in which *cis* used one subset and *trans* used the ASE estimate from the complementary subset. Correlations were calculated across genes with informative estimates in both complementary subsets. Because the same five biological replicates were reused across the exhaustive partitions, the 20 reciprocal comparisons were summarized by their medians and ranges and were treated as a robustness distribution rather than as 20 statistically independent experiments.

Allele-specific support and sensitivity analyses

For each gene, we summarized total allele-specific read depth across the five F1 replicates, the number of replicates with allele-specific support, and the number of coding SNPs in the biallelic *S. cerevisiae*/*S. paradoxus* SNP set. Genes lacking sufficient allele-specific support were left unclassified rather than assigned to the conserved class.

To assess whether regulatory-class frequencies depended on allele resolution, class summaries were recalculated after applying increasingly stringent thresholds for coding-SNP number and total allele-specific read depth. Coding-SNP thresholds were based on the intersection of the biallelic SNP set with annotated coding regions. Allele-specific depth was calculated as the summed *S. cerevisiae*-allele and *S. paradoxus*-allele counts across the F1 replicates. The numbers of conserved, *cis*-only, *trans*-only, compensatory, and reinforcing genes were compared across the thresholded gene sets.

Hybrid expression inheritance classification

To classify hybrid expression relative to the parents, we modeled the two diploid parents and the hybrid jointly in DESeq2 and extracted the contrasts hybrid vs *S.c* and hybrid vs *S.p*, again using LFC shrinkage for stability. We then assigned each gene to an inheritance category using a scheme that required both statistical significance and logical placement relative to parental means.

Nondifferential genes were those for which hybrid mean expression lay within the parental range and did not differ significantly from either parent. Additive genes were those for which the hybrid mean lay strictly between the two parents and differed significantly from both, with opposite-sign LFCs relative to each parent. Dominant genes were those for which the hybrid did not differ significantly from one parent but did differ significantly from the other, consistent with expression matching one parental state. Overdominant or underdominant (transgressive) genes were those for which hybrid expression was significantly above both parents or below both parents, respectively, with $P_{adj} < 0.005$ and $|LFC| \geq 0.58$ in both hybrid-vs-parent contrasts.

If a hybrid mean fell outside the parental range but failed the significance criteria against one or both parents, we labeled the gene as outside but not significant rather than calling it transgressive. This avoided overcalling overdominance or underdominance and ensured that only robust cases of hybrid misexpression were placed in transgressive classes.

Hybrid expression inheritance categories were assigned independently of the *cis*-/*trans*-regulatory classes. After both analyses were completed, the two annotations were intersected to determine how genes assigned to each *cis*-/*trans*-regulatory class were distributed across hybrid expression modes.

Gene ontology enrichment analysis

To test functional bias across regulatory classes, we performed Gene Ontology (GO) enrichment analysis focused on Biological Process and Molecular Function annotations using the GO resource (The Gene Ontology Consortium 2019). For each regulatory class, the query set consisted of genes assigned to that class, and the background set consisted of all 5,243 tested genes. Enrichment was assessed with one-sided Fisher's exact tests and corrected for multiple testing using the Benjamini–Hochberg procedure. Terms with FDR < 0.01 were considered significant, and we report representative enrichments together with their fold-enrichment values and adjusted *P* values. For the follow-up analysis of the *trans*-only class, we further separated *trans*-only genes by hybrid inheritance mode into *S.p*-dominant and *S.c*-dominant subsets and repeated GO enrichment using the same 5,243 tested genes as the background.

Mitochondrial sequence-support and replicate-level analyses

Species-diagnostic mitochondrial SNPs in the RNA-seq data were used to summarize *S. cerevisiae* and *S. paradoxus* mitochondrial sequence support in each F1 replicate. Replicates 1, 2, and 3 showed mixed parental support and were designated Group A, whereas 4 and 5 showed predominantly *S. cerevisiae*-like support and were designated Group B. Because these estimates were obtained from bulk cultures, mixed sequence support cannot distinguish intracellular heteroplasmy, mixtures of cells carrying different mitochondrial genomes, or recombinant mitochondrial genomes.

We examined 158 unique *trans*-only genes contributing to enriched mitochondrial and related GO terms. Replicate-level *trans*-only-like calls were generated using the same allele-specific support, significance, and effect-size thresholds as the main analysis but were used only to evaluate consistency across samples. Total-expression comparisons used normalized expression summed across both parental alleles within each F1 replicate. Differences in expression-profile similarity between the three

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

Group A and two Group B replicates were evaluated by an exact permutation test over all possible three-vs-two sample partitions.

Genomic divergence and intron architecture

To provide genomic context for the ASE and intron-retention analyses, we quantified SNP density and intron architecture across the *S. cerevisiae*/*S. paradoxus* one-to-one ortholog set. SNPs were counted from the biallelic *S. cerevisiae*/*S. paradoxus* SNP VCF used for allele-specific analyses. Coding regions were defined from annotated coding sequence (CDS) features in the merged gene annotation. Promoter regions were defined as the 600 bp upstream of each annotated gene, relative to strand. SNP density was calculated in species-specific coordinates and summarized as SNPs per gene, SNPs per promoter, and SNPs per kb. Because this analysis used a SNP VCF, these values describe SNP density and do not include indels, alignment gaps, or total sequence divergence.

To summarize the intron architecture, species-specific exon annotations were used to infer intron number for each one-to-one ortholog. Orthologs were classified as intronless in both species, intron-containing in both species, *S. cerevisiae*-only intron-containing, or *S. paradoxus*-only intron-containing. Orthologs were also classified according to whether they had the same intron number in both species. For orthologs with a single intron annotated in both species, intron lengths were compared between species.

Intron retention quantification and allele-resolved comparisons

Intron retention (IR) was quantified with IRFinder v1.3 in directional mode (Middleton et al. 2017). IRFinder estimates an intron-level IR ratio by comparing reads that support intron retention with reads that support splicing.

We compared intron retention between each diploid parent and the corresponding allele in the F1 hybrid using the replicate structure described above. The *S. cerevisiae* parent was compared with the *S. cerevisiae* allele in the hybrid, and the *S. paradoxus* parent was compared with the *S. paradoxus* allele in the hybrid. To focus on reliable intron-retention estimates, we retained introns with sufficient read support and consistent quantification across biological replicates. After filtering, 178 introns were analyzed in the *S. cerevisiae* parent-to-hybrid allele comparison, and 149 introns were analyzed in the *S. paradoxus* parent-to-hybrid allele comparison.

Differential intron retention between each parent and the corresponding hybrid allele was tested using Fisher's exact tests comparing retained and spliced read support, followed by Benjamini–Hochberg correction within each allele-specific comparison. Effect size was defined as $\Delta IR = IR_{hybrid\ allele} - IR_{parent}$. Differential intron-retention events were defined as those with absolute $\Delta IR > 0.1$ and FDR < 0.05. One additional event that narrowly missed the effect-size cutoff but showed strong statistical support was reported separately as a near-threshold event.

LYS2 promoter comparison and nucleosome prediction

To compare the promoter architecture at the *LYS2* locus across *Saccharomyces* species, we identified the annotated *LYS2* ortholog in each species and extracted the 600 bp region upstream of the coding sequence. Promoter sequences were extracted from the corresponding genome FASTA files while accounting for gene orientation so that all sequences were analyzed in the same transcriptional direction. The *S.c* and *S.p* promoters were included as the focal comparison, and three additional sensu stricto species, *S. mikatae* (*S. m*), *S. kudriavzevii* (*S. k*), and *S. uvarum* (*S. u*), were included to place these differences in a broader evolutionary context.

Promoter sequences were globally aligned and compared using three summary measures: percent identity across aligned nongap positions, total insertion/deletion burden, and the number of separate insertion/deletion events. These measures were used to distinguish broad promoter sequence divergence from more localized architectural differences.

Because TATA boxes are sequence-variable, promoters were scanned for TATA-like elements using the degenerate consensus TATAWAWA, where W represents A or T. Overlapping matches were allowed, and motif positions were recorded relative to the coding-sequence start site so that negative values indicate upstream promoter positions. Motif scanning was performed with FIMO (Find Individual Motif Occurrences) (Grant et al. 2011).

Intrinsic nucleosome occupancy was predicted for each promoter using NuPoP with the yeast model (Xi et al. 2010). NuPoP provides a predicted occupancy score at each promoter position, with higher values indicating greater predicted nucleosome occupancy. To compare promoter-proximal chromatin architecture across species, we summarized mean predicted occupancy in three regions of interest: an upstream indel-rich region near −390 bp, a broader core-promoter region from −150 to −50 bp, and a TATA-proximal region from −120 to −80 bp. These summaries were used to evaluate whether sequence differences at *LYS2* were associated with differences in predicted promoter-proximal nucleosome organization.

S. cerevisiae NucleoATAC-Seq data are from Schep et al. (2015a; PMID 26314830; GEO accession GSE66386). We used high-confidence nucleosome dyad calls from the pooled lin_all.nucpos.bed file together with the continuous nucleosome occupancy track, lin_all.occ.bw, and the NucleoATAC positioning-signal track, lin_all.nucleoatac_signal.bw. Genomic coordinates were converted to positions relative to the annotated *LYS2* start codon, with negative values indicating upstream positions.

Results

Regulatory divergence is pervasive and largely compensatory

To reveal the regulatory architecture underlying gene expression in *S. cerevisiae* and *S. paradoxus*, we measured gene expression in the diploid parents and the diploid F1 hybrid. Gene expression differences between species can arise from *cis*- or *trans*-regulatory effects. *Cis* effects can be estimated from ASE in the hybrid, where both parental alleles are exposed to the same *trans*-regulatory environment (Fig. 1a). *Trans* effects can then be inferred by comparing ASE in the hybrid with total-expression differences between the parental strains (Fig. 1a). Among 5,386 one-to-one orthologous genes, 5,243 genes (97.3%) satisfied expression-filtering criteria and were included in the expression analysis. Of these, 4,164 genes had sufficient allele-specific support for strict *cis*/*trans* classification, including ASE support in at least three F1 hybrid biological replicates. The remaining expression-filtered genes lacked sufficient allele-specific support and were not assigned a strict regulatory class.

Among the 4,164 strictly classifiable genes, 2,030 showed no detectable *cis* or *trans* effect meeting both the significance and effect-size thresholds and were classified as conserved. The remaining 2,134 genes were assigned to resolved divergent regulatory classes (Fig. 1, b and c, Supplementary Fig. 1). These included 586 *cis*-only genes, 621 *trans*-only genes, 875 compensatory *cis* + *trans* genes,

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

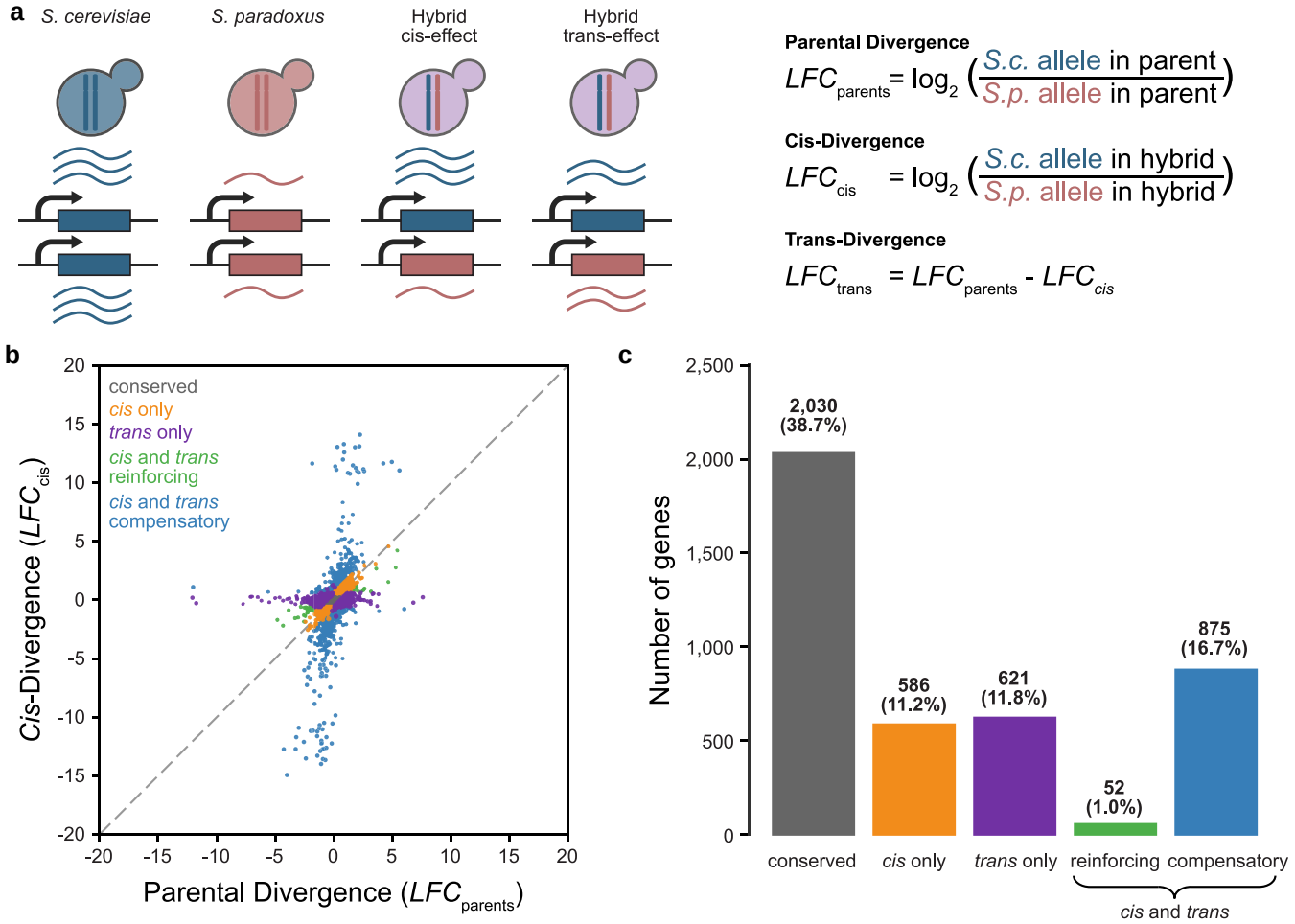


Fig. 1. Experimental framework and genome-wide classification of *cis*- and *trans*-regulatory divergence in *S. cerevisiae* × *S. paradoxus*. a) Conceptual schematic for partitioning regulatory divergence using parental expression and ASE in the F1 hybrid. In the *cis* example, the two parental alleles differ in the hybrid, indicating allele-linked regulatory divergence. In the *trans* example, parental total expression differs, whereas the two alleles are similarly expressed in the hybrid, indicating divergence in the shared *trans* environment. The equations used to calculate parental divergence, the *cis* component, and the inferred *trans* component are shown at right. *S. cerevisiae* is shown in blue, *S. paradoxus* in red, and the hybrid in purple. b) Scatter plot of regulatory divergence across genes. The x-axis shows parental-expression divergence and the y-axis shows the *cis* component measured by ASE in the hybrid. Points are colored by regulatory class: conserved, *cis*-only, *trans*-only, *cis* + *trans* reinforcing, and *cis* + *trans* compensatory. The dashed diagonal indicates equality between parental divergence and the *cis* component. Apparent outliers on the y-axis, all classified as compensatory, have strong support with median informative depths exceeding 4,000 reads and support in at least three replicates. Apparent outliers on the x-axis, mostly classified as *trans*-only, include some low-count transcripts. Raw transcript counts are included in Supplementary Dataset 1. c) Numbers and percentages of the 5,243 expression-tested genes assigned to each regulatory category. Strict regulatory classes required sufficient allele-specific support and were defined using an adjusted $P \leq 0.005$ and an absolute log2 fold change ≥0.58 for the corresponding *cis* and/or *trans* component. A total of 4,164 genes satisfied the allele-resolution criteria for strict classification, including 2,030 conserved, 586 *cis*-only, 621 *trans*-only, 52 *cis* + *trans* reinforcing, and 875 *cis* + *trans* compensatory genes. The remaining 1,079 genes (20.6%) passed the expression filter but lacked sufficient allele-specific support for strict classification and were therefore designated unclassified. Percentages shown above the bars are calculated relative to all 5,243 expression-tested orthologs.

and 52 reinforcing *cis* + *trans* genes. Expressed relative to all 5,243 expression-filtered genes, these categories correspond to 38.7% conserved, 11.2% *cis*-only, 11.8% *trans*-only, 16.7% compensatory, and 1.0% reinforcing genes.

Among the 2,134 genes with detectable regulatory divergence, 927 had both *cis* and *trans* contributions. Of these, 875 had effects in opposite directions, whereas only 52 had effects in the same direction, corresponding to an approximately 16.8-fold excess of compensatory over reinforcing classifications. Because inferred *trans* effects are calculated as parental-expression divergence minus hybrid ASE, shared-ASE measurement error can induce a negative relationship between *cis* and *trans* estimates and inflate the apparent frequency of compensatory effects (Fraser 2019; Zhang and Emerson 2019). To test whether the predominance of opposite-sign *cis*/*trans* effects depended on reuse of the same hybrid ASE estimate, we performed an exhaustive cross-replicate sensitivity analysis using all 10 possible three-vs-two divisions of the five F1 hybrid replicates in both reciprocal orientations. In each of the 20 comparisons, *cis* was estimated from one replicate subset, whereas *trans* was calculated using the ASE estimate from the complementary, nonoverlapping subset.

Across the 20 reciprocal comparisons, opposite-sign effects exceeded same-direction effects in every comparison. The proportion of genes with opposite-sign effects ranged from 84.1% to 90.8%, with a median of 88.6%, and the corresponding opposite-to-same-direction ratios ranged from 5.30:1 to 9.82:1, with a median of 7.76:1 (Supplementary Table 2). Regulatory-class assignments were also robust to stricter allele-resolution filters (Supplementary Table 3). With increasingly stringent coding-SNP thresholds, the fraction of genes with a detectable *cis* component increased from

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

36.3% in the full strict set to 42.1% at ≥100 coding SNPs, while compensatory classifications remained more frequent than reinforcing classifications across all SNP and ASE-depth thresholds tested.

Compensatory divergence conceals large and opposing regulatory effects

Given the strong bias toward compensatory divergence, we investigated whether compensatory and reinforcing changes differ in the magnitude of their underlying regulatory components. If stabilizing selection permits larger regulatory effects when they are offset by opposing changes, compensatory genes should exhibit larger absolute *cis* or *trans* effects than reinforcing genes (Metzger et al. 2017). To test this, we compared the absolute LFC magnitudes of net parental divergence, *cis* effects, and *trans* effects across regulatory classes (Fig. 2).

Consistent with this expectation, compensatory genes showed larger underlying regulatory component magnitudes than reinforcing genes (Fig. 2). For *cis* effects, the median absolute effect was 1.53 in compensatory genes and 0.83 in reinforcing genes, corresponding to approximately 2.9-fold and 1.8-fold differences, respectively. Thus, the typical *cis* effect was approximately 1.8 times larger in compensatory genes than in reinforcing genes on the log2 effect-size scale. *Trans* effects showed a similar but more modest pattern, with median absolute effects of 1.28 in compensatory genes and 1.03 in reinforcing genes, corresponding to approximately 2.4-fold and 2.0-fold differences, respectively. These differences were statistically significant, with Mann–Whitney *U* tests confirming larger *cis* and *trans* effects in compensatory genes than in reinforcing genes (*cis* effects: $P = 7.00 \times 10^{-11}$; *trans* effects: $P = 0.0078$).

Importantly, these larger *cis* and *trans* components did not produce larger net parental-expression divergence. Instead, compensatory genes had much lower net parental divergence than reinforcing genes, with median |LFCnet| = 0.43 for compensatory genes compared with 2.06 for reinforcing genes (Fig. 2). This pattern is consistent with large *cis* and *trans* effects acting in opposite directions and partially canceling each other, thereby buffering total mRNA output despite substantial underlying regulatory divergence. In contrast, reinforcing genes showed larger net parental divergence because *cis* and *trans* effects act in the same direction. For comparison, *cis*-only and *trans*-only genes showed closer agreement between their single regulatory component and net parental divergence, as expected when divergence is dominated by one regulatory mode (Fig. 2).

Thus, compensatory genes carried larger opposing *cis* and *trans* components than reinforcing genes while showing substantially smaller net parental-expression differences. Similar parental transcript abundance can therefore conceal extensive divergence in the underlying regulatory architecture. We next asked whether these different regulatory architectures were associated with distinct patterns of total-expression inheritance in the F1 hybrid.

Hybrid expression patterns differ across *cis–trans*-regulatory architectures

After estimating *cis*- and *trans*-regulatory components, we compared total expression in the F1 hybrid with expression in the

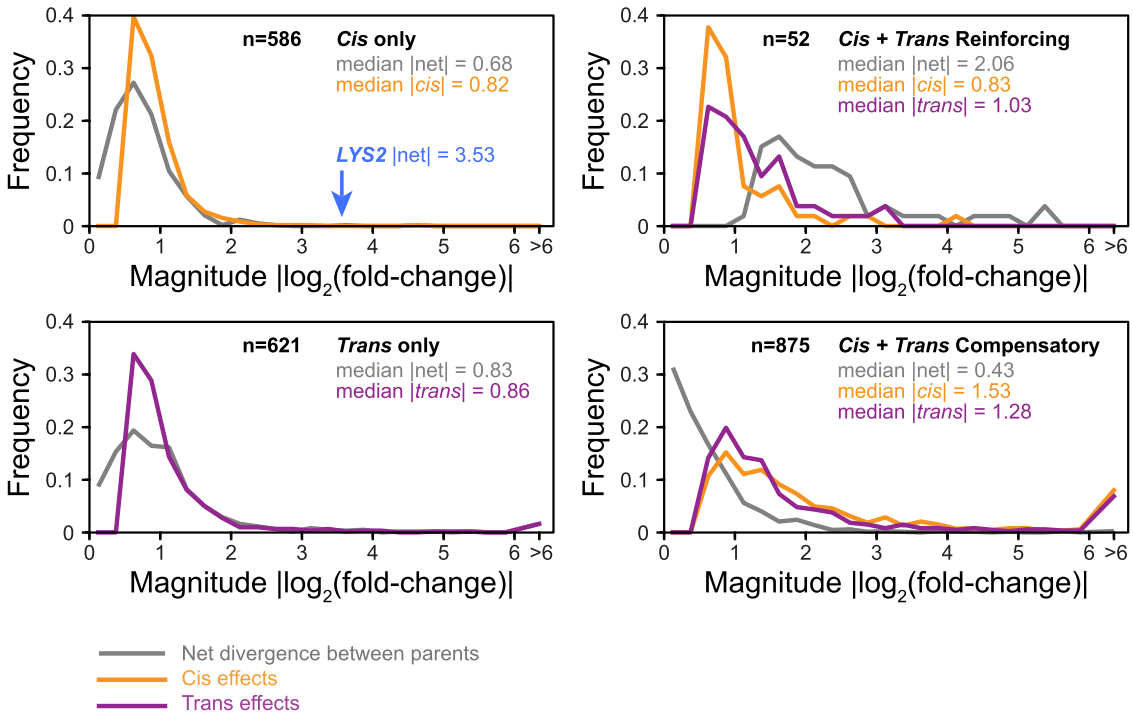
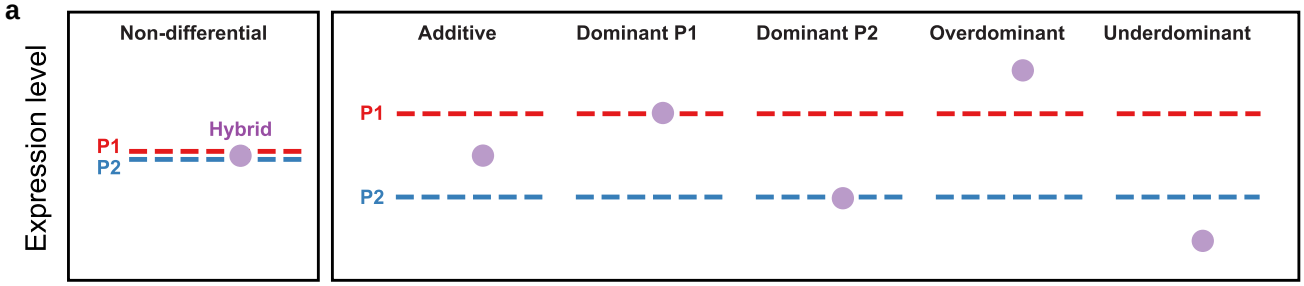


Fig. 2. Compensatory *cis–trans* divergence permits large underlying regulatory changes while stabilizing net expression differences. Each panel shows the distribution of effect-size magnitudes, plotted as frequency vs |log2 (fold-change)|, for genes assigned to one regulatory class. Gray curves show the magnitude of net divergence between parents (parent–parent expression fold change). Orange curves show the magnitude of *cis* effects (allelic imbalance in the F1 hybrid, reflecting allele-linked local regulatory differences in a shared *trans* environment). Purple curves show the magnitude of *trans* effects (differences attributable to the regulatory environment after accounting for *cis*). Panel headers report the number of genes (*n*) and the median magnitude for each component. Regulatory classes are defined as follows: *cis*-only genes show significant cis effects with no detectable *trans* component; *trans*-only genes show significant *trans* effects with no detectable *cis* component; *cis* + *trans* reinforcing genes show *cis* and *trans* effects in the same direction, increasing net divergence; and *cis* + *trans* compensatory genes show *cis* and *trans* effects in opposite directions, yielding small net divergence despite large component magnitudes. *LYS2* is highlighted (blue arrow) as an extreme example within the *cis*-only distribution. Values exceeding 6 are grouped in the rightmost ">6" bin.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

a



b

Regulatory Effect	Non-differential	Additive	Dom. (*S. cer*)	Dom. (*S. par*)	Overdominant	Underdominant	Total
Conserved	1971 (97%)	0 (<1%)	13 (<1%)	43 (2%)	3 (<1%)	0 (<1%)	2030
Cis only	428 (73%)	28 (5%)	41 (7%)	88 (15%)	1 (<1%)	0 (<1%)	586
Trans only	367 (59%)	70 (11%)	24 (4%)	160 (26%)	0 (<1%)	0 (<1%)	621
Cis & Trans, Compensatory	702 (80%)	73 (8%)	36 (4%)	57 (7%)	4 (<1%)	3(<1%)	875
Cis & Trans, Reinforcing	2 (4%)	32 (62%)	8 (15%)	10 (19%)	0 (<1%)	0 (<1%)	52

Fig. 3. Hybrid expression modes and their distribution across regulatory classes. a) Schematic of hybrid expression categories relative to the two parents. Hybrid expression was classified as nondifferential when it fell within the parental range and was not significantly different from either parent; additive when it fell between the parents and differed from both; dominant when it matched one parent more closely than the other; and transgressive when it fell outside the parental range, either above both parents as overdominant or below both parents as underdominant. Purple circles denote hybrid expression. b) Distribution of hybrid expression categories within each regulatory class. Values are shown as the number of genes, with percentages in parentheses, for nondifferential, additive, dominance toward *S. cerevisiae*, dominance toward *S. paradoxus*, overdominant, and underdominant categories.

two parental species (Fig. 3a). Across 5,386 genes in the hybrid expression analysis, hybrid expression was predominantly conservative. A total of 4,389 genes (81.5%) were classified as nondifferential, meaning that hybrid expression was not significantly different from either parent. In contrast, 300 genes (5.6%) showed additive expression, 679 genes (12.6%) showed expression biased toward one parent, and only 18 genes (0.5%) showed significant transgressive expression outside the parental range. Thus, even in the presence of widespread regulatory divergence, most hybrid transcripts showed conservative expression.

This conservative pattern was strongest for genes with no detectable *cis* or *trans* divergence. Among the 2,030 genes in the conserved regulatory class, 1,971 genes (97.1%) were nondifferential, as expected for genes without detectable regulatory divergence. Genes with only *cis* or only *trans* divergence were more likely to deviate from this pattern, but in different ways. In the *cis*-only class, 428 of 586 genes (73.0%) remained nondifferential, while the remaining genes were distributed mainly between *S. cerevisiae*-dominant and *S. paradoxus*-dominant expression. In the *trans*-only class, 367 of 621 genes (59.1%) were nondifferential, but genes showing parental bias were strongly skewed toward *S. paradoxus*. A total of 160 *trans*-only genes (25.8%) were classified as *S. paradoxus*-dominant, compared with only 24 genes (3.9%) classified as *S. cerevisiae*-dominant (Fig. 3b).

The clearest contrast emerged among genes carrying both *cis* and *trans* divergence. Most compensatory genes remained buffered, with 702 of 875 genes (80.2%) classified as nondifferential and another 73 genes (8.3%) classified as additive. Only a minority of compensatory genes showed dominant or transgressive expression patterns. By contrast, reinforcing genes were much less likely to remain nondifferential. Only 2 of 52 reinforcing genes (3.8%) were nondifferential, whereas 32 genes (61.5%) were additive, 8 genes (15.4%) were *S. cerevisiae*-dominant, and 10 genes (19.2%) were *S. paradoxus*-dominant (Fig. 3b).

Therefore, opposite-sign *cis*–*trans* changes are usually associated with conservative hybrid expression, whereas same-direction *cis*–*trans* changes are more likely to produce additive inheritance. This extends the previous result that compensatory genes can harbor large underlying regulatory differences while still producing stable total transcript abundance.

This comparison also helps explain the rare cases of transgressive expression, where hybrid expression falls significantly outside the parental range. Transgressive genes were strongly enriched in the compensatory class. Among the 11 genes with transgressive expression and a resolved regulatory architecture, 7 belonged to the compensatory class (Fig. 3b). Although most compensatory genes remained well buffered, these rare failures of buffering were therefore disproportionately drawn from this class. Specifically, compensatory genes were 6.62-fold more likely than non-compensatory resolved genes to show overdominant or underdominant expression in the hybrid (Fisher's exact test, $P = 0.0026$). Thus, although most compensatory genes did not show transgressive expression, compensatory genes were overrepresented among the small number of resolved transgressive cases. This association suggests that opposing regulatory effects are generally associated with buffering but can also accompany rare failures to maintain hybrid expression within the parental range.

Conserved and *trans*-only genes show distinct functional enrichment

The distinct effect magnitudes and hybrid inheritance patterns associated with the regulatory classes raised the question of whether these architectures were also concentrated in particular biological pathways. We therefore performed GO enrichment analysis for each regulatory class using the 5,243 expression-tested orthologs as the background set (The Gene Ontology Consortium 2019; Fig. 4a–c). Significant biological-process enrichment was detected only in the conserved and *trans*-only classes, whereas the *cis*-only, compensatory, and reinforcing classes showed no significant biological-process enrichment at the same threshold (Fig. 4a). This suggests that pathway-level organization is strongest among genes with conserved regulation and genes with *trans*-only divergence, while *cis*-containing classes are more functionally heterogeneous.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

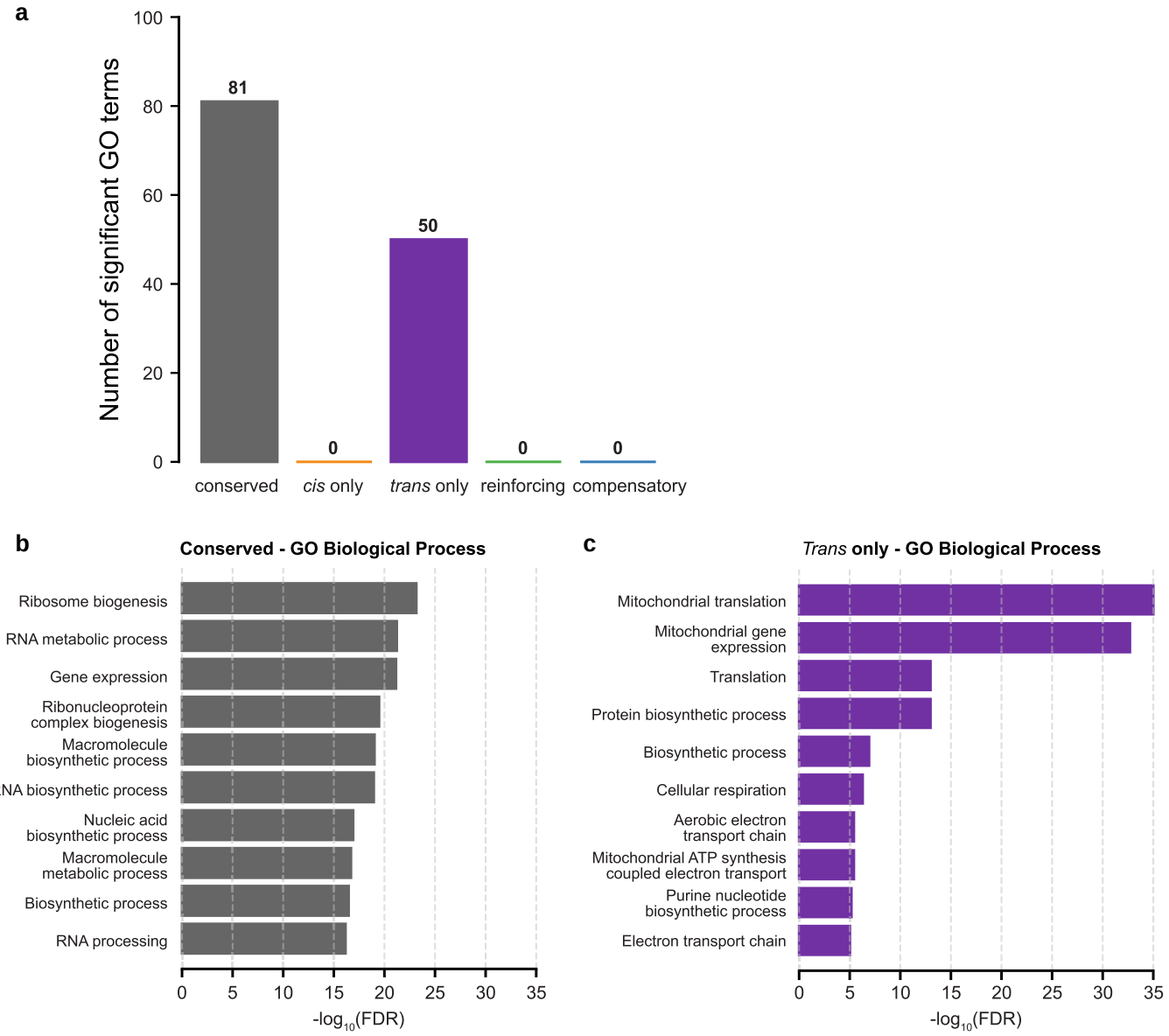


Fig. 4. GO biological-process enrichment reveals pathway-level organization concentrated in conserved and *trans*-only regulatory classes. a) Number of significantly enriched GO biological-process terms identified in each regulatory class using the same ortholog-restricted background of 5,243 tested genes. Enrichment was assessed for all classes under the same statistical framework, but only the conserved and *trans*-only classes yielded significant terms, with 81 and 50 enriched terms, respectively. The *cis*-only, *cis* + *trans* reinforcing, and *cis* + *trans* compensatory classes showed no significant biological-process enrichment at the same threshold. b) Top 10 enriched GO biological-process terms for conserved genes, ranked from most to least significant. Conserved genes were enriched primarily for ribosome biogenesis, RNA metabolism, gene expression, ribonucleoprotein complex biogenesis, and related biosynthetic and RNA-processing functions, consistent with strong conservation of core housekeeping pathways. The terms shown are ribosome biogenesis (GO: 0042254; FDR = 7.01×10^{-24}; $n = 229$), RNA metabolic process (GO: 0016070; FDR = 5.95×10^{-22}; $n = 597$), gene expression (GO: 0010467; FDR = 7.16×10^{-22}; $n = 801$), ribonucleoprotein complex biogenesis (GO: 0022613; FDR = 3.37×10^{-20}; $n = 253$), macromolecule biosynthetic process (GO: 0009059; FDR = 9.51×10^{-20}; $n = 863$), RNA biosynthetic process (GO: 0032774; FDR = 1.12×10^{-19}; $n = 530$), nucleic acid biosynthetic process (GO: 0141187; FDR = 1.23×10^{-17}; $n = 540$), macromolecule metabolic process (GO: 0043170; FDR = 2.01×10^{-17}; $n = 1070$), biosynthetic process (GO: 0009058; FDR = 3.43×10^{-17}; $n = 1026$), and RNA processing (GO: 0006396; FDR = 6.97×10^{-17}; $n = 286$). c) Top 10 enriched GO biological-process terms for *trans*-only genes, ranked from most to least significant. *Trans*-only genes were enriched for mitochondrial translation, mitochondrial gene expression, cellular respiration, electron transport, and related energy-metabolism functions, indicating that *trans*-regulatory divergence is concentrated in coordinated mitochondrial and respiratory modules. The terms shown are mitochondrial translation (GO: 0032543; FDR = 2.64×10^{-37}; $n = 77$), mitochondrial gene expression (GO: 0140053; FDR = $2.04 \times 10{-}33$; $n = 80$), translation (GO: 0006412; FDR = 1.04×10^{-13}; $n = 102$), protein biosynthetic process (GO: 0160307; FDR = 1.04×10^{-13}; $n = 102$), biosynthetic process (GO: 0009058; FDR = 1.22×10^{-7}; $n = 338$), cellular respiration (GO: 0045333; FDR = 5.23×10^{-7}; $n = 36$), aerobic electron transport chain (GO: 0019646; FDR = 3.87×10^{-6}; $n = 14$), mitochondrial ATP synthesis coupled electron transport (GO: 0042775; FDR = 3.87×10^{-6}; $n = 14$), purine nucleotide biosynthetic process (GO: 0006164; FDR = 6.80×10^{-6}; $n = 21$), and electron transport chain (GO: 0022900; FDR = 9.52×10^{-6}; $n = 14$). In b) and c), bar length indicates enrichment significance as –log10(FDR).

Conserved genes were enriched for core housekeeping processes, including RNA metabolism, ribosome biogenesis, gene expression, ribonucleoprotein complex biogenesis, biosynthesis, and RNA processing (Fig. 4b). These enrichments indicate that genes with conserved regulatory architecture are concentrated in essential cellular programs related to RNA production, ribosome assembly, and general biosynthetic capacity.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

In contrast, *trans*-only genes showed a distinct functional signature dominated by mitochondrial and respiratory processes (Fig. 4c). Enriched terms included mitochondrial translation, mitochondrial gene expression, cellular respiration, electron transport, oxidative phosphorylation-related processes, and nucleotide biosynthesis. These results suggest that *trans*-regulatory divergence is not randomly distributed across the genome but is concentrated in coordinated mitochondrial and energy-metabolism modules.

To determine the extent to which mitochondrial genotype impacts the *trans*-only module, we genotyped the mitochondria in the five hybrid replicates. Three replicates show mixed *S. cerevisiae* and *S. paradoxus* mitochondrial sequence support, whereas two replicates showed clear *S. cerevisiae* mitochondria. Because the measurements were obtained from bulk RNA-seq, we cannot distinguish intracellular heteroplasmy or mixtures of cells carrying different mitochondrial genomes. The *trans*-only classification of the mitochondrial module was broadly supported in both groups. Of the 158 mitochondrial *trans*-only genes, 140 (88.6%), 126 (79.7%), and 87 (55.1%) are supported by three, four, and five replicates, respectively.

Intron retention is highly conserved between *S. cerevisiae* and *S. paradoxus*

Introns are rare in budding yeast, with only a few hundred introns across the genome (Spingola et al. 1999). Of the 5,366 one-to-one orthologs, 5,133 (95.7%) lack annotated introns in both species, 196 (3.7%) contain introns in both species, 21 (0.4%) contain an intron only in *S. cerevisiae*, and 16 (0.3%) contain an intron only in *S. paradoxus*. Given the highly conserved intron architecture and extensive sequence divergence between *S. cerevisiae* and *S. paradoxus*, we sought to determine the extent to which splicing remained compatible within the shared hybrid *trans* environment.

Most introns showed low retention in both parents and clustered near the origin in the parent–parent comparison in Fig. 5b, indicating broadly similar IR levels under the conditions tested. A small number of introns showed baseline differences between the species, including higher retention of *SFT1* in the *S. cerevisiae* parent and higher retention of *GPI15* in the *S. paradoxus* parent. These differences were restricted to individual loci rather than reflecting a genome-wide shift in intron retention.

We next tested whether placement of each parental allele in the shared hybrid environment altered its intron retention relative to the corresponding parent. Most introns again remained close to the diagonal in the parent-vs-hybrid comparisons in Fig. 5c, indicating that IR was generally stable after hybridization. Using an IR ratio greater than 0.1 as a descriptive threshold, 22 of 178 introns, or 12.4%, exceeded this threshold in both the *S. cerevisiae* parent and the hybrid *S. cerevisiae* allele. In the *S. paradoxus* comparison, the number of introns exceeding this threshold increased modestly from 17 of 149, or 11.4%, in the parent to 25 of 149, or 16.8%, in the hybrid allele. Strong retention, defined as IR greater than or equal to 0.5, was uncommon. It was driven primarily by the canonical retained intron in *HAC1*, which showed near-complete retention in both the *S. cerevisiae* parent and the hybrid *S. cerevisiae* allele and showed no significant parent-to-hybrid change. No intron in the *S. paradoxus* parent reached an IR value greater than or equal to 0.5, whereas the *GPI15* intron was the only event in the hybrid *S. paradoxus* allele to exceed this threshold. Thus, strong intron retention remained rare, and the overall distributions provided little evidence for broad disruption of intron processing in the hybrid, although the *S. paradoxus* allele set showed a modest increase in IR values.

Against this broadly conserved background, only a small number of introns showed clear hybrid-associated changes. Using an effect-size cutoff of absolute ΔIR greater than 0.1 and FDR less than 0.05, we identified four primary differential introns and one near-threshold event with strong statistical support as shown in Fig. 5d. In the *S. paradoxus* allele, intron retention increased for *YDL012C* and decreased for *NBL1*, while *YML6* showed a near-threshold increase. In the *S. cerevisiae* allele, intron retention decreased for *SNI445* and *HNT2*. For each of these genes, the corresponding allele from the other species showed little or no comparable change. The strongest shifts were therefore predominantly allele-specific rather than part of a shared genome-wide response to the hybrid background. Additional statistically significant events with absolute ΔIR less than or equal to 0.1 were detected but were treated as secondary because their effect sizes were small.

The five loci with the strongest IR departures (*SNI445*, *HNT2*, *YDL012C*, *NBL1*, and *YML6*; Supplementary Table 4) do not share a common GO Biological Process; however, they are expressed at a lower level than other intron-containing genes (median DESeq2 baseMean of 586.4 and 3,379, respectively; $P = 0.0051$, Wilcoxon rank-sum test). These events represent rare locus-specific departures from an otherwise conserved pattern of intron processing rather than evidence for widespread splicing disruption.

Promoter architecture suggests candidate mechanisms for *cis*-regulatory divergence at *LYS2*

We examined the strongly *cis*-diverged gene *LYS2* as a case study for how promoter evolution may generate species-specific regulation (Fig. 6). The *S.c* promoter contains a 9-bp AT-rich insertion, TATGTATGA, located near −390 bp relative to the start codon (Fig. 6b). Because A/T-rich DNA is intrinsically less favorable for nucleosome formation, this insertion suggested that *cis* divergence at *LYS2* might reflect altered promoter architecture and local chromatin organization rather than simple gain or loss of a canonical activator-binding site.

Motif analysis supported this interpretation. Both species retained a conserved canonical Gcn4 site with the same match sequence, TGTGACTCAAT, at similar promoter positions, −196 bp in *S.c* and −191 bp in *S.p*. This argues against wholesale turnover of the major lysine-pathway regulatory input. In contrast, the best Fkh2-like candidate differed in both sequence and position. In *S.c*, the best match occurred at −455 bp with the sequence GTAAACA, whereas in *S.p*, the best match occurred at −391 bp with the sequence ATAAACA. TATA-like scanning further highlighted a species-specific architectural difference. Both promoters contained an upstream TATA-like sequence near the indel-rich region, TATATATA at −398 bp in *S.c* and TATATAAA at −394 bp in *S.p*, but only *S.c* carried an additional promoter-proximal TATA-like element at −88 bp, TATAAAAA (Fig. 6b). Thus, the key difference at *LYS2* is not the appearance of a new strong upstream activator motif, but a reorganization of promoter features in space.

These sequence changes produced marked differences in spacing among candidate regulatory elements. In *S.c*, the best Fkh2-like candidate was located approximately 57 bp upstream of the upstream TATA-like element. In *S.p*, the corresponding features were separated by only approximately 3 bp. Together with the additional downstream TATA-like element in *S.c*, these spacing shifts support a promoter-evolution model in which *cis* divergence arises through altered arrangement and phasing of promoter features rather than binary motif gain or loss.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

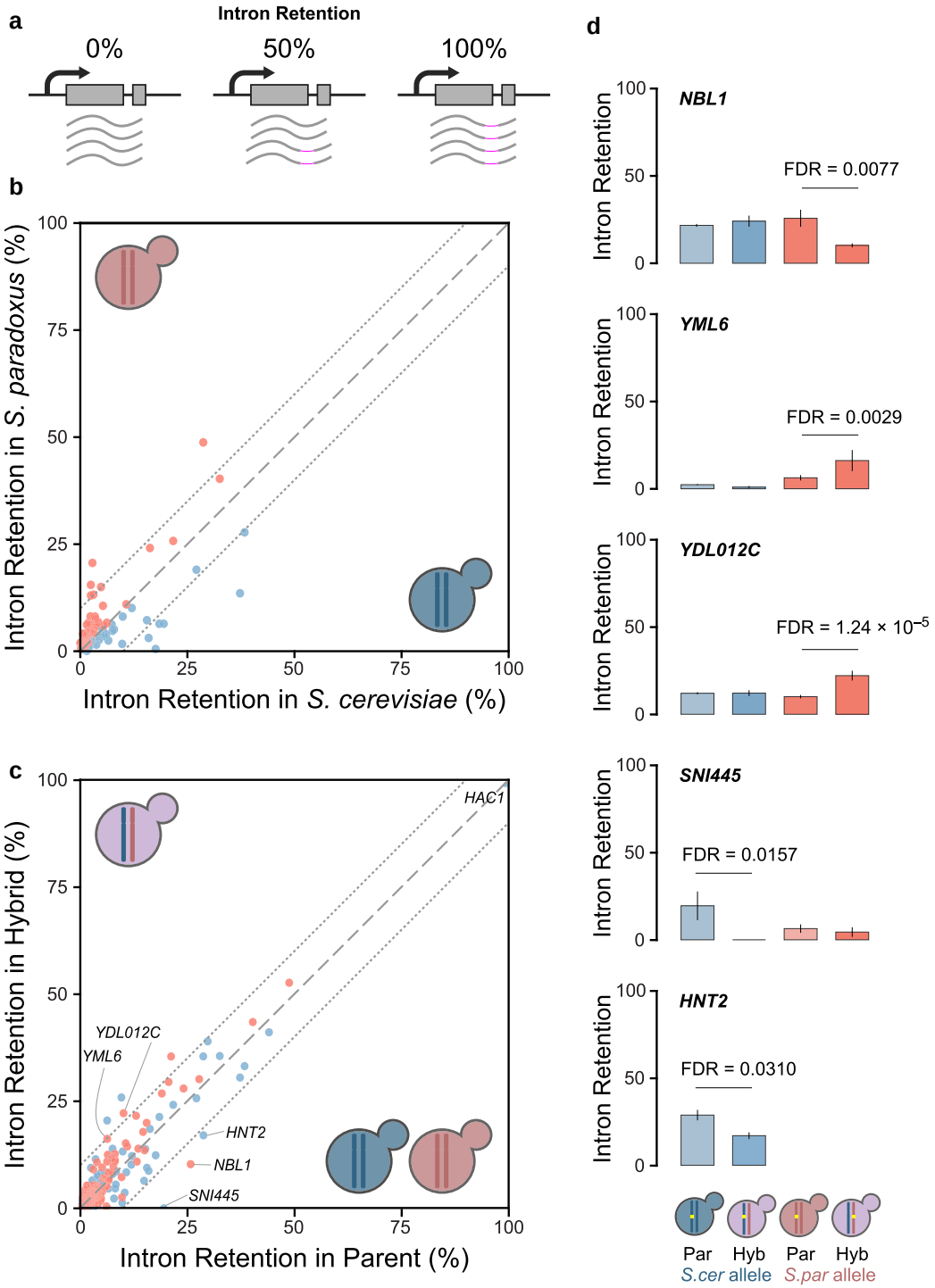


Fig. 5. Intron retention is broadly conserved between *S. cerevisiae*, *S. paradoxus*, and their F1 hybrid, with rare allele-specific shifts. a) Schematic of intron retention (IR) at 0%, 50%, and 100%. Gray arcs represent spliced reads, and magenta segments represent unspliced intron-containing reads. In panels b–d, blue denotes *S.c* or the *S.c* allele, and red denotes *S.p* or the *S.p* allele. Dashed lines in scatterplots indicate equal IR between axes, with flanking reference bounds where shown. b) Parent-vs-parent comparison of mean IR. Each point represents one gene, summarized by the maximum mean IR among analyzed introns. Labeled genes are representative examples highlighting loci with higher retention in one parent or notable retention patterns. For example, *SFT1* showed higher retention in *S.c* than *S.p* (IR = 0.373 vs 0.135; ΔIR = −0.238), whereas *GPI15* showed higher retention in *S.p* than *S.c* (IR = 0.488 vs 0.287; ΔIR = + 0.201). c) Parent-vs-hybrid comparison of allele-resolved IR. Each point represents one analyzed intron comparing the parent with the corresponding hybrid allele. Labeled genes are representative examples of stable introns, modest shifts, or candidate hybrid-associated changes. Strong retention (IR ≥ 0.5) was rare across all comparisons. *HAC1* served as a retained-intron positive control and showed near-complete retention in the *S.c* parent and hybrid *S.c* allele (0.99 ± 0.000 vs 0.98 ± 0.002), with no significant parent-to-hybrid shift (ΔIR = −0.0043; FDR = 0.175). d) Representative hybrid-associated allele-specific IR shifts. Bars show mean IR ± SEM for the parent (Par) and corresponding hybrid allele (Hyb). For the *S. paradoxus* allele, IR increased for *YDL012C* (0.101 ± 0.010 to 0.221 ± 0.028; ΔIR = + 0.120), decreased for *NBL1* (0.258 ± 0.049 to 0.102 ± 0.011; ΔIR = −0.155), and increased for the near-threshold event *YML6* (0.063 ± 0.015 to 0.162 ± 0.060; ΔIR = + 0.099). For the *S. cerevisiae* allele, IR decreased for *SNI445* (0.195 ± 0.082 to 0.000 ± 0.000; ΔIR = −0.195) and *HNT2* (0.288 ± 0.030 to 0.170 ± 0.019; ΔIR = −0.118). Horizontal brackets indicate the parent-to-corresponding-hybrid-allele comparison tested for each locus. Differential IR was assessed using Fisher's exact tests of retained and spliced read support followed by Benjamini–Hochberg correction.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

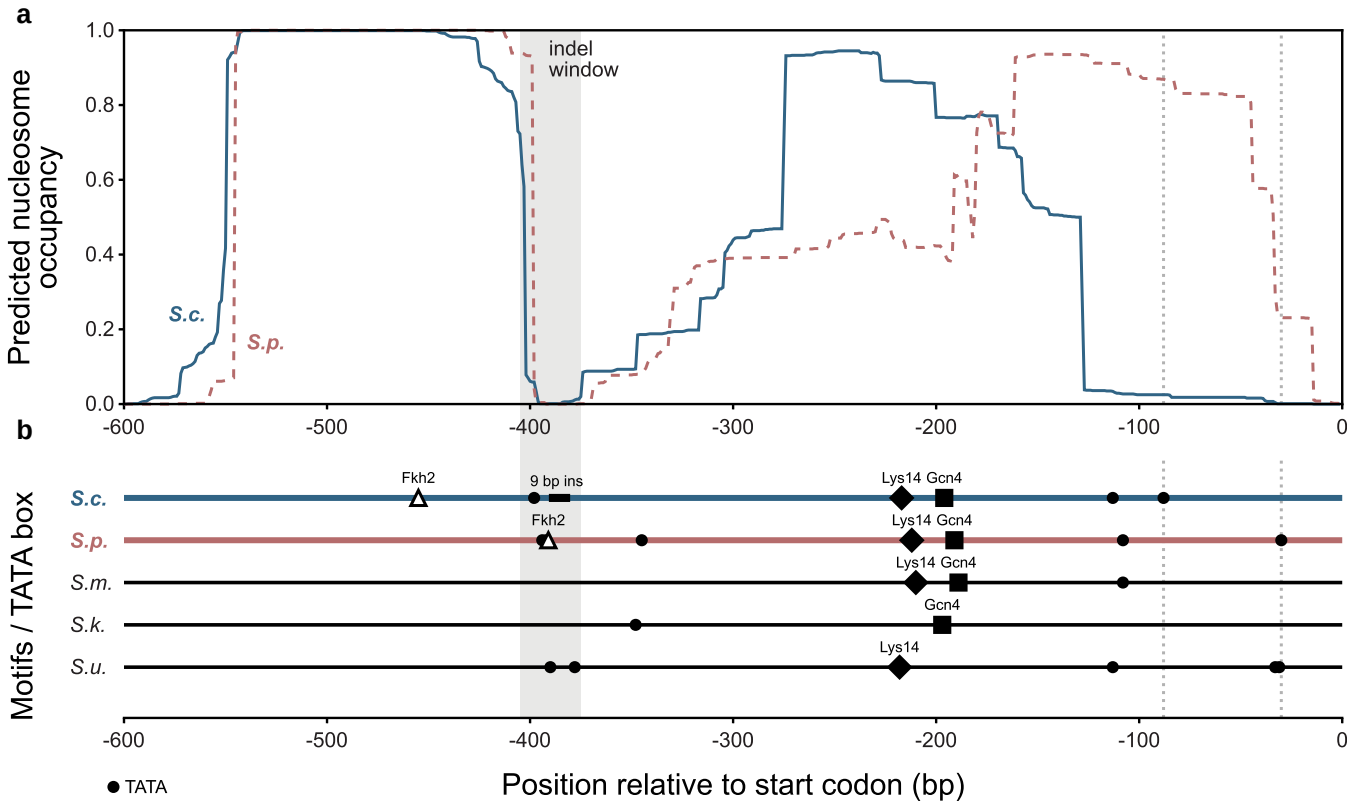


Fig. 6. *LYS2* promoter architecture and predicted intrinsic nucleosome occupancy in *S. cerevisiae* and *S. paradoxus*. a) NuPoP-predicted nucleosome occupancy across the 600-bp region upstream of the *LYS2* start codon in *S. cerevisiae* (blue solid line) and *S. paradoxus* (red dashed line). Occupancy is shown on a scale from 0 to 1, with lower values indicating lower predicted nucleosome occupancy. The gray shaded region marks the indel-centered window from −405 to −375 bp relative to the start codon. Vertical dotted lines mark promoter-proximal reference positions near −88 and −33 bp highlighted in the comparative motif analysis. b) Positions of candidate *cis*-regulatory sequence features mapped onto the same coordinate system for *LYS2* promoters from *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. uvarum*. Circles indicate TATA-like sequence matches, triangles indicate Fkh2-like matches, diamonds indicate Lys14 candidate sites, and squares indicate Gcn4 candidate sites. The gray shaded region marks the *S. cerevisiae*-specific 9-bp AT-rich insertion centered near −390 bp.

Sequence-based nucleosome prediction reinforced this view. Predicted nucleosome occupancy is reported on a 0-to-1 scale, where values near 0 indicate low predicted occupancy and values near 1 indicate high predicted occupancy (Fig. 6a). The predicted nucleosome occupancy is qualitatively consistent with NucleoATAC-Seq data (Schep et al. 2015a; Supplementary Fig. 3). Around the upstream insertion itself, mean predicted occupancy in the −405 to −375 bp window was low in both species, but lower in *S.c* than in *S.p* (0.078 vs 0.213). By contrast, predicted divergence in the core promoter was much stronger. In the −150 to −50 bp window, predicted occupancy was 0.132 in *S.c* compared with 0.882 in *S.p*. In the −120 to −80 bp window, predicted occupancy was 0.027 in *S.c* compared with 0.884 in *S.p*. These values predict a much broader promoter-proximal nucleosome-depleted region in *S.c*, coincident with the species-specific downstream TATA-like element.

An important control was that *LYS2* was not unusually diverged overall at the sequence level. Relative to 20 conserved-expression control promoters, *LYS2* showed high gap-free identity, 0.842, corresponding to approximately the 90th percentile among controls, but only typical indel burden, 34 indel bp, approximately the 60th percentile, and typical indel event count, 14 events, approximately the 65th percentile. Global A/T content was also similar between species, 0.617 in *S.c* and 0.615 in *S.p*, indicating that the relevant difference is not broad promoter degeneration or an overall base-composition shift. Instead, *LYS2* stood out specifically for local promoter architecture. Across the 20 conserved-expression controls, TATAWAWA matches were found only in far-upstream regions, −599 to −337 bp, and were absent from the promoter-proximal windows, −150 to −50 bp and −120 to −80 bp, in both species. In contrast, *LYS2* uniquely carried a proximal TATA-like element at −88 bp in *S.c*. Likewise, predicted promoter-proximal nucleosome divergence at *LYS2* was more extreme than in any control promoter. For the difference in predicted occupancy between *S.p* and *S.c*, the control maxima were 0.604 in the −150 to −50 bp window and 0.769 in the −120 to −80 bp window, whereas *LYS2* reached 0.750 and 0.857, respectively. By contrast, the difference around the upstream indel region was only 0.135 and remained within the control range. Thus, *LYS2* is not an outlier for overall promoter divergence, but it is an outlier for predicted promoter-proximal chromatin divergence.

Extending the comparison across additional sensu stricto species strengthened the same conclusion. TATA-like sequences near the upstream AT-rich region occurred in several species, but only *S.c* carried the promoter-proximal TATA-like element at −88 bp. *S. uvarum* had a downstream TATA-like feature near −33 bp, and *S.p* contained a short TATAAA sequence near −30 bp, but neither reproduced the *S.c* promoter configuration (Fig. 6b). Predicted promoter-proximal nucleosome occupancy followed these architectural differences. In the −150 to −50 bp and −120 to −80 bp windows, *S.c* remained low, 0.132 and 0.027, whereas *S.p* was high, 0.882 and 0.884. *S. kudriavzevii* and *S. uvarum*

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

showed intermediate values, 0.462 and 0.401 for *S. kudriavzevii* and 0.551 and 0.480 for *S. uvarum*. *S. mikatae* showed near-maximal predicted occupancy, 0.997 and 0.999. Although the downstream TATA-like feature in *S. uvarum* was associated with a narrow local occupancy dip near −33 bp, 0.006 compared with 0.291 in *S.p*, it did not reproduce the broad promoter-proximal depletion predicted in *S.c* (Supplementary Fig. 2).

Discussion

Here, we show that *cis*- and *trans*-regulatory divergence between *S. cerevisiae* and *S. paradoxus* is common, but these individual effects tend to cancel out and do not manifest as differences in steady-state transcript abundance between species. Within the F1 hybrid, 99.5% of genes remained within the parental-expression range, whereas only 0.5% showed transgressive expression. For genes with both *cis* and *trans* contributions, compensatory cases exceeded reinforcing cases by approximately 16.8-fold. We also find that compensatory genes tend to carry larger *cis*- and *trans*-effect sizes than reinforcing genes, consistent with stabilizing selection acting on total transcript output. This interpretation agrees with hybrid-based and comparative studies that have repeatedly identified compensatory or antagonistic *cis–trans* divergence as a common feature of regulatory evolution (Landry et al. 2005; Tirosh et al. 2009; Metzger et al. 2017; Signor and Nuzhdin 2018; Hill et al. 2021).

A key issue in interpreting compensatory *cis/trans* divergence is that inferred *trans* effects are calculated by subtracting hybrid ASE from parental-expression divergence. Shared-ASE measurement error can therefore create a spurious negative relationship between *cis* and *trans* effects and inflate apparent compensation (Fraser 2019; Zhang and Emerson 2019). Whereas Fraser (2019) reported a strong attenuation in a highly replicated mouse cross (mean $r = -0.71$ to -0.079), the median correlation in our matched replicate partitions changed only from −0.880 with shared subset-specific ASE to −0.674 with nonoverlapping subsets. Opposite-sign effects also remained predominant in all 20 comparisons, indicating that shared-ASE error inflates the conventional compensatory excess but does not fully explain the tendency for opposing effects in this interspecific cross. These conservative cross-replicate analyses still show an approximately seven-fold excess of opposite-sign over same-direction *cis/trans* effects in this interspecific yeast cross. Because RNA-seq measures steady-state transcript abundance, genes with compensatory *cis/trans* effects may have differences in transcript mRNA turnover or gene expression noise that do not affect mean steady-state levels.

Having established that compensatory divergence can buffer parental-expression differences, the hybrid expression analysis allowed us to ask whether this buffering persists in the hybrid. *Cis*-only genes remained mostly within the parental range, but when they deviated, they more often did so through additive or parent-biased outcomes. By contrast, *trans*-only genes were more strongly skewed toward dominance, particularly toward the *S.p* state. Similar relationships between *cis* divergence and additive inheritance, and between *trans* divergence and dominant or recessive inheritance, have been described in yeast and *Drosophila* hybrids (Wittkopp et al. 2004; McManus et al. 2010; Schaefke et al. 2013). Of the 11 transgressive genes with resolved regulatory architectures, 7 belonged to the compensatory class (Fig. 3b). This suggests that opposing regulatory components can occasionally fail to maintain expression within the parental range when combined in the F1. Antagonistic or novel *cis/trans* interactions have previously been associated with hybrid misexpression (Landry et al. 2005; Tirosh et al. 2009; McManus et al. 2010; Herbst et al. 2017).

Regulatory classes also differed in their functional composition. Conserved genes were enriched for central functions related to RNA metabolism, ribosome biogenesis, and RNA processing, whereas *trans*-only genes were enriched for mitochondrial translation, mitochondrial gene expression, respiration, and electron transport (Fig. 4, b and c). These class-specific patterns are consistent with previous evidence that regulatory evolution across *Saccharomyces* species can be organized at the level of biological pathways (Schraiber et al. 2013). This enrichment is not due to mitochondrial genotype. Of the five replicates of the F1 hybrid, three appear to have both *S. cerevisiae* and *S. paradoxus* mitochondria and two show clear *S. cerevisiae* mitochondria, yet the *trans*-only-like pattern of mitochondrial and respiratory GO enrichments was broadly supported across both groups.

Within the same *trans*-only class, genes showing parent-biased hybrid expression were substantially more likely to be *S. paradoxus*-dominant than *S. cerevisiae*-dominant (Fig. 3b). The *S.p*-dominant *trans*-only genes carried a strong mitochondrial and respiratory signature, whereas the smaller *S.c*-dominant subset showed weaker functional enrichment. Our results suggest that there are one or more dominant *trans*-acting factors in *S. paradoxus* that regulate the expression of mitochondrial genes. This result agrees with yeast expression quantitative trait loci (eQTL) studies showing that *trans*-acting loci and hotspots can coordinately influence large gene modules (Brem et al. 2002; Yvert et al. 2003; Curtis et al. 2013; Albert et al. 2018).

Allele-resolved intron retention provided a complementary test of regulatory stability in the hybrid transcriptome (Fig. 5b–d). In contrast to the widespread *cis/trans* divergence detected for steady-state transcript abundance, intron retention remained largely stable when the *S. cerevisiae* and *S. paradoxus* alleles were processed in the same hybrid *trans* environment. Although only a small fraction of budding-yeast genes contains introns, these genes are not distributed randomly across biological functions. In our analysis, the conserved and species-specific IR-tested gene sets were strongly enriched for translation, ribosome biogenesis, and rRNA-processing functions, consistent with the known concentration of budding-yeast introns in highly expressed and ribosome-associated genes (Spingola et al. 1999; Juneau et al. 2006). Beyond this broad functional bias, individual yeast introns can also perform specific regulatory roles. The intron in *HAC1* controls translation during the unfolded protein response, whereas the *YRA1* intron contributes to the autoregulation of an essential mRNA-export factor (Cox and Walter 1996; Rüegsegger et al. 2001; Rodriguez-Navarro et al. 2002; Preker and Guthrie 2006). Yeast introns have also been implicated in starvation responses and other dynamic physiological programs, and intron retention is recognized more broadly as a regulated mode of gene control (Gonzalez-Hilarion et al. 2016; Jacob and Smith 2017; Parenteau et al. 2019; Gomez-Montalvo et al. 2024). In this cross, extensive nucleotide divergence provided sufficient sequence variation for allele resolution, while the strong conservation of intron architecture allowed direct comparison of orthologous introns between each parent and its corresponding hybrid allele. Most introns showed similarly low retention in the parents and hybrid alleles, and only four primary events and one near-threshold event showed substantial hybrid-associated shifts. The shifted genes did not share a single functional or architectural feature and remained measurably expressed. Thus, despite the divergence observed in the regulatory mechanisms underlying transcript abundance, placing both parental alleles in the shared hybrid *trans* environment did not

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

broadly alter intron retention. Instead, the hybrid showed a largely conserved splicing pattern, as measured by intron retention, with only rare allele-specific departures at individual loci.

Finally, the *LYS2* analysis extends the genome-wide classification to a specific hypothesis for one strong *cis* effect (Fig. 6b and Supplementary Fig. 3). Yeast promoter activity depends not only on individual transcription-factor-binding sequences but also on TATA elements, sequence composition, spacing, and nucleosome organization. TATA-containing promoters constitute a distinct regulatory class, promoter nucleosome architecture is associated with transcriptional plasticity, and promoter sequence can encode differences in intrinsic nucleosome preference (Basehoar et al. 2004; Tirosh and Barkai 2008; Hornung et al. 2012; Rosin et al. 2012). AT-rich and poly(dA:dT)-like sequences can disfavor nucleosome formation and contribute to local chromatin organization, while sequence-based predictions can provide mechanistic context when matched chromatin measurements are unavailable (Segal et al. 2006; Kaplan et al. 2009; Segal and Widom 2009; Xi et al. 2010).

At *LYS2*, the two promoters differed in the sequence and spacing of several candidate features, including an *S. cerevisiae*-specific AT-rich insertion, an Fkh2-like candidate, upstream TATA-like sequences, and a promoter-proximal TATA-like element observed only in *S. cerevisiae* among the species examined. In contrast, the candidate Gcn4 site was conserved. This pattern is consistent with broader models in which *cis*-regulatory divergence can arise from multiple localized changes in promoter sequence and organization rather than from the gain or loss of a single regulatory site (Wray 2007; Wittkopp and Kalay 2012; Shih and Fay 2021). The *S. cerevisiae* promoter was predicted to have substantially lower promoter-proximal nucleosome occupancy. Relative to the conserved-expression controls, *LYS2* was an outlier for this localized predicted occupancy difference but not for overall promoter sequence divergence or indel burden. Public *S. cerevisiae* NucleoATAC data were qualitatively consistent with several major predicted occupancy transitions but did not reproduce every feature of the sequence-based profile (Supplementary Fig. 3). Because those data were generated in a different strain and lacked a matched *S. paradoxus* comparison, the analysis identifies localized promoter architecture as a plausible contributor to the *cis* effect but does not establish an interspecies chromatin difference or a causal sequence variant. Direct promoter editing and matched chromatin profiling in the parental strain backgrounds would be needed to test the proposed mechanism.

These interpretations should be considered within the scope of the study. We examined one growth condition and one F1 hybrid combination, and buffering at the RNA level does not necessarily imply equivalent protein output. The mitochondrial association remains suggestive rather than causal, while the *LYS2* analysis provides a mechanistic hypothesis that will require direct experimental testing. Even within this scope, the broader pattern is clear. Regulatory divergence between *S. cerevisiae* and *S. paradoxus* is widespread but often hidden because opposing *cis* and *trans* effects stabilize total transcript abundance, and this conclusion remained supported after exhaustive cross-replicate analyses that avoided reuse of the same hybrid ASE estimate and after sensitivity analyses accounting for variation in allele-specific resolution. In contrast, intron retention remained largely conserved when both parental alleles were processed in the shared hybrid *trans* environment, while *LYS2* illustrates how localized promoter architecture can generate a testable hypothesis for a strong *cis* effect. The interspecific hybrid therefore appears less like a transcriptomic failure state than a buffered intermediate, stable in overall output but carrying cryptic regulatory divergence.

Data availability

All strains used or generated in this study are available from the corresponding author upon request. The short-read sequencing data are available in the NCBI BioProject database and can be accessed with accession number PRJNA1469180. Supplementary Dataset 1 contains species-resolved raw featureCounts for *S. cerevisiae*, *S. paradoxus*, and F1 hybrid samples, together with gene coordinates and feature lengths. Supplementary Dataset 2 contains raw counts for the 5,386 one-to-one orthologs used for total gene-expression analyses, using the species-matched parental counts and summed *S. cerevisiae* and *S. paradoxus* counts in the F1 hybrid. Supplementary Dataset 3 contains DESeq2-normalized gene counts for these 5,386 orthologs across all parental and F1 biological replicates. Supplementary Dataset 4 contains per-replicate F1 ASE counts for the *S. cerevisiae* and *S. paradoxus* alleles, total allele-specific read counts, and the number of informative SNPs used for each gene. Supplementary Dataset 5 contains the integrated gene-level *cis-/trans*-regulatory classifications and hybrid-expression inheritance classifications, together with parental and hybrid expression summaries, log2 fold changes, and associated statistical results. Supplementary Dataset 6 contains intron-retention measurements comparing the *S. cerevisiae* parent with the *S. cerevisiae* allele in the F1 hybrid, including mean intron retention, standard errors, ΔIR, supporting spliced and unspliced reads, P values, and FDR values. Supplementary Dataset 7 contains the corresponding intron-retention measurements comparing the *S. paradoxus* parent with the *S. paradoxus* allele in the F1 hybrid.

Supplemental material available at GENETICS online.

Acknowledgements

We thank Patricia Wittkopp, David Zappulla, and members of the Lang Lab for comments on the manuscript (Rojas and Lang 2026; Schep et al. 2015b).

Funding

This work was supported by a grant from the National Institutes of Health (R35GM149540). Portions of this research were conducted on Lehigh University's Research Computing infrastructure partially supported by the National Science Foundation (Award 2019035).

Conflicts of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: GIL. Experiments: DR. Analysis: DR. Writing: DR, GIL. Editing: DR, GIL. Funding Acquisition: GIL.

Literature cited

Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. 2018. Genetics of *trans*-regulatory variation in gene expression. eLife. 7:e35471. https://doi.org/10.7554/eLife.35471.

Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA Box-containing genes. Cell. 116: 699–709. https://doi.org/10.1016/S0092-8674(04)00205-3.

Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. Science. 296: 752–755. https://doi.org/10.1126/science.1069516.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

Carroll SB. 2005. Evolution at two levels: on genes and form. PLoS Biol. 3:e245. https://doi.org/10.1371/journal.pbio.0030245.
Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. 2015. Tools and best practices for data processing in allelic expression analysis. Genome Biol. 16:195. https://doi.org/10.1186/s13059-015-0762-6.
Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. 2002. Detection of regulatory variation in mouse genes. Nat Genet. 32:432–437. https://doi.org/10.1038/ng992.
Cox JS, Walter P. 1996. A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell. 87:391–404. https://doi.org/10.1016/S0092-8674(00)81360-4.
Curtis RE, Kim S, Woolford JL, Xu W, Xing EP. 2013. Structured association analysis leads to insight into *Saccharomyces cerevisiae* gene regulation by finding multiple contributing eQTL hotspots associated with functional gene modules. BMC Genomics. 14:196. https://doi.org/10.1186/1471-2164-14-196.
Dobin A *et al.* 2013. STAR: ultrafast universal RNA-Seq aligner. Bioinformatics. 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354.
Fraser HB. 2019. Improving estimates of compensatory *cis–trans* regulatory divergence. Trends Genet. 35:3–5. https://doi.org/10.1016/j.tig.2018.09.003.
Gomez-Montalvo J *et al.* 2024. Replicative aging in yeast involves dynamic intron retention patterns associated with mRNA processing/export and protein ubiquitination. Microbial Cell. 11:69–78. https://doi.org/10.15698/mic2024.02.816.
Goncalves A *et al.* 2012. Extensive compensatory *cis–trans* regulation in the evolution of mouse gene expression. Genome Res. 22:2376–2384. https://doi.org/10.1101/gr.142281.112.
Gonzalez-Hilarion S *et al.* 2016. Intron retention-dependent gene regulation in *Cryptococcus neoformans*. Sci Rep. 6:32252. https://doi.org/10.1038/srep32252.
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics. 27:1017–1018. https://doi.org/10.1093/bioinformatics/btr064.
Herbst RH *et al.* 2017. Heterosis as a consequence of regulatory incompatibility. BMC Biol. 15:38. https://doi.org/10.1186/s12915-017-0373-7.
Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. Nat Rev Genet. 22:203–215. https://doi.org/10.1038/s41576-020-00304-w.
Hornung G, Oren M, Barkai N. 2012. Nucleosome organization affects the sensitivity of gene expression to promoter mutations. Mol Cell. 46:362–368. https://doi.org/10.1016/j.molcel.2012.02.019.
Jacob AG, Smith CWJ. 2017. Intron retention as a component of regulated gene expression programs. Hum Genet. 136:1043–1057. https://doi.org/10.1007/s00439-017-1791-x.
Juneau K, Miranda M, Hillenmeyer ME, Nislow C, Davis RW. 2006. Introns regulate RNA and protein abundance in yeast. Genetics. 174:511–518. https://doi.org/10.1534/genetics.106.058560.
Kaplan N *et al.* 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature. 458:362–366. https://doi.org/10.1038/nature07667.
King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science. 188:107–116. https://doi.org/10.1126/science.1090005.
Krieger G, Lupo O, Levy AA, Barkai N. 2020. Independent evolution of transcript abundance and gene regulatory dynamics. Genome Res. 30:1000–1011. https://doi.org/10.1101/gr.261537.120.
Landry CR *et al.* 2005. Compensatory *cis–trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. Genetics. 171:1813–1822. https://doi.org/10.1534/genetics.105.047449.
Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 30:923–930. https://doi.org/10.1093/bioinformatics/btt656.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol. 15:550. https://doi.org/10.1186/s13059-014-0550-8.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10–12. https://doi.org/10.14806/ej.17.1.200.
Martínez AA, Lang GI. 2026. Weak pervasive incompatibilities and compensatory adaptation drive hybrid genome evolution in yeast. Nat Commun. https://doi.org/10.1038/s41467-026-77469-1.
McKenna A *et al.* 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. https://doi.org/10.1101/gr.107524.110.
McManus CJ *et al.* 2010. Regulatory divergence in *Drosophila* revealed by mRNA-Seq. Genome Res. 20:816–825. https://doi.org/10.1101/gr.102491.109.
Metzger BPH, Wittkopp PJ, Coolon JD. 2017. Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. Genome Biol Evol. 9:843–854. https://doi.org/10.1093/gbe/evx035.
Middleton R *et al.* 2017. IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol. 18:51. https://doi.org/10.1186/s13059-017-1184-4.
Parenteau J *et al.* 2019. Introns are mediators of cell response to starvation. Nature. 565:612–617. https://doi.org/10.1038/s41586-018-0859-7.
Preker PJ, Guthrie C. 2006. Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. RNA. 12:994–1006. https://doi.org/10.1261/rna.6706.
Rodriguez-Navarro S, Strässer K, Hurt E. 2002. An intron in the YRA1 gene is required to control Yra1 protein expression and mRNA export in yeast. EMBO Rep. 3:438–442. https://doi.org/10.1093/embo-reports/kvf091.
Rojas DS, Lang GI. 2026. Gene expression is stable despite widespread cis and trans regulatory divergence in Saccharomyces yeasts. National Center for Biotechnology Information BioProject. PRJNA1469180.
Rosin D, Hornung G, Tirosh I, Gispan A, Barkai N. 2012. Promoter nucleosome organization shapes the evolution of gene expression. PLoS Genet. 8:e1002579. https://doi.org/10.1371/journal.pgen.1002579.
Rüegsegger U, Leber JH, Walter P. 2001. Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell. 107:103–114. https://doi.org/10.1016/S0092-8674(01)00505-0.
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. https://doi.org/10.1093/molbev/mst114.
Schep AN *et al.* 2015a. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. 25:1757–1770. https://doi.org/10.1101/gr.192294.115.

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026

Schep AN *et al.* 2015b. High-resolution nucleosome positioning from ATAC-seq chromatin accessibility data. NCBI Gene Expression Omnibus. GSE66386.

Schraiber JG, Mostovoy Y, Hsu TY, Brem RB. 2013. Inferring evolutionary histories of pathway regulation from transcriptional profiling data. PLoS Comput Biol. 9:e1003255. https://doi.org/10.1371/journal.pcbi.1003255.

Segal E *et al.* 2006. A genomic code for nucleosome positioning. Nature. 442:772–778. https://doi.org/10.1038/nature04979.

Segal E, Widom J. 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol. 19:65–71. https://doi.org/10.1016/j.sbi.2009.01.004.

Shih CH, Fay JC. 2021. *Cis*-regulatory variants affect gene expression dynamics in yeast. eLife. 10:e68469. https://doi.org/10.7554/eLife.68469.

Signor SA, Nuzhdin SV. 2018. The evolution of gene expression in *cis* and *trans*. Trends Genet. 34:532–544. https://doi.org/10.1016/j.tig.2018.03.007.

Spingola M, Grate L, Haussler D, Ares M. 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA. 5:221–234. https://doi.org/10.1017/S1355838299981682.

The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47:D330–D338. https://doi.org/10.1093/nar/gky1055.

Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. Genome Res. 18:1084–1091. https://doi.org/10.1101/gr.076059.108.

Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. Science. 324:659–662. https://doi.org/10.1126/science.1169766.

van de Geijn B, McVicker G, Gilad Y, Pritchard JK. 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Methods. 12:1061–1063. https://doi.org/10.1038/nmeth.3582.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. Nature. 430:85–88. https://doi.org/10.1038/nature02698.

Wittkopp PJ, Kalay G. 2012. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 13:59–69. https://doi.org/10.1038/nrg3095.

Wray GA. 2007. The evolutionary significance of *cis*-regulatory mutations. Nat Rev Genet. 8:206–216. https://doi.org/10.1038/nrg2063.

Xi L *et al.* 2010. Predicting nucleosome positioning using a duration hidden Markov model. BMC Bioinformatics. 11:346. https://doi.org/10.1186/1471-2105-11-346.

Yue JX *et al.* 2017. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. Nat Genet. 49:913–924. https://doi.org/10.1038/ng.3847.

Yvert G *et al.* 2003. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat Genet. 35:57–64. https://doi.org/10.1038/ng1222.

Zhang X, Emerson JJ. 2019. Inferring compensatory evolution of *cis*- and *trans*-regulatory variation. Trends Genet. 35:1–3. https://doi.org/10.1016/j.tig.2018.11.003.

Zhu A, Ibrahim JG, Love MI. 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 35:2084–2092. https://doi.org/10.1093/bioinformatics/bty895.

Editor: K. Arndt

Downloaded from academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyag233/8776363 by Lehigh University, Fairchild Martindale Library-Serials user on 03 October 2026