OXFORD GENETICS

GENETICS, 2021, 217(3), iyaa051
DOI: 10.1093/genetics/iyaa051
Advance Access Publication Date: 3 February 2021
Investigation
Highlighted Article

Trans-acting genetic variation affects the expression of adjacent genes

Krisna Van Dyke[1], Sheila Lutz[1], Gemechu Mekonnen[1], Chad L. Myers[2], and Frank W. Albert[1,*]

[1]Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN 55455, USA
[2]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA

*Corresponding author: Department of Genetics, Cell Biology, & Development, University of Minnesota, 6-160 Jackson Hall, 321 Church St SE, Minneapolis, MN 55455, USA. falbert@umn.edu

Abstract

Gene expression differences among individuals are shaped by *trans*-acting expression quantitative trait loci (eQTLs). Most *trans*-eQTLs map to hotspot locations that influence many genes. The molecular mechanisms perturbed by hotspots are often assumed to involve "vertical" cascades of effects in pathways that can ultimately affect the expression of thousands of genes. Here, we report that *trans*-eQTLs can affect the expression of adjacent genes via "horizontal" mechanisms that extend along a chromosome. Genes affected by *trans*-eQTL hotspots in the yeast *Saccharomyces cerevisiae* were more likely to be located next to each other than expected by chance. These paired hotspot effects tended to occur at adjacent genes that also show coexpression in response to genetic and environmental perturbations, suggesting shared mechanisms. Physical proximity and shared chromatin state, in addition to regulation of adjacent genes by similar transcription factors, were independently associated with paired hotspot effects among adjacent genes. Paired effects of *trans*-eQTLs can occur at neighboring genes even when these genes do not share a common function. This phenomenon could result in unexpected connections between regulatory genetic variation and phenotypes.

Keywords: trans-eQTLs; gene expression; coexpression; genetic variation; adjacent

Introduction

Genetic variation among individuals influences phenotypic traits (Lynch and Walsh 1998). Many of the DNA variants that shape phenotypes do so by altering gene expression (Zheng *et al.* 2011; Maurano *et al.* 2012; Gusev *et al.* 2014; Albert and Kruglyak 2015; Signor and Nuzhdin 2018). Genomic regions that contain variants that influence the abundance of the RNA of a given gene are called expression quantitative trait loci (eQTLs) (Brem 2002; Sun and Hu 2013; Albert and Kruglyak 2015). eQTLs can be classified according to their location in the genome relative to the genes they affect, as well as by their mechanism of action. Local eQTLs alter the expression of genes in their physical vicinity, usually by perturbing *cis*-acting regulatory mechanisms. By contrast, distant eQTLs act in *trans* by changing the abundance, cellular localization, or activity of a diffusible factor. These factors can act on genes throughout the genome, and most *trans*-eQTLs influence genes that are distant from the eQTL, typically on different chromosomes (Albert and Kruglyak 2015).

Trans-acting variation has been extensively studied in crosses in model organisms (Brem 2002; Hubner *et al.* 2005; West *et al.* 2007; Smith and Kruglyak 2008; Rockman *et al.* 2010; Orozco *et al.* 2012; Lewis *et al.* 2014; Cesar *et al.* 2018; Everett *et al.* 2020). For example, genetic mapping of mRNA levels in 1012 recombinant offspring from a cross between a laboratory strain (BY, a close relative of the genome reference strain S288C) and a vineyard isolate (RM) of the yeast *Saccharomyces cerevisiae* yielded tens of thousands of eQTLs (Albert *et al.* 2018). These eQTLs accounted for more than half of the genetic variation in mRNA abundance (a median across genes of 72%), suggesting that the dataset comprises the great majority of eQTLs with at least modest effect sizes. Almost all genes were influenced by multiple *trans*-eQTLs. The summed *trans*-eQTL effects accounted for 2.6-fold more expression variation than those of local eQTLs. In agreement with similar estimates in human populations (Grundberg *et al.* 2012; Wright *et al.* 2014), the BY/RM cross shows that *trans*-eQTLs are the major source of genetic variation in gene expression.

Trans-eQTLs tend to cluster at certain locations in the genome. These *trans*-eQTL "hotspots" can affect the expression of large numbers of genes, and have been observed in yeasts (Brem 2002; Smith and Kruglyak 2008; Clément-Ziza *et al.* 2014), nematodes (Rockman *et al.* 2010), plants (West *et al.* 2007; Fu *et al.* 2009), rodents (Bystrykh *et al.* 2005; Chesler *et al.* 2005; Orozco *et al.* 2012; Langley *et al.* 2013), and cattle (Cesar *et al.* 2018), and may also exist in humans (Brynedal *et al.* 2017). In the BY/RM yeast cross, over 90% of the *trans*-eQTLs mapped to 102 hotspot regions (Albert *et al.* 2018). Several of these hotspots have been resolved to single causal genes or variants (Brem 2002; Yvert *et al.* 2003; Smith and Kruglyak 2008; Zhu *et al.* 2008; Lutz *et al.* 2019). The BY/RM hotspots can affect hundreds and in some cases thousands of genes.

Identifying the mechanistic connection between an eQTL hotspot and the genes it affects in *trans* can be challenging.

Received: October 13, 2020. **Accepted:** December 16, 2020

Typically, the causal variant or variants in the hotspot are assumed to alter the sequence or abundance of a protein product, such as a transcription factor, that then acts on functionally related genes. In turn, the changes in expression of these genes are propagated through cellular networks to indirectly affect many other genes. In this paper, we refer to this type of mechanistic connection as "vertical" propagation.

The expression of eukaryotic genes is partially shaped by their location on the chromosome. Specifically, adjacent genes tend to show correlated expression changes in response to a broad range of stimuli (Cohen *et al.* 2000; Hurst *et al.* 2004; Michalak 2008; Kustatscher *et al.* 2017; Sun and Zhang 2019). In part, this coexpression is due to clustering of genes with similar function and regulation by similar sets of transcription factors (TFs) (Cohen *et al.* 2000; Allocco *et al.* 2004; Arnone *et al.* 2012). In addition to such vertical coregulation, "horizontal" processes including shared local chromatin states that encompass adjacent genes may facilitate coexpression even for genes that do not share a common function and that are not regulated by the same TFs (Batada *et al.* 2007; Ebisuya *et al.* 2008; Arnone *et al.* 2014; Sun and Zhang 2019). Coexpression of neighboring genes decreases with physical distance (Cohen *et al.* 2000; Quintero-Cadena and Sternberg 2016). In addition to chromatin state, non-specific promiscuous effects of *cis*-regulatory elements have been proposed to contribute to this distance-dependent, horizontal coexpression (Cohen *et al.* 2000; Quintero-Cadena and Sternberg 2016).

Here, we asked if the mechanisms that cause coexpression of adjacent genes may also shape the effects of *trans*-eQTL hotspots. We made use of the extensive set of *trans*-eQTLs identified in the BY/RM cross (Albert *et al.* 2018) to show that *trans*-eQTL hotspots affect the expression of adjacent genes more frequently than expected by chance. These paired hotspot effects occur primarily at adjacent genes that show correlated gene expression responses to a broad variety of perturbations. Paired hotspot effects can be partially explained by vertical regulation of adjacent genes by similar transcription factors. However, paired hotspot effects at adjacent genes were independently associated with features indicative of horizontal mechanisms: the physical proximity of the genes in a pair as well as similar chromatin states. These results suggest that the effects of *trans*-eQTLs can extend from an affected gene to adjacent genes via horizontal mechanisms, providing an additional mode by which *trans*-acting regulatory variation can affect gene regulatory networks and cell biology.

Materials and methods

Unless otherwise specified, analyses were conducted in R 3.6.0 (R Development Core Team 2010), using the following packages. C++ code was run using Rcpp 1.0.4 (Eddelbuettel and François 2011). Data processing was performed using tidyverse 1.3.0 (Wickham *et al.* 2019). Genomic range data were imported using Rtracklayer 1.46.0 (Lawrence *et al.* 2009) and processed using Genomic Ranges 1.38.0 (Lawrence *et al.* 2013) and Biostrings 2.54.0 (H. Pagès 2017). Plot color schemes were provided by the Viridis color package originally developed as part of Matplotlib (J. D. Hunter 2007: 200).

Gene annotations

Gene models were obtained from yeastmine (Fisk *et al.* 2006; Balakrishnan *et al.* 2012). The BY and RM strains differ in the presence or absence of stretches of genes close to chromosome ends. *Trans*-eQTL effects on these subtelomeric genes may appear to be highly correlated simply due to their shared presence or absence in the BY/RM segregants. Therefore, we removed subtelomeric genes, defined as all genes whose start or stop codon was located outside of the genetic markers forming the linkage map in Source Data 3 from Albert *et al.* (2018). We retained only "verified" open reading frames (ORFs) (Fisk *et al.* 2006) because some genes not annotated to be in this category (such as "dubious" ORFs or noncoding genes) overlap verified ORFs (sometimes on the antisense strand), complicating analyses of intergenic distances and transcription factor binding site (TFBS) annotations.

Hotspot effects

Hotspot effects were represented in a "hotspot matrix." This matrix was generated from previously published data [Source Data 9 from Albert *et al.* (2018)], which cataloged the effects of all 102 hotspots (columns) on 5629 expressed genes (rows). We removed 734 genes that were not verified ORFs (Fisk *et al.* 2006) and introduced a row of zeros for 17 genes that were expressed in the segregant data but that were not affected by any of the 102 hotspots. Genes were defined as expressed if they were present in the $\log_2$(TPM) expression value table in Source Data 1 from (Albert *et al.* 2018). The rows of the hotspot matrix were ordered by chromosome and the position of the genes on each chromosome. We removed subtelomeric genes and retained only verified ORFs as described above. After these filters, the hotspot matrix contained 4912 genes, of which 4878 were affected by at least one hotspot.

Doublet definition

A doublet was defined as two adjacent genes with nonzero effects of equal sign from a given hotspot. A single gene can be involved in up to two doublets, one with each of its two adjacent genes. Genes at the ends of chromosomes can only be involved in one doublet. For each hotspot, we counted the number of doublets among the genes affected by the hotspot.

Permutation tests

To assess whether the observed doublet count exceeded random chance, we used a permutation test. For each hotspot, we performed 100,000 permutations, in which the given hotspot effects were randomly shuffled across the genome. In each permutation, we counted the number of doublets. To determine *P*-values, we counted the number of permutations with a number of doublets greater than or equal to those observed in the real data for the corresponding hotspot and divided this count by the number of permutations. Multiple testing correction across 102 hotspots was performed by dividing the significance threshold of $P = 0.05$ by 102. This process was repeated for triplets, quadruplets, quintuplets, and sextuplets. For permutations that excluded divergently oriented gene pairs, we retained all genes and their hotspot effects but did not count doublets that occurred at divergent gene pairs. If doublets tend to occur primarily at divergent gene pairs, this strategy will tend to inflate the number of doublets observed at non-divergent pairs in the permutations compared to the real data, resulting in a conservative test of excess pairing of hotspot effects. To test whether a small number of unique gene pairs affected by many hotspots drove the overall enrichment, we performed permutations excluding the 20 adjacent gene pairs affected by the most hotspots. Due to ties, 23 unique pairs were excluded, resulting in a more conservative test. The number of hotspots affecting these excluded gene pairs ranged from 12 to 23. For context, 81 of the hotspots (39 significant hotspots) affected an excess of 23 or fewer adjacent gene pairs, such that exclusion of the top 23 gene pairs would be expected to drastically

reduce the number of hotspots that affect an excess of adjacent gene pairs if all hotspots were to affect the same set of unique adjacent gene pairs. Pair exclusion was performed as described above.

Coexpression matrices

Expression datasets (Supplementary File S1) were ordered by chromosome and gene order, with genes in rows and perturbations in columns. We universe-normalized each dataset by removing genes not present in the BY/RM expression dataset (Source Data 1 from (Albert *et al.* 2018)) and by adding a null-row for any genes that were present in the BY/RM expression dataset but absent from the given expression dataset. Quantile normalization was performed using the normalize.quantiles function from preprocessCore (Bolstad 2017), and used to generate Spearman correlation matrices via Hmisc's Rcorr function (Harrell 2020). The datasets from Albert *et al.*, 2018, Brem & Kruglyak 2005, Fleming *et al.* 2002, Hughes *et al.* 2000, and Schurch *et al.*, 2016 were treated similarly, with the following changes:

Fleming et al. (2000) and Hughes et al. (2000): Data had been preprocessed following the procedures described in (Huttenhower *et al.* 2006). Universe normalization and quantile normalization were performed as above.

Brem and Kruglyak (2005): Data had been preprocessed following the procedures described in (Huttenhower *et al.* 2006). We used the mean of the values of the microarrays using both dyes as the expression measurement for each gene. Universe normalization and quantile normalization were performed as above.

Schurch et al. (2016): Processed counts were obtained from GitHub (https://github.com/bartongroup/profDGE48/blob/master/Preprocessed_data/WT_countdata.tar.gz). Raw data, from which these processed counts were derived, is available at the European Nucleotide Archive repository (http://www.ebi.ac.uk/ena/data/view/ERX425102) under the accession number PRJEB5348, where they had been deposited by (Gierliński *et al.* 2015). Counts for each gene were universe-normalized as above and filtered to retain only genes with at least a single read in at least half of the samples. Read counts were normalized by dividing them by the length of the gene. A value of 0.5 was added to each count, followed by taking the natural log. The resulting values were centered within each sample by subtracting the mean expression of all expressed genes in the given sample. Data then underwent quantile normalization as above.

Albert et al. (2018): Log(TPM) reads in Source Data 1 from (Albert *et al.* 2018) were processed to remove the effect of genetic variation, processing batch, and growth phase. As a proxy for the growth phase, we used the optical density at the time the cultures had been collected (Source Data 2 from (Albert *et al.* 2018)). For each gene, we built a fixed-effects linear model using processing batch, optical density, and peak markers of each significant *cis*- and *trans*-eQTL reported for this gene in Source Data 3 from (Albert *et al.* 2018). The residuals from these models were centered by subtracting the mean of all residuals for the given segregant. We then performed quantile normalization as above. To gauge if removal of eQTL effects had been successful, we examined genes with strong eQTLs. Prior to removal, such eQTLs cause the expression values of their target gene to show a bimodal distribution. Specifically, the gene *STE2* is affected by a strong *trans*-effect arising from the mating-type locus, and the gene *HO* is affected by a strong local effect caused by an engineered deletion of this gene in the RM strain. Both genes had bimodal distributions prior to correction and unimodal residuals of their respective linear models (Supplementary Figure S1).

Comparison of similarity matrices

The various similarity matrices were compared using Spearman rank-based correlations between the upper triangles of the two given matrices.

Quantifying paired hotspot effects across hotspots

To capture the degree to which both genes in a given pair were affected by hotspots, we constructed a matrix summarizing hotspot effect. We initially set every possible gene pair (including adjacent and nonadjacent gene pairs) to a value of zero. If a hotspot affected both genes in a pair in the same direction, the score was increased by one. If both genes in a pair were affected in different directions, the score was decreased by one. These summarized hotspot effects are able to capture gene pairs that are in repulsion, such that they tend to be affected by multiple hotspots in opposite directions. We chose this metric for comparison to coexpression because gene pairs with anticorrelated expression may also show pairing of hotspot effects with opposite direction.

In our linear models of possible mechanisms underlying paired hotspot effects at adjacent genes, we counted, for each adjacent gene pair, the number of hotspots that affected both members of the pair in the same direction. Although this metric, which ranged from zero to 23, ignores cases in which a hotspot affects the two genes in a pair in opposite direction, we chose it for two reasons: (1) the absence of negative values allowed us to use a count-based generalized linear model (see below), and (2) conceptually, it is unclear how to relate negative values for the 28% of adjacent gene pairs that tend to be affected by hotspots in opposite direction to features that cannot take on negative values (*e.g.*, TFBS similarity or distance between genes).

Regulation of genes by similar sets of transcription factors

We used a curated collection of transcription factor regulation in the yeast genome (Monteiro *et al.* 2020). These data were binary, with a value of zero indicating that a given TF is not annotated as regulating a given gene, and a value of one indicating that it is. TF profile similarity between all gene pairs was calculated using the Jaccard index (Veerla and Höglund 2006). Adjacent genes had higher similarity in their TFBS profiles than nonadjacent genes (median 0.294 *vs* 0.275, Wilcoxon test $P < 2.2 \times 10^{-16}$), as expected (Hershberg *et al.* 2005). We only used the values for adjacent genes in our linear models.

Nucleosome occupancy measures

We used two nucleosome occupancy datasets: an ATAC-seq dataset (Schep *et al.* 2015) and a dataset based on a chemical cleavage assay (Chereji *et al.* 2018). Specifically, we used the nucleosome occupancy files available at the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE66386 and GSE97290, respectively. Data had been pre-processed with scores indicating low to high nucleosome occupancy for a given genomic range. For the chemical cleavage data, we used means for occupancy at each base pair across three replicates.

In each dataset, we calculated two metrics. (1) Mean occupancy across adjacent gene bodies was defined as the mean nucleosome occupancy of the region from 500 bp upstream of the first gene to 500 bp downstream of the second gene. (2) Similarity

in occupancy of adjacent genes was computed as the log of the inverse of the difference in mean nucleosome occupancy score of both gene bodies.

Similarity in chromatin marks at baseline and during stress response

Chromatin similarity at baseline and change in chromatin similarity in response to application of the cellular stressor diamide were calculated using ChIP-seq data published by (Weiner *et al.* 2015). Specifically, we used the nucleosome maps (Supplementary Table S2 in Weiner *et al.*) and normalized marks (Supplementary Table S3 in Weiner *et al.*) available separately at (https://www.cell.com/molecular-cell/fulltext/S1097-2765(15)00094-5) or combined at the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE61888. Supplementary Table S2 from Weiner *et al.* contains the mapped positions of 66,360 nucleosomes in wild-type yeast (strain BY4741) and Supplementary Table S3 from Weiner *et al.* contains the levels of all 26 chromatin modifications normalized to the data in their Supplementary Table S2. Thus, the values in their Supplementary Table S3 represent estimates of the log ratio of ChIP coverage with respect to each sample input, with quantile normalization performed within the time series for each chromatin mark.

To calculate similarity in baseline chromatin state between a given pair of adjacent genes, we created a vector with the normalized levels of the 26 chromatin marks on the +1 nucleosomes for each of the two genes in mid-log growth, before the application of the diamide stressor. Baseline chromatin similarity was then calculated as the spearman correlation between the vectors of the two genes. Similarity in chromatin change was calculated as the spearman correlation between vectors that contained the difference in normalized levels of the 26 chromatin marks on the +1 nucleosomes of adjacent genes between baseline and 15 minutes after the cells were exposed to diamide.

Features influencing the number of hotspots that affect gene pairs

To gauge the relative importance of various features on paired hotspot effects at adjacent genes, we fitted a negative binomial generalized linear model, using the glm.nb() function from MASS 7.3-51.6 (Venables *et al.* 2002). The response variable was the number of times adjacent gene pairs were affected by the same hotspot in the same direction. Pairs never affected by the same hotspot in the same direction were omitted. Some genes were affected by more hotspots than others, which trivially increases the number of times they could form a doublet with their adjacent gene. To control for this, we included two exposure variables as offsets, which gave the natural log of the number of times the given gene was affected by hotspots.

As predictor variables, our models included: (1) the orientation of adjacent genes with respect to each other (*i.e.*, convergent, tandem, or divergent), (2) the similarity in how gene pairs were regulated by transcription factors, (3) physical proximity on the chromosome, which was the natural log of the inverse of distance in bp between the start codons of the two genes, (4) similarity in chromatin marks between +1 nucleosomes, and (5) similarity in stress-induced changes in chromatin marks between +1 nucleosomes. To calculate the significance of each of these features, we fit a full model including all these features, as well as reduced models that omitted the given feature. Model comparison between the full model and each reduced model was performed using type III anova as implemented by the Anova() function in the car 3.0-8 package (Fox and Weisberg 2019). We also fit equivalent models to subsets of the data, separately for each of the three possible orientations of adjacent gene pairs.

Functional similarity between genes

Gene ontology method: We used the GoSemSim package (Yu *et al.* 2010) with the Wang method (Wang *et al.* 2007) and best match average scoring to deal with the semantic similarity scores of multiple GO biological process terms. GO data were obtained via the org. Sc.sgd.db 3.10.0 package (Carlson 2017).

Genetic interaction similarity method: These data were preprocessed and derived from Pearson correlations provided at [https://thecellmap.org/costanzo2016/ (Costanzo *et al.* 2016)]. This matrix was universe-normalized by removing any columns and rows corresponding to genes not present in the BY/RM data (Albert *et al.* 2018) and adding null columns for genes present in the BY/RM data that were not present in the interaction dataset.

Genes regulated by Oaf1

All genes directly regulated by Oaf1 (SGD: S000000048) were obtained from previously published data [(Bergenholm *et al.* 2018) Supplementary DATA SET S1]. Genes were considered to be regulated by Oaf1 if their promoters were bound by Oaf1 in any of the four conditions studied by (Bergenholm *et al.* 2018).

Data availability

Expression datasets used in this work can be found at the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo) with the accession numbers: GSE1990, GSE2583, GSE11452, GSE25909, GSE42536, GSE24771, GSE66386, GSE97290, and GSE61888. A full inventory of datasets used in this work, including those without an accession number, are detailed in Supplementary File S1. Code and datasets used in this work are available at (https://github.com/Krivand/Trans-acting-genetic-variation-affects-the-expression-of-adjacent-genes). All figures, intermediary processing objects, tables, and analyses can be generated directly from the code on GitHub by following the instructions at the top of the R script. Supplementary File S2 contains summary statistics for hotspot permutations, real doublet counts, and *P*-values. *P*-values listed as "0" indicate that no count greater than or equal to that of the real data was observed in the permutations. Thus, although the corresponding *P*-value is $<1 \times 10^{-5}$, we report these cases as "0" to avoid Excel formatting cells with a "<" sign as text. Supplementary File S3 contains correlation coefficients and *P*-values among all predictors used in our negative binomial generalized linear model. Supplementary File S4 contains the output of type III ANOVAs for generalized linear models. Supplementary Figure S1 shows that removal of eQTL effects from the Albert *et al.* (2018) expression data were successful. Supplementary Figure S2 contains the untransformed data used in Figure 1. Supplementary Figure S3 contains the difference in doublets counted for each hotspot and the median of matched permutations for that hotspot. Supplementary Figure S4 contains the distributions of Rho values for each coexpression matrix used in this work. Supplementary Figure S5 contains the distributions of predictor and outcome variables used in the linear models detailed in Figure 4B, separated by the orientation of each gene pair. Supplementary Material is available at figshare: https://doi.org/10.25386/genetics.13413518.

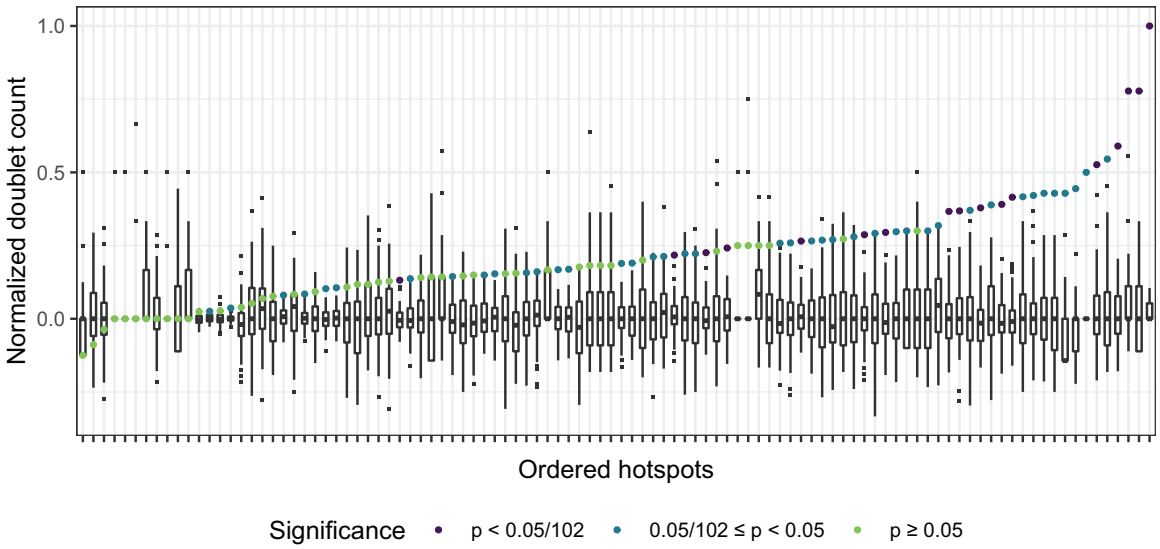

Figure 1 The number of doublets in real data (colored circles) are plotted against the number of doublets in permuted data (white boxes). Boxes extend from the 25th to 75th percentile. Whiskers extend from each box to largest or smallest values within 1.5 times the difference in the 25th to 75th percentile. Data beyond the whiskers are considered outliers and plotted as small black squares. Circles are colored by significance of the excess of doublets at the given hotspot. To aid visualization, the doublet count for each hotspot and its matched permutations were normalized by subtracting the median and dividing by the maximum value. See Supplementary Figure S2 for non-normalized results. Hotspots are ordered along the x-axis based on the value of the normalized doublet count in the real data.

Results

Trans-eQTL hotspots affect more adjacent gene pairs than expected by chance

We analyzed eQTLs mapped in the BY and RM cross (Albert *et al.* 2018). Briefly, the dataset comprises 36,498 eQTLs, of which 33,529 act in *trans*. These *trans*-eQTLs cluster at 102 hotspot locations, which affect 12 to 4093 of 4912 verified open reading frames (ORFs) with detectable expression. Genes are affected by a median of eleven hotspots, and almost all genes (98%) are affected by at least one hotspot.

We asked whether hotspots affected genes located next to each other on the chromosome more often than expected by chance. To address this question, we focused on "doublets", which we defined as pairs of adjacent verified ORFs whose mRNA abundances were both increased or both decreased by a given hotspot. Genes close to chromosome ends were excluded from the analysis (*Materials and Methods*). For each hotspot, we constructed 100,000 permutations by randomly assigning the observed effects of the hotspot to genes, irrespective of their location in the genome. This permutation strategy preserved the number and magnitude of hotspot effects, but broke any relationship between hotspot effects and a gene's location relative to its neighbors. We compared the number of observed doublets to the distribution of doublets in the hotspot-specific permutations.

Of the 102 hotspots, 91 displayed more doublets than their matched permutation median. This number was significantly higher than expected by chance (Binomial Test, $P < 2.2 \times 10^{-16}$). At 58 hotspots, the observed number of doublets exceeded that seen in 95% of the permutations (Figure 1, Supplementary File S2, Figure S2), which was more than the five hotspots expected to reach this threshold by chance (Binomial Test, $P < 2.2 \times 10^{-16}$). After Bonferroni correction, 17 hotspots showed a significant excess of doublets (Table 1, Figure 1, and Supplementary File S2 and Figure S2). At the 58 nominally significant hotspots, the number of excess doublets ranged from one to 80, with a median excess of 18 doublets (Supplementary Figure S3).

Coexpression of adjacent genes might occur due to shared promoter sequences when the genes are transcribed in divergent orientations. For example, in the yeast genome, genes in the ribosome and rRNA biosynthesis regulons tend to occur in adjacent, divergently expressed pairs that show highly correlated expression (Kraakman *et al.* 1989; Wade *et al.* 2006; Arnone and McAlear 2011; Arnone *et al.* 2014; Eldabagh *et al.* 2018). To test if the excess of paired hotspot effects was only due to gene pairs with divergent orientation, we excluded all divergent gene pairs from our permutation analysis (*Materials and Methods*). The number of doublets remained significantly higher than expected by chance (Table 1, Supplementary File S2). Further, the excess of doublets was not due to a small number of unique adjacent gene pairs that were affected by many hotspots (Table 1).

The number of sets of three adjacent genes affected by hotspots in the same direction was also higher than expected by chance ("triplets" in Table 1; Supplementary File S2), as were sets of four adjacent genes ("quadruplets") and sets of five adjacent genes ("quintuplets"), albeit at weaker significance. No enrichment was found for sets of six adjacent genes ("sextuplets" in Table 1). Thus, the mechanisms causing paired hotspot effects extend beyond immediate gene neighbors and eventually taper off as a function of distance. Overall, these results show that the genes affected by hotspots are not randomly located throughout the genome, suggesting that the effects of *trans*-acting variation may be shaped by spatial patterns of gene regulation along chromosomes.

Gene coexpression patterns are shared across diverse transcriptome datasets

Trans-eQTLs could affect the expression of adjacent genes via the same mechanisms that lead to gene coexpression (Michalak 2008; Brown *et al.* 2013; Sun and Zhang 2019). To quantitatively test this hypothesis, we first examined coexpression patterns in eleven transcriptome datasets (Figure 2) (Hughes *et al.* 2000; Fleming *et al.* 2002; Klevecz *et al.* 2004; Brem and Kruglyak 2005; Knijnenburg *et al.* 2009; Simola *et al.* 2010; Lenstra *et al.* 2011;

Table 1 Enrichment of adjacent genes affected by *trans*-eQTL hotspots

	Number of hotspots that pass multiple test correction*	P-value** for hotspots that pass multiple test correction	Number of hotspots with $P<0.05$	P-value** for hotspots with $P<0.05$	Number of hotspots that exceed median permutation	P-value** for hotspots that exceed median permutation
Doublets	17	5×10^{-38}	58	6×10^{-48}	91	4×10^{-17}
Doublets without divergent pairs	17	5×10^{-38}	57	1×10^{-46}	90	3×10^{-16}
Doublets without top 23 most affected adjacent gene pairs	14	3×10^{-30}	54	2×10^{-42}	91	4×10^{-17}
Triplets	7	1×10^{-13}	30	2×10^{-15}	66	0.002
Quadruplets	4	2×10^{-7}	18	3×10^{-6}	40	0.99
Quintuplets	2	0.002	12	0.005	19	1
Sextuplets	1	0.049	3	0.89	10	1

* The multiple-test corrected *P*-value was obtained by dividing $P = 0.05$ by the number of hotspots (102).
** *P*-values were computed using a binomial test of the alternative hypothesis of a greater number of hotspots than expected by chance. The number of hotspots expected by chance are 0.05 hotspots with multiple testing correction (*i.e.*, $P \leq 0.05/102$), 5.1 hotspots at $P < 0.05$, and 51 hotspots at $P < 0.5$.

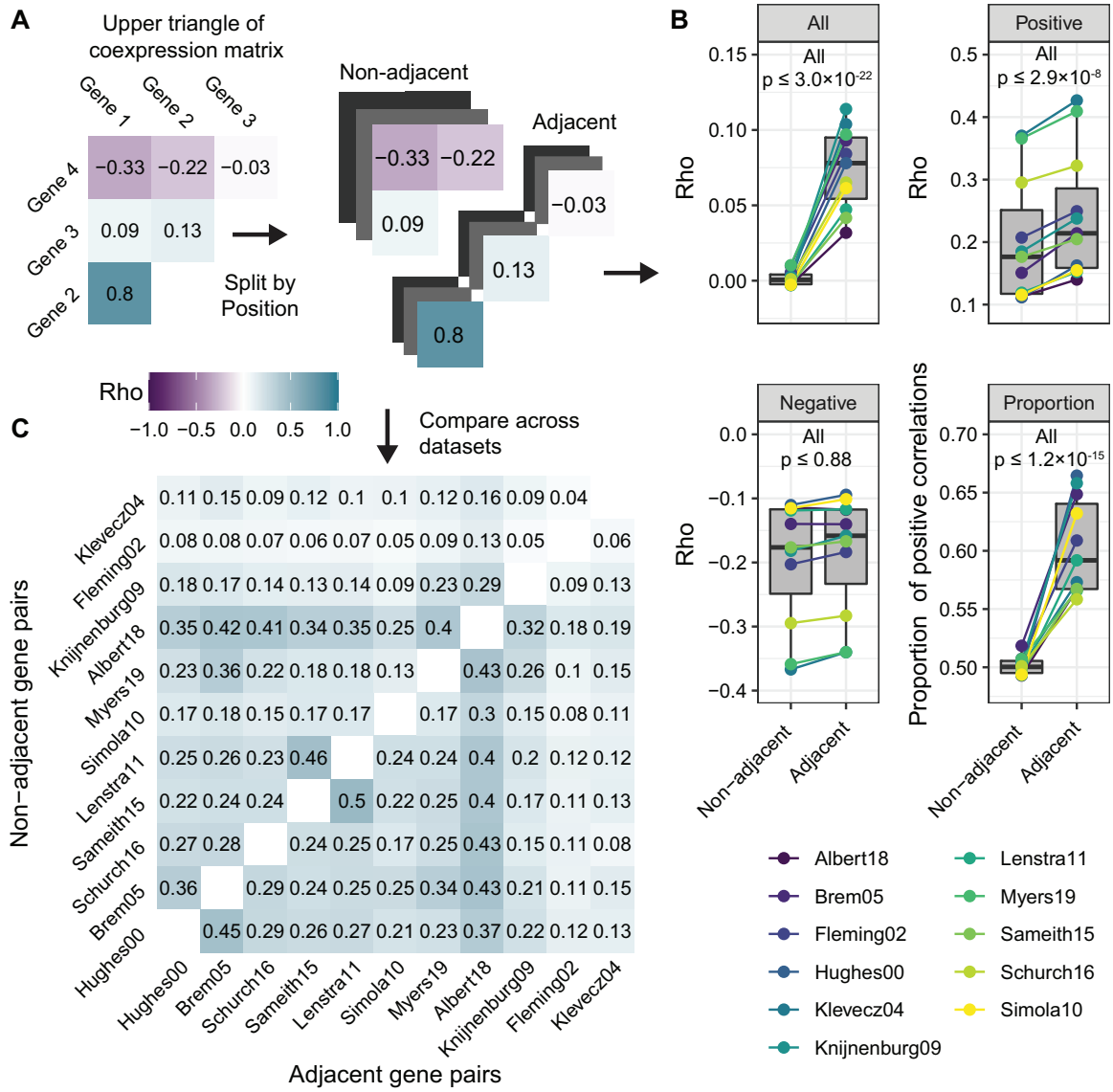

Figure 2 Coexpression analysis. (A) Schematic illustrating the separation of gene–gene coexpression matrices by adjacent versus nonadjacent genes. (B) Comparison of coexpression for adjacent genes versus nonadjacent genes in each of the eleven datasets, which are indicated by different colors. The four panels are, clockwise from top left: coexpression values (Rho) for all gene pairs, gene pairs with positive coexpression, gene pairs with negative coexpression, and the proportion of gene pairs with positive coexpression. (C) Correlations between coexpression datasets for neighboring *vs* non-neighboring genes. All correlations between datasets were statistically significant ($P \leq 8.6 \times 10^{-5}$). The color scale is the same as in panel (A), and rho values are indicated in each cell.

Sameith *et al.* 2015; Schurch *et al.* 2016; Albert *et al.* 2018; Myers *et al.* 2019). Each dataset comprised multiple genome-wide transcriptome experiments ($n = 32$–1012, median = 162) collected in yeast strains exposed to a wide range of perturbations, including engineered gene deletions and natural genetic variation, environmental stressors such as chemical treatments, and deprivation of specific nutrients. These datasets were independently collected by different laboratories using various microarray or RNA-seq techniques. We also included the expression data from which the hotspots studied here were identified (Albert *et al.* 2018). To avoid possible circularity in these data, in which the perturbations caused by the hotspots could induce correlations among adjacent genes, we used linear models to remove the effects of eQTLs on segregant expression data (*Materials and Methods*).

To measure gene–gene coexpression in each of these transcriptome datasets, we built Spearman rank-based correlation matrices (Figure 2A). In all datasets, gene pairs showed a broad range of coexpression values, ranging from strongly positive to strongly negative correlations (Supplementary Figure S4). Adjacent genes were more strongly correlated than nonadjacent genes in each dataset ($P \leq 3 \times 10^{-22}$, Wilcoxon test, Figure 2B), consistent with earlier reports (Cohen *et al.* 2000; Raj *et al.* 2006; Batada *et al.* 2007; Ebisuya *et al.* 2008; Michalak 2008). This overall correlation among adjacent genes was mostly driven by stronger positive as opposed to negative correlations, and there was a significant excess of positive correlations among adjacent compared to nonadjacent genes (Figure 2B).

Beyond these general patterns, we asked if the coexpression values for individual pairs of genes are quantitatively similar between the eleven different datasets. To do so, we computed the correlation between the coexpression relationships observed in each pair of datasets for either adjacent genes (Figure 2C, lower triangle) or nonadjacent genes (Figure 2C upper triangle). Both showed significant nonzero correlation in all comparisons (adjacent: $P < 9 \times 10^{-5}$; nonadjacent: $P < 2.2 \times 10^{-16}$). The correlation of coexpression of adjacent gene pairs between datasets was stronger than that of nonadjacent gene pairs (median rho = 0.24 *vs* 0.18, paired Wilcoxon test $P < 2.2 \times 10^{-16}$). These results show that coexpression relationships are broadly consistent across diverse environmental contexts and perturbations, especially for adjacent genes.

Trans-eQTL hotspots tend to affect adjacent gene pairs with correlated expression

We tested if *trans*-eQTL hotspots tend to affect both genes in gene pairs that show coexpression. To do so, we summarized hotspot effects on a given gene pair into a score (*Materials and Methods*). Positive scores indicated that both genes in a pair are affected in the same direction (either both genes have increased or decreased expression) by multiple hotspots. Negative scores indicated pairs in which both genes were affected in opposite directions by multiple hotspots. A score near zero indicated pairs that were targeted by multiple hotspots but without a preference for the same or opposite direction, or gene pairs that were not affected by many hotspots. Adjacent gene pairs had higher scores (Wilcoxon test, $P \leq 1.1 \times 10^{-16}$) and a higher proportion of positive scores (Test of proportions, $P \leq 8 \times 10^{-7}$) than nonadjacent gene pairs, showing that this metric captured the excess of paired hotspot effects at adjacent genes reported above.

We compared these summarized hotspot effects on gene pairs to each of the eleven coexpression matrices, separately for pairs of adjacent genes and nonadjacent genes. The degree of hotspot effect pairing was significantly correlated with coexpression in all eleven datasets (Figure 3A). This agreement was present for adjacent genes (median of eleven Rhos = 0.21, all $P < 2.2 \times 10^{-16}$) and for nonadjacent genes (median of eleven Rhos = 0.18, all $P < 2.2 \times 10^{-16}$, Figure 3A), with stronger agreement for adjacent genes (paired Wilcoxon test: $P = 5 \times 10^{-4}$). The correlation between hotspot effect pairing and coexpression arose primarily, but not exclusively, from gene pairs with stronger coexpression (Figure 3B). Thus, hotspots tend to affect both genes in pairs of genes that show stronger coexpression, particularly among adjacent genes. A parsimonious explanation for this agreement is that the molecular processes that cause coexpression are also responsible for spatial coupling of the effects of *trans*-eQTL hotspots.

Paired hotspot effects on adjacent genes are driven by regulation by similar sets of transcription factors, physical proximity, and similar chromatin dynamics

We asked which mechanisms might underlie paired hotspot effects on adjacent genes. Specifically, we considered molecular features previously identified to affect adjacent gene coexpression, including vertical regulation by similar sets of transcription factors, as well as horizontal features including shared chromatin state, orientation of the genes relative to one another, and the distance between the genes. To relate hotspot effects on adjacent genes to these features, we counted the number of hotspots that affected both genes in adjacent pairs in the same direction. These counts ranged from zero to 23 (median = 2).

In *S. cerevisiae*, adjacent genes tend to be regulated by similar sets of transcription factors (Hershberg *et al.* 2005), likely because genes with similar functions tend to be located together in the genome [see below and (Cohen *et al.* 2000; Eldabagh *et al.* 2018)]. To examine whether paired hotspot effects may result from regulation by similar sets of transcription factors at some adjacent genes, we gathered annotated TFBS (Monteiro *et al.* 2020) in the intergenic region upstream of each gene and calculated the similarity of these TFBS annotations among adjacent gene pairs. TFBS similarity was weakly but significantly correlated with hotspot effect pairing (rho = 0.033, $P = 0.02$). Thus, some paired hotspot effects at adjacent genes are due to regulation by similar sets of transcription factors. This mechanism is consistent with traditional vertical propagation of hotspot effects to transcriptional target genes. However, the weak magnitude of the correlation between TFBS similarity and paired hotspot effects suggests that other mechanisms also play a role.

A second possible mechanism that could create paired hotspot effects is shared chromatin state across neighboring genes, such that alterations to chromatin at one gene spread outward and also affect the other gene (Raj *et al.* 2006; Brown *et al.* 2013; Arnone *et al.* 2014). We first considered two independent measures of nucleosome occupancy across the bodies of adjacent genes (Schep *et al.* 2015; Chereji *et al.* 2018), but neither was correlated with paired hotspot effects (Chereji mean occupancy across adjacent gene bodies: $P = 0.70$; Chereji similarity in occupancy of adjacent genes: $P = 0.06$, Schep mean occupancy, $P = 0.16$; Schep similarity in occupancy, $P = 0.36$). These chromatin data, which were measured in normally growing, isogenic cultures, may not well reflect chromatin changes in response to a perturbation, such as a *trans*-eQTL. Therefore, we obtained data on 26 chromatin marks measured as a time course during the response to exposure to diamide, a stressful stimulus (Weiner *et al.* 2015). We focused on "+1" nucleosomes because of their important roles in transcriptional regulation (Jansen and Verstrepen 2011; Rando

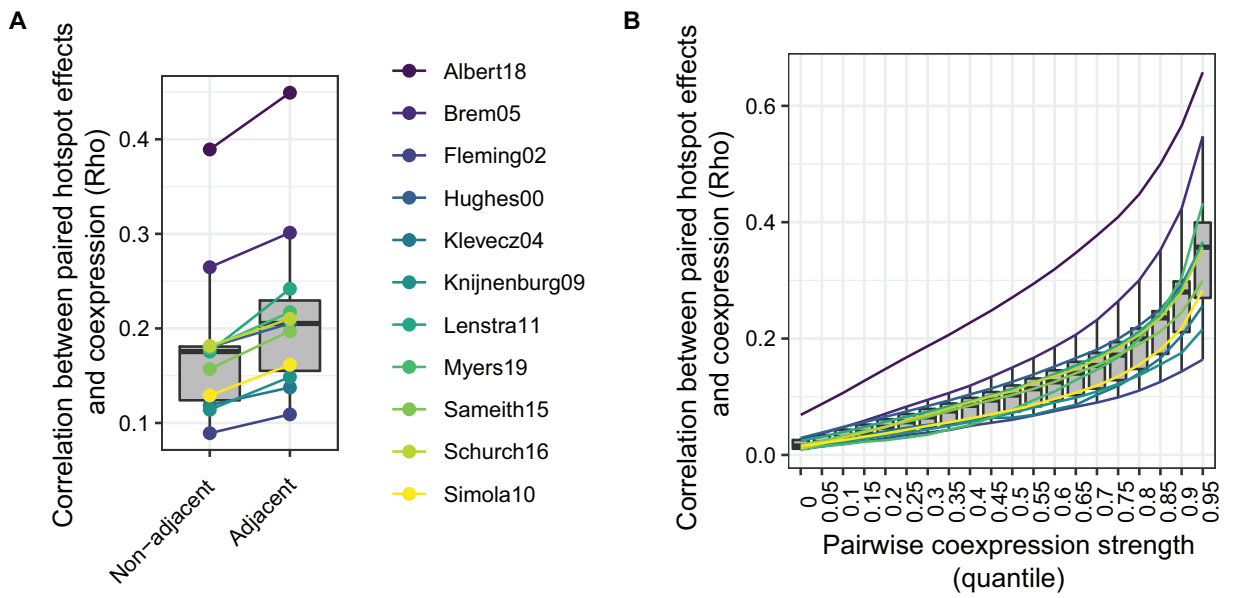

Figure 3 Agreement between paired hotspot effects and gene–gene coexpression. (A) Agreement between paired hotspot effects and the eleven coexpression datasets (shown in different colors), for nonadjacent and adjacent gene pairs. (B) Agreement between paired hotspot effects and coexpression datasets as a function of increasing coexpression. All Rho values, in all bins and for all datasets, were significantly different from zero ($P \leq 1.3 \times 10^{-16}$).

and Winston 2012). We computed two metrics: 1) similarity in chromatin state at baseline, which we defined as the correlation between the levels of chromatin marks on the +1 nucleosomes of adjacent genes before application of diamide, and 2) similarity in chromatin dynamics, which we defined as the correlation between the *change* in +1 nucleosome chromatin marks of adjacent genes following cellular stress. Baseline chromatin state similarity showed no relationship with paired hotspot effects ($P = 0.17$). By contrast, similarity in chromatin dynamics showed a weak but significant association (rho = 0.03, $P = 0.02$).

Third, we compared genes expressed in different orientations (divergent, tandem, and convergent). Analysis of variance (ANOVA) showed a small but significant association between orientation and the number of paired hotspot effects (fraction of variation explained = 0.9%, $P = 2 \times 10^{-10}$). Divergent gene pairs showed the most hotspot effect pairing (mean = 2.7 doublets), followed by pairs oriented in tandem (mean = 2.4) and convergent pairs (mean = 2.1; Supplementary Figure S5; Wilcoxon tests: divergent *vs* tandem $P = 0.014$; divergent *vs* convergent $P = 9 \times 10^{-8}$; tandem *vs* convergent = 2×10^{-4}).

Fourth, we asked how physical distance related to paired hotspot effects. Distance may be a proxy for a number of molecular mechanisms, including promiscuous effects of *cis*-regulatory elements that decay as a function of distance (Quintero-Cadena and Sternberg 2016). Although adjacent genes are all close to each other by definition, the distance between the start codons of the verified ORFs we analyzed here ranges from 46 bp to 19,141 bp, with a median of 1924 bp (Figure 4A). We found that paired hotspot effects tended to occur at adjacent genes that were closer to each other (Rho = 0.05, $P < 2.6 \times 10^{-6}$; proximity was quantified as the log of the inverse of distance in bp).

The features considered above tend to be correlated with each other (Supplementary File S3). For example, divergently expressed genes are close to each other and also share an intergenic region where the same TFBSs may influence their expression. To quantify the contributions of these features relative to each other, we included them as predictor variables in a generalized linear model of the number of paired hotspot effects for adjacent gene pairs. To gauge the significance of a given feature, we compared the full model, which contained all features, to models that did not contain the given feature. Overall, the full model explained a modest (6.4%) but highly significant ($P < 2.2 \times 10^{-16}$) fraction of the variation in paired hotspot effects among adjacent gene pairs. Several features had significant effects (Figure 4B), but no feature was highly predictive of how often a given gene pair is affected by hotspots in the same direction. These modest feature effects are consistent with the magnitude of the correlations reported above, and are expected given the small effects of individual *trans*-eQTLs, which create inescapable noise in our quantification of paired hotspot effects. Nevertheless, these analyses can quantify the importance of features relative to each other while controlling for correlations among features.

TFBS similarity explained the greatest amount of variation, such that adjacent gene pairs with similar sets of TFBSs had more paired hotspot effects (Figure 4B and Supplementary Figure S5B). Chromatin baseline similarity and similarity in chromatin change each had smaller but significant effects (Figure 4B). There was no independent effect of gene pair orientation, presumably because the differences between orientations reported above mostly reflect physical distance (Figure 4A) (Cohen *et al.* 2000). Across all adjacent gene pairs, physical proximity was the second most important feature, such that genes whose start codons are located closer to each other had more paired hotspot effects (Figure 4B, Supplementary File S4). Importantly, although chromatin baseline similarity (rho = –0.24; $P = 2.2 \times 10^{-16}$) and chromatin change similarity (rho = –0.04; $P = 0.006$), but not TFBS similarity (rho = –0.02; $P = 0.12$), are negatively correlated with distance, our model controls for these features, such that they cannot account for the independent effect of distance.

Dividing gene pairs by their orientation showed that the overall distance effect was almost exclusively driven by variation among the divergently oriented gene pairs (Figure 4B). Most divergent gene neighbors are separated by less than 1000 bp (Figure 4A). This is the same range within which coexpression of adjacent genes is most

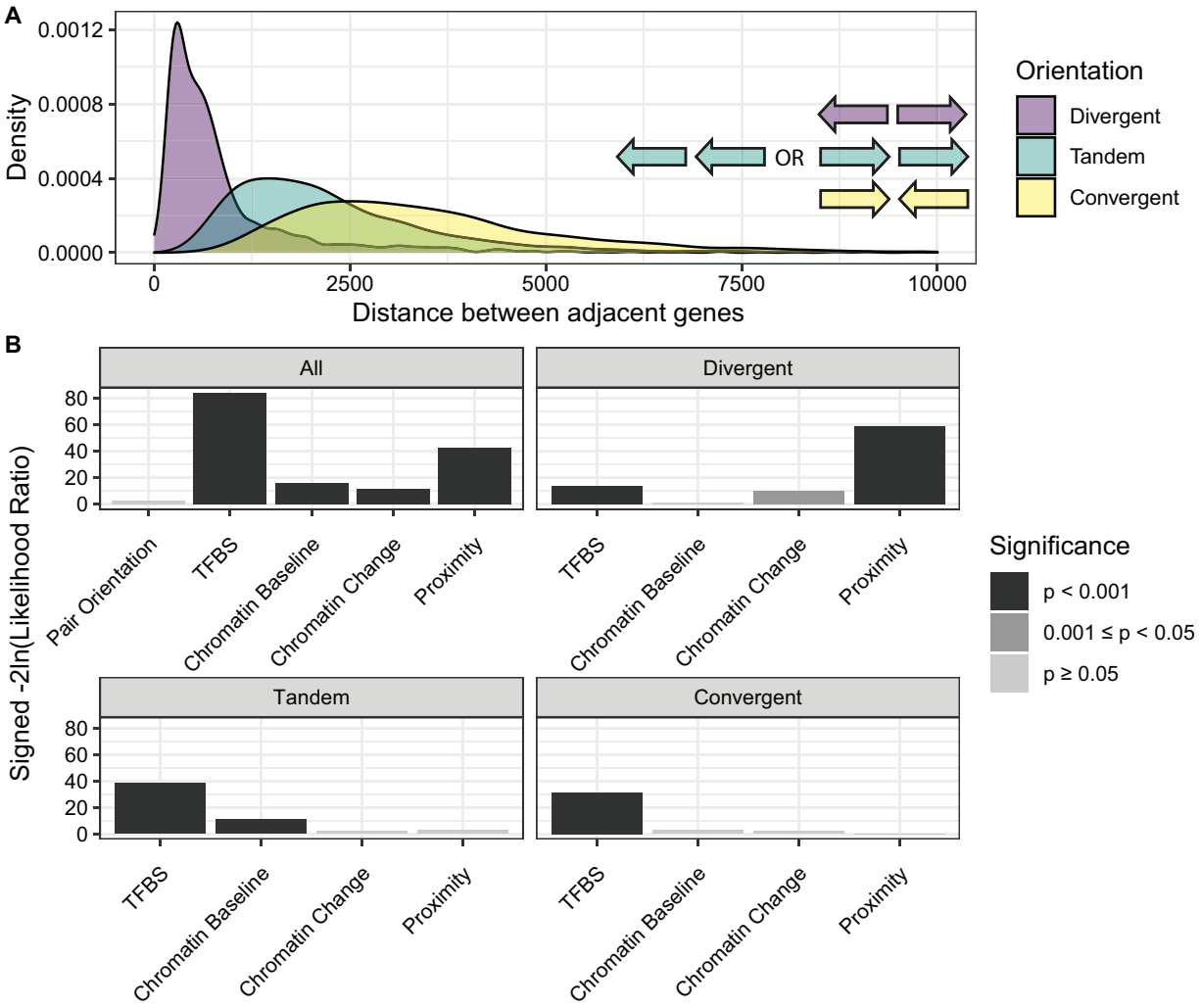

Figure 4 (A) Distribution of intergenic distances for gene pairs in each of the three possible orientations (divergent, tandem, convergent). (B) Influence of various features on paired hotspot effects. Feature importance is shown as the difference in fit between a full model with all features and a model without the given feature. Effect directions were plotted based on the sign of the feature's slope in the linear model; note that all effects happened to have a positive sign. The categorical feature of pair orientation, for which no overall sign exists for its three levels, was plotted with a positive effect.

prominent in yeast, an observation that has been attributed to non-specific, promiscuous interactions between upstream activating sequences and promoters. (Cohen *et al.* 2000; Quintero-Cadena and Sternberg 2016). Thus, proximity-based promiscuity of *cis*-regulatory elements could explain why *trans*-eQTL hotspots tend to affect both genes in close, adjacent gene pairs.

In sum, there are at least three separate reasons for the observation that *trans*-eQTL hotspots tend to affect both members of adjacent gene pairs. First, adjacent genes tend to be regulated by similar transcription factors, which results in paired hotspot effects through vertical mechanisms. Second, for genes in close physical proximity, the effect of a *trans*-eQTL on one gene can also change the expression of the other gene via horizontal mechanisms that scale as a function of distance. Third, similarity in chromatin state, both at baseline and during chromatin dynamics, also increases the probability that adjacent genes tend to be affected horizontally by the same hotspots.

Many adjacent genes with paired hotspot effects do not share a common function

Functionally related genes tend to be clustered along the genomes of humans (Caron *et al.* 2001; Al-Shahrour *et al.* 2010; Andrews *et al.* 2015), yeasts (Cohen *et al.* 2000; Pál and Hurst 2003; Poyatos and Hurst 2007; Eldabagh *et al.* 2018), flies (Spellman and Rubin 2002), mice (Li *et al.* 2005), worms (Kamath *et al.* 2003), and zebrafish (Ng *et al.* 2009). This clustering has been suggested to facilitate coexpression of genes in pathways [but see Kustatscher *et al.* (2017)]. We asked whether paired hotspot effects tend to occur at adjacent genes with related functions.

To quantify functional similarity among gene pairs, we used two metrics. First, we computed semantic similarity based on the "biological process" annotations in the Gene Ontology database (Ashburner *et al.* 2000; The Gene Ontology Consortium 2019). Second, genes with similar biological function tend to have similar synthetic genetic interaction profiles, as previously quantified (Costanzo *et al.* 2016). Although not all genes have high-quality Gene Ontology annotations, the genetic interaction profiles cover the vast majority of gene pairs in a systematic manner. In both metrics, adjacent genes showed modestly but significantly higher functional similarity than nonadjacent genes (semantic similarity: median 0.199 *vs* 0.195, Wilcoxon test $P = 0.002$; genetic interaction similarity: median 0.02 *vs* 0.005, $P < 2.2 \times 10^{-16}$).

We asked how these two measures of functional similarity relate to the molecular features of adjacent genes we compiled

above. Adjacent genes with higher semantic similarity were more likely to be regulated by similar sets of TFs (rho = 0.07, $P = 1 \times 10^{-6}$; Supplementary File S4; see also (Hershberg *et al.* 2005)) and, to a weaker extent, have similar chromatin dynamics (rho = 0.04, $P = 1 \times 10^{-5}$). By contrast, genetic interaction similarity was not influenced by these two features ($P > 0.4$). Instead, adjacent genes with similar genetic interaction patterns tended to be physically close to each other (Rho = 0.06, $P = 0.0003$). In line with these results, semantic similarity and genetic interaction similarity were neither correlated with each other for adjacent genes (rho = 0.006, $P = 0.7$) nor for all genes (rho = 0.009, $P = 0.6$). Thus, these two measures capture different aspects of functional similarity.

We tested the relationship between functional similarity and paired hotspot effects at adjacent genes. Genetic interaction similarity was weakly but significantly correlated with paired hotspot effects (rho = 0.05, $P = 0.003$), whereas semantic similarity was not (rho = 0.02, $P = 0.10$). Because functional similarity and paired hotspot effects are both shaped by similar molecular features, their correlation could be confounded by these shared influences. To control for this, we computed residuals of each metric, removing the influence of the molecular features. Correlations between the residuals of paired hotspot effects and functional similarity remained weak, and became less significant (genetic interaction similarity: rho = 0.04, $P = 0.02$; semantic similarity: rho = 0.02, $P = 0.23$). The weak magnitude of these relationships suggests that many adjacent genes with paired hotspot effects do not share a common function.

An example of paired hotspot effects and their relation to gene function is provided by a hotspot that is caused by a nonsynonymous variant in the Oaf1 transcription factor (Lutz *et al.* 2019), which regulates the transcription of genes involved in fatty acid metabolism (Baumgartner *et al.* 1999). The *OAF1* hotspot affects 39 doublets, a number that was never observed in our permutations, which showed a median (and mean) of 16 doublets and a maximum of 36 doublets. This hotspot alters the expression of genes involved in fatty acid metabolism (Lutz *et al.* 2019), including many annotated transcriptional target genes of Oaf1 (Bergenholm *et al.* 2018). For example, the hotspot alters the expression of the gene encoding the beta unit of fatty acid synthetase, *FAS1* (SGD: S000001665), which is transcriptionally regulated by Oaf1 (Bergenholm *et al.* 2018). However, the promoters of 74% of genes affected by the *OAF1* hotspot were not found to be bound by Oaf1 (Bergenholm *et al.* 2018), suggesting that the hotspot could alter the expression of some of these genes via horizontal mechanisms. For example, the hotspot affects the expression of the *PRS1* (SGD: S000001664) gene immediately downstream of, and in tandem orientation with regards to, *FAS1*. *PRS1* is not a known transcriptional target of Oaf1 (Bergenholm *et al.* 2018, also Yeastract), and encodes an enzyme involved in nucleotide, histidine, and tryptophan biosynthesis. These processes are not obviously involved in fatty acid metabolism. In summary, horizontal propagation of the effects of *trans*-eQTLs holds the potential of creating unexpected links between the genes that cause a *trans*-eQTL and the genes whose expression they affect.

Discussion

In this work, we revealed a previously unrecognized mode by which *trans*-acting genetic variation can affect the expression of multiple genes. Typically, *trans*-eQTL effects are thought to propagate vertically through regulatory pathways that are functionally related to the causal eQTL gene. Here, we showed that the effects of a *trans*-eQTL can propagate horizontally along a given chromosomal region and affect functionally unrelated genes (Figure 5).

We showed that *trans*-eQTL hotspots in a yeast cross affected pairs of neighboring genes more frequently than expected by chance. At significantly enriched hotspots, this excess corresponded to tens of adjacent gene pairs among hundreds of genes that these hotspots affected. The discovery of this subtle effect required the well-powered eQTL dataset we analyzed here. The effect of the hotspots on adjacent genes tended to involve gene pairs that showed stronger coexpression in response to a broad range of perturbations. Thus, a straightforward explanation for the effects of eQTL hotspots on adjacent genes is that the DNA variants that underlie the hotspots engage the same genomic mechanisms that also shape gene–gene coexpression.

We examined molecular features that could drive pairing of hotspot effects at adjacent genes. The strongest determinant was regulation of neighboring genes by similar sets of TFs. This result is consistent with prior observations that genes with similar function, which tend to be regulated by similar TFs as part of transcriptional regulons, tend to be physically colocated in the yeast genome (Hershberg *et al.* 2005). Pairing of hotspot effects due to regulation by similar TFs is expected under conventional, vertical propagation of hotspot effects in a pathway, if pathway members are located next to each other in the genome. However, the second strongest determinant of paired hotspot effects was the proximity of adjacent genes on the chromosome. Crucially, this analysis controlled for shared vertical regulation by similar TFs. Thus, *cis*-acting mechanisms that extend horizontally along a chromosome in a manner that is dependent on physical distance must be responsible for this effect.

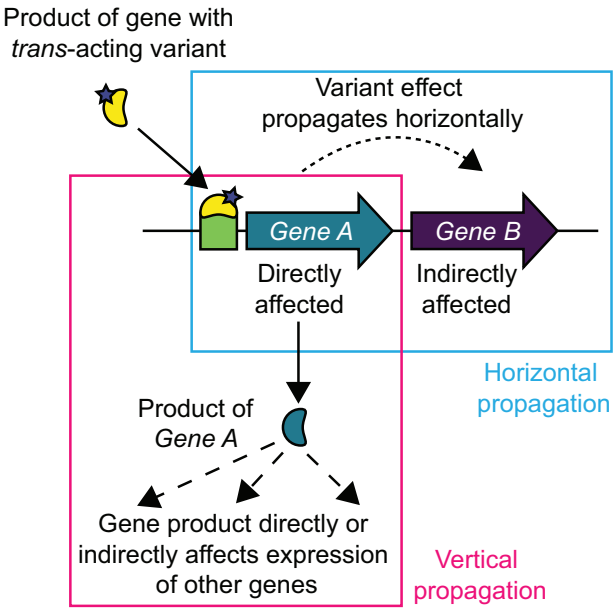

Figure 5 A schematic illustrating two modes by which *trans*-acting sequence variation affects direct and indirect target genes. Left (pink box), the effects of a *trans*-acting variant percolate "vertically" among genes by directly targeting some genes for which altered abundance of the gene product influences other genes in turn. Right (blue box), effects of *trans*-acting variation propagate "horizontally" to genes located proximally to a direct target gene. Although the schematic shows two adjacent genes oriented in tandem, the same processes can affect genes in other orientations.

A possible mechanism involves chromatin states that extend across adjacent genes, as suggested by studies of individual gene pairs (Raj *et al.* 2006; Batada *et al.* 2007; Ebisuya *et al.* 2008). Indeed, adjacent genes that had correlated chromatin marks in normally growing cultures and during response to cellular stress were more likely to be affected by hotspots as a doublet. This relationship suggests that shared chromatin state is one mechanism by which hotspot effects propagate horizontally.

Proximity-dependent promiscuous effects of *cis*-regulatory elements are another possible horizontal mechanism that could result in paired hotspot effects. In eukaryotes, transcriptional regulation is shaped by *cis*-regulatory elements bound by sequence-specific transcription factors (Wittkopp and Kalay 2012). These elements, which include yeast upstream activating elements and enhancers in higher eukaryotes, interact with the core promoters of specific genes (Hahn and Young 2011). Under proximity-dependent promiscuity, which has also been called "enhancer-promoter (EP) theory" (Quintero-Cadena and Sternberg 2016; Cera *et al.* 2019), *cis*-regulatory elements can also regulate transcription from any promoter within a certain physical distance (Butler 2001; Ebisuya *et al.* 2008). This non-specific activity decays as distance between the element and a gene increases. Distance-dependent coexpression has been described in species ranging from yeast to human, and the reach of this phenomenon scales as a function of genome size (Quintero-Cadena and Sternberg 2016). In the compact yeast genome, distance-dependent coexpression is detectable within a range of about 1000 bp. In our data, distance only influenced pairing of hotspot effects among divergently expressed gene pairs. Most of these pairs are located less than 1000 bp apart, well within the reach of non-specific EP processes. Thus, pairing of hotspot effects may be partly driven by proximity-dependent promiscuity of *cis*-regulatory elements.

The molecular features above accounted for a modest portion of the variation in hotspot pairing among adjacent genes, raising the question of which factors account for the remaining variation. Although we cannot rule out that molecular mechanisms other than those we considered could play larger roles, a likely explanation are imprecise estimates of the strength of hotspot pairing at individual gene pairs. Most *trans*-eQTLs explain 1–2% of the variance in the mRNA abundance of a given gene, and many hotspot effects of this magnitude remain to be discovered (Albert *et al.* 2018). Future inclusion of additional *trans*-eQTLs would improve the quantification of paired hotspot effects, and could further clarify the relative importance of the molecular features responsible for the excess of *trans*-eQTL hotspot effects at adjacent genes.

Although the precise mechanistic basis of horizontal propagation of hotspot effects remains unclear, its potential implications are profound. Under this proposed model, a *trans*-eQTL first alters the expression of a distant gene via conventional, vertical regulation. Altered expression of the target gene causes molecular changes in the gene's vicinity on the chromosome, which in turn change the expression of other genes in the same area. This horizontal extension of hotspot effects has the potential to create unexpected links between causal DNA variants and complex traits. If the *trans* effect extends to a neighboring gene with a function that is unrelated to the causal gene in the hotspot, the altered expression of this neighboring gene could alter cell physiology in ways that are not obvious from the function of the causal eQTL gene. Such indirect consequences could contribute further complexity to the genetic basis of quantitative traits (Boyle *et al.* 2017).

Compared to yeast, our knowledge of human *trans*-eQTLs remains limited (Heinig *et al.* 2010; Fehrmann *et al.* 2011; Small *et al.* 2011; Grundberg *et al.* 2012; Battle *et al.* 2014; Lee *et al.* 2014; Wright *et al.* 2014; Aguet *et al.* 2017; Brynedal *et al.* 2017; Yao *et al.* 2017). Nevertheless, the prevalence of EP effects across species including humans (Quintero-Cadena and Sternberg 2016) means that horizontal propagation of *trans*-effects is expected to also exist in other species.

Acknowledgments

We thank Christian Brion and Kristine Trotta for critical feedback, as well as members of the Albert lab (Randi Avery, Mahlon Collins, Kelsey Johnson, and Kaushik Renganaath) for comments on the manuscript.

Funding

This work was supported by the National Institutes of Health, Alfred P. Sloan Foundation, and the Pew Charitable Trusts.

Literature cited

Aguet F, Brown AA, Castel SE, Davis JR, He Y, eQTL manuscript working group, *et al.* 2017. Genetic effects on gene expression across human tissues. Nature. 550:204–213. doi:10.1038/nature24277.

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

Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 16:197–212. doi:10.1038/nrg3891.

Allocco DJ, Kohane IS, Butte AJ. 2004. Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics. 5: 18.

Al-Shahrour F, Minguez P, Marqués-Bonet T, Gazave E, Navarro A, *et al.* 2010. Selection upon genome architecture: conservation of functional neighborhoods with changing genes. PLoS Comput Biol. 6:e1000953.doi:10.1371/journal.pcbi.1000953.

Andrews T, Honti F, Pfundt R, de Leeuw N, Hehir-Kwa J, *et al.* 2015. The clustering of functionally related genes contributes to CNV-mediated disease. Genome Res. 25:802–813. doi:10.1101/gr.184325.114.

Arnone JT, Arace JR, Soorneedi AR, Citino TT, Kamitaki TL, *et al.* 2014. Dissecting the *cis* and *trans* elements that regulate adjacent-gene coregulation in saccharomyces cerevisiae. Eukaryot Cell. 13:738–748. doi:10.1128/EC.00317-13.

Arnone JT, McAlear MA. 2011. Adjacent gene pairing plays a role in the coordinated expression of ribosome biogenesis genes *MPP10* and *YJR003C* in *Saccharomyces cerevisiae*. Eukaryot Cell. 10:43–53. doi:10.1128/EC.00257-10.

Arnone JT, Robbins-Pianka A, Arace JR, Kass-Gergi S, McAlear MA. 2012. The adjacent positioning of co-regulated gene pairs is widely conserved across eukaryotes. BMC Genomics. 13:546.doi:10.1186/1471-2164-13-546.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, *et al.* 2000. Gene ontology: tool for the unification of biology. Nat Genet. 25:25–29. doi:10.1038/75556.

Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, *et al.* 2012. YeastMine—an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database. 2012. doi:10.1093/database/bar062. (Accessed: 2021 January 12). https://aca

demic.oup.com/database/article/doi/10.1093/database/bar062/429965.

Batada NN, Urrutia AO, Hurst LD. 2007. Chromatin remodelling is a major source of coexpression of linked genes in yeast. Trends Genet. 23:480–484. doi:10.1016/j.tig.2007.08.003.

Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, *et al.* 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. 24:14–24. doi:10.1101/gr.155192.113.

Baumgartner U, Hamilton B, Piskacek M, Ruis H, Rottensteiner H. 1999. Functional analysis of the Zn(2)Cys(6) transcription factors Oaf1p and Pip2p. Different roles in fatty acid induction of beta-oxidation in *Saccharomyces cerevisiae*. J Biol Chem. 274:22208–22216. doi:10.1074/jbc.274.32.22208.

Bergenholm D, Liu G, Holland P, Nielsen J. 2018. Reconstruction of a global transcriptional regulatory network for control of lipid metabolism in yeast by using chromatin immunoprecipitation with lambda exonuclease digestion. mSystems. 3:e00215–17. doi:10.1128/mSystems.00215-17.

Bolstad B. 2017. preprocessCore. Bioconductor. (Accessed: 2021 January 12). https://bioconductor.org/packages/preprocessCore.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell. 169:1177–1186. doi:10.1016/j.cell.2017.05.038.

Brem RB. 2002. Genetic dissection of transcriptional regulation in budding yeast. Science. 296:752–755. doi:10.1126/science.1069516.

Brem RB, Kruglyak L. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci USA. 102:1572–1577. doi:10.1073/pnas.0408709102.

Brown CR, Mao C, Falkovskaia E, Jurica MS, Boeger H. 2013. Linking stochastic fluctuations in chromatin structure and gene expression. PLoS Biol. 11:e1001621.doi:10.1371/journal.pbio.1001621.

Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, *et al.* 2017. Large-scale trans -eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. Am J Hum Genet. 100:581–591. doi:10.1016/j.ajhg.2017.02.004.

Butler JEF. 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev. 15:2515–2519. doi:10.1101/gad.924301.

Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, *et al.* 2005. Uncovering regulatory pathways that affect hematopoietic stem cell function using "genetical genomics". Nat Genet. 37:225–232. doi:10.1038/ng1497.

Carlson M. 2017. org.Sc.sgd.db. Bioconductor. (Accessed: 2021 January 12). https://bioconductor.org/packages/org.Sc.sgd.db.

Caron H, Schaik B, van Mee M, van der Baas F, Riggins G, *et al.* 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science. 291:1289–1292. doi:10.1126/science.1056794.

Cera A, Holganza MK, Hardan AA, Gamarra I, Eldabagh RS, *et al.* 2019. functionally related genes cluster into genomic regions that coordinate transcription at a distance in *Saccharomyces cerevisiae*. mSphere. 4:e00063–19. doi:10.1128/mSphere.00063-19.

Cesar ASM, Regitano LCA, Reecy JM, Poleti MD, Oliveira PSN, *et al.* 2018. Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. BMC Genomics. 19:499. doi:10.1186/s12864-018-4871-y.

Chereji RV, Ramachandran S, Bryson TD, Henikoff S. 2018. Precise genome-wide mapping of single nucleosomes and linkers in vivo. Genome Biol. 19:19.doi:10.1186/s13059-018-1398-0.

Chesler EJ, Lu L, Shou S, Qu Y, Gu J, *et al.* 2005. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet. 37:233–242. doi:10.1038/ng1518.

Clément-Ziza M, Marsellach FX, Codlin S, Papadakis MA, Reinhardt S, *et al.* 2014. Natural genetic variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. Mol Syst Biol. 10:764. doi:10.15252/msb.20145123.

Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet. 26:183–186. doi:10.1038/79896.

Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, *et al.* 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science. 353:aaf1420–aaf1420. doi:10.1126/science.aaf1420.

Ebisuya M, Yamamoto T, Nakajima M, Nishida E. 2008. Ripples from neighbouring transcription. Nat Cell Biol. 10:1106–1113. doi:10.1038/ncb1771.

Eddelbuettel D, François R. 2011. Rcpp: Seamless *R* and *C*. ++ Integration. J Stat Softw. 40. (Accessed: 2021 January 12). http://www.jstatsoft.org/v40/i08/. doi:10.18637/jss.v040.i08.

Eldabagh RS, Mejia NG, Barrett RL, Monzo CR, So MK, Foley JJ, *et al.* 2018. Systematic identification, characterization, and conservation of adjacent-gene coregulation in the budding yeast Saccharomyces cerevisiae. mSphere. 3(3):e00220–18, /msphere/3/3/mSphere220-18.atom. doi:10.1128/mSphere.00220-18.

Everett LJ, Huang W, Zhou S, Carbone MA, Lyman RF, *et al.* 2020. Gene expression networks in the *Drosophila* genetic reference panel. Genome Res. 30:485–496. doi:10.1101/gr.257592.119.

Fehrmann RSN, Jansen RC, Veldink JH, Westra H-J, Arends D, *et al.* 2011. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 7:e1002197. doi:10.1371/journal.pgen.1002197.

Fisk DG, Ball CA, Dolinski K, Engel SR, Hong EL, TheSaccharomyces Genome Database Project, *et al.* 2006. Saccharomyces cerevisiae S288C genome annotation: a working hypothesis. Yeast. 23:857–865. doi:10.1002/yea.1400.

Fleming JA, Lightcap ES, Sadis S, Thoroddsen V, Bulawa CE, *et al.* 2002. Complementary whole-genome technologies reveal the cellular response to proteasome inhibition by PS-341. Proc Natl Acad Sci. 99:1461–1466. doi:10.1073/pnas.032516399.

Fox J, Weisberg S. 2019. An R Companion to Applied Regression. 3rd ed. Los Angeles: Sage.

Fu J, Keurentjes JJB, Bouwmeester H, America T, Verstappen FWA, *et al.* 2009. System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. Nat Genet. 41:166–167. doi:10.1038/ng.308.

Gierliński M, Cole C, Schofield P, Schurch NJ, Sherstnev A, *et al.* 2015. Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. Bioinforma Oxf Engl. 31:3625–3630. doi:10.1093/bioinformatics./btv425.

Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, The Multiple Tissue Human Expression Resource (MuTHER) Consortium, *et al.* 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 44:1084–1089. doi:10.1038/ng.2394.

Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, *et al.* 2014. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet. 95:535–552. doi:10.1016/j.ajhg.2014.10.004.

Hahn S, Young ET. 2011. Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics. 189:705–736. doi:10.1534/genetics.111.127019.

Harrell FEJ. 2020. Hmisc: Harrell Miscellaneous. https://CRAN.R-project.org/package=Hmisc.

Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Cardiogenics Consortium, *et al.* 2010. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature. 467:460–464. doi:10.1038/nature09386.

Hershberg R, Yegerlotem E, Margalit H. 2005. Chromosomal organization is shaped by the transcription regulatory network. Trends Genet. 21:138–142. doi:10.1016/j.tig.2005.01.003.

Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, *et al.* 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nat Genet. 37:243–253. doi:10.1038/ng1522.

Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, *et al.* 2000. Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet. 25:333–337. doi:10.1038/77116.

Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 5:299–310. doi:10.1038/nrg1319.

Huttenhower C, Hibbs M, Myers C, Troyanskaya OG. 2006. A scalable method for integration and functional analysis of multiple microarray datasets. Bioinformatics. 22:2890–2897. doi:10.1093/bioinformatics/btl492.

Hunter JD. 2007. Matplotlib: a 2D graphics environment. Comput Sci Eng. 9:90–95. doi:10.1109/MCSE.2007.55.

Jansen A, Verstrepen KJ. 2011. Nucleosome positioning in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev. 75:301–320. doi:10.1128/MMBR.00046-10.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, *et al.* 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature. 421:231–237. doi:10.1038/nature01278.

Klevecz RR, Bolen J, Forrest G, Murray DB. 2004. A genomewide oscillation in transcription gates DNA replication and cell cycle. Proc Natl Acad Sci USA. 101:1200–1205. doi:10.1073/pnas.0306490101.

Knijnenburg TA, Daran J-MG, van den Broek MA, Daran-Lapujade PA, de Winde JH, *et al.* 2009. Combinatorial effects of environmental parameters on transcriptional regulation in *Saccharomyces cerevisiae*: a quantitative analysis of a compendium of chemostat-based transcriptome data. BMC Genomics. 10:53.doi:10.1186/1471-2164-10-53.

Kraakman LS, Mager WH, Maurer KT, Nieuwint RT, Planta RJ. 1989. The divergently transcribed genes encoding yeast ribosomal proteins L46 and S24 are activated by shared RPG-boxes. Nucl Acids Res. 17:9693–9706. doi:10.1093/nar/17.23.9693.

Kustatscher G, Grabowski P, Rappsilber J. 2017. Pervasive coexpression of spatially proximal genes is buffered at the protein level. Mol Syst Biol. 13:937.doi:10.15252/msb.20177548.

Langley SR, Bottolo L, Kunes J, Zicha J, Zidek V, *et al.* 2013. Systems-level approaches reveal conservation of trans-regulated genes in the rat and genetic determinants of blood pressure in humans. Cardiovasc Res. 97:653–665. doi:10.1093/cvr/cvs329.

Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 25:1841–1842. doi:10.1093/bioinformatics/btp328.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, *et al.* 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol. 9:e1003118.doi:10.1371/journal.pcbi.1003118.

Lee MN, Ye C, Villani A-C, Raj T, Li W, *et al.* 2014. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science. 343:1246980–1246980. doi:10.1126/science.1246980.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers Nach Margaritis T, *et al.* 2011. The specificity and topology of chromatin interaction pathways in yeast. Mol Cell. 42:536–549. doi:10.1016/j.molcel.2011.03.026.

Lewis JA, Broman AT, Will J, Gasch AP. 2014. Genetic architecture of ethanol-responsive transcriptome variation in *Saccharomyces cerevisiae* strains. Genetics. 198:369–382. doi:10.1534/genetics.114.167429.

Li Q, Lee BT, Zhang L. 2005. Genome-scale analysis of positional clustering of mouse testis-specific genes. BMC Genomics. 6:7.doi:10.1186/1471-2164-6-7.

Lutz S, Brion C, Kliebhan M, Albert FW. 2019. DNA variants affecting the expression of numerous genes in trans have diverse mechanisms of action and evolutionary histories. PLOS Genet. 15:e1008375.doi:10.1371/journal.pgen.1008375.

Lynch M, Walsh B. 1998. Genetics and Analysis of Quantitative Traits. Sunderland, MA: Sinauer.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, *et al.* 2012. Systematic localization of common disease-associated variation in regulatory DNA. Science. 337:1190–1195. doi:10.1126/science.1222794.

Michalak P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics. 91:243–248. doi:10.1016/j.ygeno.2007.11.002.

Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, *et al.* 2020. YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic Acids Res. 48:D642–D649. doi:10.1093/nar/gkz859.

Myers KS, Riley NM, MacGilvray ME, Sato TK, McGee M, *et al.* 2019. Rewired cellular signaling coordinates sugar and hypoxic responses for anaerobic xylose fermentation in yeast. PLOS Genet. 15:e1008037.doi:10.1371/journal.pgen.1008037.

Ng YK, Wu W, Zhang L. 2009. Positive correlation between gene coexpression and positional clustering in the zebrafish genome. BMC Genomics. 10:42.doi:10.1186/1471-2164-10-42.

Orozco LD, Bennett BJ, Farber CR, Ghazalpour A, Pan C, *et al.* 2012. Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. Cell. 151:658–670. doi:10.1016/j.cell.2012.08.043.

Pagès Pa H. 2017. Biostrings. Bioconductor. (Accessed: 2021 January 12). https://bioconductor.org/packages/Biostrings.

Pál C, Hurst LD. 2003. Evidence for co-evolution of gene order and recombination rate. Nat Genet. 33:392–395. doi:10.1038/ng1111.

Poyatos JF, Hurst LD. 2007. The determinants of gene order conservation in yeasts. Genome Biol. 8:R233.doi:10.1186/gb-2007-8-11-r233.

Quintero-Cadena P, Sternberg PW. 2016. Enhancer sharing promotes neighborhoods of transcriptional regulation across eukaryotes. G3 (Bethesda). 6:4167.doi:10.1534/g3.116.036228.

R Development Core Team 2010. R: A Language and Environment for Statistical Computing: reference Index. Vienna: R Foundation for Statistical Computing. (Accessed: 2020 August 28). http://www.polsci.wvu.edu/duval/PS603/Notes/R/fullrefman.pdf.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA Synthesis in mammalian cells. PLoS Biol. 4:e309.doi:10.1371/journal.pbio.0040309.

Rando OJ, Winston F. 2012. Chromatin and transcription in yeast. Genetics. 190:351–387. doi:10.1534/genetics.111.132266.

Rockman MV, Skrovanek SS, Kruglyak L. 2010. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. Science. 330:372–376. doi:10.1126/science.1194208.

Sameith K, Amini S, Groot Koerkamp MJA, van Leenen D, Brok M, *et al.* 2015. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential

mechanisms for genetic interactions. BMC Biol. 13:112. doi:10.1186/s12915-015-0222-5.

Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, *et al.* 2015. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. 25:1757–1770. doi:10.1101/gr.192294.115.

Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, *et al.* 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA. 22:839–851. doi:10.1261/rna.053959.115.

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

Simola DF, Francis C, Sniegowski PD, Kim J. 2010. Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes. Genome Biol. 11:R105.doi:10.1186/gb-2010-11-10-r105.

Small KS, Hedman ÅK, Grundberg E, Nica AC, Thorleifsson G, MuTHER Consortium, *et al.* 2011. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet. 43:561–564. doi:10.1038/ng.833.

Smith EN, Kruglyak L. 2008. Gene–environment interaction in yeast gene expression. PLoS Biol. 6:e83.doi:10.1371/journal.pbio.0060083.

Spellman PT, Rubin GM. 2002. Evidence for large domains of similarly expressed genes in the Drosophila genome. J Biol. 5.1:

Sun M, Zhang J. 2019. Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. PLOS Genet. 15: e1008389.doi:10.1371/journal.pgen.1008389.

Sun W, Hu Y. 2013. eQTL mapping using RNA-seq data. Stat Biosci. 5: 198–219. doi:10.1007/s12561-012-9068-3.

The Gene Ontology Consortium 2019. The gene ontology resource: 20 years and still going strong. Nucleic Acids Res. 47:D330–D338. doi:10.1093/nar/gky1055.

Veerla S, Höglund M. 2006. Analysis of promoter regions of co-expressed genes identified by microarray analysis. BMC Bioinformatics. 7:384.doi:10.1186/1471-2105-7-384.

Venables WN, Ripley BD, Venables WN. 2002. Modern Applied Statistics with S. New York: Springer,. 4th ed: (Statistics and computing).

Wade CH, Umbarger MA, McAlear MA. 2006. The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes. Yeast. 23:293–306. doi:10.1002/yea.1353.

Wang JZ, Du Z, Payattakool R, Yu PS, Chen C-F. 2007. A new method to measure the semantic similarity of GO terms. Bioinformatics. 23:1274–1281. doi:10.1093/bioinformatics/btm087.

Weiner A, Hsieh T-HS, Appleboim A, Chen HV, Rahat A, *et al.* 2015. High-resolution chromatin dynamics during a yeast stress response. Mol Cell. 58:371–386. doi:10.1016/j.molcel.2015.02.002.

West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, *et al.* 2007. Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in arabidopsis. Genetics. 175:1441–1450. doi:10.1534/genetics.106.064972.

Wickham H, Averick M, Bryan J, Chang W, McGowan L, *et al.* 2019. Welcome to the tidyverse. JOSS. 4:1686. doi:10.21105/joss. 01686.

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

Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, *et al.* 2014. Heritability and genomics of gene expression in peripheral blood. Nat Genet. 46:430–437. doi:10.1038/ng.2951.

Yao C, Joehanes R, Johnson AD, Huan T, Liu C, *et al.* 2017. Dynamic role of trans regulation of gene expression in relation to complex traits. Am J Hum Genet. 100:571–580. doi:10.1016/j.ajhg.2017.02.003.

Yu G, Li F, Qin Y, Bo X, Wu Y, *et al.* 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 26:976–978. doi:10.1093/bioinformatics/btq064.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, *et al.* 2003. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet. 35:57–64. doi:10.1038/ng1222.

Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. 2011. Regulatory variation within and between species. Annu Rev Genom Hum Genet. 12:327–346. doi:10.1146/annurev-genom-082908-150139.

Zhu J, Zhang B, Smith EN, Drees B, Brem RB, *et al.* 2008. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat Genet. 40:854–861. doi:10.1038/ng.167.

Communicating editor: P. Wittkopp