1

1 ***Trans-acting genetic variation affects the expression of adjacent genes***

2 Krisna Van Dyke, Gemechu Mekonnen, Chad L. Myers, Frank W. Albert

3

4 Department of Genetics, Cell Biology, and Development

5 University of Minnesota

6 Minneapolis, MN, U.S.A.

7 Correspondence to: F.W.A. (falbert@umn.edu)

8 Abstract

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

23 Introduction

24 Genetic variation among individuals influences phenotypic traits (Lynch & Walsh, 1998). Many
 25 of the DNA variants that shape phenotypes do so by altering gene expression (Albert &
 26 Kruglyak, 2015; Gusev et al., 2014; Maurano et al., 2012; Signor & Nuzhdin, 2018; Zheng et al.,
 27 2011). Genomic regions that contain variants that influence the abundance of the RNA of a given
 28 gene are called expression quantitative trait loci (eQTLs) (Albert & Kruglyak, 2015; Brem,
 29 2002; W. Sun & Hu, 2013). eQTLs can be classified according to their location in the genome
 30 relative to the genes they affect, as well as by their mechanism of action. Local eQTLs alter the
 31 expression of genes in their physical vicinity, usually by perturbing *cis*-acting regulatory
 32 mechanisms. By contrast, distant eQTLs act in *trans* by changing the abundance, cellular
 33 localization, or activity of a diffusible factor. These factors can act on genes throughout the
 34 genome, and most *trans*-eQTLs influence genes that are distant from the eQTL, typically on
 35 different chromosomes.

36 *Trans*-acting variation has been extensively studied in crosses in model organisms (Brem,
 37 2002; Cesar et al., 2018; Everett et al., 2020; Hubner et al., 2005; Lewis et al., 2014; Orozco et
 38 al., 2012; Rockman et al., 2010; Smith & Kruglyak, 2008; West et al., 2007). For example,
 39 genetic mapping of mRNA levels in 1,012 recombinant offspring from a cross between a
 40 laboratory strain (BY, a close relative of the genome reference strain S288C) and a vineyard
 41 isolate (RM) of the yeast *Saccharomyces cerevisiae* yielded tens of thousands of eQTLs. These
 42 eQTLs accounted for the majority of genetic variation in mRNA abundance (a median across
 43 genes of 72%), suggesting that the dataset comprises the great majority of eQTLs with at least
 44 modest effect sizes. Almost all genes were influenced by multiple *trans*-eQTLs. The summed
 45 *trans*-eQTL effects accounted for 2.6-fold more expression variation than those of local eQTLs
 46 (Albert et al., 2018). In agreement with similar estimates in human populations (Grundberg et al.,
 47 2012; Wright et al., 2014), the BY / RM cross shows that *trans*-eQTLs are the major source of
 48 genetic variation in gene expression.

49 *Trans*-eQTLs tend to cluster at certain locations in the genome. These *trans*-eQTL
 50 “hotspots” can affect the expression of large numbers of genes, and have been observed in yeasts

51 (Brem, 2002; Clément-Ziza et al., 2014; Smith & Kruglyak, 2008), nematodes (Rockman et al.,
52 2010), plants (Fu et al., 2009; West et al., 2007), rodents (Bystrykh et al., 2005; Chesler et al.,
53 2005; Langley et al., 2013; Orozco et al., 2012), and cattle (Cesar et al., 2018), and may also
54 exist in humans (Brynedal et al., 2017). In the BY / RM yeast cross, over 90% of the
55 *trans*-eQTLs mapped to 102 hotspot regions (Albert et al., 2018). Several of these hotspots have
56 been resolved to single causal genes or variants (Brem, 2002; Lutz et al., 2019; Smith &
57 Kruglyak, 2008; Yvert et al., 2003; Zhu et al., 2008). The BY / RM hotspots can affect hundreds
58 and in some cases thousands of genes.

59 Identifying the mechanistic connection between an eQTL hotspot and the genes it affects
60 in *trans* can be challenging. Typically, the causal variant or variants in the hotspot are assumed to
61 alter the sequence or abundance of a protein product, such as a transcription factor, that then acts
62 on functionally related genes. In turn, the changes in gene expression of these genes are
63 propagated through cellular networks to indirectly affect many other genes. In this paper, we
64 refer to this type of mechanistic connection as “vertical propagation”.

65 The expression of eukaryotic genes is partially shaped by their location on the
66 chromosome. Specifically, adjacent genes tend to show correlated expression changes in
67 response to a broad range of stimuli (Cohen et al., 2000; Hurst et al., 2004; Kustatscher et al.,
68 2017; Michalak, 2008; M. Sun & Zhang, 2019). In part, this coexpression is due to clustering of
69 genes with similar function and regulation by transcription factors (TFs) (Allocco et al., 2004;
70 Arnone et al., 2012; Cohen et al., 2000). In addition, shared local chromatin states that
71 encompass adjacent genes may facilitate coexpression even for genes that do not share a
72 common function and that are not regulated by the same TFs (Arnone et al., 2014; Batada et al.,
73 2007; Ebisuya et al., 2008; M. Sun & Zhang, 2019). Coexpression of neighboring genes
74 decreases with physical distance (Cohen et al., 2000; Quintero-Cadena & Sternberg, 2016). In
75 addition to chromatin state, non-specific promiscuous effects of *cis*-regulatory elements have
76 been proposed to contribute to this distance-dependent coexpression (Cohen et al., 2000;
77 Quintero-Cadena & Sternberg, 2016).

78 Here, we asked if the effects of *trans*-eQTL hotspots propagate “horizontally” along a

79 chromosome. We made use of the extensive set of *trans*-eQTLs identified in the BY / RM cross
80 (Albert et al., 2018) to show that *trans*-eQTL hotspots affect the expression of adjacent genes
81 more frequently than expected by chance. This effect occurs primarily at adjacent genes that
82 show correlated gene expression responses to a broad variety of perturbations. Paired hotspot
83 effects can be partially explained by regulation of adjacent genes by similar transcription factors.
84 However, paired hotspot effects were independently associated with the physical proximity of
85 the genes in a pair as well as similar chromatin states, suggesting that the effects of *trans*-eQTLs
86 can extend from an affected gene to adjacent genes via “horizontal” mechanisms that spread
87 along the chromosome. This work demonstrates the existence of an additional mode by which
88 *trans*-acting regulatory variation can affect gene regulatory networks and cell biology.

89 Results

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

91 We analyzed eQTLs mapped in the BY and RM cross (Albert et al., 2018). Briefly, the dataset
92 comprises 36,498 eQTLs, of which 33,529 act in *trans*. These *trans*-eQTLs cluster at 102 hotspot
93 locations, which affect 12 to 4,093 of 4,912 verified open reading frames (ORFs) with detectable
94 expression. Genes are affected by a median of eleven hotspots, and almost all genes (98%) are
95 affected by at least one hotspot.

96 We asked whether hotspots affected genes located next to each other on the chromosome
97 more often than expected by chance. To address this question, we focused on “doublets”, which
98 we defined as pairs of adjacent verified ORFs whose mRNA abundances were both increased or
99 both decreased by a given hotspot. Genes close to chromosome ends were excluded from the
100 analysis (Methods). For each hotspot, we constructed 10,000 permutations by randomly
101 assigning the observed effects of the hotspot to genes, irrespective of their location in the
102 genome. This permutation strategy preserved the number and magnitude of hotspot effects, but
103 broke any relationship between hotspot effects and a gene’s location relative to its neighbors. We
104 compared the number of observed doublets to the distribution of doublets in the hotspot-specific
105 permutations.

106 Out of the 102 hotspots, 91 displayed more doublets than their matched permutation
107 median. This number was significantly higher than expected by chance (Binomial Test, $p <$
108 2.2×10^{-16}). At 58 hotspots, the observed number of doublets exceeded that seen in 95% of the
109 permutations (Fig 1, Suppl Table 1, Suppl Fig 1), which was more than the five hotspots
110 expected to reach this threshold by chance (Binomial Test, $p < 2.2 \times 10^{-16}$). After Bonferroni
111 correction, 17 hotspots exhibited a significant excess of doublets (Table 1, Fig 1, Suppl Table 1,
112 Suppl Fig 1). Coexpression of adjacent genes might occur due to shared promoter sequences
113 when the genes are transcribed in divergent orientations. For example, in the yeast genome,
114 genes in the ribosome and rRNA biosynthesis regulons tend to occur in adjacent, divergently
115 expressed pairs that show highly correlated expression (Arnone et al., 2014; Arnone & McAlear,

116 2011; Eldabagh et al., 2018; Kraakman et al., 1989; Wade et al., 2006). To test if the excess of
 117 paired hotspot effects was only due to gene pairs with divergent orientation, we excluded all
 118 divergent gene pairs from our permutation analysis. The number of doublets remained
 119 significantly higher than expected by chance (Table 1, Suppl Table 1).

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

129 **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 n	p-value** for hotspots that exceed median permutation n
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}
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

130 * The multiple-test corrected p-value was obtained by dividing $p = 0.05$ by the number of
131 hotspots (102).

132 ** P-values were computed using a binomial test of the alternative hypothesis of a greater
133 number of hotspots than expected by chance. The number of hotspots expected by chance are
134 0.05 hotspots with multiple testing correction (i.e., $0.05 / 102$), 5.1 hotspots at $p < 0.05$, and 51
135 hotspots at $p < 0.5$.

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 Suppl Fig 1 for non-normalized results. Hotspots are ordered along the x-axis based on the value of the normalized doublet count in the real data.

146 **Gene coexpression patterns are shared across diverse transcriptome datasets**

147 *Trans*-eQTLs could affect the expression of adjacent genes via mechanisms that lead to gene
 148 coexpression (Brown et al., 2013; Michalak, 2008; M. Sun & Zhang, 2019). To quantitatively
 149 test this hypothesis, we first examined coexpression patterns in eleven transcriptome datasets
 150 (Fig 2) (Albert et al., 2018; Brem & Kruglyak, 2005; Fleming et al., 2002; Hughes et al., 2000;
 151 Klevecz et al., 2004; Knijnenburg et al., 2009; Lenstra et al., 2011; Myers et al., 2019; Sameith et
 152 al., 2015; Schurch et al., 2016; Simola et al., 2010). Each dataset comprised multiple
 153 genome-wide transcriptome experiments ($n = 32 - 1,012$, median = 162) collected in yeast
 154 strains exposed to a wide range of perturbations, including engineered gene deletions and natural
 155 genetic variation, environmental stressors such as chemical treatments, and deprivation of
 156 specific nutrients. These datasets were independently collected by different laboratories using
 157 various microarray or RNA-seq techniques. We also included the expression data from which the
 158 hotspots studied here were identified (Albert et al., 2018). To avoid possible circularity in these
 159 data, in which the perturbations caused by the hotspots could induce correlations among adjacent
 160 genes, we used linear models to remove the effects of eQTLs on segregant expression data
 161 (Methods).

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

171 Beyond these general patterns, we asked if the coexpression values for individual pairs of
 172 genes are quantitatively similar between the eleven different datasets. To do so, we computed the
 173 correlation between the coexpression relationships observed in each pair of datasets for either

174 adjacent genes (Fig 2C, lower triangle) or non-adjacent genes (Fig 2C upper triangle). Both
 175 exhibited significant non-zero correlation in all comparisons (adjacent: $p < 9 \times 10^{-5}$; non-adjacent:
 176 $p < 2.2 \times 10^{-16}$). The correlation of coexpression of adjacent gene pairs between datasets was
 177 stronger than that of non-adjacent gene pairs (Median Rho = 0.24 vs 0.18, paired Wilcoxon test p
 178 $< 2.2 \times 10^{-16}$). These results show that coexpression relationships are broadly consistent across
 179 diverse environmental contexts and perturbations, especially for adjacent genes.

Figure 2: Coexpression analysis. (A) Schematic illustrating the separation of gene-gene coexpression matrices by adjacent versus non-adjacent genes. (B) Comparison of coexpression for adjacent genes versus non-adjacent 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.

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

191 We tested if paired effects of *trans*-eQTL hotspots tend to occur at coexpressed gene pairs. To do
 192 so, we summarized hotspot effects on gene pairs across all hotspots into a score (Methods).
 193 Positive scores indicated that both genes in a pair are affected in the same direction (either both
 194 genes have increased or decreased expression) by multiple hotspots. Negative scores indicated
 195 pairs in which both genes were affected in opposite directions by multiple hotspots. A score near
 196 zero indicated pairs that were targeted by multiple hotspots but without a preference for the same
 197 or opposite direction, or gene pairs that were not affected by many hotspots. Adjacent gene pairs
 198 had higher scores (Wilcoxon test, $p \leq 1.1 \times 10^{-16}$) and a higher proportion of positive scores (Test
 199 of proportions, $p \leq 8 \times 10^{-7}$) than non-adjacent gene pairs, showing that this metric captured
 200 pairing of hotspot effects.

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

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 non-adjacent 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 ($p \leq 1.3 \times 10^{-16}$).

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

We asked which mechanisms might drive paired hotspot effects on adjacent genes. Specifically, we explored molecular features previously identified to affect adjacent gene coexpression, including regulation by similar sets of transcription factors, 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 a pair in the same direction. These counts ranged from zero to 41 (median = 2), with higher values for adjacent compared to non-adjacent genes ($p = 1.1 \times 10^{-14}$).

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 shared regulation by transcription factors, we gathered annotated transcription factor binding sites (TFBS; (Monteiro et al., 2020)) in the intergenic region upstream of each gene and calculated the similarity of these TFBS annotations among all gene pairs. Adjacent genes had higher similarity in their TFBS profiles than non-adjacent genes (Median 0.294 vs 0.275, Wilcoxon test $p < 2.2 \times 10^{-16}$), as expected (Hershberg et al., 2005). TFBS similarity was weakly but significantly correlated with hotspot effect pairing for both adjacent genes ($Rho = 0.033$, $p = 0.02$) and non-adjacent genes ($Rho = 0.027$, $p < 2.2 \times 10^{-16}$). Thus, some paired hotspot effects are due to regulation of adjacent genes by shared 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 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 (Arnone et al., 2014; Brown et al., 2013; Raj et al., 2006). We first considered two independent measures of nucleosome occupancy across the bodies of adjacent genes (Chereji et al., 2018; Schep et al., 2015), 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 & Verstrepen, 2011; Rando & 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 dynamic chromatin change, 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 significant association between orientation and the number of paired hotspot effects ($p = 2 \times 10^{-10}$). Divergent gene pairs showed the most hotspot effect pairing, followed by pairs oriented in tandem and convergent pairs (Suppl Fig 3; Wilcoxon tests: divergent vs. tandem $p = 0.014$; divergent vs. convergent $p = 9 \times 10^{-8}$; tandem vs. convergent $p = 2 \times 10^{-4}$).

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 & Sternberg, 2016). While 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 1,924 bp (Fig 4A). We found that paired hotspot effects tended to occur at adjacent genes that were closer to each other ($\rho = 0.05$, $p < 2.6 \times 10^{-6}$; closeness was quantified as the log of the inverse of distance in bp).

The features considered above tend to be correlated with each other. 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 gene pairs. Several features had significant effects (Fig 4), but no feature was highly predictive of whether a given gene pair is often affected by hotspots. These modest feature effects are consistent with the magnitude of the correlations reported above, and are expected given the small effects of individual

287 *trans*-eQTLs, which create inescapable noise in our quantification of paired hotspot effects.
 288 Nevertheless, these analyses can quantify the importance of features relative to each other, while
 289 controlling for correlations among features.

290 TFBS similarity explained the greatest amount of variation, such that gene pairs with
 291 shared TFBSs had more paired hotspot effects (Fig 4B, Suppl Fig 3B). Chromatin baseline
 292 similarity and similarity in chromatin change each had smaller but significant effects (Fig 4B).
 293 There was no independent effect of gene pair orientation, presumably because the differences
 294 between orientations we noted above mostly reflect physical distance (Fig 4A) (Cohen et al.,
 295 2000). Across all adjacent gene pairs, physical proximity was the second most important feature,
 296 such that genes whose start sites are located closer to each other had more paired hotspot effects
 297 (Fig 4B, Suppl Table 3). Importantly, while chromatin baseline similarity ($Rho = -0.24$; $p =$
 298 2.2×10^{-16}) and chromatin change similarity ($Rho = -0.04$; $p = 0.006$), but not TFBS similarity
 299 ($Rho = -0.02$; $p = 0.12$), are negatively correlated with distance, our model controls for these
 300 features, such that they cannot account for the independent effect of distance.

301 Dividing gene pairs by their orientation showed that the overall distance effect was
 302 almost exclusively driven by variation among the divergently oriented gene pairs (Fig 4B). Most
 303 divergent gene neighbors are separated by less than 1,000 bp (Fig 4A). This is the same range
 304 within which coexpression of adjacent genes is most prominent in yeast, an observation that has
 305 been attributed to non-specific, promiscuous interactions between upstream activating sequences
 306 and promoters. (Cohen et al., 2000; Quintero-Cadena & Sternberg, 2016). Thus, proximity-based
 307 promiscuity of *cis*-regulatory elements could explain why *trans*-eQTL hotspots tend to affect
 308 both genes in close, adjacent gene pairs.

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

that adjacent genes tend to be affected by the same hotspots.

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. Effect directions were plotted based on the sign of the feature's slope in the linear model; all effects had a positive sign. Pair orientation is plotted with a positive effect.

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 (Al-Shahrour et al., 2010; Andrews et al., 2015; Caron et al., 2001), yeasts (Cohen et al., 2000; Eldabagh et al., 2018; Pál & Hurst, 2003; Poyatos & Hurst, 2007), flies (Spellman & 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 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). While 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 non-adjacent 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. Genes with higher semantic similarity were more likely to be regulated by similar TFs ($p = 6 \times 10^{-10}$; Suppl Table 2; see also (Hershberg et al., 2005)) and, to a weaker extent, have similar chromatin dynamics ($p = 1 \times 10^{-5}$). By contrast, genetic interaction similarity was not influenced by these two features ($p > 0.4$). Instead, genes with similar genetic interaction patterns were likely to be physically close to each other ($p = 0.0003$). In line with these results, semantic similarity and genetic interaction similarity were not correlated with each other ($\text{Rho} = 0.009$, $p = 0.6$). Thus, these two measures capture different aspects of functional similarity.

We tested the relationship between functional similarity and paired hotspot effects. Genetic interaction similarity was weakly but significantly correlated with paired hotspot effects ($\text{Rho} = 0.05$, $p = 0.003$), while semantic similarity was not ($\text{Rho} = 0.02$, $p = 0.10$). Because

functional similarity and paired hotspot effects are both shaped by similar sets of 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: $\text{Rho} = 0.04$, $p = 0.02$; semantic similarity: $\text{Rho} = 0.02$, $p = 0.23$). 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 non-synonymous 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* (Bergenholt et al., 2018). For example, the hotspot alters the expression of the beta unit of fatty acid synthetase, *FASI*, which is transcriptionally regulated by *Oaf1*. However, the promoters of 74% of genes affected by the *OAF1* hotspot are not known to be bound by *Oaf1* (Bergenholt et al., 2018), and the hotspot appears to alter the expression of some of these genes via horizontal mechanisms. For example, the hotspot affects the expression of the *PRSI* gene immediately downstream of *FASI*. *PRSI* is not a known transcriptional target of *Oaf1*, 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.

373 Discussion

374 In this work, we revealed a previously unrecognized mode by which *trans*-acting genetic
 375 variation can affect the expression of multiple genes. Typically, *trans*-eQTL effects are thought
 376 to propagate through regulatory pathways that are functionally related to the causal eQTL gene.
 377 In the alternative mode we describe here, the effects of a *trans*-eQTL can spread along a given
 378 chromosomal region and affect potentially unrelated genes (Fig 5).

380 Figure 5: A schematic illustrating two modes by which *trans*-acting sequence variation affects
 381 direct and indirect target genes. Left (pink box), the effects of a *trans*-acting variant percolate
 382 “vertically” among genes by directly targeting some genes for which altered abundance of the
 383 gene product influences other genes in turn. Right (blue box), effects of *trans*-acting variation
 384 propagate “horizontally” to genes located proximally to a direct target gene.

385 We showed that *trans*-eQTL hotspots in a yeast cross affected pairs of neighboring genes more
 386 frequently than expected by chance. The effect of the hotspots on adjacent gene pairs tended to
 387 involve genes that showed stronger coexpression in response to a broad range of perturbations.
 388 Thus, a straightforward explanation for the effects of eQTL hotspots on adjacent genes is that the
 389 genetic variants underlying the hotspots trigger the same genomic mechanisms that also shape

390 gene-gene coexpression.

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

402 A possible mechanism involves chromatin states that extend across adjacent genes, as
403 suggested by studies of individual gene pairs (Batada et al., 2007; Ebisuya et al., 2008; Raj et al.,
404 2006)). Indeed, adjacent genes that had correlated chromatin marks in normally growing cultures
405 and during response to cellular stress were more likely to be affected by hotspots as a doublet.
406 This relationship suggests that shared chromatin state is one mechanism by which hotspot effects
407 spread horizontally.

408 Proximity-dependent promiscuous effects of *cis*-regulatory elements are another possible
409 mechanism that could result in paired hotspot effects. In eukaryotes, transcriptional regulation is
410 shaped by *cis*-regulatory elements bound by sequence-specific transcription factors (Wittkopp &
411 Kalay, 2012). These elements, which include yeast upstream activating elements and enhancers
412 in higher eukaryotes, interact with the core promoters of specific genes (Hahn & Young, 2011).
413 Under proximity-dependent promiscuity, which has also been called “enhancer-promoter (EP)
414 theory” (Cera et al., 2019; Quintero-Cadena & Sternberg, 2016), *cis*-regulatory elements can also
415 regulate transcription from any promoter within a certain physical distance (Butler, 2001;
416 Ebisuya et al., 2008). This non-specific activity decays as distance between the element and a
417 gene increases. Distance-dependent coexpression has been described in species ranging from

418 yeast to human, and the reach of this phenomenon scales as a function of genome size
 419 (Quintero-Cadena & Sternberg, 2016). In the compact yeast genome, distance-dependent
 420 coexpression is detectable within a range of about 1,000 bp. In our data, distance only influenced
 421 pairing of hotspot effects among divergently expressed gene pairs. Most of these pairs are
 422 located less than 1,000 bp apart, well within the reach of non-specific EP processes. Thus,
 423 pairing of hotspot effects may be partly driven by proximity-dependent promiscuity of
 424 *cis*-regulatory elements.

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

435 Compared to yeast, our knowledge of human *trans*-eQTLs remains limited (Aguet et al.,
 436 2017; Battle et al., 2014; Brynedal et al., 2017; Fehrmann et al., 2011; Grundberg et al., 2012;
 437 Heinig et al., 2010; Lee et al., 2014; Small et al., 2011; Wright et al., 2014; Yao et al., 2017).
 438 Nevertheless, the prevalence of EP effects across species including humans (Quintero-Cadena &
 439 Sternberg, 2016) means that horizontal spreading of *trans*-effects is expected to also exist in
 440 other species.

441 **Materials and methods**

442 Code and datasets used in this work are available at
 443 (<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
 445 from the code on GitHub by following the instructions at the top of the R script. Figures in this
 446 table were paneled and combined in a separate image processing package.

447 Unless otherwise specified, analyses were conducted in R 3.6.0 (R Development Core Team,
 448 2010), using the following packages. C++ code was run using Rcpp 1.0.4 (Eddelbuettel &
 449 François, 2011). Data processing was performed using tidyverse 1.3.0 (Wickham et al., 2019).
 450 Genomic range data were imported using Rtracklayer 1.46.0 (M. Lawrence et al., 2009) and
 451 processed using Genomic Ranges 1.38.0 (Michael Lawrence et al., 2013) and Biostrings 2.54.0
 452 (H. Pagès, 2017). Plot color schemes were provided by the Viridis color package originally
 453 developed as part of Matplotlib (J. D. Hunter, 2007).

454

455 **Datasets**

456 All datasets used in this work are detailed in Suppl Table 3.

457

458 **Gene annotations**

459 Gene models were obtained from yeastmine (Balakrishnan et al., 2012; Fisk et al., 2006). The
 460 BY and RM strains differ in the presence or absence of stretches of genes close to the ends of the
 461 chromosome ends. *Trans*-eQTL effects on these subtelomeric genes may appear to be highly
 462 correlated simply due to their shared presence or absence in the BY / RM segregants. Therefore,
 463 we removed subtelomeric genes, defined as all genes whose start or stop codon was located
 464 outside of the genetic markers forming the linkage map in Source Data 3 from (Albert et al.,
 465 2018). We retained only “verified open reading frames” (ORFs) (Fisk et al., 2006) because some
 466 genes not annotated to be in this category (such as “dubious” ORFs or noncoding genes) overlap

467 verified ORFs (sometimes on the antisense strand), complicating analyses of intergenic distances
468 and TFBS annotations.

469

470 **Hotspot effects**

471 The effects of hotspots were represented in the “hotspot matrix”. This matrix was generated from
472 previously published data (Source Data 9 from (Albert et al., 2018)), which cataloged the effects
473 of all 102 hotspots (columns) on 5,629 expressed genes (rows). We removed 734 genes which
474 were not verified ORFs (Fisk et al., 2006) and introduced a row of zeros for 17 genes that were
475 expressed in the segregant data but that were not affected by any of the 102 hotspots. Genes were
476 defined as expressed if they were present in the $\log_2$ (TPM) expression value table in Source Data
477 1 from (Albert et al., 2018). The rows of the hotspot matrix were ordered by chromosome and
478 the position of the genes on each chromosome. We removed subtelomeric genes and retained
479 only verified ORFs as described above. After these filters, the hotspot matrix contained 4,912
480 genes, of which 4,878 were affected by at least one hotspot.

481

482 **Doublet definition**

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

487

488 **Permutation Tests**

489 To assess whether the observed doublet count exceeded random chance, we used a permutation
490 test. For each hotspot, we performed 10^5 permutations, in which the given hotspot effects were
491 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. Multiple testing correction across 102 hotspot was performed by dividing the significance threshold of $p = 0.05$ by 102. This process was repeated for doublets excluding divergent genes, as well as for triplets, quadruplets, quintuplets, and sextuplets.

Coexpression matrices

Expression datasets (Suppl Table 3) 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 & Bloom *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 & Kruglyak 2005: Data had been preprocessed following the procedures described in (Huttenhower et al., 2006). We used the mean of both microarray dye swaps 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

519 accession number PRJEB5348, where they had been deposited by (Gierliński et al., 2015).
 520 Counts for each gene were universe-normalized as above and filtered to retain only genes with at
 521 least a single read in at least half of the samples. Read counts were normalized by dividing them
 522 by the length of the gene. A half-count was added to each count before taking the natural log of
 523 each value. Counts were centered within each sample by subtracting the mean expression of all
 524 expressed genes in the sample from all expression values in that sample. Data then underwent
 525 quantile normalization as above.

526 Albert & Bloom *et al.*, 2018: Log(TPM) reads in Source Data 1 from (Albert et al., 2018)
 527 were processed to remove the effect of genetic variation, processing batch, and growth phase. As
 528 a proxy for the growth phase, we used the optical density at the time the cultures had been
 529 collected (Source Data 2 from (Albert et al., 2018)). For each gene, we built a fixed-effects linear
 530 model using processing batch, optical density, and peak markers of each significant *cis*- and
 531 *trans*-eQTL reported for this gene (Source Data 3 from (Albert et al., 2018)). The residuals from
 532 these models were centered by subtracting the mean of all expression measurements for that
 533 segregant from all expression values for that segregant. We then performed quantile
 534 normalization as above. To gauge if removal of eQTL effects had been successful, we examined
 535 genes with strong eQTLs. Prior to removal, such eQTLs cause the expression values of their
 536 target gene to show a bimodal distribution. We examined the gene *STE2*, which is affected by a
 537 strong *trans*-effect arising from the mating type locus, and the gene *HO*, which is affected by a
 538 strong local effect caused by an engineered deletion of this gene in the RM strain. Both genes
 539 had bimodal distributions prior to correction, which became unimodal in the residuals of their
 540 respective linear models (Suppl Fig 4).

541

542 **Comparison of similarity matrices**

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

545

546 **Quantifying paired hotspot effects across hotspots**

547 To capture the degree to which both genes in a given pair were affected by hotspots, we
 548 constructed a matrix containing hotspot effect summaries. We initially set every possible gene
 549 pair to a value of zero and then examined the 102 hotspots. If a hotspot affected both genes in a
 550 pair in the same direction, the score was increased by one. If both genes in a pair were affected in
 551 different directions, the score was decreased by one. Thus, these summarized hotspot effects are
 552 able to capture gene pairs that are in repulsion, such that they tend to be affected by multiple
 553 hotspots in opposite directions. We chose this metric for comparison to coexpression because it
 554 was able to capture negative values for gene pairs with anticorrelated expression, which may also
 555 show pairing of hotspot effects with opposite direction.

556 In our linear models of possible mechanisms underlying paired hotspot effects, we used a matrix
 557 that counted the number of hotspots affecting both members of a pair in the same direction.
 558 While this metric ignores pairs of genes with hotspot effects in opposite direction, we chose it for
 559 two reasons: 1) the absence of negative values allowed us to use a count-based generalized linear
 560 model (see below), and 2) many of the features we used to model paired hotspot effects cannot
 561 take on or are not meaningful for negative values (e.g. TFBS similarity, or distance between
 562 genes).

563

564 **Shared regulation of genes by transcription factors**

565 We used a curated collection of transcription factors annotated to regulate each gene in the yeast
 566 genome (Monteiro et al., 2020). These data were binary, with a value of zero indicating that a
 567 given TF is not known to regulate a given gene, and a value of one indicating that it does. TF
 568 profile similarity between gene pairs was calculated using the Jaccard index (Veerla & Höglund,
 569 2006).

570 **Nucleosome occupancy measures**

571 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, means for occupancy at each base pair across three replicates was used for our calculations.

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 were calculated using ChIP-seq data published by (Weiner et al., 2015). Specifically, we used the nucleosome maps (Table S2) and normalized marks (Table S3) available separately at ([https://www.cell.com/molecular-cell/fulltext/S1097-2765\(15\)00094-5](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. Table S2 from Weiner et al. contains the mapped positions of 66,360 nucleosomes in wild-type yeast (strain BY4741) and Table S3 from Weiner et al. contains the levels of all 26 chromatin modifications normalized to the data in their Table S2. Thus, the values in their 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.

We calculated similarity in baseline chromatin state between a pair of genes as the spearman correlation between the normalized levels of the 26 chromatin marks on the +1 nucleosomes of the two genes in mid-log growth, before the application of the stressor. Similarity in chromatin change was calculated as the spearman correlation between the difference in normalized levels of

the 26 chromatin marks on the +1 nucleosomes of neighboring genes between baseline and 15 minutes after the cells were exposed to diamide.

601

602 **Features influencing the number of hotspots that affect gene pairs**

To gauge the relative importance of various features on paired hotspot effects, 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 with zero paired effects 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, each giving the natural log of the number of times a given gene in a pair was affected by hotspots.

As predictor variables, our models included: 1) the orientation of 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 “closeness” on the chromosome, which was the natural log of the inverse of distance in bp, 4) similarity in chromatin marks between +1 nucleosomes, and 5) similarity in cell stress-induced changes in chromatin marks between +1 nucleosomes. To calculate the significance of each of these features, we fit a full model with all predictors, 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 & Weisberg, 2019). We also fit equivalent models to subsets of the data, separately for each of the three possible gene orientations.

621 **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

625 package (Carlson, 2017).

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

631

632 **Genes regulated by Oaf1**

633 All genes directly regulated by Oaf1 were obtained from previously published data
634 ((Bergenholt et al., 2018) DATA SET S1). Genes were considered to be regulated by Oaf1 if
635 their promoters were bound by Oaf1 in any of the four conditions studied by (Bergenholt et al.,
636 2018).

637 **Acknowledgements**

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

641 **Funding**

642 This work was funded by NIH grant R35GM124676 and a Sloan Research Fellowship
643 (FG-2018-10408) to FWA. FWA is a Pew Scholar in the Biomedical Sciences, supported by The
644 Pew Charitable Trusts.

645 References

- 646 Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., Mohammadi, P., Park, Y.,
647 Parsana, P., Segre, A. V., Strober, B. J., Zappala, Z., Cummings, B. B., Gelfand, E. T.,
648 Hadley, K., Huang, K. H., Lek, M., Li, X., Nedzel, J. L., ... Biospecimen Collection
649 Source Site—NDRI. (2017). Genetic effects on gene expression across human tissues.
650 *Nature*, 550(7675), 204–213. <https://doi.org/10.1038/nature24277>
- 651 Albert, F. W., Bloom, J. S., Siegel, J., Day, L., & Kruglyak, L. (2018). Genetics of
652 trans-regulatory variation in gene expression. *ELife*, 7, e35471.
653 <https://doi.org/10.7554/eLife.35471>
- 654 Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and
655 disease. *Nature Reviews Genetics*, 16(4), 197–212. <https://doi.org/10.1038/nrg3891>
- 656 Allocco, D. J., Kohane, I. S., & Butte, A. J. (2004). Quantifying the relationship between
657 co-expression, co-regulation and gene function. *BMC Bioinformatics*, 10.
- 658 Al-Shahrour, F., Minguéz, P., Marqués-Bonet, T., Gazave, E., Navarro, A., & Dopazo, J. (2010).
659 Selection upon Genome Architecture: Conservation of Functional Neighborhoods with
660 Changing Genes. *PLoS Computational Biology*, 6(10), e1000953.
661 <https://doi.org/10.1371/journal.pcbi.1000953>
- 662 Andrews, T., Honti, F., Pfundt, R., de Leeuw, N., Hehir-Kwa, J., Vulto-van Silfhout, A., de Vries,
663 B., & Webber, C. (2015). The clustering of functionally related genes contributes to
664 CNV-mediated disease. *Genome Research*, 25(6), 802–813.
665 <https://doi.org/10.1101/gr.184325.114>
- 666 Arnone, J. T., Arace, J. R., Soorneedi, A. R., Citino, T. T., Kamitaki, T. L., & McAlear, M. A.
667 (2014). Dissecting the *cis* and *trans* Elements That Regulate Adjacent-Gene Coregulation
668 in *Saccharomyces cerevisiae*. *Eukaryotic Cell*, 13(6), 738–748.
669 <https://doi.org/10.1128/EC.00317-13>
- 670 Arnone, J. T., & McAlear, M. A. (2011). Adjacent Gene Pairing Plays a Role in the Coordinated
671 Expression of Ribosome Biogenesis Genes *MPP10* and *YJR003C* in *Saccharomyces*
672 *cerevisiae*. *Eukaryotic Cell*, 10(1), 43–53. <https://doi.org/10.1128/EC.00257-10>
- 673 Arnone, J. T., Robbins-Pianka, A., Arace, J. R., Kass-Gergi, S., & McAlear, M. A. (2012). The
674 adjacent positioning of co-regulated gene pairs is widely conserved across eukaryotes.
675 *BMC Genomics*, 13(1), 546. <https://doi.org/10.1186/1471-2164-13-546>
- 676 Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P.,
677 Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L.,

- 678 Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., &
679 Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*,
680 25(1), 25–29. <https://doi.org/10.1038/75556>
- 681 Balakrishnan, R., Park, J., Karra, K., Hitz, B. C., Binkley, G., Hong, E. L., Sullivan, J., Micklem,
682 G., & Michael Cherry, J. (2012). YeastMine—An integrated data warehouse for
683 *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database*, 2012.
684 <https://doi.org/10.1093/database/bar062>
- 685 Batada, N. N., Urrutia, A. O., & Hurst, L. D. (2007). Chromatin remodelling is a major source of
686 coexpression of linked genes in yeast. *Trends in Genetics*, 23(10), 480–484.
687 <https://doi.org/10.1016/j.tig.2007.08.003>
- 688 Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C.,
689 Haudenschild, C. D., Beckman, K. B., Shi, J., Mei, R., Urban, A. E., Montgomery, S. B.,
690 Levinson, D. F., & Koller, D. (2014). Characterizing the genetic basis of transcriptome
691 diversity through RNA-sequencing of 922 individuals. *Genome Research*, 24(1), 14–24.
692 <https://doi.org/10.1101/gr.155192.113>
- 693 Baumgartner, U., Hamilton, B., Piskacek, M., Ruis, H., & Rottensteiner, H. (1999). Functional
694 analysis of the Zn(2)Cys(6) transcription factors Oaf1p and Pip2p. Different roles in fatty
695 acid induction of beta-oxidation in *Saccharomyces cerevisiae*. *The Journal of Biological*
696 *Chemistry*, 274(32), 22208–22216. <https://doi.org/10.1074/jbc.274.32.22208>
- 697 Bergenholm, D., Liu, G., Holland, P., & Nielsen, J. (2018). Reconstruction of a Global
698 Transcriptional Regulatory Network for Control of Lipid Metabolism in Yeast by Using
699 Chromatin Immunoprecipitation with Lambda Exonuclease Digestion. *MSystems*, 3(4),
700 e00215-17, /msystems/3/4/msys.00215-17.atom.
701 <https://doi.org/10.1128/mSystems.00215-17>
- 702 Bolstad, B. (2017). *PreprocessCore*. Bioconductor.
703 <https://doi.org/10.18129/B9.BIOC.PREPROCESSCORE>
- 704 Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From
705 Polygenic to Omnigenic. *Cell*, 169(7), 1177–1186.
706 <https://doi.org/10.1016/j.cell.2017.05.038>
- 707 Brem, R. B. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science*,
708 296(5568), 752–755. <https://doi.org/10.1126/science.1069516>
- 709 Brem, R. B., & Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene
710 expression traits in yeast. *Proceedings of the National Academy of Sciences*, 102(5),

- 711 1572–1577. <https://doi.org/10.1073/pnas.0408709102>
- 712 Brown, C. R., Mao, C., Falkovskaia, E., Jurica, M. S., & Boeger, H. (2013). Linking Stochastic
713 Fluctuations in Chromatin Structure and Gene Expression. *PLoS Biology*, 11(8),
714 e1001621. <https://doi.org/10.1371/journal.pbio.1001621>
- 715 Brynedal, B., Choi, J., Raj, T., Bjornson, R., Stranger, B. E., Neale, B. M., Voight, B. F., &
716 Cotsapas, C. (2017). Large-Scale trans -eQTLs Affect Hundreds of Transcripts and
717 Mediate Patterns of Transcriptional Co-regulation. *The American Journal of Human*
718 *Genetics*, 100(4), 581–591. <https://doi.org/10.1016/j.ajhg.2017.02.004>
- 719 Butler, J. E. F. (2001). Enhancer-promoter specificity mediated by DPE or TATA core promoter
720 motifs. *Genes & Development*, 15(19), 2515–2519. <https://doi.org/10.1101/gad.924301>
- 721 Bystriykh, L., Weersing, E., Dontje, B., Sutton, S., Pletcher, M. T., Wiltshire, T., Su, A. I.,
722 Vellenga, E., Wang, J., Manly, K. F., Lu, L., Chesler, E. J., Alberts, R., Jansen, R. C.,
723 Williams, R. W., Cooke, M. P., & de Haan, G. (2005). Uncovering regulatory pathways
724 that affect hematopoietic stem cell function using “genetical genomics.” *Nature Genetics*,
725 37(3), 225–232. <https://doi.org/10.1038/ng1497>
- 726 Carlson, M. (2017). *Org.Sc.sgd.db*. Bioconductor.
727 <https://doi.org/10.18129/B9.BIOC.ORG.SC.SGD.DB>
- 728 Caron, H., Schaik, B. van, Mee, M. van der, Baas, F., Riggins, G., Sluis, P. van, Hermus, M.-C.,
729 Asperen, R. van, Boon, K., Voûte, P. A., Heisterkamp, S., Kampen, A. van, & Versteeg,
730 R. (2001). The Human Transcriptome Map: Clustering of Highly Expressed Genes in
731 Chromosomal Domains. *Science*, 291(5507), 1289–1292.
732 <https://doi.org/10.1126/science.1056794>
- 733 Cera, A., Holganza, M. K., Hardan, A. A., Gamarra, I., Eldabagh, R. S., Deschaine, M.,
734 Elkamhawy, S., Sisso, E. M., Foley, J. J., & Arnone, J. T. (2019). Functionally Related
735 Genes Cluster into Genomic Regions That Coordinate Transcription at a Distance in
736 *Saccharomyces cerevisiae*. *MSphere*, 4(2), e00063-19,
737 [/msphere/4/2/mSphere063-19.atom. https://doi.org/10.1128/mSphere.00063-19](https://doi.org/10.1128/mSphere.00063-19)
- 738 Cesar, A. S. M., Regitano, L. C. A., Reecy, J. M., Poleti, M. D., Oliveira, P. S. N., de Oliveira, G.
739 B., Moreira, G. C. M., Mudadu, M. A., Tizioto, P. C., Koltes, J. E., Fritz-Waters, E.,
740 Kramer, L., Garrick, D., Beiki, H., Geistlinger, L., Mourão, G. B., Zerlotini, A., &
741 Coutinho, L. L. (2018). Identification of putative regulatory regions and transcription
742 factors associated with intramuscular fat content traits. *BMC Genomics*, 19(1), 499.
743 <https://doi.org/10.1186/s12864-018-4871-y>

- 744 Chereji, R. V., Ramachandran, S., Bryson, T. D., & Henikoff, S. (2018). Precise genome-wide
745 mapping of single nucleosomes and linkers in vivo. *Genome Biology*, 19(1), 19.
746 <https://doi.org/10.1186/s13059-018-1398-0>
- 747 Chesler, E. J., Lu, L., Shou, S., Qu, Y., Gu, J., Wang, J., Hsu, H. C., Mountz, J. D., Baldwin, N.
748 E., Langston, M. A., Threadgill, D. W., Manly, K. F., & Williams, R. W. (2005). Complex
749 trait analysis of gene expression uncovers polygenic and pleiotropic networks that
750 modulate nervous system function. *Nature Genetics*, 37(3), 233–242.
751 <https://doi.org/10.1038/ng1518>
- 752 Clément-Ziza, M., Marsellach, F. X., Codlin, S., Papadakis, M. A., Reinhardt, S.,
753 Rodríguez-López, M., Martin, S., Marguerat, S., Schmidt, A., Lee, E., Workman, C. T.,
754 Bähler, J., & Beyer, A. (2014). Natural genetic variation impacts expression levels of
755 coding, non-coding, and antisense transcripts in fission yeast. *Molecular Systems Biology*,
756 10(11), 764. <https://doi.org/10.15252/msb.20145123>
- 757 Cohen, B. A., Mitra, R. D., Hughes, J. D., & Church, G. M. (2000). A computational analysis of
758 whole-genome expression data reveals chromosomal domains of gene expression. *Nature*
759 *Genetics*, 26(2), 183–186. <https://doi.org/10.1038/79896>
- 760 Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W.,
761 Usaj, M., Hanchard, J., Lee, S. D., Pelechano, V., Styles, E. B., Billmann, M., van
762 Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J. S., ... Boone,
763 C. (2016). A global genetic interaction network maps a wiring diagram of cellular
764 function. *Science*, 353(6306), aaf1420–aaf1420. <https://doi.org/10.1126/science.aaf1420>
- 765 Ebisuya, M., Yamamoto, T., Nakajima, M., & Nishida, E. (2008). Ripples from neighbouring
766 transcription. *Nature Cell Biology*, 10(9), 1106–1113. <https://doi.org/10.1038/ncb1771>
- 767 Eddelbuettel, D., & François, R. (2011). **Rcpp**: Seamless R and C++ Integration. *Journal of*
768 *Statistical Software*, 40(8). <https://doi.org/10.18637/jss.v040.i08>
- 769 Eldabagh, R. S., Mejia, N. G., Barrett, R. L., Monzo, C. R., So, M. K., Foley, J. J., & Arnone, J.
770 T. (2018). Systematic Identification, Characterization, and Conservation of
771 Adjacent-Gene Coregulation in the Budding Yeast *Saccharomyces cerevisiae*. *MSphere*,
772 3(3), e00220-18, /msphere/3/3/mSphere220-18.atom.
773 <https://doi.org/10.1128/mSphere.00220-18>
- 774 Everett, L. J., Huang, W., Zhou, S., Carbone, M. A., Lyman, R. F., Arya, G. H., Geisz, M. S., Ma,
775 J., Morgante, F., St. Armour, G., Turlapati, L., Anholt, R. R. H., & Mackay, T. F. C.
776 (2020). Gene expression networks in the *Drosophila* Genetic Reference Panel. *Genome*

- 777 *Research*, 30(3), 485–496. <https://doi.org/10.1101/gr.257592.119>
- 778 Fehrmann, R. S. N., Jansen, R. C., Veldink, J. H., Westra, H.-J., Arends, D., Bonder, M. J., Fu, J.,
779 Deelen, P., Groen, H. J. M., Smolonska, A., Weersma, R. K., Hofstra, R. M. W.,
780 Buurman, W. A., Rensen, S., Wolfs, M. G. M., Platteel, M., Zhernakova, A., Elbers, C.
781 C., Festen, E. M., ... Franke, L. (2011). Trans-eQTLs Reveal That Independent Genetic
782 Variants Associated with a Complex Phenotype Converge on Intermediate Genes, with a
783 Major Role for the HLA. *PLoS Genetics*, 7(8), e1002197.
784 <https://doi.org/10.1371/journal.pgen.1002197>
- 785 Fisk, D. G., Ball, C. A., Dolinski, K., Engel, S. R., Hong, E. L., Issel-Tarver, L., Schwartz, K.,
786 Sethuraman, A., Botstein, D., Michael Cherry, J., & TheSaccharomyces Genome
787 Database Project. (2006). *Saccharomyces cerevisiae* S288C genome annotation: A
788 working hypothesis. *Yeast*, 23(12), 857–865. <https://doi.org/10.1002/yea.1400>
- 789 Fleming, J. A., Lightcap, E. S., Sadis, S., Thoroddsen, V., Bulawa, C. E., & Blackman, R. K.
790 (2002). Complementary whole-genome technologies reveal the cellular response to
791 proteasome inhibition by PS-341. *Proceedings of the National Academy of Sciences*,
792 99(3), 1461–1466. <https://doi.org/10.1073/pnas.032516399>
- 793 Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third edition). SAGE.
- 794 Fu, J., Keurentjes, J. J. B., Bouwmeester, H., America, T., Verstappen, F. W. A., Ward, J. L.,
795 Beale, M. H., de Vos, R. C. H., Dijkstra, M., Scheltema, R. A., Johannes, F., Koornneef,
796 M., Vreugdenhil, D., Breitling, R., & Jansen, R. C. (2009). System-wide molecular
797 evidence for phenotypic buffering in Arabidopsis. *Nature Genetics*, 41(2), 166–167.
798 <https://doi.org/10.1038/ng.308>
- 799 The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing
800 strong. *Nucleic Acids Research*, 47(D1), D330–D338.
801 <https://doi.org/10.1093/nar/gky1055>
- 802 Gierliński, M., Cole, C., Schofield, P., Schurch, N. J., Sherstnev, A., Singh, V., Wrobel, N.,
803 Gharbi, K., Simpson, G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2015).
804 Statistical models for RNA-seq data derived from a two-condition 48-replicate
805 experiment. *Bioinformatics (Oxford, England)*, 31(22), 3625–3630.
806 <https://doi.org/10.1093/bioinformatics/btv425>
- 807 Grundberg, E., Small, K. S., Hedman, Å. K., Nica, A. C., Buil, A., Meduri, E., Barrett, A.,
808 Nisbett, J., Sekowska, M., Wilk, A., Shin, S.-Y., Glass, D., Travers, M., Min, J. L., Ring,
809 S., Ho, K., Thorleifsson, G., Kong, A., Thorsteindottir, U., ... The Multiple Tissue
810 Human Expression Resource (MuTHER) Consortium. (2012). Mapping cis- and

- 811 trans-regulatory effects across multiple tissues in twins. *Nature Genetics*, 44(10),
812 1084–1089. <https://doi.org/10.1038/ng.2394>
- 813 Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjálmsson, B. J., Xu, H., Zang, C., Ripke, S.,
814 Bulik-Sullivan, B., Stahl, E., Kähler, A. K., Hultman, C. M., Purcell, S. M., McCarroll, S.
815 A., Daly, M., Pasaniuc, B., Sullivan, P. F., Neale, B. M., Wray, N. R., ... Sullivan, P. F.
816 (2014). Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11
817 Common Diseases. *The American Journal of Human Genetics*, 95(5), 535–552.
818 <https://doi.org/10.1016/j.ajhg.2014.10.004>
- 819 H. Pagès, P. A. (2017). *Biostrings*. Bioconductor.
820 <https://doi.org/10.18129/B9.BIOC.BIOSTRINGS>
- 821 Hahn, S., & Young, E. T. (2011). Transcriptional Regulation in *Saccharomyces cerevisiae*:
822 Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of
823 Activators and Coactivators. *Genetics*, 189(3), 705–736.
824 <https://doi.org/10.1534/genetics.111.127019>
- 825 Harrell, F. E. J. (2020). *Hmisc: Harrell Miscellaneous*.
826 <https://CRAN.R-project.org/package=Hmisc>
- 827 Heinig, M., Petretto, E., Wallace, C., Bottolo, L., Rotival, M., Lu, H., Li, Y., Sarwar, R., Langley,
828 S. R., Bauerfeind, A., Hummel, O., Lee, Y.-A., Paskas, S., Rintisch, C., Saar, K., Cooper,
829 J., Buchan, R., Gray, E. E., Cyster, J. G., ... Cardiogenics Consortium. (2010). A
830 trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk.
831 *Nature*, 467(7314), 460–464. <https://doi.org/10.1038/nature09386>
- 832 Hershberg, R., Yegerlotem, E., & Margalit, H. (2005). Chromosomal organization is shaped by
833 the transcription regulatory network. *Trends in Genetics*, 21(3), 138–142.
834 <https://doi.org/10.1016/j.tig.2005.01.003>
- 835 Hubner, N., Wallace, C. A., Zimdahl, H., Petretto, E., Schulz, H., Maciver, F., Mueller, M.,
836 Hummel, O., Monti, J., Zidek, V., Musilova, A., Kren, V., Causton, H., Game, L., Born,
837 G., Schmidt, S., Müller, A., Cook, S. A., Kurtz, T. W., ... Aitman, T. J. (2005). Integrated
838 transcriptional profiling and linkage analysis for identification of genes underlying
839 disease. *Nature Genetics*, 37(3), 243–253. <https://doi.org/10.1038/ng1522>
- 840 Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow,
841 S., Ward, T. R., Kidd, M. J., Friend, S. H., & Marton, M. J. (2000). Widespread
842 aneuploidy revealed by DNA microarray expression profiling. *Nature Genetics*, 25(3),
843 333–337. <https://doi.org/10.1038/77116>

- 844 Hurst, L. D., Pál, C., & Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene
845 order. *Nature Reviews Genetics*, 5(4), 299–310. <https://doi.org/10.1038/nrg1319>
- 846 Huttenhower, C., Hibbs, M., Myers, C., & Troyanskaya, O. G. (2006). A scalable method for
847 integration and functional analysis of multiple microarray datasets. *Bioinformatics*,
848 22(23), 2890–2897. <https://doi.org/10.1093/bioinformatics/btl492>
- 849 J. D. Hunter. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science &*
850 *Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- 851 Jansen, A., & Verstrepen, K. J. (2011). Nucleosome Positioning in *Saccharomyces cerevisiae*.
852 *Microbiology and Molecular Biology Reviews*, 75(2), 301–320.
853 <https://doi.org/10.1128/MMBR.00046-10>
- 854 Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot,
855 N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P., & Ahringer, J. (2003).
856 Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi.
857 *Nature*, 421(6920), 231–237. <https://doi.org/10.1038/nature01278>
- 858 Klevecz, R. R., Bolen, J., Forrest, G., & Murray, D. B. (2004). A genomewide oscillation in
859 transcription gates DNA replication and cell cycle. *Proceedings of the National Academy*
860 *of Sciences*, 101(5), 1200–1205. <https://doi.org/10.1073/pnas.0306490101>
- 861 Knijnenburg, T. A., Daran, J.-M. G., van den Broek, M. A., Daran-Lapujade, P. A., de Winde, J.
862 H., Pronk, J. T., Reinders, M. J., & Wessels, L. F. (2009). Combinatorial effects of
863 environmental parameters on transcriptional regulation in *Saccharomyces cerevisiae*: A
864 quantitative analysis of a compendium of chemostat-based transcriptome data. *BMC*
865 *Genomics*, 10(1), 53. <https://doi.org/10.1186/1471-2164-10-53>
- 866 Kraakman, L. S., Mager, W. H., Maurer, K. T., Nieuwint, R. T., & Planta, R. J. (1989). The
867 divergently transcribed genes encoding yeast ribosomal proteins L46 and S24 are
868 activated by shared RPG-boxes. *Nucleic Acids Research*, 17(23), 9693–9706. PubMed.
869 <https://doi.org/10.1093/nar/17.23.9693>
- 870 Kustatscher, G., Grabowski, P., & Rappsilber, J. (2017). Pervasive coexpression of spatially
871 proximal genes is buffered at the protein level. *Molecular Systems Biology*, 13(8), 937.
872 <https://doi.org/10.15252/msb.20177548>
- 873 Langley, S. R., Bottolo, L., Kunes, J., Zicha, J., Zidek, V., Hubner, N., Cook, S. A., Pravenec,
874 M., Aitman, T. J., & Petretto, E. (2013). Systems-level approaches reveal conservation of
875 trans-regulated genes in the rat and genetic determinants of blood pressure in humans.
876 *Cardiovascular Research*, 97(4), 653–665. <https://doi.org/10.1093/cvr/cvs329>

- 877 Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: An R package for interfacing
878 with genome browsers. *Bioinformatics*, 25(14), 1841–1842.
879 <https://doi.org/10.1093/bioinformatics/btp328>
- 880 Lawrence, Michael, Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.
881 T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges.
882 *PLoS Computational Biology*, 9(8), e1003118.
883 <https://doi.org/10.1371/journal.pcbi.1003118>
- 884 Lee, M. N., Ye, C., Villani, A.-C., Raj, T., Li, W., Eisenhaure, T. M., Imboywa, S. H., Chipendo,
885 P. I., Ran, F. A., Slowikowski, K., Ward, L. D., Raddassi, K., McCabe, C., Lee, M. H.,
886 Frohlich, I. Y., Hafler, D. A., Kellis, M., Raychaudhuri, S., Zhang, F., ... Hacohen, N.
887 (2014). Common Genetic Variants Modulate Pathogen-Sensing Responses in Human
888 Dendritic Cells. *Science*, 343(6175), 1246980–1246980.
889 <https://doi.org/10.1126/science.1246980>
- 890 Lenstra, T. L., Benschop, J. J., Kim, T., Schulze, J. M., Brabers, N. A. C. H., Margaritis, T.,
891 van de Pasch, L. A. L., van Heesch, S. A. A. C., Brok, M. O., Groot Koerkamp, M. J. A.,
892 Ko, C. W., van Leenen, D., Sameith, K., van Hooft, S. R., Lijnzaad, P., Kemmeren, P.,
893 Hentrich, T., Kobor, M. S., Buratowski, S., & Holstege, F. C. P. (2011). The Specificity
894 and Topology of Chromatin Interaction Pathways in Yeast. *Molecular Cell*, 42(4),
895 536–549. <https://doi.org/10.1016/j.molcel.2011.03.026>
- 896 Lewis, J. A., Broman, A. T., Will, J., & Gasch, A. P. (2014). Genetic Architecture of
897 Ethanol-Responsive Transcriptome Variation in *Saccharomyces cerevisiae* Strains.
898 *Genetics*, 198(1), 369–382. <https://doi.org/10.1534/genetics.114.167429>
- 899 Li, Q., Lee, B. T., & Zhang, L. (2005). Genome-scale analysis of positional clustering of mouse
900 testis-specific genes. *BMC Genomics*, 6(1), 7. <https://doi.org/10.1186/1471-2164-6-7>
- 901 Lutz, S., Brion, C., Kliebhan, M., & Albert, F. W. (2019). DNA variants affecting the expression
902 of numerous genes in trans have diverse mechanisms of action and evolutionary histories.
903 *PLOS Genetics*, 15(11), e1008375. <https://doi.org/10.1371/journal.pgen.1008375>
- 904 Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer.
- 905 Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P.,
906 Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kuttyavin, T., Stehling-Sun,
907 S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., ... Stamatoyannopoulos, J. A.
908 (2012). Systematic localization of common disease-associated variation in regulatory
909 DNA. *Science (New York, N.Y.)*, 337(6099), 1190–1195.

- 910 <https://doi.org/10.1126/science.1222794>
- 911 Michalak, P. (2008). Coexpression, coregulation, and cofunctionality of neighboring genes in
912 eukaryotic genomes. *Genomics*, 91(3), 243–248.
913 <https://doi.org/10.1016/j.ygeno.2007.11.002>
- 914 Monteiro, P. T., Oliveira, J., Pais, P., Antunes, M., Palma, M., Cavalheiro, M., Galocha, M.,
915 Godinho, C. P., Martins, L. C., Bourbon, N., Mota, M. N., Ribeiro, R. A., Viana, R.,
916 Sá-Correia, I., & Teixeira, M. C. (2020). YEASTRACT+: A portal for cross-species
917 comparative genomics of transcription regulation in yeasts. *Nucleic Acids Research*,
918 48(D1), D642–D649. <https://doi.org/10.1093/nar/gkz859>
- 919 Myers, K. S., Riley, N. M., MacGilvray, M. E., Sato, T. K., McGee, M., Heilberger, J., Coon, J.
920 J., & Gasch, A. P. (2019). Rewired cellular signaling coordinates sugar and hypoxic
921 responses for anaerobic xylose fermentation in yeast. *PLOS Genetics*, 15(3), e1008037.
922 <https://doi.org/10.1371/journal.pgen.1008037>
- 923 Ng, Y. K., Wu, W., & Zhang, L. (2009). Positive correlation between gene coexpression and
924 positional clustering in the zebrafish genome. *BMC Genomics*, 10(1), 42.
925 <https://doi.org/10.1186/1471-2164-10-42>
- 926 Orozco, L. D., Bennett, B. J., Farber, C. R., Ghazalpour, A., Pan, C., Che, N., Wen, P., Qi, H. X.,
927 Mutukulu, A., Siemers, N., Neuhaus, I., Yordanova, R., Gargalovic, P., Pellegrini, M.,
928 Kirchgessner, T., & Lusa, A. J. (2012). Unraveling Inflammatory Responses using
929 Systems Genetics and Gene-Environment Interactions in Macrophages. *Cell*, 151(3),
930 658–670. <https://doi.org/10.1016/j.cell.2012.08.043>
- 931 Pál, C., & Hurst, L. D. (2003). Evidence for co-evolution of gene order and recombination rate.
932 *Nature Genetics*, 33(3), 392–395. <https://doi.org/10.1038/ng1111>
- 933 Poyatos, J. F., & Hurst, L. D. (2007). The determinants of gene order conservation in yeasts.
934 *Genome Biology*, 8(11), R233. <https://doi.org/10.1186/gb-2007-8-11-r233>
- 935 Quintero-Cadena, P., & Sternberg, P. W. (2016). Enhancer Sharing Promotes Neighborhoods of
936 Transcriptional Regulation Across Eukaryotes. *G3: Genes|Genomes|Genetics*, 6(12),
937 4167. <https://doi.org/10.1534/g3.116.036228>
- 938 R Development Core Team. (2010). *R: a language and environment for statistical computing*:
939 *Reference index*. R Foundation for Statistical Computing.
940 <http://www.polsci.wvu.edu/duval/PS603/Notes/R/fullrefman.pdf>
- 941 Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA
942 Synthesis in Mammalian Cells. *PLoS Biology*, 4(10), e309.

- 943 <https://doi.org/10.1371/journal.pbio.0040309>
- 944 Rando, O. J., & Winston, F. (2012). Chromatin and Transcription in Yeast. *Genetics*, 190(2),
945 351–387. <https://doi.org/10.1534/genetics.111.132266>
- 946 Rockman, M. V., Skrovanek, S. S., & Kruglyak, L. (2010). Selection at Linked Sites Shapes
947 Heritable Phenotypic Variation in *C. elegans*. *Science*, 330(6002), 372–376.
948 <https://doi.org/10.1126/science.1194208>
- 949 Sameith, K., Amini, S., Groot Koerkamp, M. J. A., van Leenen, D., Brok, M., Brabers, N.,
950 Lijnzaad, P., van Hooff, S. R., Benschop, J. J., Lenstra, T. L., Apweiler, E., van
951 Wageningen, S., Snel, B., Holstege, F. C. P., & Kemmeren, P. (2015). A high-resolution
952 gene expression atlas of epistasis between gene-specific transcription factors exposes
953 potential mechanisms for genetic interactions. *BMC Biology*, 13(1), 112.
954 <https://doi.org/10.1186/s12915-015-0222-5>
- 955 Schep, A. N., Buenrostro, J. D., Denny, S. K., Schwartz, K., Sherlock, G., & Greenleaf, W. J.
956 (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin
957 architecture within regulatory regions. *Genome Research*, 25(11), 1757–1770.
958 <https://doi.org/10.1101/gr.192294.115>
- 959 Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N.,
960 Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2016). How
961 many biological replicates are needed in an RNA-seq experiment and which differential
962 expression tool should you use? *RNA*, 22(6), 839–851.
963 <https://doi.org/10.1261/rna.053959.115>
- 964 Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans.
965 *Trends in Genetics: TIG*, 34(7), 532–544. <https://doi.org/10.1016/j.tig.2018.03.007>
- 966 Simola, D. F., Francis, C., Sniegowski, P. D., & Kim, J. (2010). Heterochronic evolution reveals
967 modular timing changes in budding yeast transcriptomes. *Genome Biology*, 11(10), R105.
968 <https://doi.org/10.1186/gb-2010-11-10-r105>
- 969 Small, K. S., Hedman, Å. K., Grundberg, E., Nica, A. C., Thorleifsson, G., Kong, A.,
970 Thorsteindottir, U., Shin, S.-Y., Richards, H. B., Soranzo, N., Ahmadi, K. R., Lindgren,
971 C. M., Stefansson, K., Dermitzakis, E. T., Deloukas, P., Spector, T. D., McCarthy, M. I.,
972 the MuTHER Consortium, the GIANT Consortium, ... the DIAGRAM Consortium.
973 (2011). Identification of an imprinted master trans regulator at the KLF14 locus related to
974 multiple metabolic phenotypes. *Nature Genetics*, 43(6), 561–564.
975 <https://doi.org/10.1038/ng.833>

- 976 Smith, E. N., & Kruglyak, L. (2008). Gene–Environment Interaction in Yeast Gene Expression.
977 *PLoS Biology*, 6(4), e83. <https://doi.org/10.1371/journal.pbio.0060083>
- 978 Spellman, P. T., & Rubin, G. M. (2002). Evidence for large domains of similarly expressed genes
979 in the Drosophila genome. *Journal of Biology*, 8.
- 980 Sun, M., & Zhang, J. (2019). Chromosome-wide co-fluctuation of stochastic gene expression in
981 mammalian cells. *PLOS Genetics*, 15(9), e1008389.
982 <https://doi.org/10.1371/journal.pgen.1008389>
- 983 Sun, W., & Hu, Y. (2013). EQTL Mapping Using RNA-seq Data. *Statistics in Biosciences*, 5(1),
984 198–219. <https://doi.org/10.1007/s12561-012-9068-3>
- 985 Veerla, S., & Höglund, M. (2006). Analysis of promoter regions of co-expressed genes identified
986 by microarray analysis. *BMC Bioinformatics*, 7(1), 384.
987 <https://doi.org/10.1186/1471-2105-7-384>
- 988 Venables, W. N., Ripley, B. D., & Venables, W. N. (2002). *Modern applied statistics with S* (4th
989 ed). Springer.
- 990 Wade, C. H., Umbarger, M. A., & McAlear, M. A. (2006). The budding yeast rRNA and
991 ribosome biosynthesis (RRB) regulon contains over 200 genes. *Yeast*, 23(4), 293–306.
992 <https://doi.org/10.1002/yea.1353>
- 993 Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., & Chen, C.-F. (2007). A new method to measure
994 the semantic similarity of GO terms. *Bioinformatics*, 23(10), 1274–1281.
995 <https://doi.org/10.1093/bioinformatics/btm087>
- 996 Weiner, A., Hsieh, T.-H. S., Appleboim, A., Chen, H. V., Rahat, A., Amit, I., Rando, O. J., &
997 Friedman, N. (2015). High-Resolution Chromatin Dynamics during a Yeast Stress
998 Response. *Molecular Cell*, 58(2), 371–386. <https://doi.org/10.1016/j.molcel.2015.02.002>
- 999 West, M. A. L., Kim, K., Kliebenstein, D. J., van Leeuwen, H., Michelmore, R. W., Doerge, R.
1000 W., & St. Clair, D. A. (2007). Global eQTL Mapping Reveals the Complex Genetic
1001 Architecture of Transcript-Level Variation in Arabidopsis. *Genetics*, 175(3), 1441–1450.
1002 <https://doi.org/10.1534/genetics.106.064972>
- 1003 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G.,
1004 Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K.,
1005 Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the
1006 Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
1007 <https://doi.org/10.21105/joss.01686>

- 1008 Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and
1009 evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13(1), 59–69.
1010 <https://doi.org/10.1038/nrg3095>
- 1011 Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R.,
1012 Chung, W., Zhou, Y.-H., Abdellaoui, A., Batista, S., Butler, C., Chen, G., Chen, T.-H.,
1013 D’Ambrosio, D., Gallins, P., Ha, M. J., Hottenga, J. J., ... Boomsma, D. I. (2014).
1014 Heritability and genomics of gene expression in peripheral blood. *Nature Genetics*, 46(5),
1015 430–437. <https://doi.org/10.1038/ng.2951>
- 1016 Yao, C., Joehanes, R., Johnson, A. D., Huan, T., Liu, C., Freedman, J. E., Munson, P. J., Hill, D.
1017 E., Vidal, M., & Levy, D. (2017). Dynamic Role of trans Regulation of Gene Expression
1018 in Relation to Complex Traits. *The American Journal of Human Genetics*, 100(4),
1019 571–580. <https://doi.org/10.1016/j.ajhg.2017.02.003>
- 1020 Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., & Wang, S. (2010). GOSemSim: An R package for
1021 measuring semantic similarity among GO terms and gene products. *Bioinformatics*,
1022 26(7), 976–978. <https://doi.org/10.1093/bioinformatics/btq064>
- 1023 Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., Mackelprang, R., &
1024 Kruglyak, L. (2003). Trans-acting regulatory variation in *Saccharomyces cerevisiae* and
1025 the role of transcription factors. *Nature Genetics*, 35(1), 57–64.
1026 <https://doi.org/10.1038/ng1222>
- 1027 Zheng, W., Gianoulis, T. A., Karczewski, K. J., Zhao, H., & Snyder, M. (2011). Regulatory
1028 variation within and between species. *Annual Review of Genomics and Human Genetics*,
1029 12, 327–346. <https://doi.org/10.1146/annurev-genom-082908-150139>
- 1030 Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., Bumgarner, R. E., &
1031 Schadt, E. E. (2008). Integrating large-scale functional genomic data to dissect the
1032 complexity of yeast regulatory networks. *Nature Genetics*, 40(7), 854–861.
1033 <https://doi.org/10.1038/ng.167>
- 1034

1035 **Supplementary Material**

1036 Supplementary Table 1: Summary statistics for hotspot permutations, real doublet counts, and
 1037 p-values. P-values listed as “0” indicate that no count greater than or equal to that of the real data
 1038 was observed in the permutations. Thus, while the corresponding p-value is $<1 \times 10^{-5}$, we report
 1039 these cases as “0” to avoid Excel formatting cells with a “<” sign as text.

1040 Supplementary Table 2: Output of type III ANOVAs for generalized linear models.

1041 Supplementary Table 3: Datasets used in this work, with details on the source of data presented
 1042 here.

Supplementary Figures

1045 Supplement Figure 1: The number of doublets in real data (colored circles) are plotted against
 1046 the number of doublets in permuted data (white boxes). Boxes extend from the 25th to 75th
 1047 percentile. Whiskers extend from each box to the largest or smallest values within 1.5 times the
 1048 difference in the 25th to 75th percentile. Data beyond the whiskers are considered outliers and
 1049 plotted as small black squares. Circles are colored by the number of permutations in which a
 1050 number of doublets was counted that was greater than or equal to the number observed in the
 1051 real data. Hotspots are ordered along the x-axis based on the value of the doublet count in the
 1052 real data. For clarity, the ten hotspots with the largest number of doublets are shown with a
 1053 different scale.

47

1054

1055

1056 Supplementary Figure 2: The distribution of Rho values for each coexpression matrix generated
1057 from the 11 expression datasets.

Supplementary Figure 3: Distributions of doublet counts per neighboring gene pair and predictor variables, separated by gene pair orientation. Distance is shown in Fig 4A. (A) Number of times gene pairs were targeted by the same hotspot. (B) Similarity in regulation by transcription factors. (C) Correlation between marks on +1 nucleosomes without perturbation. (D) Correlation in change in marks in response to cell stress on +1 nucleosomes.

Supplementary Figure 4: Distribution of residual gene expression for two genes after correction for either only batch and optical density (OD) or batch, OD, and genetic variation (A) Residuals for *STE2*, whose expression is strongly affected by the mating type locus before correction for genetic variation. (B) Residuals for *HO*, which is deleted in the RM but not the BY strain used in Albert & Bloom *et al.*, 2018, resulting in a strong local eQTL in the BY / RM progeny.