Trans-eQTL hotspots shape complex traits by modulating cellular states

Kaushik Renganaath, Frank W. Albert

Department of Genetics, Cell Biology, & Development, University of Minnesota, Minneapolis, MN 55455, USA

Abstract

Regulatory genetic variation shapes gene expression, providing an important mechanism connecting DNA variation and complex traits. The causal relationships between gene expression and complex traits remain poorly understood. Here, we integrated transcriptomes and 46 genetically complex growth traits in a large cross between two strains of the yeast *Saccharomyces cerevisiae*. We discovered thousands of genetic correlations between gene expression and growth, suggesting potential functional connections. Local regulatory variation was a minor source of these genetic correlations. Instead, genetic correlations tended to arise from multiple independent *trans*-acting regulatory loci. *Trans*-acting hotspots that affect the expression of numerous genes accounted for particularly large fractions of genetic growth variation and of genetic correlations between gene expression and growth. Genes with genetic correlations were enriched for similar biological processes across traits, but with heterogeneous direction of effect. Our results reveal how *trans*-acting regulatory hotspots shape complex traits by altering cellular states.

Keywords

Genetic variation, eQTL, QTL, quantitative genetics, mediation, pleiotropy, IRA2

Introduction

Genetic variation among individuals shapes phenotypic variation , but the mechanistic basis of how DNA variants affect most traits remains elusive. The vast majority of genetic associations found by genome-wide association studies (GWAS) of human complex traits reside in non-coding DNA and are not in linkage disequilibrium (LD) with protein-coding exons [1]. The causal variants at these GWAS loci likely act by altering the expression of at least one gene whose abundance is critical for the given trait [2](Albert & Kruglyak, 2015). Catalogs of expression quantitative trait loci (eQTLs; regions of the genome that contain regulatory DNA variation) obtained in tissues from reference populations show enrichment of eQTLs at GWAS loci, and *vice versa* [3–6,6–13]. There is a growing list of validated causal connections between the expression of a specific gene and a complex trait in humans [14–16] and model organisms [17–20].

Recent results have added nuance to straightforward models of how regulatory variation alters gene expression and complex traits. In particular, many GWAS loci in humans still lack colocalized eQTLs, even though reference catalogs from dozens of tissues contain eQTLs for nearly every gene [21]. Many known reference eQTLs at genes likely to affect a given complex trait do not colocalize with GWAS hits at the same genes [22]. The expression changes caused by known reference eQTLs only mediate a small fraction (on average 11%) of complex trait heritability [23]. These limited contributions of known reference eQTLs to complex traits may be due to systematic differences between the variants that are detectable in current human eQTL versus GWA studies [24].

Here, we investigated the intersection of regulatory variation and complex traits by integrating published eQTLs and QTLs that shape complex organismal traits from two well-powered datasets in the yeast *Saccharomyces cerevisiae* [25,26] (Figure 1A). Both datasets were gathered in the same set of about 1,000 haploid, fully sequenced, recombinant progeny (“segregants”) obtained by crossing the laboratory strain “BY” (a close relative of the reference genome strain S288C) and the vineyard strain “RM”. These two strains differ at about 1 out of every 200 base pairs (bp), providing a rich reservoir of genetic variation for the dissection of complex traits.

The eQTLs were mapped using RNA-sequencing data from segregants growing exponentially in a common laboratory condition (yeast nitrogen base, “YNB”, medium without amino acids at 30˚C). The dataset includes 36,498 eQTLs affecting the expression of 5,643 genes [25]. The eQTLs fall into two classes. First, more than half of the genes are influenced by one “local” eQTL per gene, which are typically caused by DNA variants in cis-regulatory elements in the gene’s immediate vicinity [27,28]. Second, the majority of the eQTLs (92%; 33,529) were located far from the genes they influence, typically on a different chromosome. These trans-acting eQTLs affect the expression of their target gene via altered expression or activity of an intermediate, diffusible factor. In this cross, trans-eQTLs are clustered at 102 “hotspot” locations that each influence the expression of numerous genes. Similar trans-eQTL hotspots exist in other yeast crosses [29] and the wider yeast population [30], in crosses of strains in other species including plants [31], nematodes [32], and rodents [33], and may also exist in human populations [34–37]. Hotspots in the BY / RM cross have been shown to arise from causal DNA variants with diverse molecular consequences in genes with a wide range of functions from transcription factors to enzymes, and include lab-derived alleles as well as alleles that are common across S. cerevisiae isolates [17,18,38,39].

As complex organismal traits, we used relative growth in a range of environmental conditions. Specifically, the segregants were grown on solid agar plates under one of 46 varied conditions, such as different temperatures or various chemicals. Growth was measured as colony size after 2-day incubation. Genetic mapping yielded 591 QTLs across the 46 conditions [26]. Below, we refer to these QTLs as “growth QTLs” (gQTLs), and use the more general “QTLs” to encompass both eQTLs and gQTLs.

These datasets provide several key advantages. First, the identified QTLs account for a median of 72% (eQTLs) and 88% (gQTLs) of narrow-sense heritability [25,26], permitting analyses that capture most genetic variation in both gene expression and growth. Individual detected QTLs account for as little as ~1% of phenotypic variance, such that any undetected loci make at most modest contributions. Second, there are multiple eQTLs for nearly all genes (a median of six, and up to 21; nearly all are trans-eQTLs) and multiple gQTLs for all traits (a median of 12, ranging from 5 to 29). This enables analyses across independent genetic loci across the genome. Third, gene expression

and growth were measured in separate experiments conducted independently, years apart, at different institutions, and with segregants grouped into different batches, minimizing shared environmental factors between datasets. The gene expression data cannot have been influenced by exposure to the 46 conditions. Therefore, associations between gene expression and growth observed across these datasets most likely arise from a shared genetic basis. In this regard, our design is analogous to comparisons between human eQTL reference panels from individuals without a given disease and GWAS results from individuals with a given disease. In contrast to these studies, our eQTLs and gQTLs were mapped in the same genotypes (for similar designs in model organisms, see [31,40–45]). Integration of these datasets promises to yield insights into how genetic variation in gene expression affects complex, organismal phenotypes.

Our analyses revealed thousands of significant correlations between the expression of individual genes and growth in a given condition. A prominent source of these genetic correlations were *trans*-eQTL hotspots that modulate central biological processes, reflected in expression change at hundreds of genes. Different traits showed positive or negative correlations with the same processes. These results shed light on the consequences of regulatory variation and serve as a resource connecting gene expression and complex growth traits.

Results

1. The expression of thousands of genes is genetically correlated with growth in dozens of conditions

To identify associations between gene expression and genetically complex organismal traits, we tested for correlations between the expression of each of 5,643 genes that have at least one eQTL in liquid YNB medium [25] and growth in 46 environmental conditions [26] (Figure 1A; Table S1). In principle, a correlation between two phenotypes (such as gene expression and growth) across datasets can be caused by environmental and genetic sources [46]. Given our datasets were collected in independent experiments with numerous differences (e.g. liquid vs. solid media, different segregant arrangements into measurement plates and plate positions, different research institutions), environmentally induced correlations are unlikely in these data. Instead, any observed

correlations likely arise from genetic variation that shapes both gene expression in the standard YNB condition and growth in the 46 environments.

Thousands of genes showed a significant correlation between expression level and growth in at least one condition at a false discovery rate (FDR; estimated as q-values) of 5% (median across traits: 2,296 range: 15 in cadmium chloride to 4,190 in YNB; Figure 1B, Table S2 & S3). These thousands of genetic correlations per growth trait were robust to correction for shared factors among growth traits such as the solid agar medium to which conditions were added (Supplementary text, Supplementary Figures 1 – 3), suggesting that they reflect trait-specific biology. The median absolute correlation coefficients of significant genetic correlations ranged from 0.09 (paraquat) to 0.18 (cadmium chloride), with a median across traits of 0.11 (Supplementary Figure 4). These effect sizes suggest that genetic effects on gene expression in a standard environment shape growth in different environments partially but not completely, resembling recent results in humans [23]. Downsampling showed that detection of significant correlations of these magnitudes was possible due to the large size of the segregant panel (Supplementary text, Supplementary Figures 5 & 6). The thousands of genetic correlations revealed here are due to unknown sets of genomic loci that each influence both gene expression and growth. Below, we first identify the responsible loci. We then explore biological processes involved in the genetic correlations.

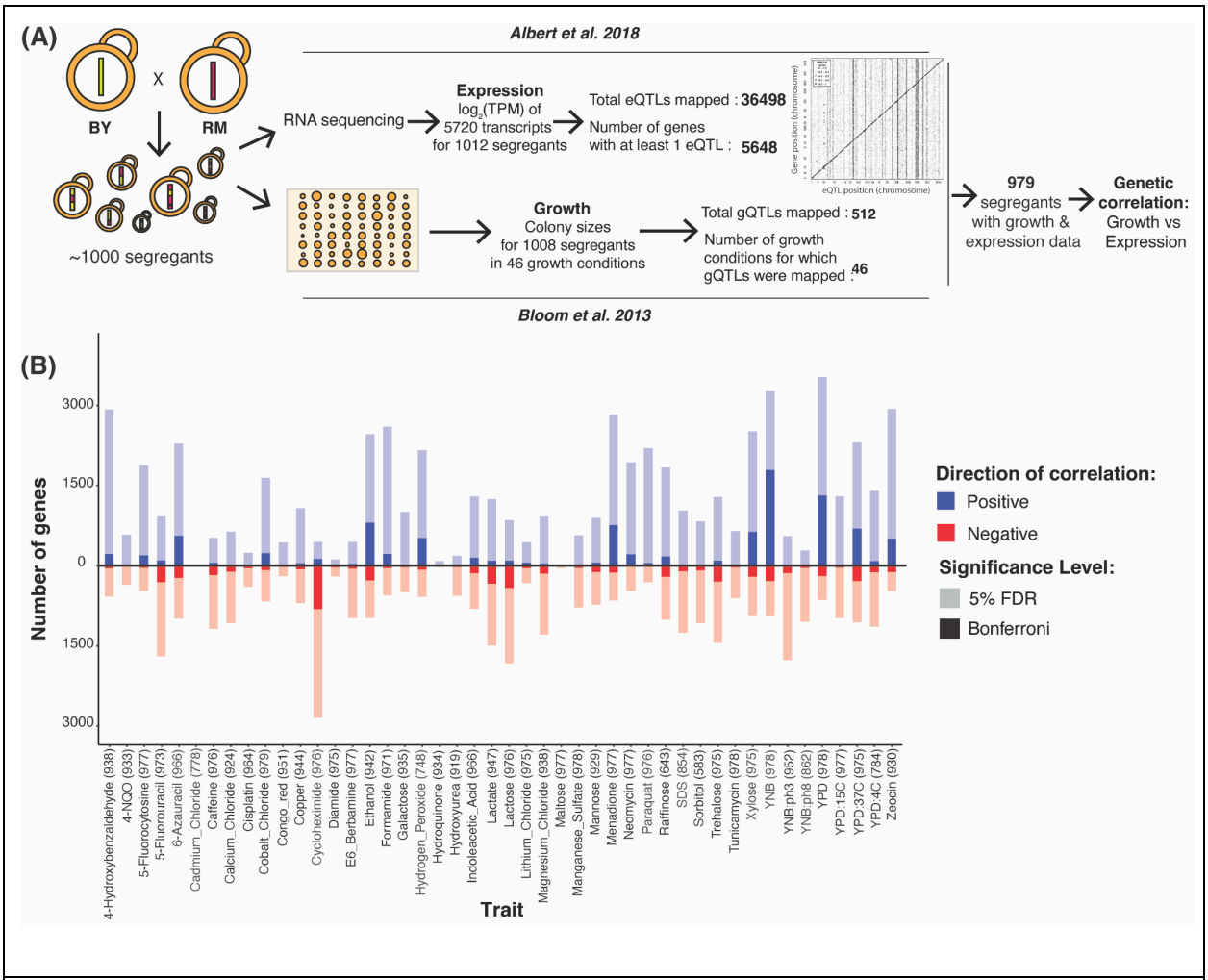

Figure 1: Genetic correlations. (A) Cartoon summarizing the datasets used to compute genetic correlations. (B) Number of genes with significant genetic correlation in each of the 46 growth conditions at 5% FDR (pale colors) and Bonferroni significance (saturated colors). Positive correlations are in blue and above the zero line; negative correlations are in red and below. For each condition, the number of segregants with both growth and expression measurements in Bloom et al. (2013) and Albert et al. (2018) used to compute the genetic correlations are indicated in parentheses.

2. Most growth QTLs are colocalized with multiple local eQTLs

137 We first focused on local eQTLs, defined as eQTLs whose physical confidence intervals
138 span the gene whose expression they affect (Figure 2A). More than half of all
139 expressed genes in this cross have a detectable local eQTL [25]. Local eQTLs tend to be
140 stronger than trans-eQTLs, intuitively suggesting that they could be a major source of
141 genetic correlations. All but one of the 591 gQTLs in our dataset overlapped with at
142 least one local eQTL, slightly but significantly more than expected by chance (median of
143 1,000 random gQTL sets: 586, $p = 0.023$). We used colocalization tests [47] to focus on
144 1,052 pairs of strong overlapping local eQTLs and gQTLs that are likely to share causal
145 variants (Figure 2A & Supplementary Text). Nearly all of the tested gQTLs (178 / 188;
146 95%; Figure 2B & C) had such a colocalized local eQTL.

147 In a straightforward model of how regulatory variation influences a trait, a gQTL is due
148 to a single strong local eQTL for a relevant gene. A likely example is a gQTL on
149 chromosome X that affects growth in the presence of tunicamycin. This gQTL overlaps
150 with three local eQTLs (Figure 2D), but only the eQTL for *CHS6* was consistent with
151 shared causal variants. Deletion of *CHS6* leads to increased resistance to tunicamycin
152 [48]. This observation is in line with the effects at this locus, where the RM allele
153 decreased *CHS6* expression ($r = -0.24 \pm 0.03$) and increased growth in tunicamycin ($r =$
154 0.2 ± 0.03).

155 In contrast with this straightforward model, 87% of tested gQTLs (163 / 188) were
156 colocalized with multiple local eQTLs (a median of 5, and up to 22; Figure 2B, Table
157 S4). Rather than suggesting causal genes that underlie the gQTL, many of these
158 colocalizations are likely due to linkage between gQTLs and the many local eQTLs in
159 this one-generation cross (Figure 2A, bottom). Inspection of individual gQTLs with
160 known causal genes confirmed this expectation (Supplementary Text, Supplementary
161 Figure 7). Further, the magnitudes of genetic correlation at gene / trait pairs were not
162 strongly dependent on whether a local eQTL for the gene was colocalized with a gQTL
163 for the trait (Supplementary Text; also see below). A conservative interpretation of these
164 results is that many and perhaps most local eQTLs are not causal for their colocalized
165 gQTLs.

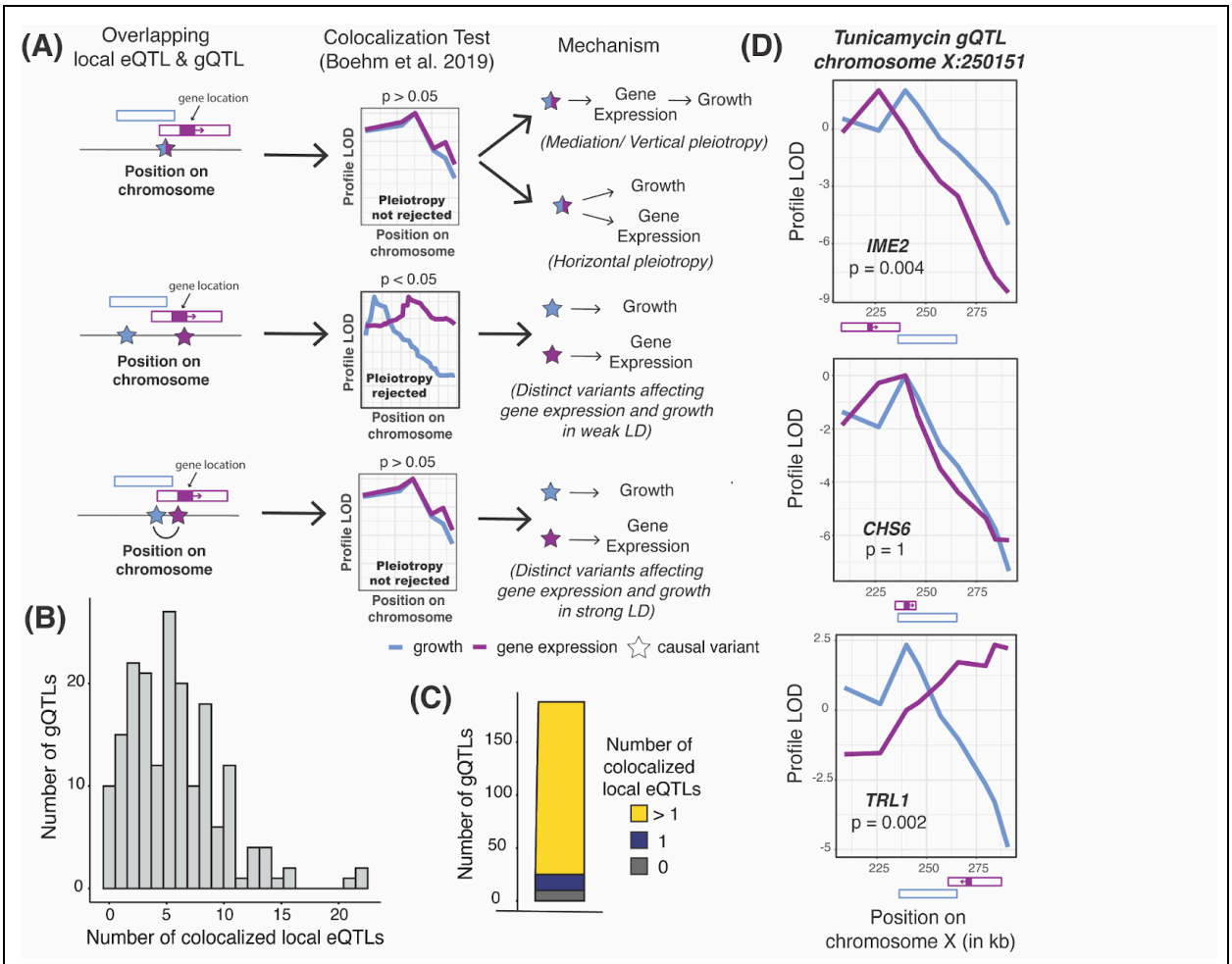

Figure 2: Local eQTL analyses. (A) Illustration of causal variants at local eQTLs that overlap a gQTL. QTL locations are shown as hollow boxes; gene position and transcriptional direction is indicated as solid purple box and arrow. A significant p-value on the colocalization test indicates that a single QTL caused by shared causal variants is rejected in favor of two separate QTLs. (B) Histogram of the number of gQTLs with a given number of colocalized local eQTLs. (C) The number of gQTLs with zero, one, and more than one colocalized local eQTL. (D) Profile LOD curves for growth in tunicamycin (blue) and the expression of indicated genes (purple) at a gQTL on chromosome X. “Profile LOD" scores are computed under the two-QTLs model as a LOD trace for one trait while keeping the QTL for the second trait fixed at its maximum-likelihood position, followed by subtracting the maximum LOD score of the single-QTL model in this region [47]. A maximum profile LOD value of zero (as for

CHS6) indicates that the two-QTLs model fits no better than the single-QTL model. Positive maximum profile LOD values (as for *TRL1* and *IME2*) indicate that the two-QTLs model fits better than the single-QTL model, with the two profile LOD curves peaking at the best positions for the two QTLs. QTL and gene positions shown as in A.

3. Multiple *trans*-eQTLs per gene contribute to genetic correlations

Local eQTLs in a biparental cross represent the effects of linked variants in a single genome region, effectively providing just a single datapoint for comparing genetic effects on gene expression and growth. In contrast, the datasets analyzed here comprise multiple gQTLs for each condition and more than one eQTL for each of 5,495 genes. Most of these eQTLs are located on different chromosomes from their target genes and influence gene expression via *trans*-acting mechanisms. These multiple QTLs per trait or gene are caused by distinct, unlinked variants, providing genetically independent observations of the relationship between a gene's expression and a trait. To examine the relationship between eQTLs and gQTLs globally, we first divided the genome into 10-kb bins and counted the number of eQTL and gQTL peak markers in each bin. eQTL and gQTL counts were significantly correlated across bins (all eQTLs: Spearman rank correlation rho = 0.31, p = 9e-28; only *trans*-eQTLs: rho = 0.30, p = 4e-27; similar results were obtained with bins of size 1 kb and 50 kb: all rho ≥ 0.16, all p ≤ 3e-16), indicating broad co-occurrence of loci that affect gene expression and growth.

If the precise abundance of a gene is important for a growth trait, its independent eQTLs should affect the trait in a consistent fashion [49]. For example, if higher expression of a gene increases growth, we would expect eQTLs that increase expression to also increase growth, while eQTLs that reduce expression to decrease growth. We searched for gene / trait pairs with such "QTL effect correlations". Briefly, we extracted the peak markers of all eQTLs for a given gene, computed their effects on gene expression and on growth, and performed weighted correlation tests on these effects (Methods; Figure 3A & B). We examined 5,186 genes with at least 3 eQTLs and chose a lenient FDR of

20% because these analyses were based on a median of 6 loci, limiting statistical power. At this threshold, 2,038 gene / trait pairs showed a QTL effect correlation, distributed over 32 traits (median: 9 pairs among these traits; up to 1,106 pairs for cycloheximide, Table S2 & S3). These significant QTL effect correlations had very high coefficients (median absolute r = 0.92), as expected for correlations based on few data points that achieve an acceptable FDR.

If genetic correlations arise from multiple independent eQTLs across the genome, we would expect correspondence between QTL effect correlations and genetic correlations. To test this, we examined the 28 traits that had at least three genes with significant QTL effect correlations. At 21 traits, genes with QTL effect correlations were enriched among genes with significant genetic correlations (FET uncorrected $p < 0.05$, Table S5). For 22 traits, there was a significant positive correlation between the magnitudes of the significant QTL effect correlations and their matched genetic correlations (Pearson correlation uncorrected $p < 0.05$, Table S5). Combining all traits and genes, there was strong correspondence between genetic correlations and the QTL effect correlations (rho = 0.69, $p < 2.2e-16$, Figure 3C).

In sum, we identified genes with multiple independent eQTLs that had consistent effects on the gene's expression and on growth in a given condition. These QTL effect correlations showed agreement with the genetic correlations, suggesting that the genetic correlations tend to arise from multiple eQTLs, most of which affect gene expression in *trans*.

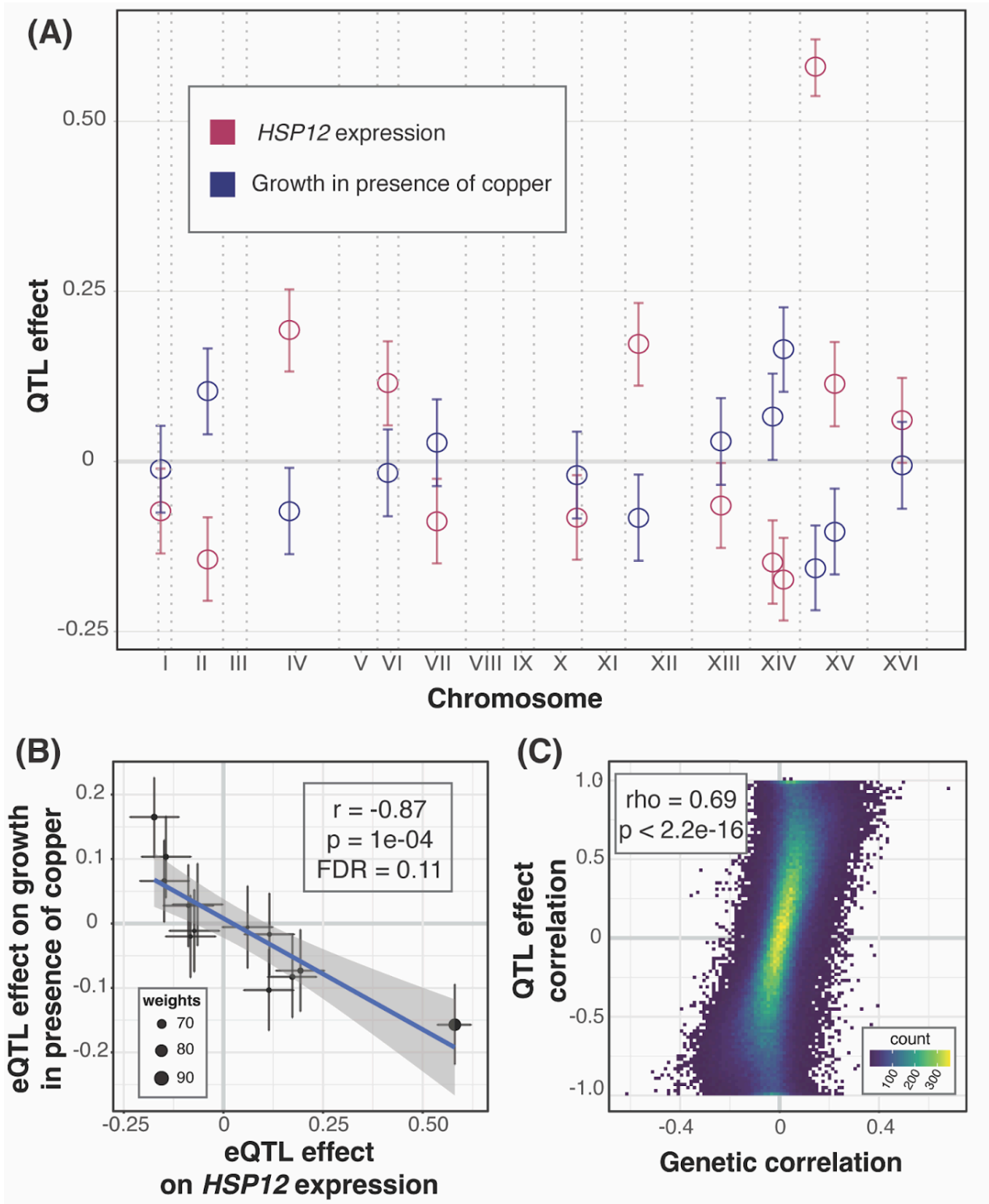

Figure 3: QTL effect correlations. (A) As an example, the plot shows the locations of eQTLs affecting the gene *HSP12* (x-axis) and their effects (y-axis; coefficient of

correlation between trait and genotype) on *HSP12* expression (red) and growth in the presence of copper (blue). 95% effect size confidence intervals are shown. *HSP12* and copper were chosen as the example due to the large number of eQTLs and strong effect size correlation. (B) Scatterplot of the effects from panel A. The line represents the weighted regression. Weights are calculated as the inverse product of the widths of the 95% effect size confidence intervals and is reflected in the size of the point. (C) Scatterplot between QTL effect correlation coefficients and genetic correlation coefficients across all genes and all growth conditions.

211

212 **4. *Trans*-eQTL hotspots influence complex growth traits**

213 The vast majority of *trans*-eQTLs in this yeast cross are clustered at 102 hotspot
214 regions that each influence the expression of dozens to thousands of genes (Figure
215 4A). Their broad effects suggest that *trans*-eQTL hotspots could also play a prominent
216 role in shaping complex traits. Indeed, about two thirds of gQTLs (398 / 591) overlapped
217 at least one *trans*-eQTL hotspot (Figure 4A), exceeding random chance ($p < 0.001$,
218 Figure 4B).

219 To quantify the influence of hotspots on growth variation, we fit mixed linear models to
220 estimate the fraction of phenotypic variance in the 46 growth traits that is explained by
221 DNA variation at *trans*-eQTL hotspots. We compared these values to the total additive
222 heritability of the traits, estimated as the fraction of phenotypic variance due to all
223 markers in the genome (Figure 5A). Across the 46 conditions, a median of 77% of the
224 heritability was captured by all 102 *trans*-eQTL hotspot markers (Table S2). To put these
225 values in context, we computed the proportion of trait variance explained by the gQTLs
226 of the given condition (Methods). The amount of variance captured by the 102 hotspot
227 markers corresponded to a median of 97% of the trait variance explained by the gQTLs
228 for the given condition (Figure 5A, Table S2). Given the 102 hotspots are distributed
229 widely across the genome, we asked if the heritability captured by the hotspots could
230 reflect random proximity to linked causal variants that are distinct from those that cause
231 the hotspots. For 30 of the 46 conditions, the additive heritability explained by the

hotspot markers was greater than that explained by 95% of 1,000 sets of 102 random genetic markers (Figure 5A). On average, the random sets explained about 28.6% of the total additive heritability, much less than the actual hotspot positions. Thus, while some of the heritability assigned to hotspots probably arises from linked non-hotspot variants, *trans*-eQTL hotspots contribute a significant amount of heritability to yeast growth traits.

The hotspots vary widely in how many genes they affect (from 26 to 4,594), suggesting that their effects on growth are also likely to vary. Indeed, while nearly all (91 of 102) hotspots overlapped a gQTL, this number was slightly but significantly lower than expected (based on 1,000 sets of randomly placed gQTLs; $p < 0.001$; Figure 4C). Thus, while gQTLs are enriched at hotspots, not all hotspots are equally relevant. To explore this observation more quantitatively, we ranked the hotspots by the number of genes they affect, formed sets comprised of an increasing number of hotspots (starting with the two hotspots affecting the most genes, and up to all 102 hotspots), and asked how much variance these sets explained in each of the 46 traits (Supplementary Figure 8). Across traits, a median of 14 hotspots (with a range of 2 to 80) was needed to explain half of the maximum variance explained by all 102 hotspots, and a median of 41 hotspots (10 to 97) was needed to explain 80% of the maximum variance. Thus, while not every trans-eQTL hotspot in the baseline liquid YNB condition contributes to growth variation, dozens of hotspots do. Further, the number of genes affected by a hotspot was positively correlated with the average proportion of variance the hotspot explained across traits (rho = 0.6, p = 2.4e-10, Figure 4D), suggesting that the average contribution of a hotspot to trait variation is related to the breadth of effects a given hotspot has on the transcriptome.

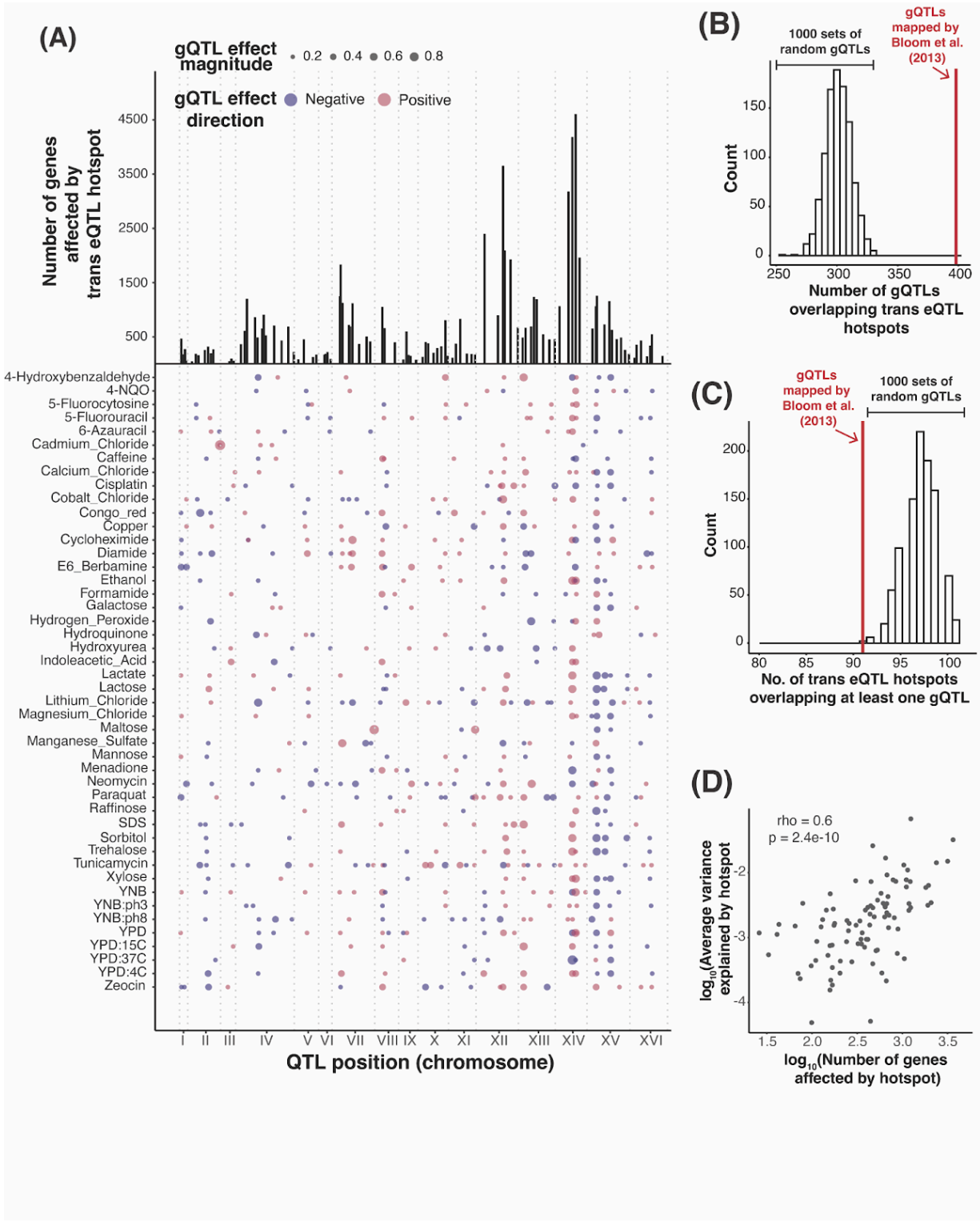

Figure 4: *Trans*-eQTL hotspot and gQTL locations. (A) The top panel shows the

number of genes affected by the 102 *trans*-eQTL hotspots (vertical lines). The bottom panel shows the locations of the 591 gQTLs as points with size scaled by effect size (coefficient of correlation between trait and genotype at the gQTL). Positive (RM allele increases growth compared to BY allele) and negative (RM allele decreases growth) gQTL effects are red and blue, respectively. (B) Histogram showing the distribution of the number of gQTLs that overlap a *trans*-eQTL hotspot in 1,000 sets of randomly placed gQTLs, compared to the actual gQTLs (red line). (C) Histogram showing the number of *trans*-eQTL hotspots that overlap at least one gQTL in 1,000 sets of randomly placed gQTLs, compared to the actual data (red line). (D) Scatter plot comparing the average growth variance explained by each hotspot and the number of genes whose expression they influence. The top two hotspots influencing the expression of most number of genes (chrXIV:466588_T/G and chrXIV:372376_G/A) are excluded from the analysis as it is not possible to estimate the average proportion of variance explained by just two markers under the linear mixed model scheme we use to estimate the proportion of variance explained by hotspots.

256

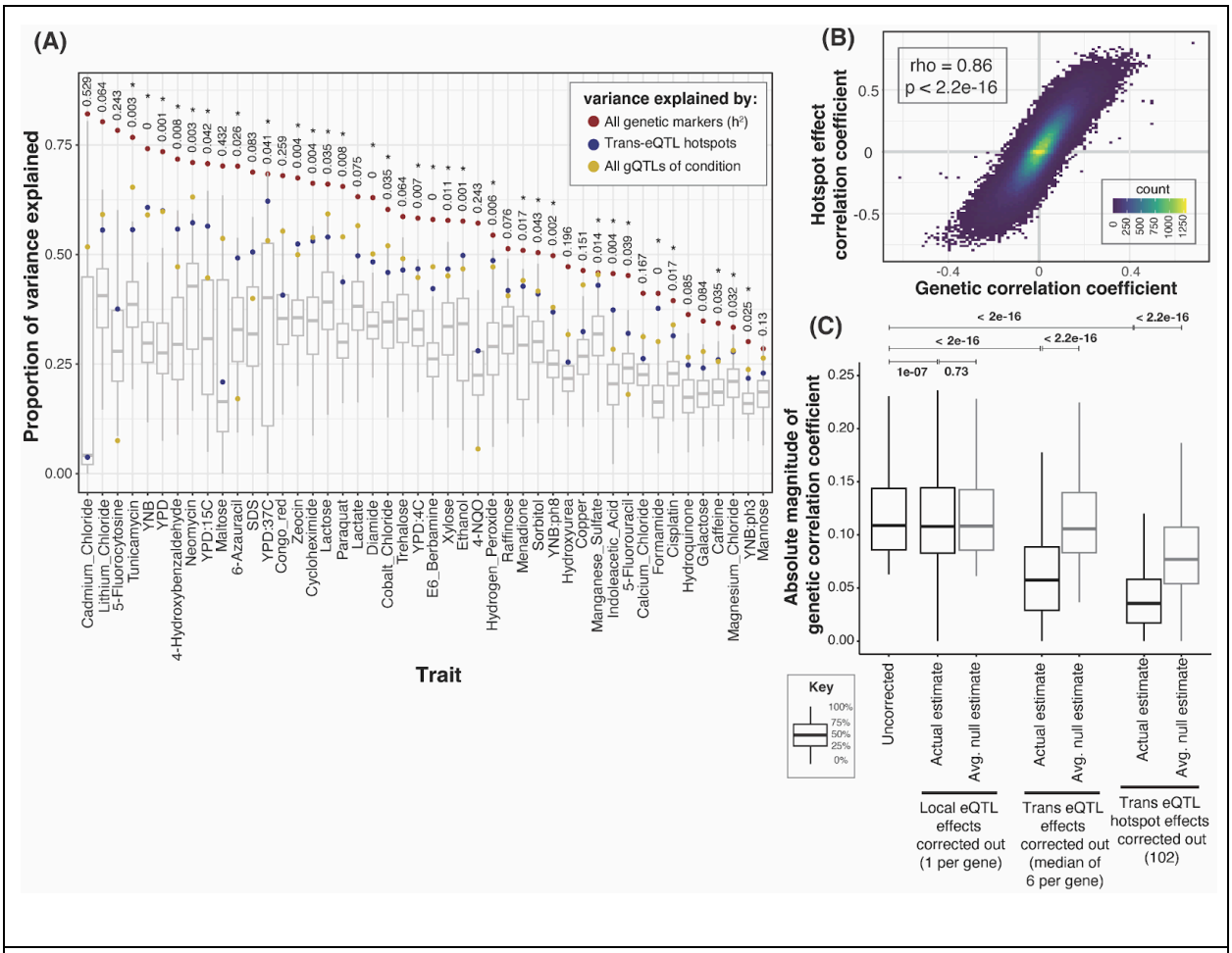

Figure 5: Phenotypic variance and genetic correlations linked to *trans*-eQTL hotspots. (A) Proportion of phenotypic variance explained by all 11,530 genetic markers (narrow-sense heritability; red dots), the 102 *trans*-eQTL hotspots (blue dots), and all gQTLs (yellow dots), in 46 growth conditions. Box plots show the distribution of variance explained by 1,000 sets of 102 random genetic markers. These distributions were used to calculate p-values (shown at top) for the proportion of variance explained by the hotspots. *: $p < 0.05$. (B) Scatterplot between hotspot effect correlation coefficients and genetic correlation coefficients across all genes and growth conditions. (C) Boxplots showing distributions of magnitudes of genetic correlation coefficients before ("uncorrected") and after removing the effects of local eQTLs (one per gene), *trans*-eQTLs (range 1 to 21 per gene, median of 6), and the 102 *trans*-eQTL hotspots (the number of markers regressed out for each category are indicated along the x-axis). This analysis only included genes with at least one local

eQTL and one *trans*-eQTL, and only uncorrected genetic correlations that were significant at 5% FDR. For each category, a random expectation was computed by calculating the decrease in the magnitude of genetic correlations after regressing out the effects of 1,000 sets of random markers of size equal to the actual set. Wilcoxon p-values for the difference in medians are indicated above the boxplots.

We asked whether the *trans*-eQTL hotspots are drivers of genetic correlations. To test this, we applied the procedure used to compute QTL effect correlations above to all *trans*-eQTL hotspot markers (Methods). Briefly, we calculated the effects of the hotspot markers on the expression of each of 5,720 genes, as well as on growth in each of the 46 conditions. We then tested for correlation between these hotspot effects on gene expression and on growth (below, “hotspot correlations”, Supplementary Figure 9).

Across the 46 conditions, a median of 1,704 genes had a significant hotspot correlation at an FDR of 5%, with a range of 4 genes (YNB at pH 8) to 3,795 (YPD) (Table S2 & S3). Notably, these hotspot correlations agreed strongly with the respective genetic correlations for the same gene / trait pairs. This agreement was seen both when combining all genes and all traits (Figure 5B) and within individual traits: at 45 / 46 traits, there was significant enrichment between genes with significant hotspot correlations and genes with significant genetic correlations (FET, uncorrected $p < 0.05$). For all traits, the coefficients of the hotspot correlations and of the genetic correlations were strongly correlated across all genes ($p < 2.2e-16$, Table S5; median $r = 0.86$). For 32 / 46 traits, these observed correlations exceeded those seen in 95% of 1,000 random sets of 102 markers, showing that the high agreement between hotspot correlations and genetic correlations is not simply due to random linkage of hotspots to other causal variants. Overall, the strength of the agreement between hotspot correlations and genetic correlations suggests that *trans*-eQTL hotspots are major drivers of genetic correlations between gene expression and growth.

To directly contrast the contribution of different sets of loci to the genetic correlations, we tested how each genetic correlation changes when the effects of the given loci are

removed. Specifically, for each gene / trait pair, we regressed out the gene's local eQTL, the gene's set of detected *trans*-eQTLs, or all *trans*-eQTL hotspots from both the gene expression and growth measurements. We then re-calculated the genetic correlations on the residual phenotypes. A drop in the strength of the genetic correlations would suggest that the removed locus was a contributor to the genetic correlation. Removal of each of the three QTL sets resulted in significant drops in the magnitude of the genetic correlations (Figure 5C). Similar patterns were seen for the individual growth traits, both for reductions in the strength and in the number of significant genetic correlations (Supplementary Figures 10 & 11). Removal of *trans*-eQTLs and especially of the hotspots resulted in much larger reductions than removal of local eQTLs (Figure 5C; median drop in absolute genetic correlation coefficient: local eQTLs: 1%, *trans*-eQTLs: 48%, hotspots: 68%). Trans-eQTLs and hotspots, but not the local eQTLs, resulted in reductions that were larger than expected by chance (Figure 5C). These results highlight the importance of *trans*-acting eQTLs for complex trait variation.

5. Major cellular states shape growth in specific conditions

To explore biological processes that connect gene expression and growth, we performed functional enrichment analyses. We focused on genetic correlations; enrichment results were broadly similar for QTL effect correlations and especially for the hotspot correlations (Table S6). Genes with significant genetic correlations were enriched for a variety of biological processes as summarized in GO-slim annotations (Figure 6). Across all 46 traits, there were 889 enrichments (p-value < 1e-5; corresponding to Bonferroni-correction for 100 GO categories tested in each of 46 traits), with individual traits having between zero and 43 enriched terms, with a median of 19.

Clustering traits based on their enrichment profiles revealed three distinct trait groups. For a group of 25 traits (shown in green in Figure 6), genes with positive genetic correlations were enriched for terms involved in ribosome biogenesis and translation, while genes with negative correlations were enriched for terms involved in cell cycle regulation, cytoskeleton organization, cellular respiration, and various metabolic

processes. In yeast, there is a well-established relationship between gene expression and growth rate in chemostat cultures [50]. Genes with ribosomal functions are expressed more highly during faster growth, while genes involved in energy metabolism are expressed more highly during slower growth. We grouped genes based on published signatures of fast and slow growth [50] and found that the expression of genes in these signatures was correlated with growth in the ‘green trait group (Figure 6). Further, genes known to increase their expression during the environmental stress response [51] showed negative genetic correlations (Figure 6), while genes that decrease their expression during the stress response showed positive genetic correlations. Thus, higher growth in the conditions in the ‘green’ trait group is associated with a genetic predisposition for higher translation and ribosome production, lower respiration, and for a gene expression pattern that is the opposite of an acute stress response.

In contrast, a group of 17 traits that formed the ‘blue’ trait group in Figure 6 tended to show enrichments for many of the same terms as the ‘green’ trait group but in the opposite direction. For example, terms related to translation tended to be enriched among genes with negative, rather than positive, genetic correlations. Some traits in this group, e.g. neomycin and hydrogen peroxide, showed strong positive correlations with the slow growth expression signature as well as with genes upregulated during the stress response. Thus, higher growth (as measured by larger colony sizes) in conditions in this trait group is associated with a genetic predisposition for lower translation, increased metabolism, and for a gene expression pattern that resembles an acute environmental stress response. The remaining four traits in the ‘orange’ group showed enrichments that combined features of the two major groups.

In sum, the genetic correlations revealed specific biological processes that are correlated with growth across several conditions. There is substantial heterogeneity in these relationships, such that growth in some conditions appears to be increased by higher expression of genes involved in a certain process (such as translation or the stress response), while growth in other conditions shows the opposite pattern.

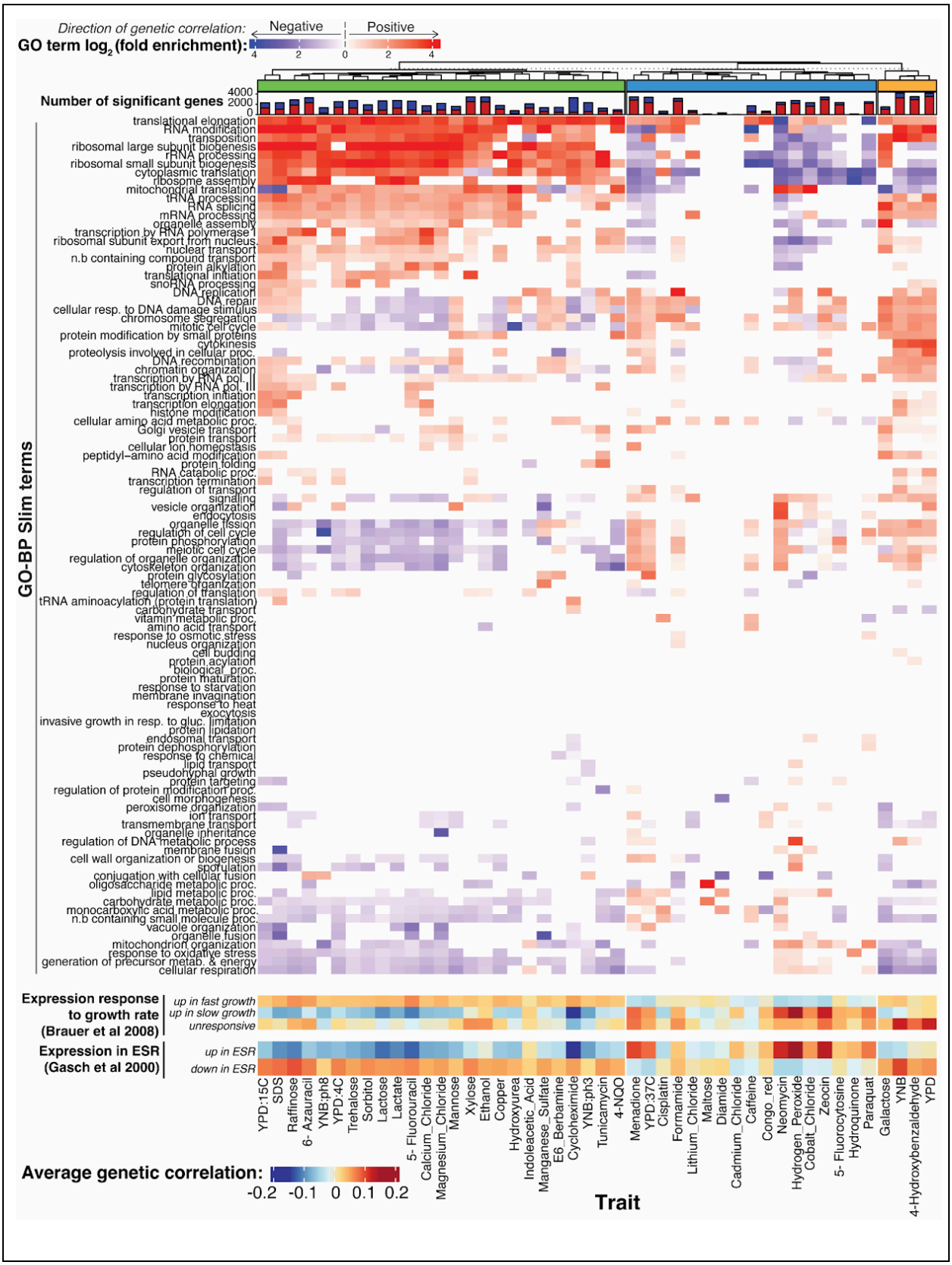
Direction of genetic correlation:
Negative
Positive
GO term log2 (fold enrichment):
4 2 0 2 4
Number of significant genes
4000
2000
0
GO-BP Slim terms
translational elongation
RNA modification
transposition
ribosomal large subunit biogenesis
rRNA processing
ribosomal small subunit biogenesis
cytoplasmic translation
ribosome assembly
mitochondrial translation
tRNA processing
RNA splicing
mRNA processing
organelle assembly
transcription by RNA polymerase I
ribosomal subunit export from nucleus
nuclear transport
n.b containing compound transport
protein alkylation
translational initiation
snoRNA processing
DNA replication
DNA repair
cellular resp. to DNA damage stimulus
chromosome segregation
mitotic cell cycle
protein modification by small proteins
cytokinesis
proteolysis involved in cellular proc.
DNA recombination
chromatin organization
transcription by RNA pol. I
transcription by RNA pol. III
transcription initiation
transcription elongation
histone modification
cellular amino acid metabolic proc.
Golgi vesicle transport
protein transport
cellular ion homeostasis
peptidyl–amino acid modification
protein folding
RNA catabolic proc.
transcription termination
regulation of transport
signaling
vesicle organization
endocytosis
organelle fission
regulation of cell cycle
protein phosphorylation
meiotic cell cycle
regulation of organelle organization
cytoskeleton organization
protein glycosylation
telomere organization
regulation of translation
tRNA aminoacylation (protein translation)
carbohydrate transport
vitamin metabolic proc.
amino acid transport
response to osmotic stress
nucleus organization
cell budding
protein acylation
biological_proc.
protein maturation
response to starvation
membrane invagination
response to heat
exocytosis
invasive growth in resp. to gluc. limitation
protein lipidation
endosomal transport
protein dephosphorylation
response to chemical
lipid transport
pseudohyphal growth
protein targeting
regulation of protein modification proc.
cell morphogenesis
peroxisome organization
ion transport
transmembrane transport
organelle inheritance
regulation of DNA metabolic process
membrane fusion
cell wall organization or biogenesis
sporulation
conjugation with cellular fusion
oligosaccharide metabolic proc.
lipid metabolic proc.
carbohydrate metabolic proc.
monocarboxylic acid metabolic proc.
n.b containing small molecule proc.
vacuole organization
organelle fusion
mitochondrion organization
response to oxidative stress
generation of precursor metab. & energy
cellular respiration
Expression response to growth rate (Brauer et al 2008)
up in fast growth
up in slow growth
unresponsive
Expression in ESR (Gasch et al 2000)
up in ESR
down in ESR
YPD:15C
SDS
Raffinose
6- Azauracil
YNB:ph8
YPD:4C
Trehalose
Sorbitol
Lactose
Lactate
5- Fluorouracil
Calcium_Chloride
Magnesium_Chloride
Mannose
Xylose
Ethanol
Copper
Hydroxyurea
Indoleacetic_Acid
Manganese_Sulfate
E6_Berbamine
Cycloheximide
YNB:ph3
Tunicamycin
4-NQO
Menadione
YPD:37C
Cisplatin
Formamide
Lithium_Chloride
Maltose
Diamide
Cadmium_Chloride
Caffeine
Congo_red
Neomycin
Hydrogen_Peroxide
Cobalt_Chloride
Zeocin
5- Fluorocytosine
Hydroquinone
Paraquat
Galactose
YNB
4-Hydroxybenzaldehyde
YPD
Average genetic correlation:
-0.2 -0.1 0 0.1 0.2
Trait

Figure 6: Biological processes enriched in genetic correlations. The top heatmap shows $\log_2$-fold enrichments for 100 GO-BP Slim terms in the set of genes with significant (5% FDR) genetic correlations. Enrichments for genes with positive and negative genetic correlations are shown in red and blue, respectively. Enrichments are displayed for terms significant at uncorrected $p < 0.001$. The stacked bar plot above the heatmap shows the number of genes with significant genetic correlation (5% FDR) for each condition, with positive genetic correlation in red and negative correlations in blue. The dendrogram clusters traits based on GO-enrichment, with the three trait groups discussed in the text indicated. The heatmaps at the bottom show average genetic correlation coefficients for genes in the indicated gene groups from Brauer et al. [50] and Gasch et al [51].

The *IRA2 trans*-eQTL hotspot mediates growth in hydrogen peroxide by modulating the expression of Msn2 target genes

Our genetic correlation analyses revealed that growth in different conditions is associated with specific biological processes. *Trans*-eQTL hotspots likely affect growth by altering the activity of these processes, as reflected by expression change at multiple genes. To explore this connection at a specific locus, we examined a hotspot caused by multiple missense variants in the *IRA2* gene [52]. This hotspot affects the expression of 1,240 genes [25] and overlaps gQTLs for 37 of the 46 conditions. Ira2 negatively regulates Ras-GTP signaling, resulting, among other consequences, in the expression of stress response genes by increasing active levels of the transcription factor Msn2. The *IRA2* RM allele is more active than the BY allele [52,53]. We therefore expected segregants carrying the RM allele to have higher expression of stress response genes regulated by Msn2 than segregants carrying the BY allele and, therefore, to be better able to grow in stress conditions including oxidative stress caused by hydrogen peroxide [54]. Indeed, the *IRA2* hotspot overlapped a gQTL for growth in hydrogen peroxide, where the RM allele increased growth compared to the BY allele ($r = 0.36$, $p = 3.5e-24$).

We asked which of the genes that are influenced by the IRA2 hotspot mediate growth in hydrogen peroxide (Figure 7A). Mediation analyses revealed 380 genes at an FDR of ≤ 5% (Figure 7B, Table S7). These 380 genes were enriched for oxidoreductase function (hypergeometric test p = 5e-7), the oxidative stress response, and several biological processes associated with energy metabolism (Table S8). The 380 genes were enriched for regulatory targets of Msn2 (FET p = 5.1e-14; 291 / 380 genes, Table S7) and included established reporters of Msn2 activity (*HSP12, TPS2* and *PNC1*) [55]. The effect of the IRA2 hotspot on growth in hydrogen peroxide was also mediated by its effect on the expression of MSN2 itself, at an FDR of 12%. The proportion of the *IRA2* hotspot effect on growth in hydrogen peroxide that was mediated by the downstream targets of Msn2p (median: 10%, range: 2.5% to 52%) was greater than that mediated by Msn2 itself (2.8%) (Figure 7, Table S7). These mediation analyses support the idea that the *IRA2* hotspot influences growth in hydrogen peroxide by influencing the expression of oxidative stress response genes regulated by Msn2. Ira2 may influence the expression of these genes primarily by altering the activity rather than the abundance of Msn2, , which reacts to stress by translocation to the nucleus upon phosphorylation [56]. The mediation effect on MSN2 suggests that the precise abundance of Msn2 might also play a role [57]. This example illustrates how hotspots may orchestrate the expression of genes in a pathway to shape growth in a given condition.

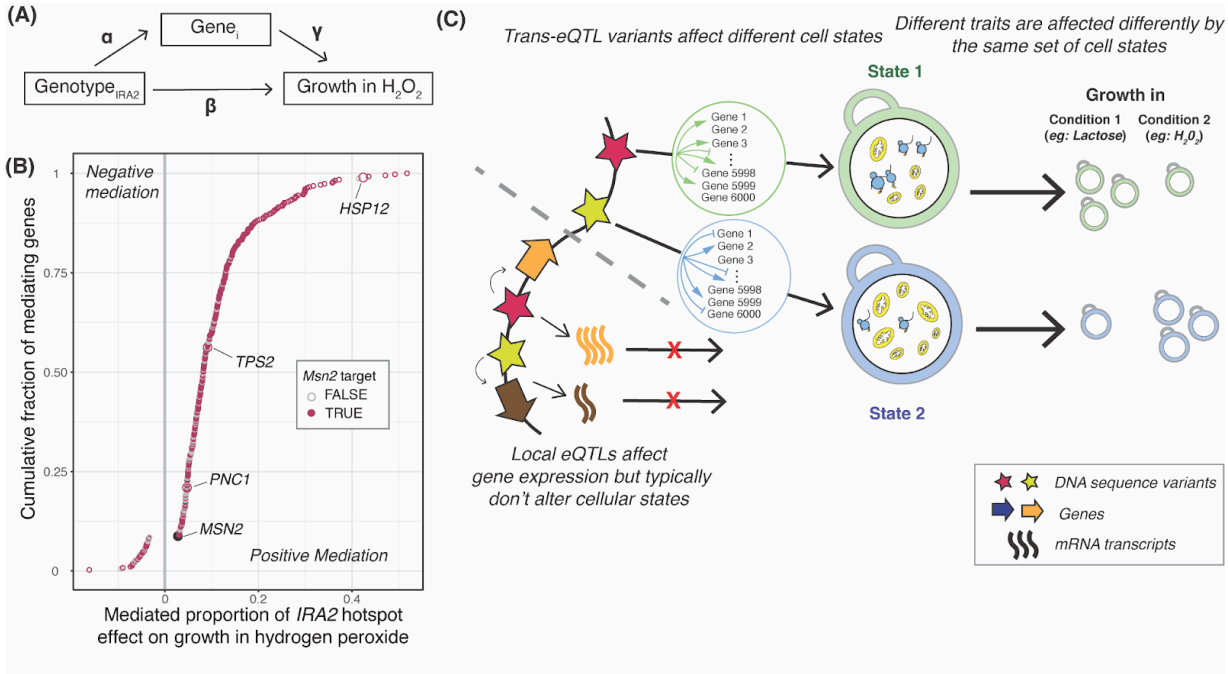

Figure 7: Mediation of growth variation at a *trans*-eQTL hotspot. (A) Conceptual model showing how the effect of the *IRA2* hotspot on growth in hydrogen peroxide may be mediated by the expression of hotspot target genes (*Gene$_i$*). The coefficients used in the mediation analysis (Methods) are indicated. (B) Mediated proportion of the *IRA2* hotspot effect on growth in hydrogen peroxide for 380 genes with significant mediation (5% FDR). Colored circles indicate whether a gene is annotated as an Msn2 target in the Yeastract database. Genes described in the text are highlighted. Positive mediation indicates genes for which increased expression increases growth, while negative mediation indicates genes for which increased expression decreases growth. (C) Conceptual model of how *trans*-eQTL hotspots affect growth by altering cellular states, as reflected in the hotspot's effect on multiple genes. The two example states roughly correspond to the 'green' and 'blue' trait clusters in Figure 6, with high translation and low respiration versus low translation and high respiration. Growth in different conditions is affected differently by these states. Most local eQTLs (shown below the dashed grey line) affect gene expression but do not go on to affect key cellular states underlying growth variation.

377

Discussion

Our integrative analyses of eQTLs that alter mRNA abundance and gQTLs that shape yeast growth revealed key insights into the interplay of gene expression and complex traits. Most prominently, there were thousands of correlations between gene expression in a commonly used laboratory condition and growth in a variety of other conditions. Because gene expression was measured separately from growth, shared environmental factors are unlikely to have induced these correlations. Instead, the correlations arise from shared genetic effects on gene expression and growth. Similarly, cellular responses to growth cannot have influenced our gene expression data. YNB medium is a possible exception given this medium was used to measure gene expression, but even here, growth and expression data were collected separately, with several experimental differences. Therefore, our results support a model in which inherited alleles shape an individual's gene expression profile in a baseline condition in ways that can influence growth when cells encounter additional environmental factors not present in the baseline condition.

The contribution of local eQTLs to genetic correlations was statistically significant but modest in magnitude (Figure 5C). While there is a growing list of local eQTLs that have been shown to underlie QTLs for complex traits (for examples in yeast, see Fay et al.[58] and Salinas et al. [20]) the large number of local eQTLs in this cross restricted our ability to systematically annotate gQTLs to specific causal local eQTLs. gQTLs can also be caused by coding variants (for examples in yeast, see Fay et al.[58] and Sadhu et al.[39]) that may not act via altered expression of the causal gene and do not involve any of the local eQTLs in the given region.

In contrast, the joint effects of multiple *trans*-eQTLs per gene was a major source of genetic correlations between gene expression and growth (Figure 5C). Enabled by the large number of mapped *trans*-eQTLs in this cross, we used these loci as genetically independent observations to search for consistent relationships between expression change at a given gene and growth in a given condition. This strategy is similar to analyses of allelic series at a given gene in human populations, in which alleles of varying molecular severity, ranging from knockouts to subtle *cis*-eQTLs, are assayed for

408 their effects on a complex trait to create dose-response curves [59,60]. Related genetic
409 strategies and synthetic approaches have been used to relate the expression level of
410 several focal genes to yeast fitness [61–64] and human gene regulation [65,66]. Here, we
411 applied this logic to independent, mostly *trans*-acting eQTLs across the genome and to
412 dozens of traits.

413 More than 90% of *trans*-eQTLs in this cross occur at hotspot positions. By applying
414 effect size correlations to *trans*-eQTL hotspots, we found that the joint effects of the
415 hotspots contribute strongly to growth trait heritability and genetic correlations between
416 gene expression and growth. Together, our results showed that genetic correlations
417 between gene expression and growth are typically not caused by single eQTLs with
418 strong effect (such as local eQTLs) but instead by the joint action of a collection of loci
419 across the genome, most of which affect mediating genes in *trans*.

420 Enrichment analyses of the genetic correlations suggested certain functional gene
421 groups as important mediators of complex trait variation. The same GO terms tended to
422 be enriched across traits (Figure 6), highlighting central biological processes with broad
423 importance for many traits. However, there were differences in the direction of these
424 enrichments.

425 We summarize these observations in a model centered on key cellular processes such
426 as translation, biomass production, and metabolism (Figure 7C). These processes are
427 in a balance, akin to a recently described cell state axis related to PKA and TOR
428 signaling, along which genetically different strains vary quantitatively [67]. The activity of
429 these key processes and the state of the cell is reflected in coordinated abundance
430 variation of hundreds of genes. Complex growth traits are shaped by variation in the key
431 processes. Depending on specific trait biology, growth in some conditions is increased
432 by a cell state that is genetically biased towards high biomass production and fast
433 growth, along with a dampened stress response. This pattern is seen in the ‘green’ trait
434 group in Figure 6, which includes several alternative carbon sources. Growth in other
435 conditions may be increased by a genetic predisposition for a stronger stress response.
436 Examples are traits in the ‘blue’ trait group, which includes stressors such as high
437 temperatures and oxidative stress caused by H_2O_2. In these conditions, genetic

predisposition for an increased baseline stress response signature may play a similar role as the phenomenon of inducible adaptive stress responses, in which exposure to low levels of one stressor protects cells from subsequent exposure to a different stressor [68–72]. Future work will need to disentangle exactly how genetic effects on the specific processes found here shape growth (Supplementary text).

Combining the above model with our QTL observations suggests a two-tiered architecture for how eQTLs shape complex traits (Figure 7C). First, in spite of their prevalence and stronger effects on gene expression [25], many local eQTLs appear to have little influence on growth traits because they do not alter the relevant central processes [21,67]. By contrast, some DNA variants do rise above this molecular background of functionally inert local variation and do affect cellular state and growth. These variants can affect cell states directly or indirectly, can be coding or non-coding, and may occur at a diverse set of genes throughout the genome [73]. When a variant does affect cell state, this effect becomes detectable as expression changes at the many genes whose abundance constitutes and reflects the state, resulting in the observation of a *trans*-eQTL hotspot. Hotspot effects on critical cell states explain the overall enrichment of *trans*-eQTL hotspots at gQTLs we observed here (Figure 4).

Our study has several limitations. Gene expression in a single baseline condition provides a good model for reference eQTL panels commonly used in other species[12,37,44,45,74] and escapes issues with reverse causation, but cannot reveal trait causation by regulatory genetic effects only present in a given condition [21,22,53,75–79]. Linkage among neighboring variants remains a key barrier for resolving loci and causal relationships. Power remains limited even in the genetically comprehensive datasets studied here. That most detected QTLs have small effects [25,26] makes it likely that large numbers of undetected eQTLs and gQTLs remain, each of very small effect. Imperfect power and linkage combine to produce wide QTL confidence intervals, which are difficult to use in colocalization analyses. Larger segregant panels are becoming available [80] but still suffer from linkage. Methods to directly engineer single variants in parallel [39,81–85] could be deployed in large samples and diverse conditions to ultimately achieve dissection of the complex interplay between gene expression and growth variation at gene-level and variant-level resolution.

In summary, we revealed a prominent role of *trans*-eQTL hotspots in shaping genetically complex traits. The *trans*-eQTL hotspots were detected in a baseline condition that was different from the tested growth conditions and they likely arise from genetic variants that alter critical cell states. These states serve as points of convergence for diverse genetic variants across the genome. When the cell is confronted with a non-baseline environment, this altered cell state provides a genetic predisposition that can help or hinder growth, depending on the specific biology of the trait.

Acknowledgements

We thank Joshua Bloom, Randi Avery, Sheila Lutz, Kelsey Johnson, Mahlon Collins, and Emma Sol Taylor-Brill for discussions and comments on the manuscript. We thank Qiuming Wu for preliminary downsampling analyses.

Funding

This work was funded by NIH grant R35GM124676 to FWA.

Materials & Methods

All analysis code is available at https://github.com/renga011/expressionGrowthRelationship.git. All analyses were conducted in R version 4.3.1 or R version 3.5.2 (https://www.r-project.org/).

Summary of data sets

The eQTL and gQTL data analyzed in this study were gathered in a panel of 1,012 *S.cerevisiae* meiotic segregants generated by crossing the "BY" prototrophic laboratory strain (a close relative of the genome reference strain S288C; MAT**a**) and the "RM" vineyard strain (*MATα hoΔ::hphMX4 flo8Δ::natMX4 AMN1-BY*). Each of the segregants had been genotyped at 42,052 markers. Markers in perfect linkage had been collapsed into a set of 11,530 unique markers [25], which we used for the analyses in this paper unless otherwise specified.

Bloom et al. measured growth of the segregants in 46 conditions [26]. Segregants were pinned to agar plates containing 50 ml of a solid base medium (YNB or YPD) with the given drug or condition (Table S1), at a density of 384 segregants per plate. After 48 hours of growth, colony size was measured by plate imaging. We used these end-point colony sizes (from the file "BYxRM_PhenoData.txt" at http://genomics-pubs.princeton.edu/YeastCross_BYxRM/) for individual segregants as relative measures of yeast growth. We sometimes refer to these growth measurements simply as "traits". Using these traits and the segregant genotypes, Bloom et al. mapped a total of 591 gQTLs across 46 conditions. gQTLs analyzed in the present work were obtained from Supplementary Table 3 in Bloom et al..

Albert, Bloom et al. grew each of the segregants in liquid YNB medium in 96-well plates, with a different arrangement of segregants into plates than in Bloom et al. RNA was extracted at an average optical density at 600 nm (OD_{600}) of ~0.4. Subsequently, they performed RNA sequencing to quantify the expression of 5,720 genes as the log2(transcripts per million + 0.5). We used these expression values (obtained from

Source Data 1 in Albert et al.: https://doi.org/10.7554/eLife.35471.021) as the gene expression measurements in the current paper.

Albert et al. used these data to map 36,498 eQTLs across 5,643 genes at 5% FDR. An eQTL was defined as local if its confidence interval overlapped the position of the gene it affected, where the gene position was considered to comprise the coding sequence plus 1,000 bases of upstream and 200 bases of downstream sequence. Using these criteria, 2,884 genes had a local eQTL. The number of eQTLs across the 5,643 genes ranged from 1 to 22, with a median of 6 eQTLs per gene. Except for a single local eQTL per gene (where present), these eQTLs all act in *trans*, comprising the vast majority of the dataset. eQTLs analyzed in the present work were obtained from Source Data 3 in Albert et al.

Data preparation

We analyzed data from 979 segregants that had both gene expression measurements and growth measurements in at least one condition. We corrected gene expression for the batch (i.e., the 96-well plate a segregant was grown in) and growth covariates (OD600 at the time of RNA extraction) reported in Albert et al., using the following linear model for each gene:

$$E = \alpha B + \beta G + R_{expression}$$

Here, E is a vector of uncorrected gene expression measurements across the segregants. B and G are vectors of segregant batch IDs (modeled as a factor) and OD_{600} measures (modeled as a numerical covariate), respectively. α and β are the estimated fixed-effect coefficients for the batch and growth covariates, respectively. $R_{expression}$ is the vector of gene expression residuals after correction for batch and growth covariates. We used $R_{expression}$ in all downstream analyses. No corrections were applied to growth measurements. We standardized the growth and corrected gene expression

measures to have a mean of 0 and standard deviation of 1, using the 'preProcess' function in the caret R package [86].

Genetic correlations

We computed the genetic correlation between the expression of a given gene and growth in a given condition as the Pearson correlation coefficient across segregants. While correlations between phenotypes (such as, in our case, gene expression and colony size) can in principle be caused by shared environmental factors in addition to shared alleles, the numerous experimental differences between the two studies analyzed here (see above) make environmental sources of these correlations extremely unlikely. We therefore consider the correlations computed here as genetic in origin. We estimated the FDR using the q-value package [87,88].

Correction for growth on YNB or YPD base medium

We removed the effects of the base medium *b* on growth in a given condition *i* using the following formula:

$$G_i = \alpha G_b + R_{growth,i}$$

Here, G_i is the vector of growth in condition *i*. G_b is the vector of growth in the given base medium *b*, where *b* can be either YNB or YPD, depending on which base medium the given condition had been added to. *α* is the estimated coefficient for the effect of the base medium on growth in the given condition. $R_{growth,i}$ is the vector of residuals, containing the growth in the given condition after correcting for the growth in the base medium. Subsequently, we re-computed the genetic correlation between the expression of each gene with the vector $R_{growth,i}$.

Shared factors across growth traits

To gauge the strength of shared factors underlying growth in the 46 conditions, we performed a principal component analysis on the matrix of growth traits. We used only the 42 / 46 growth traits with measurements for at least 20% of total segregants; the excluded growth conditions were cadmium chloride, hydrogen peroxide, raffinose and sorbitol. This analysis considered 457 / 979 segregants with growth measurements for all of these 42 conditions.

Effect of segregant panel size

To investigate how the size of the segregant panel affects the number and magnitude of detected significant genetic correlations, we randomly subsampled without replacement from the total 979 segregants to create subsets of 250, 500, and 750 segregants, with 5 random subsets for each sample size. For each subsampled segregant panel, we computed genetic correlations as above. For each subset size, we recorded the median number and magnitude of significant (at 5% FDR) genetic correlations for each of the 46 conditions across the five random subsets.

Colocalization tests

We performed colocalization tests for experimental crosses as implemented in Boehm et al[47]. Their test conducts two-dimensional scans for models with a single pleiotropic QTL or two distinct QTLs in a given interval known to contain QTLs for two traits. Here, these two traits were growth in a given condition and the expression of a given gene that is physically located in the same region (i.e., we analyzed gQTLs that overlapped local eQTLs). The test identifies the QTL positions that maximize the likelihoods under the assumption that the two traits are influenced by (a) a single pleiotropic QTL that arises from shared causal variants, or (b) two distinct QTLs arising from different causal variants. The log-ratio of the likelihoods of these two models is the test statistic, whose statistical significance is determined by parametric bootstrap. A significant p-value in this test rejects the null hypothesis of one pleiotropic QTL in favor of two distinct QTLs for the two traits. Therefore, a p-value greater than 0.05 implies pleiotropy in that a

model of shared variants could not be rejected. We ran these colocalization tests using the package *qtl2pleio* [47].

We applied the test to instances of local eQTLs overlapping a gQTL. QTL overlap was judged based on the genomic confidence intervals of the QTLs, extended by 5 kb on each side. The colocalization test is better able to detect distinct QTLs when the QTLs have larger effects. Therefore, we focused our analyses on cases where the overlapping local eQTLs and gQTLs both have a LOD score ≥ 10, to increase the chance that a failure to reject the null hypothesis of pleiotropy indeed suggests pleiotropy rather than an underpowered test.

Boehm et al. 's test has high power to detect two traits as having distinct QTLs when their peaks are separated by more than 1 centiMorgan (cM), which in our data corresponds to a correlation of neighboring marker genotypes of $r < 0.98$. To preserve power while reducing run time, we performed our colocalization tests with a reduced set of 1,496 markers for which consecutive marker genotypes had $r < 0.95$. We expected this reduced marker set to decrease the power to detect distinct QTLs, perhaps increasing the number of “false pleiotropy” assignments. To examine the severity of this bias, we re-tested an independent set of 122 QTLs that had all been called to be pleiotropic based on the reduced marker set ($r < 0.95$) with a denser set of 3,025 markers with $r < 0.98$. Even with this denser set of more highly correlated markers, pleiotropy was rejected at only 5 of these 122 QTLs (~4%). Given the majority of pleiotropy calls were thus the same when using the set of markers with $r < 0.95$ and those with $r < 0.98$, we chose to run our colocalization tests with the set of genomic markers at $r < 0.95$ to save computational time.

In the colocalization tests, we considered the smallest genomic interval that contained the confidence intervals for both the local eQTL and the gQTL, padded by 5 kb. For p-value estimation, we ran 1,000 parametric bootstraps. We considered two QTLs to be pleiotropic when the null hypothesis of pleiotropy was not rejected ($p > 0.05$) and when there were at least three genomic markers in the examined interval. We did not perform multiple-test correction for these analyses because our goal was to identify QTLs that are pleiotropic (i.e., those for which the test is not significant). A liberal criterion for

626 rejecting pleiotropy (without correcting for multiple testing) is conservative for this
627 purpose.

628

629 **QTL effect correlations**

630 We estimated the magnitudes of genetic effects on expression and growth at eQTLs
631 detected in Albert et al. as follows. For each gene, we extracted the peak markers of its
632 detected eQTLs, as reported in Source data 4 (https://doi.org/10.7554/eLife.35471.024)
633 from Albert et al. At each of these peak markers, we obtained QTL effects on gene
634 expression by computing the Pearson correlation coefficient between the expression of
635 the gene in the segregants and their genotypes at the peak marker. Similarly, at the
636 same markers, we obtained QTL effects on growth by computing Pearson correlation
637 coefficients with growth in each of the 46 conditions.

638 We then computed the weighted Pearson correlation between these QTL effects on
639 gene expression and QTL effects on growth, for each of the 46 conditions as follows.
640 For computing the weights, we first calculated the widths of the 95% confidence
641 intervals around the expression and growth effects using the *cor.test* function in R and
642 then took the inverse of their product. We subsequently passed these weights to the
643 *wtd.cor* function in the *weights* R package
644 (https://CRAN.R-project.org/package=weights) to get the weighted Pearson correlation
645 coefficients between the QTL effects on expression and growth. We estimated the FDR
646 using the "fdr" method in the *p.adjust* function , which implements the Benjamini &
647 Hochberg correction [89].

648

649 **Overlapping *trans*-eQTL hotspots and gQTLs**

650 We counted the number of overlaps between the confidence intervals of the gQTLs for
651 a given condition with the confidence intervals of the 102 *trans*-eQTL hotspots identified
652 in Albert et al., as reported in their Source Data 8
653 (https://doi.org/10.7554/eLife.35471.028). For each condition, we computed the random
654 expectation of these values as follows.

For a condition with *N* gQTLs, we selected *N* random non-overlapping genomic intervals of the same sizes as the actual gQTLs. Subsequently, we counted overlaps between these random gQTLs with the observed (i.e., non-randomized) hotspots. We repeated this 1,000 times per condition to obtain a trait-specific distribution of the expected number of gQTLs that overlap at least one *trans*-eQTL hotspot by chance. Using this distribution, we calculated p-values as the fraction of random sets that matched or exceeded the observed number of gQTL / hotspot overlaps. We considered an uncorrected $p < 0.05$ to be significant in this analysis.

Heritability estimates

For each of the 46 conditions, we computed the proportion of growth variance explained by a given set of genomic markers using the following linear mixed model, fit using the *lme4qtl* package [90]:

$$Growth_i \sim (1 \mid G_n)$$

Here, '$Growth_i$' is the growth of 979 segregants in condition *i*, and G_n is the genetic relatedness matrix generated from the given set of *n* genomic markers. We then estimated the fraction of growth variance explained by G_n. We implemented this model for the following sets of genomic markers: (a) all 11,530 genetic markers with between-genotype correlations less than 1; (b) 102 markers corresponding to the peak markers of the *trans*-eQTL hotspots from Albert et al.; (c) for each condition, the markers corresponding to the peak markers of their respective gQTLs. To estimate the heritability captured by sets of 102 random genomic markers by chance, we fit the above model at 1,000 sets of 102 random markers from the total set of 11,530 markers, for each of the 46 conditions.

For the analyses in Supplementary Figure 8, we ordered the 102 hotspots by the number of genes they affect, from many to few affected genes. We then fit the mixed linear models above on genetic relatedness matrices estimated from hotspots of

increasing size, starting with the two hotspots that affected most genes, and up to all 102 hotspots.

Hotspot effect correlations

We computed the effects of the 102 *trans*-eQTL hotspots from Albert et al. on expression of each of 5,720 genes and growth in 46 conditions from Bloom et al. using the following stepwise approach.

For each gene or growth trait, we computed the correlation coefficients between gene expression or growth and the genotype at each of the 102 hotspot peak markers. We retained the strongest significant correlation among the set with a correlation p-value < 0.05, and residualized the gene expression or growth measures to remove the effect of this strongest marker. We then used the residualized gene expression or growth values to calculate the correlation coefficients with each of the remaining 101 markers. We repeated this procedure until there were no more markers with correlation $p < 0.05$, and set the remaining effect estimates to zero. Because the hotspot locations had been previously identified using a rigorous detection process Albert et al., our procedure here did not attempt to correct for false discoveries. Instead, our goal was to assign effect estimates to any hotspot markers with a non-zero effect, based on a relaxed significance criterion.

Subsequently, we computed the hotspot effect correlation for a given gene / trait pair as the Pearson correlation coefficient between the effects of the hotspot markers on the expression of the gene and on the growth in the given growth condition. We estimated the FDR using the “fdr” method in the *p.adjust* function.

To estimate the distribution of hotspot effect correlations expected by chance, we used 1,000 sets of 102 random genetic markers as in the ‘Heritability estimates’ section and repeated the procedure above.

Quantifying the contribution of local eQTLs, *trans*-eQTLs and *trans*-eQTL hotspots to genetic correlations

713 To quantify the contribution of different sets of eQTLs to the genetic correlations, we
714 regressed out the effects of the respective loci from the expression and growth data
715 using the following linear models.

716

717 $$G_i = \sum_{k=1}^{n} \alpha_k * Genotype_k + R_{growth,\, i}$$

718 $$E_j = \sum_{k=1}^{n} \beta_k * Genotype_k + R_{expression,\, j}$$

719

720 G_i and E_j are the vectors of standardized growth measurements in condition *i* and the
721 standardized expression measurements for gene *j*, respectively. *Genotype* represents
722 the vectors of genotypes at *n* loci. *n* is either the number of local eQTLs for gene *j*
723 (usually here *n* = 1), the number of *trans*-eQTLs for gene *j*, or *n* = 102 for the
724 *trans*-eQTL hotspots. *α* and *β* are the estimated fixed effect coefficients for the
725 corresponding genotypes. $R_{growth,i}$ and $R_{expression,j}$ are the vectors of the residuals,
726 containing the growth and expression values for the given condition and gene
727 respectively after regressing out the effects of the given marker set.

728 We then used $R_{growth,i}$ and $R_{expression,j}$ to re-compute the genetic correlation coefficients.
729 Genetic correlation coefficients for all 5,643 genes and 46 conditions before and after
730 regressing out locus effects were compared using paired Wilcoxon tests.

731

732 **GO enrichment analyses and trait clustering**

733 For each of the 46 conditions, we performed GO enrichment analysis on the genes with
734 significant genetic correlations at 5% FDR using the following approach. First, we
735 divided the genes with significant genetic correlations into two groups based on the
736 directions of the correlations ('*positive*': r > 0, '*negative*': r < 0). We then performed
737 Fisher's exact tests on these gene lists to test for significant enrichment for 100 GO
738 Slim- biological process (BP) terms [91]. We used either the "positive" or the "negative"
739 enrichment in downstream analyses, based on the significance of the GO-term at p <

740 0.001. When a GO-term was significant in both the “positive” and “negative” enrichment
741 analyses, we retained the GO-term in the analysis with the greater magnitude of log-fold
742 enrichment.

743 We clustered the different growth conditions into trait groups using the 100 GO-BP
744 terms. For each condition, we calculated the fold-enrichment of each of the 100 GO-BP
745 terms as the ratio of the number of genes in our gene list annotated to the given GO
746 term and the number of genes expected to be observed for this GO term. We clustered
747 the traits based on fold-enrichments for GO-BP terms that were significant at nominal p
748 < 0.001. We performed k-means clustering using the *ComplexHeatmap* package [92] and
749 chose to cluster the traits into 3 groups based on visual inspection of the patterns of
750 enrichments across all GO term / trait pairs (Figure 6).

751

752 **Comparison with published datasets**

753 We compared our genetic correlations with results from two prior gene expression
754 studies. First, Brauer et al. studied the relationship between gene expression and
755 growth rate in chemostat cultures [50]. They identified three groups of genes based on
756 their expression response to different growth rates: genes whose expression is
757 upregulated in fast growth (n = 258), genes upregulated in slow growth (n = 367), and
758 genes that are unresponsive to growth rate (n = 980). For each trait, we computed the
759 average genetic correlation coefficients in our analyses for the genes belonging to these
760 three groups and plotted them along with the GO term fold enrichment signatures for all
761 46 growth traits. We similarly computed the average genetic correlation coefficients for
762 the groups of genes that were found in Gasch et al.[51] to be upregulated (n = 269) and
763 downregulated (n = 570) during the environmental stress response, and plotted them
764 along with the GO term fold enrichment signatures.

765

766 **Mediation**

767 We performed mediation analyses using the *mediation* package [93] for each of the 1,240
768 genes regulated by *IRA2* hotspot (as reported in Source Data 9 in Albert et al.:

https://doi.org/10.7554/eLife.35471.029), using the following rationale. If the *IRA2* hotspot's effect on growth in the presence of hydrogen peroxide is partitioned into the effect due to the expression level of a given gene *i*, and the effect independent of the expression of Gene *i* (Figure 7A), we can use the following system of linear models:

$$Gene_i = \alpha * Genotype_{IRA2} + e1$$
$$Growth_{H2O2} = \beta * Genotype_{IRA2} + \gamma * Gene_i + e2$$

Here, $Gene_i$ is the vector containing the expression values of gene *i* for 979 segregants. $Genotype_{IRA2}$ is the vector containing the genotype of 979 segregants at the peak marker of the *IRA2* hotspot. $Growth_{H2O2}$ is the vector containing the colony growth measurements of 979 segregants in presence of hydrogen peroxide. α is the estimated effect of the *IRA2* hotspot on the expression of $Gene_i$. *β* is the estimated direct effect of the *IRA2* hotspot on growth in presence of hydrogen peroxide that is independent of the expression level of the given gene, and *γ* is the estimated effect of the expression level of $Gene_i$ on growth in hydrogen peroxide. *e1* and *e2* are the residual errors. From these equations, the growth as a function of the genotype at the *IRA2* hotspot marker can be expressed as follows.

$$Growth_{H2O2} = \beta * Genotype_{IRA2} + \gamma * \alpha * Genotype_{IRA2} + e2 + e1$$
$$Growth_{H2O2} = (\beta + \alpha * \gamma) * Genotype_{IRA2} + e2 + e1$$

Hence, the estimated total effect of the genotype at the *IRA2* hotspot marker on growth in the presence of hydrogen peroxide is $\beta + \alpha*\gamma$. This includes the estimated direct effect of the genotype, β, which is independent of the expression of $Gene_i$, and the effect $\alpha*\gamma$, which is mediated by the expression level of $Gene_i$. The proportion of the hotspot's effect mediated by the expression of the gene is therefore the ratio of the mediated and total effect.

$$Proportion\ mediated = \frac{Estimated\ mediated\ effect}{Estimated\ total\ effect} = \frac{\alpha^{*}\gamma}{\beta + \alpha^{*}\gamma}$$

We calculated the p-value for the presence of a mediation effect of the gene's expression on growth using 1,000 non-parametric bootstraps, the default number of simulations in the package. The FDR was estimated using the 'fdr' method in the *p.adjust* function.

References

1. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science *337*, 1190–1195. https://doi.org/10.1126/science.1222794.
2. Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. *16*, 197–212. https://doi.org/10.1038/nrg3891.
3. Aguirre-Gamboa, R., de Klein, N., di Tommaso, J., Claringbould, A., van der Wijst, M.G., de Vries, D., Brugge, H., Oelen, R., Võsa, U., Zorro, M.M., et al. (2020). Deconvolution of bulk blood eQTL effects into immune cell subpopulations. BMC Bioinformatics *21*, 243. https://doi.org/10.1186/s12859-020-03576-5.
4. Brown, A.A., Viñuela, A., Delaneau, O., Spector, T.D., Small, K.S., and Dermitzakis, E.T. (2017). Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. Nat. Genet. *49*, 1747–1751. https://doi.org/10.1038/ng.3979.
5. Cusanovich, D.A., Billstrand, C., Zhou, X., Chavarria, C., De Leon, S., Michelini, K., Pai, A.A., Ober, C., and Gilad, Y. (2012). The combination of a genome-wide association study of lymphocyte count and analysis of gene expression data reveals novel asthma candidate genes. Hum. Mol. Genet. *21*, 2111–2123. https://doi.org/10.1093/hmg/dds021.
6. Donovan, M.K.R., D'Antonio-Chronowska, A., D'Antonio, M., and Frazer, K.A. (2020). Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. Nat. Commun. *11*, 955. https://doi.org/10.1038/s41467-020-14561-0.
7. Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al. (2020). Cell type–specific genetic regulation of gene expression across human tissues. Science *369*, eaaz8528. https://doi.org/10.1126/science.aaz8528.
8. Lin, H., Wang, F., Rosato, A.J., Farrer, L.A., Henderson, D.C., and Zhang, H. (2020). Prefrontal cortex eQTLs/mQTLs enriched in genetic variants associated with alcohol use disorder and other diseases. Epigenomics *12*, 789–800. https://doi.org/10.2217/epi-2019-0270.
9. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLOS Genet. *6*, e1000888. https://doi.org/10.1371/journal.pgen.1000888.
10. O'Brien, H.E., Hannon, E., Hill, M.J., Toste, C.C., Robertson, M.J., Morgan, J.E., McLaughlin, G., Lewis, C.M., Schalkwyk, L.C., Hall, L.S., et al. (2018). Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. Genome Biol. *19*, 194. https://doi.org/10.1186/s13059-018-1567-1.
11. Patel, D., Zhang, X., Farrell, J.J., Chung, J., Stein, T.D., Lunetta, K.L., and Farrer, L.A. (2021). Cell-type-specific expression quantitative trait loci associated with Alzheimer disease in blood and brain tissue. Transl. Psychiatry *11*, 1–17. https://doi.org/10.1038/s41398-021-01373-z.
12. The GTEx Consortium, Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G.A., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., et al. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
13. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.

https://doi.org/10.1038/ng.3538.
14. Claussnitzer, M., Dankel, S.N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N. Engl. J. Med. *373*, 895–907. https://doi.org/10.1056/NEJMoa1502214.
15. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M., et al. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature *466*, 714–719. https://doi.org/10.1038/nature09266.
16. Smemo, S., Tena, J.J., Kim, K.-H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–375. https://doi.org/10.1038/nature13138.
17. Collins, M.A., Mekonnen, G., and Albert, F.W. (2022). Variation in ubiquitin system genes creates substrate-specific effects on proteasomal protein degradation. eLife *11*, e79570. https://doi.org/10.7554/eLife.79570.
18. Collins, M.A., Avery, R., and Albert, F.W. (2023). Substrate-specific effects of natural genetic variation on proteasome activity. PLOS Genet. *19*, e1010734. https://doi.org/10.1371/journal.pgen.1010734.
19. Gerke, J., Lorenz, K., and Cohen, B. (2009). Genetic Interactions Between Transcription Factors Cause Natural Variation in Yeast. Science *323*, 498–501. https://doi.org/10.1126/science.1166426.
20. Salinas, F., de Boer, C.G., Abarca, V., García, V., Cuevas, M., Araos, S., Larrondo, L.F., Martínez, C., and Cubillos, F.A. (2016). Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. Sci. Rep. *6*, 21849. https://doi.org/10.1038/srep21849.
21. Umans, B.D., Battle, A., and Gilad, Y. (2021). Where are the disease-associated eQTLs? Trends Genet. TIG *37*, 109–124. https://doi.org/10.1016/j.tig.2020.08.009.
22. Connally, N.J., Nazeen, S., Lee, D., Shi, H., Stamatoyannopoulos, J., Chun, S., Cotsapas, C., Cassa, C.A., and Sunyaev, S.R. (2022). The missing link between genetic association and regulatory function. eLife *11*, e74970. https://doi.org/10.7554/eLife.74970.
23. Yao, D.W., O'Connor, L.J., Price, A.L., and Gusev, A. (2020). Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. *52*, 626–633. https://doi.org/10.1038/s41588-020-0625-2.
24. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. Nat. Genet., 1–10. https://doi.org/10.1038/s41588-023-01529-1.
25. Albert, F.W., Bloom, J.S., Siegel, J., Day, L., and Kruglyak, L. (2018). Genetics of trans-regulatory variation in gene expression. eLife *7*, e35471. https://doi.org/10.7554/eLife.35471.
26. Bloom, J.S., Ehrenreich, I.M., Loo, W., Võ Lite, T.-L., and Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. Nature *494*, 234–237. https://doi.org/10.1038/nature11867.
27. Rockman, M.V., and Kruglyak, L. (2006). Genetics of global gene expression. Nat. Rev. Genet. *7*, 862–872. https://doi.org/10.1038/nrg1964.
28. Renganaath, K., Cheung, R., Day, L., Kosuri, S., Kruglyak, L., and Albert, F.W. (2020). Systematic identification of cis-regulatory variants that cause gene expression differences in a yeast cross. eLife *9*, e62669. https://doi.org/10.7554/eLife.62669.
29. Boocock, J., Alexander, N., Tapia, L.A., Walter-McNeill, L., Munugala, C., Bloom, J.S., and Kruglyak, L. (2024). Single-cell eQTL mapping in yeast reveals a tradeoff between growth and reproduction. Preprint, https://doi.org/10.7554/eLife.95566.1

https://doi.org/10.7554/eLife.95566.1.
30. Caudal, É., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonnière, É., Caradec, C., Friedrich, A., Hou, J., and Schacherer, J. (2024). Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. Nat. Genet., 1–10. https://doi.org/10.1038/s41588-024-01769-9.
31. Fu, J., Keurentjes, J.J.B., Bouwmeester, H., America, T., Verstappen, F.W.A., Ward, J.L., Beale, M.H., de Vos, R.C.H., Dijkstra, M., Scheltema, R.A., et al. (2009). System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat. Genet. *41*, 166–167. https://doi.org/10.1038/ng.308.
32. Rockman, M.V., Skrovanek, S.S., and Kruglyak, L. (2010). Selection at Linked Sites Shapes Heritable Phenotypic Variation in *C. elegans*. Science *330*, 372–376. https://doi.org/10.1126/science.1194208.
33. van Nas, A., Ingram-Drake, L., Sinsheimer, J.S., Wang, S.S., Schadt, E.E., Drake, T., and Lusis, A.J. (2010). Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. Genetics *185*, 1059–1068. https://doi.org/10.1534/genetics.110.116087.
34. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., and Knight, J.C. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. *44*, 502–510. https://doi.org/10.1038/ng.2205.
35. Brandt, M., Kim-Hellmuth, S., Ziosi, M., Gokden, A., Wolman, A., Lam, N., Recinos, Y., Daniloski, Z., Morris, J.A., Hornung, V., et al. (2021). An autoimmune disease risk variant: A trans master regulatory effect mediated by IRF1 under immune stimulation? PLoS Genet. *17*, e1009684. https://doi.org/10.1371/journal.pgen.1009684.
36. Small, K.S., Hedman, A.K., Grundberg, E., Nica, A.C., Thorleifsson, G., Kong, A., Thorsteindottir, U., Shin, S.-Y., Richards, H.B., GIANT 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. https://doi.org/10.1038/ng.833.
37. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet. *53*, 1300–1310. https://doi.org/10.1038/s41588-021-00913-z.
38. Gou, L., Bloom, J.S., and Kruglyak, L. (2019). The Genetic Basis of Mutation Rate Variation in Yeast. Genetics *211*, 731–740. https://doi.org/10.1534/genetics.118.301609.
39. Sadhu, M.J., Bloom, J.S., Day, L., and Kruglyak, L. (2016). CRISPR-directed mitotic recombination enables genetic mapping without crosses. Science *352*, 1113–1116. https://doi.org/10.1126/science.aaf5124.
40. Chick, J.M., Munger, S.C., Simecek, P., Huttlin, E.L., Choi, K., Gatti, D.M., Raghupathy, N., Svenson, K.L., Churchill, G.A., and Gygi, S.P. (2016). Defining the consequences of genetic variation on a proteome–wide scale. Nature *534*, 500–505. https://doi.org/10.1038/nature18270.
41. Heyne, H.O., Lautenschläger, S., Nelson, R., Besnier, F., Rotival, M., Cagan, A., Kozhemyakina, R., Plyusnina, I.Z., Trut, L., Carlborg, Ö., et al. (2014). Genetic Influences on Brain Gene Expression in Rats Selected for Tameness and Aggression. Genetics *198*, 1277–1290. https://doi.org/10.1534/genetics.114.168948.
42. Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. *45*, 43–50. https://doi.org/10.1038/ng.2484.
43. Munro, D., Wang, T., Chitre, A.S., Polesskaya, O., Ehsan, N., Gao, J., Gusev, A., Woods, L.C.S., Saba, L.M., Chen, H., et al. (2022). The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats. Nucleic Acids Res. *50*, 10882–10895. https://doi.org/10.1093/nar/gkac912.

44. Pang, J., Fu, J., Zong, N., Wang, J., Song, D., Zhang, X., He, C., Fang, T., Zhang, H., Fan, Y., et al. (2019). Kernel size-related genes revealed by an integrated eQTL analysis during early maize kernel development. Plant J. *98*, 19–32. https://doi.org/10.1111/tpj.14193.
45. Zhang, G., Roberto, N.M., Lee, D., Hahnel, S.R., and Andersen, E.C. (2022). The impact of species-wide gene expression variation on Caenorhabditis elegans complex traits. Nat. Commun. *13*, 3462. https://doi.org/10.1038/s41467-022-31208-4.
46. Falconer, D.S., and Mackay, T. (2009). Introduction to quantitative genetics 4. ed., [16. print.]. (Pearson, Prentice Hall).
47. Boehm, F.J., Chesler, E.J., Yandell, B.S., and Broman, K.W. (2019). Testing Pleiotropy vs. Separate QTL in Multiparental Populations. G3 GenesGenomesGenetics *9*, 2317–2324. https://doi.org/10.1534/g3.119.400098.
48. Kapitzky, L., Beltrao, P., Berens, T.J., Gassner, N., Zhou, C., Wüster, A., Wu, J., Babu, M.M., Elledge, S.J., Toczyski, D., et al. (2010). Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action. Mol. Syst. Biol. *6*, 451. https://doi.org/10.1038/msb.2010.107.
49. He, X., Fuller, C.K., Song, Y., Meng, Q., Zhang, B., Yang, X., and Li, H. (2013). Sherlock: Detecting Gene-Disease Associations by Matching Patterns of Expression QTL and GWAS. Am. J. Hum. Genet. *92*, 667–680. https://doi.org/10.1016/j.ajhg.2013.03.022.
50. Brauer, M.J., Huttenhower, C., Airoldi, E.M., Rosenstein, R., Matese, J.C., Gresham, D., Boer, V.M., Troyanskaya, O.G., and Botstein, D. (2008). Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast. Mol. Biol. Cell *19*, 352–367. https://doi.org/10.1091/mbc.e07-08-0779.
51. Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Mol. Biol. Cell *11*, 4241–4257. https://doi.org/10.1091/mbc.11.12.4241.
52. Lutz, S., Van Dyke, K., Feraru, M.A., and Albert, F.W. (2022). Multiple epistatic DNA variants in a single gene affect gene expression in *trans*. Genetics *220*, iyab208. https://doi.org/10.1093/genetics/iyab208.
53. Smith, E.N., and Kruglyak, L. (2008). Gene–Environment Interaction in Yeast Gene Expression. PLOS Biol. *6*, e83. https://doi.org/10.1371/journal.pbio.0060083.
54. Hasan, R., Leroy, C., Isnard, A.-D., Labarre, J., Boy-Marcotte, E., and Toledano, M.B. (2002). The control of the yeast H2O2 response by the Msn2/4 transcription factors. Mol. Microbiol. *45*, 233–241. https://doi.org/10.1046/j.1365-2958.2002.03011.x.
55. Sadeh, A., Movshovich, N., Volokh, M., Gheber, L., and Aharoni, A. (2011). Fine-tuning of the Msn2/4–mediated yeast stress responses as revealed by systematic deletion of Msn2/4 partners. Mol. Biol. Cell *22*, 3127–3138. https://doi.org/10.1091/mbc.E10-12-1007.
56. Görner, W., Durchschlag, E., Martinez-Pastor, M.T., Estruch, F., Ammerer, G., Hamilton, B., Ruis, H., and Schüller, C. (1998). Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. Genes Dev. *12*, 586–597. https://doi.org/10.1101/gad.12.4.586.
57. Chapal, M., Mintzer, S., Brodsky, S., Carmi, M., and Barkai, N. (2019). Resolving noise–control conflict by gene duplication. PLOS Biol. *17*, e3000289. https://doi.org/10.1371/journal.pbio.3000289.
58. Fay, J.C. (2013). The molecular basis of phenotypic variation in yeast. Curr. Opin. Genet. Dev. *23*, 672–677. https://doi.org/10.1016/j.gde.2013.10.005.
59. Dendrou, C.A., Cortes, A., Shipman, L., Evans, H.G., Attfield, K.E., Jostins, L., Barber, T., Kaur, G., Kuttikkatte, S.B., Leach, O.A., et al. (2016). Resolving TYK2 locus genotype-to-phenotype differences in autoimmunity. Sci. Transl. Med. *8*, 363ra149. https://doi.org/10.1126/scitranslmed.aag1974.
60. Plenge, R.M., Scolnick, E.M., and Altshuler, D. (2013). Validating therapeutic targets

through human genetics. Nat. Rev. Drug Discov. *12*, 581–594. https://doi.org/10.1038/nrd4051.

61. Arita, Y., Kim, G., Li, Z., Friesen, H., Turco, G., Wang, R.Y., Climie, D., Usaj, M., Hotz, M., Stoops, E.H., et al. (2021). A genome-scale yeast library with inducible expression of individual genes. Mol. Syst. Biol. *17*, e10207. https://doi.org/10.15252/msb.202110207.
62. Duveau, F., Hodgins-Davis, A., Metzger, B.P., Yang, B., Tryban, S., Walker, E.A., Lybrook, T., and Wittkopp, P.J. (2018). Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. eLife *7*, e37272. https://doi.org/10.7554/eLife.37272.
63. Keren, L., Hausser, J., Lotan-Pompan, M., Slutskin, I.V., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2016). Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. Cell *166*, 1282-1294.e18. https://doi.org/10.1016/j.cell.2016.07.024.
64. Metzger, B.P.H., Yuan, D.C., Gruber, J.D., Duveau, F., and Wittkopp, P.J. (2015). Selection on noise constrains variation in a eukaryotic promoter. Nature *521*, 344–347. https://doi.org/10.1038/nature14244.
65. Freimer, J.W., Shaked, O., Naqvi, S., Sinnott-Armstrong, N., Kathiria, A., Garrido, C.M., Chen, A.F., Cortez, J.T., Greenleaf, W.J., Pritchard, J.K., et al. (2022). Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. Nat. Genet. *54*, 1133–1144. https://doi.org/10.1038/s41588-022-01106-y.
66. Naqvi, S., Kim, S., Hoskens, H., Matthews, H.S., Spritz, R.A., Klein, O.D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J.K., et al. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. Nat. Genet. *55*, 841–851. https://doi.org/10.1038/s41588-023-01366-2.
67. Weith, M., Großbach, J., Clement-Ziza, M., Gillet, L., Rodríguez-López, M., Marguerat, S., Workman, C.T., Picotti, P., Bähler, J., Aebersold, R., et al. (2023). Genetic effects on molecular network states explain complex traits. Mol. Syst. Biol. *19*, e11493. https://doi.org/10.15252/msb.202211493.
68. Berry, D.B., Guan, Q., Hose, J., Haroon, S., Gebbia, M., Heisler, L.E., Nislow, C., Giaever, G., and Gasch, A.P. (2011). Multiple Means to the Same End: The Genetic Basis of Acquired Stress Resistance in Yeast. PLOS Genet. *7*, e1002353. https://doi.org/10.1371/journal.pgen.1002353.
69. Berry, D.B., and Gasch, A.P. (2008). Stress-activated Genomic Expression Changes Serve a Preparative Role for Impending Stress in Yeast. Mol. Biol. Cell *19*, 4580–4587. https://doi.org/10.1091/mbc.e07-07-0680.
70. Guan, Q., Haroon, S., Bravo, D.G., Will, J.L., and Gasch, A.P. (2012). Cellular Memory of Acquired Stress Resistance in Saccharomyces cerevisiae. Genetics *192*, 495–505. https://doi.org/10.1534/genetics.112.143016.
71. Morano, K.A., Grant, C.M., and Moye-Rowley, W.S. (2012). The Response to Heat Shock and Oxidative Stress in Saccharomyces cerevisiae. Genetics *190*, 1157–1195. https://doi.org/10.1534/genetics.111.128033.
72. Zakrzewska, A., van Eikenhorst, G., Burggraaff, J.E.C., Vis, D.J., Hoefsloot, H., Delneri, D., Oliver, S.G., Brul, S., and Smits, G.J. (2011). Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. Mol. Biol. Cell *22*, 4435–4446. https://doi.org/10.1091/mbc.E10-08-0721.
73. Lutz, S., Brion, C., Kliebhan, M., and Albert, F.W. (2019). DNA variants affecting the expression of numerous genes in trans have diverse mechanisms of action and evolutionary histories. PLOS Genet. *15*, e1008375. https://doi.org/10.1371/journal.pgen.1008375.
74. Potokina, E., Druka, A., Luo, Z., Wise, R., Waugh, R., and Kearsey, M. (2008). Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. Plant J. *53*, 90–101. https://doi.org/10.1111/j.1365-313X.2007.03315.x.

75. Elorbany, R., Popp, J.M., Rhodes, K., Strober, B.J., Barr, K., Qi, G., Gilad, Y., and Battle, A. (2022). Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. PLoS Genet. *18*, e1009666. https://doi.org/10.1371/journal.pgen.1009666.
76. Kariuki, S.N., Maranville, J.C., Baxter, S.S., Jeong, C., Nakagome, S., Hrusch, C.L., Witonsky, D.B., Sperling, A.I., and Di Rienzo, A. (2016). Mapping Variation in Cellular and Transcriptional Response to 1,25-Dihydroxyvitamin D3 in Peripheral Blood Mononuclear Cells. PLoS ONE *11*, e0159779. https://doi.org/10.1371/journal.pone.0159779.
77. Smirnov, D.A., Morley, M., Shin, E., Spielman, R.S., and Cheung, V.G. (2009). Genetic analysis of radiation-induced changes in human gene expression. Nature *459*, 587–591. https://doi.org/10.1038/nature07940.
78. Strober, B.J., Elorbany, R., Rhodes, K., Krishnan, N., Tayeb, K., Battle, A., and Gilad, Y. (2019). Dynamic genetic regulation of gene expression during cellular differentiation. Science *364*, 1287–1290. https://doi.org/10.1126/science.aaw0040.
79. Taylor, D.L., Knowles, D.A., Scott, L.J., Ramirez, A.H., Casale, F.P., Wolford, B.N., Guan, L., Varshney, A., Albanus, R.D., Parker, S.C.J., et al. (2018). Interactions between genetic variation and cellular environment in skeletal muscle gene expression. PLoS ONE *13*, e0195788. https://doi.org/10.1371/journal.pone.0195788.
80. Nguyen Ba, A.N., Lawrence, K.R., Rego-Costa, A., Gopalakrishnan, S., Temko, D., Michor, F., and Desai, M.M. (2022). Barcoded bulk QTL mapping reveals highly polygenic and epistatic architecture of complex traits in yeast. eLife *11*, e73983. https://doi.org/10.7554/eLife.73983.
81. Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human variants with base editor screens. Cell *184*, 1064-1080.e20. https://doi.org/10.1016/j.cell.2021.01.012.
82. Lopez, S.C., Crawford, K.D., Lear, S.K., Bhattarai-Kline, S., and Shipman, S.L. (2022). Precise genome editing across kingdoms of life using retron-derived DNA. Nat. Chem. Biol. *18*, 199–206. https://doi.org/10.1038/s41589-021-00927-y.
83. Martin-Rufino, J.D., Castano, N., Pang, M., Grody, E.I., Joubran, S., Caulier, A., Wahlster, L., Li, T., Qiu, X., Riera-Escandell, A.M., et al. (2023). Massively parallel base editing to map variant effects in human hematopoiesis. Cell *186*, 2456-2474.e24. https://doi.org/10.1016/j.cell.2023.03.035.
84. Morris, J.A., Caragine, C., Daniloski, Z., Domingo, J., Barry, T., Lu, L., Davis, K., Ziosi, M., Glinos, D.A., Hao, S., et al. (2023). Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. Science *380*, eadh7699. https://doi.org/10.1126/science.adh7699.
85. Sharon, E., Chen, S.-A.A., Khosla, N.M., Smith, J.D., Pritchard, J.K., and Fraser, H.B. (2018). Functional genetic variants revealed by massively parallel precise genome editing. Cell *175*, 544-557.e16. https://doi.org/10.1016/j.cell.2018.08.057.
86. Kuhn, M. (2008). Building Predictive Models in *R* Using the **caret** Package. J. Stat. Softw. *28*. https://doi.org/10.18637/jss.v028.i05.
87. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. *100*, 9440–9445. https://doi.org/10.1073/pnas.1530509100.
88. Storey, John.D., Bass, Andrew. J., Dabney, Alan, and Robinson, David (2023). qvalue: Q-value estimation for false discovery rate control. Version 2.34.0 (Bioconductor). https://doi.org/10.18129/B9.BIOC.QVALUE https://doi.org/10.18129/B9.BIOC.QVALUE.
89. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
90. Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., and

Soria, J.M. (2018). lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC Bioinformatics *19*, 68. https://doi.org/10.1186/s12859-018-2057-x.

91. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, D700–D705. https://doi.org/10.1093/nar/gkr1029.
92. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.
93. Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). mediation: R Package for Causal Mediation Analysis. J. Stat. Softw. *59*, 1–38. https://doi.org/10.18637/jss.v059.i05.