DNA variants affecting the expression of numerous genes in *trans* have diverse mechanisms of action and evolutionary histories

Sheila Lutz[1], Christian Brion[1], Margaret Kliebhan[1], Frank W. Albert[1,*]

[1]Department of Genetics, Cell Biology, & Development, University of Minnesota, Minneapolis, MN, USA

* Corresponding author

E-mail: falbert@umn.edu (FWA)

Abstract

DNA variants that alter gene expression contribute to variation in many phenotypic traits. In particular, *trans*-acting variants, which are often located on different chromosomes from the genes they affect, are an important source of heritable gene expression variation. However, our knowledge about the identity and mechanism of causal *trans*-acting variants remains limited. Here, we developed a fine-mapping strategy called CRISPR-Swap and dissected three expression quantitative trait locus (eQTL) hotspots known to alter the expression of numerous genes in *trans* in the yeast *Saccharomyces cerevisiae*. Causal variants were identified by engineering recombinant alleles and quantifying the effects of these alleles on the expression of a green fluorescent protein-tagged gene affected by the given locus in *trans*. We validated the effect of each variant on the expression of multiple genes by RNA-sequencing. The three variants were strikingly different in their molecular mechanism, the type of genes they reside in, and their distribution in natural populations. While a missense leucine-to-serine variant at position 63 in the transcription factor Oaf1 (L63S) was almost exclusively present in the reference laboratory strain, the two other variants were frequent among *S. cerevisiae* isolates. A causal missense variant in the glucose receptor Rgt2 (V539I) occurred at a poorly conserved amino acid residue and its effect was strongly dependent on the concentration of glucose in the culture medium. A noncoding variant in the conserved fatty acid regulated (FAR) element of the *OLE1* promoter influenced the

29 expression of the fatty acid desaturase Ole1 in *cis* and, by modulating the level of this essential
30 enzyme, other genes in *trans*. The *OAF1* and *OLE1* variants showed a non-additive genetic
31 interaction, and affected cellular lipid metabolism. These results revealed remarkable diversity in
32 the molecular basis of *trans*-regulatory variation, highlighting the challenges in predicting which
33 natural genetic variants affect gene expression.

Author summary

Differences in the DNA sequence of individual genomes contribute to differences in many traits, such as appearance, physiology, and the risk for common diseases. An important group of these DNA variants influences how individual genes across the genome are turned on or off. In this paper, we describe a strategy for identifying such "*trans*-acting" variants in different strains of baker's yeast. We used this strategy to reveal three single DNA base changes that each influences the expression of dozens of genes. These three DNA variants were very different from each other. Two of them changed the protein sequence, one in a transcription factor and the other in a sugar sensor. The third changed the expression of an enzyme, a change that in turn caused other genes to alter their expression. One variant existed in only a few yeast isolates, while the other two existed in many isolates collected from around the world. This diversity of DNA variants that influence the expression of many other genes illustrates how difficult it is to predict which DNA variants in an individual's genome will have effects on the organism.

Introduction

DNA variants that alter gene expression are an important source of genetic variation for many traits [1], including common disease in humans [2], agricultural yield [3,4] and evolutionary change [5]. To map gene expression variation in the genome, expression levels are measured in a population of individuals and related to the genotype of each individual. This approach identifies expression quantitative trait loci (eQTLs) – genomic regions that each contain one or more variants that affect gene expression.

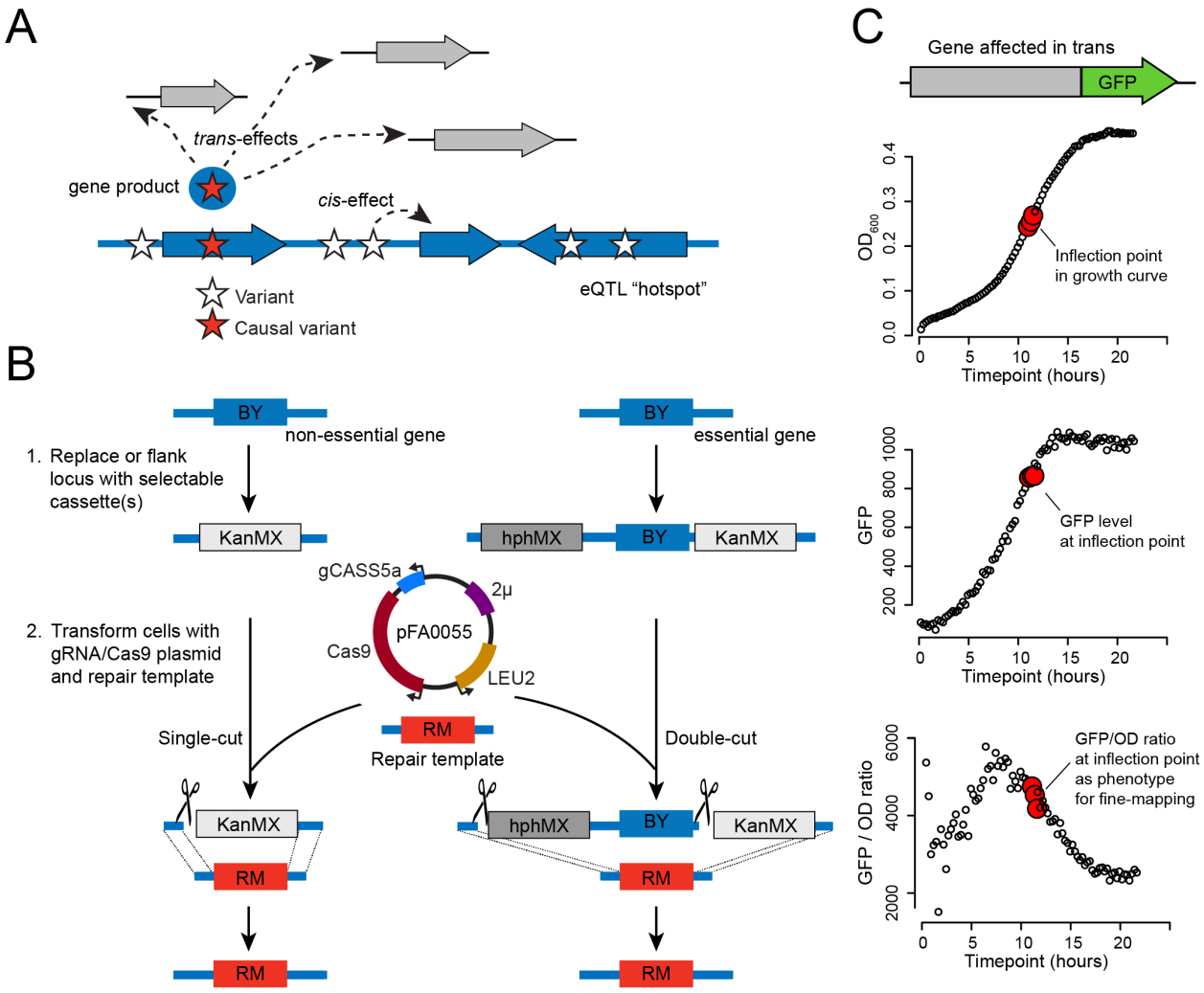
eQTLs can be classified into two types based on their mechanism of action. *Cis* eQTLs arise from variants that alter the expression of genes on the same DNA molecule, for example by changing the sequence of a regulatory element in a promoter. Most *cis*-acting variants are located close to or within the genes they influence, such that *cis* eQTLs can be detected as "local" eQTLs that overlap the given gene. By contrast, *trans* eQTLs arise from variants that change the activity and/or abundance of a diffusible factor which in turn alter the expression of other genes. *Trans* eQTLs can be located anywhere in the genome relative to the genes they affect. While they can be local (e.g., if a gene encoding a transcription factor resides next to a gene targeted by that factor), most *trans* eQTLs are "distant" from the genes they affect, usually on different chromosomes.

As in genetic mapping of other traits, identifying the specific DNA variants that have causal effects in eQTL regions is challenging. Recent studies have made progress in identifying *cis* acting variants (e.g. [6–10]). However, few *trans*-eQTLs have been resolved to single variants. This is although most of the heritable contribution to gene expression variation arises from *trans* rather than *cis* eQTLs [11–16], and although *trans* acting variation is likely to play pivotal roles in shaping diseases and phenotypes within species [11,17]. Identifying the molecular nature of *trans*-acting variants and the mechanisms by which they alter gene expression is key to understanding the connection between genotypic and phenotypic variation.

Natural isolates of the yeast *Saccharomyces cerevisiae* have provided fundamental insights into the genetics of gene expression variation ([12,18–34], reviewed in [35]). Particularly intensive efforts have been directed at the comparison between a laboratory strain (the genome reference strain S288C, or "BY") and a wine strain (RM11a, "RM"), whose genomes differ at about 40,000 variants. eQTL mapping in recombinant progeny from a cross between these strains revealed the existence of eQTL hotspots that each influence the expression of numerous genes in *trans* [18].

77 Many of these hotspots also affect protein levels [22,36,37]. Recently, an analysis of mRNA levels
78 in an expanded set of 1,012 BY/RM progeny provided a more comprehensive view of regulatory
79 variation in this cross [12]. Specifically, *trans*-acting variation arose almost exclusively from 102
80 hotspots, some of which affected the expression of thousands of genes (Fig 1A). A small number
81 of hotspots in this and other crosses have been resolved to their causal genes and nucleotide
82 variants [18,20,23,30,31,34,38–41], but progress towards a more complete view of hotspot
83 variants has been hampered by the challenges of engineering and measuring the expression effects
84 of candidate variants.

85



86

Fig 1. Hotspot fine mapping strategy. A. Illustration of the possible *cis* and *trans* effects at an eQTL hotspot. A genomic region, shown in blue with several genes depicted as wide arrows, harbors multiple variants (stars). Of these, one causal variant (red star) alters the activity and/or abundance of a gene product (blue circle with red star), which alters the expression of multiple genes (gray arrows) in *trans*. Another variant is shown affecting a neighboring gene in *cis*, but has no *trans* effect. B. A schematic showing two examples of engineering a BY allele (blue) to an RM allele (red) using CRISPR-Swap. In step one, a non-essential gene is replaced by the G418 resistance cassette (KanMX) or an essential gene is flanked by the hygromycin resistance (hphMX) and KanMX cassettes. In step two, the strain is transformed with the CRISPR-Swap plasmid pFA0055 that expresses Cas9 and the guide RNA gCASS5a and the auxotrophic marker LEU2; and a PCR-generated repair template containing the desired RM allele. The gCASS5a/Cas9 complex directs cleavage (scissors) of the cassettes. Either a single-cut or a double-cut occurs depending on the number of cassettes present. Selection of transformants for leucine prototrophy and loss of G418, or hygromycin and G418, resistance identifies strains with the desired RM allele replacement. C. Quantifying the expression of a representative gene affected in *trans* by a hotspot. Fluorescence of the protein expressed from the GFP-tagged gene and optical density (OD_{600}) of the culture are measured in 15 minute intervals during growth in a plate reader. The inflection point, the point where the culture exits logarithmic growth, and two flanking points are used to determine the GFP/OD_{600} ratio for phenotyping the effect of the engineered alleles.

Here, we describe a strategy for the identification of causal eQTL hotspot variants that combines a genome engineering approach with high-throughput quantification of fluorescently tagged protein expression as a phenotypic readout. We used this strategy to identify causal variants underlying three *trans*-acting eQTL hotspots in the BY and RM strains: a common missense variant in the glucose receptor Rgt2, a rare missense variant in the oleate-activated transcription factor Oaf1, and a common variant in the promoter of *OLE1* that alters the expression of this essential fatty acid desaturase gene. We studied the effects of these variants in more detail and discovered that the effect of the *RGT2* variant is influenced by the environment, and that the *OAF1* and *OLE1* variants interact in a non-additive fashion and lead to changes in cellular lipids. These results show that variants underlying *trans*-acting hotspots are highly diverse. They can be common or rare in the population, different in their evolutionary conservation, and located in the coding or noncoding region of genes encoding functionally diverse proteins.

Results

The CRISPR-Swap strategy for engineering allelic series

To assist fine-mapping of hotspot intervals, we devised "CRISPR-Swap" (Fig 1B), a two-step strategy for efficient allele exchange that combines advantages of "insert-then-replace" methods [42,43] with CRISPR/Cas9 engineering [44–46]. In the first step, a given locus is replaced with a selectable marker cassette. Second, the strain is transformed with a plasmid that expresses Cas9 and a guide RNA (gRNA) that targets the cassette, along with a DNA repair template containing terminal homology to sequences flanking the cassette in the genome. We designed the "gCASS5a" gRNA to specifically target a sequence shared by several selectable marker cassettes used for gene deletions in the popular pFA6a series (e.g., KanMX6, natMX4 and hphMX4; [47–50]; S1 Fig) such that the same gRNA can be used to replace each of these cassettes. By inserting two different cassettes flanking a genomic region, this gRNA can be used to exchange both cassettes along with the intervening sequence. This "double-cut" CRISPR-Swap method enables allele exchange even when the region contains sequences essential for survival (Fig 1B). Additionally, we designed a "gGFP" that specifically targets cassettes used for C-terminal tagging of open reading frames with GFP.

The gRNA/Cas9 complex is constitutively expressed from the CRISPR-Swap plasmid and will continue to cleave at the cassette(s) in the genome until a repair is made that abolishes the gRNA recognition sequence or the cell dies. Consequently, after transformation, all colonies that form on media lacking leucine have undergone a repair that blocks further cleavage by the gRNA/Cas9 complex. We designed the gCASS5a and gGFP gRNAs to cleave in the cassette but outside of the selectable marker gene, such that marker expression should be maintained if repair occurs without exchange of the selectable marker (Figure S1). Thus, transformants with the desired allele swap can be easily identified by screening for the loss of the resistance or auxotrophy conferred by the selectable marker.

We analyzed the results of 40 independent CRISPR-Swaps we performed to determine the efficiency of CRISPR-Swap. After transformation with the CRISPR-Swap plasmid expressing either gCASS5a or gGFP and a PCR-generated repair template, a median of 87.5% of the transformants no longer expressed the cassette selectable marker. We screened over 100 of these transformants for integration of the desired allele by colony PCR, restriction digestion or

sequencing and found that all had the correct allele exchange. We also sequenced the guide RNA recognition site in 13 transformants that remained G418 resistant and found that 2 had the hallmarks of repair by non-homologous end-joining, while the remaining 11 were repaired using the homologous sequence present in the GFP-His3MX cassette in these strains (see below) to repair the locus. We observed no difference in CRISPR-Swap efficiency among the two strain backgrounds (BY and RM), the genomic loci we targeted, or the gCASS5a and gGFP gRNAs. In summary, CRISPR-Swap readily creates allele replacements with high efficiency.

Fine-mapping of hotspot regions using GFP-tagged proteins to measure *trans*-gene expression

We leveraged the ability of CRISPR-Swap to rapidly engineer allelic series at a given locus to dissect three *trans*-acting hotspot regions to the causal variant. We focused on these hotspots because they have strong effects on many genes in *trans*, and because earlier fine-mapping in the 1,012 BY/RM segregant panel had located them to only a few genes. Before engineering each hotspot, we selected one abundantly expressed gene strongly affected by the hotspot in *trans* to monitor the effects of our genome edits on its gene expression (Fig 1C). We C-terminally tagged the open reading frame of this gene with GFP, engineered the hotspot locus with CRISPR-Swap, and measured GFP fluorescence in each engineered strain during growth on a 96-well plate reader (Methods). This approach provided high-throughput measurements of gene expression for the statistically powerful dissection of the hotspot loci.

A missense variant in *RGT2* influences expression of *HXT1* in trans

A hotspot locus on chromosome IV affects the expression of many genes at the mRNA and protein levels in crosses between BY and RM [12]. In the expanded BY/RM segregant panel, this locus had been mapped to a region containing three genes: *RGT2*, *ARF2*, and *RPL35-B* (Fig 2A). There mRNA levels of 1,400 genes are affected by this hotspot in *trans*, and these genes are enriched for roles in ATP synthesis and carbohydrate derivative metabolism. The strongest *trans* effect was on *HXT1* mRNA (LOD = 182), which encodes a low affinity glucose transporter whose expression is affected by Rgt2 [51], suggesting *RGT2* is likely the causal gene at this hotspot. In support of this,

177 genetic variation within the *RGT2* coding region was previously shown to influence protein levels
178 of the high affinity glucose transporter Hxt2p [29]. *RGT2* encodes a low affinity glucose sensor
179 located in the plasma membrane (Fig 2B; reviewed in [52]). The BY and RM alleles of *RGT2*
180 differ at five missense single nucleotide variant (SNVs). Three of these SNVs are located in the
181 long cytoplasmic-facing C-terminal tail required for glucose signaling, and the remaining two are
182 within the 12 predicted transmembrane helices (Fig 2B). *RGT2*, as well as *ARF2*, is also influenced
183 by a local eQTL.

184

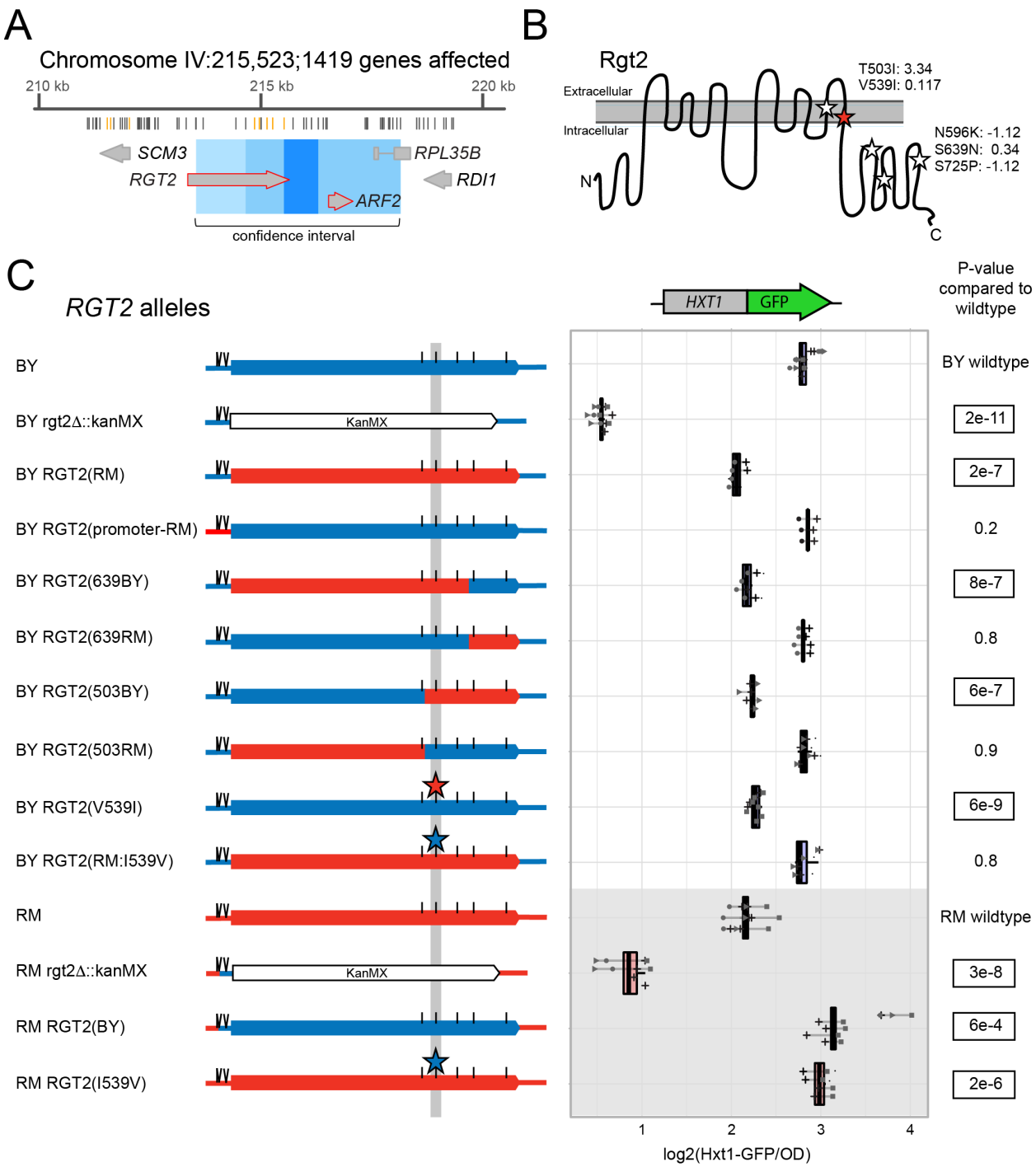


Fig 2. Fine mapping the causal variant in *RGT2*. A. Schematic of the eQTL hotspot on Chromosome IV. From top to bottom, the figure shows the hotspot location on the chromosome and the number of genes it affects, the positions of the BY and RM sequence variants (gray lines mark synonymous and intergenic variants and orange lines mark missense variants), the genes and their positions in the hotspot (large gray arrows; red outlines denote genes with a local eQTL) The 90% confidence interval (light blue), 95%

confidence interval (medium blue) and the position of markers perfectly linked to the hotspot peak marker (dark blue). B. Schematic of the Rgt2 protein in the plasma membrane. BY / RM missense variants are indicated by stars, with the identified causal variant in red. The missense variant amino acid positions, their resulting amino acid change and their PROVEAN scores [53] are shown. The synonymous variants are not depicted. C. On the left are schematics of *RGT2* alleles with BY sequences in blue and RM sequences in red. Positions of SNVs are marked with a straight line and INDELs are marked with two diagonal lines. The variants in the *RGT2* 3'UTR and synonymous variants are not depicted. On the right are the corresponding Hxt1-GFP fluorescence levels for each allele. P-values are for tests comparing each allele to its respective wildtype. Significant p-values are outlined. Blue boxplots indicate alleles in the BY background and red boxplots and gray background shading indicate alleles in the RM background. Lines group measurements of the same clone. Different symbols (circles, squares, etc.) denote different plate reader runs. Small dots next to a clone indicate a 5'UTR indel with an allele that does not match the one indicated in the allele schematics (see Methods).

To determine if *RGT2* is the causal gene at this hotspot and identify the responsible variant, we C-terminally tagged Hxt1 with GFP and created a series of allele replacements at the *RGT2* locus using CRISPR-Swap. In both BY and RM backgrounds, deletion of *RGT2* drastically reduced Hxt1-GFP levels, confirming that *RGT2* is required for proper Hxt1-GFP expression. Replacing the BY *RGT2* promoter region with the RM allele had no measurable effect on Hxt1-GFP. Reciprocal allele replacement of the entire *RGT2* coding region showed that the *RGT2* RM allele resulted in lower Hxt1-GFP expression compared to the BY allele (Fig 2C), which is the direction of effect expected from eQTL data [12].

By engineering a series of chimeric *RGT2* alleles spanning the coding region we systematically narrowed the causal region to two missense variants, neither of which were predicted to be deleterious (Fig 2C; S1 Table). Of these, a G-to-A SNV at 241,965 bp, resulting in a valine to isoleucine change at amino acid position 539 (V539I), recapitulated the effect of swapping the entire coding region. In a strain that carried all *RGT2* RM coding variants *except* V539I, Hxt1-GFP expression was indistinguishable from the BY wildtype, suggesting that V539I is the single causal variant in this gene. The effect of V539I on Hxt1-GFP expression was present in both BY and RM (Fig 2C), with no evidence that the strain background influences the effect of this variant

(interaction p-value = 0.1). The variant caused a minor effect on growth rate in the RM background, but no effect in the BY background (Table 1).

Table 1. Effects of causal hotspot variants on growth rates

Variant	Growth rate in BY (RM allele relative to BY wildtype allele)	p-value in BY	Growth rate in RM (BY allele relative to RM wildtype allele)	p-value in RM
RGT2(V539I)	98%	0.5	93%	0.02
OAF1(S63L)	96%	0.09	100%	>0.99
OLE1(FAR)	92%	0.3	92%	0.005

The *RGT2* V539I variant affects Hxt1-GFP expression in a glucose-dependent manner

Increasing glucose concentrations results in an increase in the expression of Hxt1 [51]. Given that the V539I variant lies within one of the predicted transmembrane helices that likely form the pocket necessary for glucose sensing by Rgt2 [54], we hypothesized that the effect of the V539I variant may change depending on the concentration of glucose in the culture medium. To test this idea, we measured Hxt1-GFP expression in BY and RM strains with their native *RGT2* alleles as well as with engineered V539I alleles in a range of glucose concentrations (Fig 3).

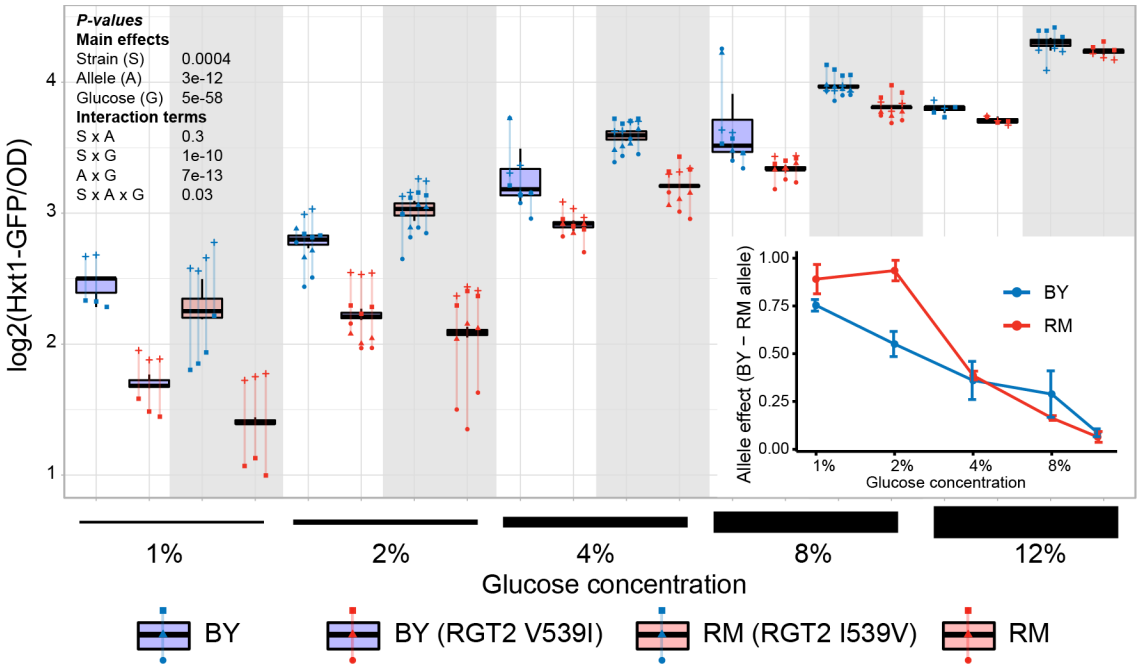


Fig 3. Effect of the *RGT2* I539V variant on Hxt1-GFP expression in different glucose concentrations. Hxt1-GFP fluorescence levels in the four different genotypes in increasing glucose concentrations are shown. Blue boxplots indicate alleles in the BY background and red boxplots and gray background shading indicate alleles in the RM background. Lines group measurements of the same clone. Different symbols (circles, squares, etc.) denote different plate reader runs. The inset graph shows fitted effect sizes as a function of glucose concentration. Error bars show standard errors of the mean. P-values are shown for ANOVA models examining the various main and interaction terms.

In both BY and RM, higher glucose concentrations increased expression of Hxt1-GFP regardless of which V539I allele was present. However, the difference in Hxt1-GFP expression between the V539I alleles showed a clear dependence on glucose levels (ANOVA test for interaction between allele effect and glucose concentration: p = 7e-13). The BY allele drove nearly 2-fold higher Hxt1-GFP expression than the RM allele at 1% glucose, while the difference between alleles became almost indistinguishable at 12% glucose. Thus, the effect size of the V539I variant underlying the *RGT2* eQTL hotspot strongly depends on the environment.

A missense variant in *OAF1* alters expression of *FAA4* in trans

Several of the eQTL hotspots identified in the BY/RM cross affect the expression of genes involved in fatty acid metabolism. They include a hotspot on chromosome I, which affects the mRNA and protein levels of 450 genes in *trans*. The locus was previously mapped to a region containing the genes *FLC2* and *OAF1* [12] (Fig 4A). Neither *FLC2* nor *OAF1* are influenced by a local eQTL, suggesting the causal variant is likely coding. *OAF1* encodes a transcription factor that activates expression of genes involved in peroxisome related functions including the β-oxidation of fatty acids [55–57], making it a promising candidate causal gene. The BY and RM alleles of *OAF1* differ at three missense SNVs (Fig 4B).

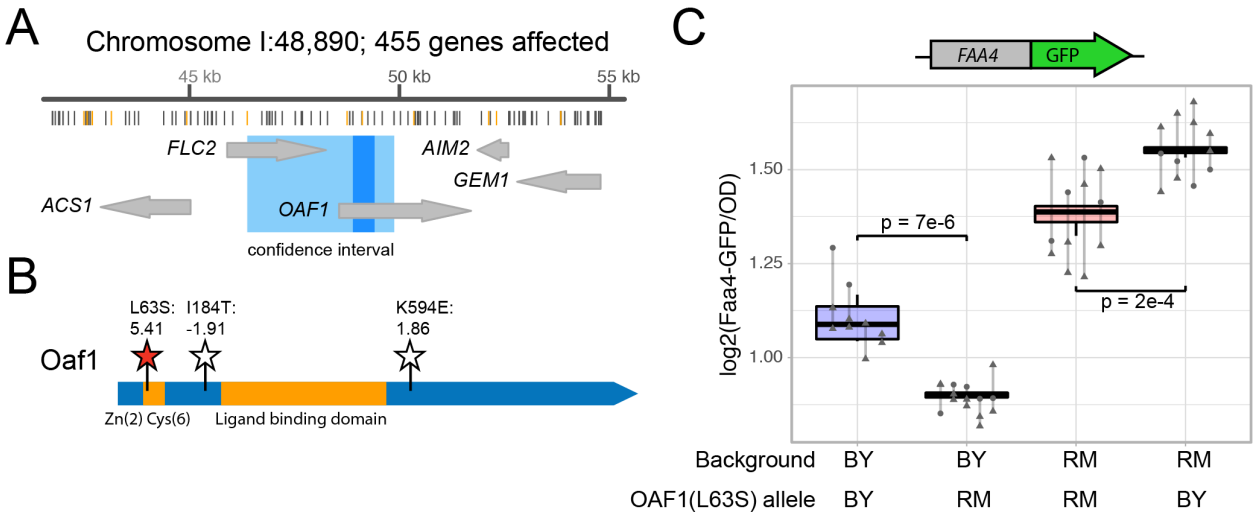


Fig 4. The *OAF1* causal variant. A. Schematic of the eQTL hotspot on Chromosome IV. From top to bottom, the figure shows the hotspot location on the chromosome and the number of genes it affects, the positions of the BY and RM sequence variants (gray lines mark synonymous and intergenic variants and orange lines mark missense variants), the genes and their positions in the hotspot (large gray arrows; red outlines denote genes with a local eQTL) The 95% confidence intervals (medium blue) and the position of markers perfectly linked to the hotspot peak marker (dark blue).B. Schematic of the Oaf1 protein domains and variants. The conserved zinc-finger domain and the ligand-binding domain [58] are in orange. BY / RM missense variants are indicated by stars, with the identified causal variant in red. The missense variant amino acid positions, their resulting amino acid change and their PROVEAN scores [53] are shown. The synonymous variants are not depicted. C. The figure shows Faa4-GFP fluorescence levels for strains with the indicated *OAF1* (L63S) alleles. P-values are for tests comparing the allele to its respective wildtype. Blue

boxplots indicate alleles in the BY background and red boxplots indicate alleles in the RM background. Lines group measurements of the same clone. Different symbols (circles, squares, etc.) denote different plate reader runs.

We engineered a series of *OAF1* alleles and measured Faa4-*GFP* expression in *trans*. *FAA4* mRNA abundance was shown to be strongly affected by this locus (LOD score = 102, [12]), and *FAA4* encodes a long chain fatty acyl-CoA synthetase, a protein with clear phenotypic connection to fatty acid metabolism. We found that a single T-to-C missense variant in *OAF1* at 48,751 bp resulted in a decrease in Faa4-GFP expression in agreement with the direction of effect observed in the eQTL data (Fig 4C, S2 Fig, S1 Table). The variant encodes a leucine at position 63 in BY and a serine in RM (L63S). In the Oaf1 protein, the L63S variant is located next to the Zn(2)Cys(6) DNA binding domain (Fig 4B) where it may alter Oaf1 binding to its regulatory targets. The effect of L63S on gene expression was consistent in both strain backgrounds (interaction p-value = 0.4; Fig 4C), and L63S had no detectable effects on growth rate (Table 1).

A noncoding promoter variant influences *OLE1* expression in *cis* and *FAA4* in *trans*

A second fatty-acid related hotspot resides on chromosome VII and affects the mRNA levels of 977 genes, which are enriched for functions in lipid biosynthetic processes as well as the response to endoplasmic reticulum stress. The confidence interval of this locus spans ~1 kb centered in the noncoding region between the *SDS23* and *OLE1* genes (Fig 5A). *OLE1* encodes the ER-bound Δ9-fatty acid desaturase, the only yeast enzyme capable of desaturating fatty acids [59,60]. *SDS23* is likely involved in regulation of the metaphase to anaphase transition of the cell cycle [61]. Given the enrichment of target genes of this hotspot involved in lipid metabolism, we reasoned that *OLE1* is the more likely causal gene. The BY and RM alleles of *OLE1* differ at one missense SNV and four non-coding variants (2 indels and 2 SNVs) in the *SDS23* and *OLE1* intergenic region.

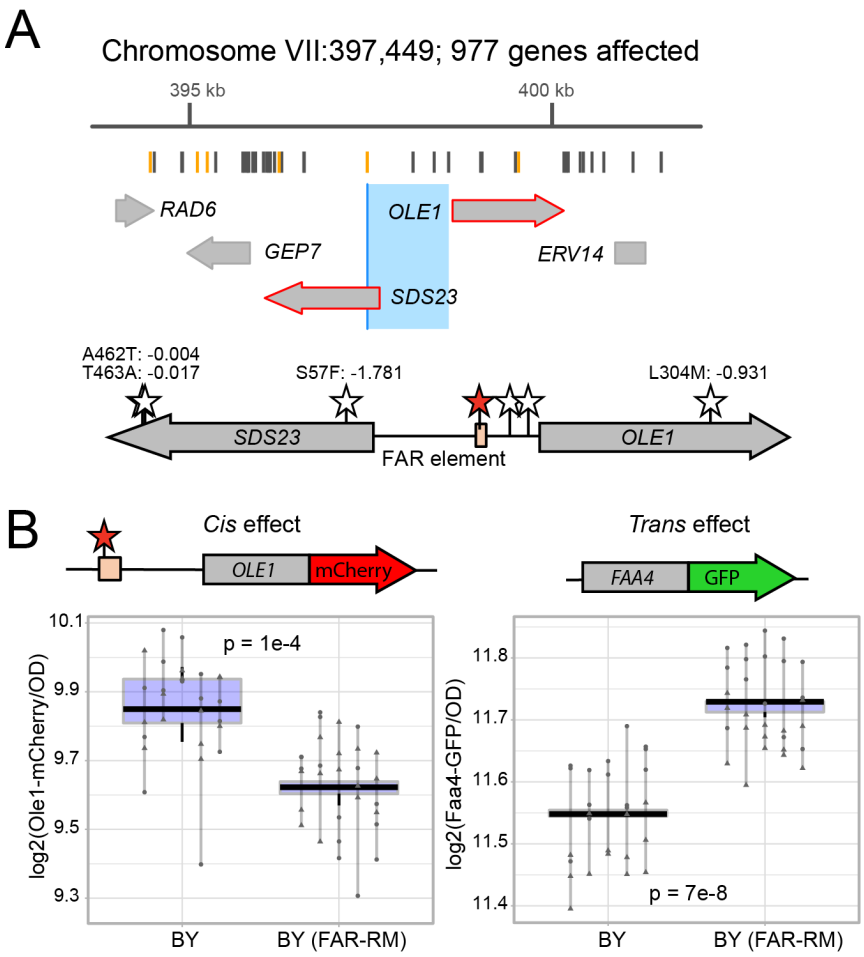


Fig 5. The *OLE1* causal variant. A. Schematic of the eQTL hotspot on Chromosome VII. From top to bottom, the figure shows the hotspot location on the chromosome and the number of genes it affects, the positions of the BY and RM sequence variants (gray lines mark synonymous and intergenic variants and orange lines mark missense variants), the genes and their positions in the hotspot (large gray arrows; red outlines denote genes with a local eQTL) The 95% confidence interval (medium blue) and the position of markers perfectly linked to the hotspot peak marker (dark blue). B. Schematic of the *SDS23* / *OLE1* region. BY / RM variants in the intergenic region and missense variants in the genes are indicated by stars, with the identified causal variant in the FAR element (peach box) in red. The missense variant amino acid positions, their resulting amino acid change and their PROVEAN scores [53] are shown. The synonymous variants are not depicted. C. Fluorescence levels of Ole1-mCherry (left panel) and Faa4-GFP (right panel). The BY strains harbor both Ole1-mCherry and Faa4-GFP and the indicated *OLE1* alleles. Lines group measurements of the same clone. Different symbols (circles, squares, etc.) denote different plate reader runs.

We engineered the essential *OLE1* locus using double-cut CRISPR-Swap, by flanking the region with the HphMX and KanMX cassettes and replacing both cassettes along with the intervening region with a series of alleles (Fig 1B). We again measured Faa4-GFP expression in the engineered strains, given *FAA4* mRNA levels are strongly affected by this locus (LOD = 78). We identified a noncoding A-to-G SNV at 398,081 bp in the intergenic region between *SDS23* and *OLE1* that affected Faa4-GFP expression in the expected direction based on the eQTL data (S3 Fig, S1 Table). This effect was consistent in both strain backgrounds (interaction p-value = 0.8). While the BY allele of this variant decreased growth rate in the RM background, the variant had no effect in the BY background (Table 1).

The causal variant is located in the fatty-acid regulated (FAR) promoter element, a region known to be important for transcriptional activation and fatty acid regulation of *OLE1* expression [62] (Fig 5B). Both *OLE1* and *SDS23* are strongly affected by local eQTLs with higher mRNA expression linked to the BY allele, suggesting that *OLE1* and/or *SDS23* expression may be affected by the FAR variant. To test the effect of the FAR variant on *OLE1* expression, we created BY strains with *OLE1* tagged with mCherry in addition to *FAA4* tagged with *GFP*. In these strains, the RM FAR allele significantly decreased Ole1-mCherry levels in *cis* and increased Faa4-GFP *levels* in *trans* (Fig 5C), which are the directions of effect expected based on the eQTL results [12].

To further examine the *cis*-regulatory activity of the FAR variant, we created reporter plasmids with multiple alleles of the *SDS23*/*OLE1* intergenic region driving expression of yeVenus. When the intergenic region drove expression of yeVenus in the direction of *SDS23*, the promoter alleles all drove similar yeVenus expression (S4 Fig). In sharp contrast, when the intergenic region drove expression of yeVenus in the direction of *OLE1*, the FAR-BY allele drove significantly higher yeVenus expression than the RM allele, fully recapitulating the difference between the all-BY and all-RM promoter alleles (S4 Fig). The effect did not depend on whether the yeVenus plasmids were expressed in the BY or RM background (interaction $p > 0.29$). Together, these results suggest that the FAR variant influences *OLE1* in *cis*, as well as *FAA4* in *trans*.

345 **Small changes in *OLE1* expression from wildtype levels are sufficient to affect *FAA4* gene**
346 **expression in *trans***

347 To explore the effect of *OLE1* and *SDS23* abundance on gene expression in *trans*, we added an
348 additional copy of *OLE1*, *SDS23* or the intergenic sequence on a single-copy plasmid in a BY
349 *FAA4-GFP* strain. The presence of an extra copy of *OLE1* resulted in significant reduction of Faa4-
350 GFP levels, consistent with eQTL data (S5 Fig). By contrast, an extra copy of *SDS23* resulted in
351 an increase in Faa4-GFP levels. An extra copy of the intergenic sequence alone did not change
352 Faa4-GFP levels (S5 Fig). These results further support that it is a change in *OLE1* and not *SDS23*
353 expression that is responsible for the *trans*-effect of the hotspot on *FAA4* expression.

354 While changes in Ole1 levels appear to be the primary cause for the *trans* effect on Faa4-GFP, the
355 noncoding FAR variant alters *OLE1* expression by only about 15% (Fig 5B). To understand the
356 relationship between *OLE1* and *FAA4* over a range of expression levels, we inserted a synthetic,
357 inducible Z_3EV promoter upstream of *OLE1*. This promoter can be activated precisely and
358 quantitatively by addition of estradiol to the culture medium [63]. We measured both Ole1-
359 mCherry and Faa4-GFP as a function of estradiol concentration and observed a clear
360 anticorrelation between Ole1-mCherry and Faa4-GFP levels (Fig 6). At an estradiol dose of 4 – 5
361 nM, the Z_3EV strain expressed Ole1-mCherry and Faa4-GFP at levels comparable to BY strains
362 expressing *OLE1-mCherry* from its native promoter with either the wildtype or FAR-RM allele.
363 Higher doses of estradiol continued to increase Ole1-mCherry levels, while Faa4-GFP dropped
364 rapidly and reached a plateau at levels well below the native BY strains. As expected for the
365 essential *OLE1* gene, low levels of induction resulted in poor growth (S6 Fig). Notably, growth of
366 the Z_3EV strain was closest to wildtype at 6 nM estradiol and higher, a concentration in which
367 *OLE1* expression levels are well above wildtype levels. The reason for this growth defect at
368 wildtype expression levels is unknown and may reflect an inability of strains with the Z_3EV
369 promoter and C-terminal mCherry fusion to finely regulate *OLE1* transcription. Taken together,
370 our results indicate that the BY strain expresses *OLE1* at a level at which even slight alterations in
371 *OLE1* expression, like that caused by the FAR variant, are sufficient to change *FAA4* expression
372 in *trans*.

373

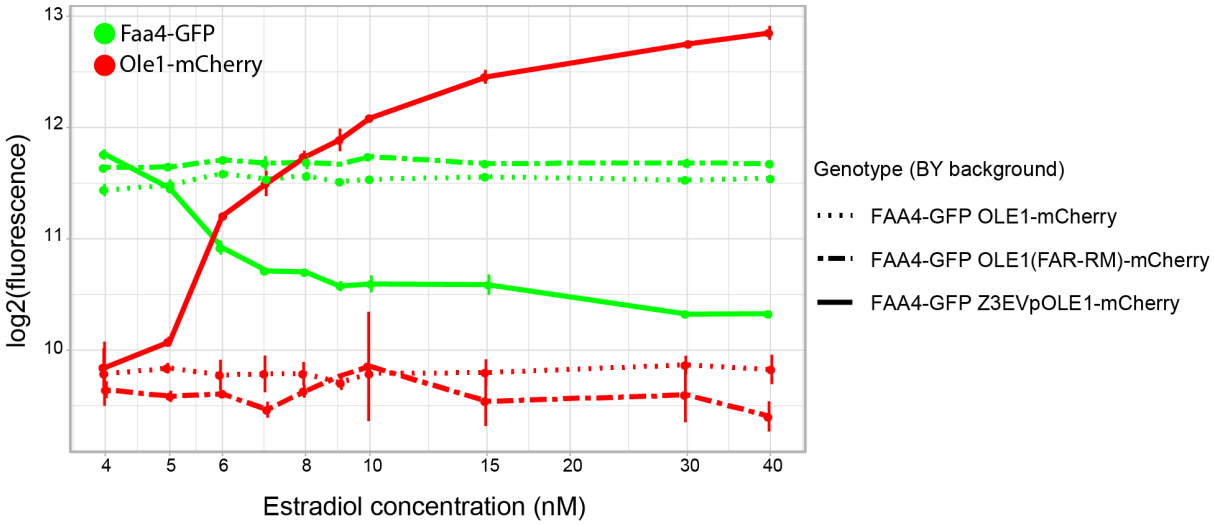


Fig 6. Expression modulation of Ole1. Fluorescence levels of Ole1-mCherry (solid red lines) under control of the estradiol-inducible Z_3EV promoter and Faa4-GFP (solid green lines) under its native promoter at various estradiol doses. Fluorescence levels of strains with the native *OLE1* promoter and the FAR-BY or FAR-RM allele respectively, are shown as dotted and dashed lines, respectively, at each estradiol concentration for comparison. All strains harbor both the mCherry tag on Ole1 and the GFP tag on Faa4.

A non-additive interaction between the *OAF1* and *OLE1* variants

The Oaf1 transcription factor binds throughout the *OLE1* promoter including in the close vicinity of the FAR variant [64]. This raises the possibility that the effects of *OAF1* L63S and the FAR variant may interact with each other genetically. While neither variant had shown significant interactions with the genetic background as a whole (see above), a specific interaction between these two variants could have been obscured by the thousands of other sequence differences between the BY and RM backgrounds. Indeed, the previous eQTL data contained a non-additive interaction affecting *FAA4* mRNA levels between two *trans* eQTLs that contained *OAF1* and *OLE1*, respectively (Fig 7A). To test if this *trans*-by-*trans* interaction could be explained by the specific causal variants we identified here, we constructed a BY strain that carried the RM alleles at both *OAF1* and the FAR variant and compared its Faa4-GFP expression to strains that carried one or the other of these edits. The combined alleles resulted in Faa4-GFP expression that differed from an additive allele combination (interaction $p = 0.002$) in a manner that mirrored the detected

eQTL interaction (Fig 7B). Specifically, in the presence of the more active RM *OAF1* allele, the FAR variant showed a greater effect than in the presence of the less active BY *OAF1* allele. As with the majority of epistatic interactions identified in this cross [65], the deviation from additivity caused by this interaction is detectable but subtle.

To test if this *trans*-by-*trans* interaction affecting *FAA4* could be mediated by a *cis*-by-*trans* interaction between the same two variants, we introduced our yeVenus plasmids with the two FAR alleles into BY strains with either the BY or RM *OAF(L63S)* alleles. We again found a significant interaction between the FAR and *OAF1* variants (Fig 7C). In the presence of the *OAF1*-L63S-RM allele, the FAR variant resulted in a larger difference in yeVenus expression than in the presence of the *OAF1*-L63S-BY allele (p = 1.5 e-7). The *OAF1* RM allele also resulted in overall higher expression of the yeVenus construct (p = 1.6 e-6). Taken together, these results show that the *OAF1* L63S variant and the *OLE1* FAR variant interact genetically such that their effects on gene expression in *cis* as well as *trans* are not simply additive.

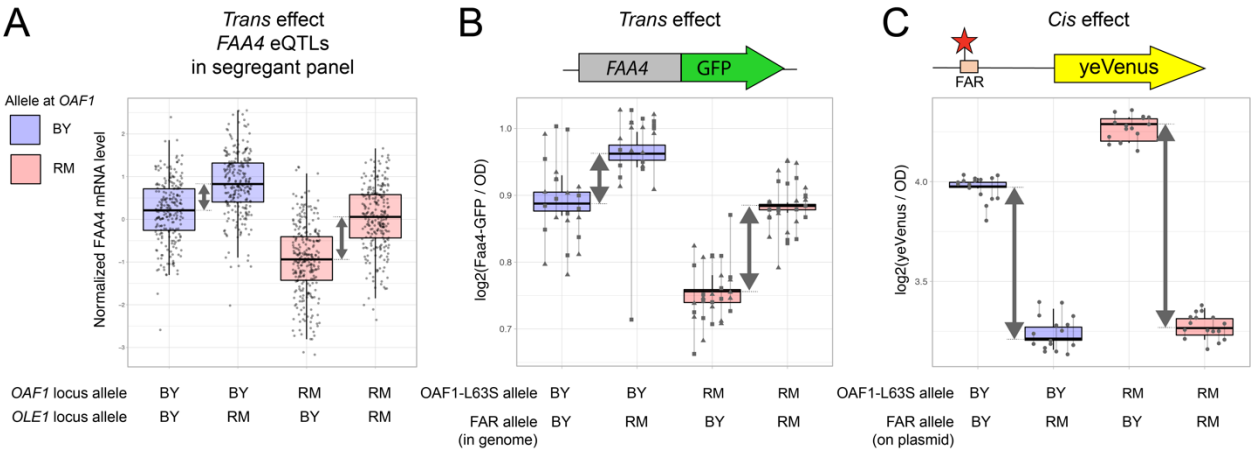
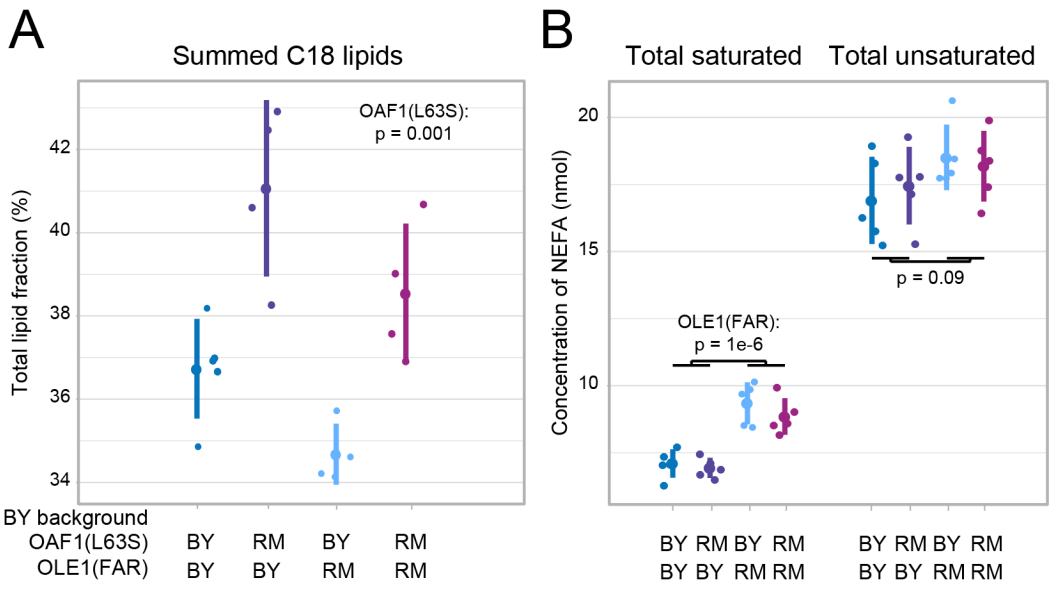


Fig 7. Nonadditive genetic interaction between the *OAF1* and *OLE1* causal variants. A. Data from [12] show levels of *FAA4* mRNA in 1,012 BY/RM segregants, divided into four groups depending on their genotypes at *OAF1* and *OLE1*. Note the slightly larger difference between *OLE1* alleles in segregants that carry the RM allele (red boxplots) compared to the BY allele (blue boxplots) at *OAF1*. B. Faa4-GFP fluorescence levels in strains engineered to carry the BY or RM allele at OAF1-L63S and OLE1(FAR). C. yeVenus fluorescence levels for strains engineered to carry the BY or RM allele at OAF1-L63S and the OLE1(FAR) promoter driving yeVenus expression on a plasmid. The difference between *OLE1* alleles in segregants (in A) or strains (in B and C) that carry the RM allele (red boxplots) compared to the BY allele (blue boxplots) at *OAF1* are highlighted by gray arrows.

418

419 **The *OAF1* and *OLE1* variants alter lipid profiles**

420 To test if the two causal variants in the fatty acid metabolism genes *OAF1* and *OLE1* alter cellular
421 phenotypes other than gene expression, we measured overall lipid composition as well as non-
422 esterified ("free") fatty acids (NEFAs) in the RM and BY strains, as well as in BY strains with
423 *OAF1(L63S)-RM*, *OLE1* FAR-RM, or both of these alleles (Fig 8 & S7 Fig). The BY and RM
424 strains differed in multiple metabolites (S7 Fig, S2 Table, S3 Table), and the presence of either of
425 the two variant alleles in the BY background also resulted in significant differences in lipid
426 metabolites. The BY *OAF1* allele decreased the fraction of longer-chain (C18) lipids (Fig 8A) and
427 also caused a decrease in the amount of C18 NEFAs (S7 Fig). This change resembles the effect of
428 an *OAF1* deletion on lipid metabolism [64], consistent with the BY *OAF1* allele having reduced
429 function (S2 Fig). The *OLE1* FAR-RM allele resulted in a significant increase in the amount of
430 saturated NEFAs but did not significantly reduce the amount of desaturated NEFAs (Fig 8B). This
431 suggests that reduced expression of the Ole1 desaturase caused by the FAR-RM allele results in
432 reduced consumption of the substrate but maintains levels of the reaction product in the reaction
433 catalyzed by Ole1. We did not detect changes in overall lipid composition due to the *OLE1* FAR-
434 RM allele. Taken together, our results show that the variants affecting *OLE1* expression and Oaf1
435 activity translate to changes in cellular lipids.



436

Fig 8. Lipid and fatty acid measurements. A. The fraction of C18 lipids in BY strains with wildtype or RM alleles at *OAF1(L63S)* or *OLE1*(FAR) or a combination of both. B. The concentration of saturated and unsaturated non-esterified fatty acids (NEFA) in the same strains. For each strain, the figure shows values for each replicate (smaller points) along with the mean (larger points) and standard deviation (vertical lines). See S7 Fig for all measurements.

The fine-mapped variants are the cause of the hotspot effects on mRNA levels

We have shown that each of the variants in *RGT2*, *OAF1*, and the *OLE1* FAR element affect the protein expression of a representative gene in *trans*. If these variants underlie the *trans*-acting hotspots, we expect them to alter the transcript levels of many genes, and that these expression changes will correlate with the known effects of the hotspots we sought to fine-map. To test this, we quantified transcript levels in BY strains edited at each variant (S4 Table – S6 Table).

Each of the three variants altered the transcript levels of dozens of genes including the representative gene we used for fine-mapping (*HXT1* or *FAA4*) with the expected direction of effect (Fig 9, Table 2). In addition, the FAR variant caused a nominally significant ($p = 0.03$) effect on *OLE1* but not on *SDS23* ($p = 0.7$), in agreement with our yeVenus reporter assay (S4 Fig). Crucially, the magnitude of expression change caused by the three variants was significantly and positively correlated with the respective hotspot effects (Fig 9) when considering all expressed genes. Like the vast majority of *trans* eQTLs [12], the three hotspots dissected here have small effects on most genes, typically explaining only a few percent of variance in mRNA levels. Our RNA-Seq experiment was not designed to detect such small effects at statistical significance, which would require dozens to hundreds of replicates. When we used a lenient significance cutoff (uncorrected $p < 0.05$) to restrict our analysis to genes with some evidence for differential expression, the correlations with hotspot effects increased at each hotspot (Table 2). These strong correlations were reflected in high directional concordance. For example, at a more stringent threshold (false discovery rate = 10%), every differentially expressed gene had concordant direction of effect with the given hotspot for *RGT2* and *OLE1*, and there was just a single discordant gene (out of 30) for *OAF1* (Fig 9). This strong agreement between expression changes caused by the three variants and known hotspot effects shows that these variants are causal variants at their respective hotspots.

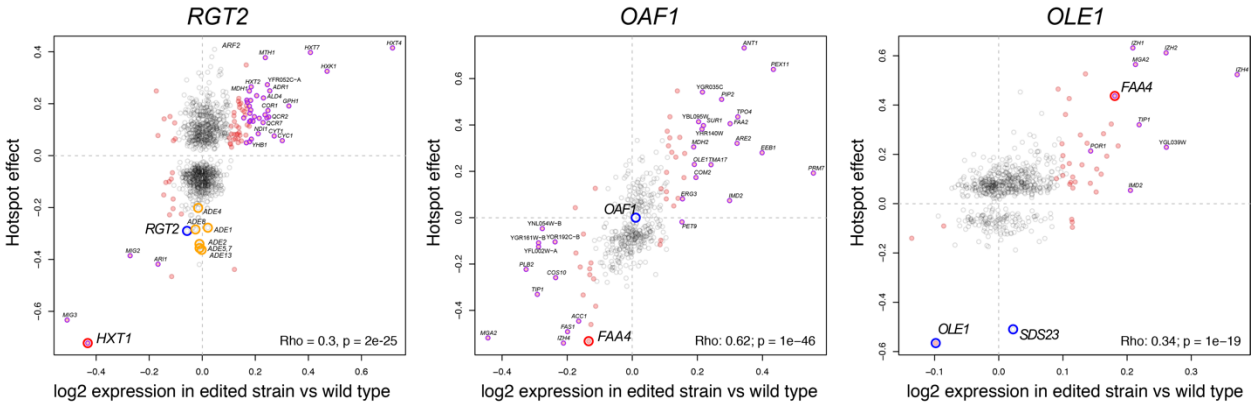


Fig 9. Comparison of the differential gene expression caused by each causal variant and the eQTL hotspot effects. The differential expression of an edited vs wildtype strain is on the x-axis and the previously determined hotspot effects are on the y-axis. All genes with non-zero hotspot effects are shown. Red points are genes with differential expression p-value < 0.05. Purple points are genes with differential expression at FDR < 0.1. The larger red circles mark the genes used for fine-mapping. The blue circles are genes we investigated in the confidence interval of each hotspot. The correlation between differential expression and hotspot effects is given in each panel. For *RGT2*, orange circles show adenine-metabolism related genes that are described in the text.

Table 2. RNA Seq results

Variant	Number of DE genes (10% FDR)	Number of DE genes ($p < 0.05$)	Correlation[1] between DE[2] and hotspot effect: all genes	Correlation[1] between DE[2] and hotspot effect: genes with DE $p < 0.05$
RGT2 (V539I)	45	183	Rho = 0.30, p = 2e-25	Rho = 0.64, $p \approx 0$
OAF1(L63S)	45	147	Rho = 0.62, p = 1e-46	Rho = 0.80, p = 6e-15
FAR	18	95	Rho = 0.34, p = 1e-19	Rho = 0.48, p = 0.0015

[1]These correlations exclude genes without a hotspot effect detected in [12].

[2]log2 fold change

If a hotspot is caused by multiple causal variants in the same or neighboring genes, our single variant edit might not account for all the effects of the hotspot. Therefore, we examined genes with

transcript levels strongly affected by the hotspot that are unaffected by our variant edits. Genes strongly affected by the hotspot at *RGT2* but not by the *RGT2 V539I* variant showed a significant enrichment for "*de novo* IMP biosynthetic process" (corrected p = 9e-9) and related terms in purine metabolism. This enrichment is driven by seven ADE genes with large hotspot effects that were all non-significant in our experiment (Fig 9). We suspect that a second variant in this region is responsible for these effects.

Population distribution and conservation of causal variants

To explore the evolutionary history of the three causal hotspot variants, we examined their distribution across a worldwide panel of 1,011 *S. cerevisiae* isolates with genomic sequence [66]. The BY allele at *OAF1* L63S is rare among yeast isolates (Fig 10). It is carried only by BY and a few close relatives while the RM allele is present in all other isolates as well as related species. Reflecting this pattern, the Protein Variation Effect Analyzer (PROVEAN) tool [53] assigned a "deleterious" score of -5.4 to the BY allele at this variant. In our experiments, the BY *OAF1* allele increased Faa4-GFP expression, which was the same direction caused by the *OAF1* knockout (S2 Fig). Thus, the *OAF1* hotspot is caused by a rare, derived missense variant almost exclusive to the BY laboratory strain that probably reduces function of the Oaf1 transcription factor.

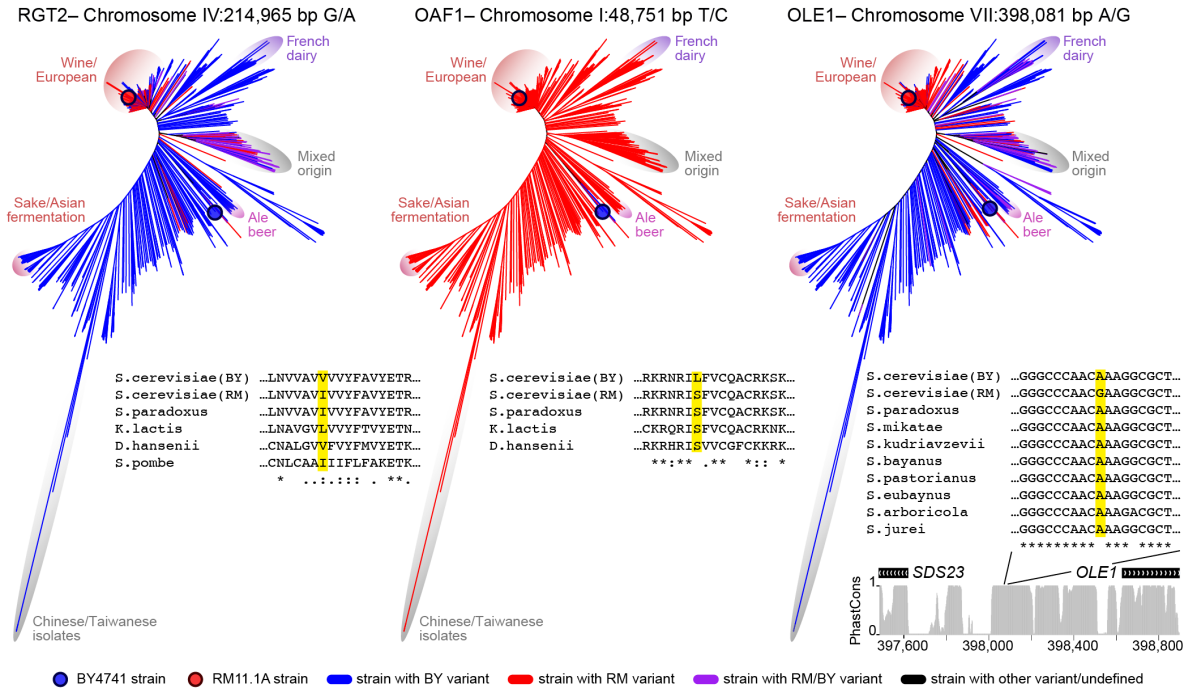


Fig 10. Population genetic features of the causal variants. Neighbor joining tree of the 1,011 *S. cerevisiae* strains sequenced in [66]. For each causal variant, the presence of the BY (blue), RM (red), BY and RM (purple), or other (black) allele is shown across the tree. For the *OAF1* and *RGT2* missense variants, amino-acid alignments with the indicated species are shown. For the intergenic FAR variant, we show nucleotide alignments as well as PhastCons conservation scores for the depicted *SDS23* / *OLE1* region from the UCSC Genome Browser.

The *RGT2* and *OLE1* variants show very different patterns. At *RGT2* V539I, both the valine in BY and the isoleucine in RM are also encoded by related yeast species (Fig 10). Evidently, the V539I variant can be tolerated without severe fitness consequences, as reflected in a "neutral" PROVEAN score of 0.12. Within *S. cerevisiae*, the highly divergent and likely ancestral group of Chinese isolates [66–68] carry the valine found in BY, suggesting that the isoleucine in RM is derived. This derived allele has ~25% frequency in the *S. cerevisiae* population, where it is predominantly found in isolates from the European wine clade, as well as in a second group of isolates with mixed origin (Fig 10). Both the derived RM variant and an *RGT2* deletion resulted in reduced induction of *HXT1* expression (Fig 2C).

518 At the *OLE1* FAR variant, the alanine found in BY is present in the ancestral Chinese isolates,
519 suggesting that the guanine in RM is derived (Fig 10). Indeed, the nucleotide sequence of the
520 noncoding region surrounding the FAR variant is conserved among *Saccharomyces sensu stricto*
521 species, and all other species carry the alanine found in BY at this position. The RM allele has
522 high frequency (46%) among yeast isolates, predominantly due to near fixation among the many
523 isolates in the European wine clade. The allele is also present in isolates from dairy, ale beer, and
524 other origins. The RM allele increased *FAA4* expression in *trans*. This is the same direction of
525 effect we observed when inserting the kanMX cassette immediately downstream of *OLE1*. Such
526 engineered alleles are commonly called "Decreased Abundance by mRNA Perturbation" (DAmP)
527 alleles and are expected to decrease gene function by lowering mRNA transcript levels ([69]; S3
528 Fig).

529 Thus, the three causal variants identified here have diverse population genetic characteristics. They
530 include a rare, lab-specific allele and two common alleles found in a quarter or more of the
531 sequenced yeast isolates. At all three variants, we observed that the likely derived allele altered
532 gene expression in the same direction as alleles that eliminate or reduce gene function.

Discussion

We fine-mapped natural DNA variants that each result in expression changes at many genes in *trans* using CRISPR-Swap, a strategy that facilitates rapid engineering of allelic series at a given locus. CRISPR-Swap is similar to other recently developed two-step engineering approaches [70,71], and has multiple advantages: 1) The great majority of clones that are transformed with the CRISPR-Swap plasmid and repair template incorporate the desired allele and clones without the desired allele are easily identified by screening for maintained expression of the cassette selectable marker. 2) gRNAs do not need to be designed and tested for each region due to the use of a common gRNA. 3) Regions without a nearby PAM site can be engineered because the gRNAs target integrated cassettes rather than the genomic region directly. 4) Larger regions, including those that contain essential genes can be engineered by flanking the region with two cassettes and using a single gRNA to cut and swap both cassettes and the intervening region. 5) Our gRNAs can be directly used to engineer existing strains that already contain cassettes e.g., strains in the *S. cerevisiae* deletion and GFP collections [72,73].

While we successfully used GFP-tagged protein abundance as phenotypic readouts amenable to high-throughput measurement, the hotspots we dissected had been identified via their effects on mRNA levels. The effects of the locus on the mRNA and protein of the gene used for phenotyping cannot always assumed to be consistent [22,37]. Further, fine-mapping using the expression of a single focal gene can only detect variants that influence this focal gene. Additional variants in the same region that specifically affect other genes would be missed. Indeed, while our RNA-Seq results were consistent with single causal variants at *OAF1* and *OLE1*, they suggested the presence of a second causal variant close to *RGT2*, which acts on genes involved in purine metabolism. Evidently, even such narrowly mapped hotspot loci as those we dissected here can be due to multiple causal variants with distinct effects but in close proximity to each other, as has been observed for QTLs for other traits in yeast [74,75].

A key result from this work is that the three causal variants we identified are strikingly different from each other. First, the variants include two coding missense variants (in *RGT2* and *OAF1*), along with the *cis*-acting noncoding FAR variant at *OLE1*. In yeast, 10 additional natural variants have been experimentally demonstrated to affect gene expression in *trans*, and there are 5 additional hotspots for which the gene but not the causal variant is known (S7 Table ,

563 [18,20,23,30,31,34,38–41], see also [76]) Coding variation underlies at least 13 of these 18 cases,
564 including missense, frameshift and transposon insertion variants. In species other than yeast,
565 information about causal *trans* eQTL variants remains extremely limited [77–80]. In human
566 genetics, searches for *trans* eQTLs often assume a model in which noncoding variants alter the
567 expression of a regulator gene in *cis*, which in turn alters the expression of other genes in *trans*
568 [81–85]. The FAR variant at *OLE1* discovered here is an example of such a mechanism. However,
569 the predominance of causal coding variants in yeast suggests that *trans* eQTLs caused by coding
570 variants may also exist in other species.

571 Second, the genes affected by the three variants encode different types of proteins: a glucose sensor
572 (Rgt2), a transcription factor (Oaf1), and the essential enzyme Ole1. While genes encoding
573 transcription factors are enriched in hotspot regions [12], hotspots clearly also arise from other
574 gene classes (S7 Table). Among these, enzymes are a particularly interesting group. Because most
575 enzymes do not directly regulate gene expression, metabolic changes caused by differential
576 enzyme activity or expression must trigger *trans* changes in gene expression indirectly [86,87].
577 The FAR variant at *OLE1* illustrates the indirect mechanisms that could underlie such indirect
578 *trans* effects. Its RM allele reduced Ole1 expression and increased saturated NEFAs. Higher lipid
579 saturation decreases membrane fluidity [88–90] which is sensed by membrane-bound dimers of
580 Mga2 or Spt23 [88,91–93]. In our data, the FAR RM allele increased *MGA2* expression in *trans*.
581 Apparently, this noncoding variant perturbs mechanisms involved in membrane homeostasis,
582 which may ultimately alter gene expression via the transcriptional regulator *MGA2*.

583 The *trans* effects of the FAR variant were caused by a decrease in Ole1 levels of only about 15%
584 (Fig 5B & 6). Such sensitivity to small expression changes may be unusual among genes. While
585 the BY and RM strains carry thousands of local eQTLs affecting at least half of the genes in the
586 genome, most of these local eQTLs do not result in detectable expression changes at other genes
587 in *trans* [12]. Little is known about whether, when, and how small expression changes at one gene
588 influence other genes in *trans*, and how some of these changes go on to influence the organism.
589 Pioneering studies have shown that even relatively small reductions in the expression of the
590 enzyme genes *TDH3* [94] and *LCB2* [95] can reduce fitness, and that the relationship between
591 gene expression and fitness is specific to each gene [96], the environment [96], and the strain
592 background [95]. Future work will explore the causal relationships among fitness and gene
593 expression changes in *trans*.

Finally, the three causal variants we discovered also differed in their population genetic characteristics. The *OAF1(S63L)* variant is rare, while the *RGT2* and *OLE1* variants are found in many isolates. These two common variants differ in the degree of evolutionary conservation of their site, with poor conservation at *RGT2(V539I)* and high conservation at the noncoding FAR variant. At least in the case of *RGT2*, simple evolutionary conservation alone would not have been sufficient to predict the causal variant.

At all three variants, the derived alleles affected *trans* gene expression in the same direction as loss of function or reduced function alleles. While this suggests that the derived alleles are detrimental, the high frequency in the population of the *RGT2* and *OLE1* alleles argues against strong negative fitness consequences. While these two variants could potentially be beneficial in some backgrounds or conditions, they could also be sufficiently mildly deleterious to have drifted to high frequency, as has been proposed to be the case for many *trans*-acting polymorphisms [97,98]. Indeed, none of the three variants resulted in consistent growth differences in both backgrounds in our culture medium (Table 1), suggesting that any fitness effects they may have are minor or occur in other environments. For example, at *RGT2*, the effect of the V539I variant on *HXT1* expression was reduced dramatically simply by altering the amount of glucose in the medium, suggesting that the long-term evolutionary consequences of this variant could be highly dependent on the environment.

With the arrival of population-scale genome sequencing [99], the functional interpretation of individual DNA variants has become a major goal of genetics, and numerous experimental and computational approaches aiming to predict phenotypic consequences of variants have been developed [53,100–107]. The diverse nature of variants that cause *trans*-eQTLs revealed here will make prediction of *trans* effects challenging. More data on the effects of variants on gene expression in *trans* will be important to understand *trans*-regulatory variation, both from high throughput approaches [108–114], and focused dissection of individual hotspots.

Materials and methods

Strains, Plasmids, Primers and Media

Experiments were performed in haploid *S. cerevisiae* strains derived from S288C (BY4741 (MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0), referred to as "BY" in the text) and RM-11a, (RM HO(BY) (MATa, his3Δ1::CloNAT, leu2Δ0, ura3Δ0 HO(BY allele) AMN1(BY allele), referred to as "RM"). All strains used in this study can be found in S8 Table. The HO(BY) allele was introduced into this RM strain by replacing the hphMX cassette at HO with the BY allele in YLK2442 (a gift from L. Kruglyak) by CRISPR-Swap. Importantly, the CloNAT resistance gene at the *HIS3* locus in RM HO(BY) is not recognized by the gCASS5a. All plasmids used in this study are in S9 Table. All primers/oligonucleotides are in S10 Table.

We used the following media (recipes are for 1L):

YNB+2% Glu +all (6.7 g yeast nitrogen base with ammonium sulfate and without amino acids, 20 g glucose, 50 mg histidine, 100 mg leucine, 50 mg methionine, 200 mg uracil and sterilized by filtration)

YNB+2% Glu -Leu (YNB+2% Glu+all (without leucine)

YPD (10 g yeast extract, 20 g peptone, 20 g glucose)

SDC-Leu (1.66 g SC -His -Leu -Ura, 50 mg histidine, 200 mg uracil, 20 g glucose)

SDC-His (1.66 g SC -His -Leu -Ura, 100 mg leucine, 200 mg uracil, 20 g glucose)

LB (10 g tryptone, 10 g NaCl, 5 g yeast extract)

Media for selection of resistance gene expression was supplemented at the following concentrations: ampicillin (100 µg/ml), nourseothricin sulfate (100 µg/ml), G418 sulfate (200 µg/ml), hygromycin B (300 µg/ml). For solid media, 20 g/L agar was added prior to autoclaving. Yeast were grown at 30°C. Bacteria were grown at 37°C.

Plasmid construction

644 To construct the CRISPR-Swap plasmids we annealed oligos OFA0185 and OFA0186 for
645 gCASS5a (pFA0055) and OFA0552 and OFA0553 for gGFP (pFA0057), and ligated them into
646 the BclI and SwaI sites of pML107, a gift of John Wyrick (Addgene #67639) as described in [45].

647 We are depositing pFA0055 and pFA0057 to Addgene under #131774 and #131784, respectively.

648 The yeVenus reporter plasmids were created by PCR fusion of the *OLE1* (-1 to -936) or *SDS23* (-
649 1 to -1090) promoter fragment with the open reading frame of yeVenus, a gift of Kurt Thorn
650 (Addgene plasmid #8714) [115]. To create plasmids pRS415-p*OLE1*(BY)-yeVenus and pRS415-
651 p*OLE1*(RM)-yeVenus, the *OLE1* promoter and yeVenus PCR fragments were digested with
652 HindIII and BglII and ligated into pRS415 [116] digested with HindIII and BamHI. To create
653 pRS415-pOLE1(BY_FAR_RM)-yeVenus and pRS415-pOLE1(RM_FAR_BY)-yeVenus the
654 *OLE1* promoter PCR fragment was digested with HindIII and PstI and ligated into the same sites
655 of pRS415-OLE1(BY)-yeVenus_1. To create pRS415-SDS23(BY)pVenus, the *SDS23* promoter
656 and yeVenus PCR fragment was digested with SalI and NdeI and cloned into the same sites of
657 pRS415-pOLE1(BY)-yeVenus_1. For details on the creation of the PCR fusions see S11 Table.

658

659 pRS415-OLE1(BY) was created by ligating the OLE1(BY) gene (-936 to+373) after PCR
660 amplification and digestion with HindIII (native in OLE1) and BamHI (on OFA0641 primer) into
661 the same sites of pRS415. pRS415-SDS23(BY) was created by ligating the SDS23(BY) gene (-
662 1033, +4730) after PCR amplification and digestion with BamHI and SacI into the same sites of
663 pRS415. pRS415-SDS23-OLE1p(BY) was created by ligating the 1009-bp intergenic region
664 between SDS23 and OLE1 after PCR amplification and digestion with BamHI and HindIII into
665 the same sites of pRS415.

666

667 **Tagging genes and cassette insertion**

668 Insertions of cassettes for genome modification were performed using a standard PCR-based
669 one-step method [50]. Selection markers used were HIS3MX6, which allows growth of *his3*-
670 mutants without exogenous histidine, KanMX4, which allows growth with G418, natMX6,
671 which allows growth with nourseothricin sulfate/CloNAT and hphMX4 or hphNT1, which
672 allows growth with hygromycin B. For C-terminal tagging of *HXT1*, we used the GFP-

673 HIS3MX6 cassette for tagging of *HXT1*, and the mCherry-hphNT1 cassette, a gift of Jiří Hašek
674 (Addgene plasmid #74635) [117] for tagging of *OLE1*. For the first step of CRISPR-Swap, we
675 used KanMX4 and hphMX cassettes for gene deletion or for cassette insertion without deletion.
676 After selection for these markers, the transformants were single colony purified and insertion
677 of the cassette in the correct location and absence of the wild-type allele were verified by PCR.

678

679 **Construction of repair templates for CRISPR-Swap**

680 Repair templates were PCR amplified from BY and RM genomic DNA with primers designed to
681 create products with termini homologous (ranging from 84 – 338 bp) to the region flanking the
682 targeted cassette and, when possible, to be free of BY/RM sequence differences. To create hybrid
683 BY and RM repair templates, we used PCR SOEing techniques [118]. See S11 Table for details
684 on construction of each template.

685

686 **CRISPR-Swap**

687 Strains were transformed with the gCASS5a or gGFP plasmid and a PCR-generated repair
688 template using a standard lithium acetate procedure [119]. For each transformation, we used 25 ml
689 of cells at OD_{600}= 0.4-0.8. To prepare 50 ml of cells for transformation, the cells were pelleted by
690 centrifugation at 3,000 g for 3 min, the supernatant removed and the cells were resuspended in 1
691 ml water and transferred to a 1.7 ml microfuge tube. The cells were pelleted at 2,500 g for 2 min,
692 the supernatant removed, and the cells were resuspended in 1 ml of Solution 1 (0.1M LiAc, 1X TE
693 buffer). The cells were pelleted once again, supernatant removed, and then resuspended in 200 µl
694 of Solution 1. For each transformation, ~125 µl of the cell mixture was transferred to a 1.7 ml
695 microfuge tube containing 100 ng of the guide RNA plasmid, 1000 ng of PCR-generated repair
696 template and 5 µl (10 µg/µl) of salmon sperm carrier DNA (Sigma #D7656) and the tube was
697 incubated on a turning wheel at 30°C for 30 min. After which, 700 µl of Solution 2 (0.1M LiAc,
698 1X TE and 40% PEG 3350) was added and the mixture was returned to the turning wheel and
699 incubated at 30°C for 30 min. Next, the mixture was incubated at 42°C for 15 min and then 500
700 µl of water was added. To wash the cells prior to plating, the cells were pelleted at 2,500 g for 2
701 min, the supernatant removed, and the cells were resuspended in 1 ml of water. The cells were

pelleted once again at 2,500 g for 2 min, the supernatant removed, and the cells resuspended in 200 µl of water, plated onto two SDC-Leu plates, and incubated at 30°C for 3 – 4 days. The median number of colonies growing on SDC-Leu after single-cut CRISPR-Swap was 38 for BY and 3 for RM strains. After double-cut CRISPR-Swap there were ~50% fewer transformants in each background. The resultant leucine prototrophic colonies were single-colony purified on SDC-Leu plates and then assayed for loss of the selectable marker cassette by identifying strains that could no longer grow on YPD with G418, or, for *OLE1* allele exchanges, YPD with G418 and/or YPD with hygromycin. We did not cure the strains of the CRISPR-Swap plasmid with the exception of the strains used for RNA sequencing and the RM HO(BY) strain YFA0254. In our experience, the plasmid is rapidly lost and therefore was likely not present during phenotyping of the strains. To cure the strains of the plasmid, the strains were single-colony streaked and then replica plated to SC-Leu to identify leucine auxotrophic colonies. We preserved a minimum of three independently derived strains for each allele swap in glycerol (15% final concentration) stored at -80°C.

Verification of allele exchange

We performed colony PCR to verify the absence of the selectable marker cassette(s) and presence of the desired allele. Further verification of variant incorporation was in some cases performed by sequencing or by restriction enzyme digestion. To verify variants by enzyme digestion, we screened for the presence of an SspI site created by the *OAF1* L63S BY allele and an NcoI site created by the *OLE1* L304M RM allele. We did not always verify the allele exchange since we found that 100% of the colonies that no longer expressed the selectable marker had the desired allele. Special cases of allele exchange verification are described below.

The *RGT2* repair templates in some allele swaps contained a single indel variant within the 5' homologous flanking region. In these strains, the variant was sequenced and only strains with the desired variant were preserved when possible. In other cases, the mismatched variant is indicated in Fig 2 and S8 Table. We found this variant to have no effect on expression of Hxt1-GFP.

The BY FAA4-GFP OLE1(L304M) strains were created using single-cut CRISPR-Swap of the KanMX cassette in BY FAA4-GFP DAmP(OLE1) (YFA0547) and a repair template made from BY genomic DNA with OFA0519, which carries the RM variant at L304M, and OFA0120. Because the site of homologous recombination can initiate anywhere along the *OLE1* gene present

in the genome, the L304M variant was verified by restriction digestion and sequencing and 4 of 10 allele swaps successfully incorporated the RM variant at L304M.

When performing double-cut CRISPR-Swap at the *OLE1* locus, we observed incorporation of unexpected BY/RM chimeric alleles in 2/36 strains in which we genotyped at least one variant. We believe this can occur because the *OLE1* repair template is homologous to the sequence between the two cassettes, allowing recombination to occur between the repair template and the intervening sequence.

Phenotyping of engineered strains

Precultures were inoculated with cells from strains freshly growing on YPD plates or from glycerol stocks and grown overnight in 800 – 1000 µl of YNB+2% Glu+all medium in a 2-ml deep-96-well-plate on an Eppendorf Mixmate shaker set at 1100 rpm at 30°C. The precultures were diluted to an OD_{600}=0.05 in 100 µl of the same medium in a 96-well flat bottom plate (Costar) and the plates were sealed with a Breathe Easy membrane (Diversified Biotech). The strains were phenotyped in a BioTek Synergy H1 (BioTek Instruments) plate reader at 30°C with readings taken every 15 min for 97 cycles with 10 sec of orbital shaking between reads and 11 – 13 min between cycles. Cell growth was characterized using absorbance readings at 600 nm and protein expression was measured using fluorescence readings taken from the bottom of the plate with the following parameters: GFP (excitation 488, emission 520 nm) and yeVenus (excitation 502, emission 532 nm) with an average of 10 reads per well and gain set on extended; and mCherry (excitation 588 nm, emission 620 nm) with an average of 50 reads per well and gain set at 150.

All plate reader measurements are available at our code repository at https://github.com/frankwalbert/threeHotspots.

Growth of *RGT2* strains in different glucose concentrations

HXT1-GFP tagged strains, BY (YFA0276-8), BY RGT2(V539I) (YFA0489-91), RM (YFA0279-81) and RM RGT2(I539V) (YFA0492-95) were precultured overnight in 1 ml of YNB+all with the indicated glucose concentration (1%-12% glucose w/vol.) The cultures were diluted in 100 µl of the same media and phenotyped as described above. We removed one strain (YFA0275, a BY

761 wild type strain present on one plate) from the analyses because it showed highly unusual *HXT1*-
762 GFP expression at high glucose.

763

764 **yeVenus reporter expression**

765 Strains BY (YFA0993), RM (YFA0254), or BY OAF1(L63S) (YFA0907) were transformed with
766 pRS415-based plasmids containing *OLE1* / *SDS23* intergenic fragments driving expression of the
767 yeVenus reporter. Phenotyping was performed as described above in YNB+2% Glu -Leu to
768 maintain the plasmids. Precultures were inoculated with transformants that were single colony
769 purified on SDC-Leu plates or using pools of transformants (~5-10 colonies) taken directly from
770 the SDC-Leu transformation plates and grown overnight, diluted, and phenotyped as described
771 above.

772

773 **Modulation of OLE1 expression using Z_3EV**

774 The BY *FAA4-GFP* strain expressing the estradiol responsive transcription activator was created
775 by inserting p*ACT1*-Z_3EV-NATMX [63] at the *HO* locus, S11 Table for more details. The *OLE1*
776 gene was then modified to have the Z_3EV responsive promoter cassette (KANMX-pZ_3EV and a
777 C-terminal fusion with mCherry-hphNT1. Cells of BY FAA4-GFP OLE1-mCherry(YFA1105),
778 BY FAA4-GFP OLE1(FAR_RM)-mCherry(YFA1140) and BY HO::pACT1_Z_3EV FAA4-GFP
779 Z_3EVpOLE1-mCherry (YFA1110; glycerol stock prepared from a culture grown in 8 nM
780 estradiol) were precultured overnight in 250 µl of YNB+2% Glu+all with 4, 5, 6, 8, 10, 15, 20, 30
781 or 40 nM of estradiol (Sigma E1024: 10 µM stock solution in ethanol), diluted in 100 µl of the
782 same media, and phenotyped as described above.

783

784 **Extraction and sequencing of mRNA**

785 Strains BY(YFA1130), BY RGT2 back to WT(YFA1131), BY OAF1(L63S) (YFA1132), BY
786 OLE1(FAR_RM) (YFA1133) and BY RGT2(V539I) (YFA1134) were precultured (7 precultures
787 of each strain) in 1 ml of YNB+2% Glu+all as described above. After ~18 hours of growth, the
788 precultures were diluted to OD_{600}=0.05 in 1 ml of the same media. After ~7 hours of growth, the
789 optical density of the cultures was measured in the plate reader and growth was continued until an

790 average plate OD_{600} = 0.37, at which time the plate was centrifuged for 3 min at 2,100 g, the
791 supernatant removed, and each cell pellet was resuspended in 1 ml of H_2O and transferred to a 1.7
792 ml microfuge tube. The tubes were centrifuged for 2 min at 16,000g, the supernatant removed, and
793 the cell pellets were immediately frozen by immersion of the tube in liquid nitrogen and stored at
794 -80°C.

795 Five cell pellets from each strain were chosen for RNA isolation so that the average OD_{600} within
796 and between strains were as uniform as possible, with an average OD_{600} = 0.37 (range of 0.34 to
797 0.40). Total RNA was isolated from cell pellets in 5 batches (each batch contained one of each of
798 the 5 strains) using the ZR Fungal/Bacterial RNA mini-prep kit including the DNase I on-column
799 digestion step (Zymo Research). Each cell pellet was resuspended in lysis buffer and transferred
800 to screw-capped tubes containing glass beads and the cells were broken open using a mini-
801 beadbeater (Biospec Products) in 5 cycles of 2 min bead beating followed by 2 min at -80°C. The
802 total RNA was eluted in 50 µl of DNAse/RNAse free water and the RNA Integrity and
803 concentration was measured using an Agilent 2200 Tape Station. RINe scores ranged from 9.8 –
804 10, with an average RNA concentration of 137 ng/µl.

805 Poly-A RNA was extracted from 550 ng of total RNA using NEBNext Poly(A) mRNA Magnetic
806 Isolation Module (NEB) and used as input into the NEB Ultra II Directional RNA library kit for
807 Illumina (NEB E7760) in two batches. NEBNext Multiplex Oligos for Illumina (Dual Index
808 Primers Set1) were used to amplify and barcode the libraries using the following cycling
809 conditions: Initial Denaturation at 98°C for 30 s and 10 cycles of: 98°C for 10 s and 65°C for 75
810 s, followed by a 65°C extension for 5 min.

811 The amplified DNA was quantified using Qubit DNA HS. Sample concentrations ranged from 47
812 – 102 ng/µl. An equal concentration of each of the 25 barcoded libraries was pooled and the
813 average fragment size of the library was 350 bp, as determined using an Agilent High Sensitivity
814 chip. High-output sequencing of 76-bp single-end reads was performed on an Illumina NextSeq
815 550 at the University of Minnesota Genomics Core. An average of 14 million reads were obtained
816 for each sample.

817 Sequencing reads are available at GEO as series GSE134169.

818

819 **Lipid Measurements**

820 Strains BY (YFA0897), RM HO(BY) (YFA0254), BY OAF1(L63S) (YFA0907), BY
821 OLE1(FAR_RM) (YFA0914) and BY OLE1(FAR_RM)OAF1(L63S)(YFA1096) were
822 precultured (6 precultures of each strain) for ~18 hours in 1 ml of YNB+2% Glu+all and then
823 diluted in 7 ml of the same media to an OD_{600} = 0.002 in a loosely capped 16 X 150 mm culture
824 tube and grown at 30°C on a turning wheel until an approximate OD_{600} = 0.34 (measured in plate
825 reader). Samples were centrifuged for 3 min at 2,100g, the supernatant was removed and the cell
826 pellets were resuspended in 1 ml of water and transferred to 1.7 ml microfuge tubes. The cells
827 were pelleted for 2 min at 16,000g, the supernatant removed, and the cell pellets were immediately
828 frozen by immersion of the tube in liquid nitrogen and stored at -80°C.

829 Cell pellets from each strain were resuspended in 1X PBS and sonicated. Fatty acids and total lipid
830 composition were measured against a standard curve on a triple quadrupole mass spectrometer
831 coupled with an Ultra Pressure Liquid Chromatography system (LC/MS) as previously described
832 [120]. Briefly, the cell pellets were spiked with internal standard prior to extraction with tert-Butyl
833 Methyl Ether (MTBE). Roughly 25% of the sample was dried down, hydrolyzed, re-extracted and
834 brought up in running buffer for the analysis of total fatty acid composition. The remaining portion
835 of the extract was dried down and brought up in running buffer prior to injecting on the LC/MS
836 for the NEFA measurement. Data acquisition was performed under negative electrospray
837 ionization condition.

838

839 **Allele frequencies of the variants in the population**

840 The phylogenetic tree used to describe the evolution of the causal variants was obtained from [66]
841 (personal communication). Briefly, the tree was formed from the analysis of 1,544,489 biallelic
842 sites across 1,011 *S. cerevisiae* strains using the R package "ape" [121] with the 'unrooted' method
843 to display the tree. For each of the three variants, the matrix of variants from [66] was used to
844 define which of the 1,011 strains carry the RM allele, the BY allele, both the BY and RM allele,
845 or another allele. The edges of the tree were colored based on the alleles each strain carries. The
846 absence of color continuity as seen with *OLE1* and *RGT2* can indicate multiple independent
847 mutation events, but is more likely to arise from out-crossing events leading to mosaic genomes
848 [122].

Alignments were performed using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) with default settings. Sequences for alignment were retrieved from the NCBI nucleotide database. Predictions of variant effect on the function of the Rgt2, Oaf1, Sds23, and Ole1 proteins were calculated using Provean (Protein Variation Effect Analyzer) (http://provean.jcvi.org).

Computational and statistical analyses

All analyses described below were conducted in R (www.r-project.org), with individual packages indicated throughout. Figures were generated using base R and ggplot2 [123]. Analysis code is available at https://github.com/frankwalbert/threeHotspots.

Quantification of gene expression from plate reader data

We used the 'growthcurver' R package [124] to fit a logistic growth curve to the OD_{600} readings in each well. Every plate had several blank wells that contained medium but no yeast, and we used these blank wells to correct for the optical density of the medium. We visually confirmed successful fit of the growth curve for every well, and additionally excluded any wells for which growthcurver indicated poor model fit. From the fitted growth model, we extracted growth rates as well as the "inflection point" of the growth curve, i.e. the time point at which the population reached half its maximum capacity. We chose this time point for our measure of expression because in practice it closely matches the OD_{600} values used to map the hotspots [12], and because while cultures are still growing exponentially at this time point, they have reached a high enough density to allow accurate quantification of fluorescence.

To obtain expression values for downstream analyses, we subtracted the mean OD_{600} and mean fluorescence of the blank wells included on each plate from all other wells on the plate. We calculated the mean of three background-subtracted time points centered on the inflection point (Fig 1C) for OD_{600} and fluorescence. Downstream analyses of gene expression used the ratio of this mean fluorescence divided by the mean OD_{600}.

Prior to statistical analysis, we $\log_2$-transformed these fluorescence ratios. Every plate that carried strains with the BY and RM backgrounds also carried untagged wild type control strains without any fluorescent markers. We subtracted the average $\log_2$(fluorescence ratio) values from these

878 untagged strains from those of the tagged strains. Thus, the fluorescent phenotypes used in
879 statistical analyses and displayed in the figures are in units of $\log_2$-fold change compared to
880 untagged strains with matched genetic background.

881

882 **Statistical analyses of plate reader data**

883 For fine-mapping, we used pairwise comparisons of genotypes to determine whether the
884 expression associated with a given edited allele differed significantly from the wild type or other
885 alleles. These pairwise tests were computed using mixed linear models whose random effect terms
886 depended on the structure of the data available for each comparison. Specifically, two random
887 terms were included where appropriate:

888 **Plate identity.** As fine-mapping progressed, most genotypes were measured on several plate runs,
889 usually along with different sets of other genotypes. This resulted in a complex data structure in
890 which genotypes were sometimes but not always included on the same plates, run over a span of
891 several months. To account for this structure, the model included plate identity as a random term.
892 To ensure that plate effects were properly accounted for in both genotypes in a given pairwise
893 comparison, each comparison was computed using only data from plates that carried both
894 genotypes under consideration. For example, while wildtype strains were run on multiple plates, a
895 given edited strain may only have been present on one of these plates. In this scenario, only this
896 one plate would be used in the statistical comparison. Note that data in figures in the paper that
897 display plate reader data from multiple plates were not corrected for plate effects. We chose to not
898 correct for plate effects in the plots because we wished to present a raw view of the data, and
899 because there is no good way to apply a common visual correction for plate identity when different
900 genotype comparisons require different plate corrections, depending on which genotypes were
901 present on each plate. In the figures, plate identity is indicated with different symbols (dots,
902 squares, triangles, etc.).

903 **Clone identity.** During strain engineering, we created at least three independent clones of each
904 strain. Clone identity was included in the model as a random effect to control for any systematic
905 differences among these clones. In the figures, we visually grouped data from different wells for a
906 clone by connecting these wells with a line. For the yeVenus reporter experiments, we collected
907 data from individual transformed colonies as well as from small transformant pools that each

908 contained (and effectively averaged) multiple colonies. For statistical analyses, we treated each
909 small transformant pool as if it were a colony.

910 In the equations below, we denote random effects in parentheses. For each genotype, we used the
911 "lmer" function in the lme4 package [125] to fit a model of the phenotype *y* with the above random
912 effect terms as appropriate:

913 $H0: y = (plate) + (clone) + \varepsilon$

914 where ε is the residual error. We fit a second model that includes a fixed effect term for genotype
915 identity:

916 $H1: y = (plate) + (clone) + genotype + \varepsilon$

917 We tested for significance of the genotype term using ANOVA comparing *H0* and *H1*. Note that
918 for a few comparisons in which only a single clone was run in replicate on a single plate, we did
919 not include random effects terms. We fit these models using the "lm" function.

920 P-values in plate reader analyses during fine mapping were not corrected for multiple testing.

921

922 **Tests for non-additive genetic interactions**

923 To test for non-additive interactions of a given allele (i.e., the BY or RM allele at a given causal
924 variant) with the strain background (BY or RM), we fit a model with a fixed effect term for strain
925 background, in addition to a term for the allele at the variant of interest:

926 $H0: y = (plate) + (clone) + strain + allele + \varepsilon$

927 and a second model that adds an interaction term between allele and strain background:

928 $H1: y = (plate) + (clone) + strain + allele + strain{:}allele + \varepsilon$

929 We used ANOVA to test if the inclusion of the *strain:allele* interaction term in *H1* significantly
930 improved model fit. These interaction tests only considered plates that contained both alleles for
931 the given variant in both strain backgrounds. We used the same models to test for interactions
932 between FAR alleles and *OAF1* alleles by replacing the "RM" factor level in *strain* with "BY
933 (OAF1-RM)".

To compare the *OAF1*/*OLE1* allele interaction to that observed earlier in eQTL data (Source Data 14 in [12]), we obtained expression values (Source Data 1 from [12]) and used a linear model to regress out effects of collection batch and OD_{600} for these data (Source Data 2 from [12]). We then used genotypes (Source Data 3 from [12]) at the marker positions corresponding to the two interacting loci at *OAF1* and *OLE1* to divide the segregants in [12] into four two-locus genotype classes and plot their *FAA4* mRNA levels in Fig 7A.

Dependence of the RGT2(V539I) effect on glucose concentration

To test if the effect of the causal variant in *RGT2* depended on glucose concentration in the medium, we extended our modeling framework to include *glucose* ($\log_2$(glucose concentration)) as a numeric covariate. We fit a model that included all possible pairwise interactions between strain, allele, and glucose:

$$H1: y = (plate) + (clone) + strain + allele + glucose + strain{:}allele + strain{:}glucose + allele{:}glucose + \varepsilon$$

and compared this model to simpler models from which we dropped the respective interaction term of interest. For example, to test for the interaction between glucose and the allele effect:

$$H0: y = (plate) + (clone) + strain + allele + glucose + strain{:}allele + strain{:}glucose + \varepsilon$$

We computed p-values comparing these models using ANOVA.

RNASeq data handling

We used trimmomatic [126] version 0.38 to trim Illumina adapters, filter out reads shorter than 36 bp, trim bases with a quality score of less than 3 from the start and end of each molecule, and perform sliding window trimming to remove bases with an average quality of less than 15 in a window of four bases. This filtering retained ≥97% of reads. We used kallisto [127] to pseudoalign these trimmed and filtered reads to the *S. cerevisiae* transcriptome obtained from Ensembl [128] build 93 based on genome version sacCer3 [129]. Following recommendations in [130], we used FastQC [131] and RSeQC [132] to examine the quality of our 25 samples and found them to all

961 be of high quality and, importantly, highly similar to each other. We retained all 25 samples for
962 downstream analysis.

963 As a measure of gene expression, we used "estimated counts" in the kallisto output for each gene
964 in each sample. To exclude genes with poor alignment characteristics, we used RSeQC to calculate
965 Transcript Integrity Numbers (TINs) per gene and sample, and also considered "effective gene
966 length" produced by kallisto. We retained genes in which no sample had a count of zero, no sample
967 had a TIN of zero, and with effective length larger than zero. This filtering retained 5,400 genes
968 (out of 6,713 annotated) for further analysis.

969

970 **RNASeq statistical analyses**

971 Statistical analyses were conducted using the DESeq2 R package [133]. During RNA isolation and
972 sequencing library generation, we had collected a number of covariates and batch identities:
973 Bioanalyzer-based RNA Integrity Number (RIN), Bioanalyzer-based RNA concentration, Qubit-
974 based RNA concentration, OD_{600} of the culture at time of flash freezing, as well as batch for cell
975 harvest, RNA isolation, and library generation. Samples from our 5 genotypes had been distributed
976 equally among these three batches. We examined the influence of these technical covariates by
977 comparing them to principal components computed on variance-stabilized data [133] and found
978 that the three batches (in particular cell harvest) appeared to influence the results. We thus included
979 these three batches as covariates in all further analyses. We used surrogate variable analysis (SVA)
980 [134] to further account for unexplained technical variation and included two SVs in our statistical
981 model. While choices about which specific technical covariates and SVs to include in the model
982 did slightly alter the significance tests for individual genes, our main result of positive correlations
983 between hotspot effects and differential expression was robust to these choices.

984 We fit the DESeq2 model to all 25 samples and conducted pairwise tests for differential expression
985 between genotypes. Specifically, we compared the edited *OAF1* and *OLE1* alleles to a BY
986 wildtype strain (YFA1130) engineered by CRISPR-Swap with gGFP to remove the GFP tag from
987 *FAA4*. We compared the edited *RGT2* variant to a BY wild type strain (YFA1131) engineered by
988 CRISPR-Swap with gCASS5a to replace rgt2Δ::kanMX6 with the *RGT2* BY allele. The main text
989 describes differential expression results based on either nominal p-values ($p < 0.05$), or based on

a multiple-testing corrected threshold computed via false discovery rate estimation using the Benjamini-Hochberg method [135] as reported by DESeq2.

To compare differential expression to hotspot effects, we used $\log_2$-fold changes estimated by DESeq2 and hotspot effects that had been estimated by fitting a lasso model to all expressed genes and all 102 hotspots, as described in [12]. These hotspot effects were obtained from Source Data 9 in [12]. We used nonparametric Spearman rank correlation to compare differential expression with the hotspot lasso coefficients. We excluded genes with a hotspot effect of zero. While inclusion of these genes slightly degraded the magnitude of the correlations between hotspot effects and differential expression, all correlations remained significant and positive.

Gene ontology enrichment analysis for genes with strong hotspot effects but no differential expression was conducted using the "Gene Ontology Term Finder" tool on SGD [136]. As the background set, we used all genes present in both the hotspot effect matrix and our RNASeq data. As the test set, we used genes with an absolute hotspot effect of at least 0.3 and a differential expression p-value larger than 0.3.

Statistical analyses of lipid data

To correct for possible technical confounders from lipid composition and NEFAs, we used a linear model to regress out effects of acquisition order and sample grouping. For NEFAs, we also regressed out total protein, which had been measured from the same samples. The residuals from these regression were used in the statistical analyses below. To obtain total saturated, unsaturated, C18, and C16 measures, we summed the measures for the respective individual lipid species in these groups.

To analyze the effects of the *OAF1* and *OLE1* alleles in the BY background, we jointly considered the measures from the four genotypes (BY, BY(*OAF1-L63S*), BY(FAR-RM), and BY with both RM alleles) by fitting a linear model to each lipid compound y that models the effects of the *OAF1* and the FAR allele at *OLE1*:

$$y = OAF1 + OLE1 + OAF1{:}OLE1 + \varepsilon$$

where ε is the residual error. We analyzed this model using ANOVA to test for main effects of each allele as well as for the interaction term. P-values were not corrected for multiple testing.

1019 The BY and RM backgrounds we compared using T-tests.

Acknowledgements

We thank Mahlon Collins, Randi A., Krisna Van Dyke, Joshua Bloom, and Meru Sadhu for critical reading and comments on the manuscript. We thank Leonid Kruglyak, David Kirkpatrick, Melissa Gardner, and Scott McIsaac for yeast strains and plasmids, and Anne Friedrich for help with the 1,011 yeast isolate genome data.

References

1. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nature Reviews Genetics. 2015;16: 197–212.

2. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science. 2012;337: 1190–1195.

3. Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, et al. Distant eQTLs and Non-coding Sequences Play Critical Roles in Regulating Gene Expression and Quantitative Trait Variation in Maize. Molecular Plant. 2017;10: 414–426. doi:10.1016/j.molp.2016.06.016

4. Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES. Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize. PLOS Genetics. 2014;10: e1004845. doi:10.1371/journal.pgen.1004845

5. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics. 2012;13: 59–69.

6. Kita R, Venkataram S, Zhou Y, Fraser HB. High-resolution mapping of cis-regulatory variation in budding yeast. Proceedings of the National Academy of Sciences. 2017;114: E10736–E10744.

7. Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, et al. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell. 2016;165: 1519–1529.

8. Arensbergen J van, Pagie L, FitzPatrick VD, Haas M de, Baltissen MP, Comoglio F, et al. High-throughput identification of human SNPs affecting regulatory element activity. Nat Genet. 2019;51: 1160–1169. doi:10.1038/s41588-019-0455-2

9. Vockley CM, Guo C, Majoros WH, Nodzenski M, Scholtens DM, Hayes MG, et al. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Research. 2015;25: 1206–1214.

10. Chang J, Zhou Y, Hu X, Lam L, Henry C, Green EM, et al. The Molecular Mechanism of a Cis-Regulatory Adaptation in Yeast. PLoS Genetics. 2013;9: e1003813.

11. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell. 2019;177: 1022-1034.e6. doi:10.1016/j.cell.2019.04.014

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

13. Price AL, Patterson N, Hancks DC, Myers S, Reich D, Cheung VG, et al. Effects of cis and trans Genetic Ancestry on Gene Expression in African Americans. PLOS Genetics. 2008;4: e1000294. doi:10.1371/journal.pgen.1000294

14. Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. PLoS Genetics. 2011;7: e1001317.

15. Grundberg E, Grundberg E, Small KS, Small KS, Hedman \AAsa K, Hedman \AAsa K, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nature Genetics. 2012;44: 1084–1089.

16. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and genomics of gene expression in peripheral blood. Nature Genetics. 2014;46: 430–437.

17. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169: 1177–1186.

18. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic Dissection of Transcriptional Regulation in Budding Yeast. Science. 2002;296: 752–755.

19. Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proceedings of the National Academy of Sciences. 2005;102: 1572–1577.

20. Smith EN, Kruglyak L. Gene–Environment Interaction in Yeast Gene Expression. PLoS Biology. 2008;6: e83.

21. Albert FW, Muzzey D, Weissman JS, Kruglyak L. Genetic Influences on Translation in Yeast. PLoS Genetics. 2014;10: e1004692.

22. Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. Genetics of single-cell protein abundance variation in large yeast populations. Nature. 2014;506: 494–497.

23. Sudarsanam P, Cohen BA. Single Nucleotide Variants in Transcription Factors Associate More Tightly with Phenotype than with Gene Expression. PLoS Genetics. 2014;10: e1004325.

24. Lewis JA, Broman AT, Will J, Gasch AP. Genetic Architecture of Ethanol-Responsive Transcriptome Variation in Saccharomyces cerevisiae Strains. Genetics. 2014;198: 369–382.

1087 25. Fraser HB, Moses AM, Schadt EE. Evidence for widespread adaptive evolution of gene
1088 expression in budding yeast. Proceedings of the National Academy of Sciences. 2010;107: 2977–
1089 2982.
1090 26. Skelly DA, Merrihew GE, Merrihew GE, Riffle M, Riffle M, Connelly CF, et al.
1091 Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast.
1092 Genome Research. 2013;23: 1496–1504.
1093 27. Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. Selection on noise
1094 constrains variation in a eukaryotic promoter. Nature. 2015;521: 344–347.
1095 doi:10.1038/nature14244
1096 28. Gagneur J, Stegle O, Zhu C, Jakob P, Tekkedil MM, Aiyar RS, et al. Genotype-
1097 Environment Interactions Reveal Causal Pathways That Mediate Genetic Effects on Phenotype.
1098 PLoS Genetics. 2013;9: e1003803.
1099 29. Parts L, Liu Y-C, Tekkedil MM, Steinmetz LM, Caudy AA, Fraser AG, et al. Heritability
1100 and genetic basis of protein level variation in an outbred population. Genome Research. 2014;24:
1101 1363–1370.
1102 30. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, et al. Trans-acting regulatory
1103 variation in Saccharomyces cerevisiae and the role of transcription factors. Nature Genetics.
1104 2003;35: 57–64.
1105 31. Fehrmann S, Bottin-Duplus H, Leonidou A, Mollereau E, Barthelaix A, Wei W, et al.
1106 Natural sequence variants of yeast environmental sensors confer cell-to-cell expression
1107 variability. Molecular Systems Biology. 2013;9: 695–695.
1108 32. Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between
1109 polymorphisms that affect gene expression in yeast. Nature. 2005;436: 701–703.
1110 33. Ronald J, Brem RB, Whittle J, Kruglyak L. Local Regulatory Variation in
1111 Saccharomyces cerevisiae. PLoS Genetics. 2005;1: e25.
1112 34. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, et al. Integrating large-scale
1113 functional genomic data to dissect the complexity of yeast regulatory networks. Nature Genetics.
1114 2008;40: 854–861.
1115 35. Thompson DA, Cubillos FA. Natural gene expression variation studies in yeast. Yeast.
1116 2017;34: 3–17. doi:10.1002/yea.3210

1117 36. Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Ruderfer DM, Bedalov A, et al.
1118 Genetic basis of proteome variation in yeast. Nature Genetics. 2007;39: 1369–1375.
1119 37. Foss EJ, Radulovic D, Shaffer SA, Goodlett DR, Kruglyak L, Bedalov A. Genetic
1120 Variation Shapes Protein Networks Mainly through Non-transcriptional Mechanisms. PLoS
1121 Biology. 2011;9: e1001144.
1122 38. Brown KM, Landry CR, Hartl DL, Cavalieri D. Cascading transcriptional effects of a
1123 naturally occurring frameshift mutation in Saccharomyces cerevisiae. Molecular Ecology.
1124 2008;17: 2985–2997.
1125 39. Kim HS, Huh J, Fay JC. Dissecting the pleiotropic consequences of a quantitative trait
1126 nucleotide. FEMS Yeast Res. 2009;9: 713–722. doi:10.1111/j.1567-1364.2009.00516.x
1127 40. Brion C, Ambroset C, Sanchez I, Legras J-L, Blondin B. Differential adaptation to multi-
1128 stressed conditions of wine fermentation revealed by variations in yeast regulatory networks.
1129 BMC Genomics. 2013;14: 681. doi:10.1186/1471-2164-14-681
1130 41. Lewis JA, Gasch AP. Natural Variation in the Yeast Glucose-Signaling Network Reveals
1131 a New Role for the Mig3p Transcription Factor. G3 - Genes|Genomes|Genetics. 2012;2: 1607–
1132 1612.
1133 42. Storici F, Resnick MA. The Delitto Perfetto Approach to In Vivo Site-Directed
1134 Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast.
1135 Methods in Enzymology. Academic Press; 2006. pp. 329–345. doi:10.1016/S0076-
1136 6879(05)09019-1
1137 43. Alexander WG, Doering DT, Hittinger CT. High-Efficiency Genome Editing and Allele
1138 Replacement in Prototrophic and Wild Strains of Saccharomyces. Genetics. 2014;198: 859–866.
1139 doi:10.1534/genetics.114.170118
1140 44. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in
1141 Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Research. 2013;41: 4336–
1142 4343.
1143 45. Laughery MF, Hunter T, Brown A, Hoopes J, Ostbye T, Shumaker T, et al. New vectors
1144 for simple and streamlined CRISPR–Cas9 genome editing in Saccharomyces cerevisiae. Yeast.
1145 2015;32: 711–720. doi:10.1002/yea.3098

1146 46. Akhmetov A, Laurent JM, Gollihar J, Gardner EC, Garge RK, Ellington AD, et al.
1147 Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. Bio-protocol. 2018;8:
1148 e2765. doi:10.21769/BioProtoc.2765
1149 47. Wach A, Brachat A, Pöhlmann R, Philippsen P. New heterologous modules for classical
1150 or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast. 1994;10: 1793–1808.
1151 doi:10.1002/yea.320101310
1152 48. Wach A, Brachat A, Alberti-Segui C, Rebischung C, Philippsen P. Heterologous HIS3
1153 Marker and GFP Reporter Modules for PCR-Targeting in Saccharomyces cerevisiae. Yeast.
1154 1997;13: 1065–1075. doi:10.1002/(SICI)1097-0061(19970915)13:11<1065::AID-
1155 YEA159>3.0.CO;2-K
1156 49. Goldstein AL, McCusker JH. Three new dominant drug resistance cassettes for gene
1157 disruption in Saccharomyces cerevisiae. Yeast. 1999;15: 1541–1553. doi:10.1002/(SICI)1097-
1158 0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K
1159 50. Longtine MS, Iii AM, Demarini DJ, Shah NG, Wach A, Brachat A, et al. Additional
1160 modules for versatile and economical PCR-based gene deletion and modification in
1161 Saccharomyces cerevisiae. Yeast. 1998;14: 953–961. doi:10.1002/(SICI)1097-
1162 0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U
1163 51. Ozcan S, Dover J, Rosenwald AG, Wölfl S, Johnston M. Two glucose transporters in
1164 Saccharomyces cerevisiae are glucose sensors that generate a signal for induction of gene
1165 expression. PNAS. 1996;93: 12428–12432. doi:10.1073/pnas.93.22.12428
1166 52. Özcan S, Johnston M. Function and Regulation of Yeast Hexose Transporters. Microbiol
1167 Mol Biol Rev. 1999;63: 554–569.
1168 53. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino
1169 acid substitutions and indels. Bioinformatics. 2015;31: 2745–2747.
1170 doi:10.1093/bioinformatics/btv195
1171 54. Scharff-Poulsen P, Moriya H, Johnston M. Genetic Analysis of Signal Generation by the
1172 Rgt2 Glucose Sensor of Saccharomyces cerevisiae. G3 (Bethesda). 2018;8: 2685–2696.
1173 doi:10.1534/g3.118.200338
1174 55. Luo Y, Karpichev IV, Kohanski RA, Small GM. Purification, identification, and
1175 properties of a Saccharomyces cerevisiae oleate-activated upstream activating sequence-binding

protein that is involved in the activation of POX1. J Biol Chem. 1996;271: 12068–12075. doi:10.1074/jbc.271.20.12068

56. Rottensteiner H, Kal AJ, Hamilton B, Ruis H, Tabak HF. A heterodimer of the Zn2Cys6 transcription factors Pip2p and Oaf1p controls induction of genes encoding peroxisomal proteins in Saccharomyces cerevisiae. Eur J Biochem. 1997;247: 776–783. doi:10.1111/j.1432-1033.1997.00776.x

57. Karpichev IV, Luo Y, Marians RC, Small GM. A complex containing two transcription factors regulates peroxisome proliferation and the coordinate induction of beta-oxidation enzymes in Saccharomyces cerevisiae. Mol Cell Biol. 1997;17: 69–80. doi:10.1128/mcb.17.1.69

58. Phelps C, Gburcik V, Suslova E, Dudek P, Forafonov F, Bot N, et al. Fungi and animals may share a common ancestor to nuclear receptors. Proc Natl Acad Sci U S A. 2006;103: 7077–7081. doi:10.1073/pnas.0510080103

59. Stukey JE, McDonough VM, Martin CE. Isolation and characterization of OLE1, a gene affecting fatty acid desaturation from Saccharomyces cerevisiae. J Biol Chem. 1989;264: 16537–16544.

60. Stukey JE, McDonough VM, Martin CE. The OLE1 gene of Saccharomyces cerevisiae encodes the delta 9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene. J Biol Chem. 1990;265: 20144–20149.

61. Goldar MM, Nishie T, Ishikura Y, Fukuda T, Takegawa K, Kawamukai M. Functional conservation between fission yeast moc1/sds23 and its two orthologs, budding yeast SDS23 and SDS24, and phenotypic differences in their disruptants. Biosci Biotechnol Biochem. 2005;69: 1422–1426. doi:10.1271/bbb.69.1422

62. Choi JY, Stukey J, Hwang SY, Martin CE. Regulatory elements that control transcription activation and unsaturated fatty acid-mediated repression of the Saccharomyces cerevisiae OLE1 gene. J Biol Chem. 1996;271: 3581–3589. doi:10.1074/jbc.271.7.3581

63. McIsaac RS, Oakes BL, Wang X, Dummit KA, Botstein D, Noyes MB. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. Nucleic Acids Research. 2013;41: e57–e57.

64. Bergenholm D, Liu G, Holland P, Nielsen J. Reconstruction of a Global Transcriptional Regulatory Network for Control of Lipid Metabolism in Yeast by Using Chromatin

1206 Immunoprecipitation with Lambda Exonuclease Digestion. mSystems. 2018;3: e00215-17.
1207 doi:10.1128/mSystems.00215-17
1208 65. Bloom JS, Kotenko I, Sadhu MJ, Treusch S, Albert FW, Kruglyak L. Genetic interactions
1209 contribute less than additive effects to quantitative trait variation in yeast. Nature
1210 Communications. 2015;6: 8712.
1211 66. Peter J, Chiara MD, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome
1212 evolution across 1,011 Saccharomyces cerevisiae isolates. Nature. 2018;556: 339–344.
1213 doi:10.1038/s41586-018-0030-5
1214 67. Wang Q-M, Liu W-Q, Liti G, Wang S-A, Bai F-Y. Surprisingly diverged populations of
1215 Saccharomyces cerevisiae in natural environments remote from human activity. Molecular
1216 Ecology. 2012;21: 5404–5417. doi:10.1111/j.1365-294X.2012.05732.x
1217 68. Duan S-F, Han P-J, Wang Q-M, Liu W-Q, Shi J-Y, Li K, et al. The origin and adaptive
1218 evolution of domesticated populations of yeast from Far East Asia. Nat Commun. 2018;9: 1–13.
1219 doi:10.1038/s41467-018-05106-7
1220 69. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman
1221 HW, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast
1222 genome. Nature Methods. 2008;5: 711–718. doi:10.1038/nmeth.1234
1223 70. Lee JT, Coradini ALV, Shen A, Ehrenreich IM. Layers of Cryptic Genetic Variation
1224 Underlie a Yeast Complex Trait. Genetics. 2019;211: 1469–1482.
1225 doi:10.1534/genetics.119.301907
1226 71. Holt S, Kankipati H, De Graeve S, Van Zeebroeck G, Foulquié-Moreno MR, Lindgreen
1227 S, et al. Major sulfonate transporter Soa1 in Saccharomyces cerevisiae and considerable substrate
1228 diversity in its fungal family. Nat Commun. 2017;8. doi:10.1038/ncomms14247
1229 72. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al.
1230 Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis.
1231 Science. 1999;285: 901–906. doi:10.1126/science.285.5429.901
1232 73. Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, et al. Global
1233 analysis of protein localization in budding yeast. Nature. 2003;425: 686–691.
1234 doi:10.1038/nature02026
1235 74. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, et al.
1236 Dissecting the architecture of a quantitative trait locus in yeast. Nature. 2002;416: 326–330.

1237 75. Sinha H, David L, Pascon RC, Clauder-Münster S, Krishnakumar S, Nguyen M, et al.
1238 Sequential Elimination of Major-Effect Contributors Identifies Additional Quantitative Trait
1239 Loci Conditioning High-Temperature Growth in Yeast. Genetics. 2008;180: 1661–1670.
1240 doi:10.1534/genetics.108.092932
1241 76. Fay JC. The molecular basis of phenotypic variation in yeast. Current opinion in genetics
1242 & development. 2013;23: 672–677.
1243 77. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—
1244 Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing
1245 GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression
1246 across human tissues. Nature. 2017;550: 204–213.
1247 78. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al.
1248 Systematic identification of trans eQTLs as putative drivers of known disease associations.
1249 Nature Genetics. 2013;45: 1238–1243.
1250 79. Small KS, Todorcevic M, Civelek M, Moustafa JSE-S, Wang X, Simon MM, et al.
1251 Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on
1252 adipocyte size and body composition. Nature Genetics. 2018;50: 572–+.
1253 80. Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Lu H, et al. A trans-acting locus
1254 regulates an anti-viral expression network and type 1 diabetes risk. Nature. 2010;467: 460–464.
1255 81. Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, et al. Dynamic Role of
1256 trans Regulation of Gene Expression in Relation to Complex Traits. The American Journal of
1257 Human Genetics. 2017;100: 571–580.
1258 82. Yang F, Wang J, Consortium TGte, Pierce BL, Chen LS, Aguet F, et al. Identifying cis-
1259 mediators for trans-eQTLs across many human tissues using genomic mediation analysis.
1260 Genome Res. 2017;27: 1859–1871. doi:10.1101/gr.216754.116
1261 83. Shan N, Wang Z, Hou L. Identification of trans-eQTLs using mediation analysis with
1262 multiple mediators. BMC Bioinformatics. 2019;20: 126. doi:10.1186/s12859-019-2651-6
1263 84. Pierce BL, Tong L, Chen LS, Rahaman R, Argos M, Jasmine F, et al. Mediation Analysis
1264 Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide
1265 Analysis among 1,800 South Asians. PLoS Genetics. 2014;10.
1266 85. Bryois J, Buil A, Evans DM, Kemp JP, Montgomery SB, Conrad DF, et al. Cis and Trans
1267 Effects of Human Genomic Variants on Gene Expression. PLoS Genetics. 2014;10: e1004461.

1268 86. Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ. Linking
1269 Metabolic QTLs with Network and cis-eQTLs Controlling Biosynthetic Pathways. PLOS
1270 Genetics. 2007;3: e162. doi:10.1371/journal.pgen.0030162
1271 87. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of
1272 gene expression in primary immune cells identifies cell type–specific master regulators and roles
1273 of HLA alleles. Nature Genetics. 2012;44: 502–510.
1274 88. Degreif D, de Rond T, Bertl A, Keasling JD, Budin I. Lipid engineering reveals
1275 regulatory roles for membrane fluidity in yeast flocculation and oxygen-limited growth.
1276 Metabolic Engineering. 2017;41: 46–56. doi:10.1016/j.ymben.2017.03.002
1277 89. Li P, Fu X, Zhang L, Li S. CRISPR/Cas-based screening of a gene activation library in
1278 Saccharomyces cerevisiae identifies a crucial role of OLE1 in thermotolerance. Microbial
1279 Biotechnology. 2018;0: 1–10. doi:10.1111/1751-7915.13333
1280 90. Fang Z, Chen Z, Wang S, Shi P, Shen Y, Zhang Y, et al. Overexpression of OLE1
1281 Enhances Cytoplasmic Membrane Stability and Confers Resistance to Cadmium in
1282 Saccharomyces cerevisiae. Appl Environ Microbiol. 2017;83: e02319-16.
1283 doi:10.1128/AEM.02319-16
1284 91. Hoppe T, Matuschewski K, Rape M, Schlenker S, Ulrich HD, Jentsch S. Activation of a
1285 Membrane-Bound Transcription Factor by Regulated Ubiquitin/Proteasome-Dependent
1286 Processing. Cell. 2000;102: 577–586. doi:10.1016/S0092-8674(00)00080-5
1287 92. Zhang S, Burkett TJ, Yamashita I, Garfinkel DJ. Genetic redundancy between SPT23 and
1288 MGA2: regulators of Ty-induced mutations and Ty1 transcription in Saccharomyces cerevisiae.
1289 Molecular and Cellular Biology. 1997;17: 4718–4729. doi:10.1128/MCB.17.8.4718
1290 93. Covino R, Ballweg S, Stordeur C, Michaelis JB, Puth K, Wernig F, et al. A Eukaryotic
1291 Sensor for Membrane Lipid Saturation. Mol Cell. 2016;63: 49–59.
1292 doi:10.1016/j.molcel.2016.05.015
1293 94. Duveau F, Toubiana W, Wittkopp PJ. Fitness Effects of Cis-Regulatory Variants in the
1294 Saccharomyces cerevisiae TDH3 Promoter. Molecular biology and evolution. 2017;34: 2908–
1295 2912.
1296 95. Rest JS, Morales CM, Waldron JB, Opulente DA, Fisher J, Moon S, et al. Nonlinear
1297 Fitness Consequences of Variation in Expression Level of a Eukaryotic Gene. Molecular biology
1298 and evolution. 2013;30: 448–456.

1299 96. Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S, et al.
1300 Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. Cell.
1301 2016;166: 1282–1294.e18.
1302 97. Schaefke B, Emerson JJ, Wang TY, Lu MYJ, Hsieh LC, Li WH. Inheritance of Gene
1303 Expression Level and Selective Constraints on Trans- and Cis-Regulatory Changes in Yeast.
1304 Molecular biology and evolution. 2013;
1305 98. Emerson JJ, Hsieh L-C, Sung H-M, Wang T-Y, Huang C-J, Lu HH-S, et al. Natural
1306 selection on cis and trans regulation in yeasts. Genome Research. 2010;20: 826–836.
1307 99. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al.
1308 DNA sequencing at 40: past, present and future. Nature. 2017;550: 345–353.
1309 doi:10.1038/nature24286
1310 100. Wagih O, Galardini M, Busby BP, Memon D, Typas A, Beltrao P. A resource of variant
1311 effect predictions of single nucleotide variants in model organisms. Molecular Systems Biology.
1312 2018;14: e8430. doi:10.15252/msb.20188430
1313 101. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general
1314 framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics.
1315 2014;46: 310–315.
1316 102. Majithia AR, Tsuda B, Agostini M, Gnanapradeepan K, Rice R, Peloso G, et al.
1317 Prospective functional classification of all possible missense variants in PPARG. Nature
1318 Genetics. 2016;48: 1570–1575.
1319 103. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning
1320 sequence-based ab initio prediction of variant effects on expression and disease risk. Nat Genet.
1321 2018;50: 1171–1179. doi:10.1038/s41588-018-0160-6
1322 104. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense
1323 mutations using PolyPhen-2. Current protocols in human genetics. 2013;Chapter 7: Unit7.20.
1324 105. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic
1325 regions by multiplex homology-directed repair. Nature. 2014;513: 120–123.
1326 106. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE, et al.
1327 Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet.
1328 2018;50: 874–882. doi:10.1038/s41588-018-0122-z

1329 107. Klein JC, Keith A, Rice SJ, Shepherd C, Agarwal V, Loughlin J, et al. Functional testing
1330 of thousands of osteoarthritis-associated variants for regulatory activity. Nat Commun. 2019;10:
1331 1–9. doi:10.1038/s41467-019-10439-y
1332 108. Sharon E, Chen S-AA, Khosla NM, Smith JD, Pritchard JK, Fraser HB. Functional
1333 Genetic Variants Revealed by Massively Parallel Precise Genome Editing. Cell. 2018;175: 544-
1334 557.e16. doi:10.1016/j.cell.2018.08.057
1335 109. Roy KR, Smith JD, Vonesch SC, Lin G, Tu CS, Lederer AR, et al. Multiplexed precision
1336 genome editing with trackable genomic barcodes in yeast. Nature Biotechnology. 2018;36: 512–
1337 520. doi:10.1038/nbt.4137
1338 110. Sadhu MJ, Bloom JS, Day L, Siegel JJ, Kosuri S, Kruglyak L. Highly parallel genome
1339 variant engineering with CRISPR–Cas9. Nat Genet. 2018;50: 510–514. doi:10.1038/s41588-018-
1340 0087-y
1341 111. Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, et al. Genome-
1342 wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome
1343 engineering. Nature Biotechnology. 2017;35: 48–55.
1344 112. Bao Z, HamediRad M, Xue P, Xiao H, Tasan I, Chao R, et al. Genome-scale engineering
1345 of Saccharomyces cerevisiae with single-nucleotide precision. Nature Biotechnology. 2018;36:
1346 505–508. doi:10.1038/nbt.4132
1347 113. Guo X, Chavez A, Tung A, Chan Y, Kaas C, Yin Y, et al. High-throughput creation and
1348 functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast. Nature
1349 Biotechnology. 2018;36: 540–546. doi:10.1038/nbt.4147
1350 114. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant
1351 Interpretation: Functional Assays to the Rescue. The American Journal of Human Genetics.
1352 2017;101: 315–325. doi:10.1016/j.ajhg.2017.07.014
1353 115. Sheff MA, Thorn KS. Optimized cassettes for fluorescent protein tagging in
1354 Saccharomyces cerevisiae. Yeast. 2004;21: 661–670. doi:10.1002/yea.1130
1355 116. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for
1356 efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989;122: 19–27.
1357 117. Malcova I, Farkasovsky M, Senohrabkova L, Vasicova P, Hasek J. New integrative
1358 modules for multicolor-protein labeling and live-cell imaging in Saccharomyces cerevisiae.
1359 FEMS Yeast Res. 2016;16. doi:10.1093/femsyr/fow027

1360 118. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without
1361 the use of restriction enzymes: gene splicing by overlap extension. Gene. 1989;77: 61–68.
1362 doi:10.1016/0378-1119(89)90359-4
1363 119. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier
1364 DNA/PEG method. Nature Protocols. 2007;2: 31–34.
1365 120. Persson X-MT, Błachnio-Zabielska AU, Jensen MD. Rapid measurement of plasma free
1366 fatty acid concentration and isotopic enrichment using LC/MS. J Lipid Res. 2010;51: 2761–
1367 2765. doi:10.1194/jlr.M008011
1368 121. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary
1369 analyses in R. Bioinformatics. 2018;35: 526–528.
1370 122. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population
1371 genomics of domestic and wild yeasts. Nature. 2009;458: 337–341.
1372 123. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag
1373 New York; 2016. Available: https://ggplot2.tidyverse.org
1374 124. Sprouffske K, Wagner A. Growthcurver: an R package for obtaining interpretable metrics
1375 from microbial growth curves. BMC Bioinformatics. 2016;17: 172. doi:10.1186/s12859-016-
1376 1016-7
1377 125. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using
1378 lme4. Journal of Statistical Software. 2015;67: 1–48. doi:10.18637/jss.v067.i01
1379 126. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
1380 data. Bioinformatics. 2014;30: 2114–2120.
1381 127. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
1382 quantification. Nature Biotechnology. 2016;34: 525–527.
1383 128. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018.
1384 Nucleic Acids Res. 2018;46: D754–D761. doi:10.1093/nar/gkx1098
1385 129. Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, et al. The
1386 Reference Genome Sequence of Saccharomyces cerevisiae: Then and Now. G3: Genes,
1387 Genomes, Genetics. 2014;4: 389–398. doi:10.1534/g3.113.008995
1388 130. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A
1389 survey of best practices for RNA-seq data analysis. Genome Biology. 2016;17: 13.
1390 doi:10.1186/s13059-016-0881-8

131. Andrews S, others. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

132. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28: 2184–2185. doi:10.1093/bioinformatics/bts356

133. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15: 550. doi:10.1186/s13059-014-0550-8

134. Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genetics. 2007;3: e161.

135. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 289–300.

136. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Research. 2012;40: D700–D705.

Supporting information captions

S1 Fig. Guide RNA recognition sequences. A. Schematic of a cassette typically used for gene deletions. The gCASS5a recognition sequence is marked with a bracket and the PAM site is underlined. Start of the TEF promoter sequence driving expression of the selectable marker is in red letters. B. Schematic of a cassette used for C-terminally tagging open reading frames with GFP. The location of the gGFP recognition sequence is marked with a bracket and the PAM site is underlined. The start of the GFP sequence is in neon green. The recognition sites for SalI (pink) and PacI (purple) are the Cas9 cleavage sites (scissors) are shown to allow easy comparison of the gRNA recognition sequences, which are specific to each cassette. Designated with arrows are the universal primer sequences, S1 or U2 and F1, used for amplification of common cassettes.

S2 Fig. *OAF1* fine mapping. On the left are schematics of *OAF1* alleles with BY sequences in blue and RM sequences in red. Missense variants are marked with a straight line. Synonymous and non-coding variants are not shown. On the right are the corresponding Faa4-GFP fluorescence levels for each allele. P-values are for tests comparing each allele to its respective wildtype. Significant p-values are outlined. Blue boxplots indicate alleles in the BY background and red boxplots and background gray shading indicate alleles in the RM background. Lines group measurements of the same clone. Different symbols (circles, squares, etc.) denote different plate reader runs.

S3 Fig. *OLE1* fine mapping. On the left are schematics of *OLE1* alleles with BY sequences in blue and RM sequences in red. Only the one missense variant and none of the synonymous variants in the open reading frame are marked. Variants in the non-coding region are maked with a single line for a SNV and a two diagonal lines for INDELs. On the right are the corresponding Faa4-GFP fluorescence levels for each allele. P-values are for tests comparing each allele to its respective wildtype. Significant p-values are outlined. Blue boxplots indicate alleles in the BY background and red boxplots and background gray shading indicate alleles in the RM background. Lines group measurements of the same clone. Different symbols (circles, squares, etc.) denote different plate reader runs.

S4 Fig. yeVenus reporter expression. On the top are schematics of the SDS23/OLE1 locus and the two orientations of the yeVenus reporter constructs. The bottom panel shows yeVenus fluorescence levels for the indicated yeVenus reporter constructs. Blue boxplots indicate alleles in the BY background and red boxplots indicate alleles in the RM background. Lines group measurements of the same clone. Different symbols (circles, squares, etc.) denote different plate reader runs.

S5 Fig. Effects of plasmid overexpression of SDS23/OLE1 sequences on Faa4-GFP expression. Faa4-GFP fluorescence levels of strains transformed with a *LEU2*-CEN plasmid containing the indicated sequence. Lines group measurements of the same clone.

S6 Fig. Growth rates as a function of estradiol dose. Error bars show standard deviations.

S7 Fig. Lipid and fatty acid measurements. All individual measurements are shown. For each genotype, the figure shows values for each replicate (smaller points) along with the mean (larger points) and standard deviation (vertical lines).

S1 Table. Fine-mapping p-values

S2 Table. Lipid measurements

S3 Table. Lipid statistics

S4 Table. RNA sequencing gene expression counts

S5 Table. RNA sequencing sample and batch information

S6 Table. RNA sequencing results

S7 Table. eQTL hotspot literature review

S8 Table. Yeast strains

1461 **S9 Table. Plasmids**

1462 **S10 Table. Oligos**

1463 **S11 Table. Yeast strain and plasmid construction**