Simultaneous quantification of mRNA and protein in single cells reveals post-transcriptional effects of genetic variation

Christian Brion, Sheila Lutz, and Frank W. Albert

Department of Genetics, Cell Biology and Development,

University of Minnesota

Minneapolis, MN, U.S.A.

Correspondence to: Frank W. Albert (falbert@umn.edu)

Keywords: eQTL, pQTL, yeast, Saccharomyces cerevisiae, proteomics, transcriptomics, gene expression regulation, complex traits, CRISPR, trans, YAK1

One sentence summary: A CRISPR-based dual reporter assay enables genetic mapping of DNA variants that specifically affect mRNA or protein levels in *trans*.

Running title: Genetics of mRNA & protein in single cells

Abstract

Trans-acting DNA variants may specifically affect mRNA or protein levels of genes located throughout the genome. However, prior work compared *trans*-acting loci mapped in studies performed separately or with limited statistical power. Here, we developed a CRISPR-based system for simultaneous quantification of mRNA and protein of a given gene via dual fluorescent reporters in single, live cells of the yeast *Saccharomyces cerevisiae*. In large populations of recombinant cells from a cross between two genetically divergent strains, we mapped 86 *trans*-acting loci affecting the expression of ten genes. Less than 20% of these loci had concordant effects on mRNA and protein of the same gene. Most loci influenced protein but not mRNA of a given gene. One such locus harbored a premature stop variant in the *YAK1* kinase gene that had specific effects on protein or mRNA of dozens of genes. These results demonstrate complex, post-transcriptional genetic effects on gene expression.

Introduction

Phenotypic variation in genetically complex traits is shaped by multiple DNA variants throughout the genome. The small effects of most of these variants pose a challenge for understanding the mechanisms through which individual variants act. Overcoming this challenge has the potential to improve our ability to understand disease, study evolutionary change, and help apply biological processes in industry and agriculture.

Many genetic variants that influence complex traits alter gene expression (Albert and Kruglyak, 2015; Maurano et al., 2012). Some of these variants are located in *cis*-regulatory elements or alter sequence features of the messenger RNA (mRNA) molecule itself. The proximity of such "*cis*-acting" variants to the genes they affect has aided their identication in numerous species (Aguet et al., 2017; Albert et al., 2018; Brem et al., 2002; Cheung et al., 2005; Clément-Ziza et al., 2014; Hasin-Brumshtein et al., 2014; Higgins et al., 2018; Ka et al., 2013; Kita et al., 2017; Morley et al., 2004; West et al., 2007). However, most genetic variation in gene expression arises from *trans*-acting variants that affect the activity or abundance of diffusible factors that in turn alter the expression of other genes (Albert et al., 2018; Grundberg et al., 2012; Wright et al., 2014). Compared to their target genes, *trans*-acting variants can be located anywhere in the genome, greatly complicating their identification in human association studies. In organisms such as yeast (Albert et al., 2018; Brem et al., 2002; Brion et al., 2020; Clément-Ziza et al., 2014; Thompson and Cubillos, 2017), plants (Fu et al., 2013; West et al., 2007; Zhang et al., 2011), worms (Snoek et al., 2017; Viñuela et al., 2010) and mouse (Gerrits et al., 2009; Hasin-Brumshtein et al., 2016), linkage analysis in recombinant progeny from experimental crosses has identified loci carrying variants affecting gene expression (expression quantitative trait loci, eQTLs), including thousands of eQTLs that affect gene expression in *trans*.

Genetic effects on gene expression can be as complex as those on organismal phenotypes. The expression of a gene can be affected by one or more *cis*-eQTLs and dozens of *trans*-eQTLs, each of which changes the expression of the gene by a small amount (Albert et al., 2018). Detecting the loci that give rise to this complex variation requires high statistical power resulting from the analysis of large numbers of individuals (Albert et al., 2018, 2014b; Bloom et al., 2013; Ehrenreich et al., 2010).

Post-transcriptional regulation plays a major role in the control of gene expression (McCarthy, 1998), and mRNA and protein levels across genes are often reported to be poorly correlated (Huh et al., 2003; Lahtvee et al., 2017; Liu et al., 2016). Nonetheless, most studies of regulatory variation measure mRNA instead of protein abundance, enabled by powerful quantification techniques such as RNA sequencing. Variants that influence mRNA abundance can act at different molecular levels, including transcription (Kilpinen et al., 2013) and mRNA degradation (Andrie et al., 2014; Pai et al., 2012). New techniques have allowed the study of gene expression variation beyond mRNA, including ribosome profiling to study mRNA translation (Albert et al., 2014a; Battle et al., 2015), and mass spectrometry to study protein abundance (Battle et al., 2015; Chick et al., 2016; Foss et al., 2011, 2007; Ghazalpour et al., 2011; Großbach et al., 2019; Picotti et al., 2013; Sun et al., 2018; Yao et al., 2018) and protein modifications such as phosphorylation (Großbach et al., 2019).

Fluorescent gene tags enable quantification of the abundance of a given protein of interest in single cells (Huh et al., 2003). In *S. cerevisiae*, fluorescence-activated cell sorting (FACS) of millions of GFP-tagged recombinant cells from a cross between genetically different strains can be used to collect populations of thousands of single cells with high or low protein expression (Albert et al., 2014b; Parts et al., 2014). Pooled, genome-wide sequencing of these populations has provided high statistical power to identify genetic loci that influence protein abundance ("protein-QTLs") (Damerval et al., 1994). This "bulk segregant" approach (Michelmore et al., 1991), which is designed to detect *trans*-acting loci, led to a 10-fold increase in the number of detected protein-QTLs (to an average of 7.2 protein-QTLs per gene) compared to analyses of mass spectrometry-based proteomics in one hundred segregants (Albert et al., 2014b).

In comparisons among different studies, many protein-QTLs did not overlap with loci that affected mRNA ("mRNA-QTLs") of the same gene, and *vice versa* (Albert et al., 2018). Further, some loci affected both mRNA and protein but in opposite directions. At such "discordant" loci, the same allele increased mRNA abundance but decreased protein abundance of the same gene. These results suggest that genetic variants can independently affect the different layers of gene expression regulation (Albert et al., 2018, 2014b; Foss et al., 2011; Großbach et al., 2019).

However, there are potential caveats to this conclusion. The QTLs under comparison were identified in experiments conducted at different times, in different laboratories, using different technologies with

unique sensitivities and biases, and often using small sample sizes that limited statistical power. These comparisons are likely to be confounded by environmental differences among studies, which existed either by design (for example, different culture media) or may have resulted from experimental inconsistencies (for example, slight differences in the precise stage of cell growth, or in temperature). These issues are especially problematic when comparing *trans*-acting QTLs with small effects, which could be particularly sensitive to environmental influences (Smith and Kruglyak, 2008). While a recent study used mRNA-sequencing and mass spectrometry of the same yeast cultures to enable a direct comparison of mRNA-QTLs and protein-QTLs (Großbach et al., 2019), its sample size limited detection of QTLs with small effects. As a result, the importance of genetic variation, especially *trans*-acting variation, that specifically affects post-transcriptional processes remains unclear.

To address this challenge, we developed a system for quantifying mRNA and protein from the same gene simultaneously, in the same, live, single cells using two fluorescent reporters. We reasoned that such an approach would equalize all environmental confounders and most of the technical biases that could obscure the relationship between mRNA-QTLs and protein-QTLs. Our assay is sensitive enough to be used in FACS, permitting the use of well-powered bulk segregant mapping in a yeast cross. Genetic mapping across ten genes revealed 86 *trans*-acting loci. The vast majority of the identified mRNA-QTLs and protein-QTLs for a given gene did not overlap or had discordant effects on mRNA and protein. These results demonstrate considerable discrepancies in the genetic basis of variation in mRNA vs protein abundance.

Results

A reporter system for quantifying mRNA and protein in single, live cells

We designed a dual reporter system for the simultaneous quantification of mRNA production and protein abundance of a given gene in single, live cells. In this system, protein abundance is measured via a fluorescent GFP tag fused to the C-terminus of the given gene of interest (Huh et al., 2003). To measure mRNA, we reasoned that a clustered regularly interspaced short palindromic repeats (CRISPR) guide RNA (gRNA) (Doudna and Charpentier, 2014) produced in equal molarity with the mRNA of interest would be able to drive proportional expression of a reporter gene via CRISPR-

activation (Gilbert et al., 2014; Konermann et al., 2015). To implement this idea, we created a gRNA tag located in the 3'UTR of the gene, downstream of the sequence encoding GFP (Figure 1A). After transcription of the mRNA along with this tag, the gRNA is released from the mRNA by two flanking self-cleaving ribozymes (Hammerhead, Hh; and Hepatitis Delta Virus, HDV) (Gao and Zhao, 2014). Because gRNA cleavage separates the mRNA from its poly-adenylated (polyA) tail, we added a synthetic polyA tail between the GFP tag and the Hh ribozyme (Gao and Zhao, 2014). Once released, the gRNA directs a catalytically deactivated CRISPR associated enzyme (dCas9) fused to a VP64 activation domain (dCas9-VP64) to drive the expression of an mCherry gene integrated in the genome (Farzadfard et al., 2013). After gRNA release, stability and half-life of the mRNA no longer affects gRNA abundance, such that mCherry expression primarily reports on mRNA production.

The reporter system is implemented as two cassettes (Figure 1A). The "GFP-gRNA tag" cassette is added at the 3' end of the gene of interest. A second cassette, which we call the "CRISPR reporter", comprises the remaining components: *dCAS9-VP64* and the *mCherry* gene under the control of an inactive *CYC1* promoter fragment. This promoter contains one recognition sequence that, when targeted by the gRNA and dCas9-VP64, drives mCherry expression (Farzadfard et al., 2013). The two cassettes are stored on two plasmids that can be used to easily construct strains for quantification of mRNA and protein of any gene of interest (Figure S1).

We tested the reporter system in diploid BY strains tagged at two genes with different expression levels: the highly expressed *TDH3,* and *GPD1*, which has an average expression level compared to other genes in the genome. Both genes gave green and red fluorescent signals in a plate-reader (Figure 1B). A strain carrying the CRISPR reporter and *TDH3* tagged with GFP but no gRNA produced no mCherry fluorescence, demonstrating that the gRNA is required for driving mCherry expression (Figure 1B). Presence of the gRNA tag increased Tdh3-GFP levels by 1.3-fold (Figure 1C). Quantitative real time reverse-transcription PCR (qPCR) confirmed expression of the gRNA and the mRNA (Figure 1D). Absence of qPCR signal from primers that spanned the ribozyme cut sites in cDNA confirmed that the ribozymes properly cleaved the gRNA (Figure 1D & S2).

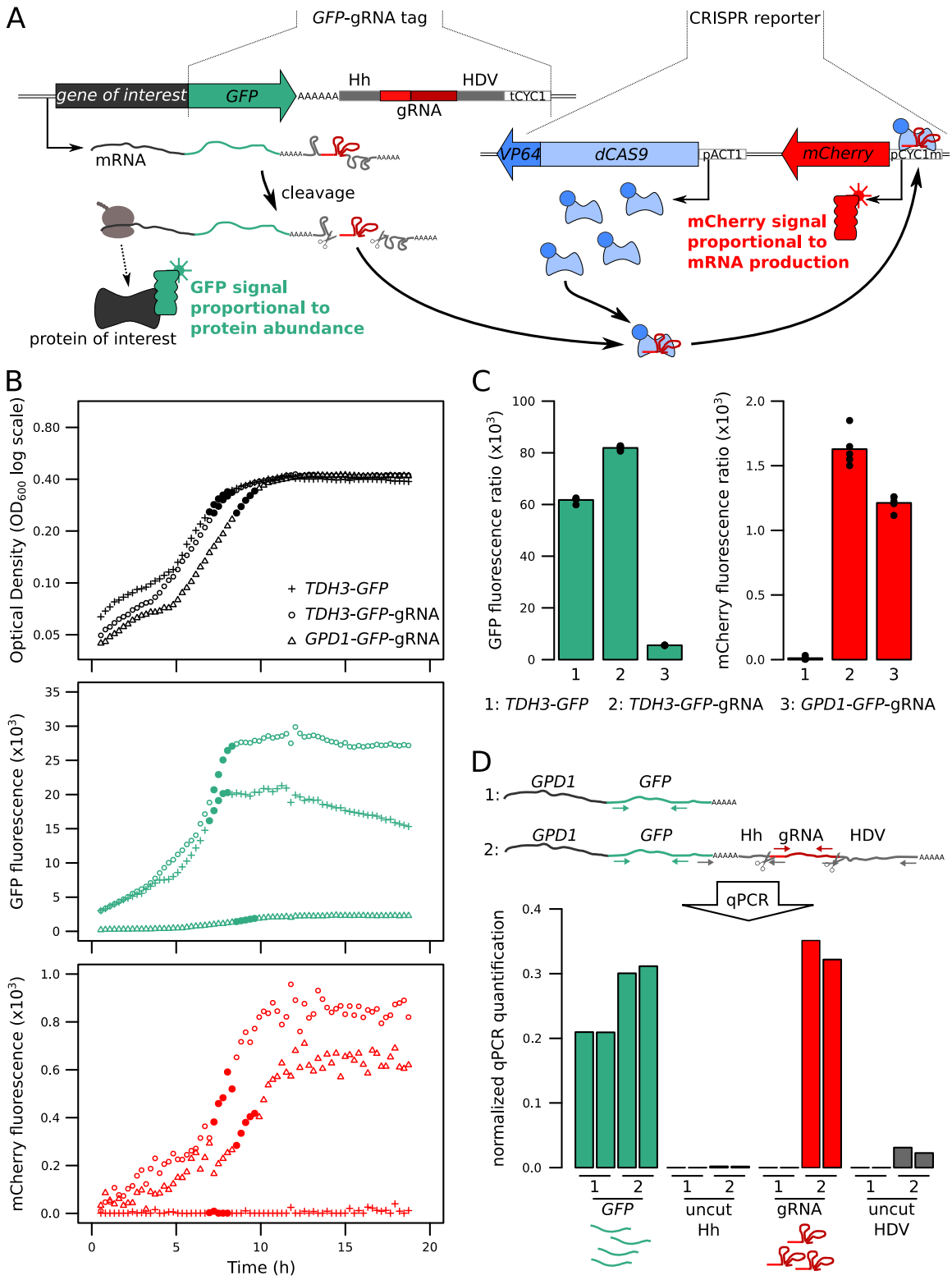

Figure 1. Fluorescence-based quantification of mRNA and protein levels. (A) Schematic of the dual quantification reporter. Hh: Hammerhead ribozyme, HDV: Hepatitis Delta Virus ribozyme, tCYC1: terminator sequence from the *CYC1* gene, *VP64*: four consecutive sequences encoding viral protein transcription activators VP16, pACT1: promoter sequence from the *ACT1* gene, pCYC1m: modified promoter sequence from the *CYC1* gene without baseline transcriptional activity. (B) Time courses of cell density and fluorescence measurements for three tagged strains during 20 h growth on a plate reader. Filled symbols correspond to five measurements at the end of the exponential growth phase that were used for

calculating fluorescence ratios for strain comparisons in the same physiological context as shown in panel C. (C) Fluorescence ratios (fluorescence / OD600) for the three strains shown in panel B. (D) RNA Quantification of the individual components of the tag, for *TDH3-GFP* and *TDH3-GFP-gRNA* by RT-qPCR. The two bars per strain show biological replicates. Normalized qPCR quantifications were calculated separately for each primer pair based on calibration with known template DNA amounts (Figure S2). Cells were grown in YNB glutamate medium.

mCherry fluorescence provides a quantitative readout of mRNA production

To characterize the quantitative response of our reporter system to a range of gene expression levels, we used the synthetic Z3EV system, which allows quantitative regulation of transcription via the concentration of estradiol in the culture medium (McIsaac et al., 2013). We cloned the Z3EV promoter upstream of a *GFP*-gRNA sequence (Figure 2A) in a strain that also contained the CRISPR reporter and grew this strain in a range of estradiol concentrations. Along with the expected increase in green fluorescence (McIsaac et al., 2013), red fluorescence increased as a monotonic function of estradiol concentration (Figure 2B). Similar results were observed in the RM11-1a strain, which has a different genetic background than BY (Figure S3). Thus, mCherry provides a quantitative readout of the expression of the tagged gene.

While green fluorescence continued to increase throughout the tested estradiol range, red fluorescence ceased to increase at concentrations of more than 4 nM estradiol (Figure 2B). qPCR quantification of the gRNA showed that mCherry fluorescence followed gRNA abundance (Figure 2C), confirming that the mCherry reporter gene is quantitatively regulated by gRNA abundance. gRNA abundance was linearly related to *GFP* mRNA and GFP fluorescence at lower doses of estradiol but stopped increasing at higher doses (Figure 2D & E). This suggests that mCherry production is limited by gRNA availability at high expression levels. Increasing the concentration of dCas9 proteins or binding sites for the gRNA had no effect on the mCherry expression plateau (Figure S4 & S5).

The linear relationship between mCherry fluorescence and mRNA abundance of the tagged gene was present up to an expression level that corresponded to half of the abundance of *ACT1* mRNA, which we had used as a reference gene in qPCR (Figure 2D). Based on previous RNAseq data (Albert et al., 2018), we estimated that 95% of *S. cerevisiae* genes fall below this threshold and can thus be

quantified by our mRNA reporter (Figure S6, Table S1). For lowly expressed genes, the GFP tag does not provide a strong enough signal to enable protein quantification (Huh et al., 2003) (Figure S6). Based on these results, we concluded that our dual reporter system can be used to simultaneously measure mRNA and protein of more than half the genes in the *S. cerevisiae* genome.

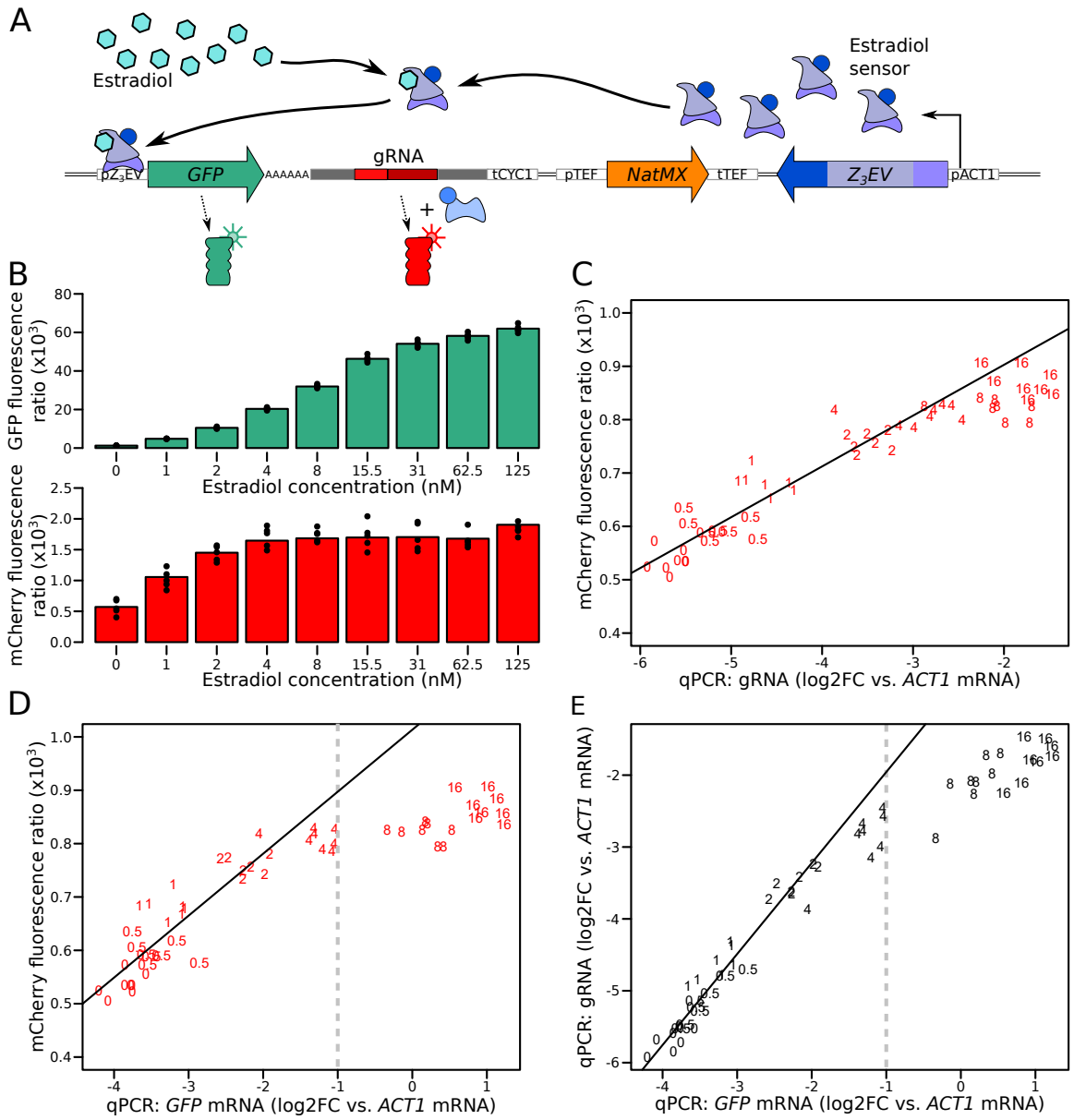

Figure 2. Characterization of the quantitative response of the reporter using an inducible transcriptional activator system to tune the expression of a *GFP* gene tagged with the gRNA. (A) Schematic of the Z_3EV system used in this experiment. (B) Fluorescence as a function of increasing estradiol concentrations. (C) Comparison of gRNA abundance (qPCR) and mCherry fluorescence in increasing estradiol concentrations. qPCR quantifications were normalised across samples using *ACT1* cDNA as a reference. log2FC: $\log_2$ of fold-change (D) Comparison of mRNA abundance (qPCR) and mCherry fluorescence in increasing estradiol concentrations. (E) Comparison of mRNA abundance to gRNA abundance in increasing

estradiol concentrations. The numbers in C to E show the concentration of estradiol in mM, with 7 to 8 biological replicates per concentration. Solid lines represent linear regressions calculated on measurements taken at up to 2 mM estradiol. Dashed vertical lines correspond to the mRNA abundance threshold, below which we deemed the reporter to be quantitative. Cells were grown in SC medium.

Simultaneous mapping of genetic variation affecting mRNA and protein levels

Our reporter system quantifies mRNA production and protein abundance at the same time, in the same live cells, exposed to the same environment. These features enable mapping of the genetic basis for variation in mRNA and protein levels, free from environmental or experimental confounders. We selected ten genes for genetic mapping (Table S2), based on several criteria. Five genes (*ARO8*, *BMH2*, *GPD1*, *MTD1*, *UGP1*) had previously been reported to have discrepant sets of mRNA-QTLs (Albert et al., 2018) and protein-QTLs (Albert et al., 2014b). Three genes (*CYC1*, *OLE1*, *TPO1*) had shown high agreement between their respective mRNA-QTLs and protein-QTLs. The remaining two genes (*CTS1* and *RPS10A*) had low protein abundance based on GFP-tag quantification (Huh et al., 2003) compared to their mRNA levels (Albert et al., 2018).

To identify genetic loci affecting mRNA production and protein abundance, we used the strains BY4741 (BY), a reference strain frequently used in laboratory experiments, and RM11-1a (RM), a vineyard isolate closely related to European strains used in wine making. These two strains differ at 47,754 variants in the yeast genome. We engineered RM to carry the CRISPR reporter inserted at the *NPR2* gene and a synthetic genetic array (SGA) marker for selection of *MAT*a haploid strains (Tong and Boone, 2007) at the neighboring *CAN1* gene. We engineered a series of BY strains, each carrying one gene tagged with the GFP-gRNA tag (Figure 3). We crossed these BY strains to the RM strain and obtained populations of recombinant haploid progeny carrying both the tagged gene and the CRISPR reporter. Flow cytometry detected a range of GFP and mCherry signals from single cells (Figure 3).

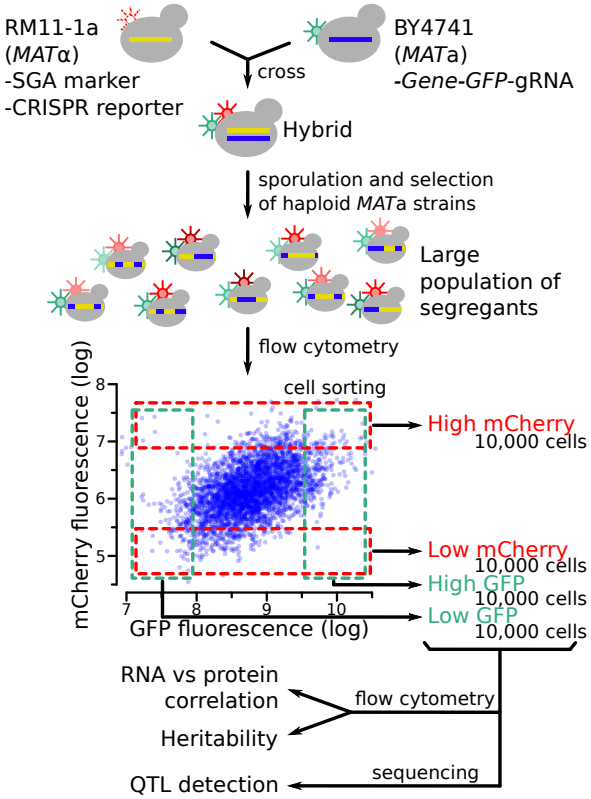

Figure 3. Schematic of the workflow for the identification of RNA-QTLs and protein-QTLs. The SGA marker allows for the selection of haploid *MAT*a strains after sporulation (Methods).

To study the relationship between mRNA and protein among single cells, we examined the cell-to-cell correlation between mCherry and GFP fluorescence in our genetically heterogeneous populations (Figure S7A). After correcting for cell size (Figure S8), mCherry and GFP were positively correlated for all tested genes (Figure S7B). The strength of the correlation varied from gene to gene. Lower correlations between mCherry and GFP were observed for the genes with high published discrepancies between mRNA-QTLs and protein-QTLs compared to those with more concordant mRNA-QTLs and protein-QTLs. Thus, different genes are influenced by mRNA-specific or protein-specific variation to different degrees.

Fluorescence-based readouts of mRNA and protein quantification in single cells enabled the use of bulk segregant analysis, a genetic mapping paradigm that gains statistical power from the analysis of millions of cells (Albert et al., 2014b; Ehrenreich et al., 2010). In each of the segregating populations, we used FACS to collect four subpopulations of 10,000 cells with high or low GFP or mCherry fluorescence, respectively, at a cutoff of 3% – 5% (Figure 3). In prior work, similarly stringent

selection provided high power for QTL mapping (Albert et al., 2014b; Ehrenreich et al., 2010; Parts et al., 2014).

To gauge the heritability of gene expression among single cells, we measured fluorescence between the high and low populations after 13 generations of growth. In almost all cases, the sorted populations showed significant (T-test, $p < 10\text{-}5$) differences in fluorescence, confirming the presence of genetic variants affecting mRNA and protein levels (Figure S9).

To map QTLs, we performed pooled whole-genome sequencing of all collected populations, computed the allele frequency of each DNA variant in each population, and calculated the difference in allele frequency (ΔAF) between high and low populations along the genome. A significant ΔAF at a locus indicated the presence of one or more genetic variants affecting protein abundance (GFP) or mRNA production (mCherry, Table S3). QTL mapping was performed in two to six biological replicates for all but one gene (*RPS10A*). Because any allele frequency differences among replicate populations sorted on the same parameters (e.g. two high GFP populations for the same gene) represent false positives, we used the replicate data to estimate false discovery rates. We chose a significance threshold (logarithms of the odds; "LOD" = 4.5) corresponding to a false discovery rate of 7% (Figure S10). Between replicates, 76% of the protein-QTLs and 78% of the mRNA-QTLs were reproducible at genome-wide significance (Figure 4A).

Across the ten genes, we detected 78 protein-QTLs and 44 mRNA-QTLs (Tables S4 & S5). By design, all detected loci were *trans*-acting, and most were located on a different chromosome than the tagged gene. One locus located at ~450 kb on chromosome XIV affected mCherry levels in the same direction in all ten genes. This region was also observed in a control experiment, in which mCherry was expressed constitutively using an *ACT1* promoter, and without a gRNA present (Figure S11). This region harbors the *MKT1* gene, which carries a variant affecting a variety of traits (Deutschbauer and Davis, 2005; Fay, 2013). While the highly pleiotropic *MKT1* locus may truly affect all ten genes we tested, it could also affect mCherry fluorescence via mCherry maturation or degradation, independently of any tagged gene. We excluded this region from further analyses.

The number of protein-QTLs per gene identified here (median = 7) agrees well with results from a previous study using the same mapping strategy (median = 8 for the same genes; (Albert et al., 2014b)), confirming that individual proteins are influenced by multiple, *trans*-acting loci. The effects

of individual protein-QTLs showed a positive correlation across studies (Pearson r = 0.73, p-value < 10-15, Figure 4B). The number of mRNA-QTLs per gene in our study (median = 3 after removing the *MKT1* locus) was lower than those from a previous study using RNA sequencing in 1,012 segregants (median = 8 for the same genes; Albert et al. 2018). This difference could be due to using our reporter in single cells with high stochastic variation compared to RNA-Seq in individually grown segregant cultures in Albert *et al.* (2018) (see Discussion). However, while the mRNA-QTLs detected by our reporter primarily reflect influences on mRNA production, the eQTLs from Albert *et al.* (2018) may reflect effects on transcription as well as mRNA degradation, which our system was not designed to capture. The effects of mRNA-QTLs were significantly correlated between studies (r = 0.44, p-value = 5×10-6, Figure 4C). Some of the QTLs we detected harbored genes known to affect expression variation. For example, a region at ~650 kb on chromosome XII that contained the gene *HAP1* affected protein abundance and / or mRNA production of *GPD1*, *CYC1*, *OLE1*, and *TPO1* (Figure 4A). In the BY strain, the *HAP1* coding sequence is interrupted by a transposon insertion, which alters the expression of thousands of mRNAs in *trans* (Albert et al., 2018; Brem et al., 2002). Overall, these agreements with previous analyses confirmed the reliability of our new reporter as a means for mapping the genetic basis of gene expression variation.

We detected several QTLs that were not shared with prior work and *vice versa* (Figure 4B – C). Most of these QTLs tended to have small effect sizes, suggesting that they could have been missed due to incomplete power in either study. Alternatively, these QTLs may reflect experimental differences between studies. For example, we observed a new, strong protein-QTL affecting Aro8 on chromosome XIV. The regulation of Aro8 expression by amino acid levels (Iraqui et al., 1998) suggests that this QTL could be due to the synthetic complete medium used here vs. YNB medium in earlier work.

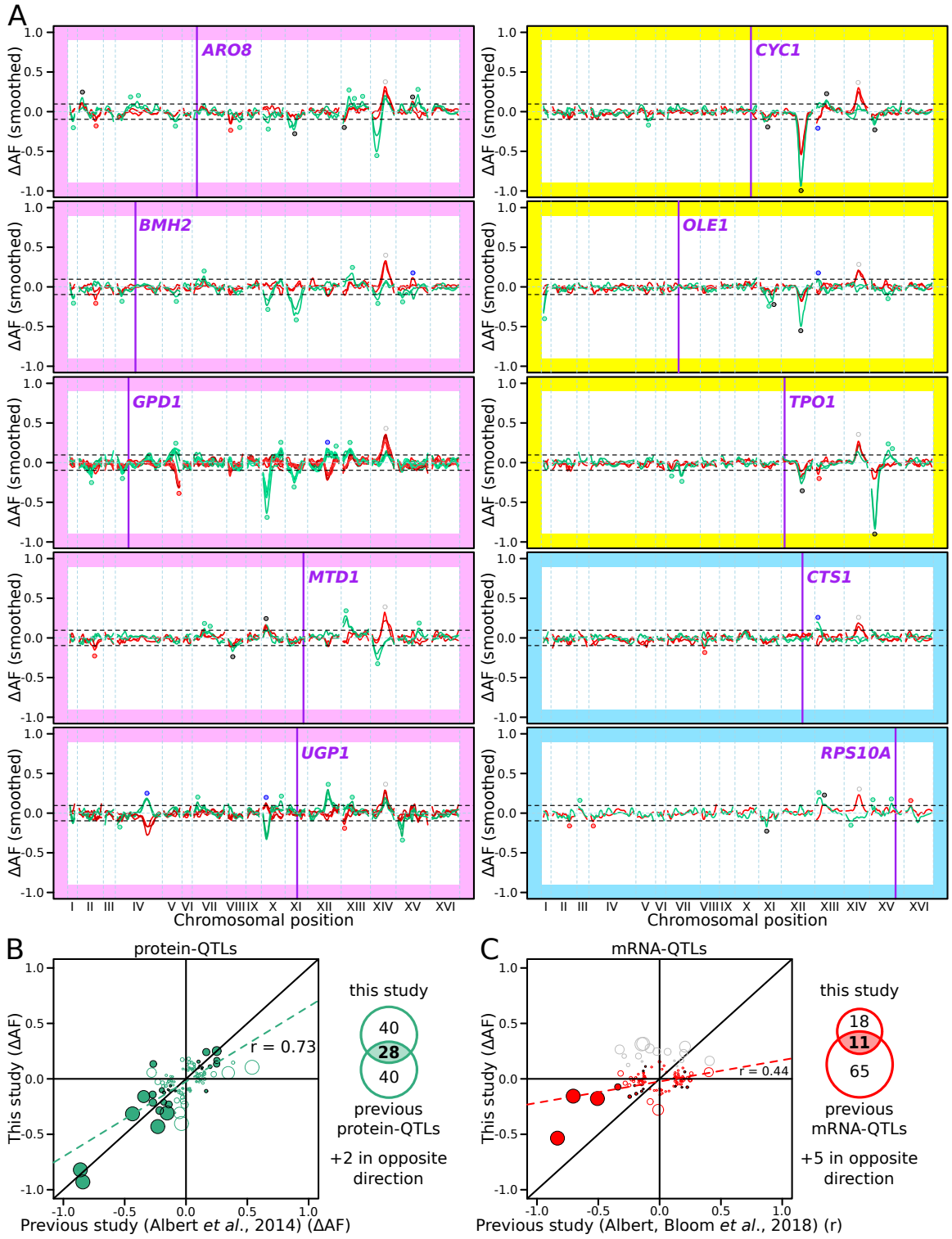

Figure 4. RNA-QTLs and protein-QTLs. (A) Allele frequency difference along the genome between the high and low population for each of the ten tagged genes, with 1 – 6 replicates per gene. Green and red curves correspond to the populations sorted on GFP and mCherry fluorescence, respectively. The colored plot borders indicate the reason for which the gene was chosen for study. Pink: high discrepancy between reported eQTLs and pQTLs, yellow: high similarity between reported eQTLs and pQTLs, blue: high mRNA level and low GFP

fluorescence. Purple vertical lines indicate the position of the tagged gene in the genome. Points indicate the location of significant QTLs, color coded based on protein or mRNA specificity (black: shared effect in same direction, green: protein specific, red: mRNA specific, blue: discordant). (B) Comparison between the effect size (ΔAF) of protein-QTLs identified in this study and pQTLs from previous work. (C) Same as (B) but comparing mRNA-QTLs identified in this study (ΔAF) to *trans*-eQTLs from previous work (eQTL effect sizes from Albert et al., 2018 are shown as a Pearson correlation coefficient between mRNA abundance and genotype at the QTL marker). Filled circles correspond to QTLs significant in both datasets. Empty circles correspond to QTLs significant in only one dataset. Grey circles correspond to QTLs located on chromosome XIV between 350 and 550 kb, which were excluded from analysis. Circle size is proportional to the LOD score of the QTL. The Venn diagrams show the total number and overlap of QTLs detected across the 10 genes between studies.

Differences between mRNA-QTLs and protein-QTLs

Genetic mapping using our reporter enabled us to compare mRNA-QTLs and protein-QTLs, free from environmental or experimental confounders. We classified 86 loci based on the presence and effect direction of their respective mRNA-QTLs and / or protein-QTLs (Figure 5A & S12, Table S6).

Of these 86 loci, 16 affected mRNA and protein of a given gene in the same direction. Such loci are expected for variants that alter a gene's mRNA production such that, in the absence of other effects, they also result in a concordant effect on protein abundance.

A majority of the loci corresponded to protein-QTLs that did not overlap an mRNA-QTL. These 52 protein-specific QTLs may arise from variants that affect translation or protein degradation, without an effect on mRNA production.

There were eleven mRNA-QTLs that did not overlap with a protein-QTL and seven loci where mRNA-QTLs and protein-QTLs overlapped but had discordant effects. These two categories may occur when protein abundance and mRNA production of the same gene are regulated separately, through two different *trans*-acting pathways. These two pathways could be affected by two distinct but genetically linked causal variants at the same locus, or by a single variant with distinct pleiotropic effects on the two pathways. Alternatively, buffering mechanisms (Battle et al., 2015; Großbach et al.,

2019) may compensate for changes in mRNA production perfectly (resulting in an mRNA-specific QTL) or may overcompensate (resulting in a discordant QTL pair) (Figure S12).

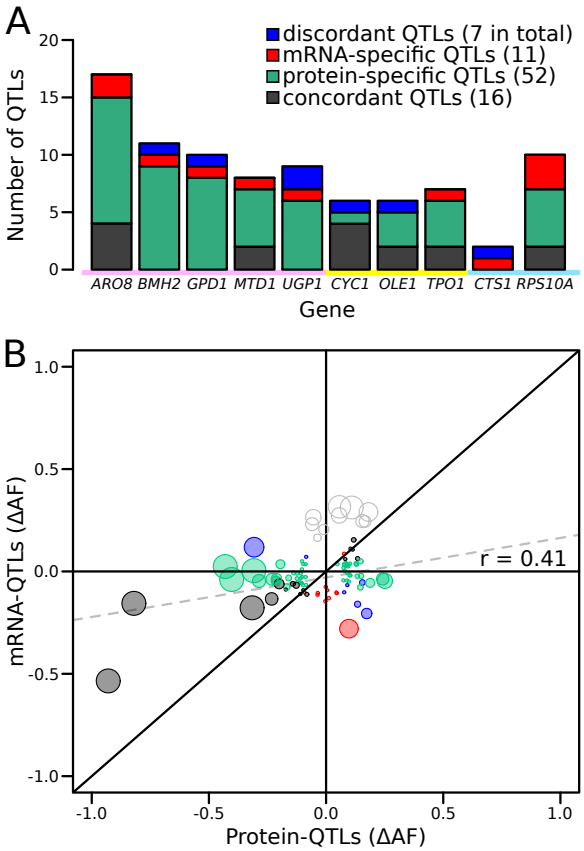

Figure 5. Comparison of RNA-QTLs and protein-QTLs. (A) Number of QTLs for each tagged gene, color coded according to type of effect on RNA and / or protein. (B) Comparison of QTL effect sizes between mRNA-QTLs and protein-QTLs. Grey circles correspond to QTLs located on chromosome XIV 350 – 550 kb, which were excluded from analysis. Circle size is proportional to the LOD score of the QTL.

Genes differed widely in the complexity and specificity of *trans*-acting loci that influenced their expression. For example, four genes (*BMH2*, *GPD1*, *UGP1*, and *CTS1*) were each influenced by multiple loci, none of which affected mRNA and protein levels in the same direction. By contrast, most of the loci influencing *CYC1* had concordant effects on mRNA and protein (Figure 5A).

While more than 73% of loci were specific for mRNA or protein, this difference might be inflated by loci that are truly concordant, but at which either the mRNA or the protein QTL narrowly failed to meet the significance threshold. To bypass this potential limitation, we compared effect sizes, expressed as ΔAF, at significant mRNA-QTLs or protein-QTLs, irrespective of the significance of the locus in the other data (Figure 5B). When considering all loci, we observed a significant, positive

correlation between mRNA and protein effect sizes (r = 0.41, p-value = 8.4×10-5, Figure 5B). This overall correlation was almost exclusively driven by the concordant QTL pairs, whose effect sizes showed a strong correlation (r = 0.88 p-value = 9×10-6). In sharp contrast, neither protein-specific QTLs (r = 0.2, p-value = 0.23) nor mRNA-specific QTLs (r = -0.05, p-value = 0.9) had correlated effects across the two data types, as expected if these loci specifically affected only mRNA or only protein. Considerable differences between mRNA-QTLs and protein-QTLs were also observed when simply considering effect directions. Overall, only 64% of QTLs affected mRNA and protein in the same direction. While this was more than the 50% expected by chance (binomial test p-value = 0.006), it left 36% of loci with discordant effects. Protein-specific QTLs showed similar directional agreement (63%) at lower significance (p-value = 0.04), while only 55% of mRNA-specific QTLs had an effect in the same direction in the protein data, which was not significantly different from chance (p-value = 0.5). Together, these results are consistent with the existence of many QTLs that specifically affect mRNA production or protein abundance.

Several loci were shared across the ten genes. Even these shared loci differed in the specificity of their effects on mRNA or protein. For example, the locus containing the *HAP1* gene had strong, concordant effects on both mRNA and protein for *CYC1* and *OLE1*, affected only the protein abundance of *UGP1*, and had significant but discordant effects on mRNA and protein for *GPD1*. Overall, these results revealed complex *trans*-acting influences on gene expression, in which genes were affected by different sets of multiple loci, with different degrees of mRNA or protein specificity.

A premature stop mutation in *YAK1* affects gene expression post-transcriptionally

The causal variants in most *trans*-acting loci are unknown, limiting our understanding of the underlying mechanisms. In particular, very few causal variants with specific *trans* effects on protein abundance are known (Chick et al., 2016; Hause et al., 2014). We noticed a region at ~155 kb on chromosome X that affected the protein abundance but not mRNA production of *ARO8*, *BMH2,* and especially *GPD1* (Figure 4A). This region spanned about 20 kb and contained 15 genes and 99 sequence variants. To identify the causal variant, we systematically divided this region into four tiles, swapped alleles in each tile using double cut CRISPR-swap, an efficient scarless genome editing

strategy (Lutz et al., 2019), and quantified the effect of these swaps on Gpd1-GFP fluorescence (Figure 6A – D).

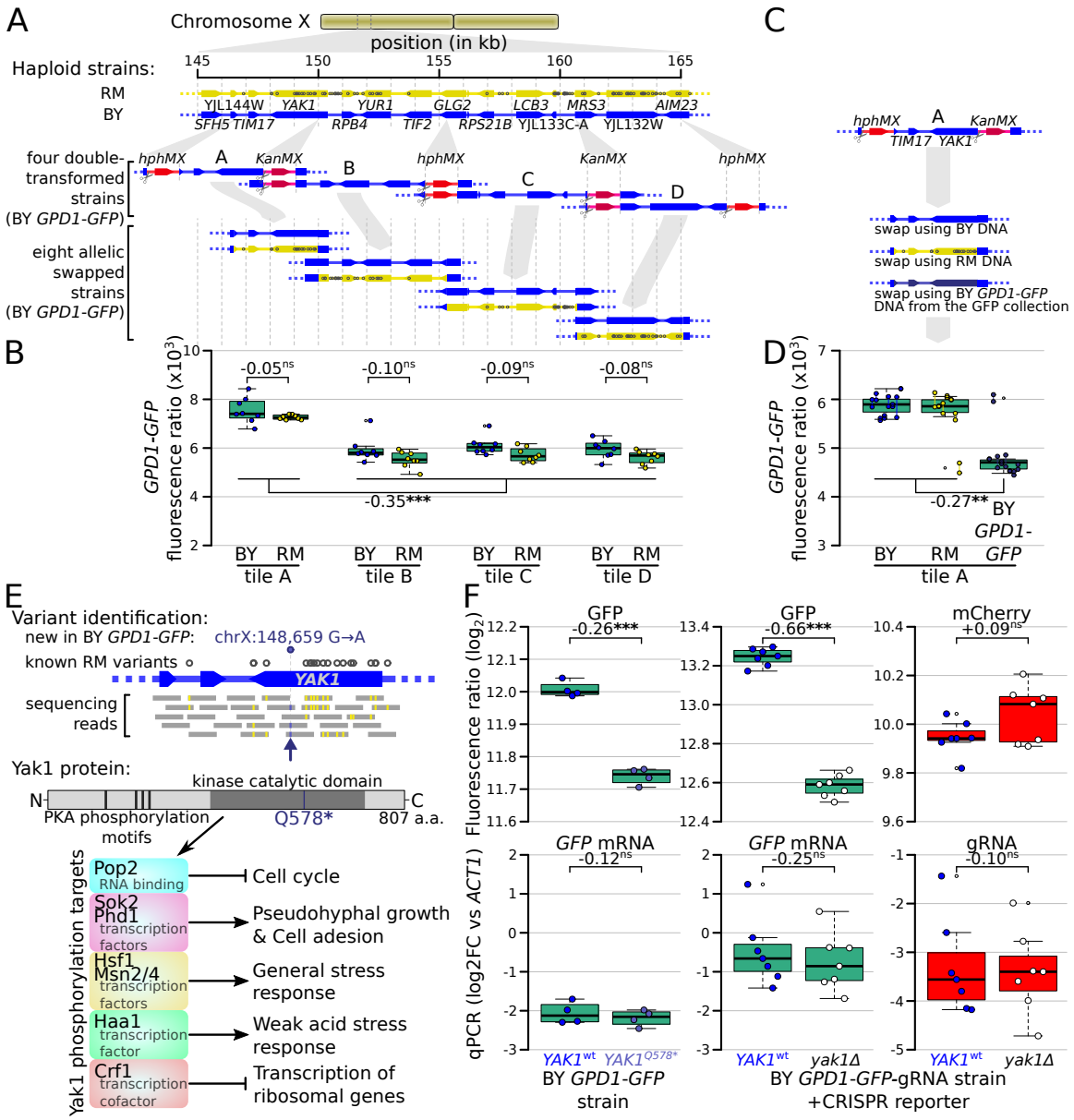

Figure 6. Identification of a causal variant influencing Gpd1-GFP protein but not *GPD1* mRNA. (A) Schematic of the investigated region and the strategy for generating tiled allele swaps across the region. Grey dots on the RM genome (yellow) indicate the positions of known BY / RM variants. (B) Boxplots comparing Gpd1-GFP fluorescence between allele swaps (6 – 8 replicates per swap). While none of the swaps resulted in a difference in fluorescence between BY and RM alleles, replacement with both backgrounds in tile A generated a significant increase in green fluorescence. Based on this result, we suspected that the BY *GPD1-GFP* strain carried a new mutation that was absent from the RM as well as BY genomic DNA used for the replacement. (C) Schematic of using BY *GPD1-GFP* DNA as a repair template for the allelic swap of tile A. (D) Boxplots comparing Gpd1-GFP fluorescence between the three swaps of tile A. The DNA repair template from the BY *GPD1-GFP* strain

resulted in low Gpd1-GFP fluorescence, suggesting a new mutation in the BY *GPD1-GFP* strain. (E) Identification of the $YAK1^{Q578*}$ mutation using sequencing data from the segregant population, and location of $YAK1^{Q578*}$ in the Yak1 kinase protein sequence. Selected known protein phosphorylation targets of Yak1 and downstream processes are indicated. (F) Effect of $YAK1^{Q578*}$ and *YAK1* knockout on Gpd1-GFP expression. Top: fluorescence, bottom: RNA quantified by qPCR. Numbers atop the boxplots correspond to log2(fold change). log2FC: $\log_2$ of fold-change. Stars indicate the significance of a t-test: ns: not significant ($p > 0.05$); ✷: $0.005 < p < 0.05$; ✷✷: $0.0005 < p < 0.005$; ✷✷✷: $p < 0.0005$. Cells were grown in SC medium.

This strategy, followed by analysis of our segregant population sequencing data, pinpointed a single G→A variant at 148,659 bp in the *YAK1* gene as the causal variant. While this variant is present in neither the BY nor RM reference genomes (Figure 6E & S13), our sequence data showed it to be present in all BY derivatives we used from the GFP collection (specifically, strains tagged at *ARO8*, *BMH2*, *GPD1*, *MTD1*, and *UGP1*; Figure S13) (Huh et al., 2003). We observed this variant in two additional strains we genotyped from the GFP collection (*FAA4* and YMR315W) and all four strains we genotyped from the tandem affinity purification (TAP)-tagged collection (*PGM1*, *NOT5*, *EMI2* and *TUB1*) (Ghaemmaghami et al., 2003). This variant was not present in a BY4741 strain that we obtained from the ATCC stock center (#201388), suggesting that the *YAK1* variant arose very recently in the specific BY4741 strain used to construct both the GFP and TAP-tagged collections. *YAK1* encodes a protein kinase involved in signal transduction in response to starvation and stress, indirectly regulating the transcription of genes involved in various pathways (Figure 6E). The causal variant changes the 578th codon (glutamine) to a premature stop codon that is predicted to disrupt translation of the Yak1 kinase domain (Figure 6E).

The *YAK1*Q578* variant led to a diminution of Gpd1-GFP fluorescence, suggesting a decrease of Gpd1-GFP protein abundance (Figure 6D). While *YAK1* may control transcription of genes in the glycerol biosynthesis pathway (Lee et al., 2008; Rep et al., 2000), which includes *GPD1*, our QTL results suggested no link between the variant and *GPD1-GFP* mRNA level. Consistent with a protein-specific *trans*-effect on *GPD1*, deletion of *YAK1* in a strain in which *GPD1* was tagged with GFP-gRNA caused a reduction of green fluorescence but had no detectable effect on mCherry fluorescence (Figure 6F). Further, qPCR indicated no difference in the level of *GPD1-GFP* mRNA in *YAK1*Q578* or *yak1Δ* compared to matched *YAK1*wt (Figure 6F).

We explored the genome wide effects of the *YAK1* variant with a differential analysis of mRNA and protein abundance, using RNA sequencing and mass spectrometry, respectively (Figure 7A, Tables S7, S8 & S9). Among 5,755 quantified mRNA transcripts, 262 were up-regulated and 310 down-regulated in the presence of *YAK1*Q578* (Benjamini-Hochberg (BH) adjusted p-value < 0.05) (Benjamini and Hochberg, 1995). The variant reduced the abundance of 82 of 2,590 quantified proteins, and increased another 82 proteins (BH adjusted p-value < 0.05). By comparing mass spectrometry and RNA sequencing results, we classified genes as affected only at the mRNA level (58 genes up, and 118 genes down-regulated), only at the protein level (60 genes up, and 50 genes down-regulated), or at both mRNA and protein (15 genes up, and 27 genes down-regulated). There was a strong enrichment for genes involved in cytoplasmic translation (q-value < 10-10) among genes with reduced mRNA abundance, which is consistent with the role of Yak1 as a regulator of transcription of ribosomal genes through Crf1 phosphorylation (Martin et al., 2004) (Figure S14, Table S10). Genes up-regulated at the mRNA level showed an enrichment in amino acid biosynthesis (q-value = 0.001). The most differentially expressed genes included known targets of the *YAK1* pathway (Figure 7A – B). Gpd1 protein was strongly reduced (BH adjusted p-value < 0.004), with a non-significant effect on *GPD1* mRNA (adjusted p-value = 0.10) (Figure 7A).

Finally, we investigated if the *YAK1*Q578* mutation affects other phenotypes. As *YAK1* and *GPD1* are involved in osmotic stress resistance (Lee et al., 2008), we grew strains carrying *YAK1*wt, *YAK1*Q578*, and *yak1Δ*, in a range of sodium chloride concentrations (Figure S15A). While this osmotic stress reduced growth, strains with *YAK1*Q578* and *yak1Δ* had a higher growth rate than wild-type, consistent with the role of Yak1 in triggering cell cycle arrest in response to stress. Gpd1-GFP abundance increased with stronger osmotic stress in *YAK1*wt and *yak1Δ*, with consistently lower expression of Gpd1-GFP in *yak1Δ* (Figure S15B-C).

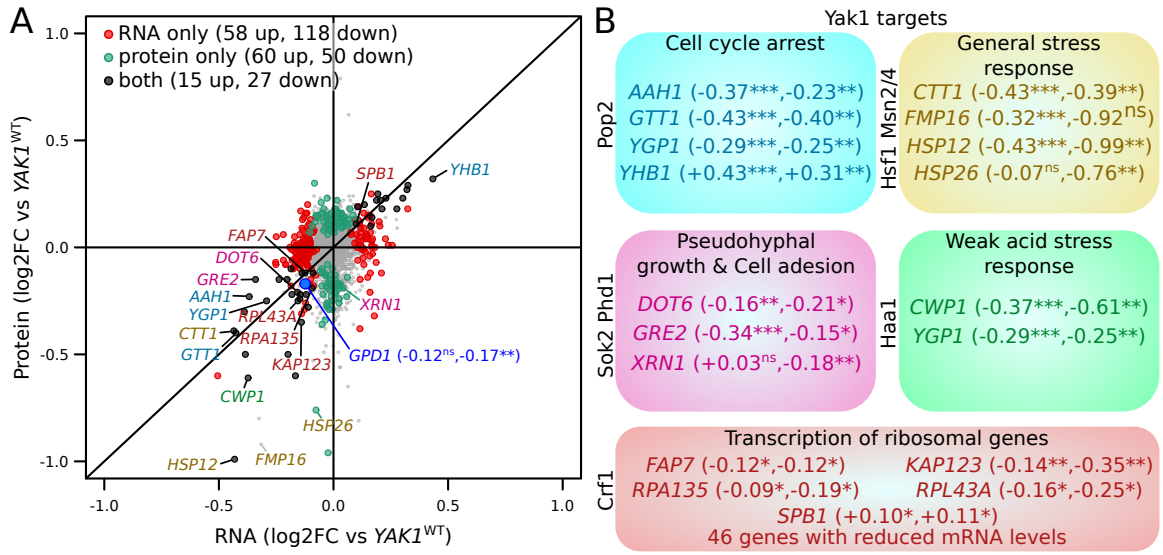

Figure 7. Effect of *YAK1*^{Q578*} on gene expression. (A) Effect on mRNA levels and protein levels quantified by RNA sequencing and mass spectrometry, respectively. Genes are colored according to their function as indicated in B). *GPD1* is highlighted in blue. (B) Examples of differentially expressed genes related to processes downstream of Yak1 phosphorylation regulation. The two numbers following gene names correspond to the $\log_2$ of fold-change (log2FC) of differential expression for mRNA and protein abundance, respectively. Stars indicate the significance of differential expression (Benjamini-Hochberg adjusted p-values). ns: not significant, $p > 0.05$; ✷: $0.005 < p < 0.05$; ✷✷: $0.0005 < p < 0.005$; ✷✷✷: $p < 0.0005$. Cells were grown in SC medium.

Discussion

We developed a fluorescence-based dual reporter system for the simultaneous quantification of mRNA and protein from a given gene in single, live cells. This system enabled the use of a statistically powerful mapping strategy to identify genetic loci that affected mRNA production or protein abundance in *trans*. Because mRNA and protein were quantified in the same exact condition, we were able to compare mRNA-QTLs and protein-QTLs without environmental confounding. Most *trans*-eQTLs have smaller effects that are more sensitive to environmental changes than *cis*-eQTLs (Smith and Kruglyak, 2008). Therefore, the high level of environmental control afforded by our method is particularly important for studying *trans*-acting variation.

We identified 86 *trans*-acting loci that contributed to variation in the expression of ten genes. The fact that 84% of these loci did not have concordant effects on mRNA production and protein abundance demonstrated the importance of variants that act on specific layers of gene expression.

The genes *ARO8*, *BMH2*, *GPD1*, *MTD1*, and *UGP1*, which we had selected for high discrepancy between previous mRNA-QTLs and protein-QTLs, showed many QTLs (89%) that were not shared between mRNA and protein in our data. In contrast, *CYC1*, *OLE1*, and *TPO1*, which we had selected for higher agreement between published QTLs, showed fewer discrepant QTLs in our data, although even for these genes the majority of QTLs was not shared between mRNA and protein (58%). These three genes showed fewer QTLs overall and all had one locus with strong effect size (Figure 4A; the *HAP1* locus for *CYC1* and *OLE1*, and the *IRA2* locus (Brem et al., 2002; Smith and Kruglyak, 2008) for *TPO1*). Based on these results, strong *trans*-acting loci may be more likely to cause concordant effects on mRNA and protein, while loci with smaller effects could be more likely to be specific to mRNA or protein.

While half of the mRNA-QTLs we detected had concordant effects on protein (16 out of 34), most protein-QTLs had no effects on mRNA (52 out of 75), in line with observations from prior work (Albert et al., 2018, 2014b; Foss et al., 2011). That 70% of our protein-QTLs had protein-specific effects suggests that the causal variants underlying many of these loci affect post-transcriptional processes. The indirect nature of our mCherry reporter and its lower signal intensity compared to GFP fluorescence are potential sources of measurement noise, which could have reduced detection power for mRNA-QTLs compared to protein-QTLs. However, our analyses of the magnitudes and directions of effects on mRNA and protein, which did not require loci in the other data to meet a significance threshold, also suggested that many protein-QTLs specifically influence protein.

We found seven loci that had discordant effects on mRNA production and protein abundance of the same gene. For example, at the *HAP1* locus, the BY allele increased Gpd1 protein abundance but decreased *GPD1* mRNA production, as had been seen when comparing QTLs across different studies (Albert et al., 2018). The highly pleiotropic effects of *HAP1* on mRNA and protein levels of many genes (Albert et al., 2018, 2014b; Smith and Kruglyak, 2008) reinforces the hypothesis that *HAP1* alleles influence Gpd1 protein abundance and mRNA production via two separate *trans*-acting mechanisms.

460 QTLs with discordant effects on mRNA and protein, as well as mRNA-specific QTLs, may be caused
461 by buffering of mRNA variation at the protein level. A well studied example of this phenomenon is the
462 regulation of expression of genes that encode members of a protein complex, in which excess protein
463 molecules that cannot be incorporated in the complex are rapidly degraded (Taggart et al., 2020).
464 Among the genes we investigated, *RPS10A* encodes a part of the ribosome small subunit complex.
465 *RPS10A* showed the highest number of mRNA-specific QTLs, possibly because Rps10a is subject to
466 buffering mechanisms.

467 The nonsense mutation (Q578*) we identified in the *YAK1* gene provides an informative example of
468 the complexity with which *trans*-acting variants can shape the transcriptome and the proteome.
469 *YAK1*Q578* changed protein abundance for many genes more strongly than mRNA abundance, but
470 also affected mRNA but not protein for many other genes. Thus, the consequences of this mutation
471 span mechanisms that affect mRNA as well as protein-specific processes. A reduction of ribosomal
472 gene transcription may account for some of these observations by reducing the translation of multiple
473 genes.

474 The *YAK1*Q578* variant likely arose as a new mutation in the BY4741 ancestor of the GFP and TAP-
475 tagged collections. Its relatively large effect and rarity in the global yeast population are consistent
476 with population genetic expectations (Eyre-Walker, 2010; Gibson, 2012) and observations (Bloom et
477 al., 2019; Fournier et al., 2019) for a deleterious variant that may have drifted to fixation in this
478 specific background, as has been suggested for many causal variants in natural yeast isolates
479 (Warringer et al., 2011). Alternatively, faster growth of a strain carrying the *YAK1*Q578* variant during
480 osmotic stress (Figure S15) may have contributed to adaptive fixation of this variant in this specific
481 copy of BY4741. While the large effect of *YAK1*Q578* aided our ability to fine-map it (Rockman,
482 2012), we suspect that its diverse, mRNA-specific as well as protein-specific mechanistic
483 consequences may be representative of more common *trans*-acting variants with smaller effects.

484 To simultaneously quantify mRNA and protein and eliminate potential environmental confounders in
485 expression QTL mapping, we developed a system in which a gRNA drives CRISPR activation of a
486 fluorescent reporter gene in proportion to a given mRNA of interest. Standard methods for mRNA
487 quantification require lysis of cell cultures or tissues, constraining sample throughput and statistical
488 power for mapping regulatory variation. Single-cell RNA sequencing (Picelli, 2017; Tang et al., 2009)

or *in situ* fluorescent hybridization (Buxbaum et al., 2014; Player et al., 2001; Rouhanifard et al., 2018) are improving rapidly, including in yeast (Gasch et al., 2017; Li and Neuert, 2019; Nadal-Ribelles et al., 2019; Wadsworth et al., 2017). However, these approaches still have throughput that is orders of magnitude below that available in FACS. Further, they involve cell lysis or fixation, precluding bulk segregant approaches that rely on growing cells after sorting. By contrast, our reporter allows quantification of mRNA production of a given gene within millions of single, live cells by flow cytometry. Because mCherry production in our system amplifies mRNA abundance signals, it is able to quantify genes with mRNA levels that would likely be hard to detect by FACS using *in situ* hybridization methods. The readout of our system is driven by a gRNA after it detaches from its mRNA. Therefore, the resulting signal is independent of the fate of the mRNA after gRNA release. Given the hammerhead ribozyme has a rate constant for self cleavage of 1.5 per minute (Wurmthaler et al., 2018), gRNA abundance is not expected to reflect the half-life and stability of most yeast mRNAs, which have a median half life of 3.6 minutes (Chan et al., 2018). By contrast, standard methods usually used in eQTL mapping quantify mRNA at steady state, which may explain some of the differences we observed between our mRNA-QTLs and known eQTLs identified by RNA sequencing.

Future versions of our reporter could be improved in several ways. The mCherry used here has a maturation time of about 40 minutes (Merzlyak et al., 2007), which limits the temporal resolution at which we can observe dynamic expression changes. Fluorescent proteins with faster maturation times could enable precise investigation of rapid temporal change in mRNA production. Using brighter fluorescent proteins or multiple copies of mCherry and its gRNA-targeted promoter could increase fluorescence and increase mRNA detection further. Finally, we observed that beyond a certain mRNA level, the abundance of gRNA no longer follows that of the tagged mRNA. Nevertheless, we estimated that our CRISPR based reporter can be used to quantify the mRNA production of most *S. cerevisiae* genes. Because CRISPR activation has been demonstrated in many organisms (Long et al., 2015; Maeder et al., 2013; Park et al., 2017), similar reporters for RNA production could be developed in other species.

Our reporter system for quantifying mRNA and protein of a given gene in the same live, single cells combined with a mapping strategy with high statistical power was deliberately designed to minimize technical or environmental confounders that may inflate differences between the genetics of mRNA

519 and protein levels. Yet, fewer than 20% of the detected loci had concordant effects on mRNA and
520 protein levels, providing strong support for the existence of discrepancies between genetic effects on
521 mRNA vs proteins. The fact that the majority of QTLs identified here were protein-specific suggests
522 that protein abundance is under more complex genetic control than mRNA abundance.

523 **Materials and methods**

524 **Yeast strains**

525 We used 160 yeast strains, 12 of which were obtained from other laboratories, including 6 from the
526 GFP collection, and 148 that were built for this study (complete list in Table S11). All strains are based
527 on two distinct genetic backgrounds: BY4741 (BY), which is closely related to the commonly used
528 laboratory strain S288c, and RM11-1a (RM), a haploid offspring of a wild isolate from a vineyard,
529 which is closely related to European strains used in wine-making. Both strains carried auxotrophic
530 markers, and RM had been engineered earlier to facilitate laboratory usage (BY: *his3Δ1 leu2Δ0*
531 *met15Δ0 ura3Δ0*; RM: *can1Δ::STE2pr-URA3 leu2Δ0 HIS3(S288C allele) ura3Δ ho::HYG AMN1(BY*
532 *allele)*; Table S11). Most strains were built using conventional yeast transformation (Gietz and
533 Schiestl, 2007) and DNA integration based on homologous recombination. Integrated DNA fragments
534 were produced by PCR (Phusion Hot Start Flex NEB M0535L, following manufacturer protocol,
535 annealing temperature: 57°C, 36 cycles, final volume: 50 µl) and gel purified (Monarch DNA Gel
536 Extraction Kit, NEB T1020L), with primers carrying 40 to 60 bp overhanging homologous sequence
537 as required. All primers are available in Table S12. For transformation, fresh cells from colonies on
538 agar plates were grown in YPD media (10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose) overnight at
539 30°C. The next day, 1 ml of the culture was inoculated in an Erlenmeyer flask containing 50 ml of
540 YPD and grown under shaking at 30°C for 3 hours to reach the late log phase. Cells were harvested by
541 centrifugation and washed once with pure sterile water and twice with transformation buffer 1 (10 mM
542 TrisHCl at pH8, 1 mM EDTA, 0.1 M lithium acetate). We resuspended the cells in 100 µl of
543 transformation buffer 1, added 50 µg of denatured salmon sperm carrier DNA (Sigma #D7656) and 1
544 µg of the DNA fragment to be integrated, and incubated for 30 minutes at 30°C. Alternatively, when
545 transforming a replicative plasmid, we used 0.1 µg of plasmid DNA and skipped this first incubation.
546 We added 700 µl of transformation buffer 2 (10 mM TrisHCl at pH8, 1 mM EDTA, 0.1 M lithium

acetate, 40% PEG 3350) and performed a second incubation for 30 minutes at 30°C. A heat shock was induced by incubating the cells at 42°C for 15 minutes. The transformed cells were then washed twice with sterile water. If the selective marker for the transformation was an antibiotic resistance gene, the cells were resuspended in 1 ml of YPD, allowed to recover for 2 hours at 30°C, and spread on a YPD plate (2% agar) containing the antibiotic (200 ng/l G418, 100 ng/l nourseothricin sulfate/CloNAT, or 300 ng/l hygromycin B). Alternatively, if the transformation was based on complementation of an auxotrophy, the cells were resuspended in 1 ml of sterile water and spread on a plate containing minimal media lacking the corresponding amino acid or nucleotide (YNB or Synthetic Complete (SC): 6.7 g/l yeast nitrogen base (VWR 97064-162), 20 g/l glucose, with or without 1.56 g/l SC -arginine -histidine -uracil -leucine (Sunrise science 1342-030), complemented as needed with amino acids: 50 mg/l histidine, 100 mg/l leucine, 200 mg/l uracil, 80 mg/l tryptophan). After two to three days of incubation at 30°C, colonies were streaked on a fresh plate containing the same selection media to purify clones arising from single, transformed cells. DNA integration in the correct location was confirmed by PCR (Taq DNA Polymerase NEB M0267L, following manufacturer protocol, annealing temperature: 50°C, 35 cycles, final volume: 25 µl, primers in Table S12). To store the constructed strains, we regrew the validated colony on a new selection media plate overnight at 30°C, scraped multiple colonies, resuspended the cells in 1.4 ml of YPD containing 20% glycerol in a 2 ml screw cap cryo tube and froze them at -80°C.

Plasmids

We constructed seven plasmids: three plasmids that do not replicate in yeast and that carry the GFP-gRNA tag, the CRISPR reporter, and Z3EV system, respectively (Figure S1), and four yeast-replicating plasmids to investigate the quantitative properties of our reporter (Figure S5). These plasmids were constructed through multiple rounds of cloning using DNA fragments from yeast DNA or plasmids acquired from Addgene (kind gifts from John McCusker: Addgene #35121-22, from Michael Nick Boddy: Addgene #41030, from Benjamin Glick: Addgene #25444, from Timothy Lu: Addgene #64381, #64389, #49013; complete list of plasmids in Table S13). Plasmids were assembled using Gibson assembly (NEBuilder HiFi DNA Assembly Cloning Kit, NEB E5520S). Fragments were either PCR amplified with a least 15 bp overlap at each end (Phusion Hot Start Flex NEB M0535L,

manufacturer protocol, annealing temperature: 57°C, 36 cycles, final volume: 50 µl, primers in Table S12) or obtained by restriction digestion of already existing plasmids (also shown in Table S12).

The fragment encoding the gRNA tag, containing the two ribozymes and the gRNA sequence itself, was purchased as a 212-bp double-stranded DNA oligo from IDT (we used the "C3" gRNA from (Farzadfard et al., 2013), as it was reported to provide the highest reporter gene expression). The synthetic polyA tail following the GFP sequence (Figure 1A) was introduced by using a PCR primer containing 45 thymines in its overhang sequence (primer OFA0038 in Table S12). Fragments for assembly were purified using agarose electrophoresis and gel extraction (Monarch DNA Gel Extraction Kit, NEB T1020L). For assembly, the given fragments were mixed at equi-molar amounts of 0.2 – 0.5 pM in 10 µl. Assembly was done by addition of 10 µl of NEBuilder HiFi DNA Assembly Master Mix and incubation at 50°C for 60 minutes. From this reaction, 2 µl of the final products were transformed into *E. coli* competent cells (10-beta Competent E.coli, NEB C3019I) through an incubation of 30 minutes on ice and a heat shock of 30 seconds at 42°C. Transformed cells were spread on LB plates (10 g/l peptone, 5 g/l yeast extract, 10 g/l sodium chloride, 2% agar) containing 100 mg/l ampicillin and grown overnight at 37°C. After cloning, the final plasmids were extracted (Plasmid Miniprep Kit, Zymo Research D4036) and verified by restriction enzyme digestion or PCR (Taq DNA Polymerase NEB M0267L, 30 cycles, 25 µl final volume, primers in Table S12). We also verified by Sanger sequencing that the gRNA tag in the plasmid had no mutation. To store the plasmids, the host bacteria were grown in LB with ampicillin overnight at 37°C and 1 ml of the culture was mixed with 0.4 ml of a sterile solution containing 60% water and 40% glycerol. The cells were stored at -80°C. The three plasmids containing the different parts of the reporter are available on Addgene (ID #157656, #157658, and #157659) along with their full DNA sequence.

Plate reader-based fluorescence measurements

Yeast fluorescence was measured in 24-hour time courses during overnight growth in a BioTek Synergy H1 plate-reader (BioTek Instruments). Fresh cells from agar plates were inoculated in 100 µl of minimal YNB media containing any complements necessary for growth of auxotrophic strains, at an initial optical density at wavelength 600 nm (OD600) of 0.05 in a 96-well flat bottom plate (Costar #3370). The plates were sealed with a Breathe Easy membrane (Diversified Biotech). Cells were

603 grown in the plate reader at 30°C and with circular agitation in between fluorescence acquisition.
604 During each acquisition, performed every 15 minutes, we recorded OD600, GFP fluorescence (read
605 from bottom, excitation 488 nm, emission 520 nm, 10 consecutive reads averaged, gain set to
606 "extended") and mCherry fluorescence (read from bottom, excitation 502 nm, emission 532 nm, 50
607 consecutive reads averaged, gain set to a value of 150). We took 97 measurements during 24 hours of
608 growth, unless individual runs were manually terminated early.

609 Raw measurements of OD600 and fluorescence were processed using R version 3.5.1 (https://www.r-
610 project.org/, scripts and raw data available at the github repository at
611 https://github.com/BrionChristian/Simultaneous_RNA_protein_QTLs). "Blank" values from wells
612 with no cells were subtracted from OD600 and fluorescence measurements of wells that had been
613 inoculated with cells. OD600 was log-transformed and manually inspected to identify the late log
614 phase, *i.e.* a time point about 3/4 into the exponential growth phase. This stage was identified
615 separately for each well, and usually corresponded to an OD600 of 0.1 – 0.3. We extracted the OD600
616 and fluorescence measurements at the five time points centered on our selected time point. The
617 mCherry and GFP fluorescence ratios were calculated as the ratio between the fluorescence and the
618 OD600 at these five time points (example in Figure 1B – C), allowing us to estimate fluorescence
619 while correcting for culture density. Focusing on the late log phase allowed measurements at higher
620 cell density to provide more robust fluorescence reads. Growth rates were estimated as the slope of a
621 linear fit of the log of OD600 over time.

622 **RNA quantification by qPCR**

623 *Cell harvest*

624 We quantified mRNA and gRNA abundance by quantitative real-time reverse-transcription PCR of
625 RNA extracted from exponentially growing cells. Cells were grown in either 50 ml of medium (YNB
626 with auxotrophic complements, results shown in Figure 1D) in shaking Erlenmeyer flasks or in 1.2 ml
627 of media (SC with auxotrophic complements and estradiol, Figure 2C – E & 6F) in a shaking 2-ml 96-
628 deep-well plate. The OD600 was monitored to identify the second half of the exponential growth
629 phase (corresponding to an OD600 of 0.35 – 0.45 OD600 in flasks, and 0.20 – 0.30 in the deep-well

plates). At this point, GFP and mCherry fluorescence ratios were recorded in a BioTek Synergy H1 plate reader (BioTek Instruments). Cells were then harvested immediately. Cells were washed with sterile water through either short centrifugation using 5 ml of culture from flasks, or vacuum-filtration through a 96-well filter plate (Analytical Sales 96110-10) using the entire remaining 1 ml of culture from the deep-well plate. Cells were then immediately flash-frozen in either isopropanol at -80°C (pellet from flask) or liquid nitrogen (filter plate) and stored at -80°C until RNA extraction.

RNA extraction from flasks

To extract the RNA from cells grown in flasks, we used the ZR Quick-RNA Kit (Zymo Research R1054). Frozen cell pellets were resuspended in 800 µl RNA Lysis Buffer 1 from the kit and transferred to a ZR BashingBead Lysis Tube. The cells were shaken in a mini-bead beater (BioSpec Products) for ten cycles of one minute in the beater, one minute on ice. Cell debris and beads were centrifuged for one minute at full speed and 400 µl of supernatants were transferred into Zymo-Spin IIIC Columns. The columns were centrifuged for one minute, and 400 µl 100% ethanol was added to the flow-through. After mixing, the flow-throughs were transferred into Zymo-Spin IIC Columns and centrifuged for one minute to bind the RNA and DNA to the columns. The columns were washed with 400 µl RNA Wash Buffer from the kit. DNA was digested in columns by adding a mixture of 5 µl DNase I and 75 µl DNA Digestion Buffer from DNase I Set kit (Zymo Research E1010) followed by a 15-minute incubation at room temperature. The columns were then washed three times with 400 µl RNA Prep Buffer, 700 µl RNA Wash Buffer, and 400 µl RNA Wash Buffer. RNA was eluted in 50 µl DNase/RNase-free water, quantified using Qubit RNA BR or HS Assay Kit (ThermoFisher Scientist Q10210 or Q52852), and stored at -20°C.

RNA extraction from 96-well plates

To extract the RNA from cells grown in 96-well plates, we used the ZR RNA in-plate extraction kit (ZR-96 Quick-RNA Kit, Zymo Research R1052), which followed the same protocol as the flask RNA extraction above, with a few minor differences. Bead-beating was done in an Axygen 1.1 ml plate (P-DW-11-C-S) with 250 µl of acid washed 425–600 µm beads (Sigma G8722) per well, sealed with an Axymat rubber plate seal (AM-2ML-RD-S). RNA purified from 200 µl of the resulting supernatant. DNA digestion and washing steps were done on Silicon-A 96-well plates from the kit. The RNA was eluted in 30 µl of DNase/RNase-free water, quantified, and stored at -20°C.

Reverse transcription and qPCR

RNA was reverse-transcribed using the GoScript RT kit (Promega A5000) following the kit protocol. We performed negative controls, no-enzyme and no-primer, which generated no qPCR signals. Quantitative PCRs were done in a 96-well plate (Bio-Rad HSP9645) using GoTaq qPCR kit (Promega A6001). Plates were sealed using a microseal 'B' Adhesive Seal (Bio-Rad MSB1001) and the reaction progress was recorded during 40 cycles using a C1000Touch plate reader (Bio-Rad). We quantified four different parts of the tag cDNA (GFP, Hh ribozyme cleavage, gRNA, and HDV ribozyme cleavage, Figure 1D), as well as *ACT1* cDNA as a reference gene. Primer sequences are in Table S12. The primers were tested and calibrated by running qPCR measurements on nuclear DNA extracts at a range of known input concentrations (Figure S2).

Segregant populations

BY strains (BY4741 background) carrying a given GFP-gRNA-tag and RM (YFA0198) carrying the CRISPR-reporter and the SGA marker were mixed for crossing on a plate with medium that allows only hybrids to grow (SC agar -leucine -histidine). Growing cells were streaked on the same medium, and a single hybrid colony was kept for storage and for generating the segregant population. For sporulation, hybrid strains were incubated in sporulation medium (2.5 g/l yeast extract, 2.5 g/l glucose, 15 g/l potassium acetate, 200 mg/l uracil, 100 mg/l methionine) at room temperature under vertical rotation in a glass tube for seven days. After verifying sporulation under a light microscope, 1 ml of medium containing the tetrads was pelleted (13,000 rpm for 5 minutes) and resuspended in 300 µl of sterile water containing about 15 µg of zymolyase. The resulting ascii were digested at 30°C for 30 minutes with agitation. Spores were separated by vortexing for about 15 seconds, and 700 µl of pure sterile water was added to the tube. We spread 250 µl of this spore suspension on a plate containing segregant selection media (SC agar, 50 mg/l canavanine, -uracil -leucine) allowing growth of haploid segregants carrying the following three alleles: (1) cells with mating type *MAT***a**, selected via the SGA-marker with *URA3* under control of the *STE2* promoter, which resulted in a ura+ phenotype only in *MAT***a** cells, (2) the SGA-marker integrated at the *CAN1* gene (whose deletion conferred canavanine resistance), which also selected for the CRISPR reporter that we had integrated at *NPR2*, the gene next to *CAN1*, (3) the given gene of interest tagged with the GFP-gRNA tag and *LEU2* selectable marker.

After three days of incubation at 30°C, segregants were harvested by scraping the entire plate in 10 ml of sterile water. Cells were centrifuged, resuspended in 3 ml of segregant selection media, and incubated at 30°C for 1.5 hours. To store these genetically diverse segregant populations, 1 ml of the culture was mixed with 0.4 ml of a sterile solution containing 60% water and 40% glycerol in a 2 ml screw-cap cryo tube and frozen at -80°C.

Cell sorting for QTL mapping

One day before cell sorting, the segregant population was thawed from the -80°C stock, mixed well, and 8 µl of culture were used to inoculate 5 ml of segregant selection media. The cells were reactivated with an overnight growth at 30°C under shaking. The next day, 1 ml of the growing culture was transferred to a new tube containing 4 ml of segregant selection media and grown for an additional two hours before cell sorting, roughly corresponding to the middle of the exponential growth phase.

Cell sorting was performed on a BD FACSAria II P0287 (BSL2) instrument at the University of Minnesota Flow Cytometry Resource (UFCR). Cells were gated to exclude doublet and cellular fragments. To focus on cells in approximately the same stage of the cell cycle, an additional gate selected cells in a narrow range of cell size as gauged by the area of the forward scatter signal (FSC). From the cells within this gate, we sorted five populations per experiment, each comprising 10,000 cells: (1) a control population from the entire gate without fluorescence selection, (2) the 3% of cells with the lowest GFP fluorescence, (3) the 3% of cells with the highest GFP fluorescence, (4) the 3% of cells with the lowest mCherry fluorescence, and (5) the 3% of cells with the highest mCherry fluorescence. Each population was collected into 1 ml of segregant selection medium. After overnight growth at 30°C, 0.9 ml of culture was mixed with 0.4 ml of a sterile solution containing 60% water and 40% glycerol, and frozen at -80°C until sequencing. The remaining 0.1 ml were inoculated into 0.9 ml of segregant selection medium and grown for 3 hours before analyzing the population using flow cytometry (see below). In total, we obtained 125 sorted populations from 25 experiments across the ten tagged genes, with 1 to 6 biological replicates per gene, as well as the untagged population (Table S2). Sorting was done in four batches on different dates. Biological replicates were performed

as independent sporulations of the stored diploid hybrids, and thus represent independent populations sorted in separate experiments.

Flow cytometry

Single cell fluorescence analysis was performed using cultures in the late log growth phase. We used a BD Fortessa X-30 H0081 flow cytometer at UFCR equipped with blue and yellow lasers and 505LP and 595LP filters to measure green (GFP) and red (mCherry) fluorescence, respectively. Forward scatter (FSC), side scatter (SSC), GFP, and mCherry fluorescence were recorded for 50,000 cells, excluding doublets and cellular debris. The voltaic gains were set as follows: 490 for FSC, 280 for SSC, 500 for GFP, and 600 for mCherry. We monitored for possible cross-contamination from cells retained in the instrument using strains expressing either only GFP or mCherry, and observed no cross-contamination. Recorded data on .fsc files were analysed using R and the flowCore package (Hahne et al., 2009). Raw data and scripts are accessible on github (https://github.com/BrionChristian/Simultaneous_RNA_protein_QTLs). The data were filtered to discard outlier cells based on unusual FSC and SSC signals. We used the fluorescence data from the sorted populations to determine correlations between red and green fluorescence, as well as heritability (Figure S7 & S8). For these analyses, fluorescence values were corrected for cell size (FSC) by calculating the residuals of a loess regression of fluorescence on FSC. Loess regression avoided the need to assume a specific mathematical relationship between the two parameters (Figure S9).

DNA extraction and sequencing

DNA extraction for whole genome sequencing was performed in 96-well plate format using E-Z-96 Tissue DNA kits (Omega D1196-01). The stored, sorted populations were thawed, mixed, and 450 µl transferred into a 2-ml 96-deep-well plate containing 1 ml of segregant selection medium for an overnight growth at 30°C. The plate was centrifuged for 5 minutes at 3700 rpm, and the supernatant was removed by quick inversion of the plate. Then, 800 µl of Buffer Y1 (182 g/l sorbitol, 0.5 M EDTA, pH 8, 14.3 mM β-mercaptoethanol, 50 mg/l zymolyase 100T) were added to the pellets, and the cells were resuspended and incubated for 2 hours at 37°C. The spheroplasts were centrifuged and

the supernatant discarded. The pellets were resuspended in 200 µl of TL buffer and 25 µl of OB Protease Solution from the kit and incubated overnight at 56°C. The next day, RNA was denatured by addition of 5 µl of RNAse A (20 mg/ml) and incubated at room temperature for 5 minutes. After addition of 450 µl of BL Buffer from the kit, the mixture was transferred onto a E-Z 96 column DNA plate and centrifuged at 3700 rpm for 3 minutes. The columns were washed once with 500 µl of HBC Buffer and three times with 600 µl of DNA Wash Buffer from the kit. After an additional centrifugation to dry the column, the DNA was eluted in 100 µl of pure sterile water, quantified using Qubit dsDNA HS Assay Kit (ThermoFisher Scientist Q32854) and stored at 4°C for library preparation the next day.

Library preparation for Illumina sequencing was performed using Nextera DNA Library Prep kit (Illumina) with modifications. The tagmentation was done on 5 ng of DNA using 4 µl of Tagment DNA buffer ("TD" in the kit) and 0.25 µl of Tagment DNA enzyme (corresponding to a 20-fold dilution of "TDE1" from the kit) and incubating for 10 minutes at 55°C. Fragments were amplified with index primers (8 Nextera primers i5 and 12 Nextera primers i7, for up to 96 possible multiplex combinations) on 10 µl tagmented DNA by adding 1 µl of each primer solution (10 µM), 5 µl of 10X ExTaq buffer and 0.375 µl of ExTaq polymerase (Takara) and water to a final volume of 50 µl. The amplification was run for 17 PCR cycles (95°C denaturation, 62°C annealing, 72°C elongation). 10 µl of each reaction were pooled for multiplexing and run on a 2% agarose gel. DNA that migrated between the 400 and 600 bp was extracted using Monarch DNA Gel Extraction Kit (NEB T1020L). The pooled library DNA concentration was determined using Qubit dsDNA BR Assay Kit (ThermoFisher Scientist Q32853), and submitted for sequencing. Sequencing was performed at the University of Minnesota Genomics Core (UMGC). Our 125 populations were processed in four batches extracted and sequenced at different times. Two were sequenced using an Illumina HiSeq 2500 (high-output mode; 50-bp paired-end) and two were sequenced using an Illumina NextSeq 500 (mid-output mode, 75-bp paired-end). Read coverage ranged from 5-fold to 24-fold coverage of the genome (median: 13-fold). The reads will be made available on NCBI SRA.

QTL Mapping

For each sorted and sequenced population, reads were filtered (MAPQ ≥ 30) and aligned to the *S. cerevisiae* reference genome (version sacCer3, corresponding to BY strain) using BWA ((Li and Durbin, 2009), command: *mem -t*). We used samtools ((Li et al., 2009), command: *view -q 30*) to generate bam files and collapse PCR duplicates using the *rmdup* command. We used 18,871 variants previously identified as polymorphic and reliable between RM and BY (Bloom et al., 2013; Ehrenreich et al., 2010) (list available on github: https://github.com/BrionChristian/Simultaneous_RNA_protein_QTLs, samtools: *mpileup -vu -t INFO/AD -l*), generating vcf files with coverage and allelic read counts at each position for each population.

The vcf files were processed in R to identify bulk segregant analysis QTLs using code adapted from Albert *et al.* 2014 (Albert et al., 2014b) (available on github: https://github.com/BrionChristian/Simultaneous_RNA_protein_QTLs). Briefly, for plotting the results, the allele frequency of the reference (that is, BY) allele was calculated at each position in each population. Random counting noise was smoothed using loess regression, and the allele frequency of a given "low" fluorescence population subtracted from its matched "high" fluorescence population to generate ΔAF. A deflection from zero indicated the presence of a QTL. To identify significant QTLs, we used an R script that implemented the MULTIPOOL algorithm (Edwards and Gifford, 2012), which calculates LOD score based on ΔAF and depth of read coverage in bins along the genome. We used MULTIPOOL output to call QTLs as peaks exceeding a given significance threshold (see below), along with confidence intervals for the peak location corresponding to a 2-LOD drop from the peak LOD value. We applied the MULTIPOOL algorithm using the following parameters: bp per centiMorgan: 2,200; bin size: 100 bp, effective pool size: 1,000. We excluded variants with extreme allele frequencies of > 0.9 or < 0.1. We initially set a permissive detection threshold of LOD > 3.0 to identify a set of candidate QTLs, which we then integrated across replicates (507 QTLs, Table S3). A second, more stringent, threshold of LOD > 4.5 was then applied to retain only significant QTLs based on our estimated false discovery rate (FDR).

To estimate FDR, we applied the multipool QTL detection algorithm to pairs of populations sorted into the same gates in different replicates. Any "QTLs" in such comparisons must be due to technical or biological noise. We restricted these analyses to replicates sequenced on the same instrument, resulting in 80 inter-replicate comparisons. From these data, we calculated the FDR as a function of the significance threshold (*thr*): FDR*thr* = (*N*repQTL*thr* / *N*rep) / (*N*fluoQTL*thr* / *N*fluo), where

797 *N*repQTL*thr* is the number of false "QTLs" from comparing the same gate across replicate populations
798 at a LOD score threshold of *thr*, *N*rep is the number of such inter-replicate comparisons (*N*rep = 80),
799 *N*fluoQTL*thr* is the number of fluorescence-QTLs at a LOD threshold of *thr*, and *N*fluo is the number
800 of high vs. low fluorescence comparisons (*N*fluo = 48; the untagged experiment was excluded). At a
801 significance threshold of LOD = 4.5, the estimated FDR was 7.3% (Figure S10), which we used to call
802 significant QTLs. For some overlap analyses (see below), we used a threshold of LOD = 3.0, which
803 corresponded to an FDR of 13%.

804 To call significant QTLs across replicates, we first scanned each replicate for QTLs at a permissive
805 threshold of LOD > 3.0. Second, at each resulting QTL peak position, we averaged ΔAF and LOD
806 scores across all available replicates without applying a LOD filter to each replicate. Third, we
807 collapsed groups of overlapping QTLs, which we defined as QTLs whose peaks were within 75,000
808 bp of each other in the different replicates. For each group of these overlapping QTLs, we averaged
809 the LOD scores, the ΔAFs, the peak positions, and the location confidence intervals to form one
810 merged QTL. Of the resulting merged QTLs, we retained those that exceeded our stringent
811 significance threshold of LOD ≥ 4.5.

812 To gauge reproducibility of these significant QTLs, we counted the number of replicates in which a
813 given QTL had been detected at the permissive LOD > 3.0, using the same definition of positional
814 overlap as above. The majority (74%) of significant QTLs were shared across all the corresponding
815 replicates. Two tagged genes had more than two replicates (*GPD1* and *UGP1*). For these genes,
816 requiring *all* replicates to be significant is conservative. Therefore, we also estimated the average
817 reproducibility of all mRNA-QTLs or all protein-QTLs by calculating the average fraction of
818 replicates that had a QTL at a given merged QTL:

819 $fraction\_overlap = mean[\ (NshareQTLij - 1) / (Nrepj - 1)\ ]$

820 Here, *NshareQTLij* is the number of replicates for which the QTL *i* is detected for the tagged gene *j* at
821 LOD > 3, and *Nrepj* is the number of replicates performed for the tagged gene *j*. Note that if only a
822 single replicate has a QTL at a given merged QTL, this fraction takes on a value of zero because in
823 such a case, there is no overlap among replicates at this QTL. The observed *fraction_overlap* was 0.76
824 for the protein-QTLs and 0.78 for the mRNA-QTLs.

825 **Comparison of mRNA-QTLs and protein-QTLs**

826 To compare mRNA-QTLs and protein-QTLs of the same gene, we first considered all merged QTLs
827 that exceeded a permissive threshold of LOD > 3.0 (after merging replicates as described above). We
828 considered an mRNA-QTL and a protein-QTL for the same gene with overlapping confidence
829 intervals as a QTL pair across mRNA and protein. We manually curated the result of this overlap
830 analysis for six cases; after curation, QTLs on chromosomes XV (*ARO8*), VIII (*MTD1*), XIII (*CYC1*)
831 and XIII (*RPS10A*) were considered to be pairs, and QTLs on chromosome V (*GPD1*) and XIV
832 (*MTD1*) were considered to be mRNA-specific.

833 From this initial set, we retained those QTL pairs at which the given mRNA and / or protein QTL met
834 a more stringent LOD score of > 4.5 (FDR = 7.2%). Applying this higher threshold only after the more
835 permissive overlap analysis allowed us to consider QTL pairs even if one of the paired QTLs did not
836 pass the strong significance threshold of LOD > 4.5. As an example, we considered overlapping QTLs
837 on chromosome XI that affected *OLE1* expression (mRNA-QTL LOD = 4.4, protein-QTL LOD =
838 15.5) to be a pair even if the mRNA LOD score was below the stringent significance threshold. In such
839 cases, we deemed it more conservative to assume that the weaker QTL exists but narrowly failed to
840 reach significance than to declare the stronger QTL as specific for mRNA or protein. We discarded all
841 QTLs located between 350 – 550 kb on chromosome XIV, as this region may affect mCherry
842 fluorescence independently of the tagged gene.

843 We distinguished four types of QTLs (Figure 5A & S12). The shared QTL pairs either had similar
844 effects on mRNA and protein abundance (16 QTL pairs, defined as having the same sign of ΔAF), or
845 discordant effects on mRNA and protein (7 QTL pairs, different sign of ΔAF). All QTLs that were not
846 part of a pair were considered to be specific (11 mRNA-specific QTLs, 52 protein-specific QTLs).

847 Finally, we conducted an analysis of mRNA or protein QTL effect sizes and directions that avoided
848 having to define potentially paired QTLs as significant or not. For each mRNA-QTL (or protein-QTL),
849 we extracted the ΔAF from the protein-QTL (or mRNA-QTL) data at the same exact position,
850 irrespective of significance in the other data. We used these values to compute correlations of effects

and to examine shared directionality of effects between mRNA-QTLs and protein-QTLs (Figure 5B, Table S6).

Allelic engineering for *YAK1* fine-mapping

To obtain strains with scarless allelic swaps in haploids, we used a strategy based on double-cut CRISPR swap (Lutz et al., 2019). We flanked each of the four tiles to be switched by two resistance markers (*hphMX* and *KanMX*) using our regular yeast transformation protocol (see above). The yeast were then transformed with 100 ng of CRISPR-Swap plasmid (pFA0055-gCASS5a , Addgene plasmid # 131774) and 1 µg of DNA repair template amplified either from BY, BY *GPD1-GFP*, or RM. The transformed cells were spread on SC -leucine plates, selecting for the presence of the plasmid expressing *CAS9* and a gRNA targeting and cleaving a sequence present in both of the two resistance cassettes. We used strains in which *GPD1* was tagged with *GFP* but not the gRNA tag, as the gRNA in our tag would likely have directed *CAS9* to cleave the mCherry promoter. Cleavage of both cassettes resulted in the region in between the resistance cassettes to be replaced by the repair template. Transformed clones were screened for the double loss of antibiotic resistance to identify those with successful editing.

We introduced the 148,659 G→A variant, which we had detected through sequence analysis (see below), in *YAK1* by single-cut CRISPR swap (Lutz et al., 2019). We replaced the *YAK1* sequence with a *hphMX* resistance cassette insertion to create *yak1Δ::hphMX*. We then delivered the CRISPR-Swap plasmid along with a repair template DNA produced by fusion PCR to carry either the G or A allele at the variant position (primers OFA0874 to OFA0881 in Table S12). Five clones of each allele, (*YAK1*wt and *YAK1*Q578*) were confirmed by Sanger sequencing (primers OFA0883 and OFA0882 in Table S12).

Sequence analyses to identify the *YAK1* and other new DNA variants

To search for new variants in our populations that were not known to be present in the BY and RM strains, we used the sequencing reads from each selected segregant population. In each of 125 populations, we analyzed bam files after collapsing PCR duplicates. We applied samtools *mpileup* (--

min-BQ 0) and bcftools *call* (*-vc*), either locally in the region affecting Gpd1-GFP (*-r chrX:146000-150000*) or on the whole genome to generate vcf files containing variant information. The vcf files were merged in R to generate matrices of polymorphic positions along with their allele frequency and coverage. The allele frequencies of the *YAK1* polymorphisms were plotted along the genome for each population (Figure S13). We excluded all 47,754 previously known BY / RM variants from the whole-genome polymorphism matrices, and also removed variants with a bcftools quality score below 30. Among the 7,624 remaining variants sites, 5,822 were identified in only one or two populations and were deemed to be sequencing errors. Only one variant was shared across most (71 out of 75) of the populations created from strains from the GFP collection (strains tagged at *ARO8*, *BMH2*, *GPD1*, *MTD1*, and *UGP1*; Table S2) and absent in all other populations. This variant was the 148,659 G-to-A SNV in *YAK1*.

***YAK1* genotyping**

The region containing the *YAK1*Q578* variant was PCR-amplified from genomic DNA isolated from strains carrying *FAA4*-GFP and YMR315W-GFP (GFP collection), *PGM1*-TAP, *NOT5*-TAP, *EMI2*-TAP, *TUB1*-TAP (TAP-tag collection), BY4741 (ATCC 201388), YLK1879 (a BY strain from the Kruglyak lab) and YLK1950 (an RM strain from the Kruglyak lab) using primers OFA0883 and OFA0882 (Table S12). The resulting PCR product was Sanger sequenced using OFA0883 to genotype the *YAK1* variant. The *YAK1*Q578* variant was observed only in the strains obtained from the GFP and TAP-tag collections (Ghaemmaghami et al., 2003; Huh et al., 2003).

Differential expression analysis by RNA sequencing and mass spectrometry

Cell harvest

We quantified RNA and protein from five biological replicates (different clones obtained after CRISPR-Swap) of *YAK1*wt and *YAK1*Q578*. For each of these 10 strains, fresh colonies were used to inoculate 5 ml of SC medium (completed with uracil, arginine, histidine, and leucine) and the culture was grown overnight at 30°C. The next day, 50 ml of SC media in an Erlenmeyer flask was inoculated with the overnight culture to an initial OD600 of 0.05 and were grown under shaking at 30°C. When

the OD600 reached 0.35 – 0.45 (late log phase), four aliquots of 7 ml of the culture was transferred to a falcon tube, centrifuged, and washed with 1 ml of sterile PBS buffer at 30°C (Phosphate Buffered Saline, pH 7.5). The pellets were immediately frozen in liquid nitrogen. For each strain, one cell pellet was used for RNA sequencing and another for protein mass spectrometry.

RNA extraction and sequencing

RNA extraction and library preparation were conducted as described in Lutz *et al*. 2019 (Lutz et al., 2019). Briefly, RNA extraction was done in two batches that each contained equal numbers of clones from the two groups (first batch: clones 1 and 2 of BY *YAK1*wt and clones 1 and 2 of BY *YAK1*Q578*, second batch: clones 3, 4 and 5 of BY *YAK1*wt and clones 3, 4 and 5 of BY *YAK1*Q578*). We used ZR Fungal/Bacterial RNA mini-prep kit with DNase I digestion (Zymo Research R2014), following the kit manual. RNA was eluted in 50 µl of RNase/DNase free water, quantified, and checked for integrity on an Agilent 2200 TapeStation. All RNA Integrity Numbers were higher than 9.5 and all concentrations were above 120 ng/µl. The RNA samples were stored at -20°C until use. Poly-A RNA selection was done using 550 ng of total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490L), processing all samples in one batch. We prepared the library using NEB Ultra II Directional RNA library kit for Illumina (NEB E7760) used dual index primers (from NEBNext Multiplex Oligos for Illumina, NEB E7600S) for multiplexing, and amplified the library for ten cycles. The libraries were quantified using Qubit dsDNA HS Assay Kit (ThermoFisher Scientist Q32854) and pooled at equal mass for sequencing using 75-bp single-end reads on Illumina NextSeq 550 at UMGC. The sequencing reads will be made available on NCBI GEO. Using the trimmomatic software (Bolger et al., 2014), reads were trimmed of adapters and low quality bases and filtered to be at least 36 bp long. Reads were then pseudo-aligned to the *S. cerevisiae* transcriptome (Ensembl build 93) and counted using kallisto (Bray et al., 2016). We used RSeQC to calculate Transcript Integrity Numbers (TIN) which provided an estimation of alignment quality for each gene of each sample. We excluded any gene with at least one read count of zero or at least one TIN of zero across the ten samples. After filtering, 5,755 genes remained for analysis. Differential expression analysis was performed using the DESeq2 R package (Love et al., 2014), using the extraction batch information as covariate. DESeq2 provided, for each gene, the log2-fold change (*YAK1*Q578* vs. *YAK1*wt) and the p-value adjusted for multiple-testing using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) (Table S7).

Protein extraction and mass spectrometry

Protein extraction and quantification using mass spectrometry was performed by the Center for Mass Spectrometry and Proteomics at the University of Minnesota. Briefly, cells from the pellets were lysed by sonication (30%, 7 seconds in Branson Digital Sonifier 250) in protein extraction buffer (7 M urea, 2 M thiourea, 0.4 M triethylammonium bicarbonate pH 8.5, 20% acetonitrile and 4 mM tris(2-carboxyethyl)phosphine). The proteins were extracted using pressure cycling (Barocycler Pressure Biosciences NEP2320, 60 cycles of 20 seconds at 20 kpsi and 10 seconds at 0 kpsi), purified by centrifugation in 8 mM iodoacetamide (10 minutes at 12000 rpm), and quantify by Bradford assay. For each sample, 40 µg of extracted proteins were fragmented by trypsin digestion (Promega Sequencing Grade Modified Trypsin V5111), and labeled by tandem mass tag isotopes (TMT10plex Isobaric Label Reagent Set, Thermo Scientific 90110). All the tagged samples were pooled. Peptides were separated first based on hydrophobicity through two consecutive liquid chromatographies, followed by separation based on mass per charge after ionisation in the first mass spectrometry step. The second mass spectrometry step after high-energy collision-induced dissociation allowed for the identification of the peptide and the quantification of the TMTs. MS-MS analysis was conducted on an Orbitrap Fusion Tribrid instrument (Thermo Scientific). Database searches were performed using Proteome Discoverer software, and post-processing and differential expression analysis was done using Scaffold 4.9. Differential expression statistics per protein were computed on mean peptide abundances after inter-sample normalisation. Normalised abundances for each of the 2,590 detected proteins are provided in Table S8. We adjusted p-values provided by Scaffold using the Benjamini-Hochberg method.

Protein and mRNA data comparison

To compare the effect of the *YAK1* variant on mRNA and protein abundance, we examined the log2 fold-change of the mRNA abundance (from DESeq2) and protein abundance (from Scaffold) for the 2,577 genes present in both datasets. We considered genes that were differentially expressed in mRNA or protein (adjusted p-value < 0.05) and that showed no evidence of difference in the other quantity (raw p-value > 0.05), to be specifically affected at the mRNA or protein level (Figure 7A, Table S9). In various categories of differentially expressed genes, we looked for gene ontology enrichment using

GOrilla (Eden et al., 2009) with the list of 2,577 genes detected in both mass spectrometry and RNAseq as the background set (Table S10).

Acknowledgments

We thank Scott McIsaac and Leonid Kruglyak for yeast strains, Joshua Bloom for implementing the MULTIPOOL algorithm in R, and Mahlon Collins for the mCherry/GFP protein fusion flow cytometry data. We thank the University of Minnesota Genomics Center, the University of Minnesota's Center for Mass Spectrometry and Proteomics, and the University of Minnesota's Flow Cytometry Resource. We thank members of the Albert lab for critical feedback on the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by NIH grant R35-GM124676 and a Sloan Research Fellowship (FG-2018-10408) to FWA.

Supplementary Material

Supplementary_Table.xlsx

Supplementary_Figures.pdf

References

978 Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park YoSon, Parsana P, Segrè
979 AV, *et al.* 2017. Genetic effects on gene expression across human tissues. *Nature* **550**:204–
980 213. doi:10.1038/nature24277
981 Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. 2018. Genetics of trans-regulatory variation in
982 gene expression. *eLife* **7**:e35471. doi:10.7554/eLife.35471
983 Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nat Rev*
984 *Genet* **16**:197–212. doi:10.1038/nrg3891
985 Albert FW, Muzzey D, Weissman JS, Kruglyak L. 2014a. Genetic influences on translation in yeast.
986 *PLoS Genet* **10**:e1004692. doi:10.1371/journal.pgen.1004692
987 Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. 2014b. Genetics of single-cell protein
988 abundance variation in large yeast populations. *Nature* **506**:494–497. doi:10.1038/nature12904
989 Andrie JM, Wakefield J, Akey JM. 2014. Heritable variation of mRNA decay rates in yeast. *Genome*
990 *Res* **24**:2000–2010. doi:10.1101/gr.175802.114
991 Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, Gilad Y. 2015. Impact of regulatory
992 variation from RNA to protein. *Science* **347**:664–667. doi:10.1126/science.1260793
993 Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful
994 Approach to Multiple Testing. *J R Stat Soc Ser B Methodol* **57**:289–300.
995 Bloom JS, Boocock J, Treusch S, Sadhu MJ, Day L, Oates-Barker H, Kruglyak L. 2019. Rare variants
996 contribute disproportionately to quantitative trait variation in yeast. *eLife* **8**:e49212.
997 doi:10.7554/eLife.49212
998 Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. 2013. Finding the sources of missing
999 heritability in a yeast cross. *Nature* **494**:234–237. doi:10.1038/nature11867
1000 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
1001 *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170
1002 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification.
1003 *Nat Biotechnol* **34**:525–527. doi:10.1038/nbt.3519
1004 Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in
1005 budding yeast. *Science* **296**:752–755. doi:10.1126/science.1069516
1006 Brion C, Caradec C, Pflieger D, Friedrich A, Schacherer J. 2020. Pervasive Phenotypic Impact of a
1007 Large Nonrecombining Introgressed Region in Yeast. *Mol Biol Evol* **msaa101**.
1008 doi:10.1093/molbev/msaa101
1009 Buxbaum AR, Wu B, Singer RH. 2014. Single β-Actin mRNA Detection in Neurons Reveals a
1010 Mechanism for Regulating Its Translatability. *Science* **343**:419–422.
1011 doi:10.1126/science.1242939
1012 Chan LY, Mugler CF, Heinrich S, Vallotton P, Weis K. 2018. Non-invasive measurement of mRNA
1013 decay reveals translation initiation as the major determinant of mRNA stability. *eLife*
1014 **7**:e32536. doi:10.7554/eLife.32536
1015 Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. 2005. Mapping
1016 determinants of human gene expression by regional and genome-wide association. *Nature*
1017 **437**:1365–1369. doi:10.1038/nature04244
1018 Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL,
1019 Churchill GA, Gygi SP. 2016. Defining the consequences of genetic variation on a proteome-
1020 wide scale. *Nature* **534**:500–505. doi:10.1038/nature18270
1021 Clément-Ziza M, Marsellach FX, Codlin S, Papadakis MA, Reinhardt S, Rodríguez-López M, Martin
1022 S, Marguerat S, Schmidt A, Lee E, Workman CT, Bähler J, Beyer A. 2014. Natural genetic

variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. *Mol Syst Biol* **10**:764.

Damerval C, Maurice A, Josse JM, de Vienne D. 1994. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* **137**:289–301.

Deutschbauer AM, Davis RW. 2005. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat Genet* **37**:1333–1340. doi:10.1038/ng1674

Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**. doi:10.1126/science.1258096

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**:48. doi:10.1186/1471-2105-10-48

Edwards MD, Gifford DK. 2012. High-resolution genetic mapping with pooled sequencing. *BMC Bioinformatics* **13 Suppl 6**:S8. doi:10.1186/1471-2105-13-S6-S8

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**:1039–1042. doi:10.1038/nature08923

Eyre-Walker A. 2010. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci* **107**:1752–1756. doi:10.1073/pnas.0906182107

Farzadfard F, Perli SD, Lu TK. 2013. Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synth Biol* **2**:604–613. doi:10.1021/sb400081r

Fay JC. 2013. The molecular basis of phenotypic variation in yeast. *Curr Opin Genet Dev* **23**:672–677. doi:10.1016/j.gde.2013.10.005

Foss EJ, Radulovic D, Shaffer SA, Goodlett DR, Kruglyak L, Bedalov A. 2011. Genetic Variation Shapes Protein Networks Mainly through Non-transcriptional Mechanisms. *PLOS Biol* **9**:e1001144. doi:10.1371/journal.pbio.1001144

Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Bedalov A, Goodlett DR, Kruglyak L. 2007. Genetic basis of proteome variation in yeast. *Nat Genet* **39**:1369–1375. doi:10.1038/ng.2007.22

Fournier T, Abou Saada O, Hou J, Peter J, Caudal E, Schacherer J. 2019. Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. *eLife* **8**:e49258. doi:10.7554/eLife.49258

Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z, Wang B, Zhai L, Dai C, Xu J, Wang W, Li X, Zheng J, Chen L, Luo L, Liu J, Qian X, Yan J, Wang J, Wang G. 2013. RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat Commun* **4**:2832. doi:10.1038/ncomms3832

Gao Y, Zhao Y. 2014. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J Integr Plant Biol* **56**:343–349. doi:10.1111/jipb.12152

Gasch AP, Yu FB, Hose J, Escalante LE, Place M, Bacher R, Kanbar J, Ciobanu D, Sandor L, Grigoriev IV, Kendziorski C, Quake SR, McClean MN. 2017. Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLOS Biol* **15**:e2004050. doi:10.1371/journal.pbio.2004050

Gerrits A, Li Y, Tesson BM, Bystrykh LV, Weersing E, Ausema A, Dontje B, Wang X, Breitling R, Jansen RC, Haan G de. 2009. Expression Quantitative Trait Loci Are Highly Sensitive to Cellular Differentiation State. *PLOS Genet* **5**:e1000692. doi:10.1371/journal.pgen.1000692

Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature* **425**:737–741. doi:10.1038/nature02046

Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, Park CC, Wen P-Z, Brewer H, Weitz K, Ii DGC, Pan C, Yordanova R, Neuhaus I, Tilford C, Siemers N, Gargalovic P, Eskin E, Kirchgessner T, Smith DJ, Smith RD, Lusis AJ. 2011. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLOS Genet* **7**:e1001393. doi:10.1371/journal.pgen.1001393

Gibson G. 2012. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**:135–145. doi:10.1038/nrg3118

Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**:31–34. doi:10.1038/nprot.2007.13

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS. 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**:647–661. doi:10.1016/j.cell.2014.09.029

Großbach J, Gillet L, Clément-Ziza M, Schmalohr CL, Schubert OT, Barnes CA, Bludau I, Aebersold R, Beyer A. 2019. Integration of transcriptome, proteome and phosphoproteome data elucidates the genetic control of molecular networks. *bioRxiv* 703140. doi:10.1101/703140

Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang T-P, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin S-Y, Glass D, Travers M, Min JL, Ring S, Ho K, Thorleifsson G, Kong A, Thorsteindottir U, Ainali C, Dimas AS, Hassanali N, Ingle C, Knowles D, Krestyaninova M, Lowe CE, Di Meglio P, Montgomery SB, Parts L, Potter S, Surdulescu G, Tsaprouni L, Tsoka S, Bataille V, Durbin R, Nestle FO, O'Rahilly S, Soranzo N, Lindgren CM, Zondervan KT, Ahmadi KR, Schadt EE, Stefansson K, Smith GD, McCarthy MI, Deloukas P, Dermitzakis ET, Spector TD, Multiple Tissue Human Expression Resource (MuTHER) Consortium. 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**:1084–1089. doi:10.1038/ng.2394

Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, Spidlen J, Strain E, Gentleman R. 2009. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* **10**:106. doi:10.1186/1471-2105-10-106

Hasin-Brumshtein Y, Hormozdiari F, Martin L, van Nas A, Eskin E, Lusis AJ, Drake TA. 2014. Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics* **15**:471. doi:10.1186/1471-2164-15-471

Hasin-Brumshtein Y, Khan AH, Hormozdiari F, Pan C, Parks BW, Petyuk VA, Piehowski PD, Brümmer A, Pellegrini M, Xiao X, Eskin E, Smith RD, Lusis AJ, Smith DJ. 2016. Hypothalamic transcriptomes of 99 mouse strains reveal trans eQTL hotspots, splicing QTLs and novel non-coding genes. *eLife* **5**:e15614. doi:10.7554/eLife.15614

Hause RJ, Stark AL, Antao NN, Gorsic LK, Chung SH, Brown CD, Wong SS, Gill DF, Myers JL, To LA, White KP, Dolan ME, Jones RB. 2014. Identification and Validation of Genetic Variants that Influence Transcription Factor and Cell Signaling Protein Levels. *Am J Hum Genet* **95**:194–208. doi:10.1016/j.ajhg.2014.07.005

Higgins MG, Fitzsimons C, McClure MC, McKenna C, Conroy S, Kenny DA, McGee M, Waters SM, Morris DW. 2018. GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. *Sci Rep* **8**:14301. doi:10.1038/s41598-018-32374-6

1115 Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global
1116 analysis of protein localization in budding yeast. *Nature* **425**:686–691.
1117 doi:10.1038/nature02026
1118 Iraqui I, Vissers S, Cartiaux M, Urrestarazu A. 1998. Characterisation of *Saccharomyces cerevisiae*
1119 ARO8 and ARO9 genes encoding aromatic aminotransferases I and II reveals a new
1120 aminotransferase subfamily. *Mol Gen Genet MGG* **257**:238–248. doi:10.1007/s004380050644
1121 Ka S, Markljung E, Ring H, Albert FW, Harun-Or-Rashid M, Wahlberg P, Garcia-Roves PM, Zierath
1122 JR, Denbow DM, Pääbo S, Siegel PB, Andersson L, Hallböök F. 2013. Expression of carnitine
1123 palmitoyl-CoA transferase-1B is influenced by a cis-acting eQTL in two chicken lines selected
1124 for high and low body weight. *Physiol Genomics* **45**:367–376.
1125 doi:10.1152/physiolgenomics.00078.2012
1126 Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E,
1127 Wiederkehr M, Gutierrez-Arcelus M, Panousis NI, Yurovsky A, Lappalainen T, Romano-
1128 Palumbo L, Planchon A, Bielser D, Bryois J, Padioleau I, Udin G, Thurnheer S, Hacker D,
1129 Core LJ, Lis JT, Hernandez N, Reymond A, Deplancke B, Dermitzakis ET. 2013. Coordinated
1130 Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription.
1131 *Science* **342**:744–747. doi:10.1126/science.1242463
1132 Kita R, Venkataram S, Zhou Y, Fraser HB. 2017. High-resolution mapping of cis-regulatory variation
1133 in budding yeast. *Proc Natl Acad Sci* **114**:E10736–E10744. doi:10.1073/pnas.1717421114
1134 Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N,
1135 Gootenberg JS, Nishimasu H, Nureki O, Zhang F. 2015. Genome-scale transcriptional
1136 activation by an engineered CRISPR-Cas9 complex. *Nature* **517**:583–588.
1137 doi:10.1038/nature14136
1138 Lahtvee P-J, Sánchez BJ, Smialowska A, Kasvandik S, Elsemman IE, Gatto F, Nielsen J. 2017.
1139 Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-
1140 Specific Translation Efficiency in Yeast. *Cell Syst* **4**:495-504.e5.
1141 doi:10.1016/j.cels.2017.03.003
1142 Lee P, Cho B-R, Joo H-S, Hahn J-S. 2008. Yeast Yak1 kinase, a bridge between PKA and stress-
1143 responsive transcription factors, Hsf1 and Msn2/Msn4. *Mol Microbiol* **70**:882–895.
1144 doi:10.1111/j.1365-2958.2008.06450.x
1145 Li G, Neuert G. 2019. Multiplex RNA single molecule FISH of inducible mRNAs in single yeast cells.
1146 *Sci Data* **6**:94. doi:10.1038/s41597-019-0106-6
1147 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
1148 *Bioinformatics* **25**:1754–1760. doi:10.1093/bioinformatics/btp324
1149 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000
1150 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and
1151 SAMtools. *Bioinforma Oxf Engl* **25**:2078–2079. doi:10.1093/bioinformatics/btp352
1152 Liu Y, Beyer A, Aebersold R. 2016. On the Dependency of Cellular Protein Levels on mRNA
1153 Abundance. *Cell* **165**:535–550. doi:10.1016/j.cell.2016.03.014
1154 Long L, Guo H, Yao D, Xiong K, Li Y, Liu P, Zhu Z, Liu D. 2015. Regulation of transcriptionally
1155 active genes via the catalytically inactive Cas9 in C. elegans and D. rerio. *Cell Res* **25**:638–
1156 641. doi:10.1038/cr.2015.35
1157 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq
1158 data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8

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

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013. CRISPR RNA–guided activation of endogenous human genes. *Nat Methods* **10**:977–979. doi:10.1038/nmeth.2598

Martin DE, Soulard A, Hall MN. 2004. TOR Regulates Ribosomal Protein Gene Expression via PKA and the Forkhead Transcription Factor FHL1. *Cell* **119**:969–979. doi:10.1016/j.cell.2004.11.047

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. 2012. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**:1190–1195. doi:10.1126/science.1222794

McCarthy JEG. 1998. Posttranscriptional Control of Gene Expression in Yeast. *Microbiol Mol Biol Rev* **62**:1492–1553.

McIsaac RS, Oakes BL, Wang X, Dummit KA, Botstein D, Noyes MB. 2013. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res* **41**:e57. doi:10.1093/nar/gks1313

Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, Gaintzeva A, Lukyanov KA, Lukyanov S, Gadella TWJ, Chudakov DM. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods* **4**:555–557. doi:10.1038/nmeth1062

Michelmore RW, Paran I, Kesseli RV. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci* **88**:9828–9832. doi:10.1073/pnas.88.21.9828

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**:743–747. doi:10.1038/nature02797

Nadal-Ribelles M, Islam S, Wei W, Latorre P, Nguyen M, de Nadal E, Posas F, Steinmetz LM. 2019. Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. *Nat Microbiol* **4**:683–692. doi:10.1038/s41564-018-0346-9

Pai AA, Cain CE, Mizrahi-Man O, Leon SD, Lewellen N, Veyrieras J-B, Degner JF, Gaffney DJ, Pickrell JK, Stephens M, Pritchard JK, Gilad Y. 2012. The Contribution of RNA Decay Quantitative Trait Loci to Inter-Individual Variation in Steady-State Gene Expression Levels. *PLOS Genet* **8**:e1003000. doi:10.1371/journal.pgen.1003000

Park J-J, Dempewolf E, Zhang W, Wang Z-Y. 2017. RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in Arabidopsis. *PLOS ONE* **12**:e0179410. doi:10.1371/journal.pone.0179410

Parts L, Liu Y-C, Tekkedil MM, Steinmetz LM, Caudy AA, Fraser AG, Boone C, Andrews BJ, Rosebrock AP. 2014. Heritability and genetic basis of protein level variation in an outbred population. *Genome Res* **24**:1363–1370. doi:10.1101/gr.170506.113

Picelli S. 2017. Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol* **14**:637–650. doi:10.1080/15476286.2016.1201618

1204 Picotti P, Clement-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW, Röst H, Sun Z, Rinner O,
1205 Reiter L, Shen Q, Michaelson JJ, Frei A, Alberti S, Kusebauch U, Wollscheid B, Moritz R,
1206 Beyer A, Aebersold R. 2013. A complete mass spectrometric map for the analysis of the yeast
1207 proteome and its application to quantitative trait analysis. *Nature* **494**:266–270.
1208 doi:10.1038/nature11835
1209 Player AN, Shen L-P, Kenny D, Antao VP, Kolberg JA. 2001. Single-copy Gene Detection Using
1210 Branched DNA (bDNA) In Situ Hybridization. *J Histochem Cytochem* **49**:603–611.
1211 doi:10.1177/002215540104900507
1212 Rep M, Krantz M, Thevelein JM, Hohmann S. 2000. The Transcriptional Response of *Saccharomyces*
1213 *cerevisiae* to Osmotic Shock Hot1p and Msn2p/Msn4p are Required for the Induction of
1214 Subsets of High Osmolarity Glycerol Pathway-Dependent Genes. *J Biol Chem* **275**:8290–
1215 8300. doi:10.1074/jbc.275.12.8290
1216 Rockman MV. 2012. The QTN Program and the Alleles That Matter for Evolution: All That's Gold
1217 Does Not Glitter. *Evol Int J Org Evol* **66**:1–17. doi:10.1111/j.1558-5646.2011.01486.x
1218 Rouhanifard SH, Mellis IA, Dunagin M, Bayatpour S, Jiang CL, Dardani I, Symmons O, Emert B,
1219 Torre E, Cote A, Sullivan A, Stamatoyannopoulos JA, Raj A. 2018. ClampFISH detects
1220 individual nucleic-acid molecules using click chemistry based amplification. *Nat Biotechnol*.
1221 doi:10.1038/nbt.4286
1222 Smith EN, Kruglyak L. 2008. Gene–Environment Interaction in Yeast Gene Expression. *PLoS Biol* **6**.
1223 doi:10.1371/journal.pbio.0060083
1224 Snoek BL, Sterken MG, Bevers RPJ, Volkers RJM, van't Hof A, Brenchley R, Riksen JAG, Cossins A,
1225 Kammenga JE. 2017. Contribution of trans regulatory eQTL to cryptic genetic variation in C.
1226 elegans. *BMC Genomics* **18**:500. doi:10.1186/s12864-017-3899-8
1227 Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E,
1228 Surendran P, Oliver-Williams C, Kamat MA, Prins BP, Wilcox SK, Zimmerman ES, Chi A,
1229 Bansal N, Spain SL, Wood AM, Morrell NW, Bradley JR, Janjic N, Roberts DJ, Ouwehand
1230 WH, Todd JA, Soranzo N, Suhre K, Paul DS, Fox CS, Plenge RM, Danesh J, Runz H,
1231 Butterworth AS. 2018. Genomic atlas of the human plasma proteome. *Nature* **558**:73–79.
1232 doi:10.1038/s41586-018-0175-2
1233 Taggart JC, Zauber H, Selbach M, Li G-W, McShane E. 2020. Keeping the Proportions of Protein
1234 Complex Components in Check. *Cell Syst* **10**:125–132. doi:10.1016/j.cels.2020.01.004
1235 Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A,
1236 Lao K, Surani MA. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat*
1237 *Methods* **6**:377–382. doi:10.1038/nmeth.1315
1238 Thompson DA, Cubillos FA. 2017. Natural gene expression variation studies in yeast. *Yeast* **34**:3–17.
1239 doi:10.1002/yea.3210
1240 Tong AHY, Boone C. 2007. High-Throughput Strain Construction and Systematic Synthetic Lethal
1241 Screening in *Saccharomyces cerevisiae* In: Stansfield I, Stark MJ, editors. Methods in
1242 Microbiology, Yeast Gene Analysis. Academic Press. pp. 369–707. doi:10.1016/S0580-
1243 9517(06)36016-3
1244 Viñuela A, Snoek LB, Riksen JAG, Kammenga JE. 2010. Genome-wide gene expression regulation as
1245 a function of genotype and age in C. elegans. *Genome Res* **20**:929–937.
1246 doi:10.1101/gr.102160.109
1247 Wadsworth GM, Parikh RY, Choy JS, Kim HD. 2017. mRNA detection in budding yeast with single
1248 fluorophores. *Nucleic Acids Res* **45**:e141. doi:10.1093/nar/gkx568

1249 Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt
1250 SW, Louis EJ, Liti G, Moses A, Blomberg A. 2011. Trait Variation in Yeast Is Defined by
1251 Population History. *PLOS Genet* **7**:e1002111. doi:10.1371/journal.pgen.1002111
1252 West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St. Clair DA.
1253 2007. Global eQTL Mapping Reveals the Complex Genetic Architecture of Transcript-Level
1254 Variation in Arabidopsis. *Genetics* **175**:1441–1450. doi:10.1534/genetics.106.064972
1255 Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, Madar V, Jansen R, Chung W, Zhou Y-H,
1256 Abdellaoui A, Batista S, Butler C, Chen G, Chen T-H, D'Ambrosio D, Gallins P, Ha MJ,
1257 Hottenga JJ, Huang S, Kattenberg M, Kochar J, Middeldorp CM, Qu A, Shabalin A, Tischfield
1258 J, Todd L, Tzeng J-Y, van Grootheest G, Vink JM, Wang Q, Wang Wei, Wang Weibo,
1259 Willemsen G, Smit JH, de Geus EJ, Yin Z, Penninx BWJH, Boomsma DI. 2014. Heritability
1260 and genomics of gene expression in peripheral blood. *Nat Genet* **46**:430–437.
1261 doi:10.1038/ng.2951
1262 Wurmthaler LA, Klauser B, Hartig JS. 2018. Highly motif- and organism-dependent effects of
1263 naturally occurring hammerhead ribozyme sequences on gene expression. *RNA Biol* **15**:231–
1264 241. doi:10.1080/15476286.2017.1397870
1265 Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, Sun BB, Laser A, Maranville JC, Wu H, Ho
1266 JE, Courchesne P, Lyass A, Larson MG, Gieger C, Graumann J, Johnson AD, Danesh J, Runz
1267 H, Hwang S-J, Liu C, Butterworth AS, Suhre K, Levy D. 2018. Genome-wide mapping of
1268 plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular
1269 disease. *Nat Commun* **9**. doi:10.1038/s41467-018-05512-x
1270 Zhang X, Cal AJ, Borevitz JO. 2011. Genetic architecture of regulatory variation in Arabidopsis
1271 thaliana. *Genome Res* **21**:725–733. doi:10.1101/gr.115337.110