Genome-wide base editor screen identifies regulators of protein abundance in yeast

Olga T. Schubert[1,2,3,4,5,*], Joshua S. Bloom[1,2,3,4], Meru J. Sadhu[1,2,3,4,6], Leonid Kruglyak[1,2,3,4,*]

[1]Department of Human Genetics, University of California, Los Angeles, Los Angeles, United States;
[2]Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, United States;
[3]Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, United States;
[4]Institute for Quantitative and Computational Biology, University of California, Los Angeles, Los Angeles, United States;
[5]Present address: Department of Environmental Systems Science, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland;
[6]Present address: National Human Genome Research Institute, National Institutes of Health, Bethesda, United States;
*Correspondence: olga.schubert@usys.ethz.ch and lkruglyak@mednet.ucla.edu

ORCIDs:
Olga T. Schubert: https://orcid.org/0000-0002-2613-0714
Joshua S. Bloom: https://orcid.org/0000-0002-7241-1648
Meru J. Sadhu: https://orcid.org/0000-0002-8636-9102
Leonid Kruglyak: https://orcid.org/0000-0002-8065-3057

Abstract

Abundance of proteins is extensively regulated both transcriptionally and post-transcriptionally. To systematically characterize how regulation of protein abundance is encoded in the genome and identify protein regulators on a genome-wide scale, we developed a genetic screen that uses a CRISPR base editor. We examined the effects of 16,452 genetic perturbations on the abundance of eleven yeast proteins representing a variety of cellular functions. Thereby, we uncovered hundreds of regulatory relationships, including a novel link between the GAPDH isoenzymes Tdh1/2/3 and the Ras/PKA pathway. Many of the identified regulators are specific to one of the eleven proteins, but we also found genes that, upon perturbation, affected the abundance of most of the tested proteins. While the more specific regulators usually act transcriptionally, broad regulators often have roles in protein translation. Our results provide unprecedented insights into the components, scale and connectedness of the protein regulatory network.

Introduction

Proteins are gene products that play many key roles in cells, serving as essential structural components, enzymes, and constituents of signaling networks, from receptors to transcription factors. Their abundance is extensively regulated transcriptionally and post-transcriptionally. For practical reasons, mRNA abundance is often studied as a proxy for protein abundance, but the correspondence between the two can be less strong than is usually implicitly assumed[1,2]. Therefore, it is important to study regulation of protein abundance directly.

CRISPR-based genetic screens using Cas9 or modified versions thereof, such as the CRISPR base editor or CRISPRi, provide an unprecedented opportunity to study the effects of thousands of genetic perturbations on gene expression in a pooled format[3–10]. The identity of genetic perturbations with phenotypic effects can be directly determined by sequencing guide RNAs, thereby eliminating the need for genome sequencing to identify causal mutations. To date, such studies have focused on the effects of genetic perturbations on mRNA expression, while the effects on protein expression remain unexplored. Here, we developed a CRISPR base editor screen to study how tens of thousands of individual mutations affect the abundance of selected proteins in yeast. Using this approach, we tested the effects of 16,452 coding mutations on the abundance of eleven yeast proteins selected to represent a variety of molecular and cellular functions. We identified hundreds of novel regulatory relationships, including both specific and broad regulators of protein abundance. The results provide general insights into how protein abundance is encoded in the genome and altered by changes in DNA sequence.

Results

A base editor screen in yeast to identify genetic effects on protein abundances

To introduce genetic perturbations across the yeast genome in a targeted fashion, we used the CRISPR base editor version 3 (BE3)[11]. This system consists of an engineered fusion of Cas9 and a cytidine deaminase that does not introduce DNA double-strand breaks but instead converts cytosine to uracil, thereby effecting a C-to-T substitution at a target site defined by the guide RNA (gRNA)[11]. We first adapted the base editor system for use in yeast (Figure S1). We then characterized base editing efficiency and editing profile of eight individual gRNAs, finding similar editing in yeast to that reported in human cells: 95% of edits were C-to-T transitions, and 89% of these occurred in a five-nucleotide region 13 to 17 base pairs upstream of the PAM sequence (Note S1, Figure S2). We next assessed base editor performance in a pooled screen format, first with a library of all 90 gRNAs targeting the CAN1 gene and then with a library of all 5430 gRNAs predicted to introduce stop codons into essential genes in yeast, along with 1500 control gRNAs (Note S1, Table S1). We found that base editor screens using both positive and negative selection can identify functionally important mutations in yeast (Figure S3, Table S2 and S3).

To identify regulators of protein abundance on a genome-wide scale using the base editor, we selected a library of 16,452 gRNAs predicted to introduce three classes of mutations: stop codons in nonessential genes, as well as missense mutations at highly conserved positions in essential genes and in nonessential genes (Methods and Table S1). We expect most genetic perturbations introduced by these gRNAs to have deleterious effects on the target genes. We introduced this gRNA library and the galactose-inducible base editor into yeast strains in which one protein of interest is endogenously tagged with green fluorescent protein (GFP)[12,13]. The GFP fluorescence level in these strains reports the abundance of the tagged protein[14]. We selected eleven strains with GFP-tagged proteins that cover a range of cellular functions. The target proteins include the Hsp70 chaperone Ssa1, which is involved in protein folding, nuclear transport and ubiquitin-dependent degradation of short-lived proteins and is a marker for heat shock; as well as the flavohemoglobin Yhb1, which is involved in oxidative and nitrosative stress responses and is a marker for oxidative stress[15]. Furthermore, we included the ribonucleotide reductase subunit Rnr2, which is involved in dNTP synthesis, the ribosomal protein Rpl9A, the histone Htb2, and several proteins involved in metabolism, including the fatty acid synthases Fas1 and Fas2, the enolase Eno2, which acts in glycolysis and gluconeogenesis, and the three glyceraldehyde-3-phosphate dehydrogenase (GAPDH) isoenzymes Tdh1, Tdh2 and Tdh3. Tdh3 is among the most highly expressed proteins in yeast and the main GAPDH enzyme, whereas the exact functional roles and regulation of Tdh1 and Tdh2 are less well understood.

After performing base editing in each of the eleven GFP strains, we isolated cells with very high and very low GFP intensity by fluorescence-activated cell sorting (Figure 1A). We next sequenced gRNA plasmids extracted from the sorted cells; the gRNAs enriched or depleted in these populations correspond to genetic perturbations that increase or decrease abundance of the protein of interest (Figure 1A). We identified 1020 gRNAs with a significant effect on the abundance of at least one of the eleven proteins ($FDR < 0.05$) (Table S4). To assess the reproducibility of the screen, we included gRNA pairs that were predicted to introduce the same mutation into the genome, and observed highly correlated effects on protein abundance, especially when both gRNAs had a significant effect (Pearson's $R = 0.9$, $p = 8.8e-13$) (Figure S4). When multiple gRNAs that introduce different mutations into the same gene had a significant effect on a protein's expression, their direction of effect was concordant 98% of the time. We therefore combined gRNA effects into a gene-level effect for all subsequent analyses, unless noted otherwise. Genetic perturbation of 710 genes had a significant effect on the abundance of at least one of the eleven proteins (FDR < 0.05) (Table S4).

Extensive genetic network underlies protein abundance regulation

The effects of genetic perturbations in 710 genes on the abundance of one or more of the 11 target proteins reveal an extensive network of regulatory relationships (Figure 1B). On average, the abundance of a protein was significantly affected by 156 gene perturbations (Figure 1C). Of these gene-protein relationships, an average of 53 (34%) are specific (defined here as gene perturbations that affect only one or two of the eleven target proteins) whereas 103 (66%) are nonspecific (the gene perturbation affects three or more of the eleven proteins). Furthermore, we found that, on average, protein abundance decreases more frequently than it increases in response to the gene perturbations tested (106 vs 50, binomial test, $p = 8.6e-6$) (Figure 1C). However, this trend is strongly dependent on the functional role of the protein. Abundance of proteins involved in core cellular functions mostly decreases, whereas abundance of stress-related proteins is just as likely to increase as it is to decrease (binomial test, $p = 2.9e-12$ and $p = 0.8$, respectively) (Figure 1C); this trend is independent of the absolute abundance of a protein (Figure S5).

The predominance of decreases in protein abundance in response to gene perturbations is driven by perturbations in essential genes, which on average decrease protein abundance in 80% of cases, whereas perturbations in nonessential genes do so in only 49% of cases (Figure 1D). Furthermore, perturbations in essential genes are generally more likely to significantly alter protein abundance than are perturbations in nonessential genes (chi-squared test, $\chi^2 = 384$, df = 1, $p < 2.2e-16$) (Figure 1D and E). For nonessential genes, premature stop codons are more likely to alter protein abundance than are missense mutations (chi-squared test, $\chi^2 = 37$, df = 1, $p = 1.2e-9$) (Figure 1E), and the average effect size of changes caused by stop codons is larger (two-sided t-test, $t = -3.0$, df = 496, $p = 0.0028$) (Figure 1F). We also find that mutations towards the end of a gene tend to have fewer and smaller effects, in line with previous observations (Figure S6)[16].

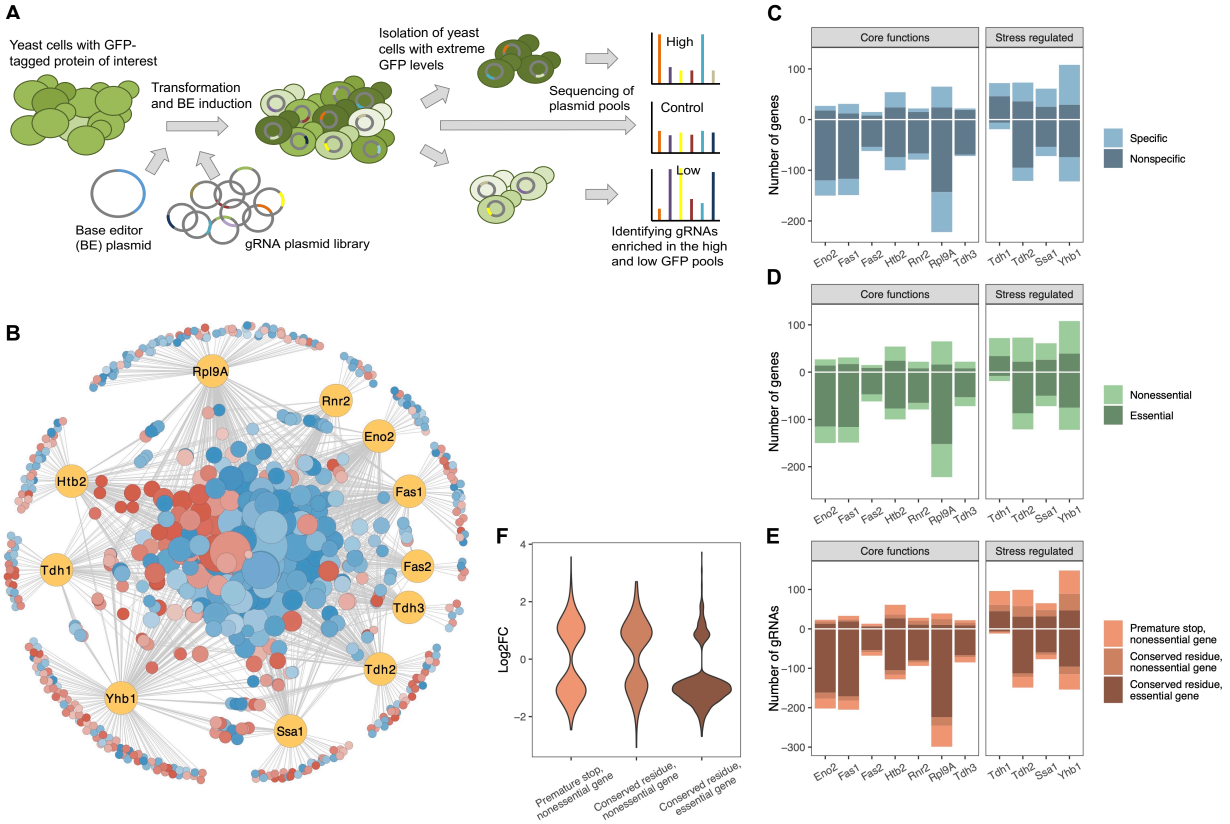

Figure 1 | A CRISPR base editor screen for protein abundance. (**A**) Schematic overview of the screen. (**B**) Network representing all identified gene-protein relationships. In yellow are the eleven proteins and in blue/red are all the gene perturbations that affect at least one of the proteins significantly (FDR < 0.05). On the outer circle are gene perturbations that affect only a single protein, whereas on the inside are those that affect two or more proteins. Node sizes correlate with the number of proteins affected. Colors indicate whether a perturbation (predominantly) increases (red) or decreases (blue) the protein(s). Figure created in Cytoscape[17]. (**C**) For each protein, the number of gene perturbations that cause a significant increase (positive vertical axis) or decrease (negative vertical axis) (FDR < 0.05). The darker shade indicates gene perturbations that affect only one or two of the eleven proteins ("specific"), whereas the lighter shade indicates gene perturbations that affect three or more of the eleven proteins ("nonspecific"). (**D**) For each protein, the number of gene perturbations that cause a significant increase or decrease (FDR < 0.05). The darker shade indicates perturbations of essential genes, whereas the lighter shade indicates perturbations of nonessential genes. (**E**) For each protein, the number of gRNAs that cause a significant increase or decrease (FDR < 0.05). The different shades reflect different types of expected mutations introduced by the particular gRNA. (**F**) Effect sizes ($\log_2$ fold-changes) of gRNAs that cause a significant change in protein abundance grouped by the type of expected mutation introduced by each gRNA (FDR < 0.05).

Base editor screen identifies known and novel regulatory relationships

The hundreds of regulatory relationships revealed by the base editor screen include many previously known regulators of the proteins under study (Figure 2A). For example, proteins involved in glycolysis (Eno2 and Tdh1/2/3) are strongly induced upon perturbation of other genes involved in glycolysis. Perturbation of Gcr1, a key transcription factor for glycolytic genes, leads to reduced Eno2 abundance, consistent with its role as a transcriptional activator[18,19]. Similarly, histone Htb2 abundance is strongly reduced upon perturbation of known regulators of histone gene transcription (Hir3, Spt10, Spt21)[20]. Ribosomal protein Rpl9A abundance is also strongly reduced upon perturbation of other ribosomal genes, as well as of genes involved in TOR signaling (Tor2, Asa1, Kog1, Rtc1, Sch9), including the TOR-regulated activators of ribosomal gene transcription, Fhl1 and Ifh1, which are known to act together with the broad transcription factor Rap1 to induce expression of ribosomal genes[21].

We also find numerous novel regulatory relationships. For example, the abundance of the stress-responsive flavohemoglobin/oxidoreductase Yhb1 is reduced by perturbation of one of its prime transcriptional activators, heme activator protein Hap1, as well as perturbations of a heme transporter and six genes involved in heme biosynthesis[22]. This is in accordance with Yhb1 being a heme protein itself[23]. A novel regulator of Yhb1 identified in our screen is the three-subunit Ssy1-Ptr3-Ssy5 (SPS) amino acid sensor. We find that perturbation of each of the three subunits, as well as of one of its two downstream effectors, the transcription factor Stp1, results in reduced Yhb1 abundance (Figure 2B). We independently confirmed the new regulatory relationship between Yhb1 and SPS by liquid-chromatography-coupled tandem mass spectrometry (LC-MS) proteomics analysis of one of the SPS mutants identified in our screen, SSY5 G638Q (Figure 2C and Table S5). Furthermore, gene ontology (GO) enrichment analysis of differentially expressed proteins in this mutant revealed that SPS perturbation leads not only to reduced Yhb1 abundance but also to reduced abundance of other mitochondrial proteins, including components of the mitochondrial ribosome, oxidative phosphorylation, and ATP metabolism (FDR = 4.5e-4, Figure 2C and Table S5). In contrast, amino acid metabolic processes are broadly induced in the mutant (FDR = 6.6e-9, Figure 2C), suggesting that SPS signaling not just activates expression of amino acid transporters but also represses amino acid biosynthesis[24]. Recently, the presence of extracellular amino acids has been found to induce Yhb1 expression in the distantly related yeast *Candida albicans*[25,26]. Our observations extend this link between amino acids and increased nitrosative and oxidative stress resistance to *S. cerevisiae* and suggest that it is mediated by the SPS amino acid sensor. We propose that SPS senses the presence of external amino acids and activates the transcription factor Stp1, which in turn induces Yhb1 expression, resulting in increased nitrosative and oxidative stress resistance (Figure 2B).

In addition to the relatively specific regulators described above, we identified two key complexes that modulate the transcriptional machinery, mediator and SAGA, as regulators of protein abundance. We find particularly large effects upon perturbation of two subunits of the Cdk8 kinase module of mediator, Ssn3 and Srb8 (Figure 2D). Moreover, we see large effects for perturbations of the two subunits forming the hook structure on mediator, Nut2 and Rox3, which supports recent reports that this domain is critical for the interaction between the core mediator complex and the Cdk8 kinase module (Figure 2D)[27,28]. The observation that perturbations of the Cdk8 module increase protein abundance is in accordance with the suggested repressor role of this transiently associating regulatory module of mediator[29]. We also find that classical post-transcriptional regulators contribute strongly to protein abundance. For example, Ssa1 and Yhb1 are significantly induced upon perturbation of a number of genes involved in proteasomal degradation (Figure 2A).

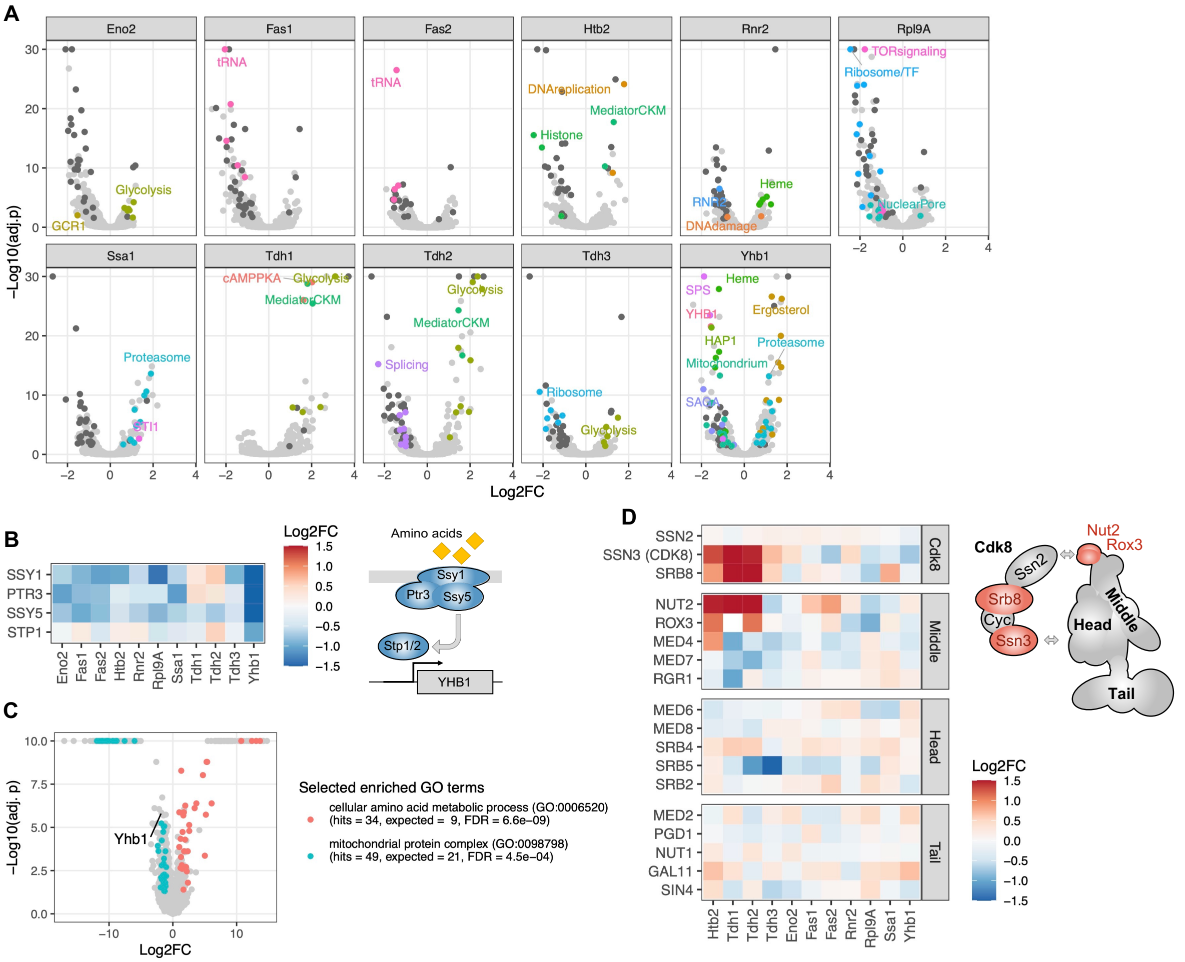

Figure 2 | Regulators of protein abundance. (**A**) Each volcano plot contains the results for a single protein, with each dot representing the effect of a gene's perturbation on that protein. The vertical axis represents the negative logarithm of the FDR-corrected p-value. Selected dots are colored to highlight functional classes. Dark gray dots indicate genes that significantly affect eight or more proteins (FDR < 0.05). (**B**) Heatmap showing effects of gene perturbations in the SPS amino acid sensor on the eleven proteins. In the scheme, all subunits with significant effect on Yhb1 are colored (FDR < 0.05). (**C**) Volcano plot showing abundance changes of proteins in the SSY5 G638Q mutant compared to wild type as measured by LC-MS proteomics. The vertical axis represents the negative logarithm of the FDR-corrected p-value. Selected enriched functional categories (gene ontology (GO) terms) are colored. (**D**) Heatmap showing effects of gene perturbations in Mediator on the eleven proteins. In the scheme, all subunits with significant effects on Htb2 are colored (FDR < 0.05); arrows indicate suggested interaction points between the Cdk8 module and mediator[27,28]. The color scale for the heatmaps is capped at -1.5 and 1.5.

Gene perturbations with broad effects on protein abundance often affect protein biosynthetic processes

About half of the 710 gene perturbations with significant effects on protein abundance affect at least two proteins, and some of these alter the abundance of a large fraction of the eleven proteins (Figure 1B and 3A). For instance, there are 29 gene perturbations that significantly alter the abundance of eight or more proteins (Figure 3A and B, upper panel). Remarkably, 21 (72%) of these 29 genes have roles in protein translation—more specifically, in ribosome biogenesis and tRNA metabolism (FDR < 8.0e-4, Figure 3C). In contrast, perturbations that affect the abundance of only one or two of the eleven proteins mostly occur in genes with roles in transcription (e.g., GO:0006351, 98 genes, 47 expected, FDR < 1.3e-5).

One of the broadest regulators is POP1, an essential gene involved in ribosomal RNA (rRNA) and transfer RNA (tRNA) maturation (Figure 3C). We used LC-MS proteomics to independently confirm that missense mutation H642Y in POP1 leads to reduced abundance of most of the eleven proteins under study (Figure 3D). Furthermore, the proteomics data revealed that perturbation of POP1 reduces the abundance of many proteins in the nucleolus (FDR = 0.013, Figure S7A), consistent with the role of POP1 in rRNA and tRNA processing. We also see the induction of a large number of proteins involved in amino acid biosynthesis (FDR = 1.4e-13, Figure S7A). This observation could represent an indirect response to perturbed ribosome metabolism or a direct role of POP1 in suppressing amino acid metabolism.

All but three of the 29 gene perturbations with broad effects on protein abundance generally lowered protein abundance (Figure 3B and C). One of the three genes whose perturbation increased protein abundance is the PP2A-like phosphatase SIT4. Perturbation of SAP155, a gene strictly required for the function of SIT4, has very similar effects on protein abundance (Figure 3B, lower panel)[30]. LC-MS proteomics of a SIT4 loss-of-function mutant (Q184*) revealed the induction of a large number of proteins involved in cellular respiration, carbohydrate metabolism, and oxidative stress (FDR < 0.01, Figure S7B and Table S5).

To better understand how the genetic effects on protein abundance relate to cellular phenotypes, we performed a fitness screen with the same gRNA library used for the protein screen (Table S6). We observed that gene perturbations with broader effects on protein abundance were more likely to impair cellular fitness (Pearson's R = -0.43, p < 2.2e-16) (Figure 3E, top panel). Consistent with this finding, these perturbations were also more likely to be in essential genes (Figure 3E, bottom panel). Because essential genes are often involved in maintaining general cellular physiology, their perturbation is expected to have more widespread effects on protein abundance than perturbation of nonessential genes that may function in niche processes. This is also in line with earlier observations that essential genes are more interconnected in protein-protein interaction networks (network hubs), and that there is a negative correlation between the connectivity of a gene product and the change in cellular growth rate upon gene deletion[31,32].

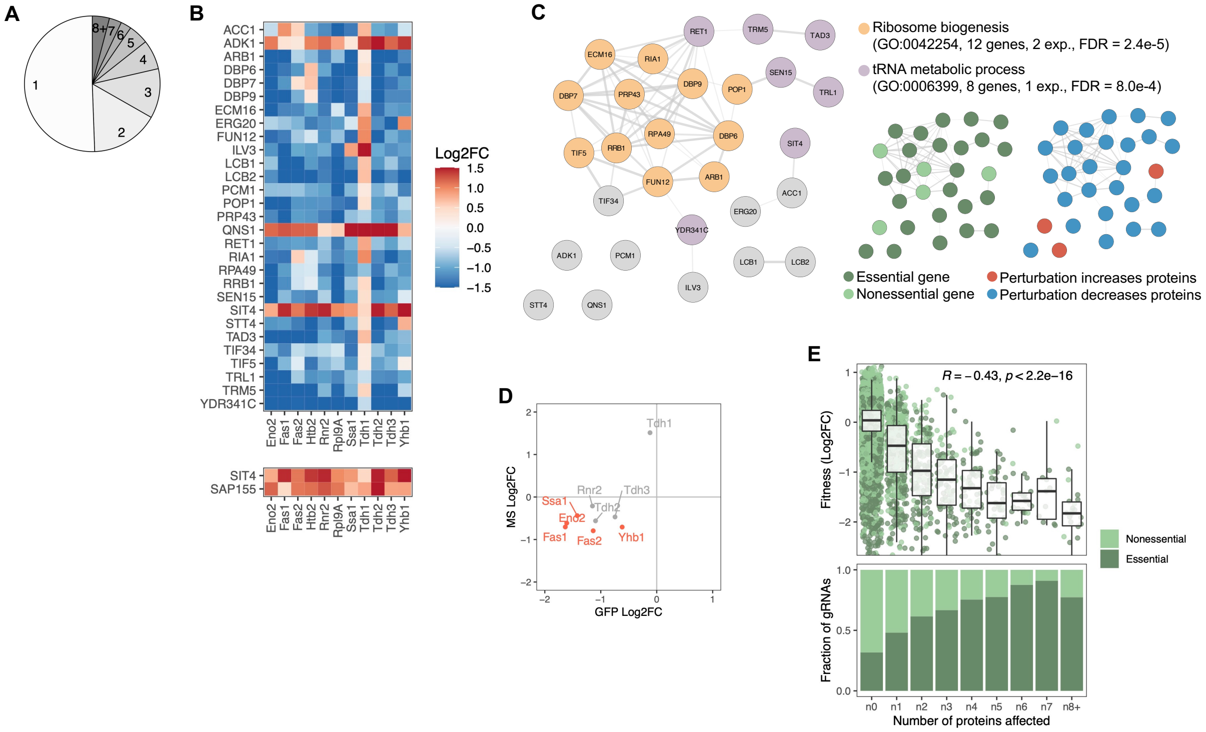

Figure 3 | Gene perturbations with broad effects on protein abundance. (**A**) Number of proteins significantly affected by each gene perturbation (FDR < 0.05). (**B**) Heatmap of protein-level effects of the 29 gene perturbations with broad effects (upper panel) and of SIT4 and SAP155 gene perturbations for direct comparison (lower panel). The color scale is capped at -1.5 and 1.5. (**C**) Network representation of the 29 genes with broad effects. Edges of the network reflect the overall confidence score for protein-protein interactions reported in the STRING database and colors represent enriched functional categories (orange: ribosome biogenesis, purple: tRNA metabolism)[35,36]. Miniature networks are colored by gene essentiality (light green: nonessential, dark green: essential) or prevalent direction of effect on protein abundance (blue: decreased abundance, red: increased abundance). (**D**) Validation of observations from the base editor screen by LC-MS proteomics for a hypomorphic mutation in POP1 (H642Y). Red data points reflect proteins that change significantly in both datasets upon POP1 perturbation (FDR < 0.1). (**E**) Upper panel: Dropout of gRNAs targeting essential and nonessential genes in a yeast culture after 24 hours in galactose media (for base editor induction) followed by 24 hours in glucose media. The horizontal axis reflects the number of proteins significantly affected (FDR < 0.05). Box plot overlays summarize the underlying data points (center line: median; box limits: first and third quartiles; whiskers: 1.5x interquartile range). Lower panel: The relative fraction of gRNAs targeting essential and nonessential genes as a function of how many proteins significantly affected (FDR < 0.05).

Several of the gene perturbations that affect multiple proteins overlap previously identified hotspots of gene expression and protein quantitative trait loci (eQTLs and pQTLs)[33,34]. Thus, the base editor screen identified candidate causal genes for eQTL and pQTL hotspots, which may provide functional insights into how natural genetic variation affects gene and protein expression (Note S2, Figure S8, Table S7).

Distinct responses of GAPDH isoenzymes to gene perturbations suggest functional diversification

We next compared the effects of gene perturbations on the three GAPDH isoenzymes, Tdh1, Tdh2 and Tdh3. GAPDH is a highly conserved enzyme best known for its central role in glycolysis and gluconeogenesis, but it has also been implicated in many non-metabolic processes[37–39]. In *S. cerevisiae*, Tdh3 is thought to be the main GAPDH enzyme in glycolysis and is one of the most highly expressed proteins[14]. None of the three GAPDH genes are essential for cell viability, but a functional copy of either TDH2 or TDH3 is required; in contrast, cells with a TDH1 deletion, either alone or in combination with a TDH2 or a TDH3 deletion, are viable[40–42]. The three isoenzymes were shown to selectively change abundance in response to available carbon sources and other environmental perturbations[43–46]. However, despite the key role of GAPDH in metabolism and other cellular processes, the specific functional roles and regulatory mechanisms of the three isoenzymes are still largely unknown.

Our base editor screen revealed that abundance of Tdh1, Tdh2 and Tdh3 responds similarly to the perturbation of many genes, including those involved in central carbon metabolism and biosynthesis of aromatic amino acids (Aro1/2) (Figure 4A). However, some gene perturbations have distinct effects on the three isoenzymes. For example, perturbation of genes encoding two members of the Mediator Cdk8 module, Ssn3 and Srb8, affect abundance of Tdh1 and Tdh2 but not Tdh3 (Figure 4A and 2D). Ssn3 and Srb8 have been implicated in carbon catabolite repression, a mechanism to repress genes involved in the use of alternate carbon sources when a preferred carbon source such as glucose is available[47]. Because Tdh3 is thought to be the main enzyme responsible for glycolytic flux under high-glucose conditions, it is plausible that the alternative enzymes Tdh1 and Tdh2 are under carbon catabolite repression and play a role when glucose is unavailable. This hypothesis is consistent with strong induction of Tdh1 in the absence of glucose[43,44].

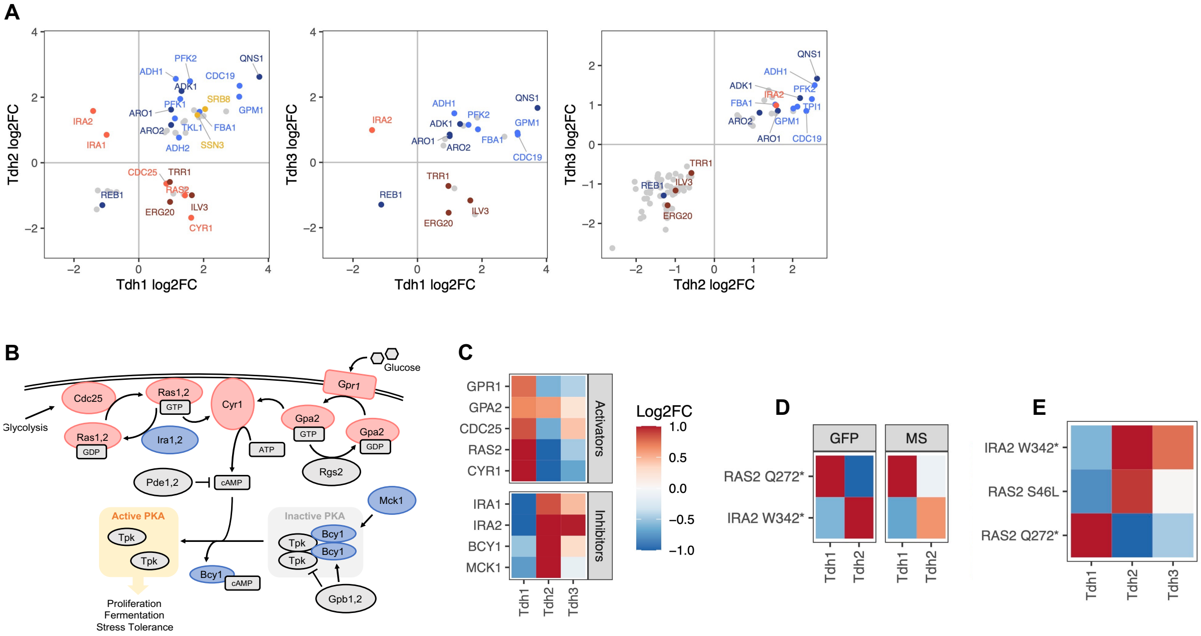

Figure 4 | Regulators of the three GAPDH isoenzymes Tdh1, Thd2 and Tdh3. (**A**) Pairwise comparison of effects of gene perturbations on Tdh1, Tdh2, and Tdh3 protein abundances. Only genes that upon perturbation affect both Tdh proteins significantly are shown (FDR < 0.05). (**B**) Ras/cAMP/PKA pathway representation with components colored if they showed a significant signal for one or more of the GAPDH isoenzymes (red: activators, blue: inhibitors). Scheme adapted from Peeters and colleagues with added Mck1[56,57]. (**C**) Heatmap representing all genes involved in the PKA pathway that upon perturbation affect at least one of the three GAPDH isoenzymes significantly (FDR < 0.1). (**D**) Effects of IRA2 W342* and RAS2 Q272* mutations on Tdh1 and Tdh2 protein abundance observed in the base editor screen (“GFP”) and by LC-MS-based proteomics on the individual mutants (“MS”). (**E**) Effects of IRA2 W342*, RAS2 Q272* and RAS2 S46L mutations on Tdh1, Tdh2 and Tdh3 protein abundance observed in the base editor screen. The heatmap color scale is identical for C, D, and E, and is capped at -1 and 1.

Tdh1 protein abundance shows the most distinct responses to gene perturbations (Figure 4A). The most striking differences occur in response to perturbation of the Ras/cAMP/PKA pathway, a key driver of cell growth and proliferation in response to glucose availability and an integrator of stress responses[48–50] (Figure 4A and B). Perturbations of activating components of this pathway (Gpr1, Ras2, Cyr1, Cdc25) increase abundance of Tdh1 but decrease abundance of Tdh2 (and, to a lesser extent, Tdh3). Perturbation of the farnesyl pyrophosphate synthetase Erg20 had the same effects. Erg20 is an essential enzyme involved in protein farnesylation. It has been shown that Ras2 farnesylation is critical for PKA activation, suggesting that the signal we see for Erg20 is also the result of a perturbed Ras/cAMP/PKA pathway[51]. Conversely, perturbations of inhibiting components of this pathway (Ira1/2, Bcy1, Mck1) decrease abundance of Tdh1 while increasing abundance of Tdh2 and Tdh3 (Figure 4C). We used LC-MS proteomics to confirm these findings in strains carrying engineered loss-of-function mutations in an inhibitor (IRA2) and an activator (RAS2) of the pathway (Figure 4D and Table S5). A functional connection between the Ras/cAMP/PKA pathway and

Tdh1/Tdh2 is further supported by the negative genetic interactions between TDH1/TDH2 and RAS2 reported in large-scale studies[52–55]. The distinct regulatory mechanisms of the three GAPDH isoenzymes uncovered by the base editor screen suggest that they play separate roles in different metabolic states of the cell.

Discordant gRNA effects point to a gain-of-function mutation in RAS2

Among the gRNAs targeting RAS2, one is predicted to introduce the missense mutation S46L and another to introduce the loss-of-function mutation Q272*. We observed that these gRNAs had discordant effects on protein abundance. Ras2 Q272* increased abundance of Tdh1 and decreased abundance of Tdh2, while Ras2 S46L had the opposite effects (Figure 4E). The effect of the Ras2 S46L mutation on Tdh1 and Tdh2 was similar to that of a loss-of-function mutation of Ira2, a direct antagonist of Ras2 (Figure 4E). Our observation that the S46L mutation activates Ras2 is consistent with the results of saturation mutagenesis of the highly conserved human K-Ras gene, which showed increased activity upon mutation of the homologous serine residue (S39 in human K-Ras) to leucine or valine[58]. Human Ras genes are important oncogenes, with activating point mutations found in as many as 20% of all human tumors[59]. Mutations of this specific serine residue have been found in cervical cancers (COSMIC database)[60]. Thus, base editor screens in yeast can recover both loss-of-function and gain-of-function mutations in genes that are relevant to human diseases.

Discussion

Systems genetics is concerned with how genetic information is integrated, coordinated, and ultimately transmitted through molecular, cellular, and physiological networks to enable the higher-order functions of biological systems[61]. Proteins are the key molecular players functionally linking genotype and phenotype of a cell. Therefore, systematically studying regulators of protein abundance is crucial to reaching a more complete, system-level understanding of molecular cell physiology. We developed a high-resolution CRISPR base editor screen and used it to investigate how the abundance of a variety of selected yeast proteins is affected by each of tens of thousands of genetic perturbations. Thereby, we gained insights into the extensive regulatory network underlying a cell's proteome, in which the abundance of any given protein is affected by a large number of genetic loci. We found that proteins involved in core cellular functions respond differently to genetic perturbations than proteins involved in stress responses. Furthermore, proteins were generally more likely to be affected by perturbations of essential genes than of nonessential genes. Genes affecting the abundance of many proteins simultaneously (i.e., hubs of the protein regulatory network) were often involved in post-transcriptional protein biosynthetic processes. In contrast, perturbations that affected only specific proteins were enriched in transcriptional regulation. In addition, we discovered numerous new regulatory relationships, including a new link between the Ras/PKA pathway and the GAPDH isoenzymes that suggests functional diversification of these homologs which enables cells to adapt to different metabolic states.

Future studies will be able to build on this experimental framework to further refine and complete our understanding of the genetic factors influencing protein abundance. Potential extensions include screening for abundance of additional proteins, examining the effects of genetic perturbations on other molecular and physiological phenotypes, increasing the number of gRNAs and using alternative base editors to achieve denser sampling of mutations across coding and non-coding regions, introducing multiple gRNAs per cell to investigate interactions between genetic perturbations, and performing screens in different genetic backgrounds or under different growth or stress conditions. A dense genome-wide sampling of mutations, together with quantitative molecular and cellular phenotypic readouts across different environmental conditions and genetic backgrounds, will provide the data needed to build holistic models that bring us closer to predicting the functional consequences of altering any base pair in the genome through natural sequence variation or engineered perturbation[62].

Acknowledgements

We would like to thank Dr. James Boocock and Dr. Frank Albert for helpful scientific discussions and Dr. Oliver F. Brandenberg for feedback on the manuscript. We would also like to thank Dr. Yasaman Jami-Alahmadi and Dr. James A. Wohlschlegel from the UCLA Proteomics Facility for proteomics measurements and Dr. Xinmin Li and his team at the UCLA Technology Center for Genomics & Bioinformatics for sequencing services. This work was supported by fellowships and grants from the Human Frontier Science Program (to OTS, LT000737/2016-L), the Swiss National Science Foundation (to OTS, P2EZP3_165280), the Howard Hughes Medical Institute (to LK) and the National Institutes of Health (to LK, R01GM102308). The authors declare no competing financial interests.

Author contributions

OTS designed the study, performed experiments and analyses, and wrote the manuscript; JSB performed analyses; JSB, MJS, and LK provided expertise and feedback; LK supervised the study. All authors reviewed the manuscript.

Online Methods

Data availability

The DNA sequencing read data is deposited in the NCBI Sequence Read Archive under BioProject ID PRJNA732764 and the mass spectrometry proteomics data is deposited in the ProteomeXchange Consortium partner repository PRIDE with the dataset identifier PXD029363 and 10.6019/PXD029363[63]. (Reviewer access: reviewer_pxd029363@ebi.ac.uk, crTJNJ9X)

Code availability

All custom code is available on GitHub without restrictions:
https://github.com/OlgaTSchubert/ProteinGenetics

Yeast strains, maintenance, and cultures

Saccharomyces cerevisiae S288C strain BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was used for all experiments except the protein abundance screen[64]. For that, we used 11 strains from the Yeast GFP Library, which is also based on BY4147. In each of these strains, a GFP cassette (GFP(S65T)-HIS3MX6, where HIS3 is from *Lachancea kluyveri*) is fused in-frame to the C-terminus of a gene of interest[12]. As a reference genome, we used the *S. cerevisiae* S288C genome sequence version R64-2-1 from the Saccharomyces Genome Database (SGD).

All yeast strains generated in this study are listed in Table S8.

Yeast cultures were grown at 30ºC with 250 rpm orbital shaking for liquid cultures. Either of the following media was used:

- Rich media (**YPD**):
    - 20 g/L Bacto Peptone (BD Biosciences #211820)
    - 10 g/L Bacto Yeast Extract (BD Biosciences #212720)
    - 2 g/L dextrose/glucose (MP Biomedicals #901521)

- Synthetic (CSM) dropout media lacking leucine and uracil (**CSM -Leu -Ura**):
    - 6.7 g/L Difco's YNB without Amino Acids (BD Biosciences #291940)
    - 2 g/L dextrose/glucose (MP Biomedicals #901521)
    - 0.67 g/L CSM powder lacking leucine and uracil (Sunrise Science Products #1038-100)
    - For non-selective version of CSM, 100 mg/L L-leucine (Sigma #L8912) and 20 mg/L uracil (Sigma #U0750) were added.

- Synthetic (SC) dropout media lacking leucine and uracil (**SC -Leu -Ura**):
    - 6.7 g/L Difco's YNB without Amino Acids (BD Biosciences #291940)
    - 2 g/L dextrose/glucose (MP Biomedicals #901521)
    - 1.74 g/L SC dropout powder lacking leucine and uracil (Sunrise Science Products #1318-030)
    - For non-selective version of SC, 100 mg/L L-leucine (Sigma #L8912) and 20 mg/L uracil (Sigma #U0750) were added.

- Canavanine media:
    - 6.7 g/L Difco's YNB without Amino Acids (BD Biosciences #291940)
    - 2 g/L dextrose/glucose (MP Biomedicals #901521)
    - 20 mg/L L-histidine hydrochloride monohydrate (Fisher Scientific #BP383100)
    - 20 mg/L L-methionine (Sigma #M9625)
    - 20 mg/L adenine sulfate salt (Sigma #A2545)
    - 60 mg/L canavanine sulfate salt (Sigma #C9758)
    - Note that canavanine media must not contain arginine for the canavanine resistance selection to work.

Agar plates were prepared with the same media formulation but with agar added at 20 g/L (BD Biosciences #214030). For induction of the base editor, which is under a galactose-inducible promoter, we used CSM/SC media with 2 g/L galactose (Sigma #G0625) instead of dextrose/glucose. Cryostocks were prepared by mixing yeast cultures or cell suspensions with sterile 40% glycerol-in-water solution to a final concentration of 13.3% glycerol (v/v) and were stored at -80ºC.

For the construction of individual plasmids and small libraries, we used a DH5α-derived *Escherichia coli* strain (NEB #C2987). For large library transformations, we used E. cloni 10G Supreme Electrocompetent Cells (Lucigen #60081). Liquid cultures were grown at 37ºC with 250 rpm orbital shaking. If not mentioned otherwise, both solid and liquid cultures were grown in LB media (20 g/L, Fisher Scientific #BP1427), which for selections was supplemented with 100 µg/ml ampicillin (Sigma-Aldrich #A0166) and/or 50 µg/ml kanamycin (Fisher Scientific #BP906). Cryostocks were prepared by mixing bacterial cultures or cell suspensions with sterile 40% glycerol-in-water solution to a final concentration of 20% glycerol (v/v) and were stored at -80ºC.

Plasmids

All plasmids generated in this study are listed in Table S9 and all primers used are listed in Table S10.

Base editor plasmid (pGal-BE3) (Figure S1A). Our yeast base editor plasmid was constructed based on the yeast Cas9 plasmid described by DiCarlo and colleagues (Addgene #43802)[65]. In addition to the Cas9 gene under a galactose-inducible GALL promoter[66], a CEN/ARS origin, and other elements, this plasmid encodes LEU2 which enables its selection in leucine-auxotroph yeast grown in media lacking leucine[67]. Using Gibson Assembly, we swapped the Cas9 gene encoded on the plasmid with the base editor v3 (BE3) described by Komor and colleagues (Addgene #73021)[11]. This plasmid is deposited in Addgene: #172409.

gRNA plasmid with 2-kb placeholder (pgRNA-backbone) (Figure S1B). Our yeast gRNA plasmid was constructed based on the yeast gRNA plasmid described by DiCarlo and colleagues (Addgene #43803)[65]. In addition to the gRNA coding sequence under an SNR52 promoter (for RNA Polymerase III transcription), a 2µ origin, and other elements, this plasmid encodes URA3 which enables its selection in uracil-auxotroph yeast grown in media lacking uracil[67]. Using Gibson Assembly, we swapped the gRNA targeting sequence encoded on the plasmid with the guide placeholder cassette from the lentiCRISPR v2 plasmid described by Sanjana and colleagues (Addgene #52961)[68]. The 2-kb guide placeholder present in this cassette and flanked by Esp3I/BsmBI restriction sites facilitates the construction of individual gRNA plasmids as well as gRNA plasmid libraries. All other Esp3I/BsmBI sites present in the plasmid backbone were removed by site directed mutagenesis (Agilent #210515). This plasmid is deposited in Addgene: #172517.

gRNA plasmids for characterization and validations. Individual gRNA plasmids were generated based on the pgRNA-backbone plasmid described above according to the protocol by Joung and colleagues[69]. Briefly, two partially complementary 25-nucleotide oligonucleotides were annealed, forming the gRNA targeting sequence. The resulting double-stranded DNA fragment with 4-nucleotide overhangs was then inserted into the pgRNA-backbone plasmid by Golden Gate Assembly[70] using the Esp3I/BsmBI restriction enzyme (NEB #E1602S). All gRNA targeting sequences for these individual base editing experiments are listed below.

gRNA plasmid libraries. The barcoded gRNA plasmid libraries were generated based on the pgRNA-backbone plasmid described above. Details are described in a separate section below. Briefly, a kanamycin resistance cassette was barcoded with 20 random nucleotides and inserted into the pgRNA-backbone plasmid, resulting in a barcoded plasmid library (Figure S1C). Subsequently, the oligonucleotide libraries with all gRNA targeting sequences were inserted into this barcoded plasmid library, resulting in barcoded gRNA plasmid libraries.

Base editor characterization with individual gRNAs

To characterize the base editor efficiency and editing window, we selected seven gRNAs predicted to introduce premature stop codons into CAN1, ADE1 and ADE2 using the Benchling (https://benchling.com). The gRNAs were cloned into the pgRNA-backbone plasmid as described above in the “Plasmids” section. In addition, we tested a gRNA-containing plasmid from DiCarlo and colleagues predicted to introduce a missense mutation into CAN1[65]. We first transformed our BY4741 yeast strain with the base editor plasmid pGal-BE3 using a high-efficiency lithium acetate-based protocol described by

Gietz and Schiestl with minor modifications[71]. The resulting strain was transformed with the eight gRNA plasmids individually (one gRNA plasmid per strain) according to the same protocol. Successful transformants growing on selective plates (CSM -Leu -Ura) were then transferred into selective galactose media to induce base editor expression and grown at 30ºC until they reached stationary phase (44 hours).

Approximately 10^8 cells per culture were harvested and genomic DNA was extracted using Qiagen's DNeasy Blood & Tissue Kit (Qiagen #69506) according to their supplementary protocol "Purification of total DNA from yeast" (DY13 Aug-06) but using 10 U/ml zymolase (amsbio #120491-1) instead of lyticase. The predicted genomic target region of each gRNA in the respective DNA sample was then amplified using PfuUltra II Fusion HS DNA Polymerase (Agilent #600670) with the following thermocycling protocol: 2 min at 95ºC, 22x (30 s at 95ºC, 30 s at 55ºC, 30 s at 72ºC), 3 min at 72ºC. Primers were designed to include overhangs that mimic the product of the Illumina Nextera transposase in order to facilitate further processing by Illumina sequencing reagents (xxx indicates the 20-nucleotide genomic binding regions of these primers):
Forward Nextera primers: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGxxx
Reverse Nextera primers: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGxxx
The resulting fragments were pooled, then purified with the MinElute PCR Purification Kit (Qiagen #28004). A second round of PCR to extend the Nextera adapters was performed using reagents from the Nextera DNA Library Prep Kit (NPM, PPC and indexes; Illumina #FC-121-1031) according to the manufacturer's instructions. This second PCR product was run on a 2% agarose gel and gel-extracted using a QIAquick Gel Extraction Kit (Qiagen #28704). The amplicon library was quantified by qPCR using the KAPA Library Quantification Kit (KAPA/Roche #KK4824). To increase diversity, 5% PhiX Sequencing Control v3 (Illumina #FC-110-3001) was added. The final library was brought to a concentration of 15 pM and sequenced on an Illumina MiSeq with paired-end 150-base pair reads.

The sequencing data was processed using CRISPResso2 (v2.0.45) [72] using the default settings, except for the following:

```
--min_average_read_quality 30
--quantification_window_size 20
--quantification_window_center -10
--base_editor_output
```

To determine the fraction of all possible base edits (Figure S2), the CRISPResso2 output file "Quantification_window_nucleotide_frequency_table.txt" was processed with a custom R script (R version 4.0.3, RStudio version 1.3).

gRNA library design

Using custom R scripts run in RStudio (R version 4.0.3, RStudio version 1.3), we first compiled a list of all possible gRNA targeting sequences (guides) suitable for base editing in yeast by searching for all possible guides in the yeast genome, defined as the 20 nucleotides upstream of an NGG PAM site. We also added 10,000 guides not targeting the yeast genome for use as controls. We predicted the expected base editing outcome assuming that all Cs will be converted to Ts within a window 13-17 base pairs upstream of the PAM site[11]. We then excluded (i) guides with a potential RNA Polymerase III terminator (TTTT), (ii) guides that target regions where two or more annotated genes overlap, and (iii) guides whose target region harbors homology to another genomic location and could therefore cause unexpected off-target effects (considering all perfect matches in the 12-nucleotide seed region directly upstream of the PAM)[65,73].

For all missense mutations expected to be introduced by the gRNAs, we determined the Provean score, which predicts the effect of amino acid substitutions based on conservation[74]. For gRNAs expected to introduce multiple missense mutations, we chose as representative value the maximal absolute Provean score. We then defined an absolute Provean score of >5 as highly conserved (18% of all gRNAs predicted to introduce missense mutations into the yeast genome and passing the above mentioned filters).

From the filtered gRNA database we extracted five different sets of gRNAs:

CA: All 170 gRNAs predicted to introduce missense or nonsense mutations in CAN1 and ADE2; only the 91 CAN1 targeting gRNAs were used for experiments reported here.

ES: All 5442 gRNAs predicted to introduce premature stop codons in essential genes

NS: 5500 gRNAs predicted to introduce premature stop codons in nonessential genes, randomly selected out of 19,772 gRNAs with that property

EP: 5500 gRNAs predicted to introduce missense mutations at highly conserved residues in essential genes, randomly selected out of 8,955 gRNAs with that property

NP: 5500 gRNAs predicted to introduce missense mutations at highly conserved residues in nonessential genes, randomly selected out of 33,357 gRNAs with that property

We also designed three sets of control gRNAs predicted to not introduce any (coding) mutations in the yeast genome: 500 gRNAs introducing synonymous mutations in yeast genes, 500 gRNAs without a C in the targeting region, and 500 gRNAs with random sequences not targeting the yeast genome.

For the synthesis of the gRNA libraries, each gRNA targeting sequence of 20 nucleotides was flanked by 30-nucleotide homology arms on either side for Gibson Assembly into the receiver plasmid (pgRNA-backbone) (Figure S1C). Additionally, at both ends of the oligonucleotides 15-nucleotide primer binding sites were added which differed for each of the five sets of gRNAs described above (control gRNAs were given the same primer binding sites as the ES gRNAs). These primer binding sites allow for specific amplification of individual sublibraries from the main library. The total length of the resulting oligonucleotides was 122 nucleotides. Note that we originally planned to remove the primer binding sites with an MluI restriction digest and therefore excluded from synthesis a small number of gRNAs that contained MluI sites in the targeting sequence. However, we ended up using an alternative strategy not dependent on MluI (see below).

All 23,541 gRNAs were synthesized as a single oligonucleotide library (Agilent #G4893A).

Construction of barcoded gRNA plasmid libraries

The barcoded gRNA plasmid library was cloned based on the pgRNA-backbone plasmid described above. First, a kanamycin resistance cassette (originally likely from the pCR-Blunt II-TOPO plasmid, Thermo Fisher Scientific #451245) was amplified using a primer pair of which the forward primer (OS123) contained an overhang with 20 random nucleotides. After removal of the plasmid template by DpnI digest (NEB #R0176S) and DNA purification (Zymo Research #D4013), this barcoded resistance cassette was introduced into the PCR-amplified, DpnI-treated, and purified pgRNA-backbone plasmid backbone by Gibson Assembly (NEB #E5510S)[75]. After DNA purification (Zymo Research #D4013), plasmid libraries were eluted in 10 µl ultra pure water and 2 µl was electroporated into five high-efficiency electrocompetent *Escherichia coli* cell aliquots each (Lucigen #60081) using a MicroPulser Electroporator (Bio-Rad Laboratories #1652100). After 60 minutes of growth in liquid recovery media (SOC outgrowth media, NEB #B9020S), cells were plated onto five 150-mm LB agar plates supplemented with ampicillin/carbenicillin (100 µg/ml) and kanamycin (50 µg/ml) and incubated overnight at 37ºC. Serial dilution platings and colony counts performed in parallel indicated a total of 150,000 independent transformants, each assumed to contain a plasmid with a different barcode. All colonies were then washed off the plates, pooled and subjected to plasmid extraction (Qiagen #12963).

The oligonucleotide library consisting of the five different sublibraries described above under “gRNA library design” arrived lyophilized and was first dissolved in Tris-EDTA buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8) to a concentration of 100 nM. An aliquot was then further diluted to 10 nM and the desired sublibraries were amplified by a 15-cycle PCR with sublibrary-specific primers (Q5 High-Fidelity 2X Master Mix by NEB #M0492S). After DNA purification (Zymo Research #D4013), a second 5-cycle PCR was performed with primers binding more proximal to the gRNA targeting sequences than the previous primers, thereby excluding the library-specific primer binding sites from the final oligonucleotide libraries. This second PCR was done with the same polymerase and again followed by DNA purification.

The amplified oligonucleotide sublibraries were then cloned into the barcoded pgRNA-backbone plasmid library described above according to the protocol by Joung and colleagues[69]. First, the barcoded plasmid library was digested with Esp3I for 60 minutes at 37ºC (Thermo Fisher Scientific #FD0454) and gel-purified (Zymo Research #D4007). The resulting linearized plasmid backbone was then combined with one of the amplified oligonucleotide libraries and assembled into a final gRNA plasmid library by Gibson Assembly (NEB #E5510S)[75]. After DNA purification (Zymo Research #D4013), plasmid gRNA libraries were eluted in 8 µl ultra pure water and 2 µl was electroporated into high-efficiency electrocompetent *Escherichia coli* (Lucigen #60081) using a MicroPulser Electroporator (Bio-Rad Laboratories #1652100). After 60 minutes of growth in liquid recovery media (Lucigen), *E. coli* cells were plated onto a 150-mm LB agar plate supplemented with ampicillin/carbenicillin (100 µg/ml) and kanamycin (50 µg/ml) and incubated overnight at 37ºC. Serial dilution platings and colony counts performed in parallel indicated a total of 1-2 million independent transformants for each sublibrary. Colonies were then washed off the plates and subjected to plasmid extraction (Qiagen #12963).

Canavanine screen

For the canavanine screen, we started with the BY4741 strain previously transformed with pGal-BE3 as described in the section “Base editor characterization with individual gRNAs”. We transformed this strain with a gRNA plasmid library based on the CA gRNA set using a large-scale high-efficiency lithium acetate-based protocol described by Gietz and Schiestl with minor modifications[76]. We scaled up the transformation 10-fold compared to a single transformation reaction and used 10 µg of plasmid DNA for cells from a 50-ml culture at an OD_{600} of 0.7 ($3.5x10^7$ cells/ml). The heat shock at 42ºC lasted 40 min. After the transformation, we plated the cells onto nine selective 100-mm plates (CSM -Leu -Ura). Across all plates, there were approximately 150,000 colonies, which were washed off, pooled and frozen in 20% glycerol at -80ºC.

About $5x10^8$ transformed cells were thawed and grown overnight in 100 ml selective glucose media (CSM -Leu -Ura). The next day, $2.5x10^8$ cells were sampled and $5x10^7$ cells were transferred into 100 ml selective galactose media (CSM/gal -Leu -Ura). After 24 hours of growth in galactose, again $2.5x10^8$ cells were sampled (“0 hours”) and 10^8 cells were transferred into 200 ml canavanine media (60 mg/L L-canavanine sulfate). From the canavanine-treated culture, $2.5x10^8$ cells were sampled after 24 and 48 hours.

Plasmid DNA was extracted from the sampled cells using the Zymoprep Yeast Plasmid MiniPrep II kit (Zymo Research #D2004) but with a modified first step where cell pellets were spheroblasted with 10 U/ml zymolase (amsbio #120491-1) in sorbitol buffer (1 M Sorbitol, 100 mM EDTA pH 8, 14 mM beta-mercaptoethanol). The plasmid region containing the gRNA target sequence and the barcode was amplified using the KAPA HiFi HotStart ReadyMix (KAPA/Roche KK2602) with the following thermocycling protocol: 45 s at 98ºC, 18x (15 s at 98ºC, 30 s at 55ºC, 30 s at 72ºC), 1 min at 72ºC. The forward (OS130-139) and reverse (OS125, PAGE-purified) primers included overhangs that mimic the product of the Illumina Nextera transposase as described above. The forward primer also included a stagger sequence of 9 to 18 nucleotides to introduce diversity in the sequencing library[69]. The resulting fragments were purified (Zymo Research #D4013) and a second round of PCR to extend the Nextera adapters was performed using reagents from the Nextera DNA Library Prep Kit (NPM, PPC and indexes; Illumina #FC-121-1031) according to the manufacturer’s instructions. Concentrations of PCR products were determined (Qubit dsDNA HS Assay Kit, Thermo Fisher Scientific #Q32851) and equal amounts of DNA from each sample were combined. The pooled sample was run on a 2% agarose gel and gel-extracted (Zymo research #D4007). The amplicon library was quantified by qPCR using the KAPA Library Quantification Kit (KAPA/Roche #KK4824). To increase diversity, 10% PhiX Sequencing Control v3 (Illumina #FC-110-3001) was added. The final library was brought to a concentration of 15 pM and sequenced on an Illumina MiSeq with paired-end 171- and 131-base pair reads. Note that some guide sequences are cut-off (18 instead of 20 nucleotides) because sequencing reads were not long enough.

Fitness screen

Note that the fitness screen using the ES gRNA set (Figure S2B) was part of a larger fitness screen using the ES, NS, EP and NP gRNA sets, which was done in duplicate. The fitness screen described later (Figure 4D) included these two replicates as well as two additional replicates of the same screen performed in parallel using just the NS, EP and NP gRNA sets (without the ES set).

The library transformations for the fitness screen were also started with the BY4741 yeast strain previously transformed with pGal-BE3 as described in the section “Base editor characterization with individual gRNAs”. We transformed this strain once with a combined gRNA plasmid library containing the ES, NS, EP and NP gRNA sets and once with a combined library containing only the NS, EP and NP gRNA sets, following a large-scale high-efficiency lithium acetate-based protocol with minor modifications[76]. We scaled up each transformation 10-fold compared to a single transformation reaction and used 10 µg of plasmid DNA for cells from a 100-ml culture at an OD_{600} of 0.6 ($3x10^7$ cells/ml). The heat shock at 42ºC lasted 90 min. After the transformation, cells were transferred into 100 ml selective media (SC -Leu -Ura) and incubated on a shaker at 30ºC. After 24 hours, cultures were diluted 1:50 and grown until they reached stationary phase. Serial dilution platings and colony counts performed in parallel indicated a total of 2-3 million independent transformants for each of the two transformations.

The fitness screen time-course was done in duplicates for each of the two library-transformed yeast strains, resulting in four quasi-replicates. Four pre-cultures inoculated with $1.3x10^8$ cells were grown overnight in 50 ml selective glucose media (SC -Leu -Ura). The next day, $1.3x10^8$ cells were transferred into 50 ml selective galactose media (SC/gal -Leu -Ura). After 24 hours in galactose, $1.3x10^8$ cells were transferred into 50 ml pre-warmed glucose media (SC -Ura). From then on, every 12 hours $1.3x10^8$ cells were transferred into 50 ml pre-warmed glucose media. Samples of $1.3x10^8$ cells were harvested at the time of transfer into galactose media as well as 0, 8, and 24 hours after transfer to glucose media. For harvesting, cells were spun down, resuspended in 200 µl PBS (Gibco/Thermo Fisher Scientific #20012027) and stored at -20ºC.

To facilitate plasmid extraction, harvested cells (in 200 µl PBS) were first spheroblasted by adding zymolase (amsbio #120491-1), DL-dithiotreitol (DTT, Sigma-Aldrich #D5545) and RNase A (PureLink, Thermo Fisher Scientific #12091021) to the following final concentrations: 180 U/ml zymolase, 20 mM DTT and 0.2 mg/ml RNase A. The cells were incubated in this solution for 4 hours at 35ºC, followed by 10 min at 99ºC to inactivate zymolase. After cool-down to room temperature, Qiagen's QIAprep Spin Miniprep Kit was used according to the manufacturer's instructions but omitting the first step (Qiagen #27104).

The plasmid region containing the gRNA target sequence and the barcode was amplified using the KAPA HiFi HotStart ReadyMix (KAPA/Roche KK2602) supplemented with EvaGreen (Biotium #31000) using the following thermocycling protocol: 45 s at 98ºC, 16x (15 s at 98ºC, 30 s at 55ºC, 30 s at 72ºC), 1 min at 72ºC. This was performed on an Agilent AriaMx Real-Time PCR System to be able to monitor and stop the amplification before it started to plateau, which was at 16 cycles. The forward (OS140-149) and reverse (OS125, PAGE-purified) primers included overhangs that mimic the product of the Illumina Nextera transposase as described above. The forward primer also included a stagger sequence of 9 to 18 nucleotides to introduce diversity in the sequencing library[69]. A second round of PCR (6 cycles) to extend the Nextera adapters was performed using the KAPA HiFi HotStart ReadyMix (KAPA/Roche KK2602) again, together with unique dual indexes (IDT for Illumina DNA/RNA UD Indexes, Integrated DNA Technologies). Concentrations of PCR products were determined by Qubit (dsDNA HS Assay Kit, Thermo Fisher Scientific Q32851) and equal amounts of DNA from each sample were combined. The pooled sample was run on a 2% agarose gel and gel-extracted (Zymo research D4007). The amplicon library was quantified by Qubit (dsDNA HS Assay Kit, Thermo Fisher Scientific Q32851). To increase diversity, PhiX Sequencing Control v3 (Illumina #FC-110-3001) was added and the final library was sequenced on an Illumina NextSeq 500 in HighOutput mode with paired-end 80- and 70-base pair reads.

Protein abundance screen

For the protein abundance screens, we selected eleven proteins that are involved in various cellular functions but also highly abundant to ensure detectability by flow cytometry. For each selected protein of interest, we streaked out the corresponding GFP-tagged strains from the GFP collection[12] on YPD plates and from there continued with a single colony from each GFP strain. We first transformed each GFP strain with the base editor plasmid pGal-BE3 using a high-efficiency lithium acetate-based protocol with minor modifications[71]. We then transformed the resulting strains with a combined gRNA library containing the NS, EP and NP gRNA sets, following a large-scale high-efficiency lithium acetate-based protocol with minor modifications[76]. We scaled up each transformation 10-fold compared to a single transformation reaction and used 10 µg of plasmid DNA for cells from a 100-ml culture at an OD_{600} of 0.4 ($2x10^7$ cells/ml). The heat shock at 42ºC lasted 90 min. After the transformation, cells were transferred into 100 ml selective media (SC -Leu -Ura) and incubated on a shaker at 30ºC for 24 hours. Subsequently, cultures were diluted ~1:25 and grown until they reached stationary phase. We prepared multiple cryostocks of $2.5x10^8$ cells for each library-transformed GFP strain and stored them at -80ºC until use. Serial dilution platings and colony counts performed in parallel indicated a total of 2-4 million independent transformants for each of the GFP strains.

The protein screen was performed as follows for each GFP strain separately (on consecutive days). First, a cryostock from a library-transformed GFP strain was inoculated in 100 ml selective glucose media (SC -Leu -Ura) and grown overnight. The next day, 7.5x10^7 cells were centrifuged (1 min at 5000 g), resuspended in 1 ml PBS (Gibco/Thermo Fisher Scientific #20012027) and transferred into 30 ml selective galactose media (SC/gal -Leu -Ura), starting at 2.5x10^6 cells/ml (OD_{600} of 0.05). After 24 hours, eight replicate cultures were set up in 30 ml pre-warmed glucose media (SC -Ura), each starting at 2.5x10^6 cells/ml (OD_{600} = 0.05) from the galactose culture. After 8-9 hours, when the cultures reached mid-exponential growth (~2x10^7 cells/ml or OD_{600} of 0.4-0.5), they were processed for fluorescence-activated cell sorting (FACS). Briefly, 3 ml of each replicate culture was centrifuged (1 min at 3000 g) and resuspended in 3 ml cold PBS (Gibco/Thermo Fisher Scientific #20012027), transferred through a 35-µm cell strainer into a 5-ml polystyrene round-bottom test tube (Falcon #352235) and kept on ice from then on. The sorting was performed on a BioRad S3e cell sorter with a cooled sample block. For the sorting, only the 40-50% most similar cells in terms of size and granularity were used (oval gate at densest area on forward scatter (FSC) vs. side scatter (SSC) plot, linear axes). From that FSC-SSC gate, we first sorted 500,000 cells as a control population. Second, we sorted 200,000 cells from the 5% of cells with the highest and the 5% of cells with the lowest GFP signal each (rectangular gates on GFP histogram plot, log axis). Sort speed was set to 10,000 events per second and sorted cells were collected in 5-ml polystyrene round-bottom test tubes (Falcon #352058). The sorted cells (in sheath fluid (PBS), BioRad #12012932) were then transferred into 1.5-ml tubes and centrifuged at full speed for 30 seconds. Supernatant was carefully removed, leaving ~200 µl behind to minimize disturbing the invisible cell pellet. The sorted and concentrated cell samples were then stored at -20ºC until all sorts for all GFP strains were completed.

To facilitate plasmid extraction, harvested cells (in 200 µl PBS) were first spheroblasted by adding zymolase (amsbio #120491-1), DL-dithiotreitol (DTT, Sigma-Aldrich #D5545) and RNase A (PureLink, Thermo Fisher Scientific #12091021) to the following final concentrations: 180 U/ml zymolase, 20 mM dithiotreitol and 0.2 mg/ml RNase A. The cells were incubated in this solution for 2-3 hours at 35ºC, followed by 10 min at 99ºC to inactivate zymolase. After cool-down to room temperature, Qiagen's QIAprep Spin Miniprep Kit was used according to the manufacturer's instructions but omitting the first step (Qiagen #27104).

The plasmid region containing the gRNA target sequence and the barcode was amplified using the KAPA HiFi HotStart ReadyMix (KAPA/Roche KK2602) supplemented with EvaGreen (Biotium #31000) using the following thermocycling protocol: 45 s at 98ºC, 22x (15 s at 98ºC, 30 s at 55ºC, 30 s at 72ºC), 1 min at 72ºC. This was performed on an Agilent AriaMx Real-time PCR system in order to be able to monitor and stop the amplification before it started to plateau, which was at 22 cycles. The forward (OS140-149) and reverse (OS125, PAGE-purified) primers included overhangs that mimic the product of the Illumina Nextera transposase as described above. The forward primer also included a stagger sequence of 9 to 18 nucleotides to introduce diversity in the sequencing library[69]. A second round of PCR (6 cycles) to extend the Nextera adapters was performed using the KAPA HiFi HotStart ReadyMix (KAPA/Roche KK2602) again, together with unique dual indexes (IDT for Illumina DNA/RNA UD Indexes, Integrated DNA Technologies). Concentrations of PCR products were determined using Qubit reagents (dsDNA HS Assay Kit, Thermo Fisher Scientific Q32851) but in 96-well format with fluorometric readings at 485/20 nm excitation and 528/20 nm emission performed on a BioTek Synergy2 plate reader using black opaque flat-bottom 96-well plates (Costar #3915). We then combined roughly equal amounts of DNA from samples of cells sorted for high or low GFP levels and double that amount for the control samples that were sorted without GFP gates. The pooled sample was run on a 2% agarose gel and gel-extracted (Zymo research D4007). The amplicon library was quantified by Qubit (dsDNA HS Assay Kit, Thermo Fisher Scientific Q32851). To increase diversity, PhiX Sequencing Control v3 (Illumina #FC-110-3001) was added and the final library was sequenced on an Illumina NovaSeq 6000 using an S1 flow cell with paired-end 100-base pair reads.

Due to underrepresentation of some of the sorted Tdh3-GFP sub-libraries in the final library mentioned above, we pooled these samples separately again, and gel-purified and quantified them as described above, followed by addition of PhiX Sequencing Control v3 (Illumina #FC-110-3001) and sequencing on an Illumina NextSeq 500 in HighOutput mode with paired-end 80- and 70-base pair reads. Read counts were then downsampled using seqtk (https://github.com/lh3/seqtk) to get comparable read numbers to the other GFP strains.

Fitness screen analysis

Fitness screen analyses were performed with custom R scripts run in RStudio (R version 4.0.3, RStudio version 1.3). For the fitness screen data shown in Figure S2B (ES gRNA set only, 2 replicates), we first extracted guide sequences from sequencing reads and tallied the read counts per guide. To determine which gRNAs change in abundance during prolonged culturing of base edited yeast, we used read counts per gRNA as a quantitative measure. We normalized the read counts across samples so that the total read counts of the control gRNAs remain constant. We then calculated for each gRNA the $\log_2$ fold change (log2FC) of the number of reads at each time point versus to the “pre” time point sampled before induction of the base editor with galactose. The reported log2FC values are averages across the two replicate experiments. To obtain gene-level effects, we used the largest absolute log2FC among all gRNAs targeting the same gene. The fitness screen data shown in Figure 4E (EP, NP and NS gRNA sets, four replicates) was processed as described above but with normalization for total read counts across samples because not all four replicates included the control gRNAs.

Protein abundance screen analysis

Protein abundance screen analyses were performed with custom R scripts run in RStudio (R version 4.0.3, RStudio version 1.3). First, for each sample, the 20-nucleotide guide and the 20-nucleotide barcode sequences were extracted from paired-end read pairs.

Next, for all barcodes paired with guides perfectly matching the library, we performed error correction to account for potential sequencing errors that would otherwise lead to inflated barcode counts. For this, read counts of barcodes with a difference of less than five nucleotides were assigned the same barcode sequence. We then removed barcodes that did not have at least two reads and gRNAs that did not have at least two barcodes in at least four of the control samples per GFP strain. We also excluded replicates with very low read counts (Eno2_A, Fas1_A , Fas1_F, Fas2_A, Fas2_D, Htb2_A). To determine which gRNAs are differentially abundant among cells with high vs low GFP levels, we used the number of unique barcodes per gRNA as a quantitative measure which we normalized across samples by total unique barcode count per sample. We then applied a generalized linear model (glm) with Poisson-distributed errors and log as a link function. The glm function outputs for each gRNA the ln("high GFP"/"low GFP") and a p-value. We converted the ln fold change to $\log_2$ fold change (log2FC). In cases where for a gRNA no barcodes were detected in any of the high or low GFP replicates, the glm function assigns an estimate of 20 or -20, respectively, which we replaced by 3 or -3, respectively. To adjust p-values for multiple testing and control the false discovery rate (FDR), we used the qvalue R package version 2.22.0[77]. Unless otherwise noted, we consider "significant" a q-value (FDR) < 0.05.

To obtain gene-level log2FC values, we used the largest absolute fold change estimate of all gRNAs targeting a gene. For gene-level q-values, we first combined the p-values of all gRNAs targeting the same gene using Fisher's method (R package poolr, version 0.8-2) and then corrected these gene-level p-values for multiple testing again with the qvalue R package. We compared this combined q-value to the q-value of the gRNA with the largest absolute fold change per gene and then took the lower of the two q-values as a final per-gene q-value. This approach raises the number of genes with significant effect (FDR < 0.05) on protein abundance from 622 to 710, compared to considering just genes targeted by significant gRNAs.

Cytoscape

For network visualizations we used Cytoscape (version 3.8.2)[17]. For the network representing the overall results from the screen (Figure 1B), the eleven proteins were defined as source nodes and the significant gene perturbations as target nodes. A diverging red-blue color gradient was chosen to reflect the log fold changes caused by the perturbation. Note that for perturbations that affect multiple proteins, the selected color depends on the order the interactions are listed in the input. The size of the target nodes correlates with the number of source nodes (proteins) it connects to. The arrangement of nodes was obtained by the yFiles Organic Layout or the Edge-weighted Spring Embedded Layout and further adjusted manually.

For the small protein-protein-interaction network of the broad-effect gene perturbations (Figure 4C), the Cytoscape StringApp was used[35,36]. The settings for STRING were as follows: the species was set to *Saccharomyces cerevisiae* and the confidence cutoff was set to 0.4. All view defaults were unselected. The type of chart to draw was set to pie chart, with two terms to chart and an overlap cutoff of 0.5. The arrangement of nodes was obtained by applying the yFiles Organic Layout and further adjusted manually.

Gene Ontology (GO) enrichment analysis

Enrichment analysis for functional annotations was performed in R using the STRINGdb package version 2.2.2[35]. This package computes the enrichment using a hypergeometric test and adjusts p-values for multiple testing using the Benjamini-Hochberg method. For the GO enrichment analysis among broad and specific regulators, the background was defined as all genes represented by our gRNA library (NS, ES, EP subsets). For the GO enrichment analysis among sets of significantly changed proteins in the proteomics data, the background was defined as the union of proteins quantified across all samples used for the enrichment analysis.

Validation experiments (RAS2, IRA2, SSY5, POP1, SIT4)

For validation of the findings from the protein screen, we generated individual mutants based on the BY4741 strain. We first cloned individual gRNA plasmids for relevant gRNAs targeting RAS2, IRA2, SSY5, POP1 and SIT4 as described above in the "Plasmids" section. These plasmids were then individually transformed into the BY4741 yeast strain previously transformed with pGal-BE3 (described in the section "Base editor characterization with individual gRNAs") using the same high-efficiency lithium acetate-based protocol with minor modifications[71]. From the successful transformants growing on selective plates (SC -Leu -Ura) we then inoculated an overnight preculture in selective glucose media (SC -Leu -Ura) followed by a 24-hour growth period in selective galactose media (SC/Gal -Leu -Ura) to induce base editor expression. Subsequently, cells were grown in non-selective liquid YPD media for 8 hours to promote plasmid loss, followed by growth on YPD plates for up to 5 days. Note that colonies on some plates grew very slowly because several of the targeted base edits affect cell fitness. From each plate we analyzed 4-8 individual colonies to see whether they acquired the expected base edits. First, a small amount of each colony was transferred into a 96-well PCR plate containing 50 µl ultra pure H_2O (Invitrogen/Thermo Fisher Scientific #10977015). Of the resulting cell solution, 25 µl was transferred into a 96-deep-well plate containing 500 µl YPD. The remaining 25 µl were boiled at 99ºC for 5 min to lyse the cells. Then 2 µl of the heat-treated cells were used to set up 50-µl PCR reactions with TaKaRa Ex Taq DNA Polymerase (TaKaRa #RR001A) and specific primer pairs designed to amplify the genomic region targeted by each gRNA. The following thermal cycling protocol was used: 1 min at 94ºC, 35x (30 s at 94ºC, 30 s at 50ºC, 1 min at 72ºC), 3 min at 72ºC. PCR products were sent for Sanger sequencing. Three independent mutant strains harboring the expected base edits were then selected and the corresponding cultures in YPD that were kept at room temperature were converted into cryo stocks and stored at -80ºC until use.

LC-MS proteomics data acquisition

For proteomics analyses, individual mutant strains described above were streaked out on YPD plates and incubated at 30ºC for several days. Overnight cultures were then inoculated from a single colony of each mutant strain in SC media. The next day, 3-ml cultures were set up at 2.5×10^6 cells/ml (OD_{600} = 0.05) and grown for several hours until they reached a density of 2.5×10^7 cells/ml (OD_{600} = 0.5). Of each culture, 2 ml were harvested by centrifugation for 1 min at 5000 g at 4ºC. Supernatant was removed and cell pellets stored at -80ºC. For cell lysis, 200 µl freshly prepared lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris-HCl at pH 8) was added to each cell pellet. The resulting solution was transferred to 2-ml screw-cap tubes containing a 100-µl volume of glass beads (500 µm, Sigma G8772). Cells were lysed by 3x 5 min bead beating at 2400 rpm (40 Hz) on a BioSpec Mini-BeadBeater-96, with 1-2 min cooling on ice between cycles. Samples were then centrifuged for 5 min at full speed (at room temperature to avoid precipitation of urea) and cleared lysates transferred to new tubes. Protein concentrations were determined with a colorimetric bicinchoninic acid (BCA) assay (Pierce / Thermo Scientific #23225). Absorbance at 562 nm was measured on a BioTek Synergy2 plate reader which indicated protein concentrations of 1-2 mg/ml.

For each sample, 25-50 µg of protein was further processed as follows. Protein disulfide bonds were reduced by the addition of 5 mM Tris (2-carboxyethyl)phosphine (TCEP) and incubation at room temperature for 30 min. Next, the free cysteine residues were alkylated by the addition of 10 mM iodoacetamide and incubation at room temperature for 30 min in the dark. Subsequently, samples were diluted with 100 mM Tris-HCl at pH 8 to reach a urea concentration of less than 2 M. For the protein digest, 0.2 µg LysC (Wako #125-05061) and 0.8 µg trypsin (Pierce / Thermo Scientific, #90057) were added and samples were incubated at 37°C overnight. The digest was quenched by adding formic acid to a final concentration of 5% (v/v). Finally, samples were desalted using C18 tips (Pierce / Thermo Scientific #87784), dried under vacuum and reconstituted in 5% formic acid (v/v).

For liquid chromatography-coupled tandem mass spectrometric (LC-MS/MS) analysis, desalted peptides were fractionated online by nano-flow reversed phase liquid chromatography using a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific). The column consisted of a 25-cm fused silica capillary with 75 µm inner diameter that was packed in-house[78] with ReproSil-Pur C18 particles (particle size 1.9 µm; pore size 120 Å; Dr. Maisch). The linear 140-min water–acetonitrile gradient from 5% to 35% acetonitrile was delivered at a flow rate of 300 nl/min. Mass spectrometric analysis was performed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) run in Data-Dependent Acquisition (DDA) mode and with an inclusion list containing a total of six peptides from Tdh1 and Tdh2 to ensure their consistent quantification (IDVAVDSTGVFK(2+), LISWYDNEYGYSAR(2+), VVDLIEYVAK(2+), HIIVDGHK(2+), DPANLPWASLNIDIAIDSTGVFK(2+), GVLGYTEDAVVSSDFLGDSNSSIFDAAAGIQLSPK(3+)). MS1 spectra were acquired at a resolution of 120,000 (FWHM at 400 m/z) over a range of 400 to 1600 m/z with a maximum injection time of 100 ms. MS1 acquisition was followed by the acquisition of MS2 spectra from precursors falling into the inclusion list windows (monoisotopic m/z +/- 25 ppm). Precursors were isolated with a window of 1.6

m/z and fragmentation was obtained by higher-energy C-trap dissociation (HCD) at a fixed collision energy of 35%. MS2 spectra were acquired at a resolution of 30,000, a maximum injection time of 54 ms and dynamic exclusion for 5 s. The remainder of the 3-s cycle time was used to acquire MS2 spectra for other precursors at a resolution of 15,000 and an exclusion window of 25 s but otherwise identical parameters as for the inclusion list precursors.

LC-MS proteomics data analysis

Peptide identification and quantification was performed with MaxQuant version 1.6.17.0[79]. The search was performed against a *S. cerevisiae* S288C protein database containing 5917 protein sequences (version R64-2-1, Saccharomyces Genome Database (SGD)). Only fully LysC- and trypsin-digested peptides with up to one missed cleavage were allowed. LysC was specified to cleave after lysine even if followed by proline and trypsin was specified to cleave after lysine and arginine if not followed by proline. Carbamidomethylation of cysteines was defined as a fixed modification and methionine oxidation and N-terminal acetylation as a variable modification. Peptide-spectrum-match (PSM) and protein-level false discovery rates (FDRs) were set to 0.01. Further analyses were performed with custom R scripts run in RStudio (R version 4.0.3, RStudio version 1.3). From the MaxQuant output (evidence.txt), all non-unique peptides were removed before further processing. Relative quantification was then performed in R using the artMS package version 1.8.3 (http://artms.org), which itself is based on the MSstats package[80]. Briefly, $\log_2$-transformed data was quantile normalized and summarized by Tukey's median polish. Missing values were imputed and p-values were adjusted for multiple testing by the Benjamini-Hochberg method. Enrichment analysis for functional annotations among sets of significantly changed proteins in different mutant strains was performed as described above with the union of proteins quantified across all samples used for the enrichment analysis as background.

Testing effect of sequence context on editing efficiency

To test whether the sequence context in the targeting region of the gRNA has an effect on editing efficiency, we made use of the data from the fitness screen, considering only gRNAs introducing stop codons into essential genes (ES subset) and the control gRNAs without a C in the target region. We considered the 20-nucleotide gRNA targeting sequence flanked on either side by an additional 5 nucleotides for a total of 30 nucleotides per gRNA. For each of the 5928 gRNAs, we then extracted all possible position-specific sequence features up to 4 nucleotides in length, resulting in a total of over 8000 sequence features across all gRNAs[81]. The gRNAs were split into a training set of 4928 gRNAs and a test set of 1000 gRNAs. An ordinary and a logistic lasso regression model was then obtained from the training set. For the logistic lasso model, we defined gRNAs with a log2FC < -0.5 as dropping out. Evaluated on the test set, the lasso-predicted and observed effects show a Pearson correlation coefficient (R^2) of 0.37 and 0.35 for the ordinary and the logistic model, respectively.

Comparison to eQTL and pQTL hotspots in natural yeast isolates

Our goal here was to identify candidate causal genes that could underlie observed eQTL and pQTL hotspots in segregants of the *S. cerevisiae* strains BY and RM11-1a[33,34]. Among the genes located in the eQTL and pQTL hotspot intervals, we identified those that affect at least three of our eleven proteins. The resulting list of gene candidates was narrowed down to those that contain at least one variant between BY and RM with a predicted medium or strong effect (i.e., missense mutation, indel, or premature stop codon) (Table S7).

References

1. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* **21**, 630–644 (2020).
2. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell **165**, 535–550 (2016).
3. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat Rev Genet 1–19 (2020) doi:10.1038/s41576-019-0209-0.
4. Hanna, R. E. *et al.* Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064-1080.e20 (2021).
5. Dixit, A. et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell **167**, 1853-1866.e17 (2016).
6. Adamson, B. et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell 167, 1867-1882.e21 (2016).
7. Jaitin, D. A. et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell 167, 1883-1896.e15 (2016).
8. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* (2017) doi:10.1038/nmeth.4177.
9. Cuella-Martin, R. et al. Functional interrogation of DNA damage response variants with base editing screens. Cell 184, 1081-1097.e19 (2021).
10. Després, P. C., Dubé, A. K., Seki, M., Yachie, N. & Landry, C. R. Perturbing proteomes at single residue resolution using base editing. *Nat Commun* **11**, 1871 (2020).
11. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature **533**, 420–424 (2016).
12. Huh, W.-K. et al. Global analysis of protein localization in budding yeast. Nature **425**, 686–691 (2003).
13. Howson, R. et al. Construction, verification and experimental use of two epitope-tagged collections of budding yeast strains. *Comparative and functional genomics* **6**, 2–16 (2005).
14. Newman, J. R. S. et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature **441**, 840–846 (2006).

15. Soste, M. et al. A sentinel protein assay for simultaneously quantifying cellular processes. Nature Methods 11, 1045–1048 (2014).
16. Sadhu, M. J. et al. Highly parallel genome variant engineering with CRISPR–Cas9. *Nature Genetics* **50**, 1–11 (2018).
17. Shannon, P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **13**, 2498–2504 (2003).
18. Baker, H. V. Glycolytic gene expression in Saccharomyces cerevisiae: nucleotide sequence of GCR1, null mutants, and evidence for expression. *Mol Cell Biol* **6**, 3774–3784 (1986).
19. Holland, M. J., Yokoi, T., Holland, J. P., Myambo, K. & Innis, M. A. The GCR1 gene encodes a positive transcriptional regulator of the enolase and glyceraldehyde-3-phosphate dehydrogenase gene families in Saccharomyces cerevisiae. Mol Cell Biol 7, 813–820 (1987).
20. Kurat, C. F. et al. Regulation of histone gene transcription in yeast. *Cell Mol Life Sci* **71**, 599–613 (2014).
21. Rudra, D., Zhao, Y. & Warner, J. R. Central role of Ifh1p–Fhl1p interaction in the synthesis of yeast ribosomal proteins. *Embo J* **24**, 533–542 (2005).
22. Cassanova, N., O'Brien, K. M., Stahl, B. T., McClure, T. & Poyton, R. O. Yeast Flavohemoglobin, a Nitric Oxide Oxidoreductase, Is Located in Both the Cytosol and the Mitochondrial Matrix EFFECTS OF RESPIRATION, ANOXIA, AND THE MITOCHONDRIAL GENOME ON ITS INTRACELLULAR LEVEL AND DISTRIBUTION*. *J Biol Chem* **280**, 7645–7653 (2005).
23. Zhu, H. & Riggs, A. F. Yeast flavohemoglobin is an ancient protein related to globins and a reductase family. *Proc National Acad Sci* **89**, 5015–5019 (19**92**).
24. Forsberg, H., Gilstring, C. F., Zargari, A., Martínez, P. & Ljungdahl, P. O. The role of the yeast plasma membrane SPS nutrient sensor in the metabolic response to extracellular amino acids. *Mol Microbiol* **42**, 215–228 (2001).
25. Williams, R. B. & Lorenz, M. C. Multiple Alternative Carbon Pathways Combine To Promote Candida albicans Stress Resistance, Immune Interactions, and Virulence. *Mbio* **11**, (2020).
26. Danhof, H. A. et al. Robust Extracellular pH Modulation by Candida albicans during Growth in Carboxylic Acids. Mbio **7**, e0**16**46-16 (2016).
27. Tsai, K.-L. et al. A conserved Mediator–CDK8 kinase module association regulates Mediator–RNA polymerase II interaction. *Nat Struct Mol Biol* **20**, 611–619 (2013).
28. Tsai, K.-L. et al. Subunit Architecture and Functional Modular Rearrangements of the Transcriptional Mediator Complex. Cell **157**, 1430–1444 (2014).
29. Elmlund, H. et al. The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. Proc National Acad Sci **103**, 15788–15793 (2006).
30. Luke, M. M. et al. The SAP, a new family of proteins, associate and function positively with the SIT4 phosphatase. Mol Cell Biol 16, 2744–2755 (1996).
31. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. Nature **411**, 41–42 (2001).
32. Batada, N. N., Hurst, L. D. & Tyers, M. Evolutionary and Physiological Importance of Hub Proteins. *Plos Comput Biol* **2**, e88 (2006).
33. Albert, F. W., Treusch, S., Shockley, A. H., Bloom, J. S. & Kruglyak, L. Genetics of single-cell protein abundance variation in large yeast populations. Nature **506**, 494–497

(2014).
34. Albert, F. W., Bloom, J. S., Siegel, J., Day, L. & Kruglyak, L. Genetics of trans-regulatory variation in gene expression. *eLife* **7**, (2018).
35. Szklarczyk, D. et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, gkaa1074- (2020).
36. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res* **18**, 623–632 (2018).
37. Ringel, A. E. et al. Yeast Tdh3 (Glyceraldehyde 3-Phosphate Dehydrogenase) Is a Sir2-Interacting Factor That Regulates Transcriptional Silencing and rDNA Recombination. *Plos Genet* **9**, e1003871 (2013).
38. White, M. R. & Garcin, E. D. D-Glyceraldehyde-3-Phosphate Dehydrogenase Structure and Function. in *Macromolecular Protein Complexes* vol. 83 413–453 (20**17**).
39. Kosova, A. A., Khodyreva, S. N. & Lavrik, O. I. Role of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in DNA repair. *Biochem Mosc* **82**, 643–654 (2017).
40. Randez-Gil, F., Sánchez-Adriá, I. E., Estruch, F. & Prieto, J. A. The formation of hybrid complexes between isoenzymes of glyceraldehyde-3-phosphate dehydrogenase regulates its aggregation state, the glycolytic activity and sphingolipid status in Saccharomyces cerevisiae. *Microb Biotechnol* **13**, 562–571 (2019).
41. McAlister, L. & Holland, M. J. Isolation and characterization of yeast strains carrying mutations in the glyceraldehyde-3-phosphate dehydrogenase genes. *J Biological Chem* **260**, 15013–8 (1985).
42. McAlister, L. & Holland, M. J. Differential expression of the three yeast glyceraldehyde-3-phosphate dehydrogenase genes. J Biological Chem 260, 15019–27 (1985).
43. Casanovas, A. et al. Quantitative analysis of proteome and lipidome dynamics reveals functional regulation of global lipid metabolism. *Chem Biol* **22**, 412–25 (2015).
44. Murphy, J. P., Stepanova, E., Everley, R. A., Paulo, J. A. & Gygi, S. P. Comprehensive Temporal Protein Dynamics during the Diauxic Shift in Saccharomyces cerevisiae. *Mol Cell Proteom Mcp* **14**, 2454–65 (2015).
45. Valadi, H. et al. NADH-reductive stress in Saccharomyces cerevisiae induces the expression of the minor isoform of glyceraldehyde-3-phosphate dehydrogenase (TDH1). *Curr Genet* **45**, 90–95 (2004).
46. Costenoble, R. et al. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. *Molecular Systems Biology* **7**, 464–464 (2011).
47. Kayikci, Ö. & Nielsen, J. Glucose repression in Saccharomyces cerevisiae. *Fems Yeast Res* **15**, fov068 (2015).
48. Dechant, R. & Peter, M. Nutrient signals driving cell growth. *Curr Opin Cell Biol* **20**, 678–687 (2008).
49. Smith, A., Ward, M. P. & Garrett, S. Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. Embo J 17, 3556–3564 (1998).
50. Conrad, M. et al. Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae. *Fems Microbiol Rev* **38**, 254–299 (2014).

51. Bhattacharya, S., Chen, L., Broach, J. R. & Powers, S. Ras membrane targeting is essential for glucose signaling but not for viability in yeast. Proc National Acad Sci 92, 2984–2988 (1995).
52. Srivas, R. et al. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. *Molecular cell* **63**, 514–525 (2016).
53. Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).
54. Costanzo, M. et al. The genetic landscape of a cell. Science **327**, 425–431 (2010).
55. Oughtred, R. et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* **30**, 187–200 (2021).
56. Peeters, K. et al. Fructose-1,6-bisphosphate couples glycolytic flux to activation of Ras. Nat Commun 8, 922 (2017).
57. Griffioen, G., Swinnen, S. & Thevelein, J. M. Feedback Inhibition on Cell Wall Integrity Signaling by Zds1 Involves Gsk3 Phosphorylation of a cAMP-dependent Protein Kinase Regulatory Subunit*. J Biol Chem **278**, 23460–23471 (2003).
58. Bandaru, P. et al. Deconstruction of the Ras switching cycle through saturation mutagenesis. *Elife* **6**, e27810 (2017).
59. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer* **3**, 11–22 (2003).
60. Tate, J. G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 47, gky1015- (2018).
61. Nadeau, J. H. & Dudley, A. M. Systems Genetics. Science **331**, 1015–1016 (2011).
62. Kinney, J. B. & McCandlish, D. M. Massively Parallel Assays and Quantitative Sequence–Function Relationships. *Annu Rev Genom Hum G* **20**, 1–29 (2019).
63. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res 47, gky1106- (2018).
64. Brachmann, C. B. et al. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)* **14**, 115–132 (1998).
65. DiCarlo, J. E. et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research* **41**, 4336–4343 (2013).
66. Mumberg, D., Muller, R. & Funk, M. Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res 22, 5767–5768 (1994).
67. Mumberg, D., Müller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene **156**, 119–122 (1995).
68. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. Nature Methods 11, 783–784 (2014).
69. Joung, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols* **12**, 828–863 (2017).
70. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *Plos One* **3**, e3647 (2008).
71. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature Protocols 2, 31–34 (2007).
72. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology* **37**, 224–2**26** (2019).

73. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337**, 816–821 (2012).
74. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688 (20**12**).
75. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 6, 343–345 (2009).
76. Gietz, R. D. & Schiestl, R. H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature Protocols 2, 38–41 (2007).
77. Storey, J. D. A direct approach to false discovery rates. *J Royal Statistical Soc Ser B Statistical Methodol* **64**, 479–498 (2002).
78. Jami-Alahmadi, Y., Pandey, V., Mayank, A. K. & Wohlschlegel, J. A. A Robust Method for Packing High Resolution C18 RP-nano-HPLC Columns. *J Vis Exp* (2021) doi:10.3791/62380.
79. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology 26, 1367–1372 (2008).
80. Choi, M. et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics (Oxford, England)* **30**, 2524–2526 (2014).
81. Rahman, M. K. & Rahman, M. S. CRISPRpred: A flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems. PLoS ONE 12, e0181943 (2017).