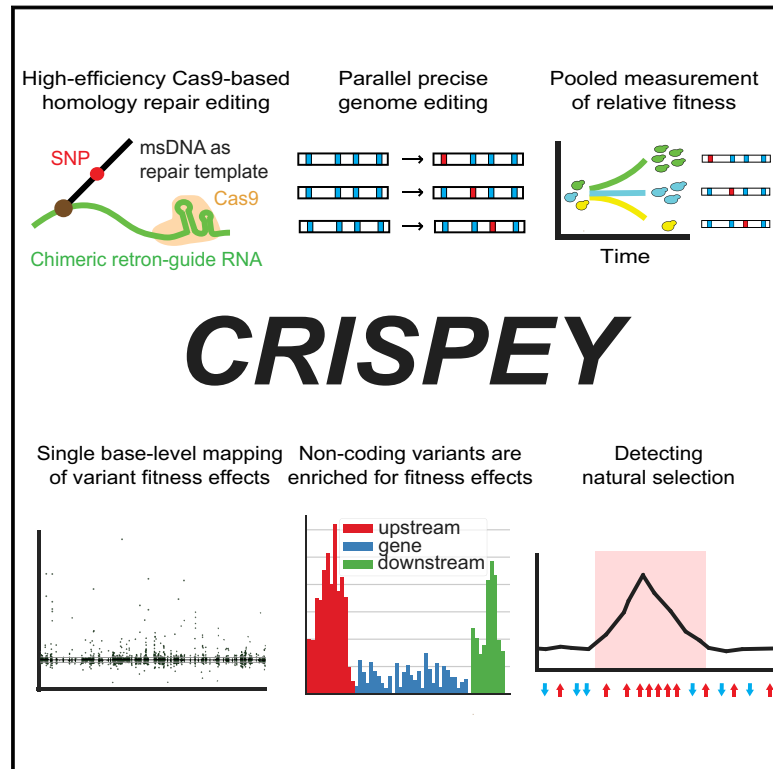


Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing

Graphical Abstract



Authors

Eilon Sharon, Shi-An A. Chen,
Neil M. Khosla, Justin D. Smith,
Jonathan K. Pritchard, Hunter B. Fraser

Correspondence

eilon@stanford.edu (E.S.),
hbfraser@stanford.edu (H.B.F.)

In Brief

A highly efficient Cas9-based method for high-throughput precise genome editing is developed and used to measure the fitness effects of thousands of natural genetic variants in yeast at single-base resolution.

Highlights

- CRISPEY—a method for highly efficient, parallel precise genome editing
- Applied to measure fitness effects of thousands of natural genetic variants in yeast
- Variants affecting fitness are enriched in promoters and TF binding sites
- Nearby variants mostly favor the same strain's alleles, indicating natural selection



Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing

Eilon Sharon,^{1,2,3,4,*} Shi-An A. Chen,^{1,4} Neil M. Khosla,^{1,4} Justin D. Smith,² Jonathan K. Pritchard,^{1,2,3} and Hunter B. Fraser^{1,5,*}

¹Department of Biology, Stanford University, Stanford, CA 94305, USA

²Department of Genetics, Stanford University, Stanford, CA 94305, USA

³Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: eilon@stanford.edu (E.S.), hbfraser@stanford.edu (H.B.F.)

<https://doi.org/10.1016/j.cell.2018.08.057>

SUMMARY

A major challenge in genetics is to identify genetic variants driving natural phenotypic variation. However, current methods of genetic mapping have limited resolution. To address this challenge, we developed a CRISPR-Cas9-based high-throughput genome editing approach that can introduce thousands of specific genetic variants in a single experiment. This enabled us to study the fitness consequences of 16,006 natural genetic variants in yeast. We identified 572 variants with significant fitness differences in glucose media; these are highly enriched in promoters, particularly in transcription factor binding sites, while only 19.2% affect amino acid sequences. Strikingly, nearby variants nearly always favor the same parent's alleles, suggesting that lineage-specific selection is often driven by multiple clustered variants. In sum, our genome editing approach reveals the genetic architecture of fitness variation at single-base resolution and could be adapted to measure the effects of genome-wide genetic variation in any screen for cell survival or cell-sortable markers.

INTRODUCTION

Traditionally, the phenotypic effects of natural genetic variation have been studied via quantitative trait locus (QTL) mapping (Mackay et al., 2009; Pickrell et al., 2010; Rockman and Kruglyak, 2006). In this approach, genetic crosses result in hundreds or even thousands of offspring, each with a distinct genotype that is measured at SNPs or other genetic variants spanning the genome. These genotypes are then correlated with a phenotype of interest to identify QTLs where genotype predicts phenotype, indicating a genomic region harboring one or more variants that affect the phenotype.

Although genetic mapping of phenotypic variation has been possible for over a century (Sturtevant, 1913), mapping resolution is limited by recombination events and few maps have

been resolved to the level of causal genetic variants (Rockman, 2012). Genome-wide association studies (GWAS) offer improved mapping resolution but still only rarely are able to unambiguously implicate specific causal variants (Schaid et al., 2018). As a result, we still do not know the variants, genes, and molecular mechanisms underlying phenotypic variation, even for most well-studied traits (Rockman, 2012).

Because pinpointing a causal variant requires laborious follow-up experiments, investigators usually prioritize which variants to test within a QTL. This has led to a well-known ascertainment bias: investigators prioritize QTLs with large effects for further study, and within these QTLs, they further prioritize missense variants within promising candidate genes (Rockman, 2012). Although each of these large-effect variants represents a fascinating example, these outliers may not be very informative for understanding the evolutionary process more generally, which is thought to be dominated by variants of much smaller effect (Pritchard et al., 2010). An unbiased method for systematically identifying causal variants is therefore required to fully understand the genes, molecular mechanisms, genetic architectures, and selection pressures underlying phenotypic variation.

In principle, QTL mapping—which can require years of effort, especially for slow-breeding species—could be circumvented if each individual genetic variant differing between the two parents could be introduced into the genome of a single reference strain and scored for the phenotype of interest. Indeed, CRISPR/Cas9 has recently transformed our ability to characterize individual genetic variants via genome editing (Cong et al., 2013; DiCarlo et al., 2013; Jinek et al., 2012; Mali et al., 2013). Following a Cas9-mediated double-strand break, precise edits can be engineered by homology-directed repair (HDR) utilizing a donor DNA containing the mutation of interest (Rouet et al., 1994). However, this process is typically low-throughput, and often has variable efficiency of successful edits (Doudna and Charpentier, 2014; Hsu et al., 2014; Liu et al., 2017).

In contrast, high-throughput genome editing has been performed by creating double-strand breaks in the absence of any donor DNA. These are repaired by non-homologous end-joining (NHEJ), often resulting in small insertions or deletions (indels) that lead to frameshifts when targeting exons. In contrast to HDR editing, NHEJ mutations are heterogeneous and thus are



not useful for studying specific variants of interest. In principle a library of donor DNAs, encoding desired mutations, could be introduced together with the single guide RNAs (sgRNAs). However, two limitations have prevented this. First, because donor DNAs are typically encoded on separate molecules from the sgRNAs, there would be no way to ensure that the corresponding sgRNA/donor DNA pairs (targeting the same site) would get into the same cells—the combinations would instead be random. Second, even if they could be delivered to the same cells, the efficiency of precise editing is often low and variable across different sites (Liu et al., 2017), introducing an unacceptable level of noise into any downstream phenotyping assay. Several methods have begun to explore high-throughput precision genome editing, but these involve HDR from high-copy plasmids that may limit their general applicability (Bao et al., 2018; Garst et al., 2017; Guo et al., 2018; Roy et al., 2018; Sadhu et al., 2018) (see Table S1).

To boost editing efficiency and enable high-throughput screens, we reasoned that generating a large number of potential donor DNA molecules within the nucleus would maximize the chance of HDR. Bacterial retrons may be ideal candidates for generating these donor DNAs. Retrongs are natural DNA elements coding for a reverse transcriptase (RT), as well as a template on which the RT acts, to create a multi-copy single-stranded DNA (msDNA) product (Hsu et al., 1990, 1992; Shimamoto et al., 1993). These msDNAs are covalently tethered to their template RNA by the RT. Following their discovery in the 1980s, retrongs were found to be active in yeast and mammalian cells, and any desired sequence can be reverse-transcribed into msDNA (Mirochnitchenko et al., 1994; Miyata et al., 1992). As we show below, coupling retrongs with CRISPR/Cas9 enables precise genome editing with very high efficiency and throughput. This allows us to pinpoint the functional variants underlying variation in fitness—the only phenotype that matters for natural selection—in any given environment.

RESULTS

High-Efficiency Genome Editing via Retron-Generated Donor DNA

To enable parallel, high-efficiency HDR editing, we devised a strategy employing a bacterial retron element for generating donor single-stranded DNAs (ssDNAs) *in vivo*. We reasoned that retrongs could generate many single-stranded donors that are more efficient in HDR than dsDNA donors (Richardson et al., 2016). We designed a chimeric RNA of *E. coli* Ec86 retron sequence joined with a sgRNA at the 3' end (Figure 1A). A portion of the bacterial retron sequence was replaced with a 100-nt donor sequence that contained the desired mutation flanked by ~50 bp homology arms (Figure 1B). Southern blot analysis showed that msDNA was produced in yeast in an RT-dependent manner (Figure 1C). We refer to this method as Cas9 Retron preClSe Parallel Editing via homologyY (CRISPEY).

We first tested CRISPEY at a single locus in haploid *Saccharomyces cerevisiae*. To visualize editing efficiency, we introduced a nonsense mutation into *ADE2* that results in pink-colored colonies. *S. pyogenes* Cas9 (SpCas9) and Ec86 reverse transcriptase (Ec86-RT) were integrated in the genome, and the retron-sgRNA was encoded on a low-copy plasmid (Figure 1B).

Remarkably, editing efficiency reached 100% after 48 hr (Figure 1D; Table S2), with 0/95 sequenced colonies having undesired mutations in *ADE2* (Table S3). As expected, editing was dependent on expression of all components: pairing of *ADE2* targeting retron-donor sequence and sgRNA, as well as expression of the RT are required for high-efficiency (Figure 1D). In addition, CRISPEY can insert a 765-nt sequence with 92% efficiency and <0.01% mutation rate within the insertion (Figure 1E; Tables S3 and S4). In summary, we found that CRISPEY is highly efficient and precise.

Genome-wide Measurement of the Fitness Effects of Natural Genetic Variants

We next turned our attention to a long-standing question in evolutionary biology: which of the genetic differences between strains, populations, or species are selectively advantageous in any given environment? In principle, CRISPEY should allow us to introduce individual variants, one at a time, onto a common genetic background, and enable comparisons even between reproductively incompatible lines. As a first test of this approach we investigated fitness differences between two well-studied *S. cerevisiae* strains: BY (a common lab strain) and RM (a vineyard isolate), in which thousands of QTLs have been mapped for a wide variety of traits (Albert et al., 2014a; Bloom et al., 2013; Ehrenreich et al., 2009), but few causal variants have been identified (Fay, 2013).

We designed 32,000 guide/donor pairs targeting 16,006 SNPs and short indels, with edits designed to introduce a single RM allele into each BY cell (Table S5). After transforming our plasmid library into BY and editing for 48 hr, the edited yeast pools competed in log-phase growth for ~25 generations in minimal media with 2% glucose. We sequenced guide/donor oligomers from six replicate flasks at each of 11 time points and estimated the abundance of each edited strain (Figures 2A and S1). Abundance measurements at similar time points were highly correlated between replicates (Pearson's $r = 0.952$ – 0.992 ; Figures S1F and S1J). We then estimated the fitness effect of each edit by measuring the change in abundance of guide/donor oligomers introducing each RM allele.

We observed a large variation in fitness across strains (i.e., cells containing a specific guide/donor oligomer), with some strains changing >30-fold in frequency during the competition (Figures 2B and S1C). Fitting a linear model to each strain's \log_2 abundance over time in generations, we estimated the relative fitness effect of each edit (approximated by the \log_2 fold change in abundance per generation; Figures S1A–S1E; STAR Methods for details) (Ritchie et al., 2015).

Fitness estimates from two independent editing experiments were only moderately correlated (Pearson's $r = 0.5$; Figures S1G and S1H), as expected if most genetic variants have no measurable fitness effect and thus no correlation between replicates. We therefore calculated the irreproducible discovery rate (IDR) (Li et al., 2011) to define a set of reproducible effects (Figures 2C and S1H). Positive fitness effects were more easily detectable, since the starting abundance of each strain leaves more dynamic range to increase than to decrease (especially for strains with relatively low starting abundance), and lower abundances are more affected by random drift and low

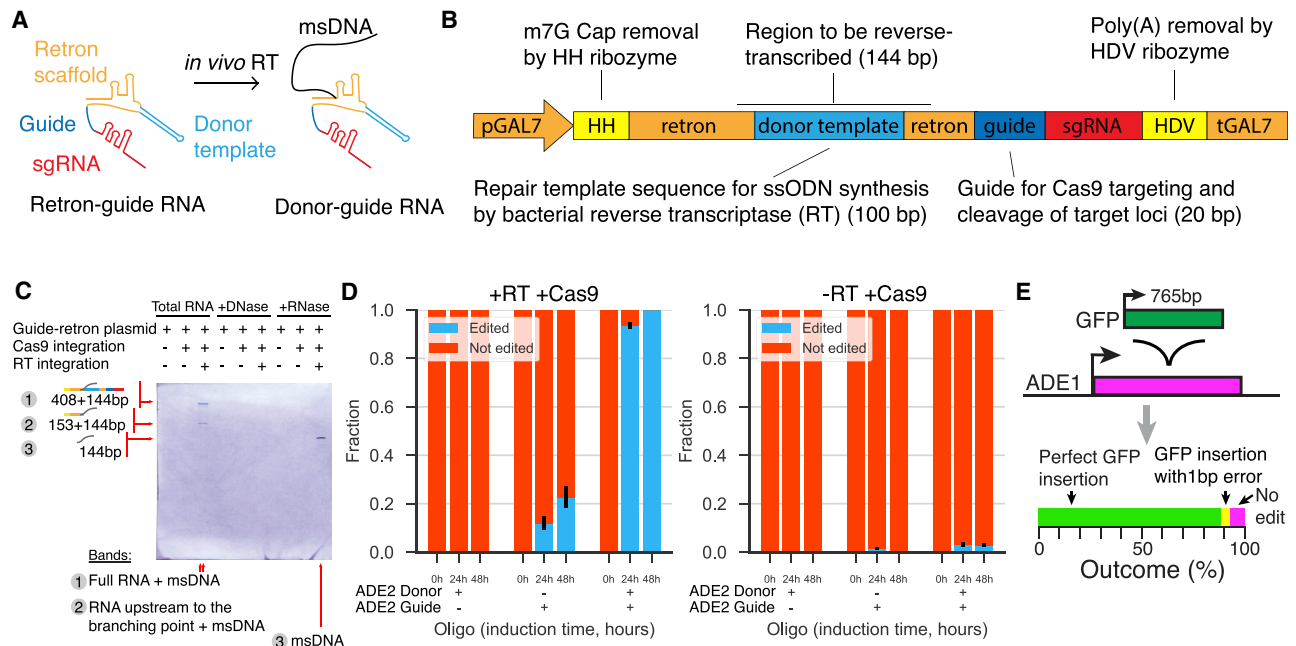


Figure 1. CRISPEY Is Highly Efficient and Precise

(A) Schematic for generation of msDNA (black) as single-stranded oligodeoxynucleotide repair donor *in vivo* through reverse transcription of a hybrid retron-guide RNA molecule. The components shown are: retron scaffold from Ec86 retron (orange), donor template to be reverse-transcribed (blue), guide sequence targeting genomic loci (dark blue), and sgRNA scaffold RNA for SpCas9 binding (red).

(B) CRISPEY construct for generating retron-guide RNA in yeast. Retron-guide RNA sequence is flanked by two self-cleaving ribozymes for removing mRNA cap and poly(A) tail after pGAL7 driven Polymerase II transcription.

(C) Southern blot analysis of msDNA species from galactose-induced yeast total RNA. Note size shift after RNase, indicating removal of retron RNA attached to msDNA. Band (2) indicates partial degradation of the retron-guide RNA.

(D) Quantification of retron-mediated CRISPR/Cas9 editing in *ADE2*. Yeast were subjected to editing for 48 hr via induced expression of Cas9, RT, and the CRISPEY construct targeting *ADE2*. The fractions (\pm STD) of edited (*ADE2* knockout phenotype expected from HDR repair) and non-edited (wild-type phenotype) are shown for combinations of *ADE2* editing retron donor and *ADE2*-targeting guide (marked by +*ADE2* Donor/+*ADE2* Guide); or “non-functional” *BFP* retron donor and GFP-targeting guide (marked by –*ADE2* Donor/–*ADE2* Guide).

(E) CRISPEY also allows insertion of a 765 bp sequence with low error rate. Top: schematic for long-insert editing of genomic DNA at the *ADE1* locus. Bottom: diagram indicating outcome of editing, with 87.4% showing the intended edit, 4.6% showing 1 bp error within the insertion (possibly due to RT-induced error), and 8% have no edit.

See also Tables S1, S2, S3, and S4.

sequencing depth. Nevertheless, both positive and negative effects were highly reproducible across independently edited biological replicates, with 572 variants below a 5% IDR (Li et al., 2011) (Figures 2C and S11).

Our library design included 8,902 variants with multiple gRNAs creating the same edit. As another test of reproducibility, we compared the fitness effects estimated for these independent gRNAs. As expected, edits with minor fitness effects showed little correlation (Pearson’s $r = 0.10$, p value $< 10^{-10}$ for edits with IDR > 0.05); however, significant edits were far more reproducible between guides (Pearson’s $r = 0.59$, p value $< 10^{-34}$ for edits with IDR < 0.05 ; Figure 2D).

As a validation, we further tested 14 edits with varying fitness effect estimates. We performed CRISPEY to create these 14 individual (non-pooled) strains, and sequence-verified each edit (at these loci we observed 96% editing efficiency and 0% NHEJ; Table S3). Competing each edited strain against its BY parent, we found that all 14 validated in both direction and approximate fitness effect, although the fitness effects estimated from pooled

competition tended to be slightly smaller than those from individual competitions (Figure 2E; Table S2). This orthogonal validation lends further support to the accuracy of our genome-wide CRISPEY results.

Characterizing Cas9 Target Mismatch Sensitivity

Next, we investigated the target specificity of Cas9 by examining CRISPEY toxicity during the editing phase. Our guides were filtered to not have highly similar secondary targets elsewhere in the genome (STAR Methods). However, both the edited target site and donor sequence on the plasmid represent possible off-target sites, differing from the intended on-target site by just one base in the case of a SNP edit. At these two potential off-target sites, the edit that is introduced by the donor will determine the mismatch between the guide and the off-target. We reasoned that since each Cas9 cut creates a fitness burden (note that plasmid loss is lethal in our growth media), guide/donor oligomers that cut their target and off-target (i.e., the donor) more efficiently may have lower fitness during the 48-hr editing phase.

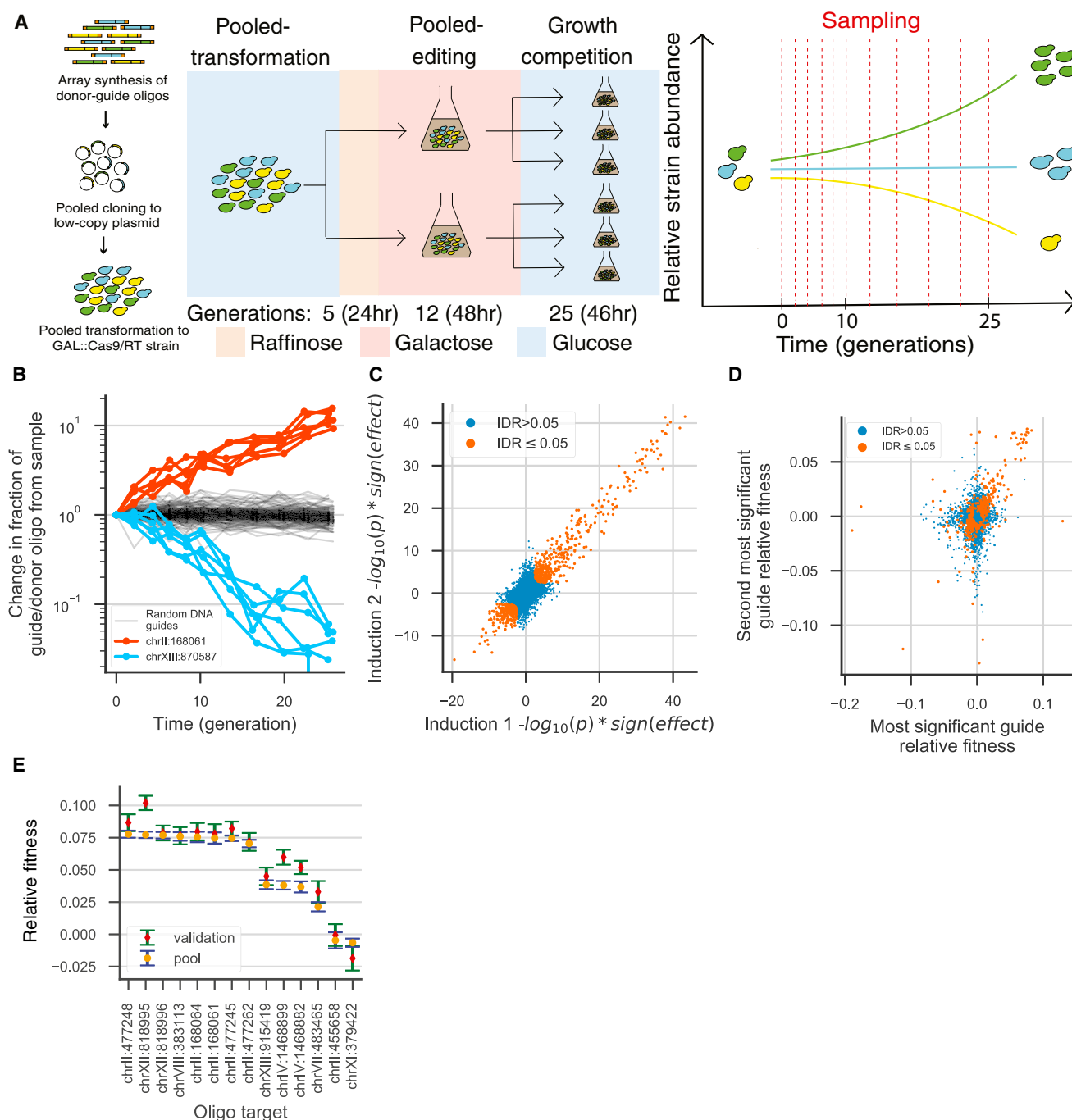


Figure 2. CRISPEY Screen for Fitness Effects of Natural Variants

(A) Schematic for experiment to assay phenotypic effects of CRISPEY libraries.

(B) Example of one edit increasing fitness (red), one decreasing fitness (blue), and random guides representing wild-type BY (black). Each guide/donor pair is shown with abundances from six biological replicate cultures.

(C) Irreproducible discovery rate analysis, based on agreement between independently edited biological replicates. The greater number of positive effects (RM allele fitter) is likely due to asymmetric power.

(D) Edits without a reproducible fitness effect (blue) show little reproducibility between independent guides, as expected if their fitness effect estimates are dominated by noise or random drift. In contrast, edits with reproducible effects (orange) are more highly correlated.

(E) We selected 14 edits for validation: 12 with positive fitness effects, 1 with no effect, and 1 with negative effect. Our estimates of \log_2 fold change per generation from pooled competition were slightly smaller than those from the validation experiments, suggesting that our fitness effects are not over-estimated due to the "winner's curse." Data are represented as mean \pm SEM.

See also Figure S1 and Tables S1, S2, S3, S4, and S5.

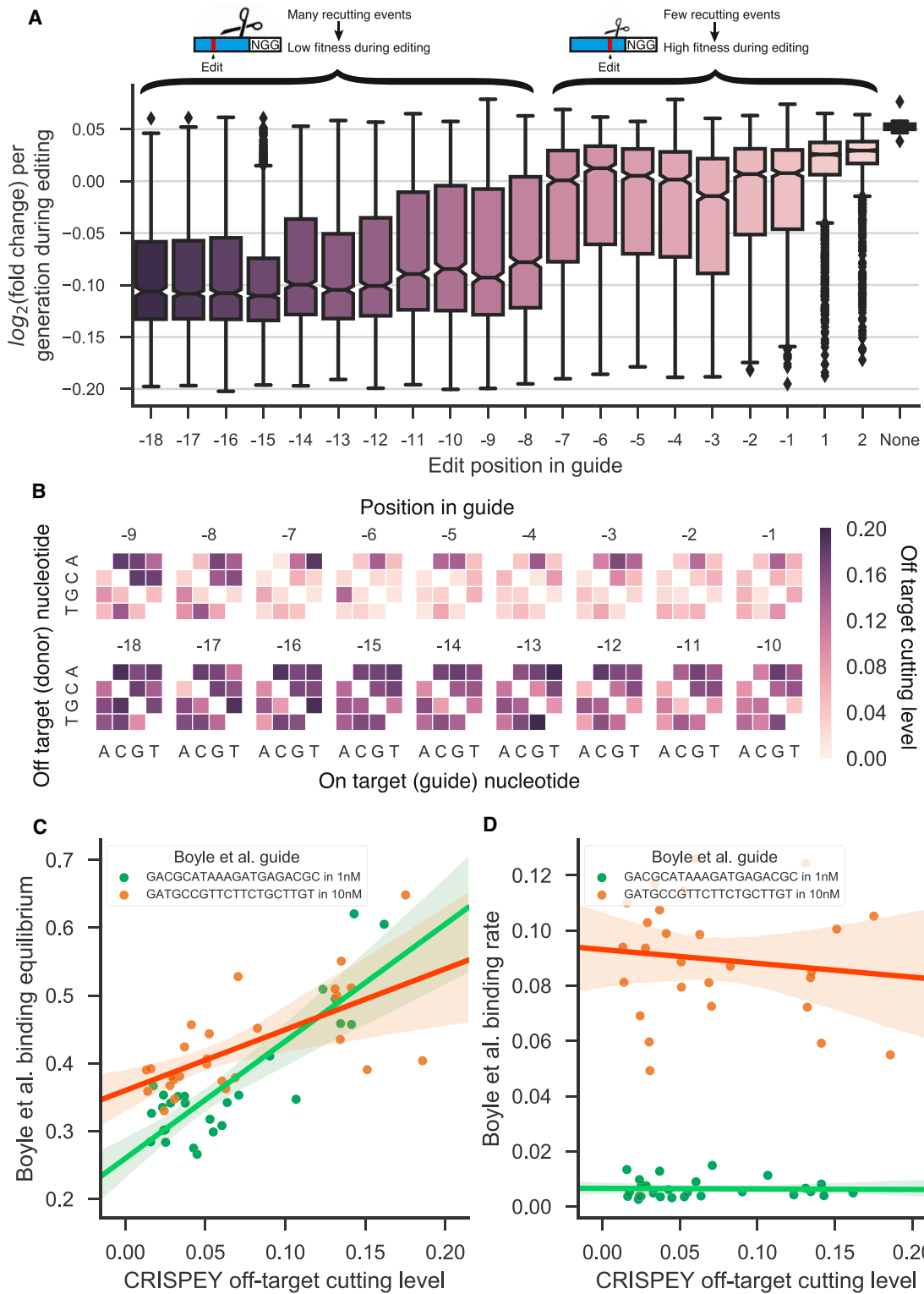


Figure 3. Cas9 Is Sensitive to Mismatches in the Seed Region and Depends on the Mismatched Nucleotides

(A) Each Cas9-induced double strand break (DSB) is potentially toxic, so edits that prevent repeated cutting—either of the genome or of the plasmid—are expected to show higher fitness during the 48-hr editing phase of our experiment. We observed the lowest toxicity for edits in the PAM or seed region (positions -1 to -7 in the guide), consistent with previous results on Cas9 mismatch-tolerance (Doench et al., 2016; Fu et al., 2016). For comparison, “None” shows the fitness of random guides that are not expected to cut anywhere. Whiskers show the distribution range (excluding outliers).

(legend continued on next page)

Therefore, examining the toxicity of CRISPEY during the editing phase as a function guide-donor mismatches can inform us about Cas9 target sensitivity. For this purpose, we defined the difference between the fitness (\log_2 fold change per generation) of cells with a specific guide/donor oligomer and the median fitness of strains with random guides (that do not promote cutting) as a proxy for the off-target cutting level.

We found that the fitness effect of each guide/donor during editing was highly reproducible (Pearson's $r = 0.97$ for guide/donor oligomer abundance \log fold change during editing), with guide/donors that edit the PAM or "seed" positions ($-7:-1$) in the guide having reduced toxicity, suggesting lower levels of off-target cutting (Figures 3A and S2A–S2D). Interestingly, the sensitivity of the guide to a mismatch depends on the base change induced by the edit, with lower sensitivity (and thus higher rates of cutting) for guanine to adenine mismatches in the seed region (Figures 3B and S2A–S2C). These results are consistent with trends previously observed for *in vitro* target specificity for four guides (Figure S3) (Pattanayak et al., 2013).

To investigate the patterns of mismatch tolerance, we compared our results to a study of dCas9 binding *in vitro*. We found a strong agreement between our off-target cutting level and equilibrium dCas9 binding to the same mismatch positions/nucleotides, but no association with dCas9 binding rates for these same mismatched guides (Figures 3C and 3D) (Boyle et al., 2017). This suggests that dissociation before cutting may be common at off-target sites *in vivo*, consistent with results observed *in vitro* and in human cell lines (Farasat and Salis, 2016; Gong et al., 2018; Ma et al., 2016a).

Overall, our results are in agreement with biochemical (Boyle et al., 2017; Fu et al., 2016) and *in vivo* (Fu et al., 2016; Morgens et al., 2017; Tsai et al., 2017) characterization of Cas9 target mismatch sensitivity (Figure S3) and extend them by testing over ten thousand on-target/off-target pairs. Note that Cas9 toxicity should not affect the growth competition because Cas9 and the gRNAs are not expressed in glucose media; consistent with this, the guide score and edit position in the guide were not correlated with fitness during growth competition (Figures S2E–S2G).

Variants Affecting Fitness Are Enriched in *cis*-Regulatory Regions

We then characterized the variants with significant fitness effect during growth competition (Figures 4A, S4A, and S4B). Many of these variants had surprisingly large effect sizes: 171 had $>1\%$ fitness difference per generation and 17 were $>5\%$. This suggests that many of these variants may involve fitness tradeoffs—where each allele confers a selective advantage in some environments—because unconditional fitness effects of this magnitude would quickly lead to fixation of the fitter allele (see Discussion).

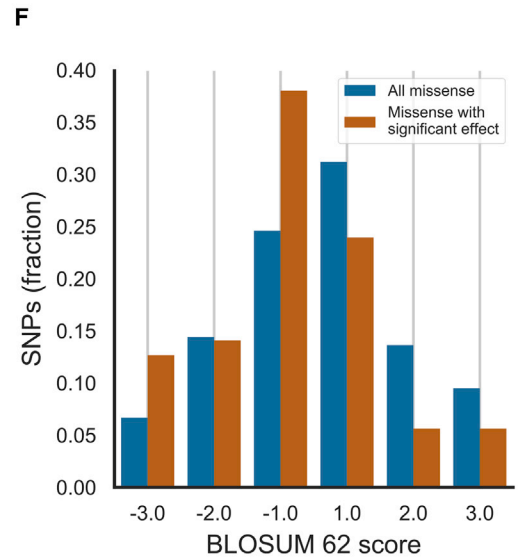
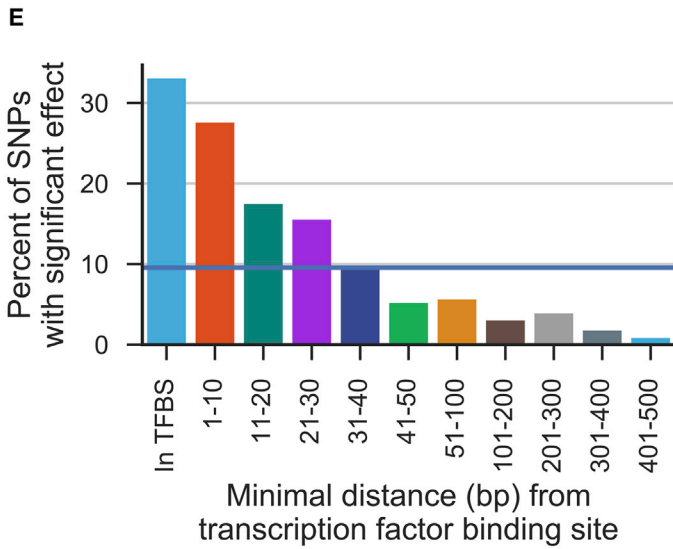
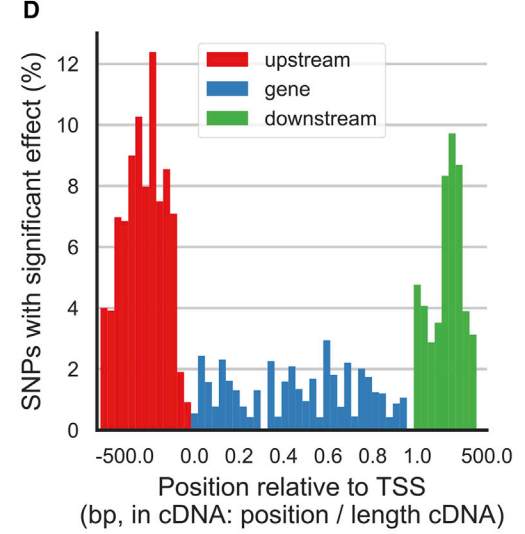
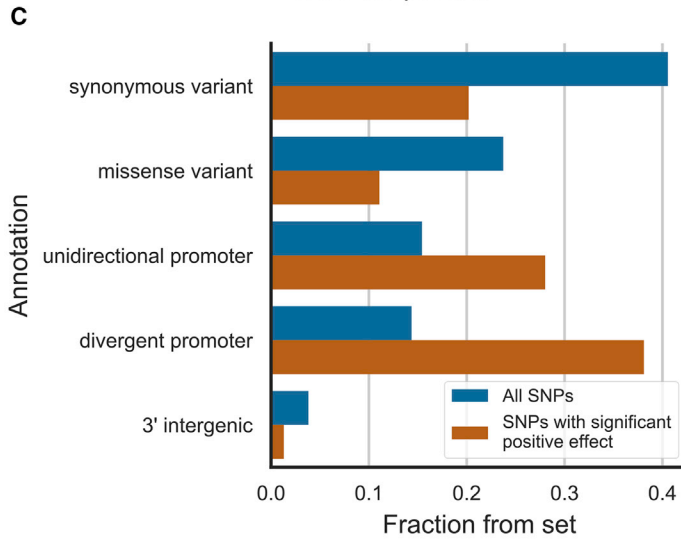
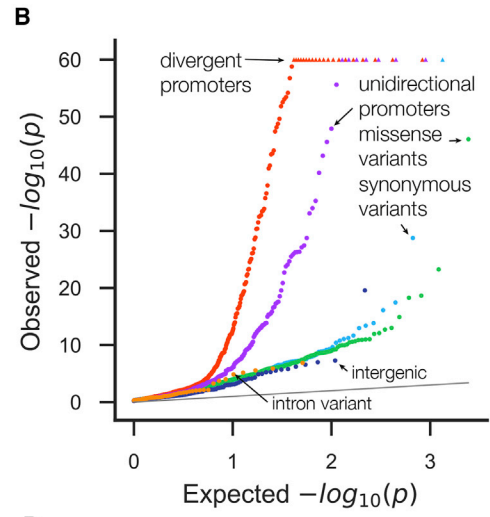
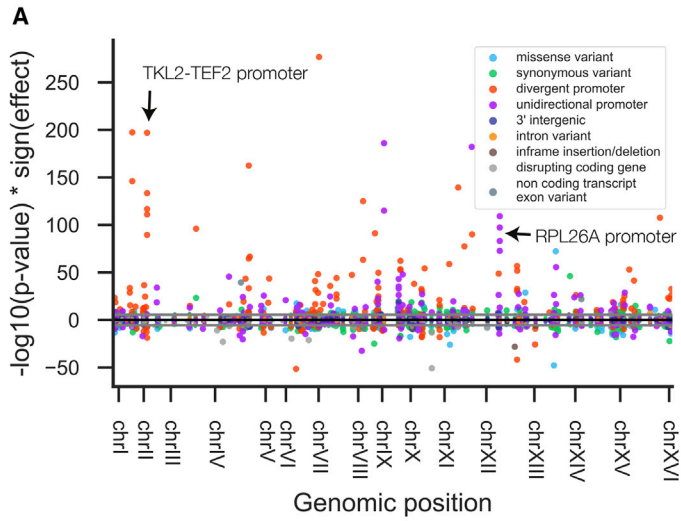
We next asked what types of genomic regions harbor these significant variants. Most previously identified variants affecting trait variation between species are within protein-coding regions (Fay, 2013; Hoekstra and Coyne, 2007; Stern and Orgogozo, 2008), although these may be influenced by ascertainment biases (Hoekstra and Coyne, 2007; Rockman, 2012). Strikingly, we found that our top 23 hits were all in promoter regions, despite promoters accounting for only 29.8% of the edits (hypergeometric p value $< 10^{-11}$). Divergent promoters—those upstream of two divergently transcribed genes—were the most enriched for significant positive effects, followed by unidirectional promoters (2.66- and 1.81-fold enrichment, binomial p value $< 10^{-15}$ and 10^{-7} respectively; Figures 4B, 4C, S4C, and S4F). This pattern persisted when controlling for chromatin accessibility, which could potentially affect editing efficiency (Figure S4D). When significant variants were visualized as a function of their position within genes or flanking noncoding regions (Figures 4D and S4E), the distribution was similar to the positions of gene expression QTLs (eQTLs) mapped in diverse *S. cerevisiae* strains (Kita et al., 2017), suggesting that *cis*-regulation may underlie the majority of these effects.

To investigate the promoter variants further, we tested whether they were enriched in transcription factor binding sites (TFBSs), which are key sequence elements in *cis*-regulation. We found a striking pattern: 33% of variants within known TFBS had significant fitness effects, compared to only 9.6% of all tested promoter variants (hypergeometric p value $< 10^{-12}$; Figure 4E). Interestingly, the fraction of significant hits showed a strong dependence on distance from the nearest TFBS, with variants within 10 bp still highly enriched—consistent with results from human studies (Farh et al., 2015; Tehrani et al., 2016)—while those >40 bp from a TFBS were depleted of fitness effects (Figures 4E, S4G, and S4H). The most highly enriched binding sites were for Reb1 (9 sites, 1.94-fold, hypergeometric p value = 0.011), Abf1 (8 sites, 2.02-fold, hypergeometric p value = 0.013), and Rap1 (8 sites, 7 of which are within ribosomal protein promoters, 2.42-fold, hypergeometric p value = 0.002).

To test whether these promoter variants impact *cis*-regulation, we examined the allele-specific expression of their neighboring genes in a BY/RM hybrid (Albert et al., 2014b). We found that genes with significant upstream variants were slightly more likely to show significant allelic imbalance in mRNA levels (Wilcoxon p value = 0.00068) and translation (Wilcoxon p value = 0.00069) in a BY/RM hybrid (Figures S5A–S5D). We also validated the effects of two variants on the mRNA levels of *RPL19B* (42% increase, unpaired t test with Welch's correction $p = 0.0082$) and *RPL26A* (21% increase, unpaired t test with Welch's correction $p = 0.0024$) using reverse-transcription qPCR (Figures S5E–S5G).

Altogether, these results strongly support a predominance of *cis*-regulatory variants affecting BY/RM fitness variation, in

(B) Off-target cutting level as a function of the edit position in the guide and the nucleotides in the guide and in the off-target sense strand. Off-target cutting level is defined as the difference between the fitness of cells with specific guide/donor and the median fitness of strains with random DNA guide/donor (i.e., guide/donor that do not promote any cutting; "None" in A). Note the lower sensitivity to guanine to adenine mismatches in the seed relative to other mismatches in this region. (C and D) Agreement of the off target cutting level shown in (B) with *in vitro* measurements of dCas9 binding equilibrium (Pearson's $r = 0.85$ and 0.65 , p value = 1.7×10^{-8} and 2×10^{-4} for the guides marked by red and green, respectively) (C) and binding rate (Pearson's $r = -0.027$ and -0.12 , p value = 0.89 and 0.52 for the guides marked by red and green, respectively) (D) from Boyle et al. (2017). See also Figures S2 and S3 and Table S5.



(legend on next page)

contrast with the predominance of coding variants impacting phenotypes that has been observed across single-locus studies in yeast, plants, and metazoans (Fay, 2013; Hoekstra and Coyne, 2007; Stern and Orgogozo, 2008). This difference may be due to the absence of ascertainment bias in our CRISPEY screen, consistent with trends observed in similarly systematic genome-wide association studies in humans (Hindorf et al., 2009).

Synonymous and Missense Variants Have Similar Impacts on Fitness

Although coding variants were under-represented among significant variants (Figure 4C), we identified 156 synonymous and 95 missense variants at IDR <0.05. Missense variants that change protein sequences were about equally likely to affect fitness as synonymous variants (Figures 4C and S4J). While this may seem counterintuitive, this does not suggest that random missense and synonymous mutations will have similar effects, because deleterious variants have already been filtered by natural selection. We hypothesized that missense variants may affect fitness via changes to proteins, while synonymous variants may affect translation via codon usage. Consistent with this, we found that significant missense variants were more likely to cause non-conservative amino acid changes (Figures 4F and S4K), and significant synonymous variants were more likely to be present in genes with strong codon usage bias (Figure S4L). It is also likely that some coding variants affect fitness via *cis*-regulation of transcription or mRNA decay, considering the abundance of eQTLs within yeast protein-coding regions (Kita et al., 2017). Our finding of widespread fitness effects of synonymous variants casts doubt on the common assumption that synonymous mutations are neutral, consistent with several other lines of evidence (Chamary et al., 2006; Lawrie et al., 2013; Plotkin and Kudla, 2011; She and Jarosz, 2018).

Polygenic Adaptation of Ribosomal Gene Promoters

We then asked whether we could detect lineage-specific selection from the patterns of fitness effects. Because BY and RM have evolved in different environments, it is likely that they have each adapted in distinct ways, and variants with large fitness effects (e.g., those above 5% in Figure 2C) represent local adaptations to each strain's typical environment. By searching over all gene ontologies (Eden et al., 2009), we found that "cytoplasmic translation" (mostly ribosomal) genes were strongly enriched for having significant RM-fitter variants in their promoters (false discovery rate [FDR] = 5.5×10^{-5}).

To test whether this enrichment was driven by selection, we applied a sign test (Fraser, 2011; Orr, 1998), which tests a null model that under neutral evolution, any gene set (e.g., pathways or protein complexes) should show a similar ratio of BY-fitter/RM-fitter alleles as the rest of the genome. In contrast, if a pathway has been under different selection pressures in the BY and RM lineages, then we may observe a bias of fitness effects in that pathway favoring one strain. We found that 25 "cytoplasmic translation" (mostly ribosomal) genes have significantly RM-fitter variants, compared to just 2 with BY-fitter variants (binomial p value < 5×10^{-9} ; Figures 5A and 5B), which is not consistent with neutral evolution and suggests polygenic adaptation of the ribosome.

To test whether these ribosomal promoter variants may affect the translation of other genes via their effects on the ribosome, we analyzed QTLs for protein levels (pQTLs) from a BY/RM cross (Albert et al., 2014a). We found that the 25 translation genes with RM-fitter variants were 2.3-fold enriched among pQTL loci whose RM alleles decrease protein levels of one or more other genes, without affecting mRNA levels (permutation p value = 0.012; STAR Methods). No enrichment was observed for pQTLs whose RM alleles increase protein levels (permutation p value = 0.81), suggesting that attenuated translation may underlie the fitness advantage of the RM alleles.

Among BY-fitter variants, promoters for genes involved in pseudohyphal growth were specifically enriched (enrichment FDR = 0.0057; binomial p value = 0.026), suggesting another target of lineage-specific selection.

Lineage-Specific Selection Leads to Clusters of Large-Effect Variants

With our screen's variant-level resolution, we can also test for selection without the need for gene set annotations. In a previous application of the sign test to BY/RM, we found that pairs of *cis*- and *trans*-acting eQTLs targeting the same gene were more likely to act in the same direction, providing evidence of selection (Fraser et al., 2010). Extending this logic to our CRISPEY results, we tested whether neighboring variants also tend to favor the same parental strain. Among intergenic variants with significant (IDR <0.05) effects, we found that variants within 50 bp of one another nearly always (98%, 88 pairs, binomial p value = 4.8×10^{-20}) favored the same parent (Figures 5C, 5D, and S4I). This bias decreased to near-background levels (65% agreement; 83 pairs, binomial p value = 0.0655) at >200 bp. This trend suggests that most clusters of variants affecting fitness resulted from lineage-specific selection, rather than neutral evolution.

Figure 4. Characterizing Variants Affecting Fitness

- (A) Distribution of significant fitness effect variants throughout the genome. "Disrupting coding gene category" includes out-of-frame indels and gain or loss of stop codon. Positive, RM-fitter; negative, BY-fitter.
- (B) Quantile-quantile plot showing the extent of significant RM-fitter hits for the major annotation categories. Values larger than 60 were set to 60 and marked by triangles (see Figure S4C for non-truncated version).
- (C) Variants with strong fitness effects (RM-fitter alleles, IDR <0.05) are enriched in promoters and depleted in coding regions.
- (D) Locations of significant variants (RM-fitter alleles, IDR <0.05) with respect to TSSs (all ORFs were scaled to the same length for visualization; 900 bp of flanking regions are also shown).
- (E) Variants in or near transcription factor binding sites are most likely to affect fitness. The blue line shows the result over all intergenic variants.
- (F) Missense variants with IDR <0.05 are enriched for smaller BLOSUM62 scores, indicating less conservative amino acid changes.

See also Figures S4 and S5.

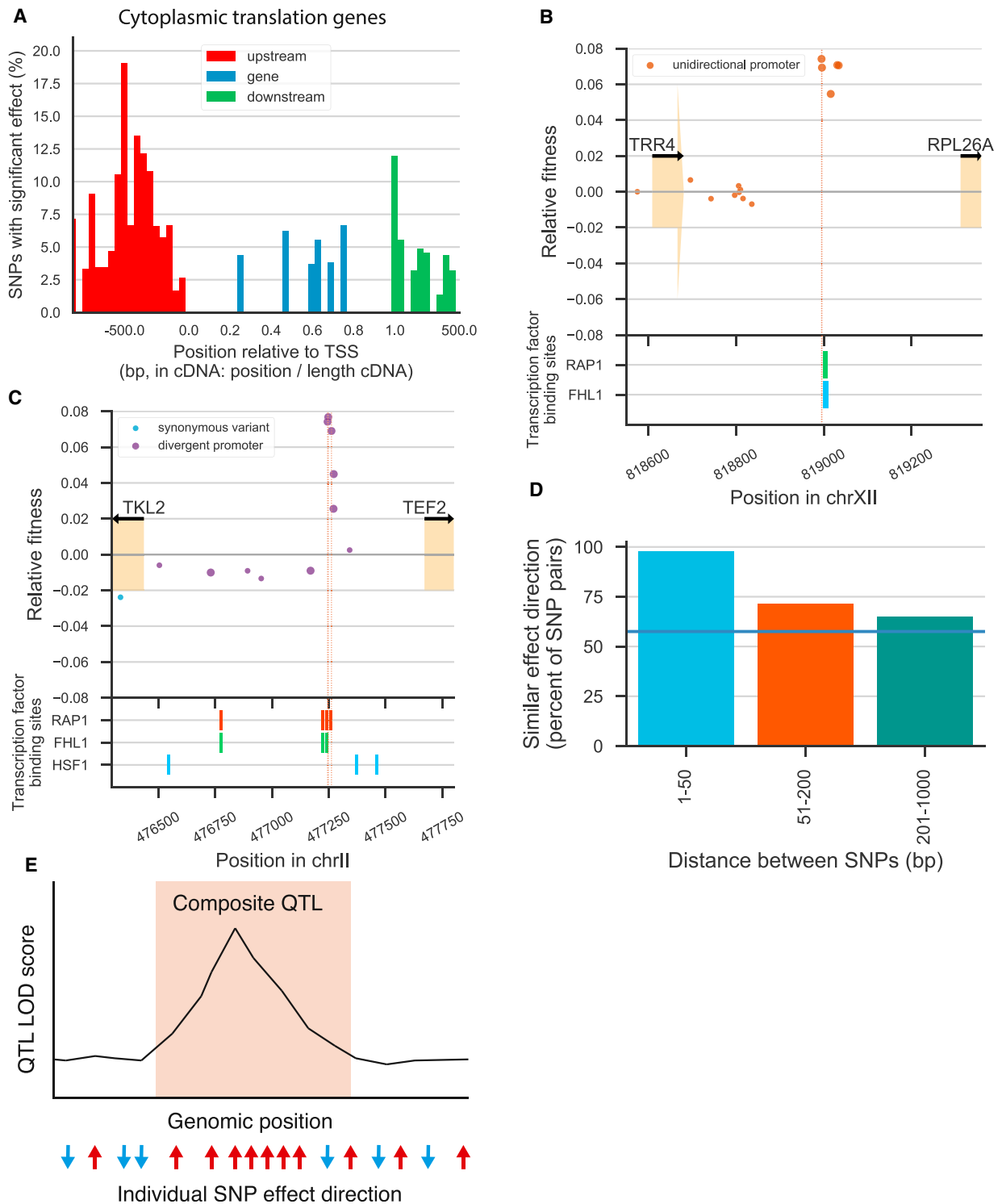


Figure 5. Detecting Lineage-Specific Selection

(A) Locations of significant variants (RM-fitter alleles, IDR <0.05) with respect to TSSs for 25 cytoplasmic translation genes. Most RM-fitter variants are in the flanking noncoding regions.

(B) Example of a ribosomal subunit with five significant variants clustered within its promoter; all five variants affect fitness in the same direction, and two (indicated by red dotted lines) are adjacent positions next to conserved Fhl1/Rap1 binding sites, both with validated fitness effects (Figure 2E). Larger dots indicate significant fitness effects.

(legend continued on next page)

This finding has important implications for the interpretation of QTLs: it has long been recognized that closely linked variants acting in the same direction could create strong QTLs (Stam and Laurie, 1996; Visscher and Haley, 1996) (Figure 5E), but the prevalence of these “composite QTLs” is unknown. If selection on closely linked variants is also common in other species, composite QTLs are likely to be quite frequent, and thus genetic architectures may be far more complex than can be inferred from QTLs alone.

DISCUSSION

In sum, our results show that genotype-phenotype relationships can now be assayed at single-base resolution in parallel, greatly boosting our ability to identify causal variants for polygenic traits. We observed that yeast genetically encoded fitness differences are mainly *cis*-regulatory and often involve multiple tightly clustered causal variants per locus affecting fitness in the same direction.

Our results raise several important questions. First, why are such large fitness effects observed among segregating variants? Theory predicts that selection should be very efficient in large populations such as *S. cerevisiae*, leading to rapid fixation or loss of alleles with fitness effects greater than $\sim 4/N_e$ (where N_e is the effective population size), which is likely to be several orders of magnitude smaller than the 1% effects we observed for 171 variants. One plausible explanation is that BY and RM are adapted to different environments, and the large fitness effects reflect these lineage-specific adaptations. This would predict that CRISPEY screens performed in environments more closely resembling RM’s historical environment would lead to greater fitness advantages for RM alleles (and vice versa for BY). Alternatively, other forms of selection could maintain large-effect variants even without local adaptation, such as balancing selection or seasonally fluctuating selection (Sellis et al., 2016; Wittmann et al., 2017).

Another question is to what extent the patterns we observed—such as the predominance of noncoding variants and clustered variants acting in the same direction—will be seen in other growth conditions, strains, and species. While there is no doubt that the quantitative fitness effects of many variants will change substantially depending on the environment and genetic background, we hypothesize that the more general patterns are unlikely to be environment- or strain-specific. This reasoning is supported by a similar predominance of noncoding variants in human phenotypic variation (Hindorf et al., 2009) and recent adaptation (Enard et al., 2014; Fraser, 2013). Nevertheless, it will be fascinating to compare results of CRISPEY screens across diverse environments to explore gene-by-environment interactions with single-variant resolution.

Our work has several limitations. First, as noted above, any pooled screen will have more power to detect strains with

increasing than with decreasing abundance. However, this effect could be minimized in the future by generating fewer edits, increasing the population size, sequencing more deeply, including more replicates, and/or performing lineage tracking with unique molecular identifier (UMI) barcodes (Michlits et al., 2017). These and other refinements in our approach should allow the detection of even smaller fitness effects, and more symmetric power to detect effects in both directions. A second limitation is that our estimated fitness effects could be underestimates if editing efficiencies at some loci are imperfect. Indeed, we observed $\sim 4\%$ unedited colonies when sequencing 14 loci (Figure 2E). However, these editing efficiencies are high enough that they should only introduce a small amount of error into fitness estimates (as seen in the validation set, Figure 2E). Moreover, we have not seen a correlation between the guide quality score or the edit position in the guide and the measured effect on fitness (Figures S2E–S3G). However, we cannot exclude the possibility of some false-negatives due to low guide efficiency. We also note that in experiments where the average fitness significantly increases over time, the analysis should account for this. Third, our method is limited by the requirement of a PAM sequence near the SNP. Approximately 70% of BY/RM SNPs can be targeted by a Cas9 protein that recognizes the “NGG” PAM sequence; performing CRISPEY with modified Cas9 proteins with different PAM sequences will increase this coverage further (Hirano et al., 2016; Hu et al., 2018). Finally, if CRISPEY editing has any unintended effects (e.g., epigenetic changes) that are heritable over many generations, this could also affect our results as well as other CRISPR-based screens.

Our approach has several advantages over other recent methods of high-throughput editing in yeast (summarized in Table S1). First, our method does not rely on high-copy plasmids as dsDNA donors, allowing adaptation to species that do not propagate plasmids. Second, ssDNA is known to be more efficient in HDR repair than dsDNA (Richardson et al., 2016), which may be critical for adapting our editing system to model systems with low HDR efficiency, such as mammalian cells. Third, our approach produces covalent tethering of the donor template to the guide RNA. This was shown to considerably increase the editing efficiency in both human and yeast (Aird et al., 2018; Lee et al., 2017; Roy et al., 2018). Fourth, retron-mediated editing is highly efficient and does not require additional genetic manipulation, such as *NEJ1* knockout or linearized vectors, which limits general applicability of editing systems. Last, although adeno-associated virus (AAV)-based genome editing approaches are very efficient in human, they are not suitable for pooled screens due to high multiplicity of infection (MOI) (Bak et al., 2017; Dever et al., 2016; Gaj et al., 2017; Nishiyama et al., 2017; Suzuki et al., 2016; Yang et al., 2016). We expect that transient expression of the retron msDNA using low MOI vectors may circumvent this limitation and enable highly parallel precise

(C) Example of a divergent promoter with five significant variants clustered around Fhl1/Rap1 binding sites (Harbison et al., 2004); all five variants affect fitness in the same direction, and three (indicated by red dotted lines) have validated fitness effects (Figure 2E).

(D) Variants at $IDR < 0.05$ favor the same parent 98% of the time when separated by up to 50 bp; the agreement decreases at larger distances, but is still greater than random (blue line) up to 1 kb.

(E) Illustration of selection within a locus leading to reinforcing variant effect directions and a resulting composite QTL.

See also Figures S4 and S5.

genome editing in human cells. To conclude, retron-mediated HDR is highly efficient and potentially adaptable to many species.

In addition to measuring fitness across diverse conditions or strain backgrounds, CRISPEY screens could easily be adapted to any trait that can lead to differential strain or allele abundance (including cell sorting based on expression of fluorescent reporter genes). Looking ahead, implementing CRISPEY in other species—including mammalian cells, in which retrons are functional (Mirochnitchenko et al., 1994)—could potentially allow rapid, base-pair level investigation of a wide range of traits and diseases.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [CONTACT FOR REAGENT AND RESOURCE SHARING](#)
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
- [METHOD DETAILS](#)
 - Construction of strains and editing vector
 - Media
 - Evaluating editing efficiency
 - Long insert integration
 - Southern blot analysis of msDNA
 - Variant Editing Library Design
 - Yeast Library Preparation
 - Pooled editing and growth competition
 - Library sequencing
 - Estimating edited variant effect on fitness
 - Genomic annotation, GO and TFBS analyses
 - Validation experiments
 - Estimating editing effect in validation strains
 - Measuring gene expression in validation strains
 - Testing enrichment of ribosomal pQTLs
 - Quantification and Statistical Analysis
- [DATA AND SOFTWARE AVAILABILITY](#)

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and five tables and can be found with this article online at <https://doi.org/10.1016/j.cell.2018.08.057>.

ACKNOWLEDGMENTS

We thank M.C. Bassik, D.A. Petrov, D.F. Jarosz, L. Qi, E.A. Boyle, A. Harpak, D.A. Knowles, N. Sinnott-Armstrong, J. Lee, Atish Agarwala, and members of the Fraser and Pritchard labs for helpful discussions. E.S. was partially funded by the European Molecular Biology Organization (LTF 646-2014). S.-A.A.C. was partially funded by the National Institutes of Health (5T32GM007276-42). This work was supported by the National Institutes of Health (2R01GM097171-05A1). We dedicate this work to the memory of Zachery R. Smith.

AUTHOR CONTRIBUTIONS

Conceptualization, H.B.F. and E.S.; Investigation, E.S., S.-A.A.C., N.M.K., and H.B.F.; Validation, S.-A.A.C. and N.M.K.; Formal Analysis, E.S. and S.-A.A.C. performed qPCR and BY-RM expression analysis, H.B.F. performed the

trans-pQTL analysis; Data Curation, E.S. and S.-A.A.C.; Methodology, Software, and Visualization, E.S.; Writing – Original Draft, E.S., S.-A.A.C., N.M.K., H.B.F., and J.K.P.; Writing – Review & Editing, E.S., S.-A.A.C., N.M.K., J.D.S., H.B.F., and J.K.P.; Supervision, H.B.F. and J.K.P.

DECLARATION OF INTERESTS

H.B.F. is a co-inventor of a patent application describing the CRISPEY approach.

Received: March 5, 2018

Revised: June 21, 2018

Accepted: August 22, 2018

Published: September 20, 2018

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Chadwick et al. (2017); Gaudelli et al. (2017); Hess et al. (2016, 2017), Kim et al. (2017a, 2017b, 2017c); Komor et al. (2016); Liang et al. (2017); Ma et al. (2016b); Nishida et al. (2016); Rees et al. (2017); Satomura et al. (2017); Zhang et al. (2017).

REFERENCES

- Aird, E.J., Lovendahl, K.N., St.Martin, A., Harris, R.S., and Gordon, W.R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* *1*, 54.
- Albert, F.W., Treusch, S., Shockley, A.H., Bloom, J.S., and Kruglyak, L. (2014a). Genetics of single-cell protein abundance variation in large yeast populations. *Nature* *506*, 494–497.
- Albert, F.W., Muzzey, D., Weissman, J.S., and Kruglyak, L. (2014b). Genetic influences on translation in yeast. *PLoS Genet.* *10*, e1004692.
- Bak, R.O., Dever, D.P., Reinisch, A., Cruz Hernandez, D., Majeti, R., and Porteus, M.H. (2017). Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *eLife* *6*, e27873.
- Bao, Z., HamediRad, M., Xue, P., Xiao, H., Tasan, I., Chao, R., Liang, J., and Zhao, H. (2018). Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nat. Biotechnol.* *36*, 505–508.
- Benatui, L., Perez, J.M., Belk, J., and Hsieh, C.-M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* *23*, 155–159.
- Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.-L.V.L.V., and Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. *Nature* *494*, 234–237.
- Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., and Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. USA* *114*, 5461–5466.
- Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* *14*, 115–132.
- Brem, R.B., and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. USA* *102*, 1572–1577.
- Chadwick, A.C., Wang, X., and Musunuru, K. (2017). In vivo base editing of PCSK9 (proprotein convertase subtilisin/kexin type 9) as a therapeutic alternative to genome editing. *Arterioscler. Thromb. Vasc. Biol.* *37*, 1741–1747.
- Chamary, J.V., Parmley, J.L., and Hurst, L.D. (2006). Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* *7*, 98–108.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* *339*, 819–823.

- Dever, D.P., Bak, R.O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C.E., Pavel-Dinu, M., Saxena, N., Wilkens, A.B., Mantri, S., et al. (2016). CRISPR/Cas9 β -globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384–389.
- DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191.
- Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096.
- Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191.
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48.
- Ehrenreich, I.M., Gerke, J.P., and Kruglyak, L. (2009). Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the BYxRM cross. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 145–153.
- Enard, D., Messer, P.W., and Petrov, D.A. (2014). Genome-wide signals of positive selection in human evolution. *Genome Res.* **24**, 885–895.
- Farasat, I., and Salis, H.M. (2016). A biophysical model of CRISPR/Cas9 activity for rational design of genome editing and gene regulation. *PLoS Comput. Biol.* **12**, e1004724.
- Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343.
- Fay, J.C. (2013). The molecular basis of phenotypic variation in yeast. *Curr. Opin. Genet. Dev.* **23**, 672–677.
- Fraser, H.B. (2011). Genome-wide approaches to the study of adaptive gene expression evolution: systematic studies of evolutionary adaptations involving gene expression will allow many fundamental questions in evolutionary biology to be addressed. *BioEssays* **33**, 469–477.
- Fraser, H.B. (2013). Gene expression drives local adaptation in humans. *Genome Res.* **23**, 1089–1096.
- Fraser, H.B., Moses, A.M., and Schadt, E.E. (2010). Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc. Natl. Acad. Sci. USA* **107**, 2977–2982.
- Fu, B.X.H., St Onge, R.P., Fire, A.Z., and Smith, J.D. (2016). Distinct patterns of Cas9 mismatch tolerance in vitro and in vivo. *Nucleic Acids Res.* **44**, 5365–5377.
- Gaj, T., Staahl, B.T., Rodrigues, G.M.C., Limsirichai, P., Ekman, F.K., Doudna, J.A., and Schaffer, D.V. (2017). Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery. *Nucleic Acids Res.* **45**, e98.
- Garst, A.D., Bassalo, M.C., Pines, G., Lynch, S.A., Halweg-Edwards, A.L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., Alexander, W.G., and Gill, R.T. (2017). Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55.
- Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471.
- Gong, S., Yu, H.H., Johnson, K.A., and Taylor, D.W. (2018). DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep.* **22**, 359–371.
- Griffith, O.L., Montgomery, S.B., Bernier, B., Chu, B., Kasaian, K., Aerts, S., Mahony, S., Sleumer, M.C., Bilenky, M., Haeussler, M., et al.; Open Regulatory Annotation Consortium (2008). ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res.* **36**, D107–D113.
- Guo, X., Chavez, A., Tung, A., Chan, Y., Kaas, C., Yin, Y., Cecchi, R., Garnier, S.L., Kelsic, E.D., Schubert, M., et al. (2018). High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR-Cas9 in yeast. *Nat. Biotechnol.* **36**, 540–546.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104.
- Hess, G.T., Frésard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042.
- Hess, G.T., Tycko, J., Yao, D., and Bassik, M.C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* **68**, 26–43.
- Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367.
- Hirano, S., Nishimasu, H., Ishitani, R., and Nureki, O. (2016). Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. *Mol. Cell* **61**, 886–894.
- Hoekstra, H.E., and Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016.
- Hsu, M.Y., Inouye, M., and Inouye, S. (1990). Retron for the 67-base multicopy single-stranded DNA from *Escherichia coli*: a potential transposable element encoding both reverse transcriptase and Dam methylase functions. *Proc. Natl. Acad. Sci. USA* **87**, 9454–9458.
- Hsu, M.Y., Eagle, S.G., Inouye, M., and Inouye, S. (1992). Cell-free synthesis of the branched RNA-linked msDNA from retron-Ec67 of *Escherichia coli*. *J. Biol. Chem.* **267**, 13823–13829.
- Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278.
- Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., and Liu, D.R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63.
- Jinek, M., Chylinski, K., Fontana, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.
- Kim, D., Lim, K., Kim, S.T., Yoon, S.H., Kim, K., Ryu, S.M., and Kim, J.S. (2017a). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480.
- Kim, K., Ryu, S.M., Kim, S.T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S., and Kim, J.S. (2017b). Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437.
- Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017c). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376.
- Kita, R., Venkataram, S., Zhou, Y., and Fraser, H.B. (2017). High-resolution mapping of *cis*-regulatory variation in budding yeast. *Proc. Natl. Acad. Sci. USA* **114**, E10736–E10744.
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.

- Lawrie, D.S., Messer, P.W., Hershberg, R., and Petrov, D.A. (2013). Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet.* *9*, e1003527.
- Lee, K., Mackley, V.A., Rao, A., Chong, A.T., Dewitt, M.A., Corn, J.E., and Murthy, N. (2017). Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *eLife* *6*, e25312.
- Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* *5*, 1752–1779.
- Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* *15*, 554.
- Liang, P., Sun, H., Sun, Y., Zhang, X., Xie, X., Zhang, J., Zhang, Z., Chen, Y., Ding, C., Xiong, Y., et al. (2017). Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein Cell* *8*, 601–611.
- Liu, Z., Liang, Y., Ang, E.L., and Zhao, H. (2017). A new era of genome integration—simply cut and paste!. *ACS Synth. Biol.* *6*, 601–609.
- Löoke, M., Kristjahan, K., and Kristjahan, A. (2011). Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques* *50*, 325–328.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Ma, H., Tu, L.-C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson, T. (2016a). CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J. Cell Biol.* *214*, 529–537.
- Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016b). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* *13*, 1029–1035.
- Mackay, T.F.C., Stone, E.A., and Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* *10*, 565–577.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* *339*, 823–826.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl variant effect predictor. *Genome Biol.* *17*, 122.
- Michlits, G., Hubmann, M., Wu, S.-H., Vainorius, G., Budusan, E., Zhuk, S., Burkard, T.R., Novatchkova, M., Aichinger, M., Lu, Y., et al. (2017). CRISPR-UMI: single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat. Methods* *14*, 1191–1197.
- Mirochnitchenko, O., Inouye, S., and Inouye, M. (1994). Production of single-stranded DNA in mammalian cells by means of a bacterial retron. *J. Biol. Chem.* *269*, 2380–2383.
- Miyata, S., Ohshima, A., Inouye, S., and Inouye, M. (1992). In vivo production of a stable single-stranded cDNA in *Saccharomyces cerevisiae* by means of a bacterial retron. *Proc. Natl. Acad. Sci. USA* *89*, 5735–5739.
- Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess, G.T., Han, K., Jeng, E.E., et al. (2017). Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* *8*, 15178.
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* *353*, aaf8729.
- Nishiyama, J., Mikuni, T., and Yasuda, R. (2017). Virus-mediated genome editing via homology-directed repair in mitotic and postmitotic cells in mammalian brain. *Neuron* *96*, 755–768.
- Orr, H.A. (1998). Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* *149*, 2099–2104.
- Pachkov, M., Balwierz, P.J., Arnold, P., Ozonov, E., and van Nimwegen, E. (2013). SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Res.* *41*, D214–D220.
- Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* *31*, 839–843.
- Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* *464*, 768–772.
- Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* *12*, 32–42.
- Pritchard, J.K., Pickrell, J.K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* *20*, R208–R215.
- Rees, H.A., Komor, A.C., Yeh, W.-H., Caetano-Lopes, J., Warman, M., Edge, A.S.B., and Liu, D.R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* *8*, 15790.
- Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* *34*, 339–344.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* *43*, e47.
- Rockman, M.V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* *66*, 1–17.
- Rockman, M.V., and Kruglyak, L. (2006). Genetics of global gene expression. *Nat. Rev. Genet.* *7*, 862–872.
- Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell Biol.* *14*, 8096–8106.
- Roy, K.R., Smith, J.D., Vonesch, S.C., Lin, G., Tu, C.S., Lederer, A.R., Chu, A., Suresh, S., Nguyen, M., Horecka, J., et al. (2018). Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat. Biotechnol.* *36*, 512–520.
- Sadhu, M.J., Bloom, J.S., Day, L., Siegel, J.J., Kosuri, S., and Kruglyak, L. (2018). Highly parallel genome variant engineering with CRISPR-Cas9. *Nat. Genet.* *50*, 510–514.
- Satomura, A., Nishioka, R., Mori, H., Sato, K., Kuroda, K., and Ueda, M. (2017). Precise genome-wide base editing by the CRISPR Nickase system in yeast. *Sci. Rep.* *7*, 2095.
- Schaid, D.J., Chen, W., and Larson, N.B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* *19*, 491–504.
- Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* *25*, 1757–1770.
- Sellis, D., Kvitek, D.J., Dunn, B., Sherlock, G., and Petrov, D.A. (2016). Heterozygote advantage is a common outcome of adaptation in *Saccharomyces cerevisiae*. *Genetics* *203*, 1401–1413.
- She, R., and Jarosz, D.F. (2018). Mapping causal variants with single-nucleotide resolution reveals biochemical drivers of phenotypic change. *Cell* *172*, 478–490.
- Shi, Z., Fujii, K., Kovary, K.M., Genuth, N.R., Röst, H.L., Teruel, M.N., and Barna, M. (2017). Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* *67*, 71–83.
- Shimamoto, T., Hsu, M.Y., Inouye, S., and Inouye, M. (1993). Reverse transcriptases from bacterial retransposons require specific secondary structures at the 5'-end of the template for the cDNA priming reaction. *J. Biol. Chem.* *268*, 2684–2692.
- Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* *122*, 19–27.

- Stam, L.F., and Laurie, C.C. (1996). Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* *144*, 1559–1564.
- Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* *62*, 2155–2177.
- Sturtevant, A.H. (1913). The linear arrangement of six sex linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* *14*, 43–59.
- Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z., et al. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* *540*, 144–149.
- Tehranchi, A.K., Myrthil, M., Martin, T., Hie, B.L., Golan, D., and Fraser, H.B. (2016). Pooled ChIP-Seq Links Variation in Transcription Factor Binding to Complex Disease Risk. *Cell* *165*, 730–741.
- Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* *14*, 607–614.
- Visscher, P.M., and Haley, C.S. (1996). Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. *Theor. Appl. Genet.* *93*, 691–702.
- Wittmann, M.J., Bergland, A.O., Feldman, M.W., Schmidt, P.S., and Petrov, D.A. (2017). Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proc. Natl. Acad. Sci. USA* *114*, E9932–E9941.
- Yang, Y., Wang, L., Bell, P., McMenamin, D., He, Z., White, J., Yu, H., Xu, C., Morizono, H., Musunuru, K., et al. (2016). A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* *34*, 334–338.
- Zhang, Y., Qin, W., Lu, X., Xu, J., Huang, H., Bai, H., Li, S., and Lin, S. (2017). Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat. Commun.* *8*, 118.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Yeast Nitrogen Base without Amino Acids and Ammonium Sulfate	Fisher Scientific	50-489-163
Drop-out Mix Synthetic Minus Histidine, Uracil, Adenine Rich w/o Yeast Nitrogen Base	US Biological	D9537-07
Drop-out Mix Synthetic Minus Uracil w/o Yeast Nitrogen Base	US Biological	D9535-100 g
Carbenicillin	Fisher Scientific	NC0753434
BD Bacto Peptone	Fisher Scientific	DF0118-17-0
BD Bacto Tryptone	Fisher Scientific	DF0123-17-3
Yeast Extract	Fisher Scientific	BP1422-500
D-(+)- Glucose	Sigma-Aldrich	G7021-5KG
D-(+)-Raffinose	Sigma-Aldrich	R-0250
D-(+)-Galactose, Anhydrous	Fisher Scientific	G1-100
PEG-3350	VWR	PO125-500GM
Lithium Acetate dihydrate, Reagent Grade	Fisher Scientific	1341730
Salmon sperm ssDNA	Invitrogen	15632-011
D-Sorbitol	Fisher Scientific	ICN10293801
Glycerol	Fisher Scientific	BP229-1
DTT	Thermo Fisher Scientific	R0861
Ammonium Sulfate	Fisher Scientific	507516983
Critical Commercial Assays		
NotI-HF restriction enzyme	New England Biolabs	R3189L
Alkaline Phosphatase, Calf Intestinal (CIP)	New England Biolabs	M0290L
NEBuilder HiFi DNA Assembly Cloning Kit	New England Biolabs	E5520S
Endura Electrocompetent cells	Lucigen	60242-2
E.Z.N.A. Plasmid Maxi Kit	Omega Biotek	D6922-02
Gene Pulser Cuvette 0.2cm electrode gap	Bio-Rad	165-2082
Gene Pulser Cuvette 0.1cm electrode gap	Bio-Rad	165-2089
Zymolyase 20T	US Biological	NC0236652
Plasmid Plus Maxi Kit	QIAGEN	12963
Q5 Hot Start High-Fidelity DNA Polymerase	New England Biolabs	M0493L
SuperScript III First-Strand Synthesis System	Invitrogen	18080051
iQ SYBR Green Supermix	Bio-Rad	1708880
Deposited Data		
Supplementary Sanger sequencing data files	https://doi.org/10.17632/jtbnrmstt.1	N/A
Pachkov et al. (2013) transcription factor binding sites	SwissRegulon	SCR_005333
Harbison et al. (2004) transcription factor binding sites	UCSC genome browser	SCR_005780
ORegAnno (Griffith et al., 2008) transcription factor binding sites	UCSC genome browser	SCR_005780
"scerevisiae_gene_ensembl" dataset	Ensembl	SCR_002344
Experimental Models: Organisms/Strains		
<i>S. cerevisiae</i> : BY4742	Brachmann et al., 1998	N/A
ZRS111 (<i>S. cerevisiae</i> BY4741 w/ pZS157 integration)	This paper	N/A
ZRS112 (<i>S. cerevisiae</i> BY4741 <i>ade2::NatMX</i> w/ pZS157 integration)	This paper	N/A
<i>S. cerevisiae</i> library: ZRS111 w/ guide/donor oligonucleotide library for variant editing cloned into pZS165	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
qPCR, genotyping, library construction and guide/donor oligonucleotides (See Table S4)	This paper	N/A
Guide/donor oligonucleotide library for variant editing (See Table S5)	This paper (Twist Bioscience)	N/A
Recombinant DNA		
p414-TEF1p-Cas9-CYC1t plasmid	DiCarlo et al., 2013	Addgene 43802
pZS157 Ec86-RT/SpCas9 expression plasmid	This paper	Addgene 114454
pZS165 CRISPEY library editing vector (Hammerhead-HDV)	This paper	Addgene 114451
pZS160 CRISPEY editing vector (HDV-HDV)	This paper	Addgene 114453
Software and Algorithms		
STAR	Dobin et al., 2013	SCR_015899
limma	Ritchie et al., 2015	SCR_010943
Variant Effect Predictor (VEP)	McLaren et al., 2016	SCR_007931
Gorilla	Eden et al., 2009	SCR_006848
biomaRt	Durinck et al., 2009	SCR_002987
IDR (Li et al., 2011)	https://cran.r-project.org/web/packages/idr/	N/A
Bowtie2	Langmead and Salzberg, 2012	SCR_005476
Azimuth	Doench et al., 2016	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact Hunter B. Fraser (hbfraser@stanford.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All strains used in this study were derivatives of *S. cerevisiae* BY4742 ([Brachmann et al., 1998](#)). Construction of strains with integrated SpCas9 and Ec86 Reverse Transcriptase is described below.

METHOD DETAILS**Construction of strains and editing vector**

CRISPEY yeast strain ZRS111 for pooled screen and validation was created by BY4742 integration of plasmid pZS157, a derivative of pRS406 (Stratagene), containing *S. cerevisiae* HIS3 gene, pGAL1::SpCas9::tCYC and pGAL10::Ec86RT::tGAL10. Ec86 SpCas9 sequence was subcloned from p414-TEF1p-Cas9-CYC1t (Addgene #43802). Reverse Transcriptase (RT) DNA fragment was yeast codon optimized and synthesized by IDT as Gblock DNA oligo. pZS157 was linearized by KpnI digestion and homologous recombination to the his3 locus. CRISPEY editing vector plasmid pZS165 was derived from (yeast centromere plasmid YCp) pRS416 ([Sikorski and Hieter, 1989](#)), containing pGAL7::Hammerhead ribozyme::Ec86retron::NotI::sgRNA::HDV ribozyme::tGAL7. Retron sequence from Ec86 was synthesized by IDT as a Gblock DNA oligo.

Media

Expression measurements and competition experiments were performed in glucose (SD -HIS/-URA: 6.7 g/L yeast nitrogen base (YNB), 1.92 g/L Drop-out Mix Synthetic Minus Histidine, Uracil, Adenine Rich w/o Yeast Nitrogen Base, 2% glucose). Editing was performed in raffinose (YNB -HIS/-URA 2% raffinose: 6.7 g/L YNB, 1.92 g/L Drop-out Mix Synthetic Minus Histidine, Uracil, Adenine Rich w/o Yeast Nitrogen Base, 2% raffinose) and galactose (YNB -HIS/-URA 2% galactose: 6.7 g/L YNB, 1.92 g/L Drop-out Mix Synthetic Minus Histidine, Uracil, Adenine Rich w/o Yeast Nitrogen Base, 2% galactose) media. *ADE2* editing efficiencies were assayed on glucose low adenine (SD -URA low-ADE: 6.7 g/L YNB, 1.92 g/L Drop-out Mix Synthetic Minus Uracil w/o Yeast Nitrogen Base, 2% glucose, 2% agar) plates.

Evaluating editing efficiency

Parental strain BY4742 was derived into two strains through plasmid integration: (1) ZRS81 contained pGAL1::SpCas9::tCYC1, and (2) ZRS82 contained both pGAL1::SpCas9::tCYC1 and pGAL10::Ec86RT::tGAL10. The parental strain BY4742, ZRS81, and ZRS82 were then transformed with pZS160 (a plasmid derived from pZS165 in which Hammerhead ribozyme was replaced with HDV ribozyme) with 3 different guide/donor inserts: *ade2hdr* (retron donor introducing non-sense mutation to *ADE2* gene, facilitating *ADE2* knockout)-sgAde2 (guide sequencing targeting *ADE2*), *gfp2bfp* (retron donor introducing point mutation to convert EGFP to BFP, irrelevant to *ADE2*)-sgAde2, or *ade2hdr*-sgGFP (guide sequence targeting EGFP) (Table S4). The oligo inserts were assembled with NotI-digested pZS160 by Gibson assembly. Nine editing experiments were conducted, with all combinations between three parental strains and three plasmids. Single colonies were inoculated into pre-editing media (YNB -HIS/-URA 2% raffinose) for 24 hours, then sub-cultured twice into editing media (YNB -HIS/-URA 2% galactose) for 24 hours, for a total of 48 hours of editing under galactose induction. Each culture was diluted and plated on YNB -URA low-ADE plates in triplicates. After 3 days of incubation at 30C, pink (*ADE2* knockout phenotype) and white (parental phenotype) colonies were scored to estimate editing efficiency of *ADE2* knockout.

To determine the mode of editing (HDR or NHEJ) and evaluate the rate of local off-target effects, we sequenced the *ADE2* locus from edited, pink colonies from a *ADE2* knockout editing assay using ZRS111 strain containing pZS165 with *ade2hdr*-sgAde2 guide/donor insert. Strains were grown in raffinose, then edited in galactose. After 48 hours of editing, cultures were plated on SD -URA low-ADE plates and incubated for 3 days at 30C. Genomic DNA from 96 pink colonies were prepared and the *ADE2* locus was amplified with PCR (Table S4). All reactions were Sanger sequenced and one failed reaction was excluded (Table S3).

Long insert integration

To test the efficiency and accuracy of CRISPEY-mediated editing to introduce large insert sequences, a guide/donor plasmid (pZS163) was generated to insert EGFP sequence into the *ADE1* locus. pZS163 was derived from pZS160 with a 902 nt donor containing *ADE1* homology arms flanking EGFP coding sequence plus linker (765 nt) (Table S4). Similar to *ADE2* knockout, cells with *ADE1* knockout exhibit pink color when grown on low adenine media, indicative of editing. ZRS111 transformants carrying pZS163 were grown in YNB -HIS/-URA 2% raffinose for 18-24 hours and YNB -HIS/-URA 2% galactose for 48 hours with 1:30 dilution every 24 hr. After 48 hours of editing, cultures were diluted, plated, and incubated at 30C until colonies were visible. Genomic DNA from 40 pink colonies were prepared and the *ADE1* locus was PCR amplified for Sanger sequencing (Tables S3 and S4).

Southern blot analysis of msDNA

Guide/donor oligonucleotide containing *gfp2bfp*-sgGFP oligo was cloned into pZS160 (Table S4). The plasmid was transformed into 3 yeast strains: BY4742, ZRS81 and ZRS82 as described above. Yeast transformants were grown in YNB -HIS/-URA 2% raffinose for 18-24 hours and YNB -HIS/-URA 2% galactose for 48 hours with 1:30 dilution every 24 hr. Cell pellets were homogenized with glass beads in Trizol (Invitrogen) and total RNA was extracted following the manufacturer's instructions. Total RNA were further treated with RQ1 DNase (Promega) or RNase A (Invitrogen) before ethanol precipitation. Untreated and nuclease-treated total RNA were loaded to Novex 10% Urea-TBE gels (Thermo Fisher Scientific) for size separation and bands were visualized by staining with SYBR GOLD before transferring to Hybond N+ nylon membranes (GE healthcare) for blotting and UV-crosslinking. For Southern blot analysis, 5'- Digoxigenin labeled DNA probe (Integrated DNA Technologies) targeting msDNA (homology to BFP) was hybridized to the membrane at 55C in DIG Easy Hyb buffer (Roche) (Table S4). Membranes were then washed with 2x SSC, 0.5% SDS twice at 55C, hybridized to anti-Digoxigenin-AP antibody (Roche) and washed 3 times in 1x PBS-T. Antibody localization was visualized by addition of 1-Step NBT/BCIP substrate (Thermo Fisher Scientific).

Variant Editing Library Design

Our library contains 31,870 20 bp guides targeting 16,006 genetic differences (variants, which we refer to as SNPs, although they also contain small insertions or deletions) between BY4742 and RM11-1a strains of *Saccharomyces cerevisiae*. To enrich for fitness differences, we focused on 121 fitness QTLs previously mapped between these strains in eight conditions (Bloom et al., 2013) (YNB, YPD, paraquat, ethanol, neomycin, diamide, cobalt-chloride, and tunicamycin). For each variant, the library contained all possible guides that do not have exact match off-targets. Off-targets were identified using Bowtie2 (Langmead and Salzberg, 2012) with the following parameters: “-f -D 20 -R 3 -N 1 -L 10 -I S,1,0.50 -gbar 3 -end-to-end -k 30 -no-head -t -rdg 10,6 -rfg 10,6.” Guides that had an off-target with 0-3 mismatches were excluded from all analysis of the growth competition. Guide quality scores were calculated using Azimuth (Doench et al., 2016). For each guide, we designed one 100 bp donor that did not contain a 10 bp homopolymer (this limitation is imposed by the oligo synthesis technology). In most cases the donor is centered around the Cas9 cut site, however to avoid 10 bp homopolymers some donor sequences were shifted up/downstream. In these cases, at least 30 bp on each side of the Cas9 cut site were required, otherwise the variant was not targeted. We refer to a combination of a guide and a donor as “oligo” (see oligo design in Figure 1B). The sequences were designed using UCSC's sacCer3 genome (Kent et al., 2002). We also design 50 pairs of oligos targeting essential genes (genes annotated as “inviable” in SGD, 5/12/2016). In each pair, one oligo knocks out the gene by introducing a nonsense mutation to a cysteine or a tyrosine codon, while the other oligo introduces a synonymous mutation in the same codon. Lastly, we designed 30 guides and donors with random sequence (referred as “Random DNA guides”). In total, we designed 32,000 guide/donor oligos. Oligonucleotide libraries were manufactured by Twist Biosciences (Table S5).

Yeast Library Preparation

Oligonucleotide libraries were amplified using primers 310 and 313 with Q5 hot-start DNA polymerase (New England Biolabs) following manufacturer's instructions, producing amplicons with flanking 20 bp sequences homologous to plasmid pZS165 (Table S4). PCR-amplified double-stranded DNA were cloned into NotI-digested pZS165 using NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs). Assembly was performed with a molar ratio of vector:insert = 1:5. The assembled plasmids were ethanol precipitated and diluted to a concentration of 80 ng/ul.

Libraries were subsequently transformed via electroporation into Endura electrocompetent cells (Lucigen). We performed 8 electroporations, each with 25ul of Endura cells and 80 ng of library plasmid DNA. Electroporations were conducted in 0.1cm-gap cuvettes in a GenePulser (Bio-Rad). Transformed bacteria was allowed to recover in recovery media (supplied with Endura competent cells) for one hour, and then plated on LB/Carbenicillin plates for incubation at 37C overnight. Dilutions were also plated to estimate cfu/transformation.

Plasmids were extracted using EZNA Plasmid Maxi Kit (Omega Biotek). Plasmids were eluted in 3 mL of kit-provided elution buffer. Samples were subsequently digested with NotI-HF and CIP (New England Biolabs) in order to linearize empty vectors. Plasmids were then ethanol precipitated to concentrate library at approximately 0.8 ug/ul.

Yeast transformation was performed via electroporation of plasmid library into yeast strain ZRS111. Electrocompetent yeast were prepared by modifying previously described methods (Benatuil et al., 2010). Yeast transformants were selected by plating on SD -HIS/-URA 2% agar plates supplemented with 1M sorbitol. Transformation efficiency was estimated by plating diluted transformation recovery culture on both YPD and SD -HIS/-URA 2% agar plates. Yeast transformants were scraped off plates using cell spreaders, resuspended in SD -HIS/-URA and frozen in 15% glycerol stocks at -80C for subsequent editing experiments. 6 million transformants were generated, with an average representation of 176 cells per guide/donor oligonucleotide.

To assess whether each cell contained only one guide/donor oligo, we performed colony PCR followed by Sanger sequencing for 48 yeast transformant clones to detect multi-vector transformation. DNA was extracted with lithium acetate/SDS and PCR was performed using Q5 hot-start DNA polymerase (New England Biolabs) following manufacturer's instructions (Lööke et al., 2011). Out of 48 clones, 45 were successfully sequenced and all of them contained only a single sequencing trace, while the remaining 3 clones had failed sequencing reactions, showing no evidence of multiple plasmids (Table S3).

Pooled editing and growth competition

For pooled genomic editing, transformed yeast libraries were inoculated in 200 mL YNB -HIS/-URA 2% raffinose at OD600 = 0.3, for 24 hours at 30C (approximately 5 generations). This is equivalent to 1.8×10^8 cells, representing ~56,000 cells per guide/donor coverage. The pre-editing cultures were then inoculated into 200 mL YNB -HIS/-URA 2% galactose in two replicates, starting from OD600 = 0.3 for two consecutive 24-hour growth periods at 30C (total of 48 hours under galactose induction, approximately 12 generations). Each editing replicate of edited yeast libraries were briefly amplified in SD -HIS/-URA media for 6 hours starting from OD600 = 0.3 and collected at OD600 = 0.8, for 3 biological replicates of growth competition at 30C. The competition culture was diluted with fresh SD -HIS/-URA media roughly every 3 hours to let cell density oscillate between OD600 = 0.2 and 0.8 for 5 cycles and then every 6 hours oscillating between OD600 = 0.1 and 0.8 for 5 cycles, and collected at every dilution for sequencing library preparation (Figure S1E; Table S5). Our library contained 32,000 guide/donor pair, therefore 200 mL of OD600 = 0.1 contain an average representation of 18,750 cells per guide/donor.

Library sequencing

2×10^9 cells were collected from each time point and digested with 250 ug/mL Zymolyase 20T in 50 mM KPi pH 7.4 buffer at 37C for 1 hour. Plasmids were extracted from yeast cells using Plasmid Plus MAXI kit (QIAGEN) with manufacturer's instructions starting from the P2 Lysis Buffer addition step. Plasmids were further concentrated with DNA Clean and Concentrator kit (Zymo Research). The guide/donor oligonucleotide libraries were amplified using primer 73 and 8 staggered reverse primers with homology to vector sequence and partial Illumina adaptor sequences (primers 443-450), using Q5 hot-start DNA polymerase (New England Biolabs) following manufacturer's instructions (Table S4). Staggered primers add 1-8 nucleotides to mitigate low complexity sequences at the 3' end for read 2. Illumina sequencing adaptors with custom barcode indices (P5:122-124; P7: 54-71, 455-460) were added to amplicons by PCR, followed by size selection, allowing multiplexed pooling of 24 samples per sequencing lane. The libraries were sequenced using 3 lanes of Illumina HiSeq 4000, 150bp paired-end (24 samples per lane), dual index workflow with a custom read 1 primer (75) (Table S4).

Estimating edited variant effect on fitness

Paired-end sequencing reads were aligned to the oligo library and the full plasmid sequence using the STAR aligner (Dobin et al., 2013) with parameters "--outFilterMultimapNmax 1" and "--outFilterMismatchNmax 0" (Table S5). Only reads that had a fragment length of 173-175, mapped to position 1 of an oligo, and that did not contain any mismatches, insertions, or deletions were used as counts for oligo abundance. At least one paired-end read mapped to 60%–70% of the designed oligos in each sample from the growth competition phase (Figures S1A and S1B). To account for the dependence of the measurement variance on the mean, each oligo's effect on growth rate was estimated using limma-voom (Ritchie et al., 2015) (for variance versus mean plot see Figure S1E; we note that DESeq2 (Love et al., 2014), which uses a similar model to MAGeCK (Li et al., 2014), yielded highly similar

results to limma-voom but with smaller p values for the significant effects. We therefore used limma-voom to be conservative). Guide/donor oligomer editing effects on fitness were estimated as the effect of time in “generations” (culture doublings), that were estimated by the optical density measurements, on the log fraction of paired-end reads that mapped to each oligo out of the total mapped reads in the sample (see reproducibility in [Figures 2](#) and [S1](#)). The model also contains an effect for replicate ID (which controls for different fractions of the oligo at the first time point in different replicates) and for sequencing lane. Thus, the full linear model was

$$\log_2(f_{i,t}) = \beta_{0,i} + \beta_{1,j}1\{\text{seq. lane} = 2\} + \beta_{2,j}1\{\text{seq. lane} = 3\} + \beta_{3,j}t$$

where $f_{i,t}$ is the fraction of reads mapped to oligomer i at time t out of all mapped reads in this sample, and t is the time in generations from the start of the growth competition. The goal was to estimate $\beta_{3,j}$ – the fitness effect (\log_2 fold change) of oligomer i . Note that this modeling approach approximates the relative fitness by assuming a fixed average fitness over time. This is only an approximation as the average fitness is expected to increase to some extent over time. However: a. According to optical density measurements the growth rate is very stable over time ([Figure S1D](#)) b. Assuming that the significant effects that we have detected are correct and real, the abundance measurements can be used to estimate how the average fitness increases over the experiment. According to this, the change in average fitness is about 5% of the relative fitness of the variant with the smallest, significantly detected effect. In conclusion, assuming fixed average fitness should not have large effect on our estimates. Detection of reproducible effects set was performed using IDR ([Li et al., 2011](#)) with similar parameters as the ENCODE project ($\mu = 0.1$, $\sigma = 1.0$, $\rho = 0.2$, $p = 0.5$, $\text{eps} = 10^{-6}$, $\text{max.ite} = 3000$) on the $-\log_{10}(p \text{ values})$. We removed seven variants whose replicates did not agree on the direction of the effect. The effects of multiple oligos per variant were combined by taking a weighted average with the inverse of the variance used as weights. The combined p value was calculated using Fisher’s method separately for p values calculated using a one-tailed positive test and a one-tailed negative test and using the more significant p value. The above data can be found in [Table S5](#).

Genomic annotation, GO and TFBS analyses

Yeast BY-RM variants were annotated using Variant Effect Predictor (VEP) ([McLaren et al., 2016](#)). The annotations were parsed to retrieve the closest annotation and the genes on both sides on the variant. For intergenic variants, the annotations were used to call whether it is in a divergent promoter, unidirectional promoter or convergent intergenic (between two gene ends). Dubious genes were filtered out from the analysis. GO enrichments were calculated using Gorilla ([Eden et al., 2009](#)). Three different sets of transcription factor binding sites were analyzed for overlap with variants: [Pachkov et al. \(2013\)](#) was download from SwissRegulon and filtered for score > 0.95 , [Harbison et al. \(2004\)](#) and ORegAnno ([Griffith et al., 2008](#)) were downloaded from UCSC genome browser sacCer3 genome version. Gene coordinates were obtained using biomaRt ([Durinck et al., 2009](#)) R package from the “scerevisiae_gene_ensembl” dataset of Ensembl.

Validation experiments

14 guide/donor oligos were selected for validation, resynthesized, and cloned into pZS165 (IDT). The *ADE2* gene was knocked out in ZRS111 to create ZRS112. ZRS112 colonies are pink when grown on low adenine media. Individual plasmids containing the validation guide-donor oligos as well as a non-editing, control plasmid (pZS165 containing *ade2*hdr-sgGFP oligo) were transformed into both ZRS111 and ZRS112. The transformants were grown in YNB -HIS/-URA 2% raffinose for 18-24 hours and YNB -HIS/-URA 2% galactose for 48 hours with 1:30 dilution every 24 hr. Individual clones were genotyped by PCR amplification of target loci and Sanger sequencing (106 clones in total, 102 edited, 4 non-edited, see [Tables S3](#) and [S4](#)). Clones with the correct edit, were selected for growth competition experiment. Equal amounts of control strain and edited strain were mixed with a starting optical density at 600 nm (OD600) of 0.025, with a volume of 1 mL in SD -HIS/-URA media. Importantly, the competition was setup in a reciprocal manner, with ZRS111 and ZRS112 alternating as control and edited strain. This would control for any genetic interactions between the edited variant and *ADE2* knockout allele in ZRS112. The cultures were grown for 5 hours shaking at 30C (3 generations) and plated to SD -HIS/-URA low-ADE plates in triplicates. The cultures were then diluted 1:32 by volume every 8 hours for 40 hours, spanning approximately 25 generations. The cultures at 40 hours were plated to SD -HIS/-URA low-ADE plates in triplicate. The plates were incubated at 30C for 2 days and moved to room temperature for 2 days. The number of white and pink colonies on each plate was recorded to determine the relative fitness of control and edited strains ([Table S2](#)).

Estimating editing effect in validation strains

In the validation experiments described above, the growth of a control strain (BY) is compared to a single strain, which was edited by a specific guide/donor oligo, by growing both in the same well. The number of cells of each of the two strains at the start and end of the growth competition is estimated by plating cells on solid media and counting colonies - one strain’s colonies are pink while the other strain’s colonies are white. The effect on growth of a guide/donor edit in these validation experiments was estimated using the following scheme. Consider M competition experiments, in which a control strain c is competing against one of the 14 validated strains a . Let S_m^0 be the total number of colonies collected at the start ($t=0$) and A_m^0 be the number of colonies out of S_m^0 that are of the edited strain. Similarly let $S_m^{t_1}$ and $A_m^{t_1}$ be the total and strain a number of colonies collected at the end ($t=t_1$) of the m ’th experiment. Let $N_{C,m}^0, N_{A,m}^0, N_{C,m}^{t_1}, N_{A,m}^{t_1}$ be the number of cells in the medium of the control and the edited strain at the start and end of the m ’th experiment. Let β_A and β_C be the exponential growth rate coefficients of the edited and control strains and

let $\Delta\beta = \beta_A - \beta_C$. We assume that β_C is independent of β_A and is thus constant across strain competitions. Our goal is to estimate $\Delta\beta$, which is expected to be equal to the \log_2 fold change in the pooled experiment. Finally, let $\alpha_m = \frac{N_{A,m}^0}{N_{C,m}^0}$. Then: $\frac{N_{A,m}^0}{N_{A,m}^0 + N_{C,m}^0} = \frac{\alpha_m}{\alpha_m + 1}$. Under our model of deterministic / non-stochastic growth,

$$\frac{N_{A,m}^{t_1}}{N_{A,m}^{t_1} + N_{C,m}^{t_1}} = \frac{N_{A,m}^0 2^{t_1 \beta_A}}{N_{A,m}^0 2^{t_1 \beta_A} + N_{C,m}^0 2^{t_1 \beta_C}} = \frac{\alpha_m 2^{t_1 \beta_A}}{\alpha_m 2^{t_1 \beta_A} + 2^{t_1 \beta_C}} = \frac{\alpha_m 2^{t_1 \Delta\beta}}{\alpha_m 2^{t_1 \Delta\beta} + 1} = \sigma(\log(\alpha_m) + \log(2) t_1 \Delta\beta),$$

where σ is the sigmoid function. Using this we can write

$$A_m^0 \sim \text{Bin}(n = S_m^0, p = \sigma(\log(\alpha_m)))$$

and

$$A_m^{t_1} \sim \text{Bin}(n = S_m^{t_1}, p = \sigma(\log(\alpha_m) + \log(2) * t_1 * \Delta\beta))$$

We estimate $\Delta\beta$ by fitting the data using the `statmodels` package in *Python*.

The selection coefficient (relative to the BY strain) is $s = 2^{\Delta\beta} - 1$. The validation experiment estimates the fitness of the edited strain relative to BY strain while the pooled experiment estimates the fitness relative to the average fitness. Therefore, we deduced the BY fitness that was estimated by the pooled experiment (median fitness of “Random DNA guides”) from the edited strain pooled estimated fitness.

Measuring gene expression in validation strains

For testing the effect of editing on expression of nearby genes in the validation strains, 3 edited clones per strain were harvested during log-phase growth and RNA was extracted with YeaStar RNA kit (Zymo Research). Yeast total RNA was treated with RQ1 DNase (Promega) and ethanol precipitated, followed by cDNA synthesis using SuperScript First-strand synthesis system (Invitrogen). Gene expression was assayed by Real-Time quantitative PCR (RT-qPCR), amplified using iQ SYBR Green Supermix (Bio-Rad) and custom qPCR primers (Table S4) on the CFX384 Real-Time PCR detection System (Bio-Rad). Expression of target genes were calculated by normalizing to reference gene *ALG9* mRNA expression. Each qPCR experiment was performed in three replicate wells. Statistical tests between validation strains (RM allele) and control strain (BY allele, labeled as WT) were analyzed using Prism GraphPad software, with unpaired t test with Welch’s correction.

Testing enrichment of ribosomal pQTLs

We sought to test the hypothesis that the fitness effects of variants in promoters of “cytoplasmic translation” genes (hereafter referred to as ribosomal) might act via changes in the ribosome—e.g. total ribosomal abundance, or differential abundance of particular subunits that affect translation of specific target genes (Shi et al., 2017). In principle, this could be tested by asking whether the presence of BY versus RM alleles at each ribosomal gene predicts the protein levels of other genes in *trans*, more than expected by chance. QTLs for protein abundance have previously been mapped in a BY/RM cross (Albert et al., 2014a); while some of these are likely to reflect translation variation, others agree in direction with eQTLs (QTLs for mRNA levels) for the same QTL/target gene pairs, suggesting that some pQTLs are a result of changes in mRNA. To define a conservative set of likely “post-transcriptional pQTLs,” we considered only pQTLs where the eQTL signal (Brem and Kruglyak, 2005) was in the opposite direction—e.g. the BY allele increasing protein level but showing no increase in mRNA level (note this does not require the eQTL to be significant, only different in direction). We further split the analysis by allelic direction, testing pQTLs where the BY allele increases protein levels separately from those where the BY allele decreases levels. Out of 1025 total reported pQTLs (Albert et al., 2014a), we classified 339 as likely post-transcriptional (with 187 BY-upregulating and 152 RM-upregulating).

Out of these 187 BY-upregulating pQTLs, we found 79 whose 2-LOD confidence intervals overlapped with one of the 25 ribosomal gene promoters with RM-fitter variants (i.e., an average of ~ 3.2 target proteins per ribosomal gene), compared to only 20 RM-upregulating. To determine whether this was different from what is expected by chance, we tested two randomization approaches. First, we chose 10,000 groups of 25 genes at random, and performed the same pQTL overlap analysis. Second, we chose 10,000 groups of 25 ribosomal genes at random (to control for the possibility that ribosomal genes are generally enriched for post-transcriptional pQTLs). We found that only 1.2% of random ribosomal gene groups and 1.0% of genome-wide gene groups had at least 79 BY-upregulating post-transcriptional pQTLs, suggesting that this enrichment is unlikely to occur by chance. In contrast, we found most random gene groups had at least 20 RM-upregulating post-transcriptional pQTLs (81% of ribosomal gene groups and 79% of genome-wide groups). Therefore, we hypothesize that the RM-fitter variants in 25 ribosomal gene promoters may result in lower translation of some genes. Since only 160 genes were tested as potential pQTL targets (Albert et al., 2014a), we had little power to find functional enrichments among the affected genes, and did not identify any enriched GO categories when accounting for the 160-gene background set.

Quantification and Statistical Analysis

Guide/donor oligo abundances were measured using Illumina HiSeq 4000, 150bp paired-end reads as described in the “Library sequencing” section. Sequencing reads were mapped to the 32,000 guide/donor oligo using STAR as described in the “Estimating edited variant effect on fitness” section to obtain guide/donor oligo counts in each sample (see Snakemake workflow in <https://github.com/eilon-s/CRISPEY>). The counts of 11 samples from each of 6 growth competition replicates (3 from each of the 2 editing replicates) were used to estimate the editing effect on fitness using limma-voom as described in the “Estimating edited variant effect on fitness” section. Effects sizes were also estimated separately for each of the 3 growth competition replicates following each editing phase replicate. These were used for the IDR analysis. Estimation of editing effects on fitness in the validation experiment is described in the “Estimating editing effect in validation strains” section. Analysis of transcription factor binding sites and gene ontology enrichment is described in the “Genomic annotation, GO and TFBS analyses” section. RT-qPCR were performed with 3 edited clones per strain or control clones as biological replicates and triplicate wells per strain for technical replication. Statistical analysis for RT-qPCR was described in the “Measuring gene expression in validation strains” section.

DATA AND SOFTWARE AVAILABILITY

The accession number for the raw sequencing data reported in this paper is NCBI Sequence Read Archive: SRP126558. All software and code use for the design and getting oligo counts are available on <https://github.com/eilon-s/CRISPEY>. Code used to create specific figures is available upon request. Sequence for oligonucleotides used in this study can be found in [Table S4](#). Raw Sanger sequencing reads can be found at: <https://doi.org/10.17632/jtbnrmstt.1>.

Figure S1. Reproducibility of Experimental Estimation of Relative Fitness Effects, Related to Figure 2

(a) and (b) show strain coverage during growth competition. (a) Each line shows the distribution of read pairs mapping to each oligo in one sample during growth competition. (b) Each line shows the distribution of estimated cells containing each oligo in one sample during growth competition right before the dilution. The number of cells was calculated by multiplying the fraction of reads mapping to each oligo by the total number of cells. The total number of cells before dilution was estimated to be $0.8 \text{ OD}_{600} \times 200 \text{ mL} \times 3 \times 10^7$ cells. The culture goes through a series of 4.5, 4 and 8-fold dilutions, so to get the size of the bottlenecks caused by the dilutions, divide the cell count prior to dilution by these numbers. (c) Abundance of strains across time points. The heatmap shows the \log_2 fold change during growth competition for strains which had reproducible ($\text{IDR} < 0.05$) effects on growth rate across two sets of three growth competition experiments each starting from independent editing experiments. (d) Showing culture doubling (which we refer to as a generation) during growth competition based on OD measurements (y axis) as a function of time. Colors mark the sample editing and growth competition replicate number. Black line shows a linear fit to the data. (e) Variance-mean trend plot produced by limma-voom (Ritchie et al., 2015). Limma-voom was used to model the effects of each guide/donor on growth rate during growth competition. (f) The heatmap shows clustering based on the Pearson correlation between raw read counts for each pair of samples. X and Y labels state the generation from growth competition start, replicate editing phase number and replicate growth competition number. Dendrogram colors match the time of the sample (i.e., sample number and dilution iteration). Notice that the samples cluster according to sampling time. (g) Correlation between relative fitness estimates using a single replicate. Note that this correlation includes both significant effects that should be correlated and many more neutral or nearly neutral effects that should not show high correlation. (h) Comparing relative fitness estimates using three growth competition replicates. Each set of three replicates follow an independent editing experiment (Pearson's $R = 0.5$). An IDR analysis of the $\log_{10}(p)$ values was used to identify a set of highly reproducible effects ($\text{IDR} < 0.05$, orange). Non-significant effects ($\text{IDR} > 0.05$, blue) show far weaker correlation, as expected. (i) IDR analysis of combined $-\log_{10}(p)$ value of each RM variant's effect on growth, multiplied by the direction of the effect. Guide/donor p values were combined using Fisher's method. Orange dots show values with $\text{IDR} < 0.05$. (j) A summary of the reproducibility analysis.

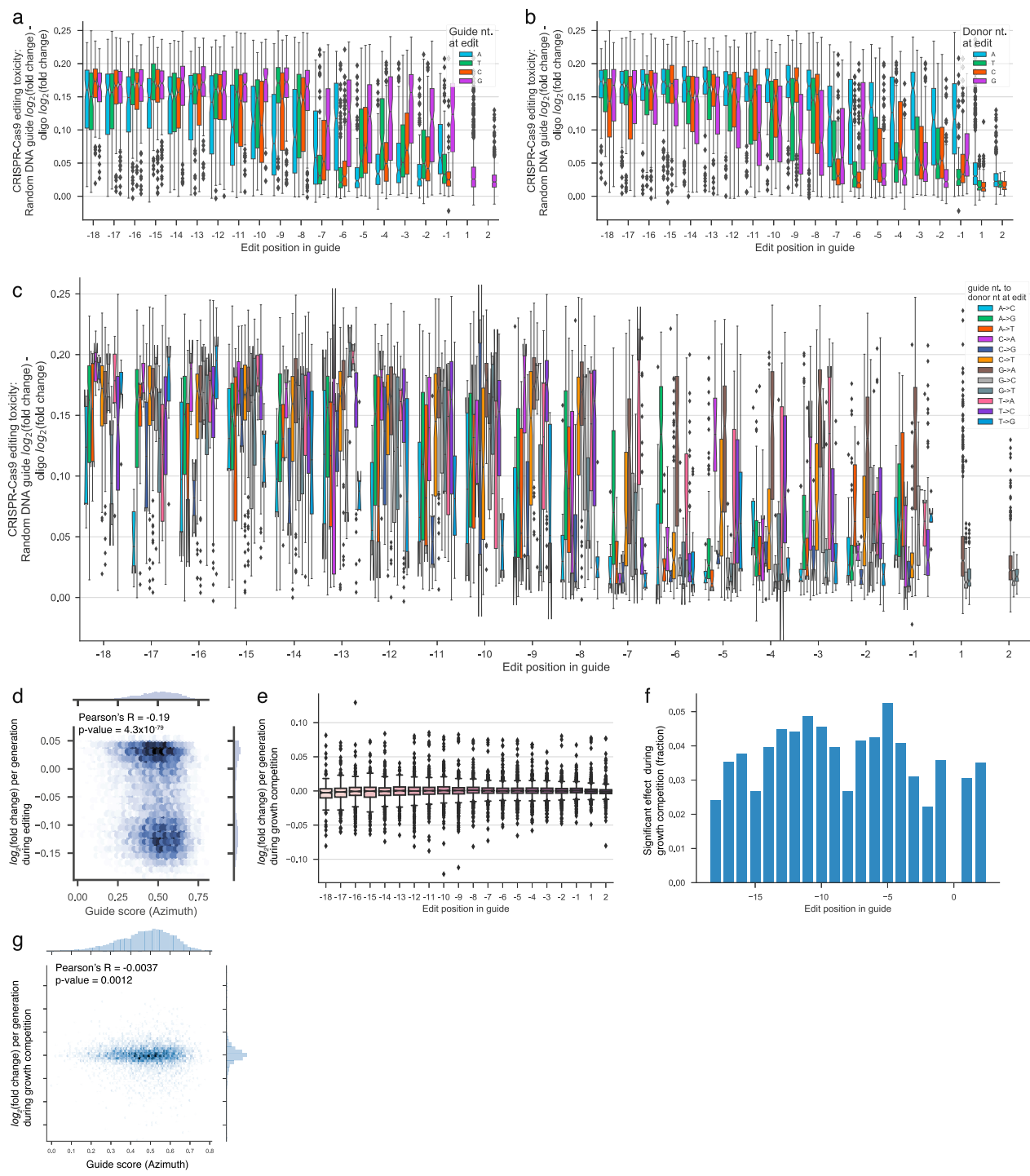


Figure S2. CRISPR-Cas9 Editing Toxicity Is Sensitive to the Target and Edited Target Sequence, Related to Figure 3

Panels (a), (b) and (c) y-axes show the difference between the mean \log_2 fold change per generation of oligos that contain a random DNA guide (and therefore will not cut the target) and the oligo \log_2 fold change per generation during editing. Higher values reflect higher toxicity and therefore more cutting events. (a) shows the guide (target) nucleotide in the edited position. (b) shows the donor (edited target) nucleotide in the edited position. (c) shows both the guide (target) and the donor (edited target) nucleotide in the edited position. Bar colors represent a specific change from guide (target) nucleotide to donor (edited target) nucleotide, as indicated in the inset. (d) Show \log_2 fold changes per generation during 48h editing by guide Azimuth (Doench et al., 2016) score. Panels (a)-(d) show results only

(legend continued on next page)

for oligos that edit a synonymous mutation by 1bp replacement. In comparison to the above, panels (e)-(g) show results for changes in guide/donor oligo abundance during the growth competition (in which Cas9 is not expressed). (e) and (g) both show the \log_2 fold change per generation and (f) shows the fraction of guide/donor oligos with significant \log_2 fold change per generation as a function of the edit position in the guide ((e) and (f)) and of guide quality score (Doench et al., 2016) (g). Whiskers in panels (a), (b), (c) and (e) show the distribution range (excluding outliers).

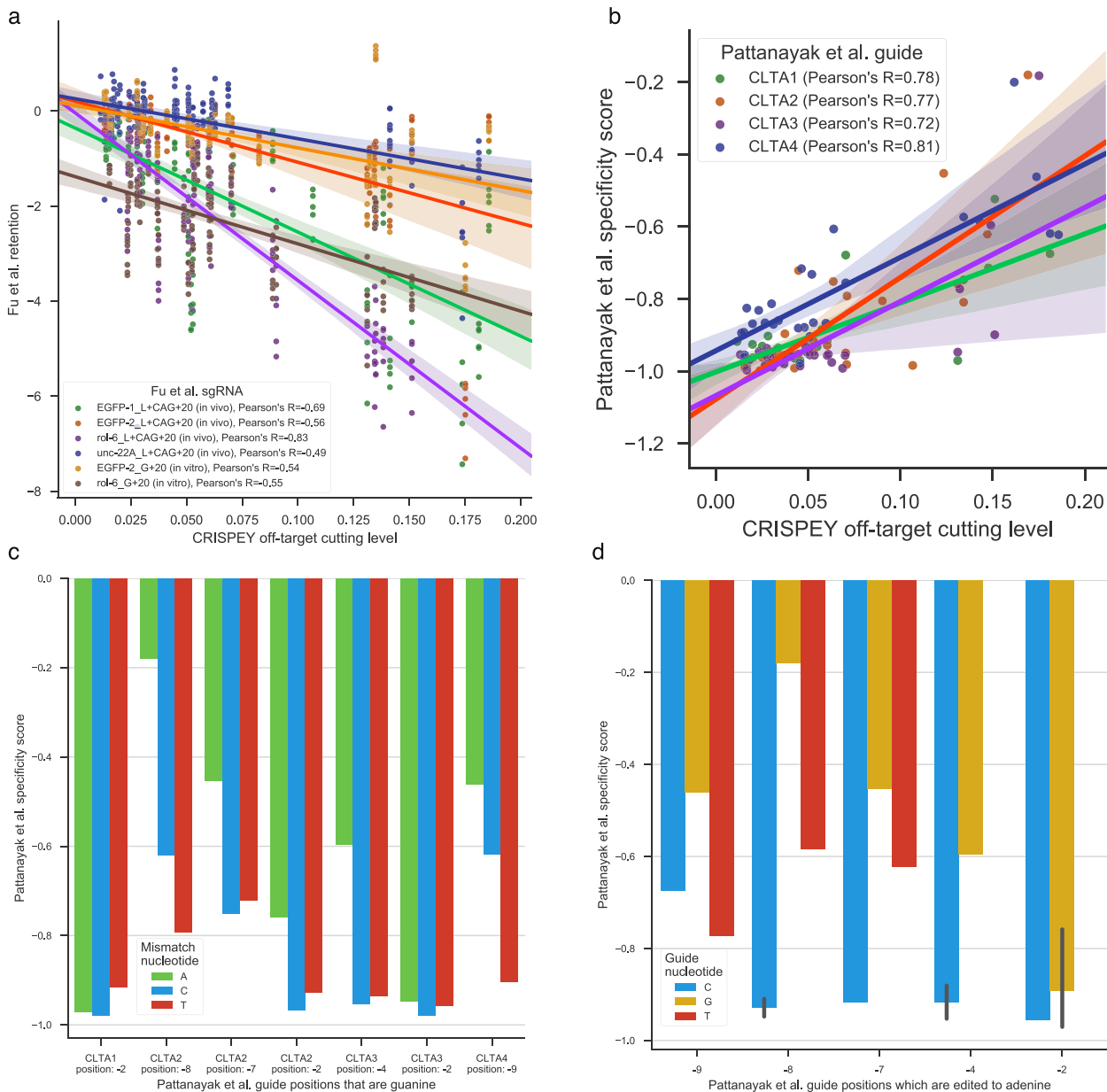


Figure S3. CRISPR-Cas9 Off-Target Cutting Level Is Correlated with Previous Cas9 Mismatch Sensitivity *In Vitro* and *In Vivo* Measurements, Related to Figure 3

(a) and (b) show an agreement between CRISPR-Cas9 off-target cutting level and previous measurement of Cas9 sensitivity to mismatches in *in vivo* (Fu et al., 2016) and *in vitro* (Fu et al., 2016; Pattanayak et al., 2013). Previous measurements were done using few guides and therefore may be affected by the specific guide-target context. Nevertheless, data from Pattanayak et al. agree with our observation of lower sensitivity to mismatches with adenine when the matching guide nucleotide is guanine (c) and to having guanine in the guide when the mismatch is with adenine (d). Data are represented as mean \pm SEM.

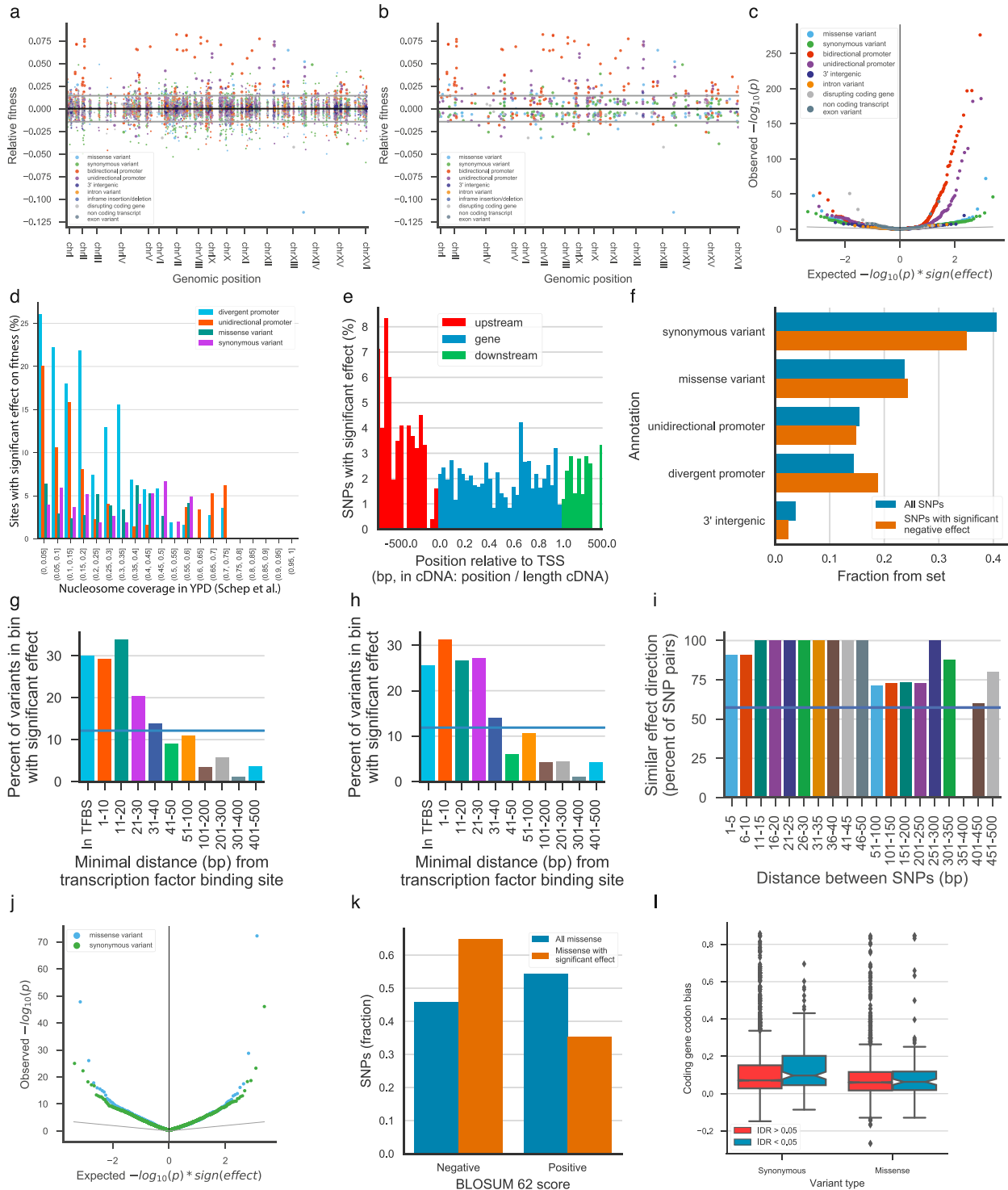


Figure S4. Characterizing Variants Affecting Fitness, Related to Figures 4 and 5

(a) Manhattan plot showing the relative fitness of editing a site from the BY variant to the RM variant. Larger dots mark IDR < 0.05. (b) similar to (a) showing only site effects with IDR < 0.05. Horizontal gray lines in (a) and (b) mark $\pm 1\%$ selection coefficient ($2^{\log_2(\text{fold change})} - 1$). (c) Quantile-quantile plot of fitness effect p values by annotation category. Positive = RM-fitter, negative = BY-fitter. (d) Enrichment of fitness effects in promoters is not caused by higher chromatin accessibility. y axis shows the fraction of variants with significant effect in each nucleosome coverage under exponential growth in YPD (Schep et al., 2015) bin by genomic annotation. (e) Alignment of the fraction of significantly negative effect variants relative to TSS. Showing the percent of variants with significant effect in upstream, gene, and downstream regions. (f) Horizontal bar chart showing the fraction of SNPs with significant negative effect across different annotations. (g) Bar chart showing the percent of variants with significant effect in different bins relative to transcription factor binding sites. (h) Similar to (g) but for a different set of bins. (i) Bar chart showing the similar effect direction (percent of SNP pairs) for different distances between SNPs. (j) Quantile-quantile plot of fitness effect p values by annotation category, similar to (c). (k) Bar chart showing the fraction of SNPs with significant effect relative to BLOSUM 62 score. (l) Box plot showing coding gene codon bias for different variant types.

(legend continued on next page)

each bin of variants with similar position relative to the translation start site. Bins outside the TSS are 50bp wide. Variants in the coding region were assigned a value according to their relative position in the coding sequence: (variant position in CDS / CDS length) * 1600 bp, and then were binned in to non-overlapping 50 bp bins. (f) Annotation enrichment among variants with significantly negative effect (IDR < 0.05). Enrichments are weak due to weaker signal overall. (g) and (h) show the percent of variants with significant effect as a function of their distance from the nearest transcription factor binding site. (g) shows the results for transcription factor binding sites from ORegAnno (Griffith et al., 2008). (h) shows the results for transcription factor binding sites from Harbison et al. (2004). The blue lines show the result over all non-genic variants. (i) Nearby sites with significant effect on fitness (IDR < 0.05) favor the same parent strain. This panel show the same analysis as Figure 5D except with finer bin resolution. Bin 351-400bp does not contain data points. (j) Quantile-quantile plot showing the p value distributions of the effects of synonymous and missense mutations. (k) Non-conservative amino acid changes are enriched in missense variants with significant effects. The plot show that missense variants with negative BLOSUM62 are enriched in the set of missense variants with significant effect (IDR < 0.05; Binomial p value < 10^{-3}). (l) Synonymous variants with significant effect on growth tend to be in genes with high codon bias. Whiskers show the distribution range (excluding outliers). Codon bias was calculated by CodonW, and downloaded from SGD. The difference between significant (IDR < 0.05) variants and the rest was significant for synonymous variants (Kolmogorov-Smirnov, p value = 0.017). No difference was observed for missense variants (Kolmogorov-Smirnov, p value = 0.50), suggesting that the difference is not due to increased importance/fitness effects of high codon-bias genes overall.

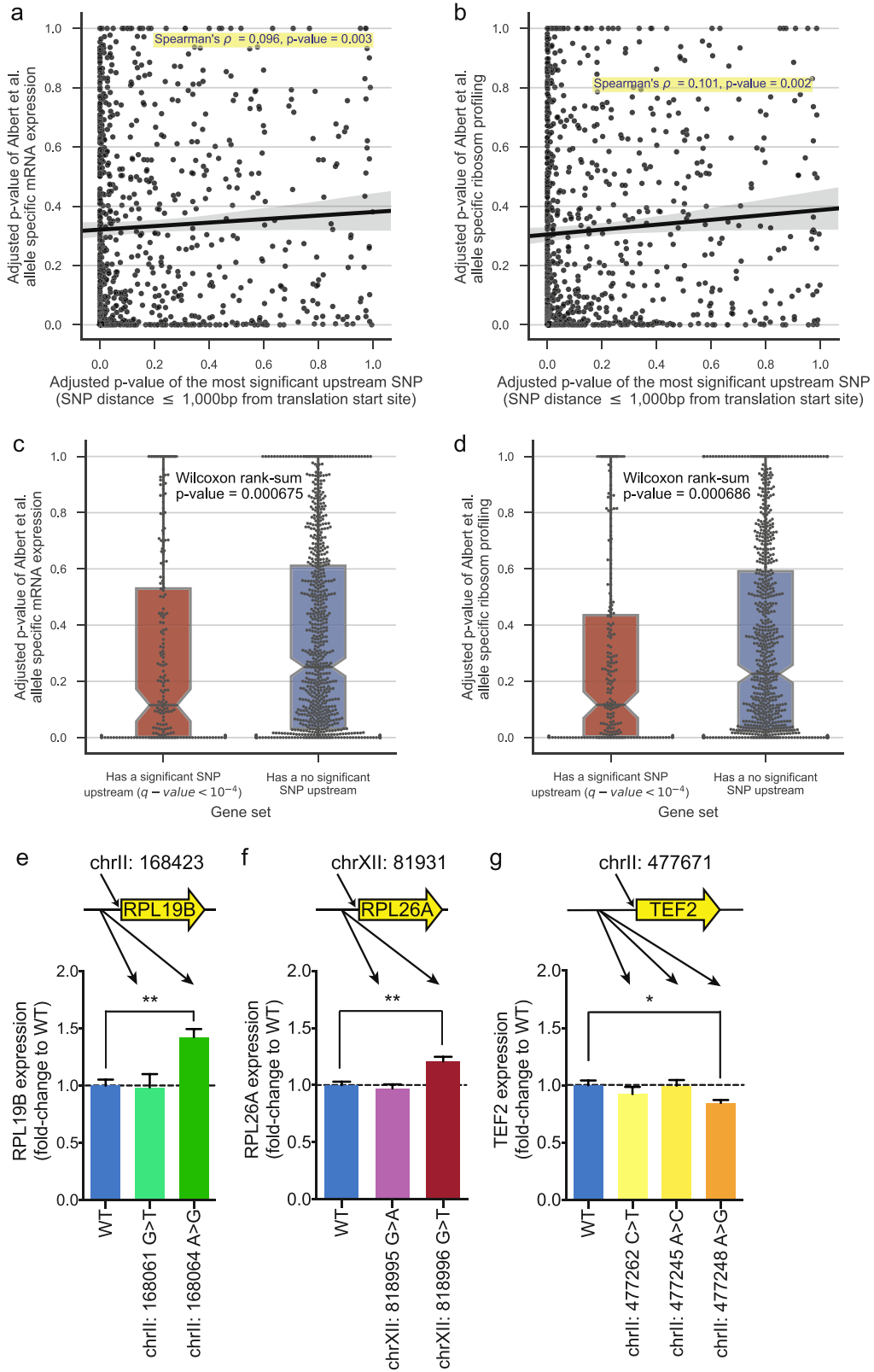


Figure S5. The Effect of BY-RM Variants on Gene Expression, Related to Figure 4

Panels (a)-(d) show that genes with significant upstream variants are more likely to exhibit allele-specific expression. (a) and (b) Scatterplots showing adjusted p value of the most significant variant upstream of the gene versus allele-specific mRNA expression (a) and ribosome footprints (b) from [Albert et al. \(2014b\)](#). (c) and (d) compare allele-specific mRNA expression (c) and ribosome footprints (d) between genes with and without significant variant upstream (significant is defined as adjusted p value $< 10^{-4}$ and variants are considered up to 1000 bp upstream). Whiskers in panels (c) and (d) show the distribution range (excluding outliers). Panels (e)-(g) show that RM-allele replacement alters expression of translation-related genes. (e) Editing chrIII:168064 site at the YBL028C-RPL19B promoter to the RM allele increases expression of RPL19B, a ribosomal subunit, by 42% (unpaired t test, Welch's corrected p value = 0.0082). (f) Editing chrXII:818996 site at the RPL26A promoter to the RM allele reduces expression of RPL26A, a ribosomal subunit, by 21% (unpaired t test, Welch's corrected p value = 0.0024). (g) Editing chrIII:477248 site at the TKL2-TEF2 promoter locus to the RM allele reduces expression of TEF2 by 15% (unpaired t test, Welch's corrected p value = 0.0267), a eukaryotic translation elongation factor. Strains that were edited to the RM allele and control strains (WT, BY genotype) were individually grown to log-phase and assayed with RT-qPCR. Data are represented as mean \pm SEM. Expression changes of edited strain relative to WT for each gene were detected using unpaired t test between RM-allele and BY-allele strains with Welch's correction (*p < 0.05; **p < 0.01).