Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Biotechnol. Author manuscript; available in PMC 2018 November 21.

Published in final edited form as:

Nat Biotechnol. 2018 July ; 36(6): 540–546. doi:10.1038/nbt.4147.

High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR/Cas9 in yeast

Xiaoge Guo[1,2,*], **Alejandro Chavez**[1,2,3,†,*,#], **Angela Tung**[1,*], **Yingleong Chan**[1,2], **Christian Kaas**[1,2,4], **Yi Yin**[5], **Ryan Cecchi**[1], **Santiago Lopez Garnier**[1], **Eric Kelsic**[1,2], **Max Schubert**[1,2], **James E. DiCarlo**[1,2,6], **James J. Collins**[1,7,8,9,10], and **George M. Church**[1,2,†]

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA

[2]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA

[3]Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA

[4]Department of Expression Technologies 2, Novo Nordisk A/S, Maaloev 2760, Denmark

[5]Department of Genome Sciences, University of Washington, Seattle, Washington, USA

[6]Department of Ophthalmology, Columbia University College of Physicians and Surgeons, New York, New York, USA

[7]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[8]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[9]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[10]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

Abstract

Construction and characterization of large genetic variant libraries is essential for understanding genome function, but remains challenging. Here, we introduce a Cas9-based approach for generating pools of mutants with defined genetic alterations (deletions, substitutions, and insertions) with an efficiency of 80–100% in yeast, along with methods for tracking their fitness *en masse.* We demonstrate the utility of our approach by characterizing the DNA helicase SGS1 with

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

†Correspondence can be addressed to: Alejandro Chavez, ac4304@cumc.columbia.edu; George M. Church, gchurch@genetics.med.harvard.edu.

*Co-first authors

#Current address: Department of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, New York, USA

COMPETING FINANCIAL INTERESTS

G.M.C. is the founder and holds leadership positions of many companies (http://arep.med.harvard.edu/gmc/tech.html). X.G., A.C., M.S. and E.K. have filed a patent application (U.S. Patent Application 62/348,438) relating to this work.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

small tiling deletion mutants that span the length of the protein and a series of point mutations against highly conserved residues in the protein. In addition, we created a genome-wide library targeting 315 poorly characterized small open reading frames (smORFs, <100 amino acids in length) scattered throughout the yeast genome, and assessed which are vital for growth under various environmental conditions. Our strategy allows fundamental biological questions to be investigated in a high-throughput manner with precision.

Libraries of cells with defined genetic alterations have proven transformative for connecting poorly understood genes to biological pathways and uncovering novel roles for previously characterized genes. However, in eukaryotes these libraries have been difficult to generate, and even in some widely used collections, such as the yeast knockout library, a majority of the members contain undesired secondary mutations[1] and suffer from the presence of selection markers[2].

In this work, we present a Cas9-based strategy for the simultaneous, seamless creation of hundreds of genetic variants without integrated selection markers in wild-type yeast cells that express the Cas9 protein along with a donor repair template. Our system is built upon CRISPR/Cas9 and its ability to stimulate homology-directed recombination (HDR) repair of a double-stranded break at a given target locus[3]. Each isogenic mutant is generated by a plasmid containing a single guide RNA (sgRNA) paired with a corresponding donor template that carries a programmed mutation (hereon referred to as the guide+donor strategy) (Figure 1a). The advantages of our concatenated guide+donor design are threefold; it enables: a) rapid cloning of all library members within one reaction, b) simultaneous delivery of both the guide and the donor in one contiguous unit thus preventing uncoupling that may result in inefficient repair and unproductive repair outcomes, and c) high-throughput molecular phenotyping using next generation sequencing (NGS) with guide +donor-containing plasmids serving as unique barcodes for tracking edited cells. A similar concept of *in cis* delivery of guide+donor was recently demonstrated in bacteria[4].

In our initial test, we integrated a copy of the *cas9* gene into the neutral *HO* locus and performed individual transformations of 34 guide+donor plasmids (Supplementary Figure 1). Upon selecting for cells with the guide+donor, however, we found that the number of colonies with the desired genetic alteration was low (0–30%), consistent with earlier attempts at *in cis* guide+donor delivery in yeast[5] (Supplementary Table 1). We sought to increase the percentage of correctly edited cells in order to enable efficient genome-scale measurements via NGS.

To test if linearization of our guide+donor plasmid would increase the efficiency of our system[6–10], we introduced our guide+donor substrate as two linear pieces of DNA. The larger DNA fragment contained the guide+donor portion of the plasmid with an internal portion of the selection marker removed. The smaller DNA fragment consisted of the missing segment of the selection marker with ~150bp of flanking homology such that HDR was required to reconstitute the full circular plasmid (Supplementary Figure 1a). With the modified approach, we observed a 6–14 fold increase in transformation efficiency (Supplementary Figure 1b) with 80–100% of the transformants containing the desired repair event, which is in stark contrast to the 0–30% proper editing observed with the unmodified

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

method[4,5] (Supplementary Table 1). No programmed edits were observed in the absence of Cas9 (Supplementary Table 1).

To begin characterizing the limitations of our system, we tested a series of vectors designed to introduce either targeted point mutations, short deletions, or sequence replacements within the *ADE2* locus. For programmed point mutations, we obtained a genome modification efficiency close to 100% for changes that occurred proximal to the Cas9-generated cut site (Supplementary Figure 2). In contrast, when the desired mutation was positioned further away from the Cas9 cut site, we noted a decrease in efficiency, with mutations 12–15bp away showing rates of editing of ~40%. While longer homology length (Supplementary Figure 3a) increases the number of colonies obtained per transformation (Supplementary Figure 3b), it does not substantially improve the proportion of correct edits (Supplementary Table 2). Of clones that fail to obtain the desired point mutation, the majority mutate the protospacer adjacent motif (PAM) as designated on the provided guide +donor to escape Cas9 cutting. Further characterization of our method for generating programmed deletions revealed that our design allows efficient removal of up to 61 contiguous bases (>90% of colonies with the desired change) but experiences a sharp decline in efficiency in creating larger deletions ((≥121bp) (Supplementary Figure 4a). Similarly, our strategy enables efficient replacement of 61bp of endogenous sequence with up to 60bp of user defined sequence (Supplementary Figure 4b).

Having gained insight into the limitation of our guide+donor strategy, we next sought to determine the generality of our method by targeting three additional loci (*SGS1, SRS2*, and *ARS214*) with a series of point mutations, deletions, and sequence replacements. Similar to our initial results, we obtained a high efficiency of genome modification (90–100%), across all targets and mutation types (Figure 1b).

To examine the targeting specificity of our Cas9-based platform, we performed whole genome sequencing on three mutant strains (*ade2Δ61bp*, *sgs1Δ60bp*, and *sgs1Δatg*) generated via our guide+donor method and observed the expected genomic edits (Supplementary Figure 5). Upon surveying all the regions in the genome that have up to 2 mismatches within the N20 guide sequence, we did not find off-target sites. Off-target effects due to Cas9 are known to result in indels. When the N20 matching parameter was further relaxed to N15+PAM, we did not observe any indels indicative of off-target Cas9 effects.

The strong correlation between the presence of a particular guide+donor plasmid and the presence of the desired genetic alteration should allow us to infer the fitness effects of these modifications by sequencing the abundance of different guide+donor pairs within a mixed pool. To test this hypothesis, we built a small library containing a mixture of guide+donor plasmids designed to modify either the non-essential *ARS214* locus or the DNA damage repair helicase *SGS1* (Figure 1c). Cells arose from the pooled transformation were grown in media with or without the genotoxic agent hydroxyurea (HU) and the abundance of various guide+donor plasmids within the population was determined by NGS. As expected, we observed a marked depletion of guide+donor pairs encoding modifications that disrupted the ATPase domain of Sgs1, which is known to play a critical role in its function (Figure

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

1d)[11–15], whereas mutating the less essential C terminus[11] lead to less depletion (Supplementary Figure 6). Introducing synonymous changes within the ATPase domain or C-terminus of *SGS1* did not show depletion, suggesting that the effects were not due to non-specific disruption of the *SGS1* locus by Cas9. Furthermore, when each of the generated strains was tested individually, the results correlated well with our pooled analysis, lending additional support for the validity of our method (Supplementary Figure 7).

In addition to exposing the mutant library to environmental perturbations, we also examined if our system could be used to observe gene-gene interactions by transforming our small library into cells defective in the structural endonuclease Mms4. In an *mms4Δ* genetic background, all *SGS1* mutants in the library exhibited ~5-fold depletion, consistent with known synthetic sickness between *SGS1* and *MMS4* (Figure 1e)[16,17].

We subsequently applied our method to perform systematic characterization of a single protein and targeted *SGS1*, a gene that encodes the yeast homologue of the human DNA helicase BLM with known roles in mitotic stability, cancer, and aging[18]. To map the critical domains within Sgs1 that provide cellular resistance to the genotoxic stressor HU, we designed a set of guide+donor constructs that generate 20 amino acid deletions with 5 amino acid sliding windows across the majority of the *SGS1* gene (see Materials and Methods for details). Among the regions showing strongest depletion within edited cells were guide +donors deleting amino acid stretches 1–85, 686–1090, and 1116–1225, which correspond to the Sgs1-Top3-binding domain, Sgs1-helicase, and RQC domains, respectively (2-tailed t-test, $P<0.0001$) (Figure 2a, Supplementary File 1)[19–22]. These results are consistent with the known mechanism by which Sgs1 functions through the recruitment of accessory proteins (through N-terminal residues)[12,14,15,19,23–27] and by resolution of DNA structural intermediates via its helicase and RecQ domains[12,28]. We performed biological replicates of our library experiments to assess reproducibility and observed a correlation of 0.86 between the $\log_2$FC observed in the two independent yeast transformations (Figure 2b). Furthermore, we performed individual phenotypic validation of seven hits from the library screen via spot assay and observed similar results (Figure 2c).

Next, we created a series of precise point mutations within Sgs1. Towards this goal, we selected a set of 9 evolutionarily conserved amino acid residues within the Sgs1 helicase domain and attempted to change them to all other possible amino acids using our guide +donor strategy. This library was exposed to increasing concentrations of HU to assay for mutant drug sensitivity. Despite targeting highly conserved residues within Sgs1, all but one tolerated alanine substitution without causing an obvious loss in resistance to our highest concentration of HU at 40mM (Figure 3a, Supplementary File 2). In the case where activity was lost, alanine was used to replace the essential helicase catalytic residue K706. Selecting one representative pair of biological replicates (40mM), we observed a correlation of 0.88 between the first and second biological replicate (Figure 3b). We individually validated six variant hits from the library screen and observed concordant results (Supplementary Figure 8). Overall, we captured expected trends of amino acid substitutions of similar charge and size being well tolerated while those with opposite properties being more detrimental to Sgs1 function.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To determine the capacity of our method to perform targeted editing across the entire yeast genome, we designed and built a guide+donor library for generating small deletions around the initiating ATG for a set of 307 randomly chosen canonical ORFs (including both essential and non-essential genes) along with 315 poorly characterized smORFs. Unlike canonical ORFs, smORFs remain largely ignored and are often missing in modern genome annotations due to their size, low conservation scores, and lack of similarity to known proteins and protein domains.

Using our genome-scale deletion library, we first performed an essentiality screen. We observed strong depletion (~8–100 fold) for all targeted essential ORFs (2-tailed t-test, $P<0.0001$) compared to <3-fold depletion for nearly all nonessential ORFs (2-tailed t-test, $P=0.01$), thus highlighting the specificity and sensitivity of our method (Figure 4a, Supplementary File 3). Out of the smORFs that were examined, 19 smORFs showed similar levels of depletion as our essential controls (2-tailed Z test, $P<0.001$), in line with previous results[29]. When we repeated our screen, we observed a correlation of 0.71 between the two independent biological replicates (Supplementary Figure 9).

Although a number of our smORF library members are located in close proximity to essential ORFs (in some cases within 132bp), our screen did not identify any of them as essential, emphasizing the specificity of our targeting method. To further demonstrate the ability of our guide+donor strategy to characterize a large number of proteins in parallel, we subjected our smORF mutant library to a series of environmental stressors including growth: at 37°C (Figure 4b), in the presence of HU (Figure 4c), or with the antifungal drug fluconazole (Figure 4d). For each of our screens, we identified nearly all of the previously known smORFs with tolerance towards each of the tested conditions, along with uncovering previously unreported roles for a large number of additional smORFs[30]. We individually validated thirteen of the hits from our library screens and observed phenotypes in agreement with the screen results (Supplementary Figure 10).

Of the 315 smORFs examined, 68 were found to play a role in cellular fitness under test conditions. This is in contrast to conventional ORFs for which 104 of 307 tested ORFs were found to be involved in growth under the same environmental conditions (Chi-squared test, $P<0.0001$). Next, we examined features (including amino acid size, gene expression level, secondary structure formation and evolutionary conservation) that could be shared by the smORFs or the ORFs exhibiting biological activity. Although smORFs show a range of sizes across the yeast genome (smallest smORF hit was 28 amino acids), we found that longer smORFs with elevated levels of RNA expression exhibited a trend of being more likely to come up as hits in our screen (Supplementary Table 3). Notably, ORFs showed no such correlation with regard to length, but maintained a similar trend with respect to expression (Supplementary Table 4). Moreover, we did not observe any difference in the prevalence of structural elements (e.g. alpha-helices and beta-sheets) within smORF hits as compared to non-hits. We did, however, observe an increased propensity for beta-sheets and a decrease in unstructured loops when smORFs as a whole were compared to the set of ORFs that were also examined in our screens (Supplementary Table 5). Finally, a large difference in the rate of gene conservation was found with 32 of the 68 smORF hits being conserved in humans as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

compared to only 43 of the 247 smORFs that showed no effect upon the examined conditions (Chi-squared test, P<0.0001) (Supplementary Table 6).

Here, we present a high-throughput method for the rapid generation and phenotypic characterization of hundreds of mutants and illustrate its potential in domain/residue mapping and functional interrogation of nearly any user-defined genomic target by introducing deletions, amino acid substitutions, and sequence replacements. This enables the creation of specific user-defined loss-of-function, gain-of-function, and altered regulation mutants *en masse*.

By editing the locus within its native context without the need for exogenous markers, we avoid artifacts from using surrogate reporter systems and false positive and negative results due to selection marker-driven positional effects[31]. The high library editing efficiency of our system (85%-95%) (Supplementary Table 7) allows users to read the guide+donor sequence on the plasmid delivered to each cell and use the sequence to identify the cell's genotype. Ultimately, this feature enables the fitness of hundreds, potentially thousands, of mutants to be tracked by sequencing the abundance of each guide+donor sequence within a population. While our method employs a similar gap-repair mechanism as reported by Horwitz *et al.* [32], our design is unique in that each guide is concatenated to a corresponding donor repair template, enabling simultaneous delivery of guide+donor.

Our tiling deletion experiment on SGS1 demonstrated our technology's ability to rapidly hone in on the critical domains required for protein function. A similar CRISPR-based protein perturbation concept to identify critical functional domains in mammalian cells[33,34] and in yeast[35] was reported previously. Of note, the underlying mechanisms of functional perturbation between these aforementioned two systems and our guide+donor platform are different in which the former ones rely on unpredictable CRISPR-induced indel and random transposase-induced insertion mutagenesis, respectively, while the variants created by our method are through programmed genetic alterations.

Deep mutational scanning (DMS) methods provided a framework for generating point mutations in a single protein of interest and functionally annotating a large fraction of these amino acid substitutions[35–39]. However, these methods are only meant to interrogate a single gene at a time, which hinders the scale of functional genomics experiments one can perform. In addition, many deep mutational scanning methods are carried out on a plasmid, thus taking the examined protein variant out of its native context[37]. Although our amino acid substitution library was not as exhaustive in its targeting scope as DMS, we are able to target hundreds of genes at a time and perform all of our genetic alterations within the native genomic locus. Previous work by Kastenmayer *et al.* [29] used labor-intensive conventional techniques to make specific gene deletions of 140 smORF mutants. In contrast, we demonstrated the ease of our guide+donor method in rapidly covering over ~79% of the 299 putative smORFs within the yeast genome, including many that had previously been neglected[29]. Given the degree of conservation between yeast and human genomes and the conservation between several smORFs and higher eukaryotes[30], it will be interesting to see if the smORFs identified in our work with roles in stress tolerance have similar functions in humans.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Our method employs the commonly used *Streptococcus pyogenes* Cas9 (SpCas9) which limits the potential target sites because of its PAM-specific requirement. Using Cas9 variants recognizing alternative PAMs[40] could greatly broaden the range of sequences that can be modified by our approach.

Although we have focused on the usage of our technology for high-throughput characterization of coding elements, we envision a broad range of additional applications such as: directed evolution, metabolic engineering, and functional interrogation of non-coding elements. Moreover, given that most clinically relevant mutations are point mutations and the high degree of gene conservation between yeast and humans, our guide+donor editing platform provides an easy way to engineer and test the effects of hundreds of currently uncharacterized single nucleotide polymorphisms that exist within human populations via their nearest yeast orthologue.

MATERIALS AND METHODS

Yeast strains and growth conditions

All strains were derived from YAC2370 (BY4741 derivative; *MATa his3Δ leu2Δ met15Δ ura3Δ*). YAC2563 was constructed by one-step integration of a PmlI-linearized plasmid carrying human codon-optimized Cas9 under the expression of *NOP1* promoter along with a linked NatMX drug selection marker (AC6218) into the *HO* locus. *MMS4* was deleted in YAC2563 background by one-step gene replacement using PCR-generated deletion cassettes (*mms4Δ::KanMX*).

Cells were grown non-selectively in YPAD (1% Bacto-yeast extract, 2% Bacto peptone, 2% dextrose; 1.5% agar for plates) supplemented with 500μg/ml adenine hemisulfate. Ura^+ colonies were selected on synthetic complete (SC) medium deficient in uracil (SC-Ura). All growth was at 30°C. For the experiments with SGS1 mutants, hydroxyurea (Sigma-Aldrich) was added to final concentrations of 5mM, 10mM, 20mM and 40mM. For the smORF library drug conditions, fluconazole (Sigma-Aldrich) and HU (Sigma-Aldrich) were added to final concentrations of 25μg/ml and 100mM, respectively.

Plasmids

Guide+donor plasmids were built in the yeast pRS426 2μm backbone containing the *URA3* selection marker[41]. The guide RNA expression cassette contained *SNR52* promoter, guide RNA sequence, chimeric single-guide RNA structural tail (sgtail), and SUP4 terminator. The donor sequence carrying the desired modification was placed immediately downstream of the terminator sequence. Individual guide+donor fragments were generated from three overlapping PCR fragments using 90-mer oligos from IDT designed to create the guide sequence and its corresponding donor sequence. The ends of the stitched PCR amplicon were designed such that they contained overlapping regions for Gibson assembly. These fragments were then assembled in combination with the plasmid backbone that was digested with NgoMIV and NheI to prepare it to accept the incoming guide+donor sequence. For library cloning described below, the plasmid backbone was further modified to remove BsmBI and SapI sites.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Guide+donor library design

Custom python scripts were used to design the libraries. Oligos were synthesized by CustomArray Inc. For the *SGS1* tiling deletion library with a sliding window of 15bp, we generated donor sequences with 80bp total homology flanking each 60bp deletion region then coupled a 20bp guide RNA that was present in each deletion region closest to the middle of the section being removed. For the Sgs1 amino acid library, we targeted the conserved residues previously reported by Kusano *et al.* (1999)[42] and also included the known catalytic residue lysine 706 (K706) as a positive control. The N20 was positioned closest to the target residue and 80bp donors were designed to change the conserved target residue to every other amino acid. Finally, the smORF deletion library was designed to delete 60bp from the 5′ terminus of each target, including the initiating ATG when possible. SapI sites were added between the guide and the donor sequence that was synthesized by CustomArray to enable downstream cloning of the sgtail and an RNA polIII terminator between these two elements. Finally, all synthesized oligos had BsmBI sites added to each end to enable the first stage of cloning in which the oligo library members were inserted into the pRS426 backbone. Library members containing restriction sites including BsmBI, SapI, NcoI and StuI were excluded from the sequence file and were not synthesized.

Cloning of the library

The CustomArray-synthesized oligo library was diluted to 1 ng/μl and 1μl of the library was amplified with Kapa SYBR FAST qPCR Kit Master Mix (Kapa Biosystems) using unique primer pairs specific to each desired library (e.g. *SGS1* tiling deletion, smORF library, etc.). Primers used for oligo library amplification were further modified to contain the necessary overlaps to enable the library to be inserted into our vector backbone via Golden Gate cloning. The PCR products were run on a gel to confirm amplicons are of the expected length. After PCR purification (Zymo Research), the amplicon is cloned into the BsmBI-containing library vector (XG128) using a standard Golden Gate protocol with BsmBI (NEB R0580S) and T4 ligase (NEB M0202S) then electroporated into 5-alpha Electrocompetent *E.coli* cells (NEB C2989). This ensuing library now contained the guide and donor sequences adjacent to the *SNR52* promoter but was still missing the sgtail and an RNA polIII terminator. To clone in the additional functional components between the guide and donor, we amplified and cloned in the sgtail and terminator sequences following the same Golden Gate cloning method as described above, but this time using SapI (NEB R0569S) and T4 ligase. The resulting Golden Gate reactions were then PCR purified and electroporated into 5-alpha Electrocompetent *E.coli* cells to create a final guide+donor library.

Transformation into yeast

Prior to transformation into yeast, each guide+donor library was double-digested with NcoI (NEB R0193T) and StuI (NEB R0187L), resulting in a linearized vector with a gap within the *URA3* selection marker. Linearized DNA containing the majority of the vector backbone, but lacking a portion of the *URA3* selection marker, was then gel extracted and purified (Zymo Research). To enable the reconstruction of the guide+donor vector within yeast via homologous recombination, a second linear fragment was generated by PCR using

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

primers that annealed to regions flanking the NcoI and StuI restriction sites, creating a PCR fragment with >100bp of overlap homology to the region removed from the guide+donor backbone. Digested DNA and PCR amplicons (1μg each per transformation) were co-transformed into yeast using standard lithium acetate transformation protocol with the addition of dimethyl sulfoxide (DMSO, 10% final concentration) before heat shock and grown on SC-URA plates for 3 days to obtain Ura^+ colonies.

For our initial library pilot experiments (Figures 1d b 1e, and Supplementary Figures 6b and 6c), 500ng of each indicated guide+donor plasmid were pooled together and double-digested with NcoI and StuI. 1μg of the linearized plasmid mix was co-transformed with 1μg of Ura3 PCR fragment (as described above) into Cas9-expressing wildtype and Mms4-inactivated strains in parallel and selected on SC-URA. Ura+ colonies were scraped off plates after 3 days. For HU sensitivity screen, cells were further diluted 1:100 in liquid media that contains no HU or 40mM HU and grown for 2 days. Cells were collected and genomic DNA was extracted for NGS. Two rounds of independent yeast library transformations were performed.

For the HU condition test of the *SGS1* mutant libraries, each library was first transformed into no-Cas9 and Cas9-expressing cells in parallel using the yeast transformation procedures as described above and selected on SC-URA. After 3 days, colonies were scraped off the plates, diluted 1:100 in liquid media that contains no HU or 40mM HU, and grown for 2 days. Cells were then collected and genomic DNA was extracted for NGS. Experiments were done in duplicates.

For the essentiality/non-essentiality test of smORF library, the library was transformed into no-Cas9 expressing cells and Cas9-expressing cells in parallel. Colonies were scraped and diluted 1:100 in liquid media and grown for 2 days. In addition, transformants from the Cas9-expressing cells were also grown in liquid media containing either 100mM HU, 25μg/ml fluconazole, or subject to 37°C for 2 days. Subsequently, cells were collected, genomic DNA extracted, and NGS was performed. All experiments were done in duplicate.

Guide+donor library preparation and sequencing

Genomic DNA was isolated from each yeast sample. Two rounds of PCR were performed using Q5® Hot Start High-Fidelity polymerase (New England Biolabs). The first round amplified each guide+donor with forward (CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNAGTGAAAGATAAATGATC) and reverse primers (GGAGTTCAGACGTGTGCTCTTCCGATCTGCGAATTGGGTACCATGT) hybridizing to common flanking regions. Subsequently, standard Illumina Truseq and/or Nextera barcodes were attached through a second round of PCR amplification. Gel purification was performed on all amplicons to confirm the amplicon size and quality before extracting and purifying the sample using the QIAquick gel extraction kit (Qiagen). DNA libraries for NGS were quantified using the KAPA Library Quantification Kit (Kapa Biosystems). Samples were pooled in equimolar amount. The final library was prepared using standard MiSeq Reagent Kit v2 (2 × 150bp) protocol with 12pM diluted DNA libraries with 15% to 25% PhiX spiked into the mixture and run on an Illumina MiSeq or NextSeq 500 Systems, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Preprocessing of library sequences and count generation

Guide RNA and donor sequences were extracted from R1 and R2 reads, respectively, and mapped to reference library members containing each guide+donor pair using a custom python script. Sequences that do not match to any of the library members were discarded from the analysis. Only sequences that contain the perfectly matched N20, sgtail (GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGTCT) and donor sequences were included in count generation. We first sequenced the plasmid libraries to determine the distribution of sequences. Reads that were severely underrepresented, i.e. less than 30 reads mapped to the guide+donor, were removed from further analysis.

Data analysis and fitness calculation

For all the conditions, the mapped reads were compared against the corresponding control experiment. The control experiment for each *SGS1* library (tiling deletion and amino acid substitutions) was the experiment performed in the absence of HU. A fold change (FC) for each guide+donor is calculated as follows:

$$FC_i = \frac{\frac{test_i}{test\ total_i}}{\frac{control_i}{control\ total_i}}$$

where $test_i$ and $control_i$ are the number of reads that mapped to guide+donor i in all the test conditions and control, respectively. The $test\ total_i$ represents the total number of reads in the test conditions and $control\ total_i$ is the number of reads in the control. If the guide+donor is enriched in the test condition, FC would be >1. If the guide+donor is depleted in the tested condition, FC would be <1. The average $\log_2$FC values of the duplicates and p-values (2-tailed Z-test) corresponding to each tested guide+donor for each library are provided in Supplementary Files 1 and 2.

The smORF library was subjected to four screens: essentiality, heat, HU, and fluconazole. While the control experiment for the last 3 test conditions was conducted in the absence of the environmental stress, the control experiment for essentiality screen was performed in a yeast strain lacking Cas9. The same FC calculation described above was carried out for each guide+donor in the smORF library. Supplementary File 3 lists the average $\log_2$FC value and p-value for each tested guide+donor.

Validation of mutants from the three libraries

Individual Ura+ transformants were picked from each library and grown overnight in 96-well plates. DNA extraction was performed followed by PCR amplification (forward primer TTCGGCGTTCGAAACTTCTCCGCA and reverse primer TAGACCGAGATAGGGTTGAGTG) and sequencing of the guide+donor on the plasmid (TTCGGCGTTCGAAACTTCTCCGCA) to determine programmed edits intended for each transformant. Individual primer pairs specific to the corresponding endogenous site were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

designed. Each endogenous site was amplified and sequenced with the forward primer to determine if the programmed edits as specified by the donor had successfully occurred.

Phenotypic validation of library hits

To validate the hits exhibiting phenotypic sensitivity and lack of sensitivity in each library screen, we picked 2–4 sensitive and 2–4 non-sensitive targets from each screen, constructed the corresponding guide+donor plasmids, and performed similar transformation experiment as described above. Individual transformants were genotyped followed by phenotyping onto the corresponding test conditions to confirm our NGS screening results. For the phenotypic growth assay, cells were grown to log phase. 3μl of each undiluted and 5-fold serially diluted culture were spotted onto SC-URA or SC-URA under tested conditions. All plates were incubated at 30°C for 48 hours and photographed.

Feature examination of hit versus non-hit between smORFs and ORFs

Comparison of target length—The amino acid length for each target in the library was obtained from YeastMine[30]. The distributions of protein sizes between the different groups, namely smORF hits versus non-hits and ORF hits versus non-hits, were compared. To determine if there is a significant difference in amino acid lengths between groups, we performed a 2-tailed t-test (summarized in Supplementary Table 3, Supplementary File 4).

Comparison of gene expression—The FPKM values for the targets were generated as follows. Raw RNA-seq data for BY4741 yeast strain was obtained from SRA (SRR3126113)[43]. FASTX Toolkit (http://www.bioinformatics. babraham.ac.uk/projects/ fastqc/) was used to remove the adapters (fastx_trimmer) and trim the ends of base pairs with a quality score lower than 30 (fastq_quality_trimmer). After quality trimming, the read pairs were intersected using an in-house pipeline. Subsequently, the reads were aligned to the S288C genome (Bioproject) using Tophat[44] and the FPKM vales were generated using Cufflinks[45]. To determine if the expression levels between the different groups were significant, we performed a 2-tailed t-test between the $\log_{10}$ FPKM values between the hits and non-hits of the smORF class and ORF class (summarized in Supplementary Table 4, Supplementary File 4).

Comparison of secondary structure—We mapped the possible presence of secondary structures (alpha-helices and beta-sheets) in each amino acid sequence using PredictProtein[46]. Several comparisons with regards to the overall distribution of secondary structures between different groups were made (summarized in Supplementary Table 5, Supplementary File 4) and examined for significant difference through Kolgomorov-Smirnov test.

Comparison of homolog conservation in human—The corresponding human homologues for the targets in the smORF library were obtained from YeastMine[30]. The number of targets containing a human homologue were counted and compared among different groups (summarized in Supplementary Table 6, Supplementary File 4). A Chi-squared test was used to test for significant difference between the different groups.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Comparison of transformation and editing efficiencies between unmodified and engineered approaches

Thirty-four guide+donor contigs were selected from the smORF library screen and were individually constructed followed by Gibson-cloning into pRS426 backbone as described above. Each smORF-targeting guide+donor plasmid construct was introduced into a Cas9-expressing yeast strain in either the unmodified or the engineered configurations. A similar transformation was also carried out side-by-side in a non-Cas9-expressing yeast strain. Colony counts for each guide+donor transformation were obtained 3 days post-transformation. A fold change in transformation efficiency was calculated based on the colony count generated from the engineered approach divided by the colony count obtained from the unmodified approach. In addition, 5 random colonies from each guide+donor transformation were PCR amplified and Sanger sequenced at the corresponding endogenous site to determine if the correct genomic edit took place. Editing efficiency was determined by the proportion of sequenced transformants with the correct genomic edit over the total number of sequenced transformants. This whole experiment was performed twice. All Sanger sequencing was performed by Genewiz, Inc.

Effect of homology length on more distant SNPs editing

ADE2-targeting guide+donor plasmids with various homology lengths (60bp, 70bp, 80bp, 90bp and 100bp) on the donor sequence to introduce genomic edits of SNPs at different PAM-distant positions (Supplementary Figure 3a) were constructed via Gibson assembly. Each guide+donor construct was transformed in the engineered configuration into both yeast strains expressing and not expressing Cas9 to examine effect of homology length on transformation efficiency. Transformation efficiency was represented by the number of transformants obtained in the presence of Cas9 over the number of transformants obtained in the absence of Cas9. Each transformation experiment was performed twice. A few colonies from each transformation were PCR amplified and Sanger sequenced at the *ADE2* target site to determine the proportion of correct genomic edits.

Whole genome sequencing to detect off-target effects of guide+donor system

Sample preparation—A Cas9-expressing parental yeast strain (YAC2563) and three yeast strains (YXG231, YXG232, YXG234) modified by guide+donor plasmids, ADE2Δ61bp, SGS1Δ60bp, and SGS1ΔATG, respectively, were grown overnight in 5ml YPAD. Genomic DNA was isolated from these cells followed by a PCR purification (Zymo Research) step to clean up the DNA. For library preparation, we used Nextera (Illumina) to fragment the genome. Roughly 35ng of genomic DNA was used for each sample, equivalent to 3 million haploid yeast genomes. After the tagmentation reaction (20μL reaction system, 55°C 15min, 70°C 30min), fragmented DNA was purified with DNA Clean-up & Concentrator-5 (Zymo Research) and used as PCR template (NEBNext® High-Fidelity 2X PCR Master Mix, NEB, 72°C 3min for Tn5 gap filling and end repair, 98°C 30s, 4 cycles of 98°C 10s, 63°C 30s, 72°C 40s, and 72°C 2min for a final extension) to add sequencing adaptors. Amplified library was cleaned up with 0.8× Ampure beads and sequenced with 2x 150bp NextSeq500/550 for a total of 28M paired-end reads.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Computational analysis—The quality of the fastq files was first evaluated using the FASTQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) followed by end trimming using FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) to obtain base pairs with a quality score lower than 30 (fastq_quality_trimmer). After quality trimming, an in-house algorithm was used to intersect the read-pairs. Subsequently, BWA (version 0.6.1-r104) was utilized to align the reads to the S288C genome downloaded from Genbank as assembly GCA_000146045.2. SNPs were detected by SAMtools mpileup and bcftools. A hard filter removing all SNPs/indels below 25% of the median depth was chosen as cutoff. The median depth was deduced using genomeCoverageBed from BEDTools (version 2.16.2) as described by Kaas *et al*[47]. Off-target analysis was carried out using Bowtie (version 0.12.7) to search the yeast genome for the guide RNA sequences corresponding to the guide +donor constructs for up to 2 mismatches. A region of 500bp surrounding each of the potential off-target sites were manually cross-referenced with the list of detected SNP/indels as previously described[48]. The expected genomic changes were manually evaluated from the aligned BAM file in Geneious (Biomatter Ltd).

Statistical analysis

Each figure description indicates the number of independent experiments. A 2-tailed Z-test was used to examine significance of depletion in each library screen in Figures 2 and 3. A 2-tailed Z-test was applied to examine depletion and enrichment in Figure 4. A Chi-squared test was used to assess statistical differences between groups in Supplementary tables 1 and 2. A 2-tailed t-test was to examine statistical significance in Supplementary tables 3 and 4. Kolgomorov-Smirnov and Chi-squared tests were used to assess the statistical differences between groups in Supplementary Tables 5 and 6, respectively.

Data availability

All NGS data generated in this study are available through NCBI SRA. Data used for amino acid length, gene expression, and human conservation comparisons are presented on Supplementary File 4 and summarized in Supplementary Tables 3–6.

Code availability

All custom scripts are available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

G.M.C was supported by NIH grants RM1 HG008525 and P50 HG005550. A.C. was funded by the National Cancer Institute grant no. 5T32CA009216-34. J.J.C was funded by the Defense Threat Reduction Agency grant HDTRA1-14-1-0006, the Paul G. Allen Frontiers Group. Y.Y. was supported by the Damon Runyon Research Foundation grant DRG-2248-16. X.G. and A.C. conceived the idea, led the study, and designed all experiments. A.C. and R.C. with input from J.E.D. demonstrated the initial feasibility of the guide+donor approach. X.G. performed majority of the experiments, including the oligo library design, library construction and analysis with significant technical contribution from A.T. Y.C. provided expertise in statistical analysis. Y.Y performed the whole genome sequencing experiment for off-target analysis. C.K. generated the RNA-seq data for BY4741 yeast strain, provided the FPKM values and analyzed the whole genome data from yeast isolates modified by guide+donor for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

off-target effects. S.L.G. assisted with oligo library design. E.K. provided insight with regard to library construction methods and analysis. M.S. provided technical expertise with regard to methods to increase guide+donor efficiency. J.J.C and G.M.C. oversaw the study. X.G. and A.C. wrote the manuscript with input from all authors.

References

1. Giaever G, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002; 418:387–391. [PubMed: 12140549]
2. Ben-Shitrit T, Yosef N, Shemesh K, Sharan R, Ruppin E, Kupiec M. Systematic identification of gene annotation errors in the widely used yeast mutation collections. Nat Methods. 2012; 9:373–378. [PubMed: 22306811]
3. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 2013; 41:4336–4343. [PubMed: 23460208]
4. Garst AD, et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat Biotechnol. 2017; 35:48–55. [PubMed: 27941803]
5. Bao Z, et al. Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. ACS Synth Biol. 2015; 4:585–594. [PubMed: 25207793]
6. Choulika A, Perrin A, Dujon B, Nicolas JF. Induction of homologous recombination in mammalian chromosomes by using the *I-SceI* system of *Saccharomyces cerevisiae*. Mol Cell Biol. 1995; 15:1968–1973. [PubMed: 7891691]
7. Brenneman M, Gimble FS, Wilson JH. Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases. Proc Natl Acad Sci USA. 1996; 93:3608–3612. [PubMed: 8622983]
8. Donoho G, Jasin M, Berg P. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Mol Cell Biol. 1998; 18:4070–4078. [PubMed: 9632791]
9. Smih F, Rouet P, Romanienko PJ, Jasin M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. Nucleic Acids Res. 1995; 23:5012–9. [PubMed: 8559659]
10. Taghian DG, Nickoloff JA. Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells. Mol Cell Biol. 1997; 17:6386–93. [PubMed: 9343400]
11. Lu J, Mullen JR, Brill SJ, Kleff S, Romeo AM, Sternglanz R. Human homologues of yeast helicase. Nature. 1996; 383:678–679. [PubMed: 8878475]
12. Ira G, Malkova A, Liberi G, Foiani M, Harber JE. Srs2 and Sgs1-Top3 suppress crossovers during double-strand break repair in yeast. Cell. 2003; 115:401–411. [PubMed: 14622595]
13. Miyajima A, et al. Different domains of Sgs1 are required for mitotic and meiotic functions. Genes and Genet Syst. 2000; 75:319–26. [PubMed: 11280006]
14. Mullen JR, Kaliraman V, Brill SJ. Bipartite structure of the SGS1 DNA helicase in Saccharomyces cerevisiae. Genetics. 2000; 154:1101–4. [PubMed: 10757756]
15. Weinstein J, Rothstein R. The genetic consequences of ablating helicase activity and the Top3 interaction domain of Sgs1. DNA Repair. 2008; 7:558–71. [PubMed: 18272435]
16. Boddy MN, Lopez-Girona A, Shanahan P, Interthal H, Heyer WD, Russell P. Damage tolerance protein Mus81 associates with the FHA1 domain of checkpoint kinase Cds1. Mol Cell Biol. 2000; 20:8758–66. [PubMed: 11073977]
17. Mullen JR, Kaliraman V, Ibrahim SS, Brill SK. Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in Saccharomyces cerevisiae. Genetics. 2001; 157:103–18. [PubMed: 11139495]
18. Chu WK, Hickson ID. RecQ helicases: multifunctional genome caretakers. Nat Rev Cancer. 2009; 9:644–54. [PubMed: 19657341]
19. Gangloff S, McDonald JP, Bendixen C, Arthur L, Rothstein R. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: A potential eukaryotic reverse gyrase. Mol Cell Biol. 1994; 14:8391–8398. [PubMed: 7969174]
20. Bennett RJ, Sharp JA, Wang JC. Purification and characterization of the Sgs1 DNA helicase activity of Saccharomyces cerevisiae. J Biol Chem. 1998; 273:9644–50. [PubMed: 9545297]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

21. Ui A, et al. The ability of Sgs1 to interact with DNA topoisomerase III is essential for damage-induced recombination. DNA Repair. 2005; 4:191–201. [PubMed: 15590327]
22. Kennedy JA, Daughdrill GW, Schmidt KH. A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1 Nucleic Acids Res. 2013; 41:10215–27.
23. Bennett RJ, Noirot-Gros MF, Wang JC. Interaction between yeast Sgs1 helicase and DNA topoisomerase III J Biol Chem. 2000; 275:26898–26905. [PubMed: 10862619]
24. Duno M, Thomsen B, Westergaard O, Krejci L, Bendixen C. Genetic analysis of the Saccharomyces cerevisiae Sgs1 helicase defines an essential function for the Sgs1-Top3 complex in the absence of SRS2 or TOP1. Mol Gen Genet. 2000; 264:89–97. [PubMed: 11016837]
25. Fricke WM, Kaliraman V, Brill SJ. Mapping the DNA topoisomerase III binding domain of the Sgs1 DNA helicase. J Biol Chem. 2001; 276:8848–8855. [PubMed: 11124263]
26. Ui A, Satoh Y, Onoda F, Miyajima A, Seki M, Enomoto T. The N-terminal region of Sgs1, which interacts with Top3, is required for complementation of MMS sensitivity and suppression of hyper-recombination in sgs1 disruptants. Mol Genet Genomics. 2001; 265:837–850. [PubMed: 11523801]
27. Onodera R, Seki M, Ui A, Satoh Y, Miyajima A, Onoda F, Enomoto T. Functional and physical interaction between Sgs1 and Top3 and Sgs1-independent function of Top3 in DNA recombination repair. Genes Genetic Syst. 2002; 77:11–21.
28. Cejka P, Plank JL, Bachrati CZ, Hickson ID, Kowalczykowski SC. Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. Nat Struct Mol Biol. 2010; 17:1377–1382. [PubMed: 20935631]
29. Kastenmayer JP, et al. Functional genomics of genes with small open reading frames (sORFs) in S cerevisiae. Genome Res. 2006; 16:365–373. [PubMed: 16510898]
30. Balakrishnan, R., et al. Database. (Oxford); 2012. YeastMine-an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit.
31. Chen X, Zhang J. The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. Cell Syst. 2016; 2:347–54. [PubMed: 27185547]
32. Horwitz AA, et al. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. Cell Syst. 2015; 1:88–96. [PubMed: 27135688]
33. Munoz DM, et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive his for highly amplified genomic regions. Cancer Discov. 2016; 6:900–13. [PubMed: 27260157]
34. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat Biotechnol. 2015; 33:661–7. [PubMed: 25961408]
35. Michel AH, et al. Functional mapping of yeast genomes by saturated transposition. Elife. 2017; 6:e23570. [PubMed: 28481201]
36. Fowler DM, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010; 7:741–746. [PubMed: 20711194]
37. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014; 11:801–807. [PubMed: 25075907]
38. Starita LM, et al. Variant interpretation: functional assays to the rescue. Am J Hum Genet. 2017; 101:315–325. [PubMed: 28886340]
39. Weile J, et al. A framework for exhaustively mapping functional missense variants. Mol Syst Bio. 2010; 13(12):957.
40. Kleinstiver BP, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015; 523:481–5. [PubMed: 26098369]
41. Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. Multifunctional yeast high-copy-number shuttle vectors. Gene. 1992; 110:119–22. [PubMed: 1544568]
42. Kusano K, Berres ME, Engels WR. Evolution of the RECQ family of helicases: a Drosophila Homolog, Dmblm, is similar to the human Bloom Syndrome gene. Genetics. 1999; 151:3.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

43. Yao W, et al. The INO80 complex requires the Arp5-les6 subcomplex for chromatin remodeling and metabolic regulation. Mol Cell Biol. 2016; 36:979–91. [PubMed: 26755556]

44. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2012; 25:1105–11.

45. Trapnell C, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010:511–515. [PubMed: 20436464]

46. Rost B, Yachdav G, Liu J. The PredictProtein server. Nucleic Acids Res. 2004; 32:W321–W326. Web Server issue. [PubMed: 15215403]

47. Kaas CS, Kristensen C, Betenbaugh MJ, Andersend MR. Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. BMC Genomics. 2015; 16:160. [PubMed: 25887056]

48. Paix A, et al. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 sites in Caenorhabditis elegans. Genetics. 2014; 198:1347–56. [PubMed: 25249454]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

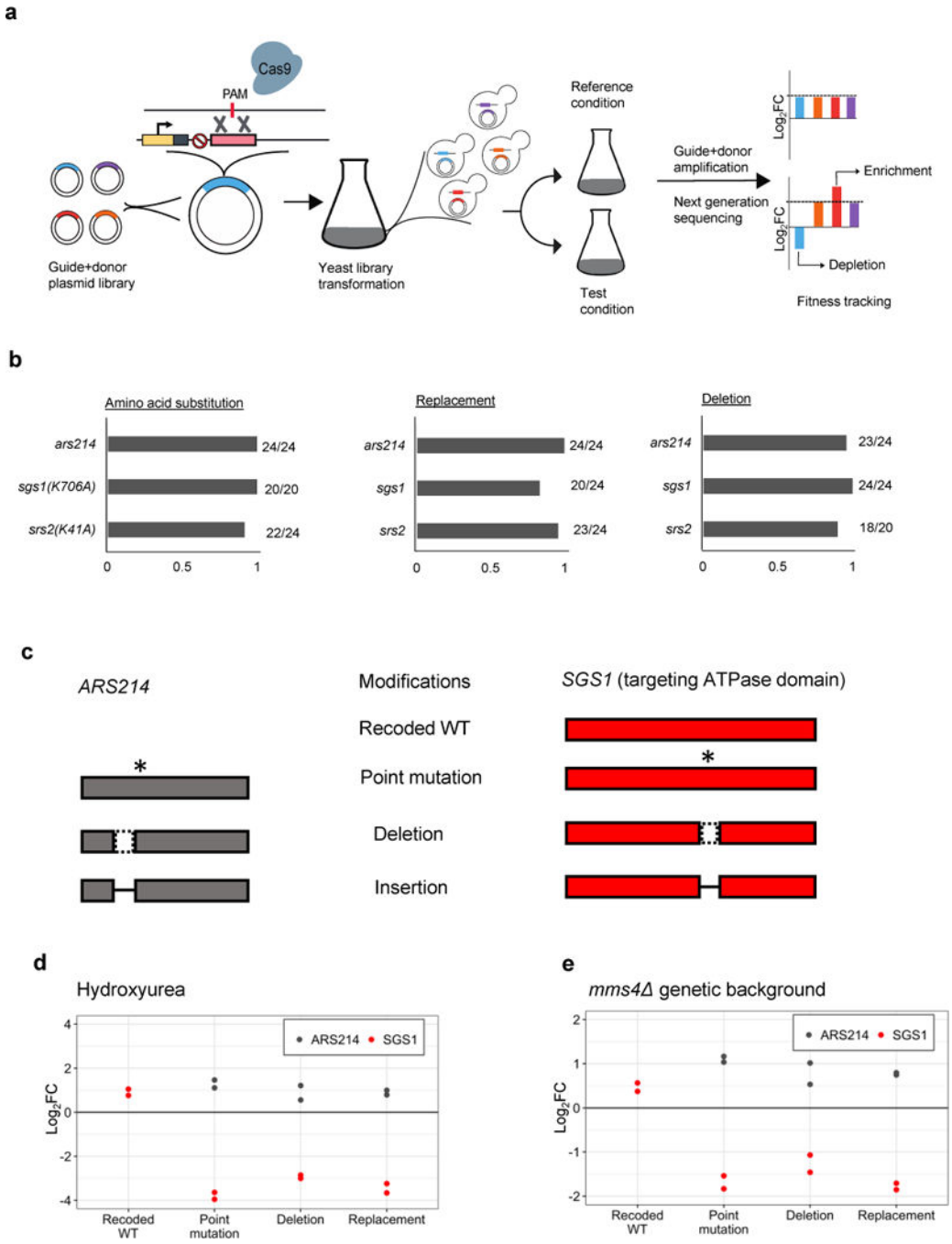

Figure 1.

Guide+donor genome-editing platform for engineering and phenotypically characterizing programmed mutations in pool. **a** Illustration of guide+donor workflow. Guide+donors targeting different genomic sites-of-interest are marked by different colors. Each guide +donor structure contains an *SNR52* promoter (yellow), an N20 sequence (dark grey), a structural sgtail (not shown), a terminator sequence (circle-backslash symbol), and a donor template with the desired mutations flanked by regions of homology (red). Pool of transformants is subject to reference and test conditions simultaneously, genomic DNA

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

extraction, and next generation sequencing of the guide+donor amplicons to determine depletion and enrichment of guide+donor targets. **b** Bar graph depicting editing efficiencies for creating programmed amino acid substitution, deletion, and sequence replacement at three endogenous sites (*ARS214*, *SGS1*, and *SRS2*). Catalytic amino acid substitutions for *SGS1* and *SRS2* and proportion of correct edits are indicated. **c** Graphical representation of guide+donor-generated *ARS214* (grey) and *SGS1* (red) variants followed by phenotypic testing in **d** and **e**. Asterisk, dotted box, and solid dash denote substitution, deletion, and replacement of an amino acid stretch with a linker sequence, respectively. Figures not drawn to scale. **d** Plot showing HU response of a guide+donor library of *ARS214* and *SGS1* mutants. X- and y-axes correspond to programmed edits encoded in the guide+donor constructs and $\log_2$ fold change, respectively. Two independent yeast library transformations were performed. **e** Dot plot displaying sensitivity of *ARS214* and *SGS1* mutants in *mms4Δ* genetic background. Genetic modifications and $\log_2$ fold change are exhibited on x- and y-axes, respectively. Two independent library transformations were performed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

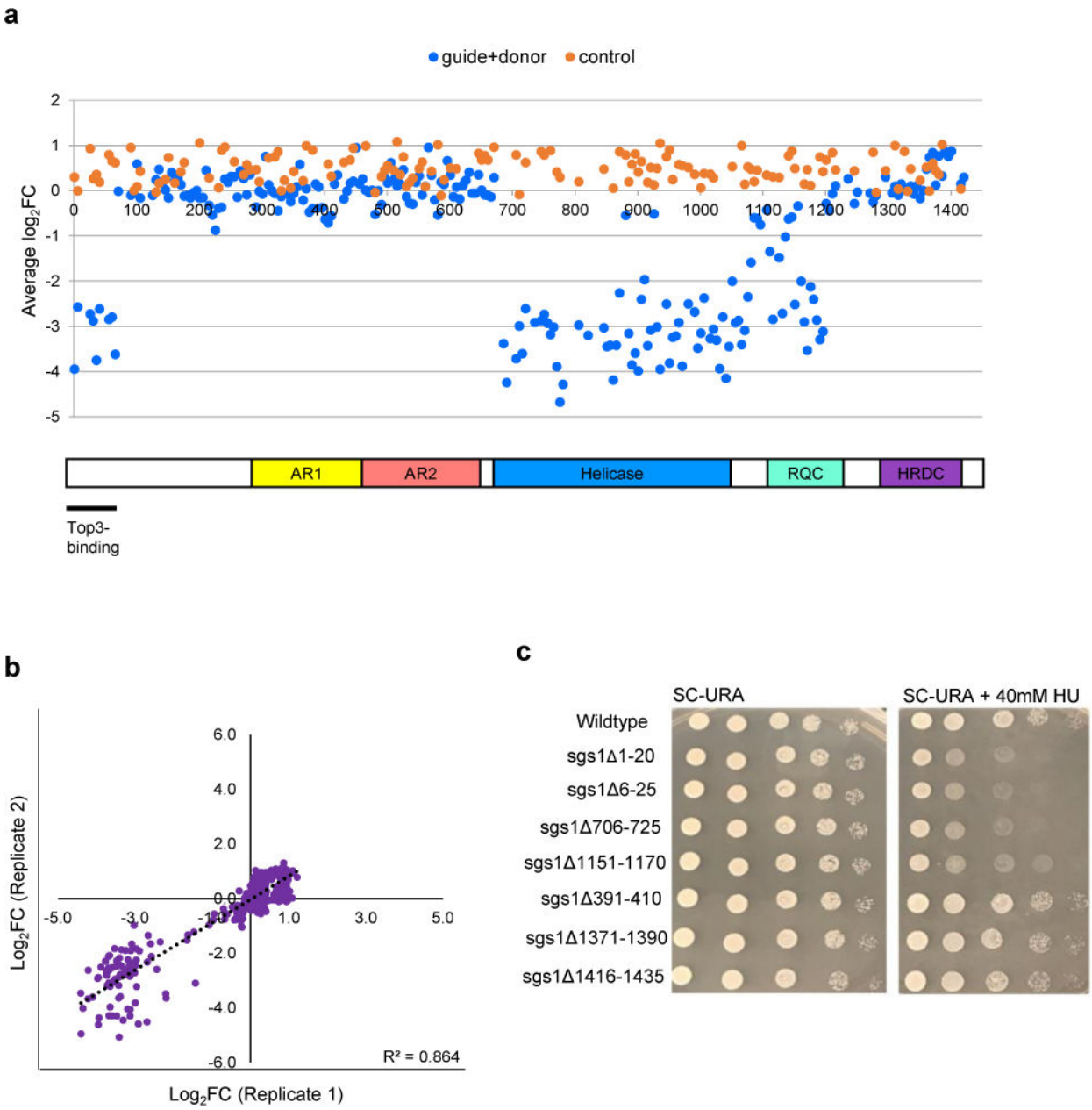

Figure 2.

Guide+donor library of *sgs1* mutants in response to HU. **a** Sgs1 tiling deletion screen. Scatterplot showing average $\log_2$ fold change in abundance of guide+donor members programmed to generate *sgs1* tiling deletion mutants across the entire *SGS1* gene in response to HU (n=2 independent yeast library transformations). Guides paired with corresponding donor sequences to generate programmed deletions are indicated in blue. Non-targeting control guides paired with sequence that lack homology regions to qualify as donors are used as controls and are marked orange. X- and y-axes denote the amino acid window along the protein and average $\log_2$ fold depletion, respectively. Schematic representation of relevant domains and motifs in Sgs1 is shown. Figures not drawn to scale. **b** Replicate analysis of $\log_2$ fold changes between two independent yeast library

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

transformations. Pearson correlation coefficient is indicated. **c** Phenotypic validation of selected sensitive and non-sensitive *sgs1* truncation mutants from the HU library screen in **a**. See Materials and Methods for details.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

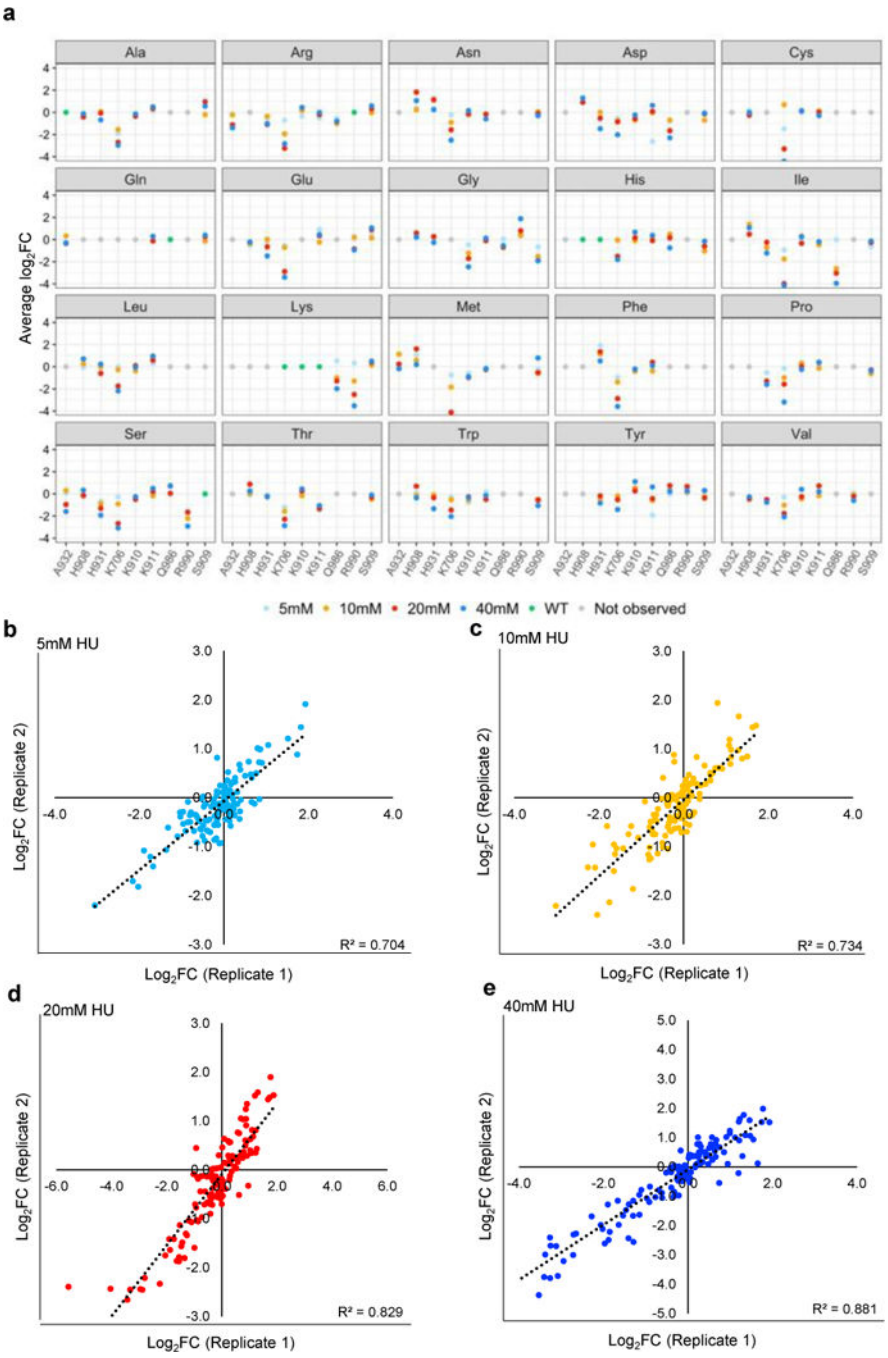

Figure 3.

Guide+donor library of amino acid substitutions of selected conserved residues in *SGS1* in response to various concentrations of HU. **a** Sgs1 amino acid residue substitution screen. Scatterplots showing average $\log_2$ fold change in abundance of guide+donor members programmed to generate precise point mutations within Sgs1 in response to HU (n=2 independent yeast library transformations). Concentrations of HU are represented by different colors and described in the legend. Selected conserved residues and average $\log_2$ fold depletion are displayed on the x- and y-axes, respectively. Each subplot shows the

corresponding amino acid to which each conserved residue was replaced. **b** Replicate analyses showing Pearson correlation of $\log_2$ fold changes between two independent yeast transformations under various drug concentrations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

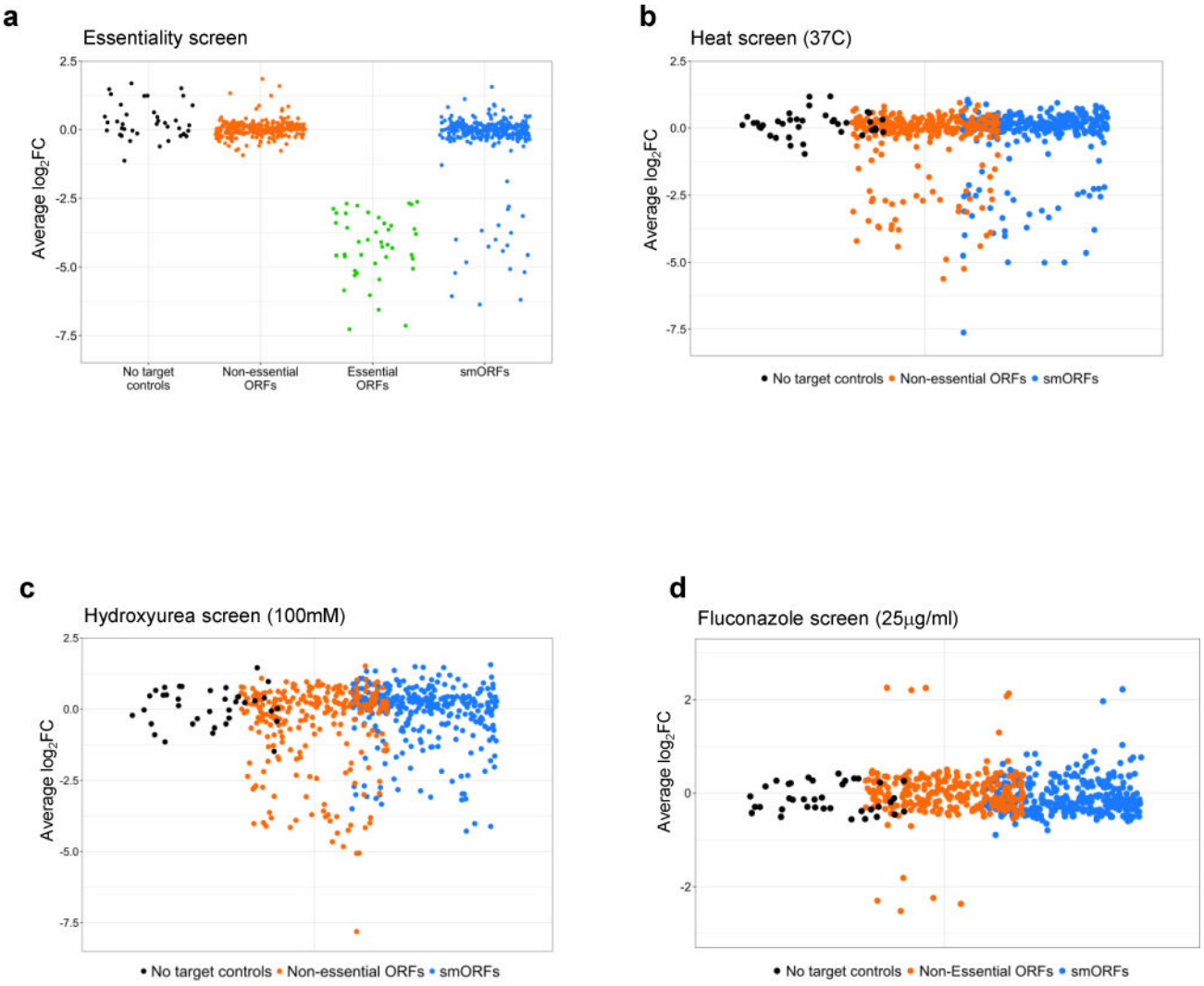

Figure 4.

smORF mutant library subject to different phenotypic screens (**a**–**d**). Two independent yeast transformations were performed for each library and subjected to different test conditions as indicated on each subplot. Shown are average $\log_2$ fold changes of guide+donor constructs in each test condition as compared to guide+donor constructs in control condition. Control guide+donors are marked in black. Guide+donors targeting essential genes, non-essential genes, and smORFs are marked in green, orange, and blue, respectively.