Highly efficient CRISPR gene editing in yeast enabled by double selection

Philippe C Després[1,2,3], Alexandre K Dubé[1,2,3,4], Lou Nielly-Thibault[2,3,4], Nozomu Yachie[5,6,7,8] and Christian R Landry*[1,2,3,4]

1. Département de Biochimie, Microbiologie et Bio-informatique, Faculté de sciences et genie, Université Laval, Québec, Québec, G1V 0A6, Canada

2. PROTEO, The Québec Research Network on Protein Function, Structure and Engineering, Université Laval, Québec, Québec, G1V 0A6, Canada

3. Centre de Recherche en Données Massives (CRDM), Université Laval, Québec, Québec, G1V 0A6, Canada

4. Département de Biologie, Faculté de sciences et Génie, Université Laval, Québec, Québec, G1V 0A6, Canada

5. Research Center for Advanced Science and Technology, Synthetic Biology Division, University of Tokyo, Tokyo, 4-6-1 Komaba, Meguro-ku, 153-8904, Japan

6. Department of Biological Sciences, Graduate School of Science, the University of Tokyo, Tokyo, Japan

7. Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan

8. PRESTO, Japan Science and Technology Agency, Tokyo, Japan

*To whom correspondence should be addressed. Tel: 1-418-656-3954, Fax 1-418-656-7176, christian.landry@bio.ulaval.ca

Abstract

CRISPR-Cas9 loss of function (LOF) screens are powerful systems biology tools that allow genome-wide perturbation studies. While currently possible in many model species, such an approach has yet to be developed in yeast because of low mutagenesis rates without donor DNA for homology directed repair. Here we develop a double selection strategy based on co-selection to dramatically increase LOF mutation rates, both for CRISPR-Cas9 and the Target-AID base editor. We constructed the pDYSCKO vector, which is amenable to high throughput double selection both for LOF by CRISPR-Cas9 and Target-AID mutagenesis. We develop and use a model of CRISPR-LOF barcode sequencing competition screens to show that this improvement provides an important increase in statistical power when attempting to detect the fitness effects of mutations. We also show that multiplex genome editing with Cas9 in yeast causes programmable chromosomal translocations at high frequency, suggesting that multiplex editing should be performed with caution. This tool will allow the development of new *S. cerevisiae* systems biology workflows and should prove compatible with most resources already available for yeast.

Introduction

Saccharomyces cerevisiae is one of the foremost model organisms in biology. The yeast genome is uniquely well suited for manipulation and comprehensive analyses. Current resources and techniques such as the gene KO collection have enabled many important discoveries (1), but still have many shortcomings. These include time and resource costs and the need for robotic platforms to perform some high throughput experiments, which can limit their use in a systems biology context. In addition, the yeast genome-wide collections contain anomalous strains bearing secondary mutations or aneuploidies that can bias experimental results at their inception (2). Furthermore, because the sharing of strain collections has led to these collections being distributed from lab to lab, the strains that end up used at the bench have had time to accumulate secondary mutations that may also bias results and affect reproducibility between labs. Being able to regenerate deletion strains *de novo* at will would be an important technological development for the yeast community.

Recently, the CRISPR-Cas9 genome editing system has been engineered to create a comprehensive experimental toolkit (3–5). One of the principal applications of the system is for gene loss of function (LOF) via DNA repair errors within the coding sequence of a gene. The scalability of this technique allowed for the development of CRISPR-cas9 LOF experiments at the genome scale (6–8), greatly facilitating systems biology experiments in models in which they were difficult so far. Despite yeast being a usual frontrunner in technological developments for systems biology, low CRISPR-Cas9 LOF efficiency in yeast (9) has so far made the application of this approach impractical. The power of the numerous resources already available for systems biology in yeast

would be greatly enhanced by the efficiency and cost-effectiveness of these new methods. For instance, it would allow for the generation of a fresh pool of LOF strains at the start of each experiment, minimizing secondary mutation accumulation, or even reconstituting a gene KO collection in a diversity of genetic backgrounds. Newly developed base editors based on nCAs9 and dCas9 fusions, such as Target-AID (10), which induces C to G and C to T changes, would also make important additions to the yeast systems biology toolkit if they could be used with high mutagenesis efficiency.

One of the crucial determinant of large-scale LOF screens is mutagenesis efficiency. Accordingly, different approaches have been used to optimize mutagenesis rate. One of them is co-selection, which relies on the selection of cells based on a marker that has mutated alongside the target loci based on a positively or negatively selected marker. This was recently shown to enhance edition rates in human cell lines (11) and drosophila (12), but has not been applied to yeast yet. Because the base efficiency of CRISPR-Cas9 KO in these model organisms is higher than in yeast by several orders of magnitudes (9), the gain of efficiency brought by co-selection in yeast would be ever greater than what it is in these other models. Finally, co-selection has also yet to be shown to increase Target-AID mutagenesis efficiency.

Here we present a high efficiency co-selection approach to easily produce LOF mutants in yeast with both CRISPR-Cas9 and Target-AID using vectors amenable to genome-wide screens. It is intuitive that a higher efficiency would allow for more efficient screening as well as the detection of more subtle effects on fitness. However, it is not clear by how much we need to improve editing efficiency to be able to perform

experiments in yeast using CRISPR-Cas9 LOF or base editing so that they would compare in terms of power with experiments currently performed with the yeast deletion collection. We therefore developed a statistical model to assess quantitatively the improvement brought by double selection on CRISPR-LOF screens and use it to show that is highly significant, particularly for fitness effects in the range in which most of gene deletion effects are observed across a range of growth conditions.

Methods

Construction of the pDYSCKO vector, pDYSCKO-*ADE1* and pDYSCKO-*VPS35*

The pDYSCKO vector was built from p426-SNR52p-gRNA.CAN1.Y-SUP4t (Addgene # 43803). Directed mutagenesis was used to create silent mutations in the AmpR gene and the *URA3* gene to remove BsaI restriction sites using the oligonucleotides mut_AmpR_For/mut_AmpR_Rev and mut_URA3_For/mut_URA3_Rev. BsaI site deletion was confirmed by analyzing the BsaI digestion profile (all oligonucleotides used in this study are presented in Table S1).

The DYSCKO cassette was synthetized as a gBLOCK (Integrated DNA Technologies, Coralville, USA) and Gibson cloned into the modified backbone amplified with Backbone_For and Backbone_Rev. Clones were confirmed by colony PCR using the DYSCKO_for and DYSCKO_rev oligonucleotides.

The g.ADE1 ssDNA oligonucleotide containing the gRNA targeting the *ADE1* gene was amplified by PCR using the gRNA_for and gRNA rev oligonucleotides. We chose this guide because it had previously been used by Nishida et al (10) to create LOFs in *ADE1*

using Target-AID, and was well suited to CRISPR-Cas9 gene disruption. Multiple amplification reactions were pooled and purified using the EZ-10 Column PCR Products Purification Kit (Biobasic, Markham, Canada). Fragment concentration was estimated using a NanoDrop (Thermofisher, Waltham, USA). The insert was cloned into the pDYSCKO vector using the Golden Gate Assembly Mix (New England Biolabs, Ipswich, USA) with the following parameters: 1 ul pDYSCKO vector (50 ng/ul), 1 ul insert (0.5 ng/ul) in the standard manufacturer 20 ul assembly reaction conditions, which was incubated for 1 hour at 37°C followed by 5 minutes at 55°C. The assembled plasmids were then transformed in bacteria and positive clones were confirmed by Sanger sequencing (CHUL sequencing platform, Québec, Canada) using the DYSCKO_rev oligonucleotide. The same process was used to assemble pDYSCKO-*VPS35*, but using the g.VPS35 ssDNA oligonucleotide.

Strains and plasmids used

All experiments were performed in the haploid *S. cerevisiae* BY4741 strain (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) or diploid BY4743 (MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0) from (13). The plasmid p415-Cas9 was obtained from Addgene (#43804). The nCas9-Target-AID plasmid was a generous gift from Dr Keiji Nishida, Kobe University, Kobe, Japan.

Media and growth conditions

Media recipes are detailed in Table S2. All growth occurred at 30°C and with shaking in the case of liquid cultures.

Cell transformations

Competent cells and transformations were performed using the standard PEG yeast transformation protocol (14), with a two-hour recovery period. Cells were plated on SC-UL and allowed to grow for 48 hours at 30°C before the start of mutagenesis experiments.

CRISPR-Cas9 and TARGET-AID mutagenesis

Multiple transformation colonies were used to inoculate a 3 ml SC-UL + 2% glucose culture in which cells were grown for 24 hours. Enough cells to inoculate at 1 OD_{600} a 3ml culture were then harvested by centrifugation and placed in SC-UL+5% glycerol for 24 hours. Cas9 or Target-AID expression was then induced by spinning down the cells and switching their media to 3 ml SC-UL+ 5% galactose for a 12-hour period. Cells were then diluted to OD of 0.1 in 3 ml SC-ULR +2% glucose + Canavanine (50 µg/ml) for a 16-hour double selection phase or 3 ml SC-UL +2% glucose at the same OD as the control condition.

Estimation of mutation rate

Mutation rates were assessed by plating cells on SC-ULR + Can and SC-ULR after galactose induction and after recovery with or without double selection at appropriate dilutions. Mutation rate was assessed by calculating the ratio of red or pink colonies over total number of colonies. The *ADE1* and *CAN1* alleles of 8 white and 8 red colonies were sequenced to confirm mutations either through Cas9 or Target-AID with CAN1_for, CAN1_rev and ADE1_for, ADE1_rev. To sequence the breakpoint of the fusion gene present in the pink colonies, the CAN1_For and ADE1_Rev oligonucleotide

combination was used to amplify the *CAN1-ADE1* breakpoint and the resulting amplicon was Sanger sequenced (CHUL sequencing platform, Québec, Canada).

Confirmation of chromosomal rearrangements

Chromosomal rearrangements of three pink colonies from independent mutagenesis experiments were confirmed by PFGE using a modified protocol based on Maringele and Lyndall(15), with BY4741, two white and two red colonies as controls. Plugs were prepared using cells harvested from overnight cultures in YPD for BY4741 and SC-ULR + canavanine for mutant strains. Migration time was 27 hours, with a switch time of 60 seconds and no ramping.

CAN1-VPS35 fusion detection

BY4741 cells were transformed with pDYSCKO-*VPS35* and p415-Cas9 as described above and went through the same mutagenesis protocol. DNA was extracted from Canavanine resistant colonies using the Yeast Quick DNA method(16). Fusions were detected by PCR using the VPS35-A and CAN1_rev oligonucleotides, and the resulting amplicons were Sanger sequenced (CHUL sequencing platform, Québec, Canada).

Sequence Alignments

Sequence alignments were performed using the MEGA 7 software (17).

CRISPR Barcode sequencing simulation

CRISPR-LOF screens have become important tools in systems biology. In these screens, populations of cells are transformed with pools of vectors bearing different gRNAs. The vectors are built in such a way that oligonucleotides can be used to amplify

the gRNA by PCR so it can be used as a unique barcode in a barcode-sequencing competition experiment (Bar-seq) (18, 19). Because each guide targets a unique locus, Bar-seq of the gRNA links LOF at the target locus with cellular fitness in the screen condition. We can model this process by adapting classical haploid selection models, with the added challenge that not all cells with a specific guide will bear a LOF at the target locus because the mutation rate will be below one.

Let n_{WT} and n_{BC} be respectively the abundance of a control guide and the abundance of a guide targeting a gene which when inactivated has a selection coefficient S_{mut}. This mutation is present in this barcoded population in a proportion μ, which represents the mutation rate after induction and recovery and thus the frequency at the beginning of the competition experiment. After a time t, the abundance ratio of these two barcodes is described by

$$\frac{n_{WT\,(t)}}{n_{BC\,(t)}} = \frac{n_{WT\,0}}{n_{BC\,0}} \cdot \frac{W_{WT\,(t)}}{W_{BC\,(t)}}$$

Were $W_{WT\,(t)}$ is the absolute fitness of the control barcode population at time t and $W_{BC\,(t)}$ the absolute fitness of the barcode associated with the mutation. Because the rate of mutation is not 1 for the strains harboring the barcode ($\mu < 1$), a certain proportion of cells ($\mu - 1$) will have the same fitness as cells bearing the control barcode so that:

$$W_{BC\,(t)} = \mu \cdot W_{mut\,(t)} + (1 - \mu) \cdot W_{WT\,(t)}$$

So that

$$\frac{n_{WT\ (t)}}{n_{BC\ (t)}} = \frac{n_{WT\ 0}}{n_{BC\ 0}} \cdot \frac{W_{WT\ (t)}}{\mu \cdot W_{mut\ (t)} + (1-\mu) \cdot W_{WT\ (t)}}$$

By dividing both the numerator and the denominator of the rightmost factor by $W_{WT\ (t)}$, we obtain

$$\frac{n_{WT\ (t)}}{n_{BC\ (t)}} = \frac{n_{WT\ 0}}{n_{BC\ 0}} \cdot \frac{1}{\mu \cdot \frac{W_{mut\ (t)}}{W_{WT\ (t)}} + (1-\mu)}$$

And then to

$$\frac{n_{WT\ (t)}}{n_{BC\ (t)}} = \frac{n_{WT\ 0}}{n_{BC\ 0}} \cdot \frac{1}{\mu \cdot \frac{W_{mut\ (t)}}{W_{WT\ (t)}} - \mu + 1}$$

Because we assume that the growth rates of the mutants and wild-type are constant over time:

$$\frac{W_{mut\ (t)}}{W_{WT\ (t)}} = \left(\frac{W_{mut\ 1}}{W_{WT\ 1}}\right)^t$$

Given that the selection coefficient S_{mut} is $(1 - \frac{W_{mut\ 1}}{W_{WT\ 1}})$ (20) we then have

$$\frac{W_{mut\ (t)}}{W_{WT\ (t)}} = (1 - S_{mut})^t$$

And therefore

$$\frac{n_{WT\ t}}{n_{BC\ t}} = \frac{n_{WT\ 0}}{n_{BC\ 0}} \cdot \frac{1}{\mu \cdot (1 - S_{mut})^t - \mu + 1}$$

Which is simplifies to

$$\frac{n_{WT\ t}}{n_{BC\ t}} = \frac{n_{WT\ 0}}{n_{BC\ 0}} \cdot \frac{1}{\mu \cdot ((1 - S_{mut})^t - 1) + 1}$$

If we assume that the control barcode and the mutant are present in the same proportion at the start of the experiment, then $\frac{n_{WT\ 0}}{n_{BC\ 0}} = 1$ and

$$\frac{n_{WT\ t}}{n_{BC\ t}} = \frac{1}{\mu \cdot ((1 - S_{mut})^t - 1) + 1}$$

This gives us a theoretical population ratio at time t for the control barcode over the barcode for a mutant, as a function of the mutagenesis rate and the selection coefficient over a generation. Because we observe these ratios by sequencing the barcode pool at a given depth, the number of reads for each barcode will be influenced by stochastic sampling. If we have k guides in the experiment, and sequence at an estimated depth of d reads on average per guide, and we assume all barcodes have equivalent abundance at the start, then the number of reads R sequenced for a barcode at $t = 0$ follows

$$R \sim B\left(k \cdot d, \frac{1}{k}\right)$$

Because of our previous assumptions, this distribution is valid at $t = 0$ for all barcodes regardless of μ or S_{mut}. For other values of t, if we assume that the distribution remains equivalent for the controls barcodes, then R for a LOF barcoded strain will follow a binomial distribution influenced by the abundance change caused by the LOF of the target allele:

$$R_{MutBC\,(t)} \sim B\left(k \cdot d, \frac{1}{k} \cdot \frac{n_{BC\ (t)}}{n_{WT\ (t)}}\right)$$

Where

$$\frac{n_{WT\ (t)}}{n_{BC\ (t)}} = \frac{1}{\mu \cdot ((1 - S_{mut})^t - 1) + 1}$$

We can therefore obtain simulated sequencing read number distributions as a function of μ, S_{mut}, t, k, and d. To assess whether a difference in read count for a barcode is significant, we can test for equality of proportion between a control barcode and the barcode of interest using the chi-square test, with the following table:

$R_{MutBC\ 0}$	$R_{MutBC\ (t)}$
$R_{wt\ 0}$	$R_{wt\ (t)}$

In CRISPR LOF barcode sequencing studies, there are multiple control guides that are assumed to be all comparable to the wild-type. If their abundance is on average equal in the screen, we can pool all controls in the same category. Most of the time, there are also multiple guides targeting the gene. However, because Target-AID can be used to create specific point mutations that may not all be equivalent to LOF and may not have the same efficiency, it is useful to consider all guides individually. If we set the number of guide per target loci to g, the number of target loci to t, and the number of control c, then the total sequencing depth for a set d changes because:

$$k_p = (g \cdot t) + c$$

Because we expect all controls to be equivalent, then we can assume that

$$R_{ctrl\ 0} \sim B\left(k_p \cdot d, \frac{c}{k_p}\right),\ R_{ctrl\ (t)} \sim B\left(k_p \cdot d, \frac{c}{k_p}\right)$$

And for a single target

$$R_{target\ 0} \sim B\left(k_p \cdot d, \frac{1}{k_p}\right),\ R_{target\ (t)} \sim B\left(k_p \cdot d, \frac{1}{k_p} \cdot \frac{n_{BC\ (t)}}{n_{WT\ (t)}}\right)$$

Using this model, we can then simulate a CRISPR-LOF screen by specifying parameters analogous to a Bar-seq screen in yeast. We draw simultaneously a sample from each distribution and test the values for significance over 10 000 iterations. The relationship between the fraction of tests that report the fitness effect as significant as a function of the fitness effect and the mutation rate can thus be examined. This was done for 100 increments of μ from 0 to 1 and 100 increments of S_{mut} from 0 to 1.

The yeast genome contains ~6,000 annotated ORFs, of which 1,156 are essential according to data from the yeast deletion collection and as such might not be interesting for a standard LOF screen. Control loci could include intron sequences, pseudogenes or predicted putative non-functional peptides, or often used controls like the HO nuclease. Therefore, to simulate a CRISPR-LOF screen in yeast, the following parameters were used $t = 4{,}800$ (number of LOF target loci), $c = 800$ (number of control guides), $g = 8$ (number of guides per target loci), $d = 200$ (expected average coverage per guide) and $t = 26$ (number of generations). The CRISPR-Cas9 LOF simulation was performed using a script written in python 2.7 using the Ipython/Jupyter interactive environment

(v4.1.1) (21), Numpy (v1.13.3) (22), Pandas (v0.20.3) (23), Matplotlib (v2.1.0) (24) and SciPy (v0.13.3) (25).

Fitness effects of gene deletions in Bar-seq experiments were retrieved from Qian et al (26). All estimates of selection coefficients from all conditions and replicates from the study were pooled. The number of generations in the study matches the number used in our simulation for most gene KO strains.

Results

pDYSCKO is a polyvalent vector for efficient gene disruption in yeast by double selection

Other study systems have shown that the activity of gene editing components could be a limiting factor in mutagenesis such that the selection of cells in which it is active could be used to enhance the proportion of mutated cells. We developed a double selection system for yeast to exploit this principle. The double selection system uses the pDYSCKO (Double Yeast Selection CRISPR-KO) plasmid, which encodes for two guide RNAs (gRNA), one targeting the gene of interest (gYFG) and the other the negative selection market *CAN1* (**Figure 1**). Another vector is used for galactose induction of the effector enzyme, which can be different variants of Cas9. LOF of *CAN1* confers resistance to the antibiotic Canavanine, which means that Canavanine media can be used to enrich the cell population for mutant cells. We estimated mutation rates using a gRNA targeting the *ADE1* gene, which leads to a red colony phenotype when inactivated in yeast, allowing for direct estimates based on colony counts.

We find that *ADE1* LOF rate is far greater for both CRISPR-Cas9 and Target-AID (**Figure 2**) when selected using the *CAN1* marker compared to non-selective media. The baseline KO efficiency without double selection compares to the initial report by DiCarlo et al (9) of mutation rates averaging 0.07% or lower and double selection greatly improves this rate. The median mutation rate for CRISPR-Cas9 shows the most dramatic improvement, with a 1200-fold gain in final mutation rate (median mutation rate: 62%). Target-AID is also improved, with median fold increase of 3 (median mutation rate: 61%,). The improvement is relatively modest but the basal rate is already relatively high, which brings the LOF well above 50%. The higher basal mutation rate is most likely is due to lower cell toxicity. This 3-fold increase appears to be critical for the efficiency of large-scale experiments. In addition, we note that the effective edition rate is most likely underestimated. Target-AID performs C to T and C to G edition (with rare C to A) (REF) and we find that white colonies after recovery systematically show C to T mutations (**figure 2b**), making them silent changed with respect to the phenotype. The 60% LOF rate could therefore correspond to about 100% editing rate.

In some context it could be useful to perform editing experiments in diploid cells. We therefore measured the LOF efficiency using pDYSCKO in diploid cells. While Target-AID mutagenesis proved efficient (median mutation rate 45%, **figure 2a**), Cas9 mediated LOF was highly toxic and the efficiency was lower than in haploid cells by several orders of magnitude. LOF screen in diploid cells should therefore use the Target-AID system.

The modified induction protocol provides multiple advantages

The induction protocol we propose is different from the original workflow proposed for yeast genome editing in yeast but offers multiple advantages in the context of a genome-wide LOF screen. The protocol proposed by DiCarlo et al consisted of a preculture round in glucose before induction in raffinose/galactose media at OD_{600} 0.3. We added a 24 hours glycerol round before galactose induction, which allows for faster expression, with peak induction being reached in as little as 6 hours (27), a lag time much shorter than what has been observed using raffinose (28). As one of the objectives of a CRISPR-LOF screen is to track fitness changes of mutants in a given condition, faster induction limits the timeframe in which de novo mutations and other stochastic effects appear in the population. This could help reduce variation in mutant abundances caused by innate fitness differences during the induction and double selection process. Finally, diluting to OD_{600} 1.0 instead of 0.3 reduces bottlenecks, which might affect strain representation in a pooled screen.

Increased mutagenesis efficiency provides greater statistical power

In a typical LOF screen, pools of gRNA are cloned and transformed in a population of cells. The fate of individual mutant cells is followed through bulk competition assays in which gRNA sequences serve as barcode to estimate the abundance of each genotype individually by deep sequencing. This approach relies on high mutation rate because a population of cells bearing a specific gRNA could be a mixture of mutated and non-mutated ones, which cannot be distinguished from sequencing the gRNA barcodes alone. A low mutation rate will thus prevent a proper measurement of fitness effects for

the mutated strains. This will be particularly critical for low fitness effect mutations. The approach of measuring fitness through gRNA sequencing is analogous to the yeast Bar-seq approach used to measure the fitness effect of gene deletions in pooled competition assays. We can therefore use existing data on Bar-seq-based fitness measurements to examine how the improved efficiency of gene KO and editing afforded by the double selection can increase the detection power of small fitness effects in this context.

Our simulations show that in experimental conditions that mimic Bar-seq experiments, the increased efficiencies allow for more sensitive detection of LOF effects (**Figure 3**), particularly in the critical zone encompassing fitness effects between 0% and 25%, which are the most frequent effects for the deletion of non-essential genes in yeast. Our double selection strategy therefore makes genome-wide screening powerful enough to meet the high standards established by the use of the yeast deletion collection. The model also predicts that the efficiency of Target-AID editing in diploid cells would be high enough to consider large-scale screening applications.

Multiplex genome editing using Cas9 causes genome rearrangements at high rates

Our experiment also revealed that multiplexed gene CRISPR-Cas9 LOF in yeast has undesirable side effects. We recurrently observed colonies exhibiting an intermediate phenotype for *ADE1* LOF (median rate in populations: 18%), with a color pattern slightly different from the standard *ade1* phenotype. We found that these cells harbored a CRISPR-Cas9 induced chromosomal rearrangement that results into a *CAN1-ADE1* fusion gene. We first identified these fusions by PCR and sanger sequencing and

confirmed that they affect chromosome size by PFGE (**Figure 4a, 4b**). We used another non-essential gene to test whether this effect was locus specific or if it could be reproduced at different positions in the genome. We find that changing the DYSCKO target site produced a different fusion at a similar rate (**Figure 4c**), showing that these fusions are programmable.

Discussion

We developed a simple protocol for high efficiency CRISPR-Cas9 or Target-AID gene disruption and editing in *S. cerevisiae*. The resulting vector, pDYSCKO, is amenable to large-scale library assembly. We provide a statistical model that describes the relationship between mutagenesis efficiency and fitness effect detection power, showing that our double selection is key to making CRISPR-cas9 KO efficient enough for genome-wide screening in yeast.

Our approach relies on the *CAN1* gene that can be used for negative selection against non-mutated cells. This marker has been used in many large-scale and systematic experiments, for instance synthetic gene arrays (SGA) (29), which means that the strains constructed are compatible in terms of genotype with other approaches routinely used by the yeast community. However, if the *CAN1* marker needs to be used for other purposes, other negatively selectable markers could be exploited, for instance *LYP1*, *FCY1* and *URA3*. Similarly, the plasmid selection markers can easily be changed to antibiotic resistance cassettes for G418 (30), Nourseothricin or Hygromycin B (31), making it a polyvalent tool for high throughput gene disruption in non-laboratory strains.

We found that the double-selection combined with CRISPR-Cas9 using the standard

Cas9 produces mutant cells with chromosomal fusions. These were apparent because they produced hypomorphic *ADE1* alleles, which develop an intermediate pink colony phenotype. We found that similar effects can be observed when targeting another gene, suggesting that this could be generalizable and even programmable. More experiments will be needed to determine whether this will be limited to loci that offer specific configurations.

The observation that multiplex genome editing causes chromosomal rearrangements at relatively high frequency opens new opportunities for the use of genome engineering in yeast and likely other organisms as well. It is however also worrisome because many recent multiplex mutagenesis workflow rely on PCR based deep sequencing of the target allele to measure success rates (i.e. CRISPR-Seq (32)) of genome editing at multiple loci. This method cannot detect chromosomal rearrangements, and as such the events may have gone unnoticed in many experiments so far. It would have been in our case if we did not have access to visible phenotypes. DSB repair resulting in chromosomal rearrangements is increasingly attributed to Alt-NHEJ (also known as microhomology-mediated end joining) (33), which is still poorly understood. Studying genetics factors that influence these Cas9 induced chromosomal rearrangements might provide new insights about Alt-NHEJ.

It is unknown if the high occurrence rate is specific to yeast, but efforts should be made to assess it in other models. Failure to do so may bias experimental results, especially in cancer cell lines where genomes are already unstable. As we have yet to observe any case of rearrangements when using Target-AID, base editors might be safer choices in most cases when attempting multiplex LOF mutagenesis. One possible

tradeoff with these editors is that some of the induced changes may lead to silent substitution, such as the ones we observed here, which effectively limit the effective LOF rate. This tradeoff can be alleviated by a careful design of gRNA in order to maximize the effects of the induced base substitution. The development of other base editors, for instance from A-T to G-C (34), offer new possibilities in terms of producing loss of function mutations without DNA cleavage, diminishing even more the need to use system that lead to chromosomal fusions.

Co-selection has been exploited in other organisms before (11, 12) for Cas9 mediated mutagenesis or homologous repair mediated gene editing. To our knowledge this is the first time it is used to enhance base editing efficiency. Understanding the molecular bases of this phenomenon could provide valuable insights to inform future genome editing efforts. Nishida et al (10), who also observed higher bi-allelic edition rates than expected by chance, had proposed Alt-NHEJ or gene conversion as a potential mechanism for this phenomenon. Our observation that co-selection is present in both diploid and haploid cells, and that it can enhance both Cas9 mediated mutagenesis and Target-AID mutagenesis suggests a complex explanation that may involve a number of DNA damage repair components.

This new method can contribute to the widespread adoption of the yeast CRISPR-Cas9 toolkit in the lab, as well as allow the development of new workflows for high throughput systems biology. The pDYSCKO vector can readily be made compatible with new base editors and other negative selection markers, making it a flexible asset for genetic perturbation at multiple scales.

Availability

The pDYSCKO vector will be available on Addgene (http://www.addgene.org/), as well as pDYSCKO-*ADE1*. The Python 2.7 Packages used are available through in the Python Package Index (https://pypi.python.org/pypi/pip) or in Conda (https://conda.io/docs/).

Supplementary material

Oligonucleotides used in this study are presented in table S1. Media recipes are presented in table S2.

Acknowledgements

The authors thank G. Charron for help with the PFGE protocol, and M. Hénault and D. Yamamoto Evans for insightful discussions.

Contributions

P.C.D and A.K.D performed the experiments. P.C.D and A.K.D designed research with support from C.R.L. and N.Y. P.C.D. developed the model with L.N.T. P.C.D analyzed the data and wrote the manuscript with contributions from all authors.

Funding

This work was supported by the Canadian Institutes of Health Research [299432, 324265 to C.R.L. 364920, 384483 and Frederick Banting and Charles Best graduate scholarship to P.C.D.), the Natural Sciences and Engineering Research Council [Alexander Graham Bell graduate fellowship to L.N.T.] and Laval University [André Darveau Fellowship to P.C.D.]

Conflict of interest

The authors declare no competing financial interests.

References

1. Giaever,G., Nislow,C., Alonso,J.M., Stepanova,A.N., Leisse,T.J., Kim,C.J., Chen,H., Altmann,M., Linder,P., Baryshnikova,A., *et al.* (2014) The yeast deletion collection: a decade of functional genomics. *Genetics*, **197**, 451–65.

2. Teng,X., Dayhoff-Brannigan,M., Cheng,W.-C., Gilbert,C.E., Sing,C.N., Diny,N.L., Wheelan,S.J., Dunham,M.J., Boeke,J.D., Pineda,F.J., *et al.* (2013) Genome-wide consequences of deleting any single gene. *Mol. Cell*, **52**, 485–94.

3. Sander,J.D. and Joung,J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.*, **32**, 347–55.

4. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P., Lim,W.A., Barrangou,R., Fremaux,C., Deveau,H., *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–83.

5. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).*, **337**, 816–821.

6. Sanjana,N.E., Shalem,O. and Zhang,F. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*, **11**, 783–784.

7. Sidik,S.M., Huet,D., Ganesan,S.M., Huynh,M.-H., Wang,T., Nasamu,A.S., Thiru,P., Saeij,J.P.J., Carruthers,V.B., Niles,J.C., *et al.* (2016) A Genome-wide CRISPR

Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell*, **166**, 1423–1435.e12.

8. Bassett,A.R., Kong,L. and Liu,J.-L. (2015) A genome-wide CRISPR library for high-throughput genetic screening in Drosophila cells. *J. Genet. Genomics*, **42**, 301–9.

9. 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.

10. 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*, **102**, 553–563.

11. Agudelo,D., Duringer,A., Bozoyan,L., Huard,C.C., Carter,S., Loehr,J., Synodinou,D., Drouin,M., Salsman,J., Dellaire,G., *et al.* (2017) Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat. Methods*, **14**, 615–620.

12. Kane,N.S., Vora,M., Varre,K.J. and Padgett,R.W. (2017) Efficient Screening of CRISPR/Cas9-Induced Events in Drosophila Using a Co-CRISPR Strategy. *G3 (Bethesda).*, **7**, 87–93.

13. Baker Brachmann,C., Davies,A., Cost,G.J., Caputo,E., Li,J., Hieter,P. and Boeke,J.D. (1998) Designer deletion strains derived fromSaccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, **14**, 115–132.

14. Amberg,D.C., Burke,D.J. and Strathern,J.N. (2005) Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition.

15. Maringele,L. and Lydall,D. (2006) Pulsed-field gel electrophoresis of budding yeast chromosomes. *Methods Mol. Biol.*, **313**, 65–73.

16. Lõoke,M., Kristjuhan,K. and Kristjuhan,A. (2017) Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques*, **62**, v.

17. Kumar,S., Stecher,G. and Tamura,K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, **33**, 1870–1874.

18. Robinson,D.G., Chen,W., Storey,J.D. and Gresham,D. (2014) Design and analysis of Bar-seq experiments. *G3 (Bethesda).*, **4**, 11–8.

19. Shalem,O., Sanjana,N.E., Hartenian,E., Shi,X., Scott,D.A., Mikkelsen,T.S., Heckl,D., Ebert,B.L., Root,D.E., Doench,J.G., *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (80-.).*, **343**, 84–87.

20. Hartl,D.L. (2000) A Primer of Population Genetics. *Sunderland, Mássachussets. U.S.A Sinauer Assoc. Inc.*

21. Perez,F. and Granger,B.E. (2007) IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.*, **9**, 21–29.

22. van der Walt,S., Colbert,S.C. and Varoquaux,G. (2011) The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.*, **13**, 22–30.

23. McKinney,W. (2010) Data Structures for Statistical Computing in Python.

24. Hunter,J.D. (2007) Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, **9**,

90–95.

25. Oliphant,T.E. (2007) SciPy: Open source scientific tools for Python. *Comput. Sci. Eng.*, **9**, 10–20.

26. Qian,W., Ma,D., Xiao,C., Wang,Z. and Zhang,J. (2012) The Genomic Landscape and Evolutionary Resolution of Antagonistic Pleiotropy in Yeast. *Cell Rep.*, **2**, 1399–1410.

27. Johnston,M., Flick,J.S. and Pextont,T. (1994) Multiple Mechanisms Provide Rapid and Stringent Glucose Repression of GAL Gene Expression in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **14**, 3834–3841.

28. Stockwell,S.R., Landry,C.R. and Rifkin,S.A. (2015) The yeast galactose network as a quantitative model for cellular memory. *Mol. Biosyst.*, **11**, 28–37.

29. Tong,A.H., Evangelista,M., Parsons,A.B., Xu,H., Bader,G.D., Pagé,N., Robinson,M., Raghibizadeh,S., Hogue,C.W., Bussey,H., *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**, 2364–8.

30. Wach,A., Brachat,A., Pöhlmann,R. and Philippsen,P. (1994) New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast*, **10**, 1793–808.

31. Goldstein,A.L. and McCusker,J.H. (1999) Three new dominant drug resistance cassettes for gene disruption inSaccharomyces cerevisiae. *Yeast*, **15**, 1541–1553.

32. Tothova,Z., Krill-Burger,J.M., Popova,K.D., Landers,C.C., Sievers,Q.L., Yudovich,D., Belizaire,R., Aster,J.C., Morgan,E.A., Tsherniak,A., *et al.* (2017)

Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell Stem Cell*, **21**, 547–555.e8.

33. Chiruvella,K.K., Liang,Z. and Wilson,T.E. (2013) Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect. Biol.*, **5**, a012757.

34. 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.

35. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**.

Figures

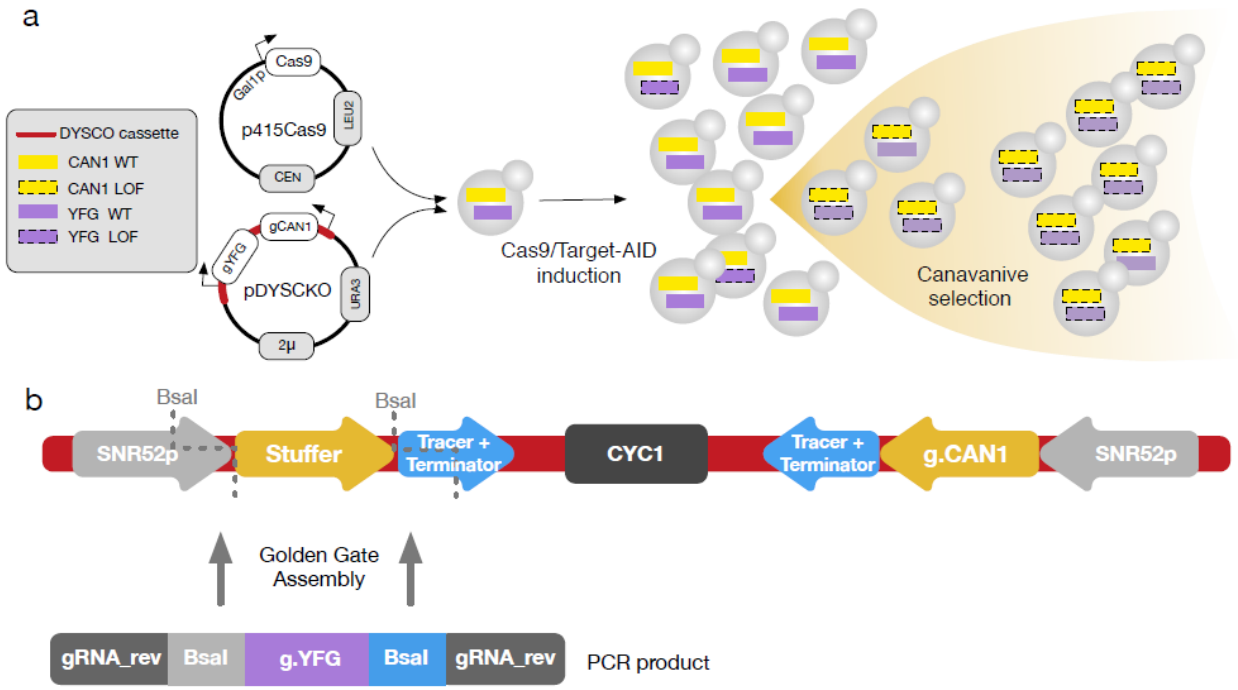

Figure 1. The pDYSCKO-*CAN1* plasmid for double selection LOF. a) Overview of co-selection using the pDYSCKO and the p415Cas9 plasmid, expressing guide RNAs and Cas9 respectively. Only a fraction of cells in the population are successfully edited during Cas9 induction (galactose media), but cells with a LOF at one loci have a greater chance of being bearing LOFs at both targeted loci. Canavanine selection allows for enrichment for these cells in the population, as successful gene edition becomes selectable. b) Overall structure of the DYSCKO cassette, which is on a plasmid also containing standard selection markers (AmpR and URA3). The two symmetrical gRNA expression units use the same promoters and terminators (from SNR52 and SUP4 respectively). The stuffer is a short sequence containing two restriction sites for BsaI that does not match any sequence in the yeast genome. Custom gRNA insertion in the vector performed through Golden Gate assembly (35) using a short dsDNA fragment containing the gRNA target the sequence of interest (g.YFG). In this study, we amplified the inserts from ssDNA oligonucleotides using the gRNA_for and gRNA rev oligonucleotides.

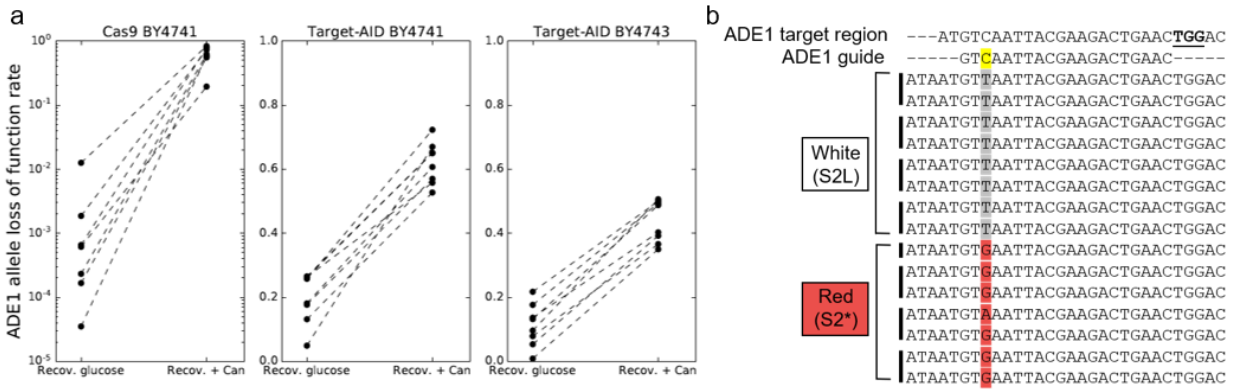

Figure 2 Double selection increases gene editing efficiency. a) The same cell cultures were placed in different recovery conditions after induction of Cas9 or Target-AID expression. Recovery in glucose only does not allow for do-selection. Recovery in canavanin select for cells that have LOF mutations at the *CAN1* locus. Selection on the *CAN1* marker increases the representation of ADE1 LOF colonies. Cells with non-functional *ADE1* alleles accumulate a red metabolic intermediate that gives a characteristic color to colonies allowing for easy identification of mutants, while mutants without LOF mutations in *ADE1* have a wild-type phenotype. The left panel shows standard Cas9 KO in haploid cells, the middle panel Target-AID base edition in haploid cells and the right panel, Target-AID edition in diploid cells. Plating was performed after the 16 hour recovery period. b) Mutations induced by Target-AID in white and red colonies from four independent double selection mutagenesis experiments. The base editing site is shown in yellow in the guide sequence. The changes induced correspond to those expected, with similar proportions of C-T and C-G and rare C-A. All white colonies have been edited as well, showing that the LOF only occurs with the C to G mutations and thus cannot reach a rate of 100% in this specific assays.

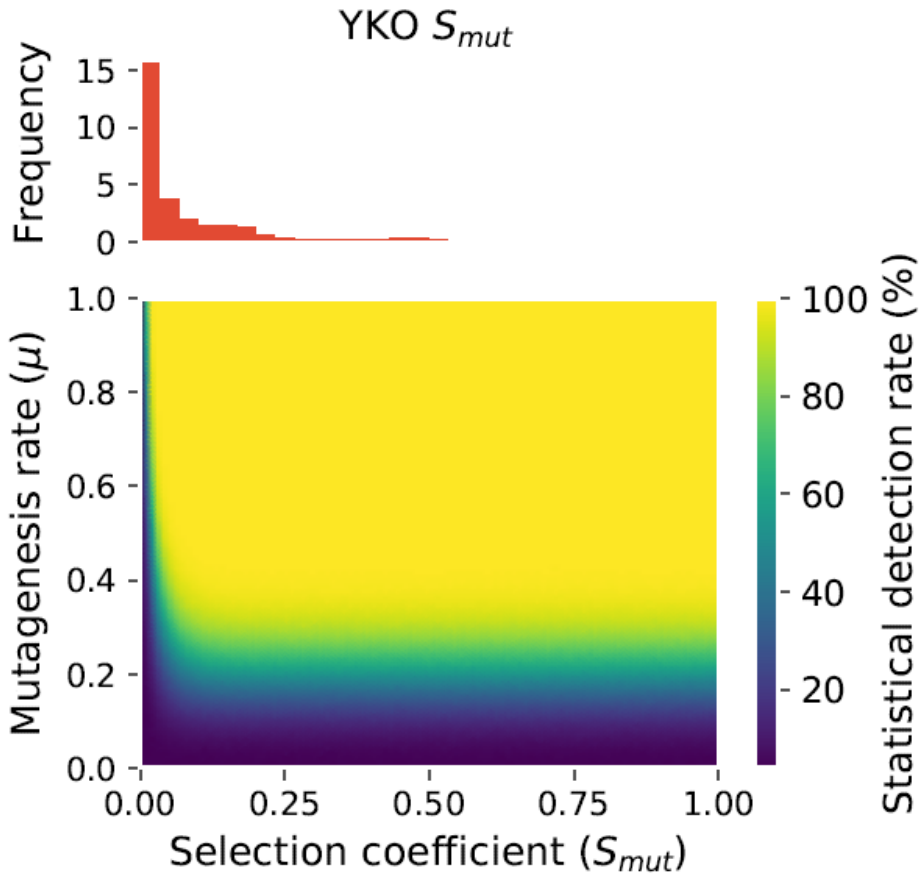

Figure 3 Improvement in mutagenesis rates enhance the power of CRISPR-LOF screens. The rate of detection of fitness defects is shown as a function of mutagenesis rate and the selection coefficient of the mutation. Improving the mutation rate above 0.3 dramatically increases the ability to detect growth defects, a requirement in many experiments including genome-wide CRISPR LOF barcode sequencing. The increase is particularly noticeable for growth rate reduction values associated to the majority of gene KO (YKO) mutants in various conditions (26), as shown in the panel above. The simulations were performed using conditions that reflect typical yeast Bar-seq experiments.

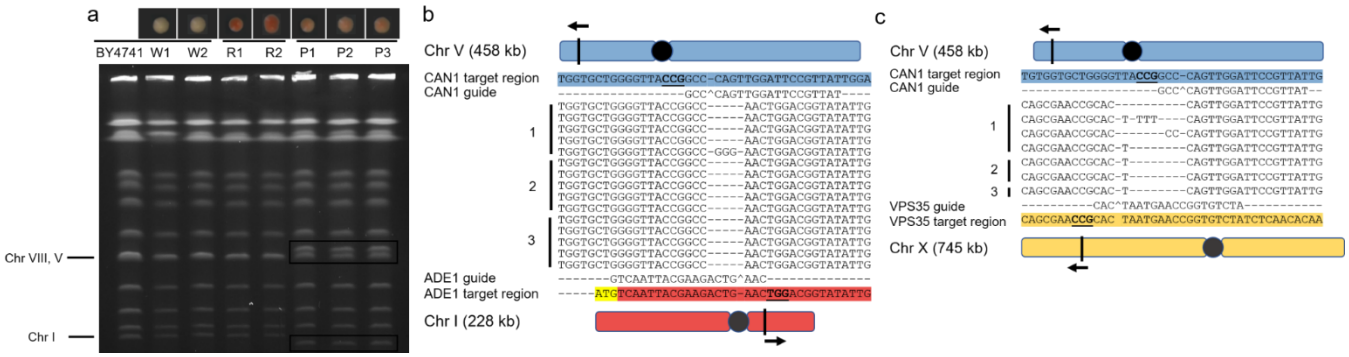

Figure 4 Multiplex Cas9 mediated genome editing causes chromosomal rearrangements in yeast. a) Chromosomal migration patterns of the different Canavanine resistant mutants generated using CRISPR-Cas9. Mutants without LOF mutations in *ADE1* have the same phenotype as the wild-type strain. Colonies formed by cells bearing the fusion have a distinct pink color. *CAN1* only LOF (W1 and W2) and red *CAN1*, *ADE1* LOF mutants show a chromosomal migration pattern that is comparable to the wild-type. Pink colonies from independent mutagenesis experiments show altered length for chromosomes I and chromosomes V, consistent with a reciprocal translocation centered on the predicted Cas9 cut sites. b) Breakpoint sequencing of 5 pink colonies from 3 independent mutagenesis experiments confirms PFGE pattern. c) The same CRISPR-Cas9 mutagenesis protocol was performed but with a gRNA targeting *VPS35*. Colonies in which cells had a fusion allele were identified by PCR as Δ*VPS35* cells have no easily identifiable phenotype. Breakpoint sequencing of 7 chromosomal translocations from three independent mutagenesis experiments confirmed the translocations. Fusion occurrences per replicate were 4 (4) /15, 3 (2) /16, and 1 (1) /16, median rate 18.75%. The number in parenthesis represents the number of detected breakpoints that were sequence