Curr Genet (2013)
DOI 10.1007/s00294-013-0406-x
Received: 1 July 2013 / Revised: 12 August 2013 / Accepted: 29 August 2013

Electronic supplementary material

Rapid, efficient and precise allele replacement in the fission yeast *Schizosaccharomyces pombe*

Jun Gao, Fengling Kan, Jacy L. Wagnon[1], Aaron J. Storey, Reine M. Protacio, Mari K. Davidson*, and Wayne P. Wahls*

Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, 4301 West Markham Street (slot 516), Little Rock, AR 72205-7199, USA

wahlswaynep@uams.edu or davidsonmarik@uams.edu

Supplementary Results and Discussion

Section labels match those in the primary text for ease of reference.

Construction of gene targeting vectors.

The primary text provides details on the design of the gene targeting (allele replacement) vector, on the choice of target, and on the types of modifications that can be engineered. Because the study employed standard recombinant DNA methods and established approaches for site-directed mutagenesis, these were described superficially in the text. Here we describe in greater detail several approaches that were employed and that might be useful to others who wish to use pop-in, pop-out methods for precise allele replacement.

Standard recombinant DNA methods are used to clone the target (homologous) region into the targeting vector either before or after mutagenesis. In some cases there will be no convenient restriction endonuclease (RE) recognition sites for this subcloning. In that case, one can amplify the target locus using PCR primers that contain on their 5' ends the desired RE recognition site, plus the minimum number of extra bases required for efficient cleavage (*e.g.,* as described in the New England Biolabs Inc. product guide). Following amplification, the PCR products are cut with that RE and subcloned. This has no impact upon subsequent steps of pop-in, pop-out allele replacement. Furthermore, the engineered RE sites will not be left in the genome (they will always be lost with the pop-out cassette).

During the construction of gene targeting vectors in *E. coli*, and at stages of allele replacement in *S. pombe*, it is necessary to screen candidate clones for those with the desired changes. Some types of DNA modifications, such as deletions or insertions of DNA sequences encoding epitope tags, can be identified readily by length polymorphisms (*e.g.,* of a PCR product). Other

modifications, such as those that encode a single amino acid substitution in a protein (*e.g.,* Fig. S1), are not readily detectable without DNA sequencing.

To avoid having to use DNA sequencing as a diagnostic tool for the screening of candidate clones, one can engineer additional, nearby base pair changes that create or ablate a RE cut site, but that are translationally silent (*e.g.,* Fig. S1). Since the added or ablated RE cut site is linked tightly to the desired modification, it is diagnostic for that modification. We refer to such diagnostic changes as restriction fragment length polymorphisms (RFLPs), although some diagnostic length polymorphisms (*e.g.,* those created by insertions or deletions) do not require digestion with a RE to be scored.

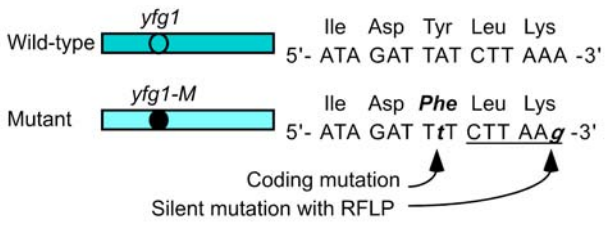

Fig. S1 Diagnostic genotyping of modified alleles. Alleles are different shades for visual reference and differ by discrete modifications (filled circle and, *e.g.,* base substitutions). For site-directed mutagenesis in *E. coli* and for subsequent allele replacement in *S. pombe*, it is necessary to screen clones for those with the desired modifications. Some types of modifications (*e.g.,* those encoding amino acid substitutions) cannot be identified without DNA sequencing. To avoid having to use DNA sequencing to screen candidate clones, one can include translationally silent changes that introduce an RFLP (*e.g.,* an *Eco*RI restriction site, underlined) along with the desired modifications. Then candidate clones can be screened rapidly and economically for the presence or absence of the RFLP in a PCR product.

The cloning and engineered modification of DNA sequences by PCR provides a powerful way to introduce desired modifications (even complex changes), but PCR is inherently mutagenic. Because spontaneous mutations arise at a certain rate in proportion to the number of base pairs replicated, about half of all mutations will arise in the last round of DNA replication. In addition, if a given mutation occurs by chance in an early round of PCR, it will be a "jackpot" mutation whose frequency representation is high in the final population of PCR products. Three strategies can help to reduce the frequency of off-target mutations and hence reduce the number of candidate clones that must be sequenced.

First, use a thermostable DNA polymerase that has proofreading activity. Second, if possible use cloned (*i.e.,* abundant) DNA as template and empirical testing to reduce the number of amplification cycles required. Third, to deal with potential jackpot mutations, split each PCR reaction mixture into multiple different tubes for amplification (we typically use five), then following PCR combine the products for subcloning (if required) and to transform *E. coli*. Rare jackpot mutations might still occur in one amplification tube, but are unlikely to occur in multiple tubes, and hence by splitting the amplification mixture one can ensure statistically that no single jackpot mutation will contaminate all of the amplification products in the pooled sample.

One can readily introduce desired modifications into gene targeting vectors by (option 1) inverse PCR. In this approach, a gene targeting vector that contains wild-type DNA sequences in the homologous portion is subjected to *in vitro* mutagenesis (Fig. S2a). This can be accomplished using well established protocols (Geiser et al. 2001) and standard laboratory reagents or commercial kits (*e.g.,* QuikChange from Stratagene). Circular plasmid DNA obtained from DNA methylation-proficient *E. coli* cells is used as template for PCR extension from overlapping primers that contain the modifications of interest. The inverse PCR copies the entire plasmid

and the resultant linear product contains complementary, overlapping ends with the desired modifications. The samples are then treated with the DNA methylation-dependent RE *Dpn*I, which degrades the methylated plasmid template and leaves intact the unmethylated PCR product. The samples are then transformed into *E. coli* cells that circularize the linear PCR product via the overlapping, complementary DNA ends. Consequently, a high proportion of the *E. coli* transformants contain an intact, circular plasmid DNA molecule (gene targeting vector) with precisely engineered modifications.

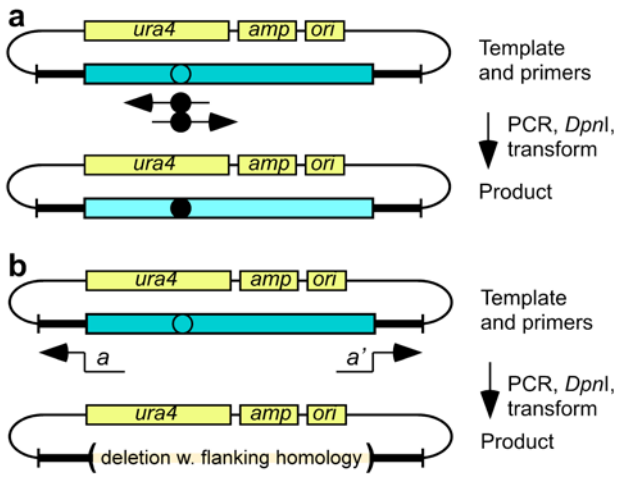

Fig. S2 Allele generation by inverse PCR. A gene targeting (allele replacement) vector with wild-type sequences is used as template. (**a**) Overlapping primers contain mutations of interest (filled circles) and flanking homology. Following inverse PCR, the template is degraded by digestion with *Dpn*I and the PCR product is transformed into *E. coli* to regenerate intact plasmid with the desired modifications. (**b**) Variations on the theme. Primers with different locations but partial overlapping complementarity (*a* and *a'*) can be used to engineer deletions. When used for allele replacement in the genome, these will produce clean deletions (*e.g.,* removal of a coding region without any additional changes). This process can also be used to make more complex changes, such as DNA fusions that encode in-frame epitope tags (see text).

Inverse PCR can also be used to engineer more complex changes. For example, one can use a pair of "hybrid" primers that anneal individually to opposite ends of an open reading frame, but that share a region of DNA sequence complementarity with each other (*e.g.,* Fig. S2b). This can be used to ablate with precision a specific region in the gene targeting cassette, such as an ORF or DNA region that encodes a discrete protein domain. There are many possible variations on this theme, some of which can be used to engineer even more complex changes. For example, we have successfully used as "primers" a single, 530 base pair-long dsDNA fragment with homology on its ends (236 and 29 base pairs, respectively) to the template plasmid. Extension from those ends introduced a "gene fusion" into the gene targeting vector. In short, many different types of allele modifications can be engineered by inverse PCR, including substitutions, deletions and insertions.

The principal advantages of the inverse-PCR approach for mutagenesis are that it is facile, rapid and precise. The principal disadvantage is that it can be difficult to obtain long PCR products, which must be about 6,000 base pairs (or greater) in length, depending on the sizes of the vector backbone, *ura4*$^+$ cassette, and target sequences. There are two options if long PCR products cannot be obtained. First, one can use a smaller plasmid (*e.g.,* one that lacks the *ura4*$^+$ insert) as template for mutagenesis by inverse PCR, then use subcloning to make the larger targeting vector. Alternatively, one can construct the targeting vector via recombinant PCR.

Another way to introduce desired modifications into the gene targeting vector is by (option 2) recombinant PCR. In this approach, one uses two (or more) sequential PCR reactions to generate desired modifications, and then subclones the resulting products into the gene

targeting vector (Fig. S3). As with inverse PCR (above), the fundamental methods and optimization procedures are well established (Krawchuk and Wahls 1999; Heckman and Pease 2007; Bryksin and Matsumura 2010). And as for inverse PCR, a wide variety of different types of allele modifications can be engineered.

The principal advantage of recombinant PCR is that it is more robust than the inverse PCR approach because the amplicon lengths can be much shorter. The disadvantages of recombinant PCR are that it requires more reagents and experimental steps (number of PCR primers, number of stages of PCR amplification, number of PCR products that must be gel purified). These factors correspondingly increases the cost, relative to inverse PCR.

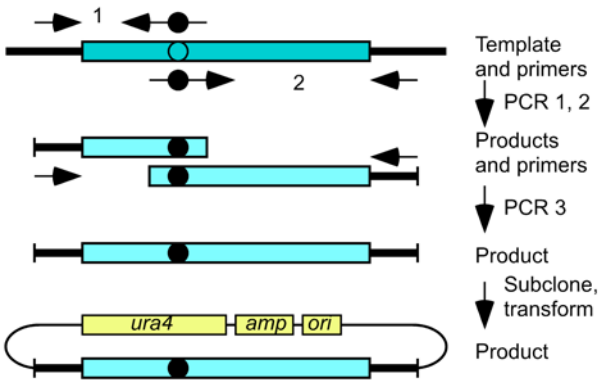

Fig. S3 Allele generation by recombinant PCR. Two independent PCR reactions produce "left" and "right" products. These are purified, combined, and mixed with additional outside primers. During subsequent PCR, the oligonucleotide-length overlap between "left" and "right" products is sufficient to prime between the two and lead to full-length product. The product is then subcloned into the targeting vector. This process can be used to make more complex changes, such as DNA fusions that encode in-frame epitope tags (not shown).

The two strategies (inverse PCR and recombinant PCR) and other strategies can be combined productively. For example, one might use a single plasmid DNA template and inverse PCR to engineer a collection of point mutations encoding different amino acid substitutions, then use recombinant PCR to fuse each coding region in frame to DNA encoding green fluorescent protein, and then subclone those DNA fragments into a plasmid bearing *ura4*$^+$ for gene targeting. Note that placing recombinant PCR before inverse PCR would also work and would involve a similar number of reagents and steps. The main point is that there is considerable flexibility in the selection and order of processes used to generate gene targeting constructs.

Development of a process for locus specific, saturating mutagenesis *in situ* (targeted forward genetics)

In section four of the "Results" (main text), we report that the pop-in, pop-out approach can be applied in population scale to an individual target locus, without genotyping, to identify clones with an altered phenotype. This suggested suitability of the approach for "targeted forward genetics" and, as with other applications of the pop-in, pop-out approach, molecular mechanisms of recombination and probability theory come into play.

At the stage of transformation, Ura$^+$ colonies can arise by pop-in homologous recombination at the correct target locus (observed mean of 83%) or by nonhomologous integration elsewhere (Table 1). However, only the correctly targeted integrations contain tandem repeats that can undergo pop-out homologous recombination (Fig. 2c-2d) to produce at high frequency FOA^r colonies (≥ 50-fold observed difference, relative to spontaneous mutations in the *ura4*$^+$ cassette or elsewhere). Therefore, in population scale the selection for FOA^r colonies enriches for clones that have had correctly targeted pop-in, pop-out events (Table S1 and data discussed below).

Table S1 Population frequencies of correctly targeted pop-in, pop-out among FOA^r colonies.

Model	Integration type	Pop-in step (fraction of Ura$^+$ colonies)	Pop-out step (fraction of FOA^r colonies)	Relative efficiency	Net pop-in, pop-out homologous targeting efficiency (% of FOA^r colonies)
Observed	**H**	**0.83**	≥ 0.98	≥ 0.8134	**≥ 99.6**
	NH	0.17	≤ 0.02	≤ 0.0034	
Calculated 1	**H**	**0.75**	≥ 0.98	≥ 0.7350	**≥ 99.3**
	NH	0.25	≤ 0.02	≤ 0.0050	
Calculated 2	**H**	**0.50**	≥ 0.98	≥ 0.4900	**≥ 98.0**
	NH	0.50	≤ 0.02	≤ 0.0100	
Calculated 3	**H**	**0.25**	≥ 0.98	≥ 0.2450	**≥ 94.2**
	NH	0.75	≤ 0.02	≤ 0.0150	

For 39 different gene targeting vectors, the observed mean frequency of homologous (H) gene targeting was 83% of Ura$^+$ colonies, with 17% being due to nonhomologous (NH) integration elsewhere (**Table 1**). The lowest observed homologous targeting frequency was 25%, so calculations here are based on representative homologous targeting efficiency values between 25% and 83%.

Among a population of Ura$^+$ colonies generated by transformation (and then subjected to a brief period without selection), those with homologous integrations produce FOA^r colonies at a frequency ≥ 50-fold higher than those with nonhomologous integrations (pop-out of the *ura4$^+$* cassette plus spontaneous mutations, versus spontaneous mutations alone).

The relative efficiency is the product of the frequency values for the pop-in step (Ura$^+$) and the pop-out step (FOA^r) among the homologous and nonhomologous integration classes.

Note that even when homologous gene targeting (pop-in) efficiencies are low (*e.g.,* 25%), the vast majority of the FOA^r colonies in the population will arise from pop-in, pop-out recombination events at the locus being targeted. Consequently, the streamlined pop-in, pop-out approach can be used for saturating mutation screens of discrete chromosomal elements (*e.g.,* a protein coding region) *in situ*.

Homologous gene targeting (pop-in) efficiencies can vary from locus to locus or from construct to construct. Nevertheless, the fact that FOA^r colonies arise more frequently from pop-out recombination events than from mutation (third section of “Results”) confers great power to population studies. For example, even if the initial homologous gene targeting efficiency is only 25% for a given locus [the lowest efficiency that we observed for 39 different gene targeting constructs (Table 1 and second section of “Results”)], then about 94% of the FOA^r colonies would be due to pop-in, pop-out events at the locus being targeted (Table S1).

We emphasize that not all FOA^r colonies that arise from pop-in, pop-out recombination events at the target locus will have an allele replacement at that locus. Pop-out excision of the targeting vector can leave either a wild-type or a modified allele in the genome (Fig. 2c-2d) and the proportions of these two alternative outcomes is dictated by the lengths of homology within which pop-out recombination events can occur (Fig. 4). This does not compromise the utility of the approach for saturating, locus-specific mutation screens *in situ*, although it does mean that one should screen more FOA^r colonies than the number of desired mutants. The actual number of colonies to be screened can be predicted (Fig. 4) from positions of desired mutations relative to lengths of homology in the targeting vector (*a, b* and *c* in Fig. 2). A more detailed modeling of mutation frequency distributions, with regard to *in situ* mutagenesis, is provided below.

Nine independent experiments, using nine different vectors targeted to *ade6*, validated the utility of pop-in, pop-out for targeted mutagenesis (allele replacement) in population scale (Table 1,

footnote “c”). Those data also demonstrated that selection for FOA^r resistance enriches the population for members that have undergone successful pop-in, pop-out recombination.

In those nine experiments, we obtained a mean homologous targeting (correct pop-in) efficiency of 69.6%. In blinded experiments (prior to knowing which of the Ura$^+$ transformants were due to homologous integration events) we analyzed populations of FOA^r colonies derived from the populations of Ura$^+$ founders. For those FOA^r populations we observed a mean allele replacement frequency of 35.9% (Table 1, footnote “c”). This observed mean frequency matches the predicted mean frequency (35.1%) based on homology length ratios for pop-out recombination events that would leave the modified allele in the genome (Fig. 4 and Table 1). If one normalizes the observed population data (35.9% of FOA^r colonies have allele replacement) to the known effects of length ratios on the outcome of pop-out recombination (35.1% of pop-out events should produce allele replacement), then 102% (means normalized) of the FOA^r colonies are attributable to correctly targeted pop-in, pop-out recombination events. This outcome was achieved even though only 69.6% (mean) of the Ura$^+$ founders in the initial populations were due to correctly targeted pop-in events.

To recapitulate, in population scale experiments the selection for FOA resistance enriches specifically for clones that arise from pop-in, pop-out recombination at the target locus. Even if the initial gene targeting efficiency is low, the vast majority of FOA^r colonies will be due to pop-in, pop-out events at the target locus of interest (Table S1). Correspondingly, the frequency of off-target events in the final population will be low.

Two additional factors must considered in designing experiments. First, for pop-in recombination, mutations located close to the DSB in the targeting vector can be lost due to gene conversion triggered by DSB end resection, or heteroduplex DNA adjacent to DSBs, or both [Fig. 3a and Table 1, footnote “d”; see also (Szankasi et al. 1988; Tatebayashi et al. 1994; Davidson et al. 2004)]. We have not defined conversion tract lengths with precision, but results for one mutation (*ctt1-ΔM26*) located close to (141 base pairs away from) the DSB suggest that the median tract length is about 150 base pairs (this presumably extends to each side of the DSB). Second, for pop-out recombination, the position of a mutation relative to ratio lengths of homology will affect the frequency with which that mutation is left in the genome (Fig. 4). These experimentally defined parameters and probability theory support rational design of *in situ* mutation screens.

A conceptual representation is provided in Fig. S4. We refer to the chromosomal DNA element that is to be subjected to saturating mutagenesis *in situ* as the “element of interest” (rather than “target”) to avoid confusion with other definitions of “target” used in the study. The homologous region of the gene targeting vector (or a subsection of the homologous region) is mutagenized *in vitro* and that mutagenized population of gene targeting vector is used for pop-in, pop-out protocols. We illustrate the effects of DSB location (pop-in recombination) and the location of excision events (pop-out recombination) on final genotypes for DSBs placed to the left (DSB1), middle (DSB2) and right (DSB3) of the element of interest (Fig. S4).

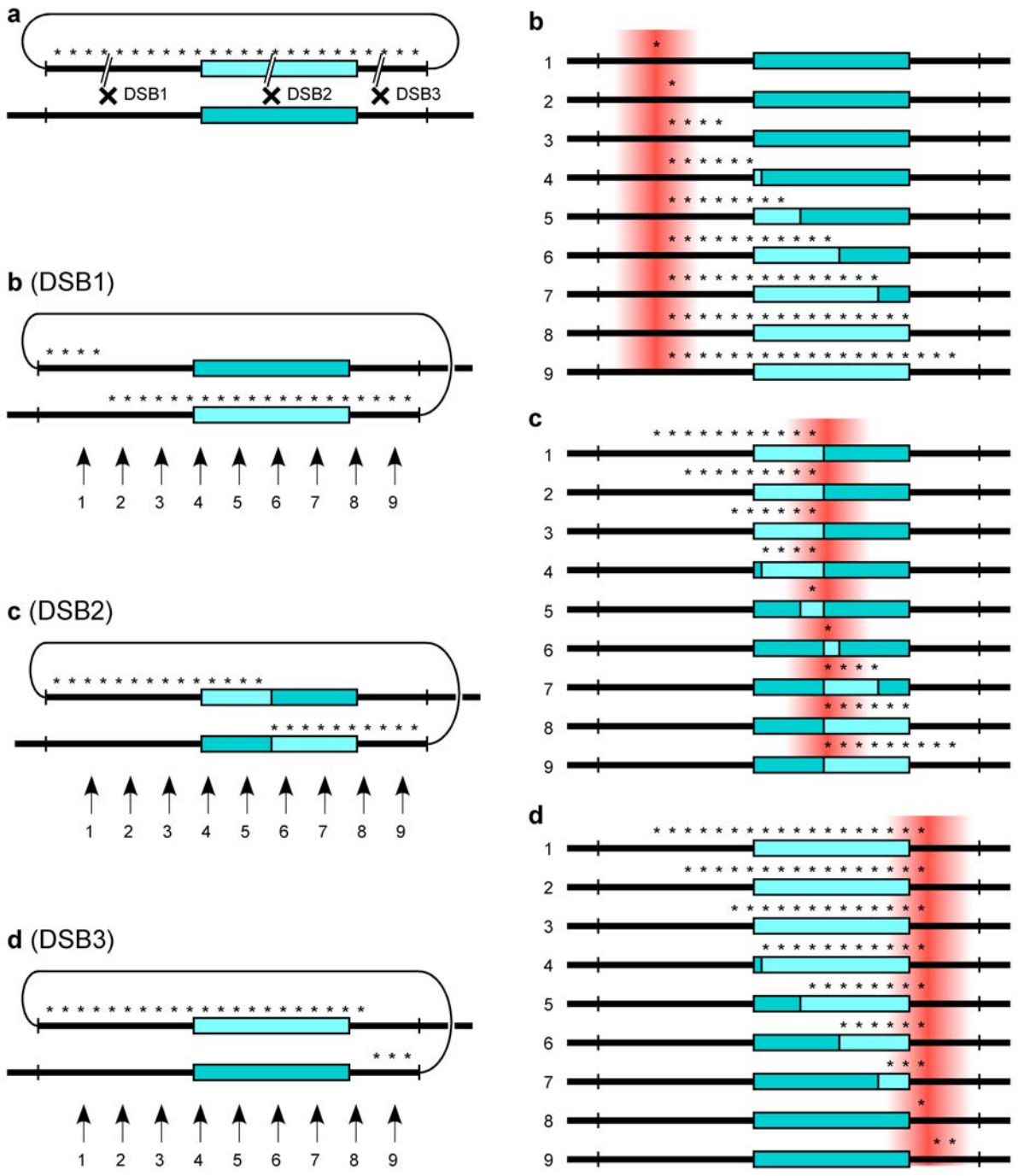

Fig. S4 Effects of DSB position on saturating *in situ* mutation screens. (**a**) Depicted are a chromosomal element of interest (box), the population average distribution of mutations (asterisks) sprinkled into the homologous region of the gene targeting vector, and the positions of DSBs used to promote pop-in recombination. Ideally, each DNA molecule in the targeting vector population will have, on average, one mutation. (**b-d**) Left panels show structures of tandem duplications created by pop-in for each DSB position. Also indicated are representative positions of subsequent pop-out recombination events (numbers). Right panels depict the population average distribution of each mutation in the genome after pop-out. (Individual clones will have only one or a few of the mutations depicted in each row, depending on the density of mutations in the gene targeting vector population.) Mutations in the "DSB zone" (red shading) are prone to loss (gene conversion) during pop-in recombination events. Consequently, they will be under represented or absent from the population of pop-out recombinants (see text for details).

At first glance, placing a DSB within the element of interest (Fig. S4a, DSB2) would seem attractive because the distribution of mutations left in the genome after pop-out would be centered on the element of interest (Fig. S4c). Unfortunately, there are two problems with this approach. First, the distribution of mutations in the final population would be polarized. Mutations towards the center of the element would be represented at high frequency and mutations towards the ends would be recovered at low frequency. Such positional bias in mutation frequency distributions is not well suited for saturating mutation screens, which are efficient only if mutations are distributed stochastically. Second, mutations flanking the DSB are lost through gene conversion during pop-in recombination, with the loss rate being an inverse function of distance from the DSB (dictated by lengths of end resection and/or heteroduplex DNA). This is a major problem—if the experimental approach precludes recovery of some mutations in the region of interest, then the genetic screen is fundamentally flawed.

The alternative is to place the DSB to the left (Fig. S4a, DSB1) or to the right (Fig. S4a, DSB3) of the element of interest. Such configurations can avoid entirely (*e.g.,* Fig. S4b) or in large part (*e.g.,* Fig. S4d) the loss of mutations (within the region of interest) due to gene conversion at the pop-in step. Unfortunately, neither approach alone is satisfactory for saturating mutagenesis because the distribution of mutations in each final population would be polarized. Many mutations would be recovered for one end of the element and few for the other (Fig. S4b or Fig. S4d), which undermines the power of the genetic screen.

The solution is simple and powerful. One can split the mutagenized gene targeting vector into two batches, digest one batch with a RE that cuts to the left of the region of interest (Fig. S4a, DSB1) and digest the other batch with a RE that cuts to the right of the region of interest (Fig. S4a, DSB3). Each batch is used for pop-in, pop-out protocols and the resulting FOA^r colonies are combined for analyses. The net outcome would be a fairly uniform frequency distribution of mutations, spanning the element of interest, in the population of pop-out (FOA^r) colonies (combined mutation frequency distributions of Fig. S4b and S4d, see also Fig. 5 in main text).

For the sake of illustration and conceptual clarity, Fig. S4 depicts the outer two DSB positions as being outside of the element of interest. Although this is the optimal approach experimentally, there is flexibility in positioning of the two DSBs used for *in situ* mutagenesis. The DSBs (one or both) can even be positioned within the element of interest (*e.g.,* Fig. 5). Except within the “DSB zones” of gene conversion (whose extent has not been precisely defined), the frequency distribution of mutations between the two DSBs will be uniform and frequency distributions outside of the DSBs will decrease in proportion to distance from each DSB (Fig. 5). And within each DSB zone, mutations will still be recovered at a frequency sufficiently high for saturation screening because they will be contributed by pop-in, pop-out mutagenesis from the opposing DSB (Figs. S4, 5).

In the Discussion of the main text, we describe how such precisely targeted, saturating mutagenesis can be used to study essential and non-essential genes *in situ*. Importantly, the mutations will be expressed from the endogenous locus (single copy), under control of native regulatory sequences, without any heterologous sequences such as selectable markers. We also describe how the population scale approach can be applied productively to characterize other chromosomal elements, such as regulatory DNA sequence motifs.

Supplementary References

Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48:463-465

Davidson MK, Young NP, Glick GG, Wahls WP (2004) Meiotic chromosome segregation mutants identified by insertional mutagenesis of fission yeast *Schizosaccharomyces pombe*; tandem-repeat, single-site integrations. Nucleic Acids Res 32:4400-4410

Geiser M, Cebe R, Drewello D, Schmitz R (2001) Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. Biotechniques 31:88-90, 92.

Heckman KL, Pease LR (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc 2:924-932

Krawchuk MD, Wahls WP (1999) High-efficiency gene targeting in *Schizosaccharomyces pombe* using a modular, PCR-based approach with long tracts of flanking homology. Yeast 14:1419-1427

Szankasi P, Heyer WD, Schuchert P, Kohli J (1988) DNA sequence analysis of the *ade6* gene of *Schizosaccharomyces pombe*. Wild-type and mutant alleles including the recombination hot spot allele *ade6-M26*. J Mol Biol 204:917-925

Tatebayashi K, Kato J, Ikeda H (1994) Structural analyses of DNA fragments integrated by illegitimate recombination in *Schizosaccharomyces pombe*. Mol Gen Genet 244:111-119