Published online 29 June 2022

Nucleic Acids Research, 2022, Vol. 50, No. 17 ***e102***
https://doi.org/10.1093/nar/gkac546

A multiplexed, three-dimensional pooling and next-generation sequencing strategy for creating barcoded mutant arrays: construction of a *Schizosaccharomyces pombe* transposon insertion library

Yanhui Li[1,2], Neil Molyneaux[2], Haitao Zhang[3], Gang Zhou[4], Carly Kerr[3], Mark D. Adams[2], Kathleen L. Berkner[5] and Kurt W. Runge[1,2,3,*]

[1]Department of Molecular Genetics, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, OH 44195, USA, [2]Department of Genetics and Genomic Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA, [3]Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, OH 44195, USA, [4]Department of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, OH 44195, USA and [5]Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, OH 44195, USA

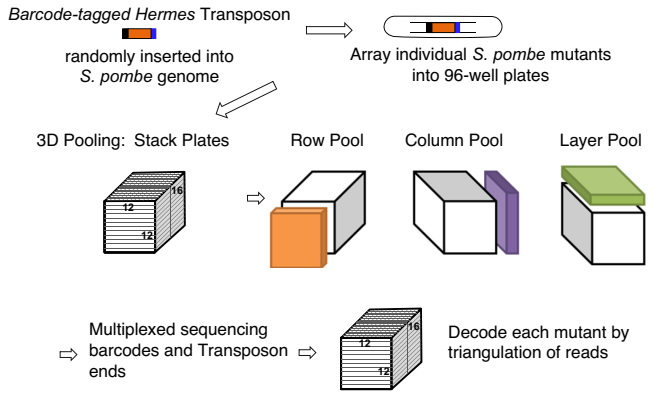
Received May 25, 2021; Revised June 02, 2022; Editorial Decision June 07, 2022; Accepted June 12, 2022

ABSTRACT

Arrayed libraries of defined mutants have been used to elucidate gene function in the post-genomic era. Yeast haploid gene deletion libraries have pioneered this effort, but are costly to construct, do not reveal phenotypes that may occur with partial gene function and lack essential genes required for growth. We therefore devised an efficient method to construct a library of barcoded insertion mutants with a wider range of phenotypes that can be generalized to other organisms or collections of DNA samples. We developed a novel but simple three-dimensional pooling and multiplexed sequencing approach that leveraged sequence information to reduce the number of required sequencing reactions by orders of magnitude, and were able to identify the barcode sequences and DNA insertion sites of 4391 *Schizosaccharomyces pombe* insertion mutations with only 40 sequencing preparations. The insertion mutations are in the genes and untranslated regions of nonessential, essential and noncoding RNA genes, and produced a wider range of phenotypes compared to the cognate deletion mutants, including novel phenotypes. This mutant library represents both a proof of principle for an efficient method to produce novel mutant libraries and a valuable resource for the *S. pombe* research community.

*To whom correspondence should be addressed. Tel: +1 216 445 9771; Fax: +1 216 444 0512; Email: rungek@ccf.org
Present addresses:
Yanhui Li, Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA.
Gang Zhou, Department of Pathology, Baylor Scott & White Health, Temple, TX, USA.
Mark D. Adams, The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.
Kathleen L. Berkner, Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, OH, USA.
Kurt W. Runge, Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, OH, USA.

GRAPHICAL ABSTRACT



INTRODUCTION

Defining gene function in the post-genomic era has benefited from the construction of collections of defined mutants in model organisms (1–5). One useful form of such collections is arrayed deletion mutants, in which each mutant is in a known location in an array (e.g. a known well of specific 96-well plate in a collection of such plates). Such arrays have allowed the rapid phenotypic screening under a wide variety of conditions to elucidate new gene functions (6–10). In the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, the arrayed deletion mutants also each contain one or two 'barcodes', unique DNA sequences specific for each mutant. Each mutant is thus tagged with a 20-bp unique sequence that allows one to assay mixed cultures of many different mutants at once and track the relative abundance of each mutant by measuring the relative proportions of each barcode (4,5,11–15).

The arrayed, barcoded, defined mutant collections are powerful tools, but targeted deletions such as those in yeasts or mammalian cells (16) have some drawbacks. Targeted deletions are labor intensive to construct and validate. Collections of deletion mutants greatly increase the labor and expense of mutant generation and validation, in some cases requiring thousands of PCR or sequencing reactions to validate the mutant collection (4,5,17–20). In addition, deletion mutations cause loss of gene function, those essential genes that are required for growth are absent from the collection and the focus on protein-coding genes means that noncoding RNAs were not targeted. The goal of this work was to devise an efficient, cost-effective method to produce a large number of uniquely barcoded mutants characterized at the sequence level that would have wide application to model organisms or arrays of genetic materials such as plasmids with random inserts. We used the fission yeast *S. pombe* in a proof-of-principle project to develop this system as *S. pombe* is an important eukaryotic model system to study aging and gene–drug interactions (21–28), has powerful molecular genetics and has processes similar to mammals, including cell cycle control, RNA splicing, RNAi-mediated gene silencing, telomere function and chromosomes with large centromeres containing repetitive sequences (29,30).

Our novel approach used a unique combination of transposon insertion mutagenesis, random barcoding, three-dimensional (3D) pooling of mutants, high-throughput sequencing and subsequent computational data analysis to leverage sequence information to identify the sequences of the insertion sites in gene regulatory and coding sequences (CDS) and barcodes in viable mutants. Each identified mutant and its barcode were validated by three independent sequencing reactions as part of the procedure. The method does not require complex pooling methods that required coding and decoding [e.g. (31)] to achieve efficiency and low cost. The analysis pipeline is applicable to any insertions that have defined sequences when inserted into the genome, and can be applied to any collection of viable cells such as mutagenized single-celled organisms or plasmid collections in bacteria. The collection of 4095 viable, uniquely barcoded, validated *S. pombe* insertion mutants generated in this project contained mutants with similar and novel phenotypes compared to cognate strains in the deletion collection. The insertions disrupted 20% of the annotated essential genes, 40% of the nonessential genes and 30% of the noncoding RNA genes. Our approach thus allows the construction of an insertion library that complements an existing gene deletion collection to serve as a valuable resource for the elucidation of gene function, and provides a means to rapidly generate defined mutation collections in other model systems.

MATERIALS AND METHODS

Construction of barcoded *Hermes* transposon plasmids

Hermes transposon donor plasmid (pHL2577) and transposase plasmid (pHL2578) were obtained from Dr Henry Levin. We replaced the *LEU2* gene with a *ura4*$^+$ marker in the backbone of pHL2578 to construct pHL2578u (32).

The 78-bp barcode oligo (5′-/5Phos/TG GCC ACC CGG GCC ANN NAN ANN NAN ANN NAN ANN NAN ANN NAN ANN NAN ANN NAG GGC CAC CCG GGC CGG CGC GCC C-3′) was annealed to oligo (5′-/5Phos/CG CGC CGG CCC GGG TGG CC-3′) under the following conditions: 1 min at 95°C, −1°C per cycle, 15 cycles; 1 min at 80°C, −0.5°C per cycle, 70 cycles; and 1 min at 45°C, −0.5°C per cycle, 66 cycles to 12°C, followed by filling in to generate double-strand (ds) barcodes using Klenow fragment (3′→5′ exo$^-$) (NEB). The ds DNA barcode fragments with 5′ blunt ends and 3′ CCC overhangs were cloned within the *Hermes* transposon in pHL2577 and then transformed into DH5α by electroporation. Ten separate transformations each produced 1–2 × 10^5 bacterial colonies per transformation. The colonies from each transformation were scraped from agar plates for plasmid preparation. The barcoded *Hermes* transposon plasmids (pHL2577 barcode) were isolated by Plasmid Midi Kit (Qiagen).

Generation of a library of *Hermes* insertion mutants

Schizosaccharomyces pombe KRP201 (*h*$^+$, *ade6-m216*, *leu1-32*, *ura4-D18*) cells were transformed with 1 μg of pHL2578u plasmid and grown on an EMM-ura plate. Frozen competent KRP201 cells were made as described (33). These cells were transformed with the pHL2577 barcode plasmids (1 μg) and plated on EMM + adenine,

leucine, histidine and uracil plates for 24 h and then replica plated on YES + G418 (200 μg/ml) + FOA (1 g/l) and grown for 3 days. About 1000 *S. pombe* colonies were picked into 96-well plates from each pHL2577 barcode plasmid library transformation. A total of 96 plates were frozen in YES + G418 + FOA + 15% glycerol and stored at −80°C.

To calculate the probability of isolating two mutants with the same barcode, we used the following formula: $P = 1 - (1 - f)^N$ (34), where $f = 1$/(number of barcoded *Hermes* transposon plasmids), N is the number of *S. pombe* clones sampled, P is the probability of getting a barcode and $1 - P$ is the probability of not getting a barcode.

This calculation predicts a >99% probability of not getting the same barcode if picking 1000 *S. pombe* colonies from each barcoded *Hermes* transposon plasmid transformation.

The same method was used to estimate the total number of *Hermes::kanMX* insertion mutants required to obtain insertions in 80% of the protein-coding genes (see the 'Discussion' section). Solving for N where $P = 0.8$ and f = (1/number of protein-coding genes) × (fraction of insertions in protein-coding genes/total number of insertions) = (1/5132) × (2273/9024) yields 14 520 mutants.

Pooling cells for sequencing, genomic DNA preparation and fragmentation

Cells from the frozen 96-well plates containing *Hermes* barcoded insertions were revived on omni YES plates. Three copies of each plate from the omni YES plates were made in YES + G418 + 5-FOA liquid. The 24 plates were stacked as 2 plates for one layer, totaling 12 layers. For each row, column and layer pool, a multichannel pipettor was used to remove 50 μl of cells from each well and transfer the cells to a sterile basin. For example, a 12-channel pipettor was used to transfer cells from the same row for each layer in the stack of 2 × 12 plates to construct that row pool. The pooled cells from the sterile basin were transferred to sterile 50-ml tubes, the cells were pelleted and the media discarded. The cell pellet was resuspended in 1.0 ml of sterile YES + G418 + FOA + 3% glucose medium and transferred to a 250-ml flask containing 50 ml of the same medium. Cells were grown with shaking at 32°C overnight, cell density was determined and 10^9 cells were transferred to a 50-ml tube. Cells were pelleted, resuspended in 1 ml sterile Milli-Q filtered water and transferred to a 1.5-ml screw cap tube, cells were pelleted again, the supernatant discarded and the cell pellets were used to prepare genomic DNA. Genomic DNA was extracted from 16 row-pooled, 12 column-pooled and 12 layer-pooled cells by resuspending 10^9 cells from each pool in 250 μl lysis buffer (100 mM Tris, 50 mM EDTA, 1% SDS) and 500 μl 0.5-mm zirconia/silica beads (BioSpec Inc.). Cells were broken in a Mini-BeadBeater (BioSpec Inc.) for 2 min. Genomic DNA was purified by phenol/chloroform and precipitated by isopropanol. After further treatment with RNase and proteinase K, genomic DNA was subjected to phenol/chloroform extraction and precipitated with ethanol.

Genomic DNA (2 μg) was fragmented by restriction enzymes MseI, ApoI or MfeI (NEB) digestions in parallel. The digestion was done at 37°C for 8 h for MseI and MfeI, or at 50°C for 8 h for ApoI. The reactions were heat inactivated for 10 min at 80°C. The digested DNAs were isolated using a Qiagen PCR Purification Kit. ApoI and MfeI digestions were mixed and isolated together.

Ligation-mediated PCR

The MseI ds linkers were generated by annealing the upper strand oligo (5Phos/TAGTCCCTTAAGCGGAG/3AmM/-amino) to the lower strand oligo (5′-GTAATACGACTCACTA TAGGGCTCCGCTTAAGGGAC-3′). ApoI and MfeI ds linkers were generated by annealing the upper strand oligo (5Phos/AATTGTCCCTTAAGCGGAG/3AmM/-amino modified) to the lower strand oligo (5′-GTAATACGACTC ACTATAGGGCTCCGCTTAAGGGAC-3′). A 20-fold molar ratio of linkers was used for ligation onto restriction enzyme-digested genome fragments. T4 DNA ligase (NEB) was added and the reaction was incubated for 16 h at 16°C, and then heat inactivated for 20 min at 65°C. The three enzyme ligation products from the same pool were mixed.

Linker ligation-mediated PCR was performed in three steps (see the 'Results' section). To amplify the *Hermes* transposon right end (HR) insertion sites, the first step was done by the HR outside primer (5′-CTTGCACTCAAAAGGCTTGACAC-3′) specific to the transposon right end using the linker primer (5′-GTAATACGACTCACTATAGGGCTC-3′) specific to the linkers using the following conditions: 2 min at 98°C, 6 cycles of 15 s at 98°C, 30 s at 65°C, 40 s at 72°C and then 24 cycles of 15 s at 98°C, 30 s at 60°C, 40 s at 72°C and a final step for 5 min at 72°C. The PCR products were diluted 20-fold.

The second step was performed using the adaptor–linker primer (5′-CAAGCAGAAGACGGCATACGAGCTCT TCCGATCTGTAATACGACTCACTATAGGGCT-3′) and an 8-bp indexed HR-nested primer (5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTX XXXXXXXTATGTGGCTTACGTTTGCCTGTGG-3′), which, respectively, adds one of the Illumina adaptors to PCR products. The conditions were as follows: 2 min at 98°C, 10 cycles of 15 s at 98°C, 30 s at 60°C, 40 s at 72°C and a final step for 5 min at 72°C. The third step was done by an adaptor–linker primer and adaptor–seq primer (5′-AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCT-3′) to add Illumina sequencing primer and another Illumina adaptor to the final product. The conditions were as follows: 2 min at 98°C, 10 cycles of 15 s at 98°C, 1 min at 72°C and a final incubation for 5 min at 72°C.

The *Hermes* transposon left end insertion sites were similarly amplified in three steps using *Hermes* transposon left end-specific primers. All PCRs were performed by Phusion High-Fidelity DNA Polymerase (NEB).

Amplification of barcodes and Illumina library preparation

Barcodes were amplified from each pool of genomic DNA by an indexed barcode primer (5′-ACACTCTTTCCCTA CACGACGCTCTTCCGATCTXXXXXXXTATCCC GGGATTTTGGCCAC-3′) and barcode reverse primer (5′-CAAGCAGAAGACGGCATACGAGCTCTTCCG ATCTCTGCAGCGAGGAGCCGTAAT-3′) using the

following conditions: 2 min at 98°C, 30 cycles of 15 s at 98°C, 30 s at 60°C, 30 s at 72°C and a final step for 5 min at 72°C. The second step was done using the adaptor–seq primer and barcode reverse primer to add Illumina adaptors and a sequencing primer to the final products. The conditions were 2 min at 98°C, 10 cycles of 15 s at 98°C, 30 s at 72°C and a final step for 5 min at 72°C. The final products of transposon left and right end insertion fragments and barcodes were gel isolated (Supplementary Figure S1). Equal molar amounts of products were mixed. Customized index tags are presented in Supplementary Table S3.

Mapping of integration sites

Single-end sequencing of multiplexed samples was performed on multiple lanes of the Illumina HiSeq 2500. Sequence reads were extracted from FASTQ files from the sequencers (Supplementary Figure S2). The raw sequence data were parsed into row, column or layer pools and read by the 8-bp index tags, followed by trimming the adaptor sequences. The data were then further sorted and trimmed by the reads preceding the barcode (5′-TATCCCGGGATTTTGGCCACCCGGGCC-3′), transposon right end (5′-TATGTGGCTTACGTTTGC CTGTGGCTTGTTGAAGTTCTCTG-3′) or left end (5′-GCGCATAAGTATCAAAATAAGCCACTTGTT GTTGTTCTCTG-3′). Genome sequences were mapped to the *S. pombe* genome using Bowtie. A customized triangulation program was used to bundle Bowtie hits that started at contiguous mapped bases; the intersected row, column and layer pool reads were assigned to barcodes and integration sites to strains (see the 'Results' section).

A list of the insertion sites and barcode sequences associated with each insertion is provided in Supplementary File S1. A list of mutated genes is provided in Supplementary File S2.

Verification of high-throughput sequencing results

Random strains were picked from the *Hermes* library. To verify *Hermes* insertion sites, inverse PCR was performed on individual mutants and the insertion points were compared to the high-throughput results (32). For some strains, a *Hermes* primer that bound the transposon end and a genome primer, which was designed based on the integration sites from the high-throughput results, were used in PCR to test for the presence of the insertion.

To verify the barcode sequences, primer 3829s (5′-CAAGACTAGGAAAAGAGCATAAG-3′) and 4171as (5′-GACTGTCAAGGAGGGTATTC-3′) were used to amplify and sequence the DNA barcodes from individual strains, which were then compared to the high-throughput results.

Examination of respiration mutants and CPT-resistant mutants

Two thousand three hundred twenty-eight unique *S. pombe* genes disrupted by *Hermes* transposon were sorted by Gene Ontology under the term 'respiratory chain complex I, II, III, IV, V and assembly proteins' (AmiGO, http://amigo.geneontology.org/cgi-bin/amigo/go.cgi). *Schizosaccharomyces pombe* mutants carrying 15 genes under this GO term were spot tested on YES and nonfermentable YEEG (0.5% yeast extract, 2% glycerol, 2% ethanol, 2 g/l casein amino acids, amino acids mix) plates. Photos were taken after 5 days. The first spot contained 2×10^4 cells. The rest were 5-fold dilutions. Growth from defective strains was inoculated in liquid YEEG at OD_{600} 0.2 and cultured for 5 days to confirm phenotypes.

CPT mutants were spot tested on YES, CPT 5, 10 and 15 μM plates. The first spot contained 3×10^6 cells. The rest were 5-fold serial dilutions. Photos were taken on the third day or until phenotypes were observed.

Construction of pooled mutants from the final library

Primary pools: Once the final library of sequenced, barcoded insertion mutants was assembled, frozen stocks of pools of the entire collection were made to allow screening of all of the mutants at once. The primary pool was made from a copy of the final library in 96-well plates grown in YES + G418 + FOA + 3% glucose at 32°C until all wells had grown to saturation. Cells from each well were harvested with a multichannel pipettor as described earlier, and transferred to a 2-l Erlenmeyer flask. YES + G418 + FOA + 3% glucose medium was added to 1 l, and the culture was grown in a shaking incubator overnight at 32°C. Cell density in the saturated culture was determined by optical density, and cells were pelleted and resuspended at 10^9 cells/ml in YES + G418 + FOA + 3% glucose. The final suspension was brought to 15% glycerol and 80–1 ml aliquots were frozen in freezer vials at −80°C.

Secondary pools: A 1 ml aliquot of the primary library pool was amplified to create multiple stocks to be used in screening for different colony phenotypes. A single 1 ml frozen primary pool aliquot was thawed on ice in a 4°C room for 15 min, and the cells transferred to 50 ml of YES + 3% glucose precooled to 4°C in a 250-ml flask. The flask was placed in a room temperature shaking water bath that was set for 32°C at 170 rpm for 5.5 h. Two 25 ml aliquots of this culture were diluted into 500 ml of YES + 200 μg/ml G418 prewarmed to 32°C and grown in a 32°C air shaker for 19 h to a density of 6×10^7 cells/ml by OD_{600}. A 10 μl aliquot of the 1 l of cells was taken, diluted into YES + 3% glucose and dilutions were plated to determine cell viability (which was 5.1×10^7 cells/ml). The 1 l of cells were then pelleted in four sterile 250-ml centrifuge bottles, resuspended in 10 ml of YES per bottle and transferred to sterile 50-ml tubes. Cells were pelleted, the supernatant discarded and resuspended in 10 ml of YES + 3% glucose + 15% glycerol per tube. The cells were pelleted, the supernatant discarded and resuspended in 10 ml of YES + 3% glucose + 15% glycerol per tube again and the 40 ml of cell suspension was pooled. This secondary amplification was divided into 40–1 ml aliquots ($\sim 1.3 \times 10^9$ cells/ml) in freezer vials and stored at −80°C.

Additional strain constructions

Schizosaccharomyces pombe strain modifications were carried out using standard methods. For fusion of the right

end and left end small open reading frames (ORFs) of *Hermes* to the *ade7*$^{+}$ ORF, 100-mer oligonucleotides with 75 bp of sequence identity to the *S. pombe* genome were used to amplify *Hermes::kanMX* from the library constructed in pHL2577 (primers a7HLE_S + A7HLE_AS for the left end and a7HRE_S + a7HRE_AS for the right end; see Supplementary Table S4). The homology to the genome was further extended by reamplification of each product with the primers A7HRLplus_S and A7HRLplus_AS_2. The PCR product was transformed into the *ade7*$^{+}$ strain (KRP387 *h*$^{-}$ *ura4-D18 leu1-32 his3-D1 arg3-D4*), selecting transformants using the kanMX marker. The correct transformants were verified by PCR using primers in *Hermes::kanMX* and the genomic sequence not present in the PCR product. The *ade7Δ* control is a strain bearing a complete *ade7* ORF deletion (KRP389, which is KRP387 but *ade7Δ::arg3*$^{+}$). For transfer of the camptothecin (CPT) resistance insertion, primers were used to amplify the *Hermes::kanMX::barcode* insertion from the genome of the resistant strain with flanking genomic DNA, and the PCR product was transformed into KRP201, selecting for the kanMX marker. Transformants were validated by colony PCR and then tested for CPT resistance.

RESULTS

Construction of a barcode-tagged *Hermes* transposon insertion mutagenesis library in *S. pombe*

An arrayed collection of sequenced, barcoded insertion mutants can greatly enhance genetic investigation of cellular processes. *Schizosaccharomyces pombe* lacked a library of viable insertion mutants. Insertions into the *S. pombe* genome can be made efficiently with the *Hermes* transposon (35). *Hermes* has been adapted to a two-plasmid system where one plasmid expresses the transposase while the other bears a modified transposon containing a selectable marker (32). Insertion of this transposon into a coding exon is predicted to disrupt gene function as the three reading frames would reach a stop codon after 31 (TAA), 75 (TGA) or 44 (TAA) bases on the right end of *Hermes*, and after 28 (TGA), 264 (TAA) and 95 (TGA) bases on the left end. The goal of this project was to produce a collection of viable insertion mutants with sequenced barcodes in unique *S. pombe* genes.

A genome-wide mutant library for phenotypic screening requires a method that allows one to monitor the relative growth of each individual mutant in mixed cultures containing all of the mutants in the collection. DNA barcodes that uniquely tag individual gene deletions in the *S. cerevisiae* (5) and *S. pombe* (4) deletion strain sets enable phenotypic analysis of the whole collection in pooled competitive growth assays (4,36–39). We therefore devised a strategy to tag each *Hermes* transposon with a unique DNA barcode.

We designed a library of DNA barcodes containing 27 random nucleotides, encoding up to 4^{27} possible barcodes. This large number of variants meant that a collection of several thousand mutants would almost certainly all have unique barcodes. These DNA barcodes were flanked by SfiI sites, so pools of barcodes can be oligomerized for sequencing in the absence of next-generation sequencing facilities as before (21). The barcode library was cloned into the *Hermes* transposon vector and transformed into *Escherichia coli DH5α* to produce ∼1–2 × 10^{5} bacterial colonies per transformation. Ten transformations were performed to generate 10 barcoded *Hermes* transposon plasmid libraries. Each library was used to generate ∼1000 *S. pombe* insertion mutants (Figure 1) with a >99% probability that each barcode was unique.

We previously established a method of efficiently generating single *Hermes* transposon insertions in *S. pombe* by modifying the system of Evertts *et al*. (Figure 1A) (32,35). The modified *Hermes* transposon bore *kanMX6*, which allowed selection of integration events by G418 resistance, and contained *URA3* as the marker on the plasmid backbone. Expression of the transposase was driven by the inducible *nmt1* promoter on a plasmid that we altered to contain the *ura4*$^{+}$ marker. Transformants that had lost both plasmids could then be selected on medium containing 5-FOA (40). *Schizosaccharomyces pombe* cells bearing the transposase plasmid were grown under inducing conditions, and then transformed with barcode-tagged transposon plasmids. Cells were allowed to grow for only two to five divisions to allow transposition, and then transferred to YES + G418 + 5-FOA to select for cells bearing a transposon integrated into the genome and against both plasmids. Surviving cells were picked and placed into 96-well plates. A total of 9024 mutant strains were picked into ninety-six 96-well plates (Figure 1A).

Development of a novel 3D pooling and high-throughput multiplexed sequencing strategy to map transposon integrations and DNA barcodes

Sequencing previous insertion libraries was costly and labor intensive (17). We developed an innovative approach that used 3D pooling and leveraged results from next-generation sequencing to identify the unique insertion sites of each transposon insertion. Cells were processed in groups of twenty-four 96-well plates for sequencing on the Illumina platform. Each plate contained 94 strains with two empty wells uniquely spaced to identify each plate. Each pool was constructed as a stack of 24 plates, with 2 plates in each layer and a total of 12 layers (Figure 2A, stack of plates). Each row pool consists of wells from the same row from each layer of plates (Figure 2A, row pools). Each column pool consists of wells from the same column from each layer of plates (Figure 2A, column pools). Finally, each layer pool consists of all of the wells in each layer of plates (Figure 2A, layer pools). Each layer therefore contained 12 rows and 16 columns (Figure 2A). Cells were collected as 16 column pools, 12 row pools and 12 layer pools. These 40 pools contained three copies of each mutant in a different row, column and layer pool. The 3D coordinates of an identified mutant are determined by sequence comparisons between each row, column and layer pool in the 24-plate sublibrary (Figure 2A). A total of four groups of plates were processed.

The 3D pooling, high-throughput sequencing strategy utilized three important principles. First, each pool was constructed with a unique sequence or customized index tag, which identifies the sequences associated with that pool (Supplementary Table S3). Second, each pool was amplified in three ways with primers specific to the *Hermes* left or

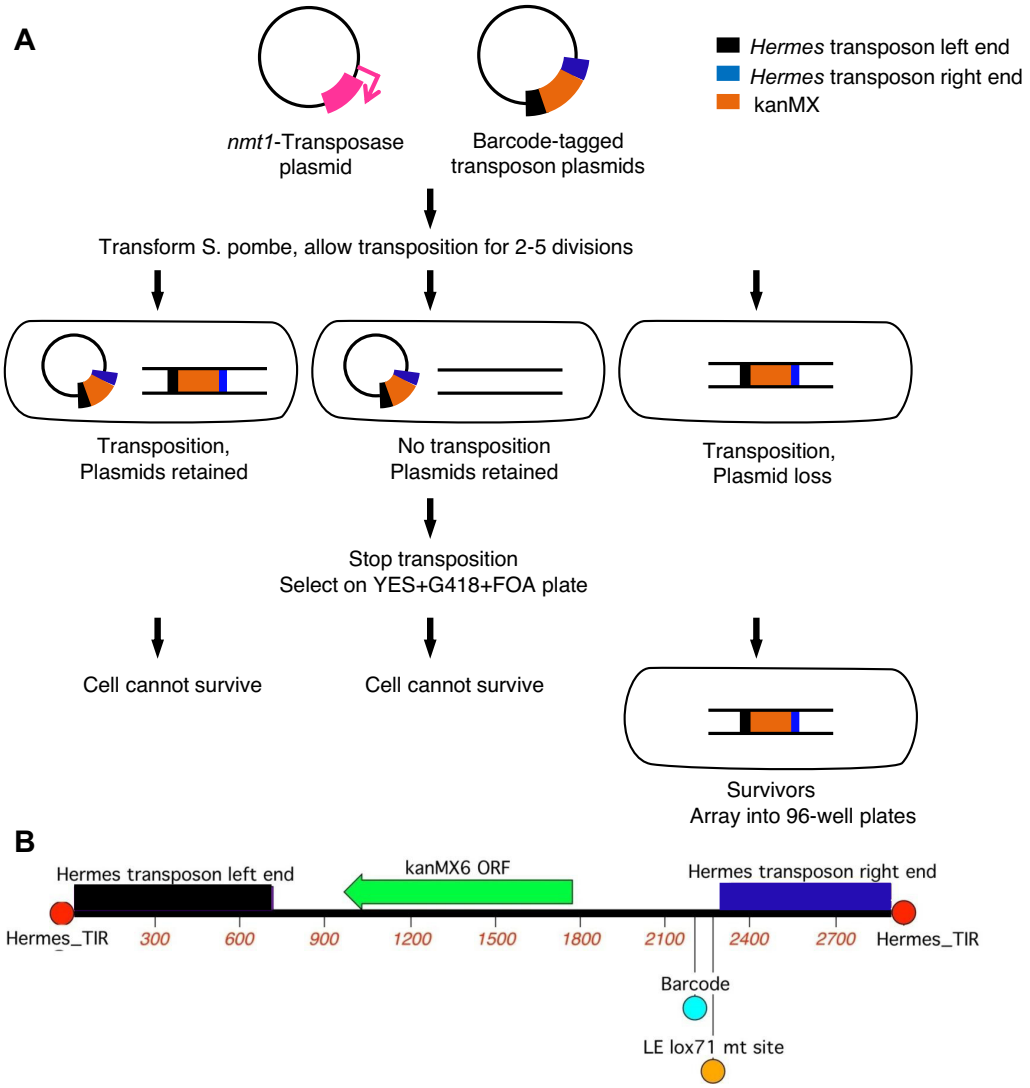


Figure 1. Overview of *Hermes* library construction. (**A**) Transposase is expressed from the *nmt1* promoter, which is active on minimal medium (EMM) but not on rich medium (YES). The transposase plasmid was transformed into *S. pombe* cells grown on EMM to preload cells with transposase. These cells were then transformed with the transposon plasmid library, and transformants were plated on EMM plates and grown for two to five divisions to allow transposition. Cells were then replica plated to the YES + G418 + 5-FOA plate to stop transposition. The transposase and transposon plasmids contained the *ura4*$^{+}$ or *URA3* marker, respectively. Cells that contained either plasmid were killed on 5-FOA medium. Therefore, only cells with a transposon inserted into the genome that did not retain the transposon or transposase plasmid could grow on the YES + G418 + 5-FOA plates. A total of 96 plates (9024 mutants) were picked into 96-well plates. (**B**) The architecture of the *Hermes* transposon. TIR means terminal inverted repeats. LE lox71 mt site is a mutant loxP site that allows Cre recombinase-mediated integration of single plasmids bearing the lox66 site (62).

right ends and to the 27-bp barcode (Figure 2B). The combination of the pool-specific index tag and specific primer sequences identified the pool and the type of sequence (barcode or *Hermes*–genome junctions). We therefore used 100-bp single-end sequencing reactions that captured the index tag, the unique primer sequence and over 41 bp of barcode or 30 bp of genomic DNA. Third, a customized triangulation script first parsed the sequences into individual row, column or layer pools, and then subdivided each pool into sequences specific for the barcode or *Hermes* right or left end. The sequences in each row pool were then compared against all of the column and layer pools to identify the individual well in the array of twenty-four 96-well plates that contained all three sequences (Figure 2C). For example, a barcode sequence from row pool 1 would be compared to the barcode sequences from all of the column pools and layer pools. This comparison led to the identification of a single column pool and single layer pool that contained this individual sequence. The location where these three pools intersected identified the individual well containing this barcode sequence (e.g. the red dot in Figure 2A). The mapping of genomic sequences adjacent to the *Hermes* right and left ends identified the genomic location of the insertion and provided an internal check that these two independent sequencing reactions had identified the same genomic locus.

To determine the number of mutants to analyze in each sequencing reaction, we calculated the number of sequences in each pool to obtain a sufficient average number of se-

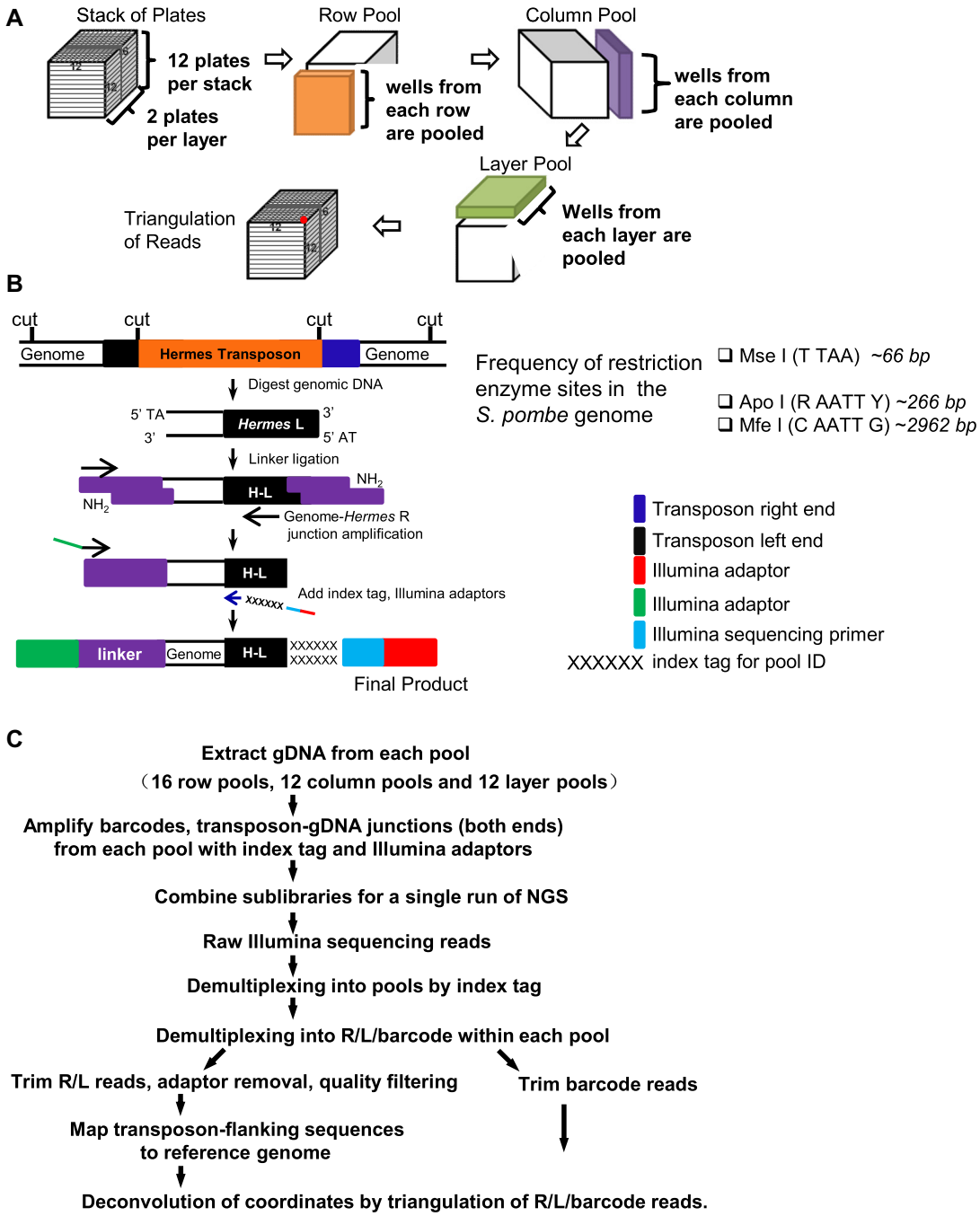


Figure 2. 3D pooling and multiplexed deep sequencing to map transposon integrations and DNA barcodes. (**A**) 3D pooling. Three copies of 24 plates (2256 mutants) were stacked as 2 plates per layer for a total of 12 layers. Cells were collected in the format of a row pool (144 strains/pool), a column pool (192 strains/pool) and a layer pool (192 strains/pool). For each row, column or layer pool, 50 μl of cells were collected with a multichannel pipettor, pooled and DNA was extracted (see the 'Materials and Methods' section). A total of 40 pools of cells were collected, including 16 column pools, 12 row pools and 12 layer pools. The red dot is described in the main text. (**B**) Amplification of *Hermes* insertion sites. The *Hermes* right side is shown as an example. Genomic DNA from a row, column or layer pool was fragmented by restriction enzyme digestion, with the predicted average fragment size shown. Double-strand DNA linkers with overhangs compatible with MseI, ApoI and MfeI were ligated to digested genome fragments. The linkers were synthesized with amine groups at the 3′ end to prevent self-ligation. The first-round PCR utilized a linker primer and a primer that specifically annealed to 19 bp of a unique *Hermes* border sequence, just inside the terminal inverted repeats. The second-round PCR re-amplified and enriched the genome–*Hermes* fragment with a nested transposon primer. To index the pools, primers were synthesized with 8-mer tags. Illumina adaptors and sequencing primers were added to the final products. The same approach was adapted to *Hermes* left end using different specific primers and index tags. The barcode tags were amplified as described in Supplementary Figure S1. (**C**) Analysis pipeline. After sequencing, flanking genomic sequences at both ends of the transposon were aligned to the *S. pombe* reference genome to map the insertion. Each intersection of insertion points from a row pool to a column pool to a layer pool decoded one mutant and identified its location in a 96-well plate. The same triangulation program was applied to decode the DNA barcodes (Supplementary Figure S2). The independent right end and left end reads served as an internal validation of the decoded transposon insertion sites.

quences to define the product. We chose 500 sequence reads per product because some products might amplify poorly and be underrepresented in the final sequencing reaction. Also, an average of 500 sequences would allow the acquisition of a sufficient number of sequences to identify the majority of products. We processed four groups of twenty-four 96-well plates. Each group has 2256 mutants, each mutant has three associated sequences (barcode and *Hermes* left and right ends with adjacent genomic DNA) and each associated sequence was performed three times (once in each row, column and layer pool). Therefore, ∼20 000 distinct products (2256 × 3 × 3) were expected. The 40 row, column and layer pools were processed in 40 individual PCR reactions to create DNA libraries for sequencing on the Illumina HiSeq 2500, which can output ∼10 million reads. This sequencing depth allowed each product to be read about at least 500 times for each group of twenty-four 96-well plates.

Ligation-mediated PCR was employed to amplify transposon-flanking DNA sequences (Figure 2B) (41–43). Genomic DNA from each pool was digested by different restriction enzymes, MseI and ApoI/MfeI, to increase the chance of capturing appropriately sized flanking genomic DNA fragments for sequencing libraries. After the ligation of ds linkers, DNAs from the same pool were mixed together and the transposon-flanking genome sequences were amplified. To link the amplified products to the pools where they originated, a unique 8-mer index tag sequence for each pool was included in the PCR primers. All index tags differed by at least two nucleotides so that the chances of mis-sorting due to a sequence miscall were minimized. Finally, two Illumina adaptors were incorporated, and the products could be directly sequenced using the Illumina platform. The DNA barcodes were directly amplified from pools of genomic DNA (Supplementary Figure S1), and the index tags and Illumina adaptors were sequentially incorporated by multiple rounds of PCR as described earlier. All PCR products were pooled and sequenced in a single lane on the Illumina HiSeq 2500. The entire insertion mutant collection of 9024 mutants assembled in 96 plates was sequenced in a total of four lanes to yield 100-bp single-end reads (examples of typical sequence reads are shown in Supplementary Figure S1B–D).

Sequence identification and assignment to individual strains

We developed a customized bioinformatics pipeline to decode transposon insertions and DNA barcodes using the 3D pooling strategy (Figure 2C and Supplementary Figure S2). The raw sequence data were trimmed of adaptor sequences and then sorted by the index tags into 40 collections of sequences using Novobarcode (Novocraft Technologies). These data were further sorted by the consensus reads into DNA barcode, and transposon left end or right end flanking sequences, also using Novobarcode. Genomic sequences were mapped onto the *S. pombe* reference genome using the Bowtie algorithm. A customized script triangulated the row, column and layer pool reads to assign the barcodes and integration sites to individual strains. We manually examined the barcodes and integration sites from 20 randomly chosen strains and found that all assignments agreed with the high-throughput mapping.

The pipeline successfully decoded insertion sites in unique regions of the genome. However, the triangulation script required unique sequences to map individual mutants to a specific microtiter plate well. Consequently, if the insertion site was within repetitive sequences such as the centromere, subtelomere, rDNA repeats and mating-type region, the insertion mutant could not be mapped to a well within the group of 96-well plates and was not reported. Strains with insertions in repeated DNAs were present in the original set of transposon insertion mutants. The genomic sequences from the row, column and layer pools were compared to the *S. pombe* genome by BLAST, and repeated DNA sequences were detected, but were not mapped to individual microtiter plate wells. The number of insertions in repetitive DNA in the original 9024 mutants is therefore unknown. We note that mapping an insertion would also be compromised if the sequence of a sample in a row, column or layer pool was of poor quality, or did not match the reference genome sequence.

Schizosaccharomyces pombe mutations generated from random insertions of the *Hermes* transposon

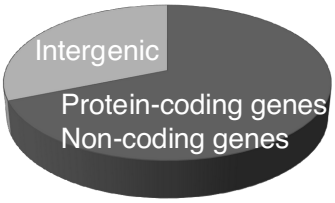
We successfully mapped *Hermes* transposon integration sites and the associated DNA barcodes in 4095 *S. pombe* mutants out of the 9024 mutants sequenced. A total of 4391 distinct insertion sites were recovered, as a fraction of the mutants contain more than one transposon insertion (Supplementary File S1). Over 90% of strains in the current collection carried a single transposon insertion, and ∼70% of transposon insertions were in protein-coding genes and noncoding RNA genes. The remaining 30% of the insertions were in the intergenic regions (Figure 3A and Supplementary File S1), as defined by the *S. pombe* genome database as of May 2013 (44). The frequency of insertion in each chromosome was proportional to chromosome size (Figure 3B). While *Hermes* insertion has a strong requirement for TNNNNA, the *S. pombe* genome is 70% A/T and so contains a large number of sites, consistent with nearly random integration of *Hermes* observed.

The *Hermes* transposon did show a bias for inserting into untranslated regions (UTRs). Of the 2753 insertions in protein-coding genes, 38% (1057) were in the coding exons and introns, and 62% (1696) were in the UTRs. The aggregate size of *S. pombe* UTRs (∼3.7 Mb) was much smaller than coding exons and introns (∼7 Mb), but significantly more UTR insertions were recovered compared to the number of insertion distributions per kb of coding exon and intron ($P < 0.01$) (Figure 3C).

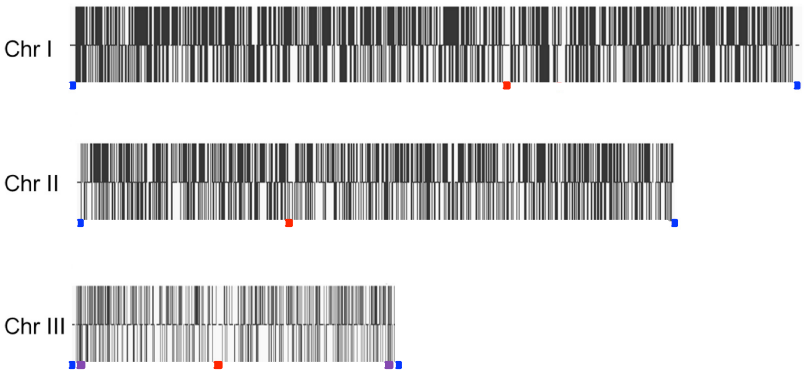
The distribution of *Hermes* insertions is different among essential genes and nonessential genes. About 59% of insertions into nonessential gene mapped to UTRs and 41% mapped to CDS. The insertions in essential genes were more enriched in UTRs (87%). Only 13% were in the protein-coding regions (Figure 4A). We further analyzed the insertions within the first and last 150 bp of CDS regions as well as in the remaining sequence in the coding regions in the essential and nonessential genes. There were more insertions in the middle of the CDS of nonessential genes than from essential genes (Figure 4B). The 47 *Hermes* insertions in the coding regions of essential genes are shown in Figure 4C.

A

No. of strains sequenced	9,024 (96 plates)
No. of strains mapped in unique region	4,095
No. of insertion sites	4,391
within genes (protein coding, non-coding genes)	3,022 (69%)
in intergenic regions	1,369 (31%)



B



C

Protein-coding gene mutations	**2,753**
UTR mutants	1,696
5'UTRs	1,113
3'UTRs	583
Coding Exon mutants	1,000
Intron mutants	57

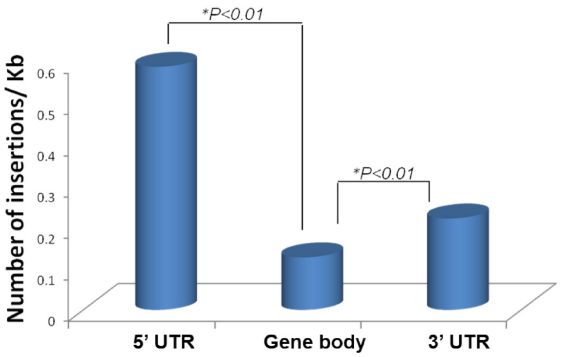


Figure 3. *Hermes* transposon collection. (**A**) *Hermes* transposon library statistics. A total of 9024 mutants were sequenced, and 4095 mutants were successfully mapped to the *S. pombe* genome. A total of 4391 *Hermes* insertions were recovered from these 4095 mutants. About 70% of the *Hermes* insertions were in protein-coding genes or noncoding RNA genes. The remaining 30% were in the intergenic regions. (**B**) *Hermes* distribution on *S. pombe* chromosomes. The frequency of insertion in each chromosome was proportional to chromosome size. Each black line represents a *Hermes* insertion. The two orientations of the insertions were represented by upward or downward lines. Red bar, centromere. Blue bar, telomere. Purple bar, site of the rDNA arrays (1225 kb in size at the left end and 240 kb at the right end, each annotated in the genome as one repeat). (**C**) UTR mutants were enriched in the collection. The bar graph was plotted by a number of *Hermes* insertions per kb of UTRs or gene body (coding exons and introns of protein-coding genes). Chi-square statistics were used to compare differences between groups. A difference was taken as significant when *P*-value was <0.01.

The insertions in the first 150 bp of the CDS of several essential genes in these viable strains were surprising. We hypothesized that the viability of these cells might be due to a cryptic promoter activity from the *Hermes::kanMX* insertion and/or fusions between small ORFs at the end of *Hermes* and genomic sequences. Examination of the terminal 120 bp of each end of *Hermes* revealed an ORF of 31 amino acids (aa) plus 1 nt at the left end, and a 13 aa + 2 nt ORF at the right end (Figure 5A). The *Hermes* insertions in the first 150 bp of these essential genes were predicted to make fusion proteins that contain the majority of the essential gene products (not shown). We therefore tested whether the ATG codon of the *Hermes* short ORF could support expression of a reporter gene, *ade7*$^{+}$. The *ade7*$^{+}$ gene was chosen because it is a small housekeeping gene that lacks introns or extensive post-transcriptional processing. Cells that do not express *ade7*$^{+}$ cannot grow on medium lacking adenine and form red colonies when adenine is limiting, while cells with low levels of *ade7*$^{+}$ expression grow slowly and form pink colonies under these two respective conditions (Figure 5D) (45–47).

Hermes::kanMX was inserted at the second codon of the *ade7*$^{+}$ gene such that the entire transposon including the ATG from the left or right end short ORF was in frame

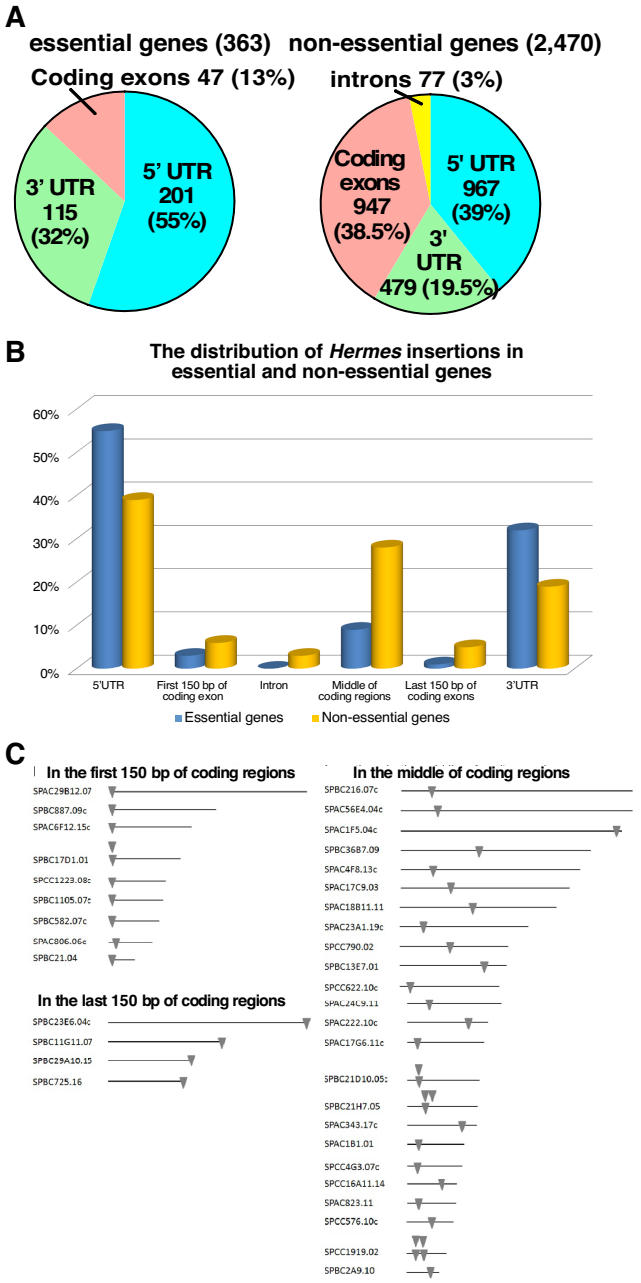


Figure 4. The distribution of *Hermes* insertions in essential genes and nonessential genes. (**A**) *Hermes* insertions were enriched in the UTRs of essential genes. The table and chart show total number of insertions in the 5′ UTR, 3′ UTR, and coding exons and introns in the essential genes and nonessential genes. (**B**) The comparison of *Hermes* insertion distribution in essential genes and nonessential genes. The coding region of each gene was divided into three parts: the first 150 bp, the last 150 bp and the regions in between (middle of coding regions). The number of insertions was plotted for each part. The lengths of *S. pombe* gene coding regions were downloaded from PomBase. (**C**) The distribution of *Hermes* insertions in essential genes. *Hermes* insertions in the coding region of essential genes are shown. Line, coding region of *S. pombe* genes. Triangle, *Hermes* transposon. Different strains with insertions in the same gene are indicated by multiple triangles over one line.

with the remaining *ade7*$^{+}$ CDS (Figure 5B). As the Ade7 N-terminus is predicted to have strong structural interactions with the full protein (48,49), the additional amino acids encoded by the short ORFs were not added (Figure 5C). Both left end and right end small ORF fusions supported *ade7*$^{+}$ expression, as cells bearing these fusions could grow on medium lacking adenine, in contrast to cells lacking the *ade7* gene (*ade7Δ*, Figure 5E). Left end +ATG-*ade7*$^{+}$ fusions grew faster than right end +ATG-*ade7*$^{+}$ fusions on medium lacking adenine and the right end fusions were slightly more pink than the left end fusions, suggesting differences between the two *Hermes::kanMX* ends in driving expression (Figure 5E).

As a control for the left end +ATG-*ade7* and right end +ATG-ade7 fusions (*LE+ATG-ade7* and *RE+ATG-ade7*, respectively), similar fusions were constructed where the ATG of the short ORFs was mutated to TTC. The entire *ade7* ORF and 100 bp of flanking sequence in these –ATG alleles were PCR amplified and sequenced to confirm that the ATG to TTC conversions were the only mutations present in and near *ade7*. For two independently isolated strains of each allele lacking an ATG, the *LE–ATG-ade7* and *RE–ATG-ade7* showed the same growth phenotypes as the *ade7Δ* strain, indicating that the –ATG constructs did not support *ade7* expression (Figure 5E and Supplementary Figure S4). Consequently, an in-frame fusion to the short ORFs at the ends of *Hermes* can provide a start codon for an in-frame fusion protein, and provides an explanation for how the insertions in the first 150 bp of essential genes produced viable cells.

These results suggest that the insertions in the first 150 bp of the CDS of the essential genes allow cell viability by means of a cryptic transcriptional activity in *Hermes::kanMX*, which may include the production of protein fusions with the small ORFs at the ends of *Hermes*. These considerations suggest that some *Hermes::kanMX* insertions may produce truncated products that may yield additional mutant phenotypes.

The heterochromatic centromere and telomere regions contain unique regions that could be mapped, but these transcriptionally silenced regions also had fewer identified transposon insertions. We recovered only four insertions within centromeres and five from chromosome I and II telomere regions (Supplementary Figure S3A and B). This result most likely reflects silencing of the *kanMX* gene, which impairs the selection for G418 resistance.

Strains bearing the 4391 mapped insertions were transferred to a final set of approximately forty-four 96-well plates, although a small fraction did not regrow and were lost. Our final *Hermes* insertion collection contains 4308 insertions with mutations in 268 essential genes (21% of *S. pombe* annotated essential genes), 1472 nonessential genes (41% of nonessential genes), 589 noncoding genes (31% of noncoding genes) and 1369 intergenic sites (Table 1, *Hermes* library). Several genes have multiple *Hermes* insertions at different sites in different strains, with a total of 363 essential gene insertions, 2470 nonessential gene insertions and 1159 noncoding gene insertions and the remaining 1369 insertions in regions with no identified genes. Some strains have more than one insertion and some insertions affect

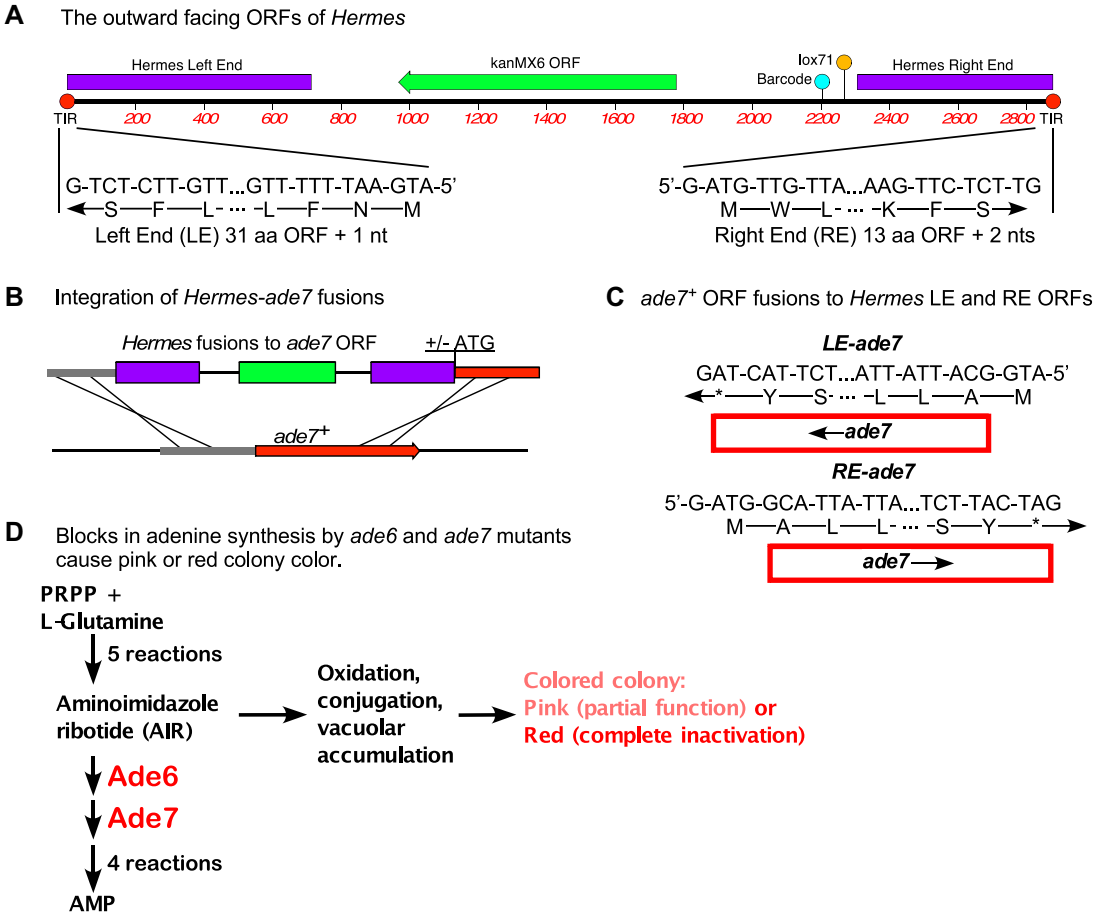


Figure 5. The *Hermes::kanMX* insertion can express gene fusions under unique circumstances. (**A**) A schematic of the *Hermes::kanMX* transposon showing the small ORFs that start at each end of the transposon and extend into the adjacent genomic DNA. TIR signifies the 17-bp terminal inverted repeats. (**B**) A schematic of the *Hermes-ade7* constructions that fuse the ATG of the *Hermes* small ORFs [in panel (A)] to the second codon of *ade7*$^+$ at the *ade7* genomic locus. *Hermes*, with and without the short ORF ATGs, is fused to *ade7* in both orientations. (**C**) An abbreviated diagram of the ORF formed by inserting the entire *Hermes::kanMX* transposon at *ade7* such that the methionine (M) of *ade7*$^+$ is replaced by the methionine of the short *Hermes* ORF from the left end (*LE+ATG-ade7*) or right end (*RE+ATG-ade7*). Because the N-terminus of the Ade7 protein forms part of the structure of the final protein (48,49) (https://alphafold.ebi.ac.uk/entry/Q9UUB4), the additional amino acids of the short ORFs were not included in the fusion. In the –ATG constructions, the *Hermes*-derived ATG is mutated to TTC. (**D**) A diagram of the adenine biosynthetic pathway showing the Ade6 and Ade7 steps where mutation results in colored colonies. Loss or reduction in Ade6 or Ade7 enzyme activity allows the accumulation of the AIR intermediate, which is subsequently oxidized, conjugated to glutathione or amino acids and concentrated in the vacuole to result in a colored colony (45–47). When grown on medium with limiting adenine, loss of activity causes formation of red colonies while reduced function causes formation of pink colonies. (**E**) *Hermes::kanMX* right end and left end ORF fusions can support gene expression. Cells lacking the *ade7*$^+$ gene (*ade7Δ*, KRP389) cannot grow on synthetic medium lacking adenine and form red colonies on medium with limiting adenine. A representative left end-*ade7* fusion with the *Hermes*-derived ATG (*LE+ATG-ade7* in KRP387) grew as well as *ade7*$^+$ cells form white or light pink colonies on medium with limiting adenine, showing that the *ade7* ORF fusion is expressed. A representative right end-*ade7*$^+$ (*RE+ATG-ade7* in KRP387) also showed expression, but the smaller colonies on medium lacking adenine and pink color in medium with limiting adenine suggest that *ade7* expression is reduced compared to the left end fusions and the wild-type *ade7*$^+$ gene. In contrast, the left end and right end fusions where *Hermes*-derived ATG was mutated to TTC (*LE–ATG-ade7* and *RE–ATG-ade7*) had the same phenotypes as the *ade7Δ* cells. Color balance and contrast of the limiting adenine pictures was adjusted to highlight the difference between white and pink colony color (48). Analysis of independently constructed +ATG-*ade7* and –ATG-*ade7* strains is shown in Supplementary Figure S4.

Table 1. Comparison of *Hermes* library with the Bioneer library

	Essential genes	Nonessential genes	Noncoding genes	Intergenic insertions
Schizosaccharomyces pombe	1260	3576	1876	
LR *Hermes* library (total mutants)	363	2470	1159	1369
LR *Hermes* library (unique genes)	268	1472	589	1369
Haploid ORF deletion library [4]	0	3308	N/A	0

Gene numbers include insertions that disrupt overlapping protein-coding and noncoding RNA genes as well as genes with more than one insertion.

more than one gene (genes and their associated insertions are summarized in Supplementary File S2).

Phenotypic characterization of the *S. pombe* insertion mutants

A major advantage of the insertion mutant library is the presence of a wider variety of phenotypes compared to a deletion library, as demonstrated by the mutants in essential genes that we isolated. To determine whether these insertion mutants showed a range of phenotypes, we examined mutants in two phenotypic categories: the growth of *S. pombe* cells on nonfermentable carbon sources (2% glycerol, 2% ethanol) and resistance to the topoisomerase inhibitor CPT. Normal growth on nonfermentable carbon sources requires an intact mitochondrial respiratory chain for carbon metabolism (50). Thus, mutants in respiratory chain complex genes are expected to have impaired growth on nonfermentable carbon sources. There were 16 insertion mutants in 15 genes in the library categorized under the GO term 'respiration chain complexes I, II, III, IV, V and assembly proteins' (Supplementary Table S1). All three mutants with insertions in coding exons, two mutants with insertions in introns and three mutants with insertions in the 5′ UTR showed defective growth (Figure 6A). In contrast, six UTR mutants (three each in the 5′ and 3′ UTRs) showed normal growth (Figure 6B and Supplementary Table S1). Thus, our insertion library can identify regions of the UTRs important for gene function.

CPT is a topoisomerase inhibitor causing replication fork breakage when the replisome encounters the topoisomerase–CPT–DNA adduct (51,52). Deshpande *et al.* screened 2662 *S. pombe* complete or partial ORF deletion mutants for growth on plates containing CPT and identified a set of 119 CPT sensitivity genes (51). We searched our insertion library for mutations in the CPT sensitivity gene set and found 54 mutants with insertions in 37 genes (Supplementary Table S2). We tested these mutants for growth on different concentrations of CPT.

Many of the insertion mutants showed sensitivity to either low or high concentrations of CPT, including 17 mutants with insertions in coding exons, 1 with an insertion in an intron and 14 with insertions in the 5′ or 3′ UTR. One coding exon mutant and 20 UTR insertion mutants showed no change in CPT sensitivity, thus identifying regions of the 20 genes that are dispensable for CPT resistance (Figure 7A, Supplementary Figure S5 and Supplementary Table S2). We observed 19 insertion mutants that were more sensitive to CPT than wild-type cells and were less sensitive

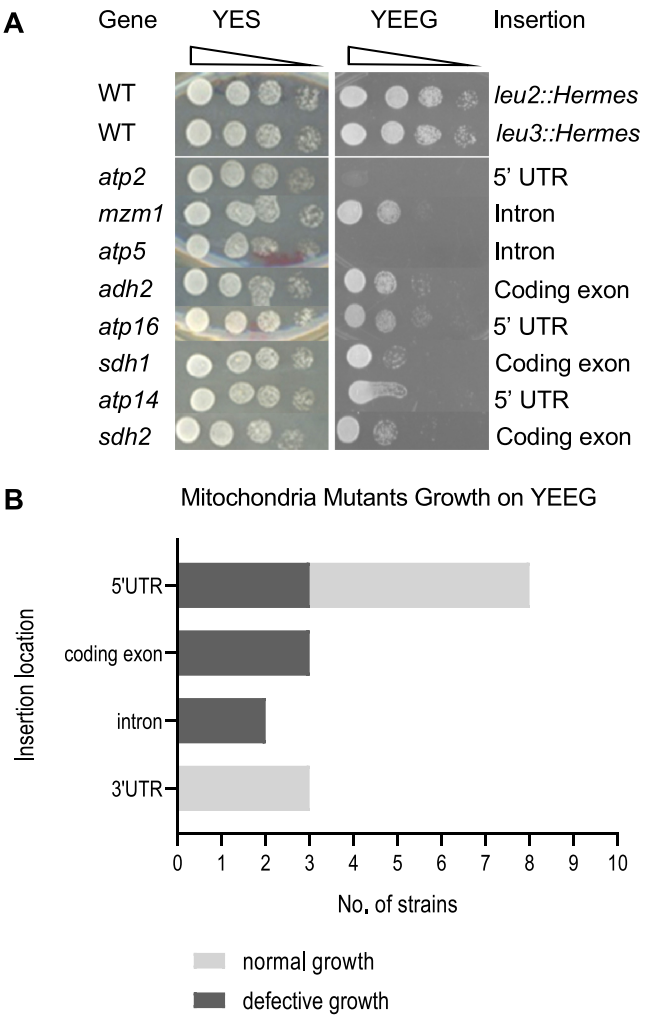


Figure 6. Defective growth of *Hermes* respiratory chain mutants on nonfermentable carbon sources. (**A**) Spot test *Hermes* mutants on YES (fermentable carbon source) and YEEG (nonfermentable carbon source). The mutants with impaired growth are shown. The *S. pombe* background strain is *leu1-32,* and the *leu2::Hermes* and *leu3::Hermes* (both *leu*$^{-}$) have the same growth characteristics as the original wild-type strain. (**B**) *Hermes* insertions in respiration chain complex genes show different effects on nonfermentable carbon source. A summary of the insertion location and number of mutants with normal or defective growth on YEEG is shown.

than the corresponding deletion mutant, showing that our insertion mutants had distinct phenotypes from the deletion library.

Importantly, one 5′ UTR mutant displayed a CPT-resistant phenotype (Figure 7B). This mutant bore a transposon insertion in the 5′ UTR of *SPBC16A3.17c*, 1278 bp

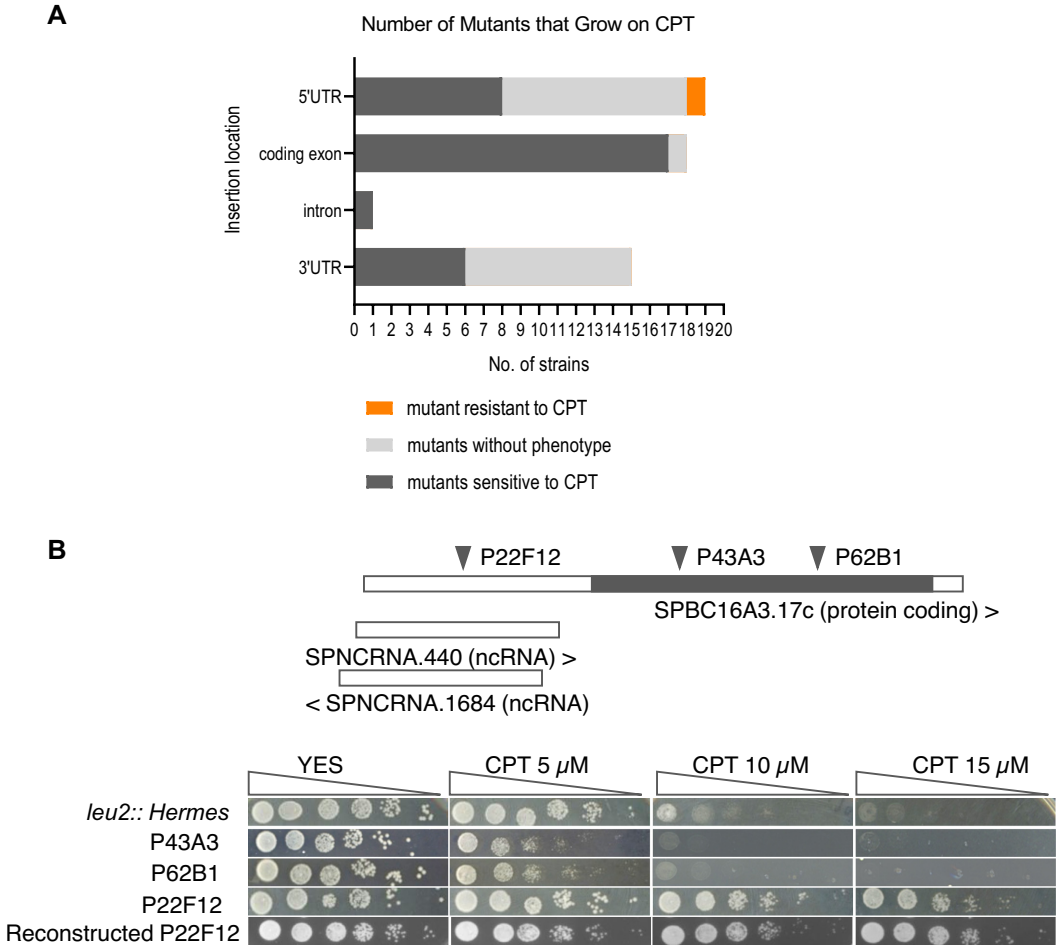


Figure 7. (**A**) The *Hermes* insertion mutants refine our understanding of CPT resistance genes. The 54 mutants bearing insertions in 37 genes required for CPT resistance identified 32 CPT-sensitive strains with phenotypes similar to the gene deletion, 21 with insertions in the coding exons, introns and UTRs with no phenotypes and 1 insertion in the 5′ UTR with the resistance phenotype (Supplementary Figure S5 and Supplementary Table S2). The different insertion phenotypes identify gene regions required and dispensable for CPT resistance. (**B**) The 5′ UTR insertion mutant is resistant to CPT. The SPBC16A3.17c gene was disrupted by three different transposon insertions, one in the 5′ UTR (strain P22F12) and two in coding exons (strains P43A3 and P62B1). Spot tests on 5, 10 and 15 μM CPT plates showed that the insertions in the gene body were CPT sensitive while the 5′ UTR insertion and reconstructed mutant were more CPT resistant than wild-type cells. Please note that all strains are *leu*$^{-}$ due to a background *leu1-32* mutation, and the wild-type strain that carries a *Hermes* insertion in the *leu2* gene has the same phenotype as the wild-type progenitor strain used to make the library.

upstream of the start codon. Two functionally unknown noncoding RNA genes overlapping with the 5′ UTR were disrupted by transposon insertion. In contrast, two separate insertions in coding exons of the same gene were sensitive to CPT as expected, showing that the 5′ UTR insertion has a novel phenotype. To exclude the possibility that the CPT-resistant phenotype was generated from unrelated genomic mutations, we reintroduced the 5′ UTR transposon insertion into a wild-type background. The resulting new mutants were still resistant to CPT (Figure 7B, reconstructed P22F12). Therefore, this novel CPT resistance phenotype was due to the transposon insertion. We noted that *SPBC16A3.17c* encodes a transmembrane transporter, orthologous to the *S. cerevisiae AZR1* gene. Different levels of *AZR1* expression in *S. cerevisiae* can cause CPT sensitivity or resistance (53,54), consistent with the different phenotypes observed in our SPBC16A3.17c insertion mutants.

Identification of mutated genes from a screen of pooled mutants using the *Hermes::kanMX*-associated barcodes

To demonstrate the utility of the random barcodes associated with the mapped *Hermes::kanMX* insertions, we constructed a mixed pool of the final library and used it to screen for phenotypes of individual colonies. Colonies with the desired phenotype were picked, and the barcodes of each colony were amplified and sequenced. The barcode sequences were then used to identify the mutation. We used two screens: one in which the desired mutation producing the phenotype was known, and a second screen that gave unexpected results but still identified a mutation expected to cause the phenotype.

The first screen used pink–red–white colony color to determine whether the two known *ade6::Hermes::kanMX* mutants in the library could be identified. The strain used to make the insertion library bears *ade6-M216*, a missense mutation that reduces Ade6 function to produce a pink

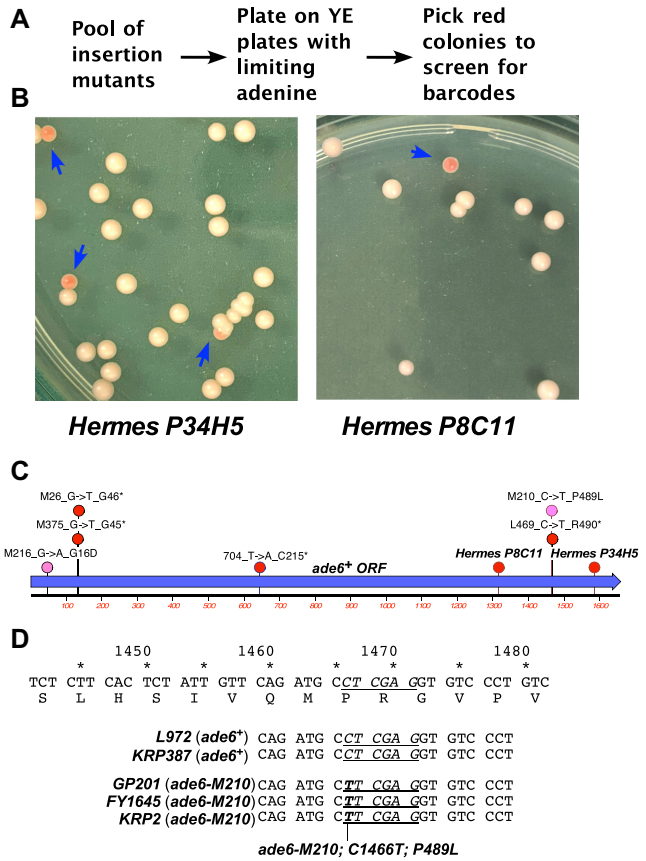


Figure 8. Identification of *ade6::Hermes::kanMX* mutants from a pool of mutants using the insertion-associated barcodes. (**A**) A schematic of the mutant screen. A pool of the insertion mutants was plated on YE medium with limiting adenine to form ∼6000 colonies. Two dark red colonies among the background of pink colonies were identified, and the barcode DNA was amplified from each colony, sequenced and compared against Supplementary File S1 to identify the insertion site. Each colony identified a different insertion in *ade6*. (**B**) Plates showing the color difference of the identified mutants. Each mutant was plated in a 1:30 ratio of the red mutant to a strain with the average pink colony color on YE + 3% glucose and incubated for 4 days at 32°C. Blue arrows highlight the red colonies, and the red insertion mutant names are shown below each picture. The barcodes ATCGACAAACAAAGAAAACGTAAATTGACATTTACAGAGA and ATCTACATATAAAATAACATTGAGATGTATAAGTACATTAA identified strains P34H5 and P8C11, respectively. (**C**) A schematic of the *ade6*$^+$ gene ORF showing the location of commonly used *ade6* mutations and the *Hermes::kanMX* insertions. The nonsense mutations (*ade6-M26, ade6-M375, ade6-704 and ade6-L469*; red circles) all form red colonies, while the missense mutations (*ade6-M216 and ade6-M210*; pink circles) form pink colonies on YE medium with limiting adenine (55,71). The sites of the two identified *Hermes::kanMX ade6* mutations present in the library that form red colonies are shown. (**D**) Validation of the *ade6-M210* mutation. The base change in the *ade6-M210* allele is known from personal communications from several labs (Wayne Wahls, Ramsay McFarlane, Susan Forsburg and Mikel Zaratiegui) but has not been published. We confirmed these earlier observations by sequencing the region of *ade6* containing this mutation from two *ade6*$^+$ genes (strain L972 from J. Kohli and KRP387 from our lab) and three *ade6-M210* alleles (strains GP201 from Gerry Smith, FY1645 from Robin Allshire and KRP2 from JoAnn Wise).

colony color (Figures 5C and 8C) (55). Nonsense mutations in *ade6* that ablate enzyme function cause the formation of red colonies (e.g. M26, M375, 704 and L469; Figure 8C) (47). As *Hermes::kanMX* insertions introduce stop codons, the *ade6::Hermes::kanMX* insertions should produce red colonies similar to the nonsense mutations. Cells were plated on the rich (YE) medium with limiting amounts of adenine, and two red colonies in a background of ∼6000 pink colonies were identified. The barcodes identified both *ade6*$^-$ strains in the collection (Figure 8B and C). The insertions were near the 3′ end of the ORF, near a nonsense mutation (*L469*) that results in a red colony and a missense mutation (*M210*) that forms a pink colony (Figure 8C and D), consistent with the *Hermes::kanMX* insertion inactivating the *ade6* gene.

The second screen was a novel application of a colony color assay used in *S. cerevisiae* to identify mutants that accumulate hydrogen sulfide (H_2S): plating cells on rich medium with lead acetate and extra ammonium sulfate (56,57). Extracellular sulfate is converted to sulfide in three enzymatic steps, and sulfide is used by the Met17 enzyme to make homocysteine during sulfur metabolism (Figure 9B). The H_2S that accumulates in *S. cerevisiae met17*$^-$ mutants reacts with the lead ions to produce a black–brown precipitate while *MET17*$^+$ colonies remain white (56,57). As the *S. pombe* insertion collection contained a *met17* insertion mutant, plates with rich medium plus lead acetate and ammonium sulfate were used to screen the mutant pool as well.

We found that sulfur metabolism in *S. pombe* was notably different from that in *S. cerevisiae*: almost all of the colonies were black–brown, except for three that were all white (Figure 9A and C). These data indicated that *S. pombe* normally produces H_2S when provided with excess ammonium sulfate. An expert in eukaryotic sulfur metabolism suggested that the white colonies were mutants that disrupted the uptake and/or the reduction of sulfate to sulfide (C. Hine, personal communication) (58,59). Amplification and sequencing of the barcodes from the three white colonies gave the same sequence, which identified strain P4B9 that has an insertion near the 5′ end of the *met10*$^+$ ORF. Met10 is a subunit of the heteromeric enzyme that converts sulfite to sulfide (60,61), suggesting that these colonies are white due to lack of sufficient sulfide production to form the black–brown precipitate. Therefore, in spite of this unexpected aspect of sulfur metabolism in *S. pombe*, screening the pool of barcoded mutants revealed a mutant consistent with the evolutionarily conserved pathways in sulfide production. Thus, the limiting adenine and lead acetate plate screens demonstrate the utility of the barcoded mutants in a pooled analysis setting.

DISCUSSION

3D cell pooling combined with a deep sequencing strategy greatly speeds characterization of an arrayed random insertion mutant library

The approach described here provides an efficient path to sequence genome-wide collections of insertion elements that generate defined boundaries upon integration into the genome, such as transposons or retroviral vectors. Efficiency was achieved by the combination of multiplexing samples in indexed pools, and using the derived sequence information to find the same sequence in different pools. This process identified the well in which the mutant resided in the collection of 96-well plates as well as identifying the

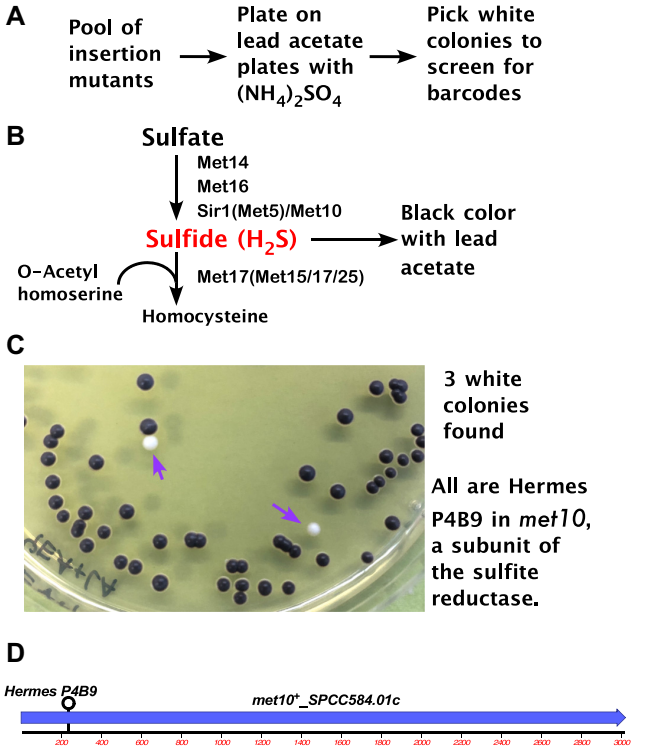


Figure 9. Identification of *met10::Hermes::kanMX* mutants from a pool of mutants using the insertion-associated barcodes. (**A**) A schematic of the mutant screen. A pool of the insertion mutants was plated on YE + adenine (225 mg/l) medium with 0.1% lead acetate and 0.02% ammonium sulfate to form ∼6000 colonies. The majority of colonies were black–brown in color. Three white colonies were identified, and the barcode DNA was amplified from each colony, sequenced and compared against Supplementary File S1 to identify the insertion site. Each colony contained the barcode AGGTAAAGTGACAATCATAATGAAATTTATATCAACAAGTA that identified the same insertion mutant in SPCC584.01c or *met10*$^+$. (**B**) A schematic of the *S. pombe* biosynthetic pathways that reduce exogenous sulfate to sulfide, which causes the dark PbS precipitate in colonies grown on plates with lead ions. The *S. pombe* enzyme names are shown, with the *S. cerevisiae* equivalents shown in parentheses if the name is different. (**C**) Plates showing the color difference of the identified mutants. Plates from the original screen for mutant with two of the three white colonies found are shown. All three colonies had the barcode that identified the P4B9 strain with an insertion in *met10*$^+$, a component of the sulfite reductase enzyme that produces sulfide. (**D**) A schematic of the *met10*$^+$ gene ORF showing the location of the *Hermes::kanMX* insertion.

barcode sequence and transposon insertion. Varshney *et al.* used a pooling approach with a large-scale zebrafish mutant project with 6-mer DNA tags arrayed in 96-well format to index each mutant sample, which identified the genome sequences flanking each integration by using an index tag (17). This strategy greatly reduced the cost of traditional capillary sequencing of individual samples by pooling all samples into a single lane of the Illumina sequencing platform.

However, the Varshney *et al.* approach still required producing thousands of individual samples for genomic DNA preparation, which was labor intensive. In addition, thousands of pre-synthesized index tags were required. By using 3D pooling, sequencing efficiency was greatly improved, allowing us to process and sequence thousands of samples at one time with only 40 index tags, greatly reducing the effort and cost required to create a defined mutant collection. Our pooling approach did not require more complex pooling methods [e.g. shifted transversal design (31)], which can be powerful but require encoding and decoding of pools that may not be universally available. Our approach becomes even more powerful with continuing improvements in technology as the entire mutant collection of ∼9200 strains (96 plates) could be processed at one time on a newer, higher throughput platform, such as NovaSeq that generates over 2×10^9 reads.

The *Hermes* insertion mutant library is a multifaceted resource

Our defined, barcode-tagged library of viable *S. pombe* mutants can be applied in high-throughput phenotypic screens and in the analysis of individual strains. The set of mutants can be assayed in pooled competitive growth experiments under different drug and stress conditions, and the abundance of an individual mutant in the pooled culture can be represented by its barcode abundance, usually by high-throughput sequencing (4,36,38). The barcodes are also flanked by SfiI sites that allow them to be cut out of amplified pools of barcodes, oligomerized by ligation, and cloned and sequenced to determine the most frequent barcodes if high-throughput sequencing is not available (62). The defined structure of the transposon insertions, the known locations and the inclusion of a *lox71* site facilitate further modification of the mutated genes (see Supplementary Materials and Supplementary Figures S6–S9 for examples). We have previously induced excision of *Hermes* transposons from the *S. pombe* genome and shown that the repaired locus contains a variety of mutations (32), so excision allows one to generate random, targeted mutations.

The *Hermes* insertion library mutants display a wide range of phenotypes

Some *Hermes* mutants revealed different phenotypes compared to the haploid deletion mutants. First of all, a total of 368 insertions in *S. pombe* essential genes yielded viable phenotypes. In contrast, the null mutants of these essential genes in the haploid deletion set cannot survive. Second, a phenotypic comparison of 55 *Hermes* insertion mutants in CPT sensitivity genes to their corresponding deletion mutants revealed that 33 (∼60%) *Hermes* mutants were CPT sensitive, including most coding exon and intron insertion mutants and ∼40% of the UTR mutants. However, 19 of these insertion mutants were more resistant to CPT than the corresponding deletion mutants, demonstrating an increased range of phenotypes for mutations in these genes (Figure 7A and Supplementary Table S2). Third, one 5′ UTR insertion mutant showed the opposite phenotype of CPT resistance. Thus, some *Hermes::kanMX* insertions may produce new phenotypes by either truncating gene products or interfering with normal gene regulation from upstream transcriptional control sequences. Our library therefore provides a useful new tool for the analysis of gene function and can reveal new phenotypes.

Our collection of viable, barcoded insertion mutants contrasts with dense transposon integration, or Tn-seq, ap-

proaches for identification of genes affecting different cellular processes. Tn-seq uses hundreds of thousands of insertion mutants to prepare DNA for next-generation sequencing, and subsequently identifies genes required for growth under specific conditions. For example, Guo *et al.* used this approach to identify essential and nonessential genes for cell growth on synthetic medium, identifying essential genes as those with few or no transposon insertions (63). Lee *et al.* recently identified genes for novel factors that promote heterochromatin formation by performing *Hermes::kanMX* integrations into heterochromatin reporter strain (43). This work identified insertions in the 3′ ends of 65 essential genes that allowed viability but impaired heterochromatin formation. Thus, the Tn-seq approach can identify regions of essential genes that are dispensable for growth and involved in specific functions. The Tn-seq approach thus generates extensive information from a specific screen, although the original mutants are never recovered. In contrast, the smaller barcoded insertion library of viable mutants in this work can be repeatedly used for different mutant screens. The barcode insertion library was made in a strain amenable to systematic genetic analysis approaches (6,64), which would allow the introduction of reporter genes for different processes. A second difference is that the Tn-seq approach requires significant bioinformatic support to process the next-generation sequencing data, while identifying existing mutants in the barcoded insertion library does not (e.g. Figures 8 and 9). Consequently, the approaches have distinct uses, with Tn-seq allowing a robust approach to gene identification in a specific process, and the barcoded insertion library allowing rapid identification of a subset of genes to identify a biological process for further investigation.

Implications for future library construction

Construction and analysis of this insertion mutant library revealed several key points for designing future approaches in *S. pombe* and other organisms. First, the insertion element should have little target site selection bias so that the insertions can be as randomly distributed throughout the genomes as possible. Second, the integration sites should be of a defined structure to allow high-throughput sequencing. Third, it is important to generate as many single insertions as possible, so the phenotype can be easily associated with a single mutation. Fourth, the insertion element should allow future modifications of the mutant collection. The *Hermes* insertion library met these criteria quite well and provides guidance for constructing similar libraries in other organisms.

The *Hermes* transposon did show minor target site preferences**.** The *Hermes* transposon was originally chosen because *Hermes* efficiently targets ORFs and regions upstream and downstream of ORF, based on a small number of samples (35). While this work was in progress, the *Hermes* transposon was reported to preferentially insert into nucleosome-free regions in *S. cerevisiae* (41). More recently, results of an analysis of 1.36 million *Hermes* transposon target sites in *S. pombe* also suggested that *Hermes* insertions prefer to insert into nucleosome-free regions *in vivo* (63). Our analysis of 9024 mutants identified insertions in 2753 protein-coding genes out of a total of 5131 (65). We calculate that obtaining a *Hermes::kanMX* library with 80% of the protein-coding genes marked by an insertion would require 14 520 mutants (see the 'Materials and Methods' section). Therefore, an alternative approach that utilizes the *Hermes* transposon in combination with other elements that can target nucleosomal DNA, similar to the multi-transposon approaches used in *Drosophila* library constructions (2), may be more efficient.

Unfortunately, transposons that target nucleosome-bound DNA in *S. pombe* have not yet been discovered. The *S. cerevisiae* retrotransposon Ty1 targets a specific surface of the nucleosome at the H2A/H2B interface to insert in nucleosome-bound DNA (66–68). The majority (∼90%) of insertions were within the predicted 5′ region of Pol III-transcribed genes (69). Recently, a rice miniature inverted repeat transposable element (MITE) was applied to *S. cerevisiae*. About 65% of the insertions were in genes (70). It is not known whether Ty1 or MITE is active in *S. pombe*, but it may be possible to adapt these systems to other organisms in the same way we adapted the *Hermes* system for the *S. pombe* insertion library.

DATA AVAILABILITY

The mutants described in Supplementary File S1 will be made available as individual strains in 96-well plates or as mixed pools of mutants through the National BioResource Project, Yeast section at Osaka City University for international distribution (https://yeast.nig.ac.jp/yeast/). The datasets generated during the current study are available in the NCBI Sequence Read Archive repository as BioProject PRJNA685113.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Valerie Wood (PomBase) for discussions on the rDNA array and *ade6-M210*, Dr Christopher Hine (Cleveland Clinic) for discussions on sulfur metabolism and two anonymous reviewers for comments and suggestions on the manuscript, resulting in the experiments in Figures 5, 8 and 9. The CWRU Genomics Core, which performed the high-throughput sequencing, was supported in part by the Case Comprehensive Cancer Center (5P30CA043703).

FUNDING

National Institutes of Health [GM050752 and AG051601 to K.W.R., HL55666 and HL081093 to K.L.B.]. Funding for open access charge: National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. Ross-Macdonald,P., Coelho,P.S., Roemer,T., Agarwal,S., Kumar,A., Jansen,R., Cheung,K.H., Sheehan,A., Symoniatis,D., Umansky,L. *et al.* (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*, **402**, 413–418.

2. Bellen,H.J., Levis,R.W., He,Y., Carlson,J.W., Evans-Holm,M., Bae,E., Kim,J., Metaxakis,A., Savakis,C., Schulze,K.L. *et al.* (2011) The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. *Genetics*, **188**, 731–743.
3. Alonso,J.M., Stepanova,A.N., Leisse,T.J., Kim,C.J., Chen,H., Shinn,P., Stevenson,D.K., Zimmerman,J., Barajas,P., Cheuk,R. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.
4. Kim,D.U., Hayles,J., Kim,D., Wood,V., Park,H.O., Won,M., Yoo,H.S., Duhig,T., Nam,M., Palmer,G. *et al.* (2010) Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat. Biotechnol.*, **28**, 617–623.
5. Winzeler,E.A., Shoemaker,D.D., Astromoff,A., Liang,H., Anderson,K., Andre,B., Bangham,R., Benito,R., Boeke,J.D., Bussey,H. *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.
6. Tong,A.H., Evangelista,M., Parsons,A.B., Xu,H., Bader,G.D., Page,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**, 2364–2368.
7. Roguev,A., Bandyopadhyay,S., Zofall,M., Zhang,K., Fischer,T., Collins,S.R., Qu,H., Shales,M., Park,H.O., Hayles,J. *et al.* (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, **322**, 405–410.
8. Collins,S.R., Miller,K.M., Maas,N.L., Roguev,A., Fillingham,J., Chu,C.S., Schuldiner,M., Gebbia,M., Recht,J., Shales,M. *et al.* (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*, **446**, 806–810.
9. Krogan,N.J., Cagney,G., Yu,H., Zhong,G., Guo,X., Ignatchenko,A., Li,J., Pu,S., Datta,N., Tikuisis,A.P. *et al.* (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, **440**, 637–643.
10. Braberg,H., Echeverria,I., Bohn,S., Cimermancic,P., Shiver,A., Alexander,R., Xu,J., Shales,M., Dronamraju,R., Jiang,S. *et al.* (2020) Genetic interaction mapping informs integrative structure determination of protein complexes. *Science*, **370**, eaaz4910.
11. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
12. Zhang,X., Paganelli,F.L., Bierschenk,D., Kuipers,A., Bonten,M.J., Willems,R.J. and van Schaik,W. (2012) Genome-wide identification of ampicillin resistance determinants in *Enterococcus faecium*. *PLoS Genet.*, **8**, e1002804.
13. Braberg,H., Moehle,E.A., Shales,M., Guthrie,C. and Krogan,N.J. (2014) Genetic interaction analysis of point mutations enables interrogation of gene function at a residue-level resolution: exploring the applications of high-resolution genetic interaction mapping of point mutations. *Bioessays*, **36**, 706–713.
14. Ooi,S.L., Shoemaker,D.D. and Boeke,J.D. (2003) DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. *Nat. Genet.*, **35**, 277–286.
15. Peyser,B.D., Irizarry,R.A., Tiffany,C.W., Chen,O., Yuan,D.S., Boeke,J.D. and Spencer,F.A. (2005) Improved statistical analysis of budding yeast TAG microarrays revealed by defined spike-in pools. *Nucleic Acids Res.*, **33**, e140.
16. Gaj,T., Gersbach,C.A. and Barbas,C.F. 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.*, **31**, 397–405.
17. Varshney,G.K., Lu,J., Gildea,D.E., Huang,H., Pei,W., Yang,Z., Huang,S.C., Schoenfeld,D., Pho,N.H., Casero,D. *et al.* (2013) A large-scale zebrafish gene knockout resource for the genome-wide study of gene function. *Genome Res.*, **23**, 727–735.
18. Jacobs,M.A., Alwood,A., Thaipisuttikul,I., Spencer,D., Haugen,E., Ernst,S., Will,O., Kaul,R., Raymond,C., Levy,R. *et al.* (2003) Comprehensive transposon mutant library of pseudomonas aeruginosa. *Proc. Natl Acad. Sci. U.S.A.*, **100**, 14339–14344.
19. Suzuki,N., Okai,N., Nonaka,H., Tsuge,Y., Inui,M. and Yukawa,H. (2006) High-throughput transposon mutagenesis of *Corynebacterium glutamicum* and construction of a single-gene disruptant mutant library. *Appl. Environ. Microbiol.*, **72**, 3750–3755.
20. Lewenza,S., Falsafi,R.K., Winsor,G., Gooderham,W.J., McPhee,J.B., Brinkman,F.S. and Hancock,R.E. (2005) Construction of a mini-Tn5-luxCDABE mutant library in *Pseudomonas aeruginosa* PAO1: a tool for identifying differentially regulated genes. *Genome Res.*, **15**, 583–589.
21. Chen,B.R., Li,Y., Eisenstatt,J.R. and Runge,K.W. (2013) Identification of a lifespan extending mutation in the *Schizosaccharomyces pombe* cyclin gene clg1^{+} by direct selection of long-lived mutants. *PLoS One*, **8**, e69084.
22. Chen,Y. and Klionsky,D.J. (2011) The regulation of autophagy—unanswered questions. *J. Cell Sci.*, **124**, 161–170.
23. Takeda,K., Mori,A. and Yanagida,M. (2011) Identification of genes affecting the toxicity of anti-cancer drug bortezomib by genome-wide screening in *S. pombe*. *PLoS One*, **6**, e22021.
24. Ivey,F.D., Wang,L., Demirbas,D., Allain,C. and Hoffman,C.S. (2008) Development of a fission yeast-based high-throughput screen to identify chemical regulators of cAMP phosphodiesterases. *J. Biomol. Screen.*, **13**, 62–71.
25. Kim,D.M., Kim,H., Yeon,J.H., Lee,J.H. and Park,H.O. (2016) Identification of a mitochondrial DNA polymerase affecting cardiotoxicity of sunitinib using a genome-wide screening on *S. pombe* deletion library. *Toxicol. Sci.*, **149**, 4–14.
26. Zuin,A., Carmona,M., Morales-Ivorra,I., Gabrielli,N., Vivancos,A.P., Ayte,J. and Hidalgo,E. (2010) Lifespan extension by calorie restriction relies on the Sty1 MAP kinase stress pathway. *EMBO J.*, **29**, 981–991.
27. Miwa,Y., Ohtsuka,H., Naito,C., Murakami,H. and Aiba,H. (2011) Ecl1, a regulator of the chronological lifespan of *Schizosaccharomyces pombe*, is induced upon nitrogen starvation. *Biosci. Biotechnol. Biochem.*, **75**, 279–283.
28. Spivey,E.C., Jones,S.K., Rybarski,J.R., Saifuddin,F.A. and Finkelstein,I.J. (2017) An aging-independent replicative lifespan in a symmetrically dividing eukaryote. *eLife*, **6**, e20340.
29. Rhind,N., Chen,Z., Yassour,M., Thompson,D.A., Haas,B.J., Habib,N., Wapinski,I., Roy,S., Lin,M.F., Heiman,D.I. *et al.* (2011) Comparative functional genomics of the fission yeasts. *Science*, **332**, 930–936.
30. Wood,V., Gwilliam,R., Rajandream,M.A., Lyne,M., Lyne,R., Stewart,A., Sgouros,J., Peat,N., Hayles,J., Baker,S. *et al.* (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature*, **415**, 871–880.
31. Thierry-Mieg,N. (2006) A new pooling strategy for high-throughput screening: the shifted transversal design. *BMC Bioinformatics*, **7**, 28.
32. Li,Y., Wang,J., Zhou,G., Lajeunesse,M., Le,N., Stawicki,B.N., Corcino,Y.L., Berkner,K.L. and Runge,K.W. (2017) Nonhomologous end-joining with minimal sequence loss is promoted by the Mre11–Rad50–Nbs1–Ctp1 complex in *Schizosaccharomyces pombe*. *Genetics*, **206**, 481–496.
33. Suga,M. and Hatakeyama,T. (2005) A rapid and simple procedure for high-efficiency lithium acetate transformation of cryopreserved *Schizosaccharomyces pombe* cells. *Yeast*, **22**, 799–804.
34. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) In: *Molecular Cloning: A Laboratory Manual*. (2nd edn), Cold Spring Harbor Laboratory, NY.
35. Evertts,A.G., Plymire,C., Craig,N.L. and Levin,H.L. (2007) The Hermes transposon of *Musca domestica* is an efficient tool for the mutagenesis of *Schizosaccharomyces pombe*. *Genetics*, **177**, 2519–2523.
36. Han,T.X., Xu,X.Y., Zhang,M.J., Peng,X. and Du,L.L. (2010) Global fitness profiling of fission yeast deletion strains by barcode sequencing. *Genome Biol.*, **11**, R60.
37. Smith,A.M., Heisler,L.E., Mellor,J., Kaper,F., Thompson,M.J., Chee,M., Roth,F.P., Giaever,G. and Nislow,C. (2009) Quantitative phenotyping via deep barcode sequencing. *Genome Res.*, **19**, 1836–1842.
38. Smith,A.M., Heisler,L.E., St Onge,R.P., Farias-Hesson,E., Wallace,I.M., Bodeau,J., Harris,A.N., Perry,K.M., Giaever,G., Pourmand,N. *et al.* (2010) Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.*, **38**, e142.
39. Pierce,S.E., Davis,R.W., Nislow,C. and Giaever,G. (2007) Genome-wide analysis of barcoded *Saccharomyces cerevisiae* gene-deletion mutants in pooled cultures. *Nat. Protoc.*, **2**, 2958–2974.

40. Boeke,J.D., Trueheart,J., Natsoulis,G. and Fink,G.R. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.*, **154**, 164–175.
41. Gangadharan,S., Mularoni,L., Fain-Thornton,J., Wheelan,S.J. and Craig,N.L. (2010) DNA transposon Hermes inserts into DNA in nucleosome-free regions *in vivo*. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 21966–21972.
42. Guo,Y. and Levin,H.L. (2010) High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*. *Genome Res.*, **20**, 239–248.
43. Lee,S.Y., Hung,S., Esnault,C., Pathak,R., Johnson,K.R., Bankole,O., Yamashita,A., Zhang,H. and Levin,H.L. (2020) Dense transposon integration reveals essential cleavage and polyadenylation factors promote heterochromatin formation. *Cell Rep.*, **30**, 2686–2698.
44. Wood,V., Harris,M.A., McDowall,M.D., Rutherford,K., Vaughan,B.W., Staines,D.M., Aslett,M., Lock,A., Bahler,J., Kersey,P.J. *et al.* (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.*, **40**, D695–D699.
45. Chaudhuri,B., Ingavale,S. and Bachhawat,A.K. (1997) *apd1*$^{+}$, a gene required for red pigment formation in *ade6* mutants of *Schizosaccharomyces pombe*, encodes an enzyme required for glutathione biosynthesis: a role for glutathione and a glutathione-conjugate pump. *Genetics*, **145**, 75–83.
46. Silver,J.M. and Eaton,N.R. (1969) Functional blocks of the ad-1 and ad-2 mutants of *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, **34**, 301–305.
47. Smirnov,M.N., Smirnov,V.N., Budowsky,E.I., Inge-Vechtomov,S.G. and Serebrjakov,N.G. (1967) Red pigment of adenine-deficient yeast *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, **27**, 299–304.
48. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
49. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A. *et al.* (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
50. Steinmetz,L.M., Scharfe,C., Deutschbauer,A.M., Mokranjac,D., Herman,Z.S., Jones,T., Chu,A.M., Giaever,G., Prokisch,H., Oefner,P.J. *et al.* (2002) Systematic screen for human disease genes in yeast. *Nat. Genet.*, **31**, 400–404.
51. Deshpande,G.P., Hayles,J., Hoe,K.L., Kim,D.U., Park,H.O. and Hartsuiker,E. (2009) Screening a genome-wide *S. pombe* deletion library identifies novel genes and pathways involved in genome stability maintenance. *DNA Repair*, **8**, 672–679.
52. Liu,L.F., Desai,S.D., Li,T.K., Mao,Y., Sun,M. and Sim,S.P. (2000) Mechanism of action of camptothecin. *Ann. N.Y. Acad. Sci.*, **922**, 1–10.
53. Tenreiro,S., Rosa,P.C., Viegas,C.A. and Sa-Correia,I. (2000) Expression of the AZR1 gene (ORF YGR224w), encoding a plasma membrane transporter of the major facilitator superfamily, is required for adaptation to acetic acid and resistance to azoles in *Saccharomyces cerevisiae*. *Yeast*, **16**, 1469–1481.
54. Svensson,J.P., Pesudo,L.Q., Fry,R.C., Adeleye,Y.A., Carmichael,P. and Samson,L.D. (2011) Genomic phenotyping of the essential and non-essential yeast genome detects novel pathways for alkylation resistance. *BMC Syst. Biol.*, **5**, 157.
55. Szankasi,P., Heyer,W.D., Schuchert,P. and Kohli,J. (1988) DNA sequence analysis of the *ade6* gene of *Schizosaccharomyces pombe*. Wild-type and mutant alleles including the recombination host spot allele *ade6-M26*. *J. Mol. Biol.*, **204**, 917–925.
56. Ono,B., Ishii,N., Fujino,S. and Aoyama,I. (1991) Role of hydrosulfide ions (HS^{-}) in methylmercury resistance in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **57**, 3183–3186.
57. Cost,G.J. and Boeke,J.D. (1996) A useful colony colour phenotype associated with the yeast selectable/counter-selectable marker MET15. *Yeast*, **12**, 939–941.
58. Hine,C., Harputlugil,E., Zhang,Y., Ruckenstuhl,C., Lee,B.C., Brace,L., Longchamp,A., Trevino-Villarreal,J.H., Mejia,P., Ozaki,C.K. *et al.* (2015) Endogenous hydrogen sulfide production is essential for dietary restriction benefits. *Cell*, **160**, 132–144.
59. Hine,C. and Mitchell,J.R. (2015) Calorie restriction and methionine restriction in control of endogenous hydrogen sulfide production by the transsulfuration pathway. *Exp. Gerontol.*, **68**, 26–32.
60. Hansen,J., Cherest,H. and Kielland-Brandt,M.C. (1994) Two divergent MET10 genes, one from *Saccharomyces cerevisiae* and one from *Saccharomyces carlsbergensis*, encode the alpha subunit of sulfite reductase and specify potential binding sites for FAD and NADPH. *J. Bacteriol.*, **176**, 6050–6058.
61. Masselot,M. and Surdin-Kerjan,Y. (1977) Methionine biosynthesis in *Saccharomyces cerevisiae*. II. Gene–enzyme relationships in the sulfate assimilation pathway. *Mol. Gen. Genet.*, **154**, 23–30.
62. Chen,B.-R., Hale,D.C., Ciolek,P.J. and Runge,K.W. (2012) Generation and analysis of a barcode-tagged insertion mutant library in the fission yeast *Schizosaccharomyces pombe*. *BMC Genomics*, **13**, 161.
63. Guo,Y., Park,J.M., Cui,B., Humes,E., Gangadharan,S., Hung,S., FitzGerald,P.C., Hoe,K.L., Grewal,S.I., Craig,N.L. *et al.* (2013) Integration profiling of gene function with dense maps of transposon integration. *Genetics*, **195**, 599–609.
64. Roguev,A., Wiren,M., Weissman,J.S. and Krogan,N.J. (2007) High-throughput genetic interaction mapping in the fission yeast *Schizosaccharomyces pombe*. *Nat. Methods*, **4**, 861–866.
65. Wood,V. (2006) *Schizosaccharomyces pombe* comparative genomics: from sequence to systems. In: Sunnerhagen,P. and Piskur,J. (eds). *Comparative Genomics: Using Fungi as Models*. Springer, Berlin, pp. 233–285.
66. Mularoni,L., Zhou,Y., Bowen,T., Gangadharan,S., Wheelan,S.J. and Boeke,J.D. (2012) Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res.*, **22**, 693–703.
67. Baller,J.A., Gao,J., Stamenova,R., Curcio,M.J. and Voytas,D.F. (2012) A nucleosomal surface defines an integration hotspot for the *Saccharomyces cerevisiae* Ty1 retrotransposon. *Genome Res.*, **22**, 704–713.
68. Bridier-Nahmias,A. and Lesage,P. (2012) Two large-scale analyses of Ty1 LTR-retrotransposon *de novo* insertion events indicate that Ty1 targets nucleosomal DNA near the H2A/H2B interface. *Mob. DNA*, **3**, 22.
69. Mularoni,L., Zhou,Y., Bowen,T., Gangadharan,S., Wheelan,S.J. and Boeke,J.D. (2012) Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res.*, **22**, 693–703.
70. Fattash,I., Bhardwaj,P., Hui,C. and Yang,G. (2013) A rice stowaway MITE for gene transfer in yeast. *PLoS One*, **8**, e64135.
71. Park,J.M., Intine,R.V. and Maraia,R.J. (2007) Mouse and human la proteins differ in kinase substrate activity and activation mechanism for tRNA processing. *Gene Expr.*, **14**, 71–81.