OXFORD UNIVERSITY PRESS | Nucleic Acids Research

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

Journal:	*Nucleic Acids Research*
Manuscript ID	NAR-03083-Met-K-2020.R2
Manuscript Type:	3 Methods Manuscript (Online Publication)
Key Words:	S. pombe library, DNA barcodes, three-dimensional pooling, multiplexed high-throughput sequencing, Hermes transposon

DATA AVAILABILITY

Does the manuscript use or report the following? If so, please provide details in a Data Availability statement below and in the manuscript.	
New genome expression or sequencing data (ChIP-seq, RNA-seq…) - Must comply with ENCODE Guidelines. - All datasets must be validated via biological replicates. - Must deposit data in GEO or an equivalent publicly available depository and provide accession numbers, private tokens, reviewer login details and/or private URLs for Referees. - Excluding RNA-Seq, data must be viewable on the UCSC (eukaryotes) or other suitable genome browsers; must provide genome browser session links (even if GEO entries are publicly available).	**No**
Novel nucleic acid sequences - Must deposit in EMBL / GenBank / DDBJ. - Must provide sequence names and accession numbers.	**No**
Illumina-type sequencing data - Must submit data to BioProject/SRA, ArrayExpress or GEO. - Must provide link for reviewers (BioProject/SRA), login details (ArrayExpress) or accession numbers and private tokens (GEO).	**Yes**
Novel protein sequences - Must deposit UniProt using the interactive tool SPIN. - Must provide sequence names and accession number.	**No**
Novel molecular structures determined by X-ray crystallography, NMR and/or CryoEM/EM - Must deposit to a member site of the Worldwide Protein Data Bank (RCSB PDB, PDBe, PDBj) and provide the accession numbers. - If structures are unreleased (i.e. status HPUB), **MUST** upload: - the validation reports (.pdf) - molecular coordinates (.pdb or .mmcif). - one of the following: • X-ray data (.mtz, .cif) • NMR restraints and chemical shift files (.mr, .tbl or .str) • CryoEM map files (.map).	**No**
Novel molecular models based on SAXS, computational modeling, or other combinations of strategies that are generally not appropriate for deposition in the PDB - Must deposit coordinates and all underlying data in appropriate databases (including but not limited to the Small Angle Scattering Database and PDB-Dev). - Must report on validation of the structure against experimental data (if available) or report on statistical validation of the structure by model quality assessment programs. If applicable, these should be uploaded as a Data file.	**No**
Molecular behaviour studies derived from biological NMR spectroscopy data (not necessarily leading to new structures)	**No**

- Must deposit NMR spectral data, including assigned chemical shifts, coupling constants, relaxation parameters (T1, T2, and NOE values), dipolar couplings, in BMRB.	
Novel nucleic acids structure - Must deposit to NDB (via PDB if possible) and provide accession numbers.	**No**
Structures of nucleosides, nucleotides, other small molecules - Must deposit in the Cambridge Crystallographic Data Centre (CCDC) and provide the structure identifiers.	**No**
Mass spectrometry proteomics - Must deposit to ProteomeXchange consortium and provide Dataset Identifier and reviewer account details. If appropriate, data and corresponding details can also be deposited in the Panorama repository for targeted mass spec assays and workflows.	**No**
Microarray data - Must comply with the MIAME Guidelines - Must deposit the data to GEO or Array Express, and provide accession numbers and private tokens (GEO) or login details (ArrayExpress).	**No**
Quantitative PCR - Must comply with the MIQE Guidelines. - Details should be supplied in Materials and Methods section of manuscript.	**No**
Synthetic nucleic acid oligonucleotides including siRNAs or shRNAs - The manuscript should include controls to rule out off-target effects, such as use of multiple siRNA/shRNAs or inclusion of cDNA rescue data. - Manuscript should provide exact sequences, exact details of chemical modifications at any position, and source of reagent or precise methods for creation. These can be included in the main text or in Supplementary Material.	**No**
Flow Cytometry experiments - Must deposit in FlowRepository. - Must provide Repository ID and URL with secret code for Referees.	**No**
Software and source codes - Must deposit in GitHub and provide link to code in GitHub or upload source code as Data file.	**No**
Gel images, micrographs, graphs, and tables - Optionally, may deposit in a general-purpose repository such as Zenodo or Dryad. If applicable, provide access details.	**No**

REFEREES – you will find data deposition details below
Data Availability: The mutants described Additional File 1 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. The link to these data is: https://dataview.ncbi.nlm.nih.gov/object/PRJNA685113?reviewer=7oqtmqklbtshimhgvjfdv1k3a6

KEY POINTS

(3 bullet points summarizing the manuscript's contribution to the field)

- We developed a novel, high-throughput method to construct an arrayed, barcode-tagged transposon insertion library by deep sequencing and computational analysis.
- The valuable approach can be applied to construction and analysis of insertion mutant libraries in a wide variety of model systems.
- This library represents an important resource for the international S. pombe community.

May 30, 2022

Dear Dr. Kimmel and Referees 1 and 2,

Thank you for the opportunity to revise our submission NAR-03083-Met-K-2020.R1 “A multiplexed, three-dimensional pooling and next generation sequencing strategy for creating barcoded mutant arrays: Construction of a *Schizosaccharomyces pombe* transposon insertion library” for consideration for publication in *Nucleic Acids Research Methods Online*.

Referee 2 was satisfied with our responses to the critiques after the first revision and supported publication of the work. Referee 1 had remaining concerns about a typographical error, some aspects of the presentation of data and a figure and an additional control for our experiments to show that the *Hermes* transposon can support expression of nearby genes under very specific circumstances. We spent two months creating the control strains that robustly address Referee 1’s concerns. These new results are now presented in an updated Fig. 5 and new Supplementary Fig. 4.

We note that the order of the authors has changed to reflect this additional work, and Dr. Haitao Zhang’s name is now in front of Dr. Gang Zhang.

We believe that we have answered all of the Referees concerns and that this submission describing our novel approach to create a barcoded, sequenced insertion library of viable mutants with a wide range of phenotypes has met the requirements for publication.

Please feel free to contact us with any additional comments regarding the manuscript.

Sincerely,
Kurt Runge for the authors

Point by point response to Referee 1.

1. Again, the number of insertion mutations varies. In the abstract the number is 4,381 but later in the manuscript the number 4,391 is used (page 22, line 19).

> We have corrected this typographical error in the abstract to 4,391.

2. Page 19, line 27 describes the mapping of genome sequences of the insertion sites before you discuss ligation mediated PCR. This seems backwards.

> Goal of this sentence was to show that the triangulation script provides an independent means of validating the Hermes insertion site. We would like to leave the sentence where it is to associate it with concepts related to mapping the insertion sites.

3. Page 23, line 41. You propose the model that ATGs in the TIRs of Hermes can function at insertion sites to translate disrupted ORFs. Unfortunately, you did not test this by mutating the ATGs in the Hermes TIR in your ade7 fusions. There are other possibilities such as activating

cryptic promoters in ade7 sequence that can produce transcripts translated from a downstream ATG. Yours is a reasonable model but the integration profiles of Guo et al (Fig 4) and Lee et al (Fig. 1B) suggest that if Hermes can produce protein fusions it's not common. In comparison the work of Michael et al (PMID: 28481201) uses the MiniD transposon in cerevisiae and found extensive complementation of disrupted genes from an internal MiniD promoter.

Our goal with this experiment was to answer the Reviewer's initial concern that these essential genes with insertions could be expressed at some level, which our ade7 fusion results showed. To address Reviewer 1's new concern, we made a second set of constructs in which the Hermes-derived ATGs are mutated to TTC to eliminate the start codon for the downstream ade7 ORF. The resulting Left End and Right End fusions missing an ATG (LE-ATG-ade7 and RE-ATG-ade7) have the same growth phenotypes as cells bearing the ade7-deletion, as now shown in Figure 5 and the new Supplemental Figure 4. We analyzed two separate isolates of each fusion, and the ade7 ORF and flanking sequences were amplified and sequenced from all four strains to show that the only mutations were the ATG to TTC changes at the initiator methionine codon. Consequently, expression of the Hermes-ade7 fusions requires and in-frame fusion with the short ORF ATGs at the right and left ends of Hermes.

We agree that such fusions are likely to be uncommon, consistent with the previously published work of Guo et al. and Lee et al from the Levin lab. We modified the existing text to be certain that the small fraction of insertions in the first 150 bp of essential genes was not overstated as a large number but as a surprising finding for a small number of genes.

4. Your description of the Ade7 fusion is not clear. One has to refer to the methods section to discover this is a genomic edit inserting a full Hermes at the second codon of ade7.

Figure 5 describing this experiment has been revised to include a new panel B depicting the integration of Hermes into the *ade7* locus.

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,a], Neil Molyneaux[2], Haitao Zhang[3], Gang Zhou[5,b], Carly Kerr[3], Mark D. Adams[2,c], Kathleen L. Berkner[5,d] and Kurt W. Runge[1,2,e*]

[1]Department of Molecular Genetics, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 44195, USA

[2]Department of Genetics and Genomic Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio, 44106, USA

[3]Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic.

[4]Department of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 44195, USA

[5]Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 44195, USA

[a]Present address: Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas;

[b]Present address: Department of Pathology, Baylor Scott & White Health, Temple, Texas;

[c]Present address: The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut;

[d]Present address: Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic;

[e]Present address: Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic.

* To whom correspondence should be address. Tel: (216) 445-9771; Fax: (216) 444-0512; Email: rungek@ccf.org

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 both reduce the number of required sequencing reactions by orders of magnitude, and were able to identify the barcode sequences and DNA insertion sites of 4,391 *S. pombe* insertion mutations with only 40 sequencing preparations. The insertion mutations are in the genes and UTRs of non-essential, essential and non-coding 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.

Keywords: three-dimensional pooling; multiplexed high-throughput sequencing; insertion mutant library; DNA barcodes; *Hermes* transposon; *S. pombe* library.

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 are 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 yeast *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 non-coding 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 *Schizosaccharomyces 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, 3-dimensional 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 and barcodes in viable mutants. Each identified mutant and its barcode are 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 4,095 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 non-essential genes and

30% of the non-coding 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 means to rapidly generate defined mutation collections in other model systems.

Material & Methods

Construction of barcoded-*Hermes* transposon plasmids

Hermes transposon donor plasmid (pHL2577) and transposase plasmid (pHL2578) were 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 condition (1 min at 95°C, -1°C per cycle, 15 cycles; 1 min at 80°C, -0.5°C per cycle, 70 cycles; 1 min at 45°C, -0.5°C per cycle, 66 cycles to 12°C), followed by filling in to generate 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 then transformed into DH5α by electroporation. Ten separate transformations each produced 1-2 x10^{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

S. 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 hours then replica plated on YES+G418 (200 µg/ml) + FOA (1 g/L) and grown for 3 days. About 1,000 *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 formula: $P = 1- (1-f)^N$ (34), where:

f = 1/ (Number of barcoded-*Hermes* transposon plasmids)

N = number of *S. pombe* clones sampled

P = the probability of getting a barcode

1-P = the probability of not getting a barcode

This calculation predicts a >99% probability of not getting the same barcode if picking 1,000 *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 Discussion). Solving for N where P = 0.8 and f = (1/number of protein coding genes) X (fraction of insertions in protein coding genes/total number of insertions) = (1/5132) X (2273/9024) to yield 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 x 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 milliQ-filtered water, transferred to a 1.5 ml screw cap tube, cells were pelleted, 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, 50Mm EDTA, 1% SDS) and 500 µl 0.5mm 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 RNAase and proteinase K, genomic DNA was subjected to phenol/chloroform extraction and precipitated with ethanol.

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

Ligation-mediated PCR

The Mse I ds linkers were generated by annealing the upper strand oligo (5Phos/TAGTCCCTTAAGCGGAG/3AmM/-amino) to the lower strand oligo (5'GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC3'). Apo I and Mfe I ds linkers were generated by annealing the upper strand oligo (5Phos/AATTGTCCCTTAAGCGGAG/3AmM/-amino modified) to the lower strand oligo (5'GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC3'). 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, 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 Results). To amplify the *Hermes* transposon right end (HR) insertion sites, the first step was done by the HR outside primer (5'CTTGCACTCAAAAGGCTTGACAC3') specific to the transposon right end using the linker primer (5'GTAATACGACTCACTATAGGGCTC3') specific to the linkers using the following condition: 2 min at 98°C, 6 cycles of 15 sec at 98°C, 30 sec at 65°C, 40 sec at 72°C and then 24 cycles of 15 sec at 98°C, 30 sec at 60°C, 40 sec 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'CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGTAATACGACTCACTAT AGGGCT3') and an 8-bp indexed HR-nested primer (5'ACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXXXXTATGTGGCTT ACGTTTGCCTGTGG3'), which respectively adds one of the Illumina adaptors to PCR products. The conditions were 2 min at 98°C, 10 cycles of 15 sec at 98°C, 30 sec at 60°C, 40 sec 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'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGATCT3') to add Illumina-sequencing primer and another Illumina adaptor to the final product. The conditions were 2 min at 98°C, 10 cycles of 15 sec at 98°C, 1min 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'ACACTCTTTCCCTACACGACGCTCTTCCGATCT**XXXXXXX**TATCCCGGGATTTTG GCCAC3') and barcode reverse primer (5'CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCTGCAGCGAGGAGCCG TAAT3') using the following conditions: 2 min at 98°C, 30 cycles of 15 sec at 98°C, 30 sec at 60°C, 30 sec 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 sec at 98°C, 30 sec 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 (Supplemental Fig. 1). Equal molar amounts of products were mixed. Customized index tags are presented in Supplemental Table 3.

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 (Supplemental Fig. 2). The raw sequence data was 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'TATCCCGGGATTTTGGCCACCCGGGCC3'), transposon right end (5'TATGTGGCTTACGTTTGCCTGTGGCTTGTTGAAGTTCTCTG3') or left end (5'GCGCATAAGTATCAAAATAAGCCACTTGTTGTTGTTCTCTG3'). 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 Results).

A list of the insertion sites and barcode sequences associated with each insertion are provided as Additional File 1. A list of mutated genes is in Additional File 2.

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'CAAGACTAGGAAAAGAGCATAAG3') and 4171as (5'GACTGTCAAGGAGGGTATTC3') 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

2,328 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). *S. pombe* mutants carrying 15 genes under this GO term were spot tested on YES and non-fermentable YEEG (0.5% yeast extract, 2% glycerol, 2% ethanol, 2g/L casein amino acids, amino acids mix) plates. Photos were taken after 5 days. The first spot contained 2 x10^4 cells. The rest were 5-fold dilutions. Growth from defective strains was inoculated in liquid YEEG at OD_{600} 0.2 and cultured 5 days to confirm phenotypes.

CPT mutants were spot tested on YES, CPT 5 µM, 10 µM and 15 µM plates. The first spot contained 3 x10^6 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 assemble, 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 above, 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 per 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 pre-cooled to 4°C in a 250 ml flask. The flask was placed in a room temperature shaking water bath at that was set for 32°C at 170 rpm for 5.5 hr. 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 hr to a density of 6 x 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 x 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 (~1.3 x 10^9 cells/ml) in freezer vials and stored at -80°C.

Additional strain constructions

S. pombe strain modifications were carried out using standard methods. For fusion of the right end and left end small 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 Supplemental Table 4). 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 *ade7*Δ control is a strain bearing a complete *ade7* ORF deletion (KRP389, which is KRP387 but *ade7*Δ*::arg3*$^{+}$). For transfer of the camptothecin-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 camptothecin 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. *S. 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 Sfi I 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 *E.*

coli DH5α to produce about 1-2 x10^5 bacterial colonies per transformation. Ten transformations were performed to generate 10 barcoded-*Hermes* transposon plasmid libraries. Each library was used to generate ~1,000 *S. pombe* insertion mutants (Fig. 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*. (Fig. 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). *S. pombe* cells bearing the transposase plasmid were grown under inducing conditions, then transformed with barcode-tagged transposon plasmids. Cells were allowed to grow for only 2-5 divisions to allow transposition, 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 9,024 mutant strains were picked into ninety six 96-well plates (Fig. 1A).

Development of a novel three-dimensional 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 three-dimensional (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 (Fig. 2A, stack of plates). Each Row pool consists of wells from the same row from each layer of plates (Fig. 2A, row pools). Each column pool consists of wells from the same column from each layer of plates (Fig. 2A, column pools). Finally, each layer pool consists of all of the wells in each layer of plates (Fig. 2A, layer pools). Each layer therefore contained 12 rows and 16 columns (Fig. 2A). Cells were collected as 16 column pools, 12 row pools and 12 layer pools. These 40 pools contained 3 copies of each mutant in a different row, column and layer pool. The three-dimensional coordinates of an identified mutant is determined by sequence comparisons between each row, column and layer pool in the 24 plate sub-library (Fig. 2A). A total of 4 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 (Supplemental Table 3). Second, each pool was amplified in three ways with primers specific to the *Hermes* left or right ends and to the 27 bp barcode (Fig. 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 (Fig. 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 Fig. 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 sequences 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 24 96-well plates. Each group has 2,256 mutants, each mutant has 3 associated sequences (barcode and *Hermes* left and right ends with adjacent genomic DNA), and each associated sequence was performed 3 times (once in each row, column and layer pool). Therefore, about 20,000 distinct products (2,256 x 3 x 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 24 96-well plates.

Ligation-mediated PCR (LM-PCR) was employed to amplify transposon flanking DNA sequences (Fig. 2B)(41-43). Genomic DNA from each pool was digested by different restriction enzymes, Mse I and Apo I/Mfe I, to increase the chance of capturing appropriately sized flanking genomic DNA fragments for sequencing libraries. After the ligation of double-strand 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 2 nucleotides so that the chances of mis-sorting due to a sequence miscall was 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 (Supplemental Fig.1), and the index tags and Illumina adaptors were sequentially incorporated by multiple rounds of PCR as described above. All PCR products were pooled and sequenced in a single lane on the Illumina Hiseq 2500. The entire insertion mutant collection of 9,024 mutants assembled in 96 plates was sequenced in a total of four lanes to yield 100 bp single-end reads (and examples of typical sequence reads are shown in Supplemental Fig. 1 B-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 (Fig. 2C; Supplemental Fig. 2). The raw sequence data was 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, 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, sub-telomere, 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 9,024 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.

S. pombe mutations generated from random insertions of the *Hermes* transposon

We successfully mapped *Hermes* transposon integrations sites and the associated DNA barcodes in 4,095 *S. pombe* mutants out of the 9,024 mutants sequenced. A total of 4,391 distinct insertion sites were recovered, as a fraction of the mutants contain more than one transposon insertion (Additional File 1). Over 90% of strains in the current collection carried a single transposon insertion, and ~70% of transposon insertions were in protein-coding genes and non-coding RNA genes. The remaining 30% of the insertions were in the intergenic regions (Fig. 3A; Additional File 1), 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 (Fig. 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 UTRs. Of the 2,753 insertions in protein-coding genes, 38% (1,057) were in the coding exons and introns, and 62% (1,696) 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)(Fig. 3C).

The distribution of *Hermes* insertions is different among essential genes and non-essential genes. About 59% of insertions into non-essential gene mapped to UTRs and 41% mapped to coding sequences (CDS). The insertions in essential genes were more enriched in UTRs (87%). Only 13% were in the protein-coding regions (Fig. 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 non-essential genes. There were more insertions in the middle of the CDS of non-essential genes than from essential genes (Fig. 4B). The 47 *Hermes* insertions in the coding regions of essential genes are shown in Fig. 4C.

The insertions in the first 150 bp of the CDS of several essential genes in these viable strains was 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 plus 1 nt at the Left End, and a 13 aa +2 nt ORF at the Right End (Fig. 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 house-keeping 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 (Fig. 5D).

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 with the remaining *ade7*$^{+}$ CDS (Fig. 5B). As the Ade7 N-terminus is predicted to have strong structural interactions with the full protein (45,46), the additional aa encoded by the short ORFs were not added (Fig. 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 (*ade7*Δ, Fig. 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 (Fig. 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 were mutated to TTC. The entire *ade7* ORF and 100 bps 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 *ade7*Δ strain, indicating that the –ATG constructs did not support *ade7* expression (Fig. 5E, Supplemental Fig. 4). 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 4 insertions within centromeres and 5 from chromosome I and II telomere regions (Supplemental Fig. 3A-3B). This result most likely reflects silencing of the *kanMX* gene, which impairs the selection for G418 resistance.

Strains bearing the 4,391 mapped insertions were transferred to a final set of ~44 96-well plates, although a small fraction did not regrow and were lost. Our final *Hermes* insertion collection contains 4,308 insertions with mutations in 268 essential genes (21% of S. pombe annotated essential genes), 1,472 non- essential genes (41% of non-essential genes), 589 non-coding genes (31% of non-coding genes) and 1,369 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, 2,470 non-essential gene insertions and 1,159 non-coding gene insertions and the remaining 1369 insertions in regions with no identified genes. Some strains have more than one insertion and some insertions affect more than one gene (genes and their associated insertions are summarized in Additional File 2).

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 non-fermentable carbon sources (2% glycerol, 2% ethanol) and resistance to the topoisomerase inhibitor Camptothecin (CPT). Normal growth on non-fermentable carbon sources requires an intact mitochondrial respiratory chain for carbon metabolism (47). Thus, mutants in respiratory chain complex genes are expected to have impaired growth on non-fermentable 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 1). All 3 mutants with insertions in coding exons, 2 mutants with insertions in introns and 3 with insertions in the 5' UTR showed defective growth (Fig. 6A). In contrast, 6 UTR mutants (3 each in the 5' and 3' UTRs) showed normal growth (Fig. 6B, Supplementary Table 1). Thus, our insertion library can identify regions of the UTRs important for gene function.

Camptothecin (CPT) is a topoisomerase inhibitor causing replication fork breakage when the replisome encounters the topoisomerase-CPT-DNA adduct (48,49). Deshpande et al. screened 2,662 *S. pombe* complete or partial ORF deletion mutants for growth on plates containing CPT and identified a set of 119 CPT-sensitivity genes (48). We searched our insertion library for mutations in the CPT-sensitivity gene set and

found 54 mutants with insertions in 37 genes (Supplementary Table 2). 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 (Fig. 7A, Supplemental Fig.5, and Supplementary Table 2). We observed 19 insertion mutants that were more sensitive to CPT than wild type cells and were less sensitive 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 (Fig. 7B). This mutant bore a transposon insertion in the 5' UTR of *SPBC16A3.17c,* 1278 bp upstream of the start codon. Two functionally unknown non-coding 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 (Fig. 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 (50,51), 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 if 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 colony color (Fig. 5C,8C)(52). Nonsense mutations in *ade6* that ablate enzyme function cause the formation of red colonies (e.g. M26, M375, 704 and L469, Fig. 8C). 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 2 red colonies in a background of ~6000 pink colonies were identified. The

barcodes identified both *ade6*$^{-}$ strains in the collection (Fig. 8B,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 (Fig. 8C,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 (53,54). Extracellular sulfate is converted to sulfide in three enzymatic steps, and sulfide is used by the Met17 enzyme to make homocysteine during sulfur metabolism (Fig. 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 (53,54). 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 than *S. cerevisiae*: almost all of the colonies were black-brown, except for three that were all white (Fig. 9A,C). These data indicated that *S. pombe* normally produces H_2S when provided with excess ammonium sulfate. An expert in eukaryotic sulfur metabolism suggested to us that the white colonies were mutants that disrupted the uptake and/or the reduction of sulfate to sulfide (C. Hine, pers. comm.)(55,56). 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 (57,58), 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

Three-dimensional 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 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 three-dimensional 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 ~9,200 strains (96 plates) could be processed at one time on a newer, higher-throughput platform, such as NovaSeq that generates over 2 x 10^9 reads.

The *Hermes* insertion mutant library is a multi-faceted 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 *Sfi I* 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 (59). 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

Supplemental Material and Supplemental Figures 6 to 9 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 (Fig. 7A, Supplementary Table 2). 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, approaches 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 non-essential genes for cell growth on synthetic medium, identifying essential genes as those with few or no transposon insertions (60). 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 were 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 (SGA) approaches (6,61), 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. Figs. 8,9). 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 quiet 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 prefers to insert into nucleosome-free regions in vivo (60). Our analysis of 9,024 mutants identified insertions in 2,753 protein coding genes out of a total of 5,131 (62). 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 Materials and Methods). 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 (63-65). The majority (~90%) of insertions were within the predicted 5' region of Pol III-transcribed genes (66). Recently, a rice miniature inverted repeat transposable elements (MITEs) was applied to *S. cerevisiae*. About 65% of the insertions were in genes (67). It is not known whether Ty1 or MITE are 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.

Declarations

Data Availability: The mutants described Additional File 1 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. The link to these data is: https://dataview.ncbi.nlm.nih.gov/object/PRJNA685113?reviewer=7oqtmqklbtshimhgvjfdv1k3a6

Supplementary Data: Supplementary data includes four tables and nine figures cited in the text and two additional files. Additional File 1 is a spreadsheet with information on each insertion mutant. Additional File 2 is a spreadsheet listing each mutated gene and the insertions in that gene as described in Additional File 1.

Competing Interests: The authors declare no competing financial interests.

Funding: This work was supported by National Institutes of Health R01's (grant numbers GM050752, AG051601 to KWR, HL55666, HL081093 to KLB). The CWRU Genomics Core, which performed the high-throughput sequencing, was supported in part by the Case Comprehensive Cancer Center (5P30CA043703).

Acknowledgements: The authors thank Dr. Valerie Wood (Pombase) for discussions of 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 Figs. 5, 8 and 9.

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**.
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**.
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) *Molecular Cloning: A Laboratory Manual*. Second ed. Cold Spring Harbor Laboratory, Cold Spring Harbor.
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 e2688.
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**, D695-699.
45. 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.
46. 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.
47. 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.
48. 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 (Amst)*, **8**, 672-679.
49. 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.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.
55. 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.
56. 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.
57. 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.

58. 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.
59. 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.
60. 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.
61. 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.
62. Wood, V. (2006) In Sunnerhagen, P. and Piskur, J. (eds.), *Comparative Genomics: Using Fungi as Models*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 233-285.
63. 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.
64. 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.
65. 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.
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. Fattash, I., Bhardwaj, P., Hui, C. and Yang, G. (2013) A rice Stowaway MITE for gene transfer in yeast. *PLoS One*, **8**, e64135.
68. Phosphoribosylaminoimidazole-succinocarboxamide synthase: AlphaFold structure prediction. DeepMind_Technologies_Limited https://alphafold.ebi.ac.uk/entry/Q9UUB4 Accessed on January 21, 2022.
69. 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.
70. 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.
71. 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.

72. 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.

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

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

	essential genes	nonessential genes	non-coding genes	intergenic insertions
S. pombe	1,260	3,576	1,876	
LR *Hermes* Library (total mutants)	363	2,470	1,159	1,369
LR *Hermes* Library (unique genes)	268	1,472	589	1,369
Haploid ORF Deletion Library*	0	3,308	N/A	0

*D Kim et al. Nat Biotechnol. 2010 Jun;28(6):617-23.

Figure Legends

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 2-5 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 (9,024 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 (59).

Figure 2. 3D pooling and multiplexed deep sequencing to map transposon integrations and DNA barcodes. (A) 3D pooling. Three copies of 24 plates (2,256 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 pools, 50 µl of cells were collected with a multi-channel pipettor, pooled and DNA was extracted (Materials and

Methods). 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 Mse I, Apo I and Mfe I 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 Supplemental Figure 1. **(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 (Supplemental Figure 2). The independent right end and left end reads served as an internal validation of the decoded transposon insertion sites.

Figure 3. *Hermes* transposon collection. (A) *Hermes* Transposon Library Statistics. A total of 9,024 mutants were sequenced, and 4,095 mutants were successfully mapped to the *S. pombe* genome. A total of 4,391 *Hermes* insertions were recovered from these 4,095 mutants. About 70% of the *Hermes* insertions were in protein-coding genes or non-coding 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 exon and introns of protein-coding genes). Chi Square statistics were used to compare differences between groups. A difference was taken as significant when a *p* value was less than 0.01.

Figure 4. The distribution of *Hermes* insertions in essential genes and non-essential 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, coding exons and introns in the essential genes and non-essential genes. **(B) The comparison of *Hermes* insertion distribution in essential gene and non-essential 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.

Fig. 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 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 (45,46,68), 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 (69-71). 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 (*ade7*Δ, 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 *ade7*Δ cells. Color balance and contrast of the limiting adenine pictures was adjusted to highlight the difference between white and pink colony color (45). Analysis of independently constructed +ATG-*ade7* and –ATG-*ade7* strains are shown in Supplemental Fig. 4.

Figure 6. Defective growth of *Hermes* respiratory chain mutants on non-fermentable carbon sources. (A) Spot test *Hermes* mutants on YES (fermentable

carbon source) and YEEG (non-fermentable 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 non-fermentable carbon source.** A summary of the insertion location and number of mutants with normal or defective growth on YEEG are shown.

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 (Supplemental Fig. 5 and Supplementary Table 2). 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 (strain P43A3, P62B1). Spot tests on 5, 10, 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.

Fig. 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 were plated on YE medium with limiting adenine to form ~6000 colonies. Two dark red colonies among the background of pink colonies were identified, the barcode DNA was amplified from each colony, sequenced and compared against Table 3 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 ATCGACAAACAAAAGAAAACGTAAATTGACATTTACAGAGA 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, -M375, -704, -L469*, red circles) all form red colonies while the missense mutations (*ade6-M216, -M210*, pink circles) form pink colonies on YE medium with limiting adenine (52,72). 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, 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).

Fig. 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 were 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, the barcode DNA was amplified from each colony, sequenced and compared against Additional File 1 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 which 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 the *Hermes::kanMX* insertion.

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,a], Neil Molyneaux[2], Haitao Zhang[3], Gang Zhou[5,b], Carly Kerr[3], Mark D. Adams[2,c], Kathleen L. Berkner[5,d] and Kurt W. Runge[1,2,e*]

[1]Department of Molecular Genetics, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 44195, USA

[2]Department of Genetics and Genomic Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio, 44106, USA

[3]Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic.

[4]Department of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 44195, USA

[5]Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 44195, USA

[a]Present address: Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas;

[b]Present address: Department of Pathology, Baylor Scott & White Health, Temple, Texas;

[c]Present address: The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut;

[d]Present address: Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic;

[e]Present address: Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic.

* To whom correspondence should be address. Tel: (216) 445-9771; Fax: (216) 444-0512; Email: rungek@ccf.org

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 both reduce the number of required sequencing reactions by orders of magnitude, and were able to identify the barcode sequences and DNA insertion sites of 4,391 *S. pombe* insertion mutations with only 40 sequencing preparations. The insertion mutations are in the genes and UTRs of non-essential, essential and non-coding 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.

Keywords: three-dimensional pooling; multiplexed high-throughput sequencing; insertion mutant library; DNA barcodes; *Hermes* transposon; *S. pombe* library.

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 are 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 yeast *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 non-coding 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 *Schizosaccharomyces 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, 3-dimensional 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 and barcodes in viable mutants. Each identified mutant and its barcode are 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 4,095 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 non-essential genes and

30% of the non-coding 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 means to rapidly generate defined mutation collections in other model systems.

Material & Methods

Construction of barcoded-*Hermes* transposon plasmids

Hermes transposon donor plasmid (pHL2577) and transposase plasmid (pHL2578) were 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 condition (1 min at 95°C, -1°C per cycle, 15 cycles; 1 min at 80°C, -0.5°C per cycle, 70 cycles; 1 min at 45°C, -0.5°C per cycle, 66 cycles to 12°C), followed by filling in to generate 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 then transformed into DH5α by electroporation. Ten separate transformations each produced 1-2 x10^{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

S. 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 hours then replica plated on YES+G418 (200 µg/ml) + FOA (1 g/L) and grown for 3 days. About 1,000 *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 formula: $P = 1- (1-f)^N$ (34), where:

f = 1/ (Number of barcoded-*Hermes* transposon plasmids)

N = number of *S. pombe* clones sampled

P = the probability of getting a barcode

1-P = the probability of not getting a barcode

This calculation predicts a >99% probability of not getting the same barcode if picking 1,000 *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 Discussion). Solving for N where P = 0.8 and f = (1/number of protein coding genes) X (fraction of insertions in protein coding genes/total number of insertions) = (1/5132) X (2273/9024) to yield 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 x 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 milliQ-filtered water, transferred to a 1.5 ml screw cap tube, cells were pelleted, 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, 50Mm EDTA, 1% SDS) and 500 µl 0.5mm 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 RNAase and proteinase K, genomic DNA was subjected to phenol/chloroform extraction and precipitated with ethanol.

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

Ligation-mediated PCR

The Mse I ds linkers were generated by annealing the upper strand oligo (5Phos/TAGTCCCTTAAGCGGAG/3AmM/-amino) to the lower strand oligo (5'GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC3'). Apo I and Mfe I ds linkers were generated by annealing the upper strand oligo (5Phos/AATTGTCCCTTAAGCGGAG/3AmM/-amino modified) to the lower strand oligo (5'GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC3'). 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, 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 Results). To amplify the *Hermes* transposon right end (HR) insertion sites, the first step was done by the HR outside primer (5'CTTGCACTCAAAAGGCTTGACAC3') specific to the transposon right end using the linker primer (5'GTAATACGACTCACTATAGGGCTC3') specific to the linkers using the following condition: 2 min at 98°C, 6 cycles of 15 sec at 98°C, 30 sec at 65°C, 40 sec at 72°C and then 24 cycles of 15 sec at 98°C, 30 sec at 60°C, 40 sec 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'CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGTAATACGACTCACTAT AGGGCT3') and an 8-bp indexed HR-nested primer (5'ACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXXXXTATGTGGCTT ACGTTTGCCTGTGG3'), which respectively adds one of the Illumina adaptors to PCR products. The conditions were 2 min at 98°C, 10 cycles of 15 sec at 98°C, 30 sec at 60°C, 40 sec 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'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGATCT3') to add Illumina-sequencing primer and another Illumina adaptor to the final product. The conditions were 2 min at 98°C, 10 cycles of 15 sec at 98°C, 1min 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'ACACTCTTTCCCTACACGACGCTCTTCCGATCT**XXXXXXX**TATCCCGGGATTTTG GCCAC3') and barcode reverse primer (5'CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCTGCAGCGAGGAGCCG TAAT3') using the following conditions: 2 min at 98°C, 30 cycles of 15 sec at 98°C, 30 sec at 60°C, 30 sec 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 sec at 98°C, 30 sec 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 (Supplemental Fig. 1). Equal molar amounts of products were mixed. Customized index tags are presented in Supplemental Table 3.

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 (Supplemental Fig. 2). The raw sequence data was 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'TATCCCGGGATTTTGGCCACCCGGGCC3'), transposon right end (5'TATGTGGCTTACGTTTGCCTGTGGCTTGTTGAAGTTCTCTG3') or left end (5'GCGCATAAGTATCAAAATAAGCCACTTGTTGTTGTTCTCTG3'). 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 Results).

A list of the insertion sites and barcode sequences associated with each insertion are provided as Additional File 1. A list of mutated genes is in Additional File 2.

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'CAAGACTAGGAAAAGAGCATAAG3') and 4171as (5'GACTGTCAAGGAGGGTATTC3') 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

2,328 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). *S. pombe* mutants carrying 15 genes under this GO term were spot tested on YES and non-fermentable YEEG (0.5% yeast extract, 2% glycerol, 2% ethanol, 2g/L casein amino acids, amino acids mix) plates. Photos were taken after 5 days. The first spot contained 2 x10^4 cells. The rest were 5-fold dilutions. Growth from defective strains was inoculated in liquid YEEG at OD_{600} 0.2 and cultured 5 days to confirm phenotypes.

CPT mutants were spot tested on YES, CPT 5 µM, 10 µM and 15 µM plates. The first spot contained 3 x10^6 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 assemble, 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 above, 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 per 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 pre-cooled to 4°C in a 250 ml flask. The flask was placed in a room temperature shaking water bath at that was set for 32°C at 170 rpm for 5.5 hr. 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 hr to a density of 6 x 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 x 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 (~1.3 x 10^9 cells/ml) in freezer vials and stored at -80°C.

Additional strain constructions

S. pombe strain modifications were carried out using standard methods. For fusion of the right end and left end small 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 Supplemental Table 4). 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 *ade7*Δ control is a strain bearing a complete *ade7* ORF deletion (KRP389, which is KRP387 but *ade7*Δ*::arg3*$^{+}$). For transfer of the camptothecin-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 camptothecin 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. *S. 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 Sfi I 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 *E.*

coli DH5α to produce about 1-2 x10^5 bacterial colonies per transformation. Ten transformations were performed to generate 10 barcoded-*Hermes* transposon plasmid libraries. Each library was used to generate ~1,000 *S. pombe* insertion mutants (Fig. 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*. (Fig. 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). *S. pombe* cells bearing the transposase plasmid were grown under inducing conditions, then transformed with barcode-tagged transposon plasmids. Cells were allowed to grow for only 2-5 divisions to allow transposition, 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 9,024 mutant strains were picked into ninety six 96-well plates (Fig. 1A).

Development of a novel three-dimensional 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 three-dimensional (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 (Fig. 2A, stack of plates). Each Row pool consists of wells from the same row from each layer of plates (Fig. 2A, row pools). Each column pool consists of wells from the same column from each layer of plates (Fig. 2A, column pools). Finally, each layer pool consists of all of the wells in each layer of plates (Fig. 2A, layer pools). Each layer therefore contained 12 rows and 16 columns (Fig. 2A). Cells were collected as 16 column pools, 12 row pools and 12 layer pools. These 40 pools contained 3 copies of each mutant in a different row, column and layer pool. The three-dimensional coordinates of an identified mutant is determined by sequence comparisons between each row, column and layer pool in the 24 plate sub-library (Fig. 2A). A total of 4 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 (Supplemental Table 3). Second, each pool was amplified in three ways with primers specific to the *Hermes* left or right ends and to the 27 bp barcode (Fig. 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 (Fig. 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 Fig. 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 sequences 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 24 96-well plates. Each group has 2,256 mutants, each mutant has 3 associated sequences (barcode and *Hermes* left and right ends with adjacent genomic DNA), and each associated sequence was performed 3 times (once in each row, column and layer pool). Therefore, about 20,000 distinct products (2,256 x 3 x 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 24 96-well plates.

Ligation-mediated PCR (LM-PCR) was employed to amplify transposon flanking DNA sequences (Fig. 2B)(41-43). Genomic DNA from each pool was digested by different restriction enzymes, Mse I and Apo I/Mfe I, to increase the chance of capturing appropriately sized flanking genomic DNA fragments for sequencing libraries. After the ligation of double-strand 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 2 nucleotides so that the chances of mis-sorting due to a sequence miscall was 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 (Supplemental Fig.1), and the index tags and Illumina adaptors were sequentially incorporated by multiple rounds of PCR as described above. All PCR products were pooled and sequenced in a single lane on the Illumina Hiseq 2500. The entire insertion mutant collection of 9,024 mutants assembled in 96 plates was sequenced in a total of four lanes to yield 100 bp single-end reads (and examples of typical sequence reads are shown in Supplemental Fig. 1 B-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 (Fig. 2C; Supplemental Fig. 2). The raw sequence data was 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, 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, sub-telomere, 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 9,024 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.

S. pombe mutations generated from random insertions of the *Hermes* transposon

We successfully mapped *Hermes* transposon integrations sites and the associated DNA barcodes in 4,095 *S. pombe* mutants out of the 9,024 mutants sequenced. A total of 4,391 distinct insertion sites were recovered, as a fraction of the mutants contain more than one transposon insertion (Additional File 1). Over 90% of strains in the current collection carried a single transposon insertion, and ~70% of transposon insertions were in protein-coding genes and non-coding RNA genes. The remaining 30% of the insertions were in the intergenic regions (Fig. 3A; Additional File 1), 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 (Fig. 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 UTRs. Of the 2,753 insertions in protein-coding genes, 38% (1,057) were in the coding exons and introns, and 62% (1,696) 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)(Fig. 3C).

The distribution of *Hermes* insertions is different among essential genes and non-essential genes. About 59% of insertions into non-essential gene mapped to UTRs and 41% mapped to coding sequences (CDS). The insertions in essential genes were more enriched in UTRs (87%). Only 13% were in the protein-coding regions (Fig. 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 non-essential genes. There were more insertions in the middle of the CDS of non-essential genes than from essential genes (Fig. 4B). The 47 *Hermes* insertions in the coding regions of essential genes are shown in Fig. 4C.

The insertions in the first 150 bp of the CDS of several essential genes in these viable strains was 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 plus 1 nt at the Left End, and a 13 aa +2 nt ORF at the Right End (Fig. 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 house-keeping 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 (Fig. 5D).

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 with the remaining *ade7*$^{+}$ CDS (Fig. 5B). As the Ade7 N-terminus is predicted to have strong structural interactions with the full protein (45,46), the additional aa encoded by the short ORFs were not added (Fig. 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 (*ade7*Δ, Fig. 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 (Fig. 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 were mutated to TTC. The entire *ade7* ORF and 100 bps 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 *ade7*Δ strain, indicating that the –ATG constructs did not support *ade7* expression (Fig. 5E, Supplemental Fig. 4). 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 4 insertions within centromeres and 5 from chromosome I and II telomere regions (Supplemental Fig. 3A-3B). This result most likely reflects silencing of the *kanMX* gene, which impairs the selection for G418 resistance.

Strains bearing the 4,391 mapped insertions were transferred to a final set of ~44 96-well plates, although a small fraction did not regrow and were lost. Our final *Hermes* insertion collection contains 4,308 insertions with mutations in 268 essential genes (21% of S. pombe annotated essential genes), 1,472 non- essential genes (41% of non-essential genes), 589 non-coding genes (31% of non-coding genes) and 1,369 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, 2,470 non-essential gene insertions and 1,159 non-coding gene insertions and the remaining 1369 insertions in regions with no identified genes. Some strains have more than one insertion and some insertions affect more than one gene (genes and their associated insertions are summarized in Additional File 2).

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 non-fermentable carbon sources (2% glycerol, 2% ethanol) and resistance to the topoisomerase inhibitor Camptothecin (CPT). Normal growth on non-fermentable carbon sources requires an intact mitochondrial respiratory chain for carbon metabolism (47). Thus, mutants in respiratory chain complex genes are expected to have impaired growth on non-fermentable 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 1). All 3 mutants with insertions in coding exons, 2 mutants with insertions in introns and 3 with insertions in the 5' UTR showed defective growth (Fig. 6A). In contrast, 6 UTR mutants (3 each in the 5' and 3' UTRs) showed normal growth (Fig. 6B, Supplementary Table 1). Thus, our insertion library can identify regions of the UTRs important for gene function.

Camptothecin (CPT) is a topoisomerase inhibitor causing replication fork breakage when the replisome encounters the topoisomerase-CPT-DNA adduct (48,49). Deshpande et al. screened 2,662 *S. pombe* complete or partial ORF deletion mutants for growth on plates containing CPT and identified a set of 119 CPT-sensitivity genes (48). We searched our insertion library for mutations in the CPT-sensitivity gene set and

found 54 mutants with insertions in 37 genes (Supplementary Table 2). 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 (Fig. 7A, Supplemental Fig.5, and Supplementary Table 2). We observed 19 insertion mutants that were more sensitive to CPT than wild type cells and were less sensitive 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 (Fig. 7B). This mutant bore a transposon insertion in the 5' UTR of *SPBC16A3.17c,* 1278 bp upstream of the start codon. Two functionally unknown non-coding 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 (Fig. 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 (50,51), 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 if 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 colony color (Fig. 5C,8C)(52). Nonsense mutations in *ade6* that ablate enzyme function cause the formation of red colonies (e.g. M26, M375, 704 and L469, Fig. 8C). 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 2 red colonies in a background of ~6000 pink colonies were identified. The

barcodes identified both *ade6*$^{-}$ strains in the collection (Fig. 8B,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 (Fig. 8C,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 (53,54). Extracellular sulfate is converted to sulfide in three enzymatic steps, and sulfide is used by the Met17 enzyme to make homocysteine during sulfur metabolism (Fig. 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 (53,54). 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 than *S. cerevisiae*: almost all of the colonies were black-brown, except for three that were all white (Fig. 9A,C). These data indicated that *S. pombe* normally produces H_2S when provided with excess ammonium sulfate. An expert in eukaryotic sulfur metabolism suggested to us that the white colonies were mutants that disrupted the uptake and/or the reduction of sulfate to sulfide (C. Hine, pers. comm.)(55,56). 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 (57,58), 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

Three-dimensional 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 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 three-dimensional 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 ~9,200 strains (96 plates) could be processed at one time on a newer, higher-throughput platform, such as NovaSeq that generates over 2 x 10^9 reads.

The *Hermes* insertion mutant library is a multi-faceted 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 *Sfi I* 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 (59). 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

Supplemental Material and Supplemental Figures 6 to 9 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 (Fig. 7A, Supplementary Table 2). 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, approaches 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 non-essential genes for cell growth on synthetic medium, identifying essential genes as those with few or no transposon insertions (60). 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 were 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 (SGA) approaches (6,61), 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. Figs. 8,9). 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 quiet 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 prefers to insert into nucleosome-free regions in vivo (60). Our analysis of 9,024 mutants identified insertions in 2,753 protein coding genes out of a total of 5,131 (62). 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 Materials and Methods). 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 (63-65). The majority (~90%) of insertions were within the predicted 5' region of Pol III-transcribed genes (66). Recently, a rice miniature inverted repeat transposable elements (MITEs) was applied to *S. cerevisiae*. About 65% of the insertions were in genes (67). It is not known whether Ty1 or MITE are 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.

Declarations

Data Availability: The mutants described Additional File 1 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. The link to these data is: https://dataview.ncbi.nlm.nih.gov/object/PRJNA685113?reviewer=7oqtmqklbtshimhgvjfdv1k3a6

Supplementary Data: Supplementary data includes four tables and nine figures cited in the text and two additional files. Additional File 1 is a spreadsheet with information on each insertion mutant. Additional File 2 is a spreadsheet listing each mutated gene and the insertions in that gene as described in Additional File 1.

Competing Interests: The authors declare no competing financial interests.

Funding: This work was supported by National Institutes of Health R01's (grant numbers GM050752, AG051601 to KWR, HL55666, HL081093 to KLB). The CWRU Genomics Core, which performed the high-throughput sequencing, was supported in part by the Case Comprehensive Cancer Center (5P30CA043703).

Acknowledgements: The authors thank Dr. Valerie Wood (Pombase) for discussions of 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 Figs. 5, 8 and 9.

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**.
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**.
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) *Molecular Cloning: A Laboratory Manual*. Second ed. Cold Spring Harbor Laboratory, Cold Spring Harbor.
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 e2688.
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**, D695-699.
45. 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.
46. 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.
47. 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.
48. 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 (Amst)*, **8**, 672-679.
49. 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.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.
55. 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.
56. 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.
57. 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.

58. 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.
59. 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.
60. 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.
61. 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.
62. Wood, V. (2006) In Sunnerhagen, P. and Piskur, J. (eds.), *Comparative Genomics: Using Fungi as Models*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 233-285.
63. 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.
64. 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.
65. 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.
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. Fattash, I., Bhardwaj, P., Hui, C. and Yang, G. (2013) A rice Stowaway MITE for gene transfer in yeast. *PLoS One*, **8**, e64135.
68. Phosphoribosylaminoimidazole-succinocarboxamide synthase: AlphaFold structure prediction. DeepMind_Technologies_Limited https://alphafold.ebi.ac.uk/entry/Q9UUB4 Accessed on January 21, 2022.
69. 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.
70. 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.
71. 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.

72. 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.

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

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

	essential genes	nonessential genes	non-coding genes	intergenic insertions
S. pombe	1,260	3,576	1,876	
LR *Hermes* Library (total mutants)	363	2,470	1,159	1,369
LR *Hermes* Library (unique genes)	268	1,472	589	1,369
Haploid ORF Deletion Library*	0	3,308	N/A	0

*D Kim et al. Nat Biotechnol. 2010 Jun;28(6):617-23.

Figure Legends

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 2-5 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 (9,024 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 (59).

Figure 2. 3D pooling and multiplexed deep sequencing to map transposon integrations and DNA barcodes. (A) 3D pooling. Three copies of 24 plates (2,256 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 pools, 50 µl of cells were collected with a multi-channel pipettor, pooled and DNA was extracted (Materials and

Methods). 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 Mse I, Apo I and Mfe I 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 Supplemental Figure 1. **(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 (Supplemental Figure 2). The independent right end and left end reads served as an internal validation of the decoded transposon insertion sites.

Figure 3. *Hermes* transposon collection. (A) *Hermes* Transposon Library Statistics. A total of 9,024 mutants were sequenced, and 4,095 mutants were successfully mapped to the *S. pombe* genome. A total of 4,391 *Hermes* insertions were recovered from these 4,095 mutants. About 70% of the *Hermes* insertions were in protein-coding genes or non-coding 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 exon and introns of protein-coding genes). Chi Square statistics were used to compare differences between groups. A difference was taken as significant when a *p* value was less than 0.01.

Figure 4. The distribution of *Hermes* insertions in essential genes and non-essential 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, coding exons and introns in the essential genes and non-essential genes. **(B) The comparison of *Hermes* insertion distribution in essential gene and non-essential 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.

Fig. 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 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 (45,46,68), 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 (69-71). 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 (*ade7*Δ, 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 *ade7*Δ cells. Color balance and contrast of the limiting adenine pictures was adjusted to highlight the difference between white and pink colony color (45). Analysis of independently constructed +ATG-*ade7* and –ATG-*ade7* strains are shown in Supplemental Fig. 4.

Figure 6. Defective growth of *Hermes* respiratory chain mutants on non-fermentable carbon sources. (A) Spot test *Hermes* mutants on YES (fermentable

carbon source) and YEEG (non-fermentable 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 non-fermentable carbon source.** A summary of the insertion location and number of mutants with normal or defective growth on YEEG are shown.

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 (Supplemental Fig. 5 and Supplementary Table 2). 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 (strain P43A3, P62B1). Spot tests on 5, 10, 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.

Fig. 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 were plated on YE medium with limiting adenine to form ~6000 colonies. Two dark red colonies among the background of pink colonies were identified, the barcode DNA was amplified from each colony, sequenced and compared against Table 3 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 ATCGACAAACAAAAGAAAACGTAAATTGACATTTACAGAGA 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, -M375, -704, -L469*, red circles) all form red colonies while the missense mutations (*ade6-M216, -M210*, pink circles) form pink colonies on YE medium with limiting adenine (52,72). 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

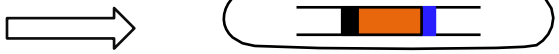
communications from several labs (Wayne Wahls, Ramsay McFarlane, Susan Forsburg, Mikel Zaratiegui) but has not been published. We confirmed these earlier observations by sequencing the region of *ade6* containing this mutation from two *ade6*$^{+}$ genes (strain L972 from J. Kohli and KRP387 from our lab) and three *ade6-M210* alleles (strains GP201 from Gerry Smith, FY1645 from Robin Allshire, and KRP2 from JoAnn Wise).

Fig. 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 were 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, the barcode DNA was amplified from each colony, sequenced and compared against Additional File 1 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 which identified the P4B9 strain with an insertion in

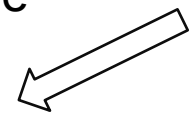
met10^{+}, a component of the sulfite reductase enzyme that produces sulfide. **D.** A schematic of the *met10*$^{+}$ gene ORF showing the location the *Hermes::kanMX* insertion.

Barcode-tagged Hermes Transposon

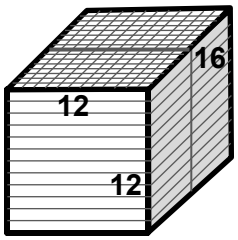
randomly inserted into
S. pombe genome



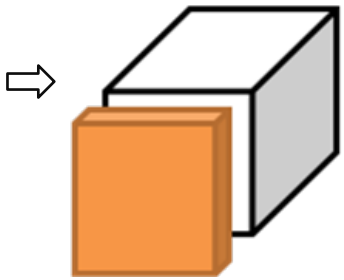
Array individual *S. pombe* mutants
into 96-well plates



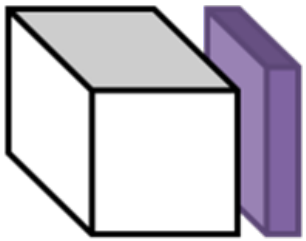
3D Pooling: Stack Plates



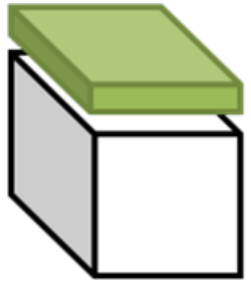
Row Pool



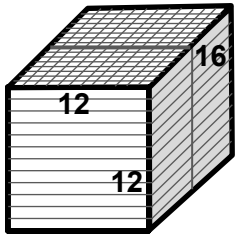
Column Pool



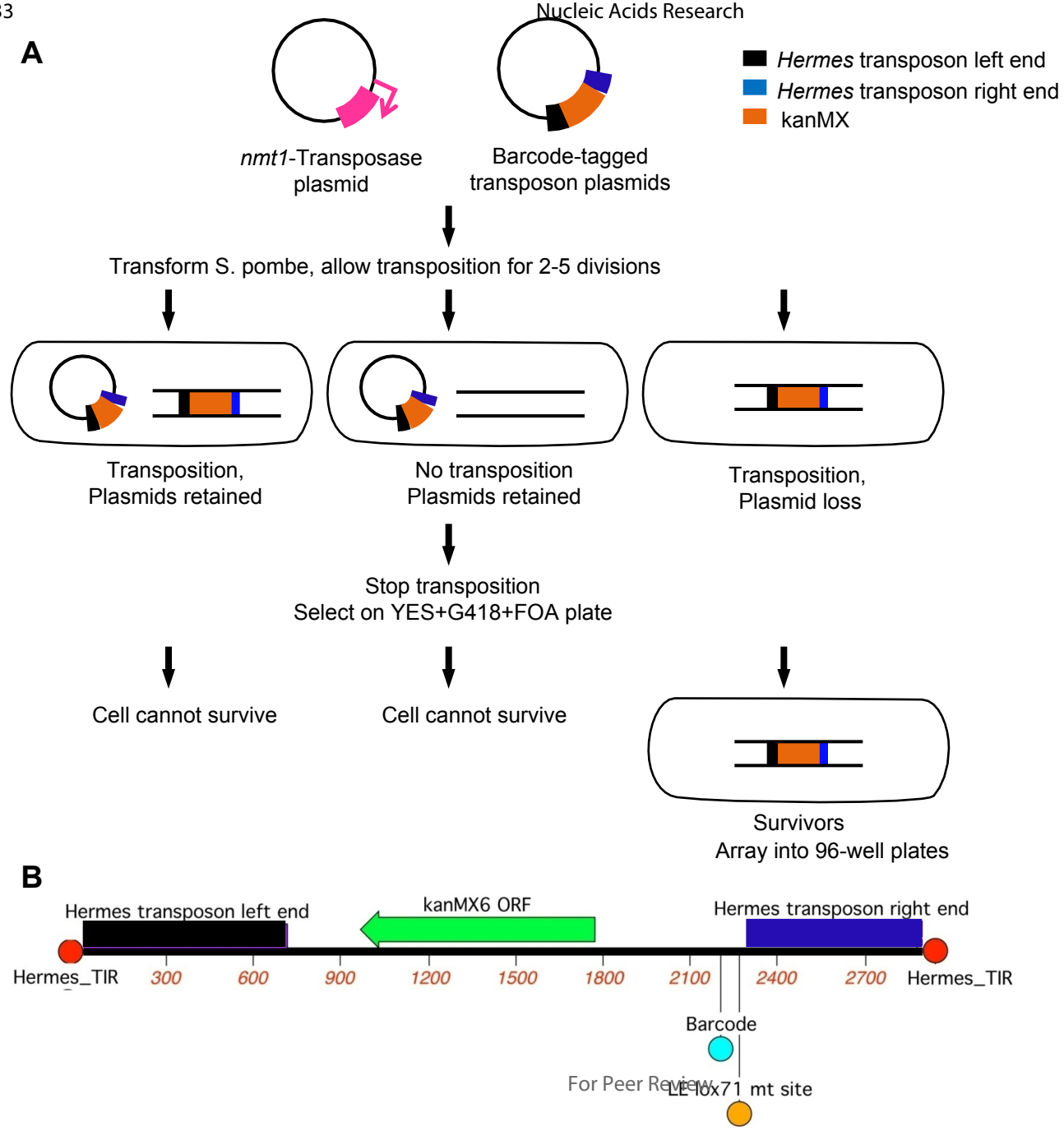
Layer Pool

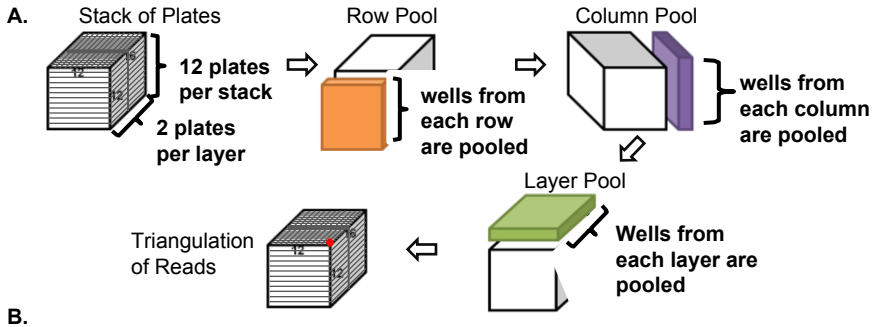


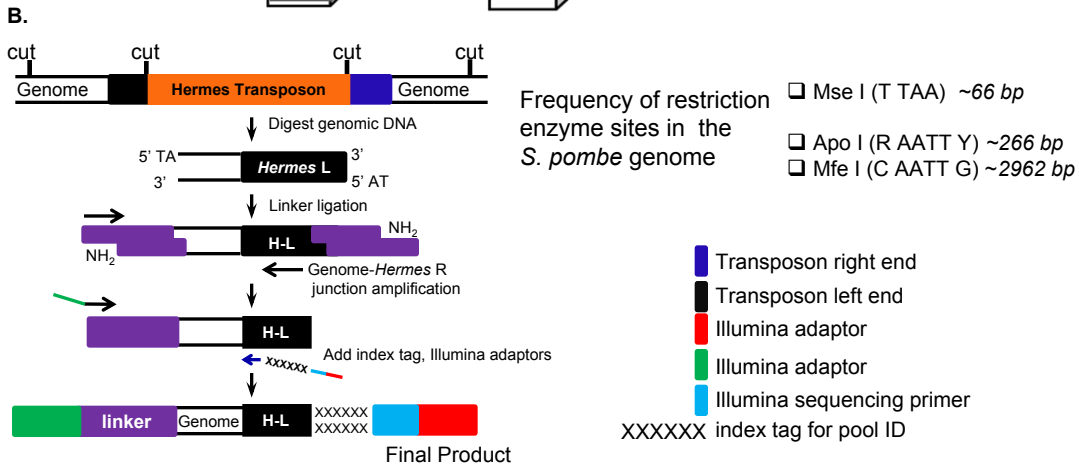
Multiplexed sequencing
barcodes and Transposon
ends

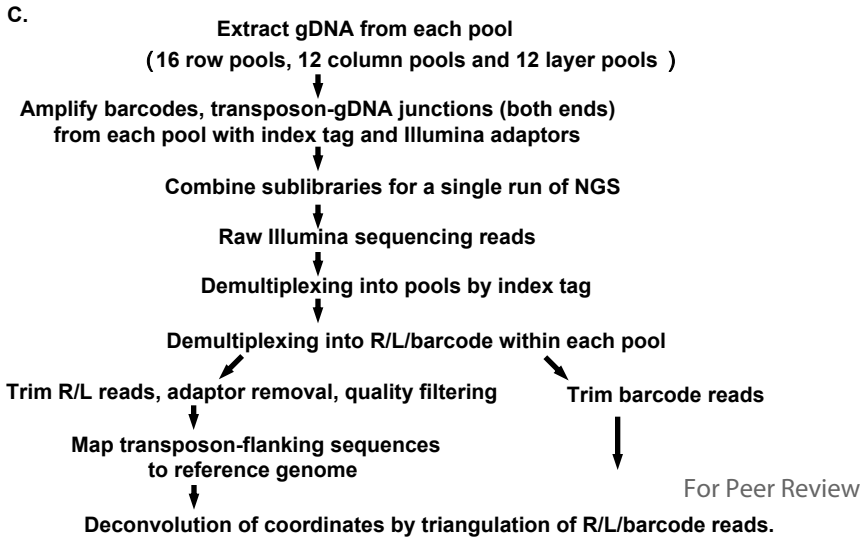


Decode each mutant by
triangulation of reads



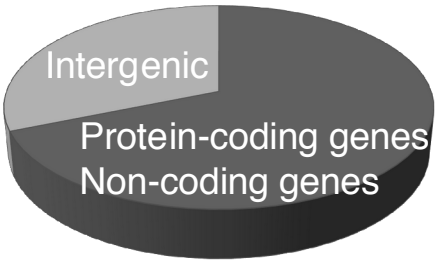




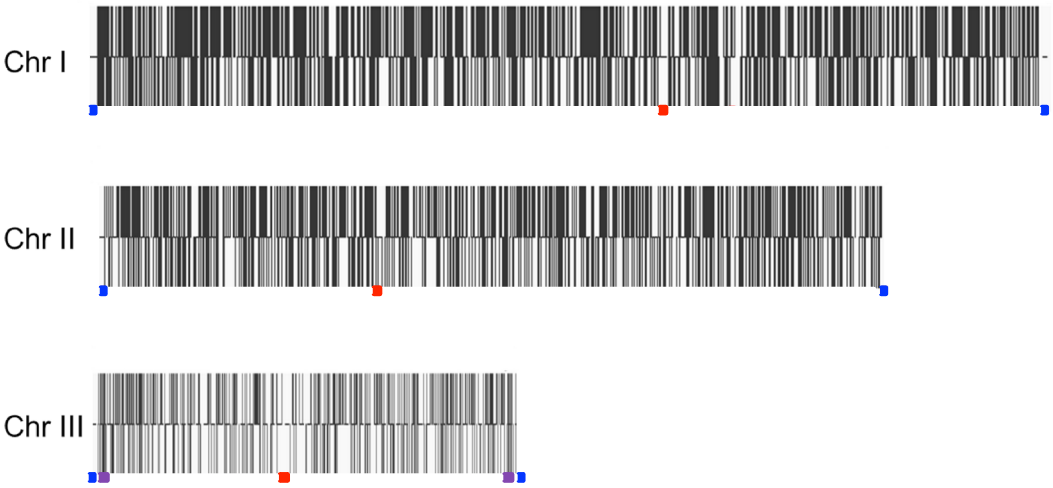


A.

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%)

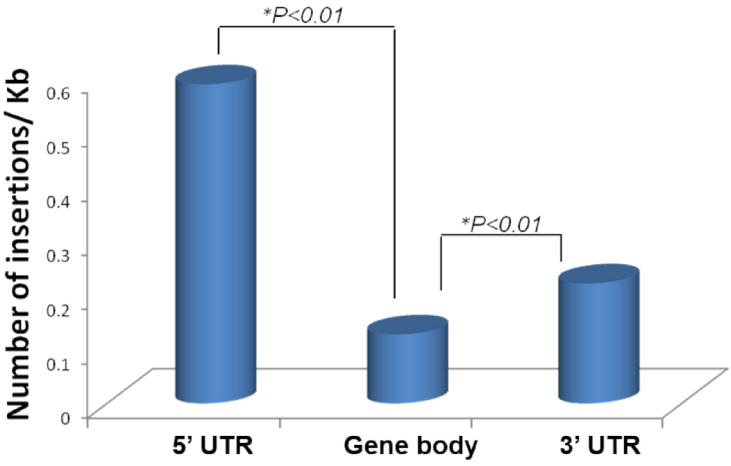


B.

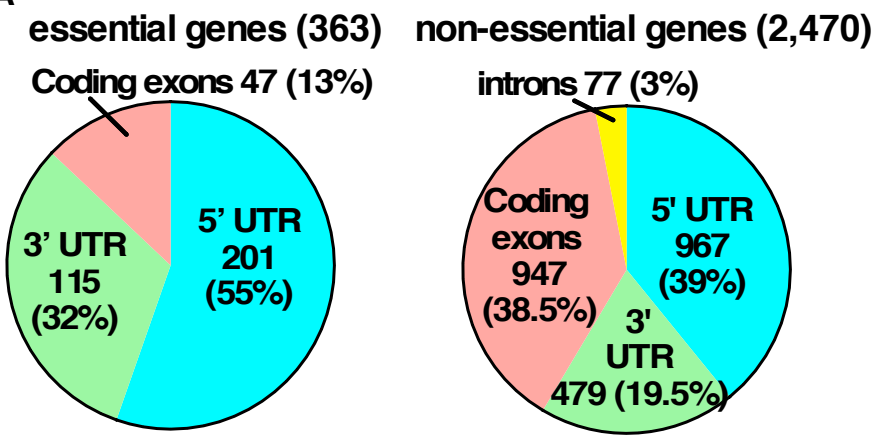


C.

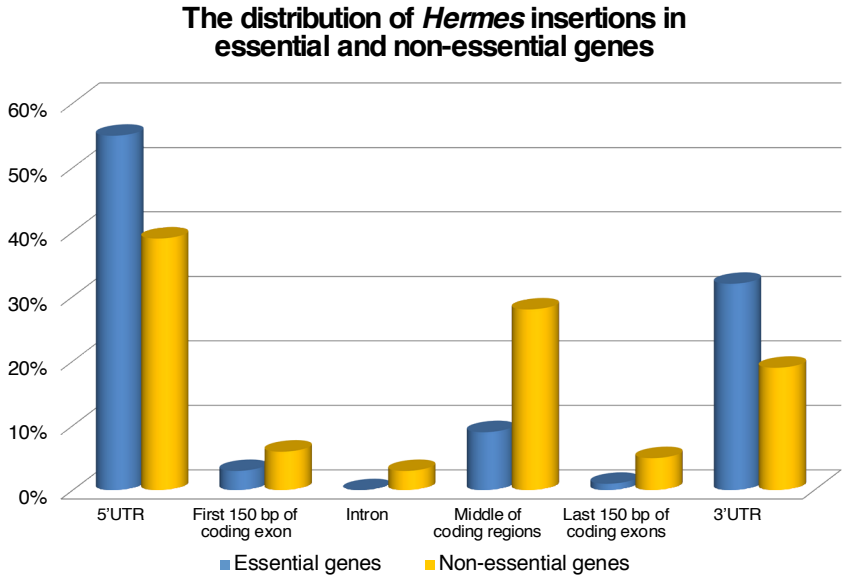
Protein-coding gene mutations	**2,753**
UTR mutants	1,696
5'UTRs	1,113
3'UTRs	583
Coding Exon mutants	1,000
Intron mutants	57



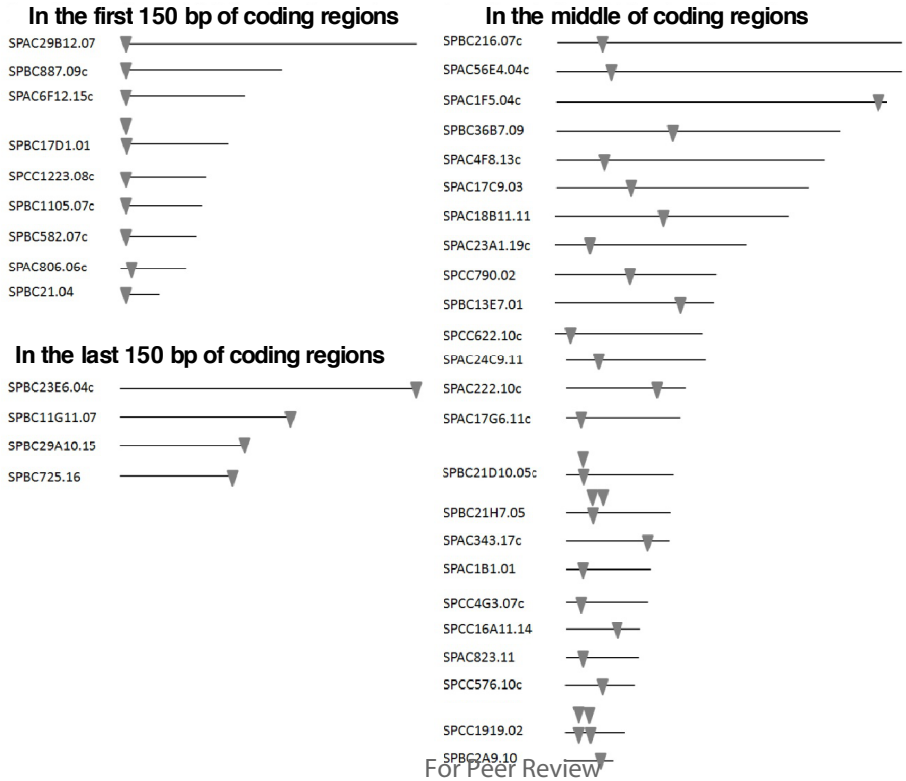
A


essential genes (363)
Coding exons 47 (13%)
3' UTR 115 (32%)
5' UTR 201 (55%)
non-essential genes (2,470)
introns 77 (3%)
Coding exons 947 (38.5%)
5' UTR 967 (39%)
3' UTR 479 (19.5%)

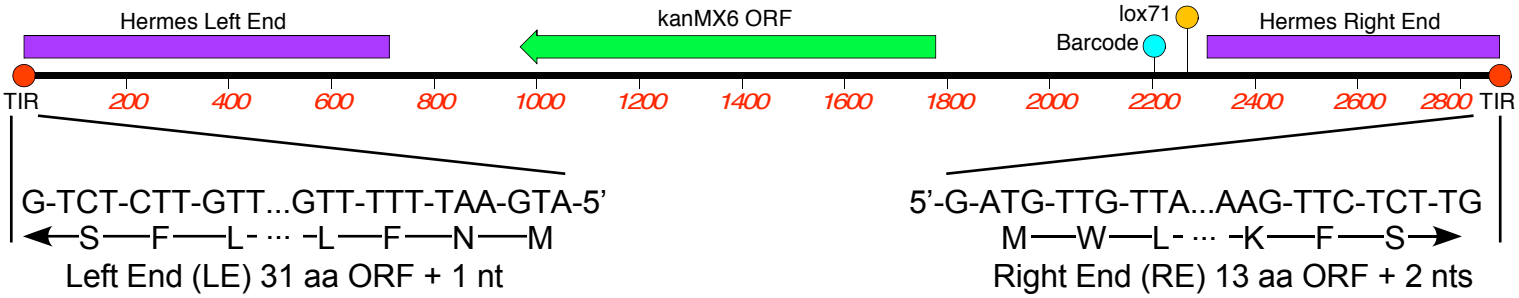
B


The distribution of Hermes insertions in essential and non-essential genes
60%
50%
40%
30%
20%
10%
0%
5'UTR
First 150 bp of coding exon
Intron
Middle of coding regions
Last 150 bp of coding exons
3'UTR
Essential genes
Non-essential genes

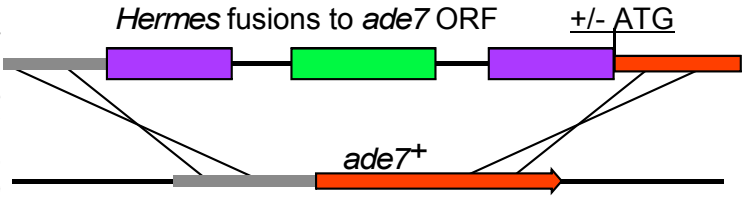
C


In the first 150 bp of coding regions
SPAC29B12.07
SPBC887.09c
SPAC6F12.15c
SPBC17D1.01
SPCC1223.08c
SPBC1105.07c
SPBC582.07c
SPAC806.06c
SPBC21.04
In the last 150 bp of coding regions
SPBC23E6.04c
SPBC11G11.07
SPBC29A10.15
SPBC725.16
In the middle of coding regions
SPBC216.07c
SPAC56E4.04c
SPAC1F5.04c
SPBC36B7.09
SPAC4F8.13c
SPAC17C9.03
SPAC18B11.11
SPAC23A1.19c
SPCC790.02
SPBC13E7.01
SPCC622.10c
SPAC24C9.11
SPAC222.10c
SPAC17G6.11c
SPBC21D10.05c
SPBC21H7.05
SPAC343.17c
SPAC1B1.01
SPCC4G3.07c
SPCC16A11.14
SPAC823.11
SPCC576.10c
SPCC1919.02
SPBC2A9.10

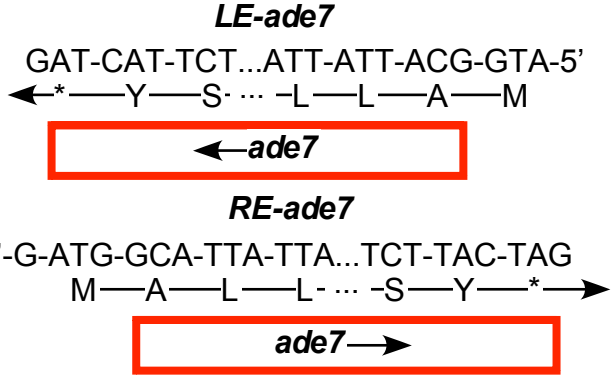
A. The outward facing ORFs of *Hermes*



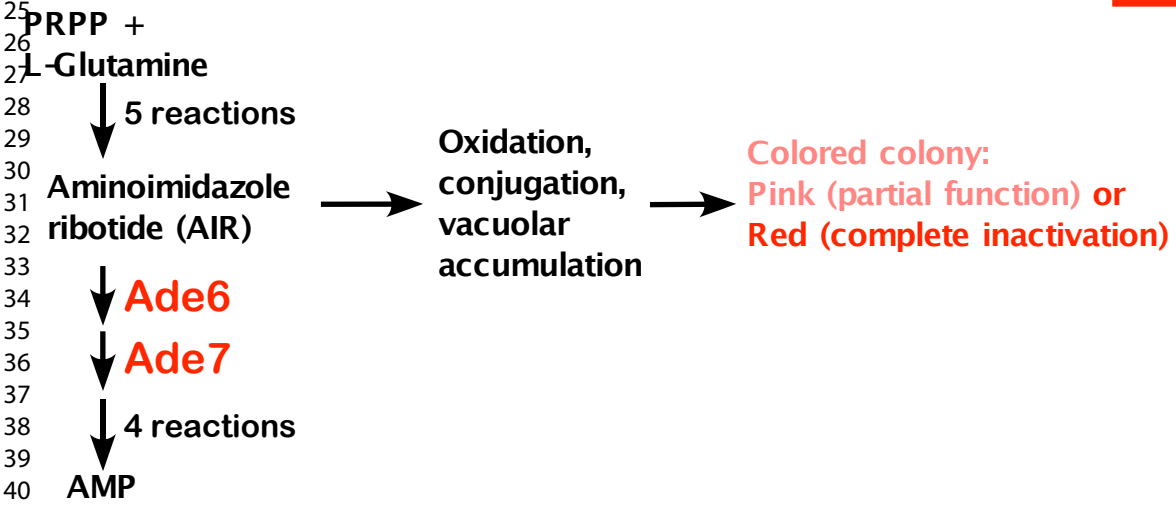
B. Integration of *Hermes-ade7* fusions



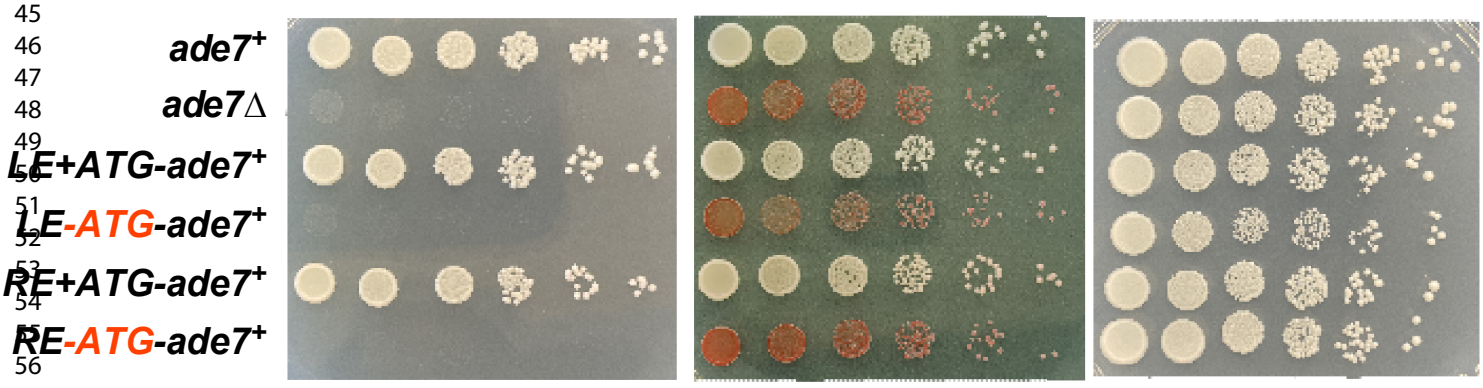
C. *ade7*+ ORF fusions to *Hermes* LE and RE ORFs

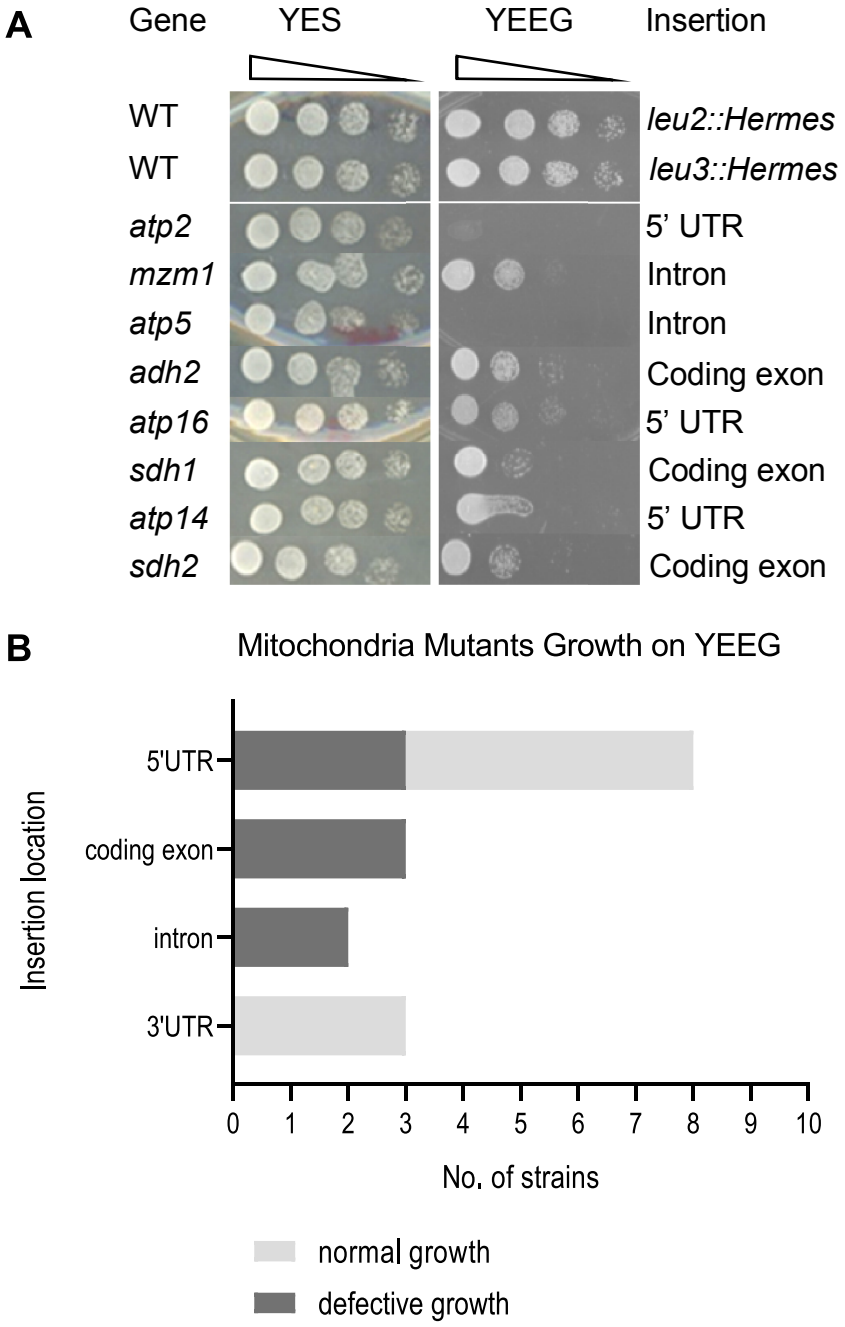


D. Blocks in adenine synthesis by *ade6* and *ade7* mutants cause pink or red colony color.

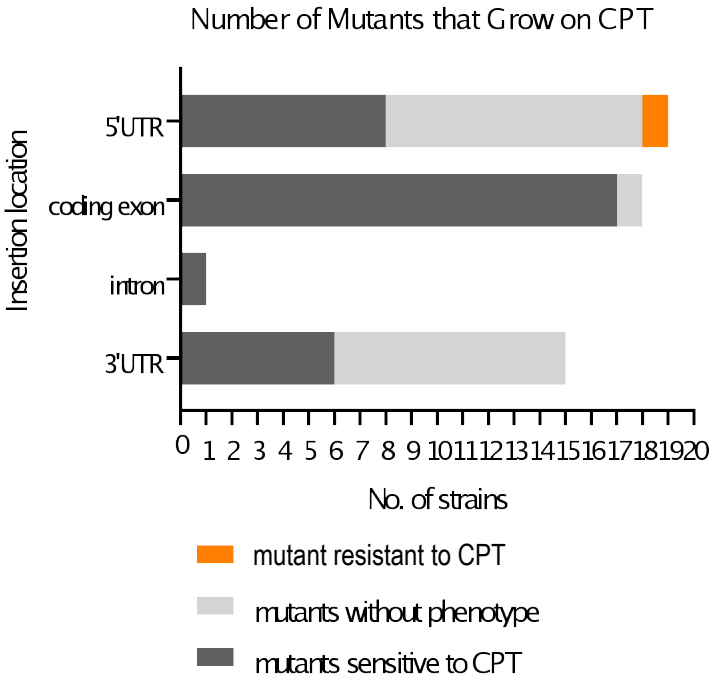


E. Phenotypes of independent fusions of *ade7* to the left end or right end of *Hermes::kanMX*

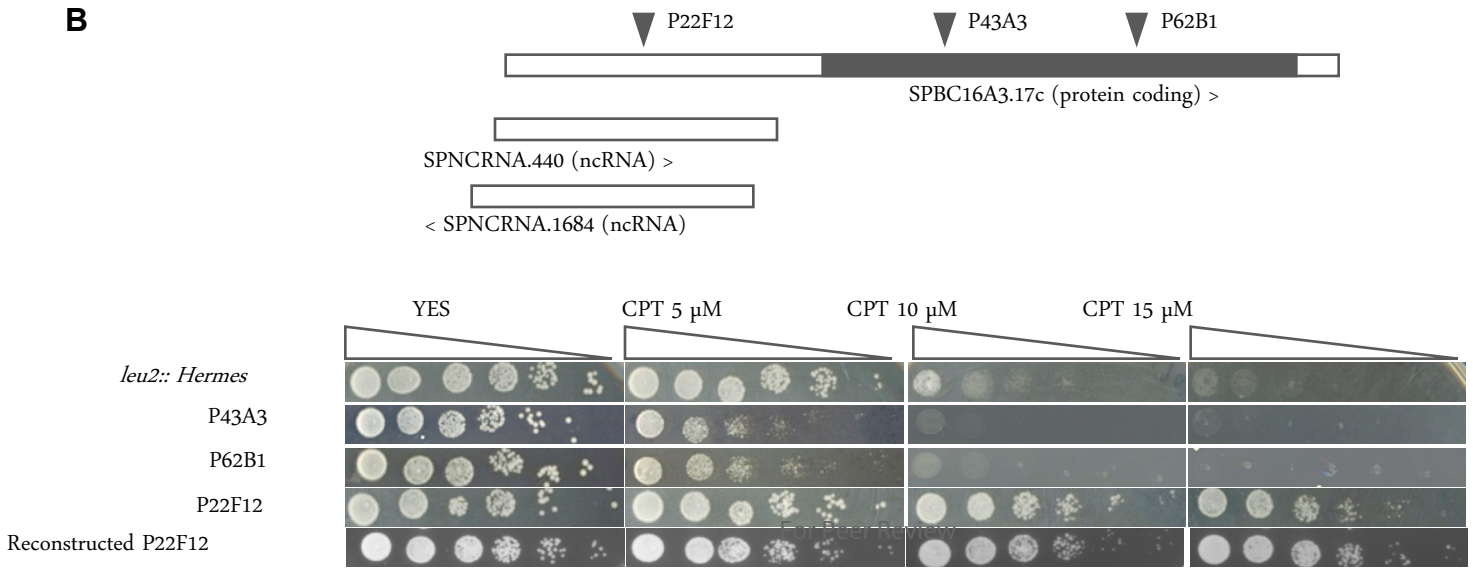



A
Gene
YES
YEEG
Insertion
WT
WT
atp2
mzm1
atp5
adh2
atp16
sdh1
atp14
sdh2
leu2::Hermes
leu3::Hermes
5' UTR
Intron
Intron
Coding exon
5' UTR
Coding exon
5' UTR
Coding exon
B
Mitochondria Mutants Growth on YEEG
Insertion location
5'UTR
coding exon
intron
3'UTR
0
1
2
3
4
5
6
7
8
9
10
No. of strains
normal growth
defective growth

A

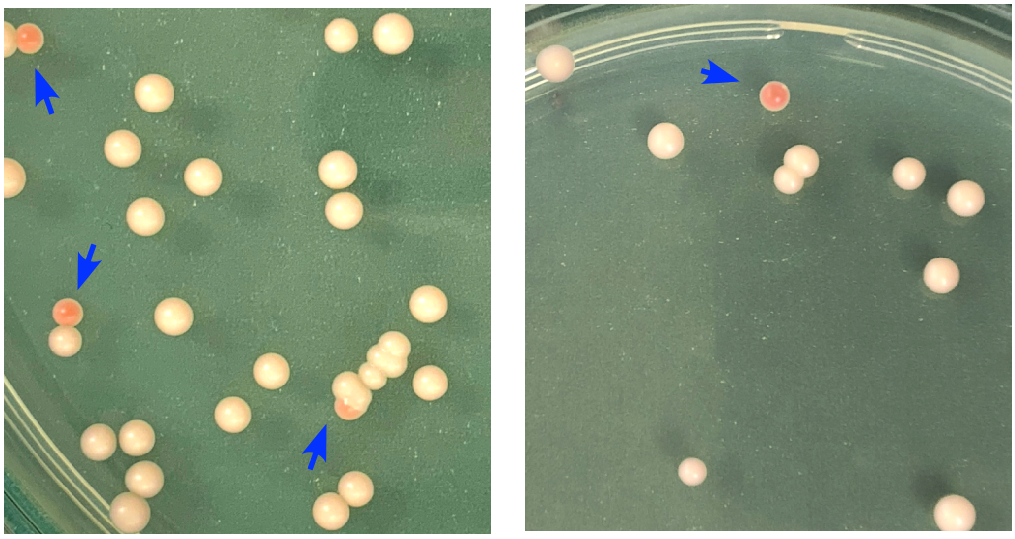

Number of Mutants that Grow on CPT
Insertion location
5'UTR
coding exon
intron
3'UTR
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
No. of strains
mutant resistant to CPT
mutants without phenotype
mutants sensitive to CPT

B


P22F12
P43A3
P62B1
SPBC16A3.17c (protein coding) >
SPNCRNA.440 (ncRNA) >
< SPNCRNA.1684 (ncRNA)
YES
CPT 5 µM
CPT 10 µM
CPT 15 µM
leu2:: Hermes
P43A3
P62B1
P22F12
Reconstructed P22F12

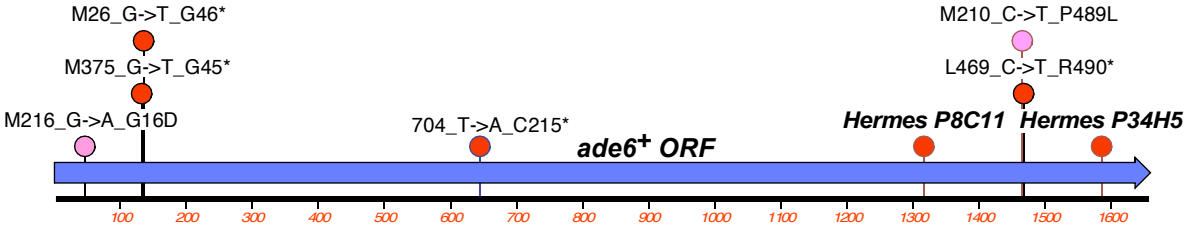
A. Pool of insertion mutants → Plate on YE plates with limiting adenine → Pick red colonies to screen for barcodes

B.



Hermes P34H5 *Hermes P8C11*

C.



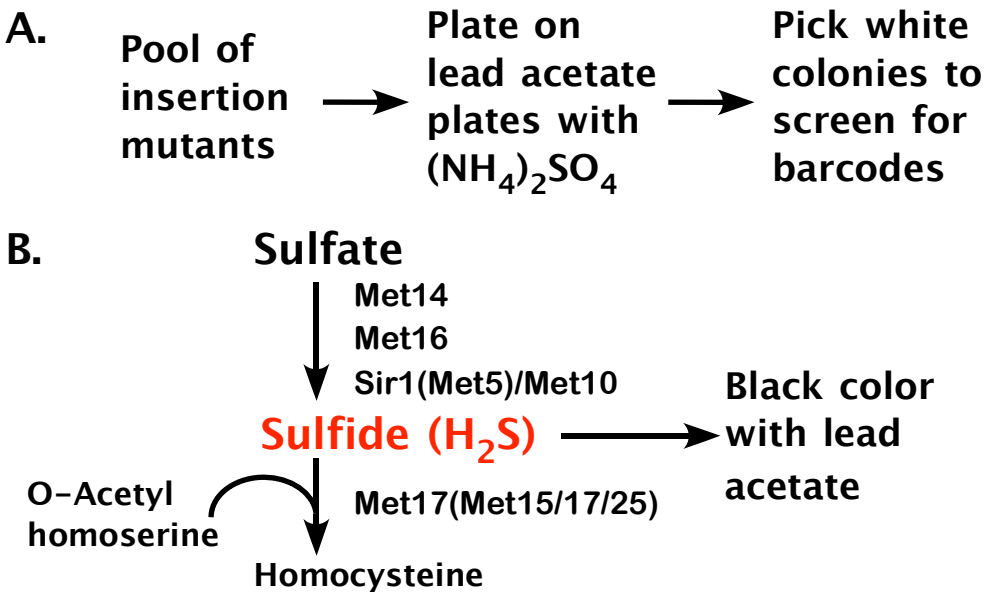
D.

```
          1450        1460        1470        1480
        *   *   *   *   *   *   *   *   *   *   *
TCT CTT CAC TCT ATT GTT CAG ATG CCT CGA GGT GTC CCT GTC
 S   L   H   S   I   V   Q   M   P   R   G   V   P   V
```

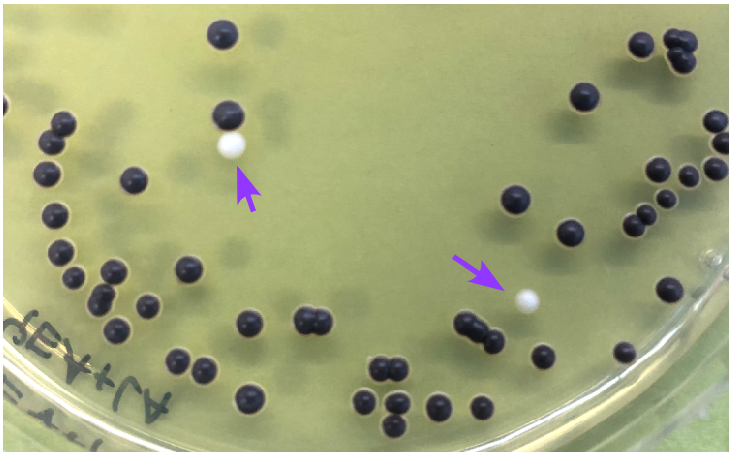
L972 (*ade6+*) CAG ATG CCT CGA GGT GTC CCT
KRP387 (*ade6+*) CAG ATG CCT CGA GGT GTC CCT

GP201 (*ade6-M210*) CAG ATG CTT CGA GGT GTC CCT
FY1645 (*ade6-M210*) CAG ATG CTT CGA GGT GTC CCT
KRP2 (*ade6-M210*) CAG ATG CTT CGA GGT GTC CCT

ade6-M210; C1466T; P489L



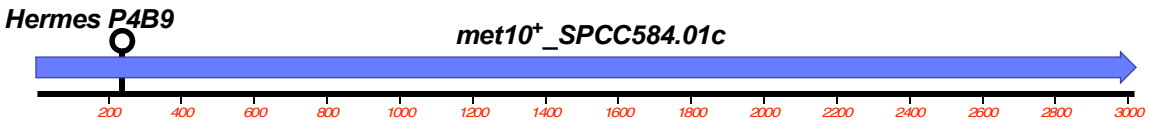
C.



3 white colonies found

All are Hermes P4B9 in *met10*, a subunit of the sulfite reductase.

D.



SUPPLEMENTAL MATERIAL

Contents:

Supplemental Tables

Supplemental Tables 1 and 2 describe the *Hermes* insertion mutants tested in Figure 6A and B, respectively. Supplemental Table 3 describes index tags used in for identifying pools for high-throughput sequencing. Supplemental Table 4 describes the oligonucleotide primers used so for PCR.

Supplemental Figures:

Supplemental Figures 1 through 4 describe additional aspects of methods and data in the Results.

Text to describe Supplemental Figures 5 to 8.

Supplemental Figures 5 to 8 describe additional ways to utilize the *Hermes* insertions as referred to in the Discussion.

Supplemental Table 1. *Hermes* insertion mutant strains used in Fig. 6

Gene name	Encoded Protein	*Hermes* Insertion	Growth on YEEG
atp14	F1-ATPase subunit H (predicted)	5' UTR	Defective
atp16	F1-ATPase delta subunit (predicted)	5' UTR	Defective
atp2	F1-ATPase beta subunit Atp2	5' UTR	Defective
atp3	F1-ATPase gamma subunit (predicted)	5' UTR	Normal
atp5	F0-ATPase delta subunit (predicted)	Intron	Defective
cox13	cytochrome c oxidase subunit VIa (predicted)	3' UTR	Normal
cox14	cytochrome c oxidase assembly protein Cox14 (predicted)	3' UTR	Normal
cox17	metallochaperone Cox17(respiratory chain complex IV assembly)	5' UTR	Normal
cox5	cytochrome c oxidase subunit V	3' UTR	Normal
cox6	cytochrome c oxidase subunit VI (predicted)	5' UTR	Normal
cox9	cytochrome c oxidase subunit VIIa (mitochondrial respiratory chain complex IV)	5' UTR	Normal
mzm1	mitochondrial respiratory chain complex III assembly protein Mzm1	Intron	Defective
sdh1	succinate dehydrogenase Sdh1(mitochondrial respiratory chain complex II)	Coding Exon	Defective
sdh2	succinate dehydrogenase (ubiquinone) iron- sulfur protein subunit	Coding Exon	Defective
		Coding Exon	Defective
tim11	F0-ATPase subunit E	5' UTR	Normal

Supplemental Table 2. *Hermes* insertion strains used in the CPT-sensitivity assay

Gene name	Insertion	Insertion point	Sensitive to CPT(μM)
tdp1	ORFΔ		5
	Coding Exon	III:530334-530335	5
	Coding Exon	III:529408-529409	5
alp14	ORFΔ		5
	5' UTR	III:1011566-1011567	10
rnc1	ORFΔ		5
	Coding Exon	III:65899-65900	5
	5' UTR	III:66958-66959	10
mug24	ORFΔ		5
	5' UTR	III:1918372-1918373	Not Sensitive
mhf2	ORFΔ		5
	Coding Exon	III:2101723-2101724	10
dad5	ORFΔ		5
	3' UTR	III:1670295-1670296	10
	3' UTR	III:1669975-1669976	15
	3' UTR	III:1670148-1670149	10
	3' UTR	III:1670121-1670122	Not Sensitive
git3	ORFΔ		5
	5' UTR	III:1545380-1545381	Not Sensitive
	5' UTR	III:1545742-1545743	Not Sensitive
	5' UTR	III:1546043-1546044	Not Sensitive
SPCC1393.11	ORFΔ		5
	5' UTR	III:821273-821274	10
sgf73	ORFΔ		5
	5' UTR	III:2123943-2123944	Not Sensitive
SPBC725.10	ORFΔ		5
	Coding Exon	II:1225168-1225169	5
kin1	ORFΔ		5
	5' UTR	II:2694295-2694296	Not Sensitive
trt1	ORFΔ		5
	Coding Exon	II:2065448-2065449	Not Sensitive
sat1	ORFΔ		5
	3' UTR	II:3858892-3858893	Not Sensitive
	3' UTR	II:3858892-3858893	Not Sensitive
	3' UTR	II:3858892-3858893	Not Sensitive
	Coding Exon	II:3856366-3856367	5
bdc1	ORFΔ		5
	5' UTR	II:2425546-2425547	10
rps402	ORFΔ		5

	5' UTR	II:1651866-1651867	10
csn1	ORFΔ		5
	3' UTR	II:4027332-4027333	10
sgf29	ORFΔ		5
	Coding Exon	II:2419576-2419577	5
	5' UTR	II:2420034-2420035	Not Sensitive
SPBC16A3.17c	ORFΔ		5
	Coding Exon	II:4265968-4265969	5
	Coding Exon	II:4266736-4266737	5
	5' UTR	II:4264819-4264820	Resistant
cbp1	ORFΔ		5
	3' UTR	II:3510683-3510684	Not Sensitive
rad26	ORFΔ		5
	Coding Exon	I:4453049-4453050	10
SPAC9.02c	ORFΔ		5
	3' UTR	I:1457123-1457124	Not Sensitive
bst1	ORFΔ		5
	Coding Exon	I:1686804-1686805	5
	Coding Exon	I:1688372-1688373	5
mrc1	ORFΔ		5
	Coding Exon	I:4209641-4209642	10
ryh1	ORFΔ		5
	3' UTR	I:1193477-1193478	Not Sensitive
	3' UTR	I:1192931-1192932	15
	3' UTR	I:1193455-1193456	Not Sensitive
pab1	ORFΔ		5
	5' UTR	I:509884-509885	5
rad57	ORFΔ		5
	5' UTR	I:2122831-2122832	10
mkh1	ORFΔ		5
	Coding Exon	I:615881-615882	5
rad1	ORFΔ		5
	Coding Exon	I:4979380-4979381	10
rho2	ORFΔ		5
	intron, splice region variant	I:4423192-4423193	5
mre11	ORFΔ		5
	3' UTR	I:439737-439738	10
	5' UTR	I:436049-436050	15
SPAC12G12.12	ORFΔ		5
	Coding Exon	I:325051-325052	10
rps3001	ORFΔ		5
	5' UTR		10
arp42	ORFΔ		5
	3' UTR	I:4352502-4352503	Not Sensitive

gpa2	ORFΔ		5
	5' UTR	I:2519102-2519103	Not Sensitive
rad9	ORFΔ		5
	Coding Exon	I:1714404-1714405	5
SPBC20F10.07	ORFΔ		5
	5' UTR	II:3295617-3295618	Not Sensitive
mug24	ORFΔ		5
	5' UTR	III:1918455-1918456	Not Sensitive

Supplemental Table 3. Barcodes (customized index tags) used in for identifying pools for high-throughput sequencing.

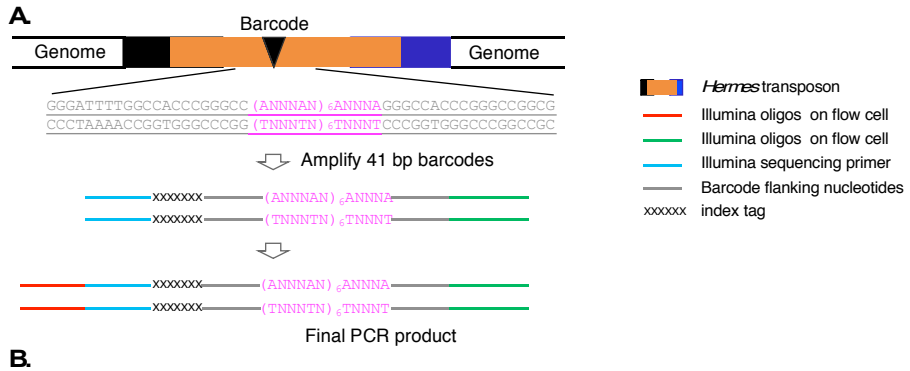
Index tag #	Index tag sequence	Pool ID
1	TTCAATG	Row1
2	CTTCCAG	Row2
3	AGCTTCA	Row3
4	TATGAAA	Row4
5	CCTTGTA	Row5
6	ATGAAAT	Row6
7	TAGAGAA	Row7
8	AGACGAG	Row8
9	GTCGTGG	Row9
10	CACCACG	Row10
11	CCACAAA	Row11
12	TTATAGC	Row12
13	GGCAAGC	Row13
14	TATACCG	Row14
15	TCAGGGG	Row15
16	TACTCTT	Row16
17	TCTGCCT	Column1
18	GCAGCCC	Column2
19	CGCGGAA	Column3
20	ACATCCG	Column4
21	AATTACT	Column5
22	GTAGGGC	Column6
23	GAAGAAC	Column7
24	AGCGAGT	Column8
25	GCCTAGT	Column9

26	CTGCTCC	Column10
27	TGGCTGG	Column11
28	CGGTCAC	Column12
29	AGTATGC	Layer 1
30	GCGCCTC	Layer 2
31	CCTCAGC	Layer 3
32	CAAACTC	Layer 4
33	ATTCGCT	Layer 5
34	AAGGCAG	Layer 6
35	CTCCGTC	Layer 7
36	GGTACTG	Layer 8
37	AAACACA	Layer 9
38	GACATAG	Layer 10
39	ATATTAC	Layer 11
40	GGGAATA	Layer 12

Supplemental Table 4. Oligonucleotides used in this study.

Name	Sequence	Purpose
a7HLE_S	TTA TTT TTA TAC CAA AAG CTA ATT GTC TTC AAG TGG AAT AGT TTA AAC CTT CAT TTC TAC ATC AAG TTC ATT TCA ACA GAG AAC TTC AAC AAG CCA CAG G	*Hermes* Left End ORF fusion
a7HLE_AS	TAC ACA CTC ATA CAG GTC TCT GAC TTT GCC GGT AGC AAT TTT AGT GAA TGG AGC ATC CAA CGA AGT TTT TAA TAA TGC CAT TGA TTC ATC GAC ACT CGG G	Row *Hermes* Left End ORF fusion 2
a7HRE_S	GTT ATT TTT ATA CCA AAA GCT AAT TGT CTT CAA GTG GAA TAG TTT AAA CCT TCA TTT CTA CAT CAA GTT CAT TTC AAC AGA GAA CAA CAA CAA GTG GCT T	*Hermes* Right End ORF fusion
a7HRE_AS	ACT CAT ACA GGT CTC TGA CTT TGC CGG TAG CAA TTT TAG TGA ATG GAG CAT CCA ACG AAG TTT TTA ATA ATG CCA TAG ATA AGC ACA AGT GTT TTG GGT G	*Hermes* Right End ORF fusion
LEa7noATG_S	CGA GTG TCG ATG AAT CAT TCG CAT TAT TAA AAA CTT CGT TGG ATG	*Hermes* Left End –ATG ORF fusion
LEa7noATG_AS	AAT GCG AAT GAT TCA TCG ACA CTC GGG TAT G	*Hermes* Left End –ATG ORF fusion2
REa7noATG_S	CAA AAC ACT TGT GCT TAT CTT TCG CAT TAT TAA AAA CTT CGT TGG ATG	*Hermes* Right End –ATG ORF fusion
REa7noATG_AS	AAT GCG AAA GAT AAG CAC AAG TGT TTT GGG TG	*Hermes* Right End –ATG ORF fusion2
A7HRLplus_S	ATC ACA CGG ATT TTC TTT AAA TAC TTT GTT ATT TTT ATA CCA AAA GCT AAT TGT CTT CAA	Extending homology to *ade7* genomic sequence
A7HRLplus_AS_2	GTT GCT ACG AAC AAA AGA TCA TCG GGA AAT TCT ACA CAC TCA TAC AGG TCT CTG ACT TTG	Extending homology to *ade7* genomic sequence
ade7_confSnew	CAT CGA AAG TTA GAG TTA ACT GG	Confirmation of *Hermes* integration at *ade7*
ade7_confAS	TGC TTC GTG ATA ACA GGA GG	Confirmation of *Hermes* integration at *ade7*
CPC3	GGC TGG CCT GTT GAA CAA GTC TGG A	Confirmation of *Hermes* integration at *ade7*
CPN10	GAT GTG AGA ACT GTA TCC TAG CAA G	Confirmation of *Hermes* integration at *ade7*
a7noATG-conf-S	ACT TAG GTT GCC ATC ATC CTC	Primers for a7HERMnoATG-LE colony PCR (with CPN10)
KanC	TGA TTT TGA TGA CGA GCG TAA T	Primers for a7HERMnoATG-LE colony PCR
A7noATG-conf-AS	GTA CGT TAA GGC CAT GAA CAG	Primers for a7HERMnoATG-LE colony PCR
a7noATG-conf-S	ACT TAG GTT GCC ATC ATC CTC	Primers for a7HERMnoATG-RE colony PCR (with KanC)
A7noATG-conf-AS	GTA CGT TAA GGC CAT GAA CAG	Primers for a7HERMnoATG-RE colony PCR (with CPN10)
a7HERM-noATG-LE-S	CAC ACT CAA GTG CAT AAG CCA CT	Primers for amplifying noATG-ade7 CDS
a7HERM-noATG-RE-S	AAT CGC ACA CGT CCA CTT GTG A	Primers for amplifying noATG-ade7 CDS

a7HERM-noATG-AS	AAT CGC CAA AAC ACC GTT AAC AGG	Primers for amplifying noATG-ade7 CDS
a7noATG-ORFseq-S	CGA AGC TAT TGT TCG TGG CT	Primers for noATG-ade7 CDS sequencing
a7noATG-ORFseq-AS	GTC TCA GCA ACT TGT TTG GC	Primers for noATG-ade7 CDS sequencing
pHL2577-3829S	AAG ACT AGG AAA AGA GCA TAA G	*Hermes* barcode amplification and sequencing
pHL2577-4171AS	GAC TGT CAA GGA GGG TAT TC	*Hermes* barcode amplification



B.

5′ **CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGTAATACGACTCACTATAGGGCTCCGCTTAAGGGACTAGENO ME**CAGAGAACTTCAACAAGCCACAGGCAAACGTAAGCCACATA**XXXXXX**AGATCGGAAGAGCGTCGTGTAGGGAAAGA GTGTAGATCTCGGTGGTCGCCGTATCATT 3′

Illumina adaptor-linker-GENOME-Hermes Right end-index tag-illumina sequencing primer-Illumina adaptor

C.

5′**AATGATACGGCGACCACCGAGATCT**ACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXX***GCGCATAAGTATCAAAA TAAGCCACTTGTTGTTGTTCTCTG***GENOME**TAGTCCCTTAAGCGGAGCCCTATAGTGAGTCGTATTAC**AGATCGGAAGAG CTCGTATGCCGTCTTCTGCTTG 3′

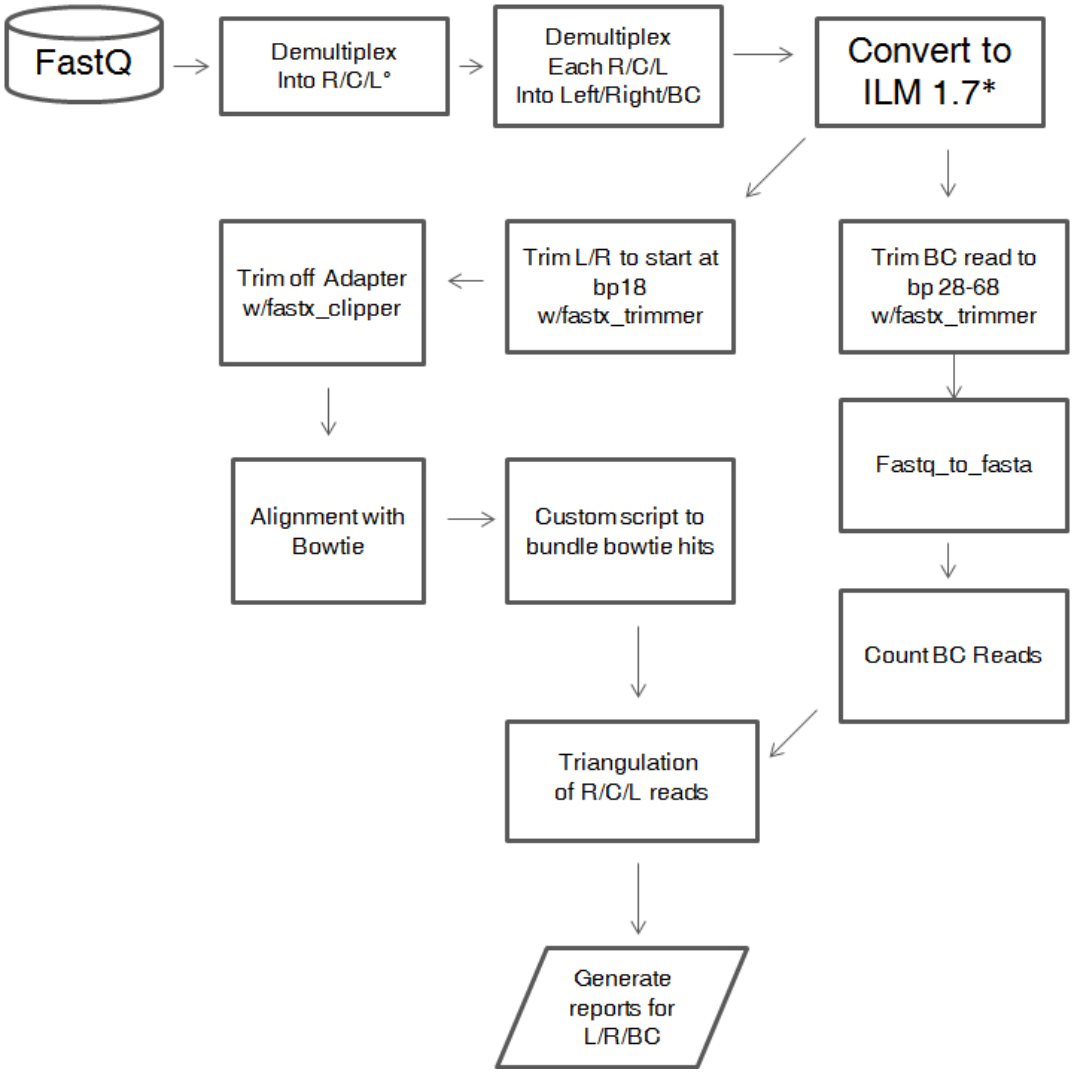
Illumina adaptor- illumina sequencing primer-index tag-Hermes Left end-GENOME-linker-Illumina adaptor

D.

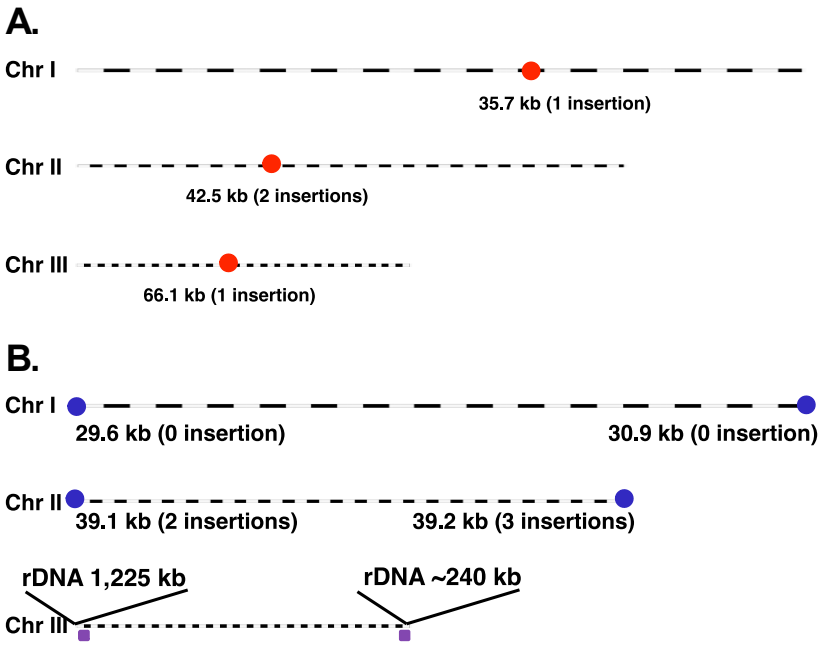
AATGATACGGCGACCACCGAGATCT**ACACTCTTTCCCTACACGACGCTCTTCCGATCT**XXXXXXTATCCC**GGGATTT**TGGCCA CCCGGGCC**ANNNNANNNNANNNNANNNNANNNNANNNNANNNNANNNNANNNNA**GGGCCACCCGGGCCGGCGCGCCC**GGGC CGGCGCGCCAGATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAGAATACCC TCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCATACATCCCCA TGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCGCGAAGCAAAA**ATTACGGCTCCTCGCTGCAG**AGA TCGGAAGAGC**TCGTATGCCGTCTTCTGCTTG

Illumina adaptor- illumina sequencing primer-index tag-barcode flanking sequence-BARCODES-barcode flanking sequence-Illumina adaptor

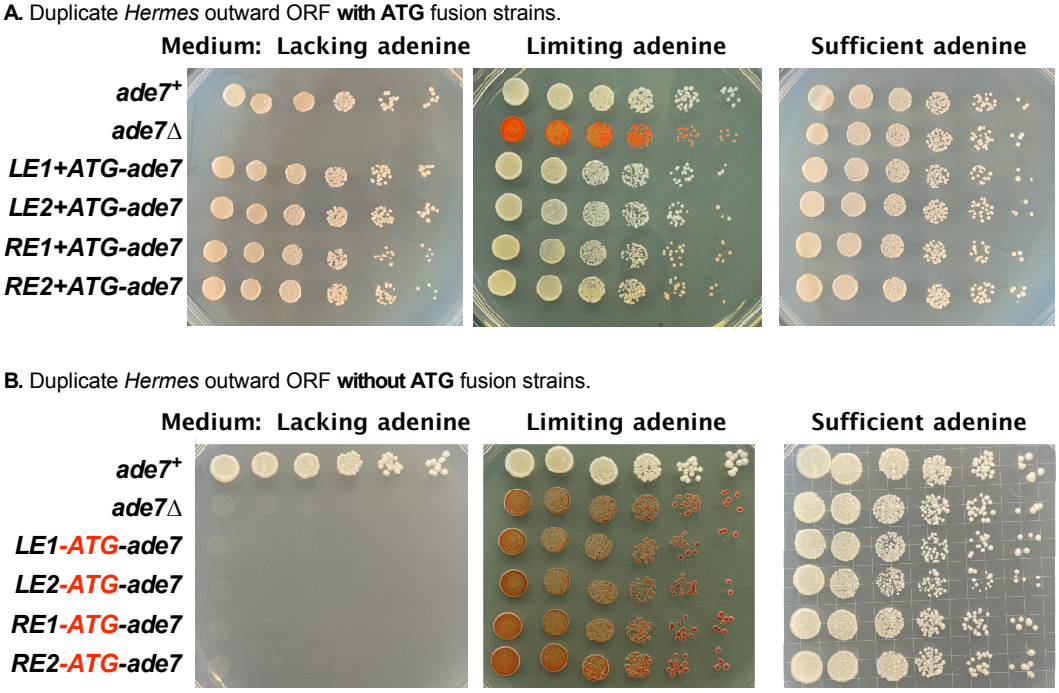
Supplemental Fig. 1. Amplification of barcodes. Genomic DNAs extracted from row, column and layer pool cultures were used as the template. Primers specific to the consensus sequence were used to amplify the barcodes and incorporate the index tags. Then Illumina adaptors and sequencing primers were added to the final PCR reaction. **A.** Amplification of the barcode sequences. **B-D.** Representative sequences of the amplification products of *Hermes* Right end (B), *Hermes* Left end (C) and barcode (D) products.



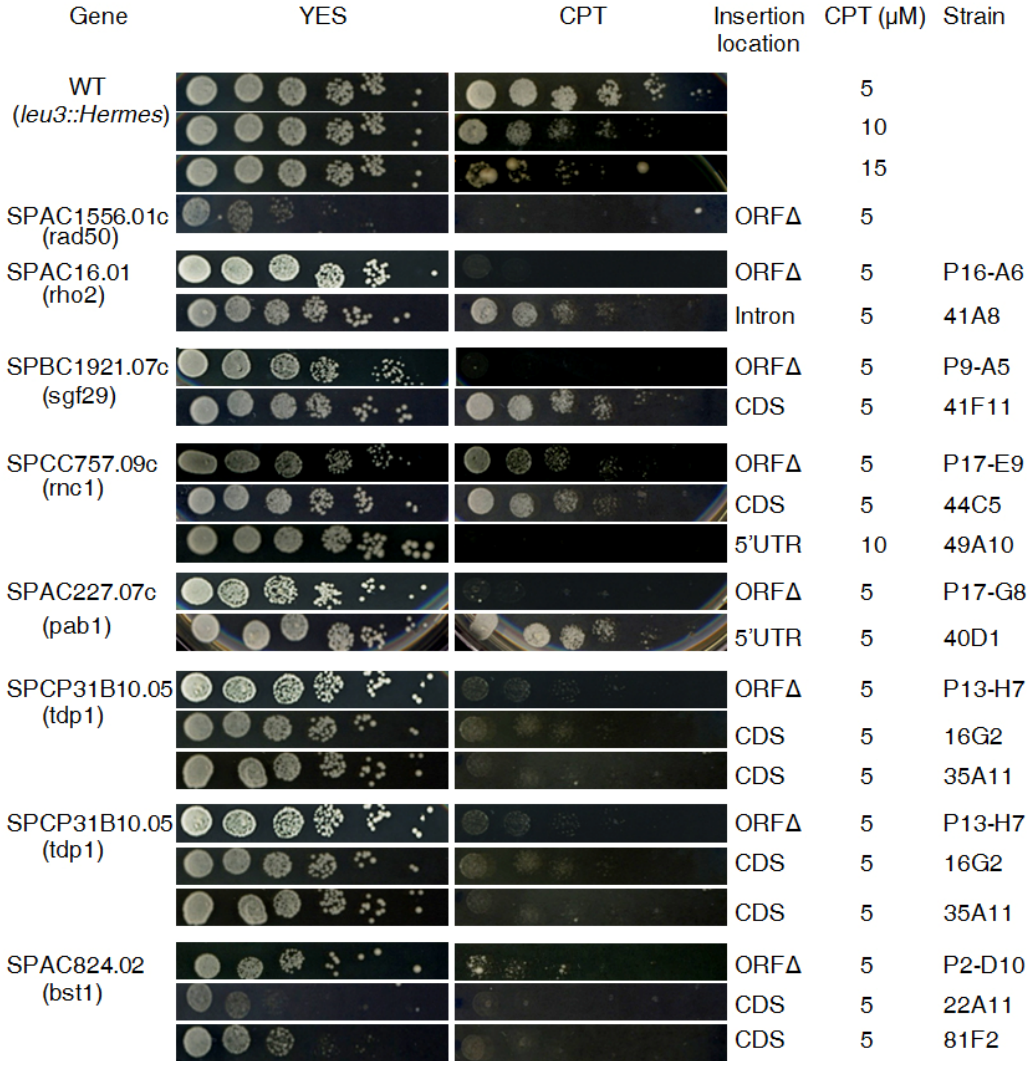
Supplemental Fig. 2. Transposon Integration sites and DNA barcode analysis pipeline. ILM1.7 is a FastQ format, which is required by the w/fastx clipper or trimmer tool. R (row); C(column); L(layer); BC(barcode).

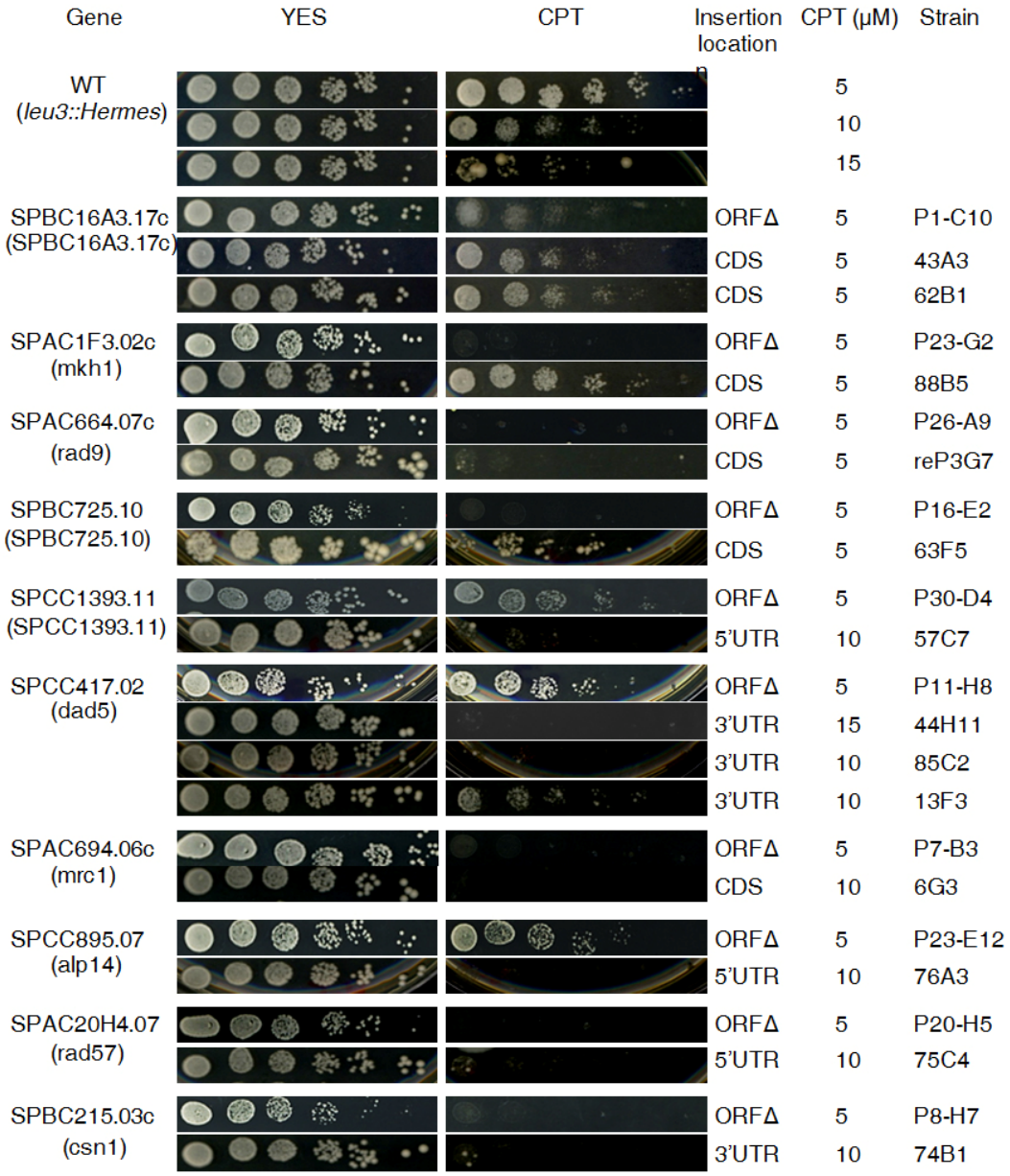


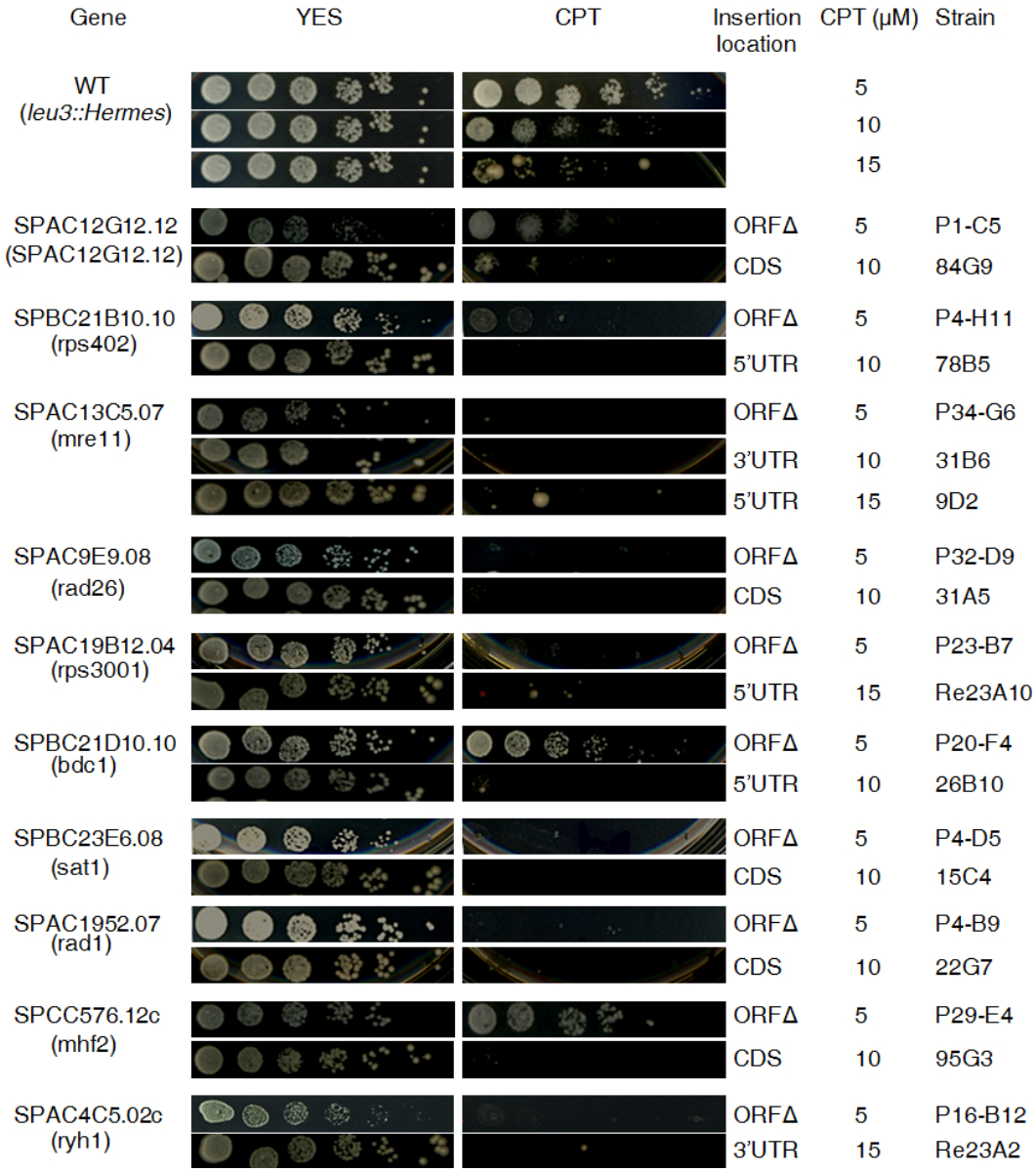
Supplemental Fig. 3. The distribution of *Hermes* mutants in heterochromatin. (A) *Hermes* distributions on *S. pombe* centromere. The *Hermes* collection contained 4 total insertions in the three centromeres. **(B) *Hermes* distributions on *S. pombe* telomeres.** The *Hermes* collection contained 5 insertions in telomeres of chromosome II. The rDNA repeats are represented as in Fig. 3B.



Supplemental Fig. 4. Analysis of independently constructed strains with *Hermes* Left and Right end ORF fusions to *ade7*. Two separate isolates of each of the four types of each of the fusions were analyzed for *ade7* expression by growth as in Fig. 5. All 4 –ATG-*ade7* fusions were amplified by PCR and sequenced to validate that the only difference between the +ATG-*ade7* and –ATG-*ade7* constructs was the mutation of the initiator ATG to TTC.


Gene
YES
CPT
Insertion location
CPT (μM)
Strain
WT (leu3::Hermes)
5
10
15
SPAC1556.01c (rad50)
ORFΔ 5
SPAC16.01 (rho2)
ORFΔ 5 P16-A6
Intron 5 41A8
SPBC1921.07c (sgf29)
ORFΔ 5 P9-A5
CDS 5 41F11
SPCC757.09c (rnc1)
ORFΔ 5 P17-E9
CDS 5 44C5
5'UTR 10 49A10
SPAC227.07c (pab1)
ORFΔ 5 P17-G8
5'UTR 5 40D1
SPCP31B10.05 (tdp1)
ORFΔ 5 P13-H7
CDS 5 16G2
CDS 5 35A11
SPCP31B10.05 (tdp1)
ORFΔ 5 P13-H7
CDS 5 16G2
CDS 5 35A11
SPAC824.02 (bst1)
ORFΔ 5 P2-D10
CDS 5 22A11
CDS 5 81F2

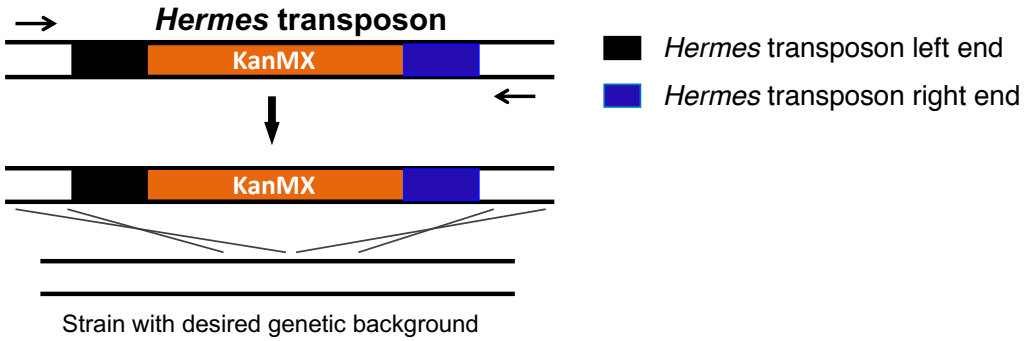

Gene
YES
CPT
Insertion location
CPT (μM)
Strain
WT (leu3::Hermes)
5
10
15
SPBC16A3.17c (SPBC16A3.17c)
ORFΔ 5 P1-C10
CDS 5 43A3
CDS 5 62B1
SPAC1F3.02c (mkh1)
ORFΔ 5 P23-G2
CDS 5 88B5
SPAC664.07c (rad9)
ORFΔ 5 P26-A9
CDS 5 reP3G7
SPBC725.10 (SPBC725.10)
ORFΔ 5 P16-E2
CDS 5 63F5
SPCC1393.11 (SPCC1393.11)
ORFΔ 5 P30-D4
5'UTR 10 57C7
SPCC417.02 (dad5)
ORFΔ 5 P11-H8
3'UTR 15 44H11
3'UTR 10 85C2
3'UTR 10 13F3
SPAC694.06c (mrc1)
ORFΔ 5 P7-B3
CDS 10 6G3
SPCC895.07 (alp14)
ORFΔ 5 P23-E12
5'UTR 10 76A3
SPAC20H4.07 (rad57)
ORFΔ 5 P20-H5
5'UTR 10 75C4
SPBC215.03c (csn1)
ORFΔ 5 P8-H7
3'UTR 10 74B1



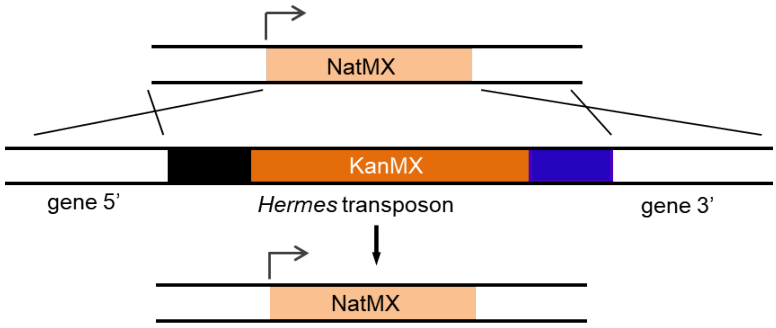
Supplemental Fig. 5. Defective growth of *Hermes* mutants and ORF deletion mutants on CPT plates. Spot tests of *Hermes* insertion mutants and ORF deletion mutants on 5, 10 and 15 μM CPT plates. CPT (μM) shows the lowest concentration where the mutant showed sensitivity. The ORF mutants from the Bioneer library are shown for comparison. Mutants that grew the same as the wild type strain at 15 μM CPT were considered not sensitive and are not shown. Note that the host strain for the *Hermes* insertions bears the *ade6-m216* mutation that causes the formation of dark pink colonies for some strains.

Engineering insertion elements to facilitate future use of the mutants

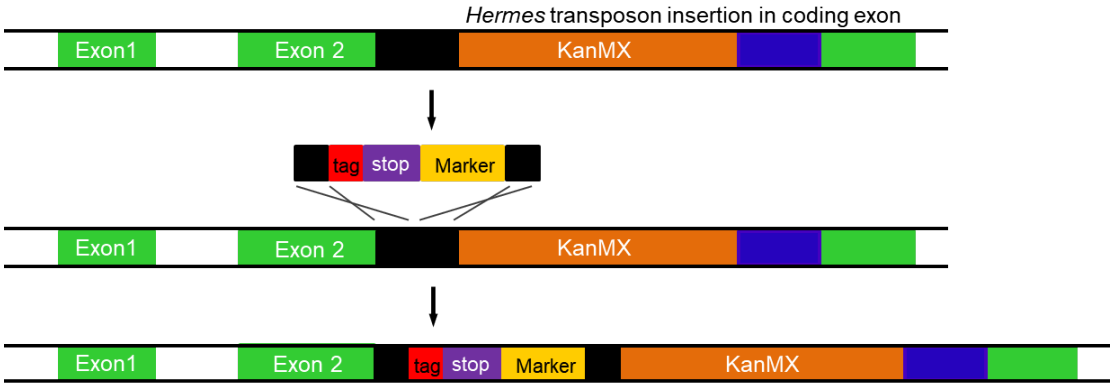
The design of the *Hermes* insertion element greatly facilitates the future utilization of individual mutant strains and the library beyond mutant screening. Following the example of the mTn3 transposon library in *S. cerevisiae*, various mutant alleles can be generated from a single insertion [1]. First of all, a given *Hermes* transposon insertion can be easily transplanted to other genetic backgrounds by transformation or mating, for verification of the mutant phenotype (Supplemental Fig.6), as was done with the CPT resistant mutant (Fig. 7B). Second, the *Hermes*-kanMX insertions can be converted to ORF deletions using another marker (e.g. natMX) with flanking homology to the genome. Integration will replace the *Hermes*-kanMX marker, so one can screen for Nat resistant and G418 sensitive strains to easily identify the deletions (Supplemental Fig. 7). This approach may be useful for transposon mutagenesis projects in strains containing complex reporter systems. Third, mutated genes of interest can be epitope-tagged for analyzing truncated protein functions generated from transposon insertion in coding exons. The entire *Hermes* transposon could be replaced by universal "epitope tag-selective marker" fragments commonly used in the yeast field [2], resulting in a C-terminal in-frame fusion to the epitope tag and replacement of the kanMX marker (Supplemental Fig. 8). Alternatively, fusions to reading frames in the *Hermes* flanking regions could be constructed to modify and *Hermes* insertions (Supplemental Fig. 8). Fourth, conditional alleles could be generated from mutants bearing the *Hermes* transposon within the promoter or 5' UTR regions. The *Hermes* transposon provides an integration site for a selective marker-conditional promoter fragment that could be used in many insertions (Supplemental Fig.9). Gene expression can then be "turned on" or "turned off" using inducible or temperature-sensitive promoters [3, 4] or over-expressed from strong promoters. The conversion could allow induction of gene silencing, followed by kinetic analysis of the effects of the loss of gene function within a few generations.



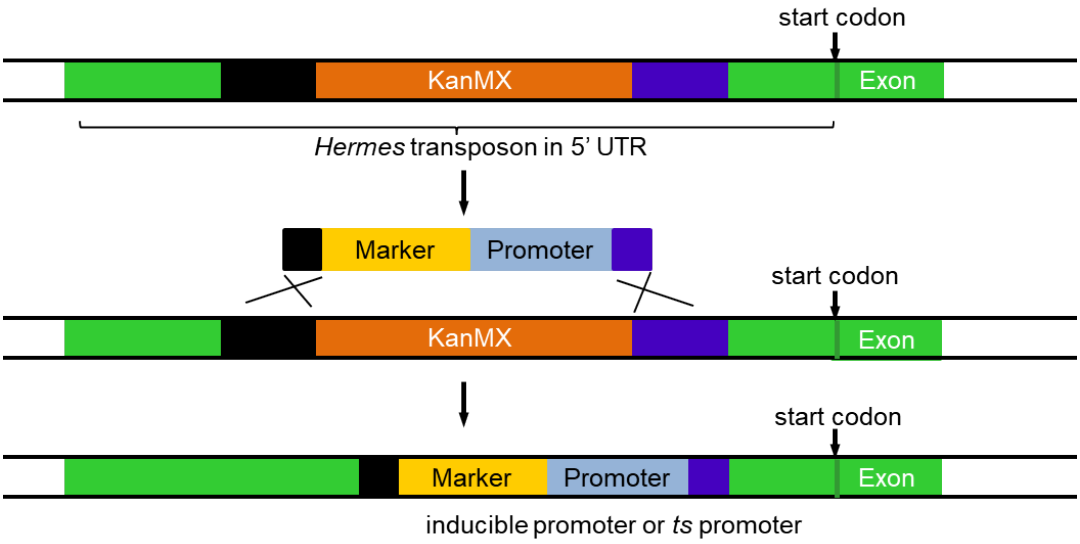
Supplemental Fig. 6. Moving a *Hermes* transposon insertion mutants to another genetic background. The transplantation element is amplified by a pair of primers surrounding the *Hermes* transposon insertion sites, followed by transformation of the element to a desired genetic background of strains.



Supplemental Fig. 7. Converting a *Hermes* transposon insertion to a complete gene deletion allele. The *Hermes* transposon insertion can be replaced by a NatMX marker flanked with homologous arms by homologous recombination. The new strain can be quickly selected by NatMX resistant and G418 sensitive phenotypes.



Supplemental Fig. 8. Epitope tagging of disrupted genes. When the *Hermes* transposon inserts coding-exons, universal tagging elements with homology to the *Hermes* transposon homologous arms, a selective marker and an epitope tag could be transformed into cells to make a C-terminal tagged strain. Vectors with the epitope tag in each reading frame would allow fusion to any disrupted protein-coding gene.



Supplemental Fig. 9. Conditional alleles generated from *Hermes* transposon insertions. When the *Hermes* transposon inserts in the 5' UTR or in front of a 5' UTR, a universal targeting element containing *Hermes* transposon homologous arms, a selective marker and a conditional promoter can be transformed into cells to make a conditional allele. The promoter can be inducible promoters such as *nmt1* or *Ts* promoters. Gene expression can then be controlled by culturing in conditional media or a restrictive/permissive temperature.

References:

1. Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung KH, Sheehan A, Symoniatis D, Umansky L, et al: **Large-scale analysis of the yeast genome by transposon tagging and gene disruption.** *Nature* 1999, **402:**413-418.
2. Longtine MS, McKenzie Ar, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR: **Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae.** *Yeast* 1998, **14:**953-961.
3. Kumar R, Singh J: **A truncated derivative of nmt 1 promoter exhibits temperature-dependent induction of gene expression in Schizosaccharomyces pombe.** *Yeast* 2006, **23:**55-65.
4. Garg A: **A lncRNA-regulated gene expression system with rapid induction kinetics in the fission yeast Schizosaccharomyces pombe.** *RNA* 2020, **26:**1743-1752.