CONSTRUCTION AND ANALYSIS OF A GENOME-WIDE INSERTION LIBRARY IN SCHIZOSACCHAROMYCES POMBE REVEALS NOVEL ASPECTS OF DNA REPAIR

By

YANHUI LI

Submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy

Dissertation Adviser: Kurt W. Runge, Ph.D.

Department of Genetics and Genome Sciences

CASE WESTERN RESERVE UNIVERSITY

January 2015
CASE WESTERN RESERVE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Yanhui Li
Candidate for the Ph.D. degree *.

Committee Chair
Helen Salz, Ph.D.

Committee Member
Kurt Runge, Ph.D.
Peter Harte, Ph.D.
Jo Ann Wise, Ph.D.

Date of Defense
10/14/2014

*We also certify that written approval has been obtained for any proprietary material contained therein.
### Table of Contents

- List of Figures 5
- List of Tables 8
- List of Abbreviations 9
- Acknowledgements 10
- Abstract 11
- Chapter 1. Background, significance and specific aims 13
  - 1.1 Genome-wide mutant collections are powerful tools to study genome functions 14
    - 1.1.1 Forward genetics and reverse genetics 14
    - 1.1.2 Genome-wide mutant collections in the budding yeast 15
    - 1.1.3 Genome-wide insertion mutant collections in multicellular organisms 21
    - 1.1.4 New tools to generate mutant collections in multicellular organisms for reverse genetic studies 28
    - 1.1.5 Genome-wide mutant collections in the fission yeast 36
  - 1.2 *Hermes* transposon excision and DNA DSB repair 39
    - 1.2.1 DNA DSB breaks cause genome instability and cancer progression 39
    - 1.2.2 Different pathways for DNA repair 43
    - 1.2.3 DNA damage repair by NHEJ 47
  - 1.3 Significance and specific aims of this study 49
- Chapter 2. The *S. pombe* Mre11-Rad50-Nbs1-Ctp1 Complex Promotes Non-Homologous End-Joining with Minimal Sequence Loss 53
  - Abstract 54
  - Introduction 55
  - Materials and Methods 59
Hermes transposon insertion
Mapping Hermes insertion sites
Hermes transposon excision induction and DSB repair assays
Reconstruction assay

Results

A transient transfection assay in S. pombe to maximize single transposon insertions
Transposon excision footprints indicate repair by NHEJ
NHEJ repair of excision events requires Ku, DNA ligase 4 and MRN
The Mre11 dimerization domain is required for efficient NHEJ
Mre11 and Ctp1 nuclease functions are dispensable for efficient NHEJ.

Chapter 3. A Schizosaccharomyces pombe transposon insertion library for high-throughput genome-wide studies
Abstract
Introduction

Results

Construction of a barcode-tagged Hermes transposon insertion mutagenesis library
Development of a novel three-dimensional pooling and high-throughput multiplexed sequencing strategy to map transposon integrations and DNA barcodes
Sequence identification and assignment to individual strains
Phenotypic characterization of the S. pombe insertion mutants

Material and Methods

Construction of barcoded-Hermes transposon plasmids
Generation of a library of Hermes insertion mutants
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooling cells, genomic DNA preparation, fragmentation</td>
<td>134</td>
</tr>
<tr>
<td>Ligation-mediated PCR</td>
<td>135</td>
</tr>
<tr>
<td>Amplification of barcodes and Illumina library preparation</td>
<td>137</td>
</tr>
<tr>
<td>Mapping of Integration sites</td>
<td>137</td>
</tr>
<tr>
<td>Verification of high-throughput sequencing results</td>
<td>139</td>
</tr>
<tr>
<td>Examination of respiration mutants and CPT-resistant mutants</td>
<td>139</td>
</tr>
<tr>
<td>Discussion</td>
<td>139</td>
</tr>
<tr>
<td>The Insertion library mutants display a wide range of phenotypes</td>
<td>139</td>
</tr>
<tr>
<td>The Hermes transposon has minor target site preferences</td>
<td>142</td>
</tr>
<tr>
<td>3D cell pooling combined with a deep-sequencing strategy greatly</td>
<td>144</td>
</tr>
<tr>
<td>speeds sequencing of a random-insertion mutant library</td>
<td></td>
</tr>
<tr>
<td>Engineering insertion elements to facilitate future use of the mutants</td>
<td>147</td>
</tr>
<tr>
<td>Chapter 4. General discussion and future directions</td>
<td>152</td>
</tr>
<tr>
<td>4.1 What nucleases open the hairpin in Mre11-mediated NHEJ?</td>
<td>153</td>
</tr>
<tr>
<td>4.2 A proposed screen for hairpin-opening nucleases</td>
<td>153</td>
</tr>
<tr>
<td>4.2 Advantages of generating a sequenced library by insertion mutagenesis for a reverse genetic screen</td>
<td>157</td>
</tr>
<tr>
<td>4.3 Future application of barcode-tagged Hermes transposon insertion mutant library in genome-wide screens</td>
<td>160</td>
</tr>
<tr>
<td>Application of the Hermes transposon insertion library in genome-wide aging studies</td>
<td>160</td>
</tr>
<tr>
<td>Application of the Hermes transposon insertion library in a genome-wide screen</td>
<td>164</td>
</tr>
<tr>
<td>4.4 Future application of Hermes transposon insertion mutagenesis in disease-gene discovery</td>
<td>165</td>
</tr>
<tr>
<td>Bibliography</td>
<td>168</td>
</tr>
</tbody>
</table>
## List of Figures

### Chapter 1

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Yeast gene deletion construction.</td>
<td>17</td>
</tr>
<tr>
<td>1.2</td>
<td>Construction of the mTn insertion mutant collection.</td>
<td>19</td>
</tr>
<tr>
<td>1.3</td>
<td>Engineering mTn to generate multiple mutant alleles from a single insertion.</td>
<td>20</td>
</tr>
<tr>
<td>1.4</td>
<td>Construction of a zebrafish random insertion library.</td>
<td>24</td>
</tr>
<tr>
<td>1.5</td>
<td>Illumina sequencing platform.</td>
<td>26</td>
</tr>
<tr>
<td>1.6</td>
<td>Identification of MLV integrations in the zebrafish genome by a high-throughput sequencing approach.</td>
<td>27</td>
</tr>
<tr>
<td>1.7</td>
<td><em>Drosophila</em> transgenic RNAi library.</td>
<td>29</td>
</tr>
<tr>
<td>1.8</td>
<td>Gene editing by Transcription activator-like effector nucleases (TALENs).</td>
<td>31</td>
</tr>
<tr>
<td>1.9</td>
<td>Zinc-finger nucleases (ZFNs).</td>
<td>33</td>
</tr>
<tr>
<td>1.10</td>
<td>CRISPR/Cas system.</td>
<td>34</td>
</tr>
<tr>
<td>1.11</td>
<td>Targeting Induced Local Lesions in Genomes (TILLING).</td>
<td>35</td>
</tr>
<tr>
<td>1.12</td>
<td>Random insertions produce various types of mutants.</td>
<td>37</td>
</tr>
<tr>
<td>1.13</td>
<td><em>Hermes</em> transposon system.</td>
<td>39</td>
</tr>
<tr>
<td>1.14</td>
<td>Origins of DSB formation.</td>
<td>41</td>
</tr>
<tr>
<td>1.15</td>
<td>DNA damage response.</td>
<td>43</td>
</tr>
<tr>
<td>1.16</td>
<td>DNA repair pathway choices.</td>
<td>46</td>
</tr>
<tr>
<td>1.17</td>
<td>DNA end recognition and core NHEJ factors.</td>
<td>49</td>
</tr>
</tbody>
</table>

### Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>A transient transfection assay to generate mutants with single transposon insertions.</td>
<td>67</td>
</tr>
<tr>
<td>2.2</td>
<td>The majority of <em>Hermes</em> transposon insertions from the transient transfection protocol are single events.</td>
<td>69</td>
</tr>
<tr>
<td>2.3</td>
<td>Removal of a <em>Hermes</em> transposon requires end-joining repair.</td>
<td>71</td>
</tr>
<tr>
<td>2.4</td>
<td>Quantitation and sequence analysis of <em>Hermes</em> transposon excision events.</td>
<td>73</td>
</tr>
</tbody>
</table>
Figure 2.5  Efficient Hermes transposon excision requires transposase expression.  

Figure 2.6  Hermes excision footprint sequences from KRP 3-3 cells indicate repair by NHEJ.  

Figure 2.7  Theoretical mechanism of transposon footprint formation suggests a role for the MRN complex in S. pombe NHEJ.  

Figure 2.8  Efficient NHEJ requires Ku70, DNA ligase 4 and MRN.  

Figure 2.9  Hermes excision footprints from different NHEJ mutants.  

Figure 2.10  Graphical summary of Hermes excision frequencies of different NHEJ mutants shown in Figure 2.8.  

Figure 2.11  Mutations in the Mre11 dimerization domain, but not the nuclease domain, greatly reduce NHEJ.  

Figure 2.12  Graphical summary of Hermes excision frequencies of different mre11 mutants from Figure 2.11.  

Figure 2.13  Sequences of mre11 nuclease mutant excision events.  

Figure 2.14  Model for the role of MRN-Ctp1 in NHEJ repair of Hermes excision sites.  

Chapter 3  

Figure 3.1 Overview of Hermes library construction.  

Figure 3.2 Overview of 3D pooling strategy and multiplexed high-throughput sequencing to map transposon integrations and DNA barcodes.  

Figure 3.3 Amplification of Hermes Insertion sites.  

Figure 3.4 Amplification of barcodes.  

Figure 3.5 Hermes transposon collection.  

Figure 3.6 The distribution of Hermes insertions in essential genes and non-essential genes.  

Figure 3.7 The distribution of Hermes mutants on heterochromatin.  

Figure 3.8 Defective growth of Hermes respiratory chain mutants on non-fermentable carbon sources.  

Figure 3.9 Defective growth of Hermes mutants and ORF deletion mutants on CPT plates.
Figure 3.10  The *Hermes* insertion mutants refine our understanding of CPT resistance genes. 131
Figure 3.11 A 5' UTR mutant is resistant to CPT. 132
Figure 3.12 Transposon Integration sites and DNA barcode analysis pipeline. 138
Figure 3.13 Increased read depth of the deep-sequencing platform allows a 3D pooling strategy to scale up and process thousands of samples in parallel. 146
Figure 3.14 Hermes transposon insertion mutants can be easily transplanted to any desired genetic background. 149
Figure 3.15 A *Hermes* transposon insertion can be easily converted to a complete gene deletion allele. 149
Figure 3.16 Epitope tagging of disrupted genes. 150
Figure 3.17 Conditional alleles generated from Hermes transposon insertions. 150

**Chapter 4**
Figure 4.1 Screen for hairpin-resolving nucleases. 154
Figure 4.2 Genome-wide screen for hairpin-resolving nucleases. 157
Figure 4.3 *S. pombe* CLS assay. 162
Figure 4.4 Identification of longevity genes by screening the Hermes transposon library using the CLS assay and high-throughput barcode sequencing. 164
Figure 4.5 Engineered sleeping beauty transposon is active in mammalian cells. 166
### List of Tables

Table 2.1 *S. pombe* strains used in this work. 60  
Table 2.2 Yield of transposon insertions 68  
Table 3.1 Comparison of the *Hermes* library with the Bioneer library. 120  
Table 3.2 *Hermes* insertion mutant strains used in Figure 3.8. 122  
Table 3.3 *Hermes* insertion strains used in the CPT-sensitivity assay. 125  
Table 4.1 *S. pombe* endonuclease candidates for hairpin-resolving screen assay. 155  
Table 4.2 Comparison of tools for genome-wide mutagenesis 158
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABRC</td>
<td><em>Arabidopsis</em> Biological Resource Center</td>
</tr>
<tr>
<td>CLS</td>
<td>Chronological life span</td>
</tr>
<tr>
<td>CPT</td>
<td>Camptothecin</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>DSBs</td>
<td>DNA double-strand breaks</td>
</tr>
<tr>
<td>GDP</td>
<td><em>Drosophila</em> Gene Disruption Project</td>
</tr>
<tr>
<td>HA</td>
<td>haemagglutinin</td>
</tr>
<tr>
<td>Hermes</td>
<td>Hermes transposon</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous recombination</td>
</tr>
<tr>
<td>LM-PCR</td>
<td>Ligation-mediated PCR</td>
</tr>
<tr>
<td>MLV</td>
<td>murine leukemia virus</td>
</tr>
<tr>
<td>MMEJ</td>
<td>Microhomology-mediated end joining</td>
</tr>
<tr>
<td>MRN</td>
<td>Mre11-Rad50-Nbs1</td>
</tr>
<tr>
<td>NASC</td>
<td>European <em>Arabidopsis</em> Stock Centre</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous End joining</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PB</td>
<td>PiggyBac</td>
</tr>
<tr>
<td>RLS</td>
<td>Replicative life span</td>
</tr>
<tr>
<td>SB</td>
<td>Sleeping Beauty</td>
</tr>
<tr>
<td>shRNA</td>
<td>short hairpin RNA</td>
</tr>
<tr>
<td>TALEN</td>
<td>transcription activator-like effector domain nucleases</td>
</tr>
<tr>
<td>TILLING</td>
<td>Targeting Induced Local Lesions in Genomes</td>
</tr>
<tr>
<td>TIRs</td>
<td>terminal inverted repeats</td>
</tr>
<tr>
<td>ts</td>
<td>temperature sensitive</td>
</tr>
<tr>
<td>TSSs</td>
<td>transcription start sites</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>ZFN</td>
<td>zinc-finger nucleases</td>
</tr>
<tr>
<td>β-gal</td>
<td>β-galactosidase</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to take this opportunity to express my sincere thanks to all the people who generously supported me during my PhD study.

First, I would like to express my deepest thanks to my supervisor, Dr. Kurt Runge. He gave me the great opportunity to work on such interesting projects. He taught me not only how to do research, but also how to present my work properly and confidently by practicing talks before my seminars. Without his guidance, patience, and encouragement, I could not have completed this dissertation.

Next, I want to thank my committee members Drs. Helen Salz, Jo Ann Wise and Peter Harte, for their insightful advice, feedback and challenges. I also thank Dr. Mark Adams and Neil Molyneaux for their precious time and help on the high-throughput data analysis. I am grateful to Dr. Tanaka Hisashi for his constructive advice on my project. I want to express my appreciation to all the current and past members of the Runge lab for their assistance, support, and friendship.

I would like to dedicate this dissertation to my family. I thank my dear parents for their unconditional love and support for so many years. I thank my husband for his invaluable love, company and enduring support.
Construction and Analysis of a Genome-Wide Insertion Library in *Schizosaccharomyces pombe* Reveals Novel Aspects of DNA Repair

Abstract

By

YANHUI LI

In the post-genomic era, utilizing yeast deletion libraries for high-throughput phenotypic screening accelerates gene function studies. However, haploid deletion collections are still insufficient to reveal all gene functions because only non-essential genes are knocked out. Mutants in essential genes and a large number of functionally uncharacterized non-coding genes are not included. Moreover, haploid deletion collections only contain null mutants. A resource that allows rapid phenotypic screening of many kinds of mutants (e.g. null, partial or altered function) in all genes is needed.

We constructed a defined, DNA barcode-tagged *Hermes* transposon insertion mutant library in *S. pombe*. The collection contained 4,095 haploid insertion mutant strains. Most strains (~90%) carry single transposon insertions. The insertions were distributed among 368 essential genes, 2,470 non-essential genes and 1,159 non-coding genes. The library contains a wide variety of mutations than the available gene deletion library. Insertions are distributed among open reading frames as well as 5’ and 3’ regulatory regions of genes. Phenotypic screening of selected mutants in the presence of the topoisomerase I inhibitor CPT revealed that some insertion mutants have the predicted
phenotypes while some have unexpected phenotypes. This library therefore represents an important resource for the international *S. pombe* community.

The construction of similar transposon insertion libraries in other organisms is labor-intensive. We developed a novel three-dimensional pooling strategy and a multiplexed high-throughput analysis pipeline to sequence the transposon insertion sites and DNA barcodes from thousands of samples at once. The approach greatly reduced the effort and was cost-effective. It can be applied to any insertion element, and will accelerate the construction of sequenced insertion mutant libraries in a wide variety of model systems.

As a first step in the development of the *Hermes* transposon as a genetic tool for large-scale mutagenesis, we analyzed its insertion and excision behavior in the context of the *S. pombe* genome, which led to novel insights into DNA damage repair by Non-homologous End joining (NHEJ) processing in *S. pombe*. Excision of the *Hermes* transposon from the *S. pombe* genome leaves a DNA double-strand break capped by hairpin ends that must be processed and ligated by cellular enzymes. We found that repair was through NHEJ and required the Mre11-Rad50-Nbs1 (MRN) and, to a lesser extent, Ctp1. The MRN complex has important roles in DNA damage signaling, homologous recombination repair and telomere replication. The role of MRN in NHEJ is unclear. We found that NHEJ in *S. pombe* did not require Mre11 nuclease function, but did require Mre11 dimerization function, suggesting that the primary role of Mre11 may be to tether the broken DNA ends in NHEJ.
Chapter 1

Background, significance and specific aims
In the post-genomic era, the availability of complete genome sequences of humans and many model organisms has led to the discovery of thousands of genes. However, the functions of many of them remain unknown. Tools have been developed to characterize gene functions in high-throughput ways. Arrayed libraries for high-throughput phenotypic screening in yeast\(^1\), worms\(^2\), and flies\(^3\) have been utilized to examine individual mutant phenotypes to provide insight into gene functions. Using these libraries for epistasis testing of mutant combinations has suggested gene or protein interactions. Analysis of mutant collections has contributed to the identification of the components of physiological pathways in a few organisms \(^4\). However, additional tools and systems are required to leverage the power of well-developed model organisms to gain a complete understanding of all gene and protein functions.

1.1 Genome-wide mutant collections are powerful tools to study genome functions

1.1.1 Forward genetics and reverse genetics

A powerful approach for elucidating gene function is to observe the phenotypes of mutants. “Genome-wide” mutation collections are essential tools for systematically defining gene functions and dissecting biological pathways \(^4\text{-}^6\). Two strategies are utilized to study gene functions systematically: forward genetics and reverse genetics. Classical forward genetics defines gene functions by generating a random collection of mutants, often by chemical mutagens or ionizing radiation, followed by screening for the phenotypes of interest and identifying the mutated genes by positional cloning or transformation and
complementation \(^7,^8\). For example, in yeast, mutagenesis can be induced by UV to generate defective growth on non-fermentable carbon sources such as glycerol and ethanol medium, which may identify gene mutations involved in aerobic respiration. These genes can be identified by complementation of the mutant phenotype with a wild-type genomic DNA library. Forward genetic screening has led to many fundamental biological discoveries. However, a major limitation of this approach is that it is time consuming and has a low efficiency in cloning mutant genes, especially in a large-scale screen, which looks for a desired phenotype to identify entire pathways or classes of genes. The nature of the mutated gene is unknown. Cloning the wild type copy is laborious and must be repeated for each mutant. Some mutant genes cannot be cloned by complementation screening because assaying the phenotype is not amenable to mass screening.

An alternate and more effective approach is to create defined mutations in all genes first and then characterize the effects of these mutations, a process called reverse genetics \(^4\). This approach requires sequenced mutant collections in which each cell carries a known mutation and does not require gene cloning following genetic screens. In recent years, projects to generate comprehensive gene disruption libraries have advanced rapidly for many organisms \(^9-13\).

1.1.2 Genome-wide mutant collections in the budding yeast

The budding yeast \textit{Saccharomyces cerevisiae} and the fission yeast \textit{Schizosaccharomyces pombe} are popular models for studying gene functions because of their ease of genetic manipulation \(^7\). The two yeasts share similar
genetic tools for creation and analysis of mutants. Both have high rates of homologous recombination, allowing precise gene knock out. In addition, culturing of haploid yeast allows for analysis of recessive mutations while diploid yeast allows for complementation tests. In budding yeast, gene deletion and random mutagenesis are the main approaches for generating mutant libraries. Homologous recombination is highly efficient in yeast. It is utilized to knock out genes and remains the fundamental strategy for deciphering protein function in vivo. Gene deletion collections are constructed by removing open reading frames (ORFs) from all protein-coding genes, and replacing them with tags to facilitate assaying their functions. This has resulted in gene collections of S. cerevisiae and S. pombe mutant genes. A transposon insertion mutant collection was constructed in S. cerevisiae. Such tools are under-developed in S. pombe compared to S. cerevisiae.

The genome-wide yeast mutant collections greatly accelerate systematic analyses of gene function. They are utilized in genetic screens, investigating molecular mechanisms, and identifying pathways that can be targeted for medical treatments. For example, a systematic functional screen using the S. cerevisiae gene-deletion pool of mutants was performed and mitochondrial respiration proteins were identified. The screen helped in the identification of human orthologs and candidate genes associated with putative human mitochondrial-related diseases. Moreover, researchers can use strains from the collection to analyze genes of interest instead of making gene deletions on their own.
The gene deletion collections in *S. cerevisiae*. A gene knockout collection was created in budding yeast *S. cerevisiae* in 1999 \(^{15}\). The method used to construct this collection was a PCR-based gene deletion strategy (Figure 1.1). Each gene disruption was replaced with a kanMX module and uniquely tagged with one or two 20-mer sequences, called “DNA barcodes.” About one-third of the ORFs in the genome were deleted when published in 1999, and the project was nearly complete 3 years later, with 90% of the yeast genes disrupted. Four different mutant collections were generated: haploids of both mating types, homozygous diploids for non-essential genes, and heterozygous diploids, which contained the essential and non essential ORFs \(^{15}\). Utilizing these collections and over 100 different culture conditions, more than 5,000 genes were assigned a phenotype \(^4\).

**Figure 1.1  Yeast gene deletion construction.** The gene deletion cassette contains the KanMX4 gene encoding G418 resistance, flanked by unique DNA barcodes (Up tag and Down tag) and homologous sequences flanking the ORF. The cassette replaces the ORF of target genes by homologous recombination.
The mTn transposon insertion collection in *S. cerevisiae*. The construction of systematic deletion sets is difficult. New approaches are being developed for the rapid mutagenesis of genomes and to assess gene function on a genomic scale. Snyder et al. generated a collection of mutants by transposon insertions \(^9\). This approach used a minitransposon (mTn), modified from the bacterial transposable element Tn3, transformed into a plasmid library, and each plasmid was subsequently integrated into the genome by homologous recombination \(^9,17\) (Figure 1.2). Individual plasmids were transformed into a diploid yeast strain in a 96-well format. By homologous recombination, each mTn integrates at its corresponding genomic locus to replace its genomic copy. Of the total 11,000 mutant strains, mTn insertion sites in the yeast genome for 6,358 strains were determined by sequencing genomic DNA plasmids carrying mTn. The insertion mutations are randomly distributed throughout the genome and a large collection of yeast mutant strains is generated, each carrying an insertion at a known genomic location. These insertions affected about one-third of the 6,200 predicted genes in *S. cerevisiae* \(^9\).
Figure 1.2 Construction of the mTn insertion mutant collection. A yeast genomic DNA plasmid library is mutagenized in *Escherichia coli* by mTn transposon. In 96-well plates, individual plasmids carrying a fragment of yeast genomic DNA and an mTn insertion are prepared. Yeast genomic DNA is released by restriction enzyme digestion and transformed into a diploid *S. cerevisiae* strain. Each genomic fragment carrying mTn should replace the corresponding genomic locus by homologous recombination.

In addition to generating a large number of mutants, mTn was modified for multiple purposes. The mTn vector allowed the generation of lacZ reporter gene fusions. It also allowed epitope-tagging of ORFs from a single mutagenic event. The mTn contained a lacZ reporter gene lacking a promoter and a start codon (Figure 1.3). Introduction of the mTn transposon into yeast results in production of β-galactosidase (β-gal) if the lacZ gene is fused in-frame with a
yeast protein-coding sequence. Using this reporter system, over 300 non-annotated *S. cerevisiae* ORFs were identified, indicating that transposon mutagenesis can be utilized to identify new protein-coding genes.

The mTn–3xHA/lacZ minitransposon contains lox P sites that allow the transposon element to be reduced to a 3xHA tag. When the epitope is fused in-frame with a yeast ORF, introduction of the Cre recombinase produces tagged protein (Figure 1.3). Over 1,300 transposon-tagged proteins were localized by indirect immunofluorescence against the 3xHA tags.

**Figure 1.3 Engineering mTn to generate multiple mutant alleles from a single insertion.** mTn contains a *lacZ* reporter gene that lacks the start codon and promoter sequence. Insertion of transposon results in production of a β-gal fusion protein in-frame fused with an ORF. Additionally, mTn also contains lox P sites. Introducing Cre recombinase will reduce the transposon to a haemagglutinin (3xHA) epitope tag. If the HA tag is fused in-frame to the gene ORF, an epitope-tagged protein will be generated. IR, transposon inverted repeat.
1.1.3 Genome-wide insertion mutant collections in multicellular organisms

Creating complete collections of gene deletions in multicellular organisms has been challenging because the rate of homologous recombination in multicellular organisms is very low. Moreover, the number of genes is much larger than in yeast. Effective methods to efficiently manipulate their genomes are lacking. Instead of gene deletion, investigators have relied on random mutagenesis although identifying the mutation sites within the genome of a mutant cell line was challenging.

The random insertion mutagenesis approach has been applied to a wide variety of organisms. Various transposon libraries were constructed and sequenced in the past decade. Genes from pathogenic bacteria that increase virulence or cause antibiotic resistance were identified from screening these collections and provided potential targets for therapy. The transposon random insertion approach has been applied to multicellular organisms such as flies, worms and plants. For example, large insertion mutagenesis collections have been generated in Drosophila melanogaster and Arabidopsis thaliana. More recently, a large-scale zebrafish mutant resource has been created. To date, these collections do not mark every gene.

The Drosophila Gene Disruption Project The Drosophila Gene Disruption Project (GDP) aims to create a collection of mutants containing single transposon insertions to help functionally analyze genes. The project was launched in 1991 when the Drosophila genome was not sequenced and little was known about genome regulatory elements. Over 1,000 mutant strains were
generated by single \textit{P-element} transposon insertions \textsuperscript{10}. Homozygous mutants that had a phenotype were saved. As the project progressed, the genome was sequenced and ORFs were annotated. By 2004, 40\% of known \textit{Drosophila} ORFs had one or more associated insertion alleles generated by single \textit{P element} or \textit{piggyBac} transposon insertions\textsuperscript{27}. Transposon insertion sites were determined by sequencing inverse PCR products flanking \textit{P element} or \textit{piggyBac} insertions.

It is difficult to disrupt every ORF with \textit{P element} and \textit{piggyBac} alone because transposon insertions in genome regions have target site preferences. Different transposable elements display distinctive patterns of ‘hot spots,’ where insertions accumulated at a higher than expected frequency, and ‘cold spots’ where insertions were not found \textsuperscript{11}. \textit{P elements} prefer to insert near promoters in \textit{Drosophila} and cause the mis-expression of proteins. \textit{PiggyBac} elements also have target hotspots. Different transposons have been used to increase novel gene hits. Recently, the \textit{Minos} transposon, a mariner family member, was applied to mutagenize the \textit{Drosophila} genome \textsuperscript{11}. The \textit{Minos} transposon has been modified to induce a recombination-based strategy for targeting genes of interest \textsuperscript{11}.

The additions of \textit{P-element}, \textit{piggyBac}, and \textit{Minos} insertions increase the number of tagged genes to 9,440, about two-thirds of all annotated \textit{Drosophila} protein-coding genes \textsuperscript{11}. In addition to gene ORFs, other \textit{Drosophila} genome structures and features were documented, such as multiple transcript isoforms, novel RNA genes, and genomic regulatory elements. The \textit{Drosophila} Gene Disruption Project evolved beyond disruption of genes. Over 14,000 strains
disrupting promoters, unannotated genes, putative regulatory sequences, and unknown aspects of genome function were generated. All mutant strains are stocked, available from the Bloomington Drosophila Stock Center.

The large Arabidopsis thaliana T-DNA collections The large collections of Arabidopsis thaliana T-DNA insertion mutants are among the most important resources for researchers to interrogate Arabidopsis thaliana gene functions. T-DNA originates from the Ti-plasmid, which encodes the causative agent in the crown gall tumor. Several groups around the world mapped T-DNA insertions through TAIL-PCR or adaptor-mediated PCR. To date, over 325,000 T-DNA insertion lines have been isolated and sequenced. These mutants cover ~88% of Arabidopsis thaliana genes. The mutant strains are stocked in the Arabidopsis Biological Resource Center (ABRC), European Arabidopsis Stock Centre (NASC), and GABI-Kat (Genomanalyse im Biologischen System Pflanzen-Kölner Arabidopsis T-DNA Lines).

A large-scale zebrafish gene mutagenesis resource Recently, a large-scale zebrafish mutant collection was constructed by inserting proviral forms of the murine leukemia virus (MLV). Mutagenesis was performed at the 1,000 to 2,000 cell blastula stage by injection of embryos with MLV. The mutant zebrafish lines are cryopreserved as sperm samples from F1 fish that carry random provirus integrations. Any desired mutant line could be recovered by in vitro fertilization with the frozen sperm sample containing the integration of interest to raise F2 fish.
Figure 1.4 Construction of a zebrafish random insertion library. The 1000-2000 cell stage blastula embryos were injected with MLV to establish mosaic founder fish. The founders were crossed with wild-type fish. Heterozygous F1 male fish carrying the provirus insertions in their genomes were used for sperm cryopreservation and preparation of genome DNA. Integrations were amplified, mapped and assigned to the corresponding sperm samples.  

Mapping insertion element integration sites in the genome has been challenging. Before the next-generation sequencing platform was developed, traditional ways of identifying integration sites of insertion elements were performed in 96-well format. Genomic DNA preparation, amplification of the insertion site containing genome fragments by various PCR strategies (such as inverse-PCR, Tail-PCR etc.), purification of amplification products and capillary sequencing were all operated in 96-well plates. The genome sequences flanking the insertion point from individual mutants were compared with the reference genome to identify the insertion sites. These
approaches resulted in high costs and were very labor-intensive. The recent development of next-generation sequencing platforms overcomes these limitations by sequencing thousands of samples simultaneously. The effort and cost of defining a library of random insertion mutants would, therefore, be reduced.

The Illumina Genome Analyzer sequencer uses sequencing by synthesis technology (Figure 1.5). Samples being sequenced have fixed adaptors at both ends. Samples are first denatured to single strands and grafted to the flow cell, where complementary adaptors are immobilized. After bridge amplification, sequences form clusters that contain clonal DNA fragments. In the process of DNA synthesis, four kinds of nucleotides (ddATP, ddGTP, ddCTP, ddTTP), each with a different cleavable fluorescent dye, would be added one base at a time. The signal could be captured to identify DNA sequences. The latest Illumina GAIIx series can attain 85 Gb/run and generate 30-40 million reads per lane that meet quality control standards. The latest outputs of Illumina HiSeq 2500 reach 10-1000 Gb per run and are able to produce 300 million-2 billion reads. The read size per fragment is up to 150 bp.
Figure 1.5 Illumina sequencing platform. 1) Single-stranded DNA samples are anchored to the Illumina flow cells by annealing to the first immobilized oligo adaptors (green bars). 2) New DNA is synthesized. Sample DNA strands are washed away. 3) Newly synthesized DNA contains an adaptor sequence complementary to the second immobilized oligo adaptors (red bars). 3) Newly synthesized DNA strands are annealed to immobilized adaptors (red bars) and serve as templates. 4) Bridge amplification generates copies of the original library immobilized on the flow cell surface. 5) DNA strands are denatured to single strands. 6) One strand is removed by periodate cleavage. 7) They are sequenced by synthesis. Clusters of templates undergo new DNA synthesis with 4 fluorescent-labeled reversible terminators. Each single base is identified one at a time by laser excitation and imaging. Blue arrow, Illumina sequencing primer.

A high-throughput mapping of MLV insertions in the zebrafish genome took advantage of the depth of sequence attainable by Illumina platforms. The mapping approach allowed the sequencing of 1,024 F1 fish provirus integrations at one time, reducing the cost of processing and sequencing (Figure 1.6). Genomic DNAs of individual F1 fish were arrayed in the 96-well format. Adjacent genomic DNA sequences of the virus insertion were amplified by linker-mediated PCR. To link an amplicon to the F1 fish from which it originated, distinctive “index tags” (6 bp double-strand DNAs), were incorporated adjacent to the ligation site of the Linker for each amplicon. A total of 1,024 nonredundant index tagged linkers were synthesized to allow multiplexed sequencing. The indexed amplicons were pooled together and sequenced on the HiSeq 2000 platform.
providing up to 200 million paired sequencing reads. A customized bioinformatics pipeline was developed to map retroviral insertions in the zebrafish genome. The sequencing of 6,144 F1 fish recovered 15,223 unique integration sites. A total of 3,054 unique genes were hit. Phenotypic examination of 41 F2 fish found 12 mutant lines showed a variety of phenotypes relevant to both developmental processes and human genetic diseases, indicating that the provirus integrations disrupted gene function.

Figure 1.6 Identification of MLV integrations in the zebrafish genome by a high-throughput sequencing approach. The genomic DNAs of individual F1 fish were arrayed in 96-well format, followed by restriction enzyme digestion. Products from the same fish were pooled and ligated with DNA linkers containing unique 6 bp index tags. MLV integration fragments were amplified by linker-mediated PCR. Products from 1,024 F1 fish were pooled together to sequence using the Illumina Hiseq 2000.
1.1.4 New tools to generate mutant collections in multicellular organisms for reverse genetic studies

In addition to random mutagenesis, new tools have been developed to study gene functions in multicellular organisms in recent years. These tools include RNA-mediated interference (RNAi), which silence target genes; gene editing tools that generate targeted genome double strand breaks and trigger DNA repair pathways to introduce mutations or new DNA fragments, such as transcription activator-like effector domain nucleases (TALEN), zinc-finger nucleases (ZFN) and CRISPR. Large-scale libraries have been generated using some of these tools, which are reviewed below.

**Gene silencing** RNAi inhibits gene activity by the introduction of sequence specific double-stranded RNA (dsRNA) of targeted genes. A genome-wide tissue-specific RNAi library was constructed in *Drosophila* in 2007. The method relied on the GAL4-UAS system to control dsRNA expression and sequence specific RNAi (Figure 1.7). In *Drosophila*, RNAi can be triggered by a long double-stranded ‘hairpin’ RNA expressed from a transgene containing an inverted repeat fragment (UAS-IR). By crossing UAS-IR to a GAL4 driver line, the inducible GAL4 system restricted RNAi expression to a selected tissue or stage of the fly’s lifespan when the GAL4 driver is expressed. The collection covered 12,088 genes, about 88% of predicted protein-coding genes.
**Figure 1.7 Drosophila transgenic RNAi library.** The GAL4/UAS system drives the expression of cloned inverted repeats (300-400 bp), forming hairpin RNAs. Hairpin RNAs are processed by Dicer into double strand siRNAs. The siRNA targets mRNA in a sequence-specific manner and induces the degradation of mRNA.

Genome-wide RNAi mutant collections have been constructed from nematodes. A library of bacterial strains was constructed, each expressing dsRNA designed to target a single *C. elegans* gene \(^{42}\). *C. elegans* that eat these bacteria incorporate the dsRNA and silence their endogenous genes \(^{42}\). The library consists of 16,757 bacterial strains covering about 86% of the 19,427 predicted *C. elegans* genes. Utilizing the RNAi library, many systematic screens have been performed, such as those identifying extended lifespans when partially silenced \(^{43}\).
RNAi libraries have been constructed for humans and mice. Moffat et al. developed lentiviral short hairpin RNA (shRNA) libraries targeting the human and murine genomes in 2006. The libraries target each of the 22,000 human and mouse genes published at that time. Availability of these libraries allows genome-wide screens for disease-related genes, such as genes that support Hepatitis C virus replication in human cell lines.

Genome-wide RNAi collections are important resources to study gene functions in multicellular organisms. However, the production of RNAi libraries requires intensive efforts, including synthesis and validation of large libraries of double stranded RNAs. Moreover, off-target effects exist and partial mRNA knockdown is not always sufficient to cause the phenotypic change of interest.

Gene editing by transcription activator-like effector domain nucleases (TALEN), zinc-finger nucleases (ZFN) and CRISPR are new tools for targeted genome modifications. They induce site-specific DSBs in a genome and trigger endogenous DNA repair. Cells use two systems to repair genome damage: Non-homologous end joining (NHEJ) and Homologous recombination (HR). The NHEJ process is error-prone, which will result in small deletions or insertions and can thus cause frameshift mutations and generate null alleles. HR repair, in contrast, can insert a transgene by cleavage-stimulated homologous recombination.

TALENs are fusion proteins that consist of DNA-binding motifs and FokI nuclease. The DNA-binding domain contains tandem, nearly identical 34 amino-acid repeats. Only two amino acids per repeat (at positions 12 and 13) are
hyper-variable (Figure 1.8). The hyper-variable amino acids associate with different DNA nucleotides. TALENs are currently used as an active tool for site-specific genome targeting in human and mouse cells 47,54,55, *C. elegans* 56, zebrafish 57, and rats 58.

Figure 1.8 Gene editing by Transcription activator-like effector nucleases (TALENs). TALEN is the fusion of the FokI nuclease domain and a DNA binding domain. TALEN is derived from TAL effector (TALE), which contains multiple approximate 34 amino acids repeats of DNA binding domain targeting a single base pair of genomes. FokI cleaves genomic DNA and induces DNA double-strand breaks. The cellular DNA repair system is then activated. If repaired by NHEJ, the broken chromosome will be rejoined imprecisely, resulting in small insertions or deletions to disrupt gene function. If repaired by HR, the repair pathway can use a transfected DNA to insert into the break site generated by the TALEN, resulting in gene replacement 59.
Engineered TALENs can be potentially applied to genome-wide mutagenesis to yield a large library of custom-designed DNA binding domains for TALENs to target each gene in a given eukaryotes. Several groups have developed protocols for simple, rapid assembly of TALEN\textsuperscript{60-62}. Most recently, a genome-scale collection of TALENs plasmids for targeting 18,740 protein-coding genes in human cells was developed\textsuperscript{63}. However, as the DNA binding domain consists of tandem repetitive structures, it is very challenging to synthesize TALENs and produce TALEN libraries across species.

ZFNs are composed of a zinc finger DNA-binding domain and FokI nuclease. The DNA-binding domains typically contain 3-6 individual zinc finger repeats that can target specific DNA sequences\textsuperscript{47}. Zinc finger domains can be engineered to target genes (Figure 1.9). Currently, no genome-wide ZFN collections are available because it is very hard to generate a library of Zinc finger domains.
Zinc-finger nucleases (ZFNs). ZFNs are created through the fusion of the DNA cleavage domain from the FokI endonuclease with zinc-finger proteins. Each zinc-finger is about 30 amino acids and contacts 3-4 base pairs of genomic DNA. Zinc-finger proteins are designed as left zinc-finger protein (ZFP) and right ZFP. ZFNs target specific DNA sequence and induce DSBs. DNA repair by NHEJ or HR will result in similar consequences as TALENs.

The CRISPR/Cas9 system is derived from Streptococcus pyogenes. The CRISPR system consists of the nuclease Cas9, the CRISPR RNA (crRNA) array that encodes the guide RNAs and a trans-activating crRNA (tracrRNA). This tracrRNA is part of the machinery that processes the crRNA array into discrete units. The crRNA unit uses a 20-nt guide sequence to bind genomic DNA and target cleavage by Cas9 (Figure 1.10). Several groups developed lentiviral-based large libraries covering over 18,000 and 7,000 human genes by cloning libraries of 20 nt oligos into lentiviral vectors. These libraries were used in functional screens for pluripotent stem cell survival and melanoma resistance to chemotherapy, and resistance to the nucleotide analog 6-thioguanine. Additional screens have identified host genes essential
for the intoxication of cells by anthrax and diphtheria toxins \(^69\) and resistance to "clostridium septicum" alpha-toxin \(^70\).

**Figure 1.10  CRISPR/Cas system.** The CRISPR/Cas9 system consists of the crRNA and tracrRNA-RNA pairs and Cas9 endonuclease protein. crRNA and tracrRNA forms a guide RNA for sequence-specific invasion of genomic DNA. Cas9 bears two endonuclease active sites to cleave DNA and introduce DSBs at the target sites of guided RNA\(^47\). The DSBs generated by TALEN, ZFN or CRISPR trigger DNA repair systems. NHEJ repair is error prone and will introduce small insertions or deletions at the repair sites. Mutations within genes will result in partial loss of function and null alleles.

**TILLING (Targeting Induced Local Lesions in Genomes)** TILLING is a method that combines random chemical mutagenesis with PCR screening of genes of interest \(^71,72\) (Figure 1.11). Chemical mutagenesis provides a variety of allele types, including missense and null mutations. TILLING has been employed to generate mutant alleles in targeted genes and is suitable for application in plants where effective mutagenesis approaches are lacking. TILLING has been
used to study *Arabidopsis thaliana* gene functions\(^ {73,74}\) and to a limited extent in other organisms such as *C. elegans*\(^ {75}\). The approach is useful for organisms lacking reverse genetic tools, where mutants with a range of phenotypes are highly desirable.

![Diagram](Image)

**Figure 1.11 Targeting Induced Local Lesions in Genomes (TILLING).** Following random chemical mutagenesis of plants, seeds of individual plants are collected and stored. Genomic DNAs are pooled from 8 plants. PCR is performed from pooled genomic DNAs with target gene-specific primers. The products are denatured by heating, followed by a slow cooling step to reanneal. This forms annealed heteroduplexes (wild type mutant strands). The products are digested with the nuclease CEL I. After denaturing, samples are electrophoresed. Short products will appear if the targeted mutation exists in the pool of genomic DNAs. Mutations are tracked down to the individual by repeating this process with DNA from the positive pool\(^ {72,76}\).
1.1.5 Genome-wide mutant collections in the fission yeast

*S. pombe* is an increasingly popular experimental system. It can be genetically manipulated almost as easily as *S. cerevisiae*. The cellular processes of *S. pombe* are similar to mammalian cells including cell-cycle control, RNA splicing, proteolysis, RNAi-mediated gene silencing, telomere function, and centromere structure\(^{14,77-79}\). After completion of the *S. pombe* genome sequence in 2002, more than 50 genes were identified that share large homology sequences with human disease genes, whose mutations cause cancer and metabolic, neurological and cardiac diseases\(^{80}\).

**S. pombe gene deletion collection** A genome-wide gene deletion set was created in 2010\(^{14,80}\). Genes were deleted and replaced using homologous recombination by a “deletion cassette” containing the KanMX marker flanked by a pair of unique molecular barcode tags (Figure 1.1). The heterozygous diploid deletion collection covered 98.4% of the 4,914 protein coding ORFs of *S. pombe* genome. The haploid deletion collection covered ~68% of protein-coding genes constructed from heterozygous diploid mutants\(^{14}\). However, a major defect limiting the use of the collection was that for ~600 genes, only part of the coding sequence (<80%) was removed\(^{14}\). It is not known whether they were true null mutants.

The haploid gene deletion collection contained non-essential genes. Essential gene functions cannot be studied in haploids. In contrast, a random insertion mutagenesis can target both essential and non-essential genes. Random insertions will generate various types of mutations (Figure 1.12). The
secondary structures in 5' UTRs are major regulatory elements for mRNA translation\textsuperscript{81}. RNA binding proteins (RBP) can recognize specific motifs in 5' UTRs and interact with the translation machinery to control gene expression by positively or negatively controlling translation of target mRNAs\textsuperscript{81}. Transposon insertions in the 5' UTR may disrupt secondary structures or RBP binding motifs and affect gene expression. Regulatory regions within the 3' UTR can influence translation efficiency, localization, and stability of the mRNA\textsuperscript{82}. Thus, transposon insertions in the 3' UTR may also alter gene expression. Insertions in the coding exons or introns may generate truncated proteins or proteins with altered functions. Different insertions in the ORFs may generate distinct mutant alleles. Moreover, random insertions not only disrupt protein-coding genes but also affect large intergenic regions where non-coding RNA genes and regulatory elements reside.

![Diagram](image_url)

**Figure 1.12** Random insertions produce various types of mutants.
Currently, there are no insertion mutant libraries for *S. pombe*. The Levin lab developed a *Hermes* transposon insertion system that can efficiently disrupt the *S. pombe* genome\(^3\) (Figure 1.13). The *Hermes* transposon is from the housefly *Musca domestica* and uses a transposase enzyme that binds the terminal inverted repeats (TIRs) at both ends of the transposon to excise an integrated copy\(^4\). The transposon then integrates itself with the genome\(^4\). The system efficiently disrupts *S. pombe* ORFs, indicating that it is able to cause gene disruptions. Assaying 26 integration sites recovered 14 (54\%) occurred in ORFs\(^3,5\). In addition, *Hermes* transposon insertion in the genome allows identification of insertion sites in a high-throughput manner. As deep-sequencing technology continues to improve, the cost to generate a sequenced transposon library will be greatly reduced. These features make the *Hermes* transposon system a good tool to generate an insertion mutagenesis library in *S. pombe*.

When the *Hermes* transposon inserts itself into the *S. pombe* genome, it is possible that the transposon can be mobilized and reinsert into another genomic locus. This “hopping behavior” has been observed in many transposons such as *PiggyBac*\(^6\) and *Sleeping beauty*\(^7\). Due to the “hopping behavior” of the transposon, we need to understand *Hermes* transposon excision and insertion in *S. pombe* before applying the system to mutagenesis on a larger scale.
Figure 1.13 Hermes transposon system. The Hermes transposon and transposase gene are carried by different plasmids. Hermes transposase expression is controlled by derepressible nmt1 promoter. Transposase releases the Hermes transposon from its donor plasmid. The transposon will randomly insert itself into the S. pombe genome.

1.2 Hermes transposon excision and DNA DSB repair

We want to utilize the Hermes transposon as a large-scale mutagenesis tool in S. pombe. It is important to understand whether the Hermes transposon can be mobilized in the S. pombe genome and whether transposon excision can cause mutations at the excision sites. Excision of the integrated transposon will leave DSBs in chromosomal DNA, which can lead to cell cycle checkpoint arrest, loss of genetic material, and genome rearrangements.

1.2.1 DNA DSB breaks cause genome instability and cancer progression

Origins of DSBs The genome is constantly under assault from exogenous and endogenous sources. Numerous DNA lesions are generated every day. The most deleterious are DNA DSBs. Introduction of a DSB can be mediated by a variety of exogenous agents such as ionizing radiation (IR) and
chemotherapeutic drugs. DSBs are also generated endogenously from normal biological processes such as meiotic chromosome recombination. DSB repair is essential for generating gametes with various allele combinations. DSBs are induced at immunoglobulin genes during V(D)J recombination. DNA repair generates class switch recombination and antigen receptor diversity for the T and B cells development. Another major endogenous source for DSBs is DNA replication stress that occurs when the replication fork collapses at a single strand DNA nick or a replication fork stalls at a interstrand crosslink (Figure 1.14).
Figure 1.14 Origins of DSB formation. DSBs arise by exogenous or endogenous causes. DSBs are generated directly from ionizing radiation (IR) (A). DSBs arise in endogenous processes such as meiotic recombination (B) and V(D)J recombination (C). Transposition induces DSBs. Hermes transposon excision from the genome generates DSBs. Chromosome ends are capped by hairpins, resembling V(D)J events (D). DNA replication stress induces DSBs. When replication forks collapse at single strand DNA nicks (E) or stall at inter-crosslink sites (F), one-ended DSBs are generated. The origins of DSBs are compiled from Chapman, et al. 2013 92.

Consequence of DSBs Unrepaired DSBs can result in senescence and trigger apoptosis or loss of genetic material. Mis-processing and repair of DSBs can lead to chromosomal aberrations associated with genomic instability and can ultimately result in carcinogenesis and human genetic syndromes. For example, chromosome translocations may activate oncoprotein expression 93. Deletions
can result in a loss of heterozygozity, which may inactivate a tumor suppressor gene.  

**DNA damage response upon DSB** Upon DNA damage, cells respond by activating checkpoints to block cell cycle progression and allow DNA repair. Cells have DNA damage checkpoints in G1/S, intra-S and G2/M phases. The MRN (MRE11-RAD50-NBS1) complex is a “first responder” to DNA breaks. The complex binds DSBs and recruits the DNA damage checkpoint kinase ataxia-telangiectasia-mutated (ATM), leading to H2AX phosphorylation and recruitment of DNA repair proteins. Activation of ATM leads to downstream effector kinase Chk2 phosphorylation and activation. In S-phase, another checkpoint kinase ATR is recruited and bound to RPA-coated single strand DNA, resulting in phosphorylation and activation effector kinases Chk1 and Chk2. Chk2 activation blocks cell cycle progression by inactivating Cdc25 phosphatases, which will inactivate cyclin-dependent kinases (CDK). (Figure 1.15).
Figure 1.15 DNA damage response. Following DNA double-strand breaks, the MRN complex first binds to the DSB then recruits and activates ATM/ATR. Chk2 is phosphorylated by ATM and inactivates CDC25, which will block cyclin dependent kinases and cell cycle progression.  

1.2.2 Different pathways for DNA repair

Cells have evolved several DSB repair mechanisms. NHEJ, Microhomology-mediated end joining (MMEJ) and HR are major DNA repair pathways and that have been extensively studied. Multiple players influence the pathway choice by competing the DSB ends and ensure that these repair pathways are executed accurately.

DNA end resection A major influence in the choice between DSB repair pathways is DNA end protection competing with DNA end resection. DNA end...
resection plays a key role in allowing the search for homologous sequences for base paring. In mammalian cells, Ku binds dsDNA ends with extremely high affinity, blocks resection and promotes NHEJ \(^{102}\). In contrast, MMEJ involves the loss of larger amounts of DNA sequence and does not require Ku \(^{103}\). Both HR and MMEJ require resection of a 5’ DNA strand at the DSB, while NHEJ does not. In humans and yeast, the MRN complex (Mre11-Rad50-Nbs1) and CtIP (called Ctp1/Sae2 in yeast) initiate DSB resection to promoting HR \(^{104}\).

MRN regulates a switch between NHEJ and HR with its nuclease activities. Mre11 endonuclease activity introduces nicks on the DNA duplexes, which triggers Exo1 dependent long-range DNA resection to promote HR \(^{105}\). Mre11 also removes protein adducts from DNA ends to promote HR. Spo11 induces DSBs in meiotic recombination and remains covalently attached to DSB ends. The Mre11 nuclease promotes the release of Spo11-bound DNA complexes to initiate DNA resection \(^{106}\). Significantly, Mre11 nuclease activity has recently been implicated in Ku removal, suggesting that MRN and Ku compete at DSB ends \(^{107,108}\).

**53BP1-chromatin Interactions promote NHEJ** The DNA damage response recruits DNA repair proteins. The tumor suppressors 53BP1 and BRCA1 are enriched at DSB sites and are key regulators of DSB repair by NHEJ or HR \(^{92}\). 53BP1 interacts with chromatin at DSB sites. 53BP1 binds dimethylated histone H4 Lys20 (H4K20me2) \(^{109}\). This interaction is crucial for limiting the resection of the DSB intermediates of immunoglobin gene class switch recombination (CSR) \(^{110}\), implying a role in inhibiting HR repair and promotion of
NHEJ (Figure 1.16). In the immune system during class switch recombination, highly repetitive DNA segments are recombined to generate the different classes of antibodies. DSBs form in multiple switch regions and are repaired by NHEJ. In the absence of 53BP1, DSB DNA end resection increases \(^{110}\). 53BP1 deficiency yields the pronounced CSR defects: in vitro CSR in 53bp-/- mice B cell levels are decreased by 90% relative to wild type \(^{111}\).

In contrast to the effect of 53BP1, BRCA1 is required for efficient HR \(^{112}\). How BRCA1 is able to counteract 53BP1 activity to suppress NHEJ remains unclear. BRCA1 may function as a scaffold protein that associates with many interaction partners such as CtIP and the MRN complex for initialization of DNA resections and promote HR \(^{113}\) (Figure 1.16).
Figure 1.16 DNA repair pathway choices. (Top panel) In G1 phase cells, oligomeric 53BP1 binds nucleosomes in DSB sites. 53BP1 binding mediates a state restrictive to DSB end processing, favoring repair by NHEJ. 53BP1 may be recruited by the presentation of H4K20me2 at DSB sites. (Lower panel) In S phase, BRCA1 proteins are enriched on chromatin, mediating a state permissive to DSB end resection. Ku and MRN compete at DNA ends. MRN binding to DSB sites initiate DNA resection, promoting HR repair. 

53BP1
H4K20me2
Ku

DNA end resection initiated,
DSB repair by HR

BRCA1
Ku
MRN-CTIP

DNA end resection is restricted,
DSB repair by NHEJ
1.2.3 DNA damage repair by NHEJ

NHEJ mediates the direct re-ligation of the broken DNA molecule 114. NHEJ is active at all phases of the cell cycle, whereas HR is only active during S/G2 phases when a homologous template is available 114. NHEJ repair includes steps: DNA end recognition, bridging broken DNA strands, DNA end processing and ligation of the broken ends 114 (Figure 1.17).

DNA end recognition and assembly of the NHEJ complex at DSB sites

The initial step in NHEJ is binding of the Ku heterodimer (Ku70 and Ku80) to the DSB 115. Crystallographic studies of Ku70/80 show that the heterodimer forms a ring-shaped structure that can slide onto the ends of the dsDNA helix 116. Ku70/80 binds to the DNA ends to protect ends from non-specific processing 116. Ku binding is important in maintaining genome integrity, as non-specific processing of DSB ends could lead to chromosomal aberrations. The bound Ku heterodimer also serves as a scaffold to recruit and assemble NHEJ complexes including DNA-PKcs 117, XRCC4 115, and DNA Ligase IV 118.

Bridging of the DNA ends

Ku recruits DNA-PKcs to the DSB ends. DNA-PKcs belongs to the phosphatidylinositol-3 (PI-3) kinase-like kinase family (PIKK) 119. In NHEJ repair, Ku and DNA-PKcs are suggested to play a role in the formation of a synaptic complex. The complex holds the broken DNA strands in close proximity 120. However, yeast lacks the ortholog of DNA PKcs and it is unclear which proteins bridge DNA ends in yeast NHEJ.
**DNA end processing** Depending on the nature of the damage, DSB DNA ends may require different enzymes to produce ligatable substrates, such as resection of DNA ends or removal of blocking groups of 5’-hydroxyls, 3’-phosphates or proteins \(^{121}\). In mammalian V (D) J joining, Artemis is implicated in resecting DNA ends for NHEJ. Artemis has a 5’ endonuclease activity, a 5’-3’ exonuclease activity on ssDNA, and the ability to remove the 3’-phosphoglycolate groups from DNA termini \(^{122}\). Artemis cleaves hairpin structure formed during V (D) J recombination \(^{122}\).

Mre11 is the core subunit of the MRN complex, which plays important roles in DNA resection to initiate HR repair. Mre11 has both 3’-5’ exonuclease activity on dsDNA duplexes and ssDNA endonucleolytic activities \(^{123}\). Whether the MRN complex is required in NHEJ DNA resection is not clear. In *S. cerevisiae*, end-joining defects associated with Mre11 deletion can be rescued by nuclease-defective Mre11, indicating that a structural rather than catalytic function of Mre11 is more important \(^{124}\).

**Ligation of the broken ends** The final step is the ligation of the broken ends by DNA Ligase IV. In mammalian cells, XRCC4 stabilizes and stimulates the ligation activity of Ligase IV \(^{125}\). *S. cerevisiae* Lig4/Lif1/Nej1 joins the aligned DNA ends \(^{126}\). In contrast, *S. pombe* uses Lig4/Xlf to rejoin the DNA ends \(^{127}\).
Figure 1.17 DNA end recognition and core NHEJ factors. In mammalian cells, when a DNA DSB is induced, Ku quickly binds to broken DNA ends. Ku (black shapes) serves as a scaffold to recruit the NHEJ core machinery. Ku and DNA-PKcs (orange crescent) tether DNA ends via formation of a synaptic complex. Artemis (not shown) may “clean up” DNA ends to generate substrates for ligation. XRCC4 and XLF aid DNA Ligase 4 (green shape) in sealing the broken DNA ends. DNA-PKcs and Artemis are absent in *S. cerevisiae* and *S. pombe*.

1.3 Significance and specific aims of this study

In the post-genomic era, the completion of genome sequences discovered thousands of genes, a majority of them with unknown functions. Genome-wide mutant collections are great tools for examining the functions of individual genes. The genome-wide yeast haploid deletion collections greatly accelerate systematic analyses of gene functions. However, the haploid deletion collections
are still insufficient for revealing gene functions because only non-essential genes are knocked out. Mutations in essential genes and a large number of non-coding genes are not included. Moreover, the haploid deletion collections only contain null mutants. A resource that allows rapid phenotypic screening of many kinds of mutants (e.g. null, partial or altered function) is needed. Such a resource does not exist for *S. pombe* yet.

We are interested in understanding the pathways that control aging. The mechanisms of aging are still unclear. To comprehensively identify aging regulators, we previously established a method to directly select long-lived mutants from a pool of random mutants. This approach requires an expanded mutant collection that goes beyond null mutants. For example, hypomorphic *daf-2* mutants of *C. elegans* exhibit a long-lived phenotype, but null mutants do not.

This dissertation shows how we created a *Hermes* transposon-based insertion mutant library in the fission yeast *S. pombe*. My further characterization of the *Hermes* transposon as a genetic tool required the analysis of *Hermes* transposon excision from the *S. pombe* genome, which led to novel discoveries regarding DNA damage repair by NHEJ. The specific aims of my dissertation were:

**Aim 1 (Chapter 2): Characterize *Hermes* transposon insertion and excision in *S. pombe*.**

To construct an insertion mutant library, we first optimized *Hermes* transposon insertion to obtain one insertion per cell. We also induced transposon
excision to determine its frequency and precision, which had not been previously characterized and would directly impact the library quality. A *Hermes* transposon excision from the *S. pombe* genome occurred at a low frequency (~$10^{-5}$-$10^{-3}$ per cell). The majority of the excisions left a frameshift mutation at the original locus. The sequence of *Hermes* excision sites and the analysis of excision in different mutants revealed that removal of *Hermes* occurred by a novel NHEJ pathway that required MRN and the dimerization domain of Mre11, but not an intact Mre11 nuclease domain.

**Aim 2 (Chapter 3): Construct and characterize a sequenced barcoded *Hermes* transposon mutant library in *S. pombe.*

The advantage of the insertion mutant library is that it can contain multiple types of known mutations: insertions in the UTRs or non-coding RNAs can alter gene regulation; insertion in the ORF can eliminate function or produce a truncated protein with altered function. Our library has the potential to include essential genes and multiple mutation forms for one gene.

I constructed a barcode-tagged *S. pombe* insertion mutant library that consisted of 9,024 transposon-insertion mutants in which >90% have a single transposon and a unique barcode. We developed a novel three-dimensional pooling strategy combined with a multiplexed high-throughput sequencing analysis pipeline to map transposon insertion sites and sequence the DNA barcodes. Our approach can be applied to other insertion-based mutation sites, allowing mapping with greatly reduced effort and cost.
Our analysis sequenced the insertion sites and barcodes of 4,095 barcode-tagged random *Hermes* transposon insertions in the *S. pombe* genome. Insertions occurred in ORFs, 5’ and 3’ UTRs and non-coding RNA genes. Mutants were arrayed in known locations in a series of 96-well plates. The current collection disrupted 268 unique essential genes, 1,472 unique non-essential genes and 589 unique non-coding RNA genes. Screening of a small subset of mutants (71) revealed that some insertions had novel phenotypes compared to the deletion library. Therefore, our *Hermes* insertion library will allow the identification of phenotypes distinct from the deletion mutants and thus has the potential to uncover new gene functions in a variety of cellular processes.
Chapter 2

The *Schizosaccharomyces pombe* Mre11-Rad50-Nbs1-Ctp1 Complex Promotes Non-Homologous End-Joining with Minimal Sequence Loss

Yanhui Li\(^1,2\), Michael Lajeunesse\(^3\), Kathleen L. Berkner\(^4\) and Kurt W. Runge\(^1,2\)*

\(^1\) 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 Biology, John Carroll University, University Heights, OH 44118\
\(^4\) Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 44195, USA

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

Author contributions:
Conceived and designed the experiments: KWR, YL,. Performed the experiments: YL, ML. Analyzed the data: YL, KWR. Wrote the manuscript: KWR, YL, KLB.
Note: This chapter is a manuscript to be submitted to *Genetics*.
Abstract

The Mre11-Rad50-Nbs1 (MRN) complex has important roles in DNA damage signaling, telomere replication, homologous recombination and microhomology-mediated end-joining. The role of MRN in non-homologous end-joining (NHEJ) is not clear. We found that MRN is required in Schizosaccharomyces pombe for NHEJ repair of non-ligatable ends arising from excision of the hAT transposon Hermes. Excision leaves a DNA double-strand break capped by hairpin ends similar to structures generated by palindromes, trinucleotide repeats and immunoglobin gene rearrangements, and requires nucleolytic processing before repair. Most repaired excision sites in S. pombe had less than 5 bp of sequence loss or mutation, implicating NHEJ. Repair levels were reduced more than 1000-fold in cells lacking the NHEJ factors Ku or DNA ligase 4, or each subunit of MRN. Loss of the MRN-associated nuclease Ctp1 caused a 10-fold reduction in NHEJ, and mutants with defective Mre11 dimerization domains reduced repair 300-fold. Surprisingly, Mre11 and Ctp1 nuclease functions were not required for repair, which contrasts with the known role of Mre11 nuclease functions in other forms of DNA repair, and indicates the requirement for an alternative nuclease in NHEJ. These results reveal a role for MRN-Ctp1 in the NHEJ repair of non-ligatable ends.
Introduction

DNA double-strand breaks (DSBs) are toxic lesions that can recombine with other chromosomal DNAs, and are therefore challenges to genomic integrity that must be repaired to allow cellular growth \(^{130-132}\). DSBs can be repaired by two major pathways in eukaryotes: homologous recombination (HR) and end-joining. HR is an error-free repair mechanism by which information lost at the DSB is replaced by copying the homologous sequence from another DNA molecule \(^{101,133,134}\). End-joining mechanisms are error-prone, and involve the direct joining of the two ends of the DSB after different amounts of DNA sequence loss. Non-Homologous End-Joining (NHEJ) fuses ends with a small sequence loss in a process that requires the Ku heterodimer (Ku70 and Ku80), while Microhomology-Mediated End-Joining (MMEJ) involves loss of larger amounts of DNA sequence and does not require Ku. Resection of a 5' DNA strand at the DSB is required for HR and MMEJ, but not for NHEJ \(^{135-139}\).

The Mre11-Rad50-Nbs1 (MRN) complex plays a central role in HR and MMEJ as well as activation of the DNA damage checkpoint and telomerase recruitment to telomeres \(^{97,123,140,141}\), is one of the first proteins to bind to DSBs and has been implicated in the processing of DNA adducts and DNA secondary structures \(^{142-144}\). Mre11 has both endo- and exonuclease activities that are important for generating the 3' single-stranded DNA overhang important for DSB repair in HR and MMEJ \(^{136,145,146}\). Rad50 is an ATPase that stimulates ATP-dependent DNA unwinding and Mre11 nuclease activation \(^{124,147}\). Nbs1 binds to Mre11 and allows recruitment of DNA damage checkpoint kinases ATM (in
mammals) or Tel1 (in fungi) to MRN-bound DSBs\(^{148-150}\). Structural studies indicate that MRN complexes can tether separated DNA ends together, which is thought to pair chromosomal sequences for HR or MMEJ repair\(^{151-155}\). Mre11 dimerization (reviewed in\(^{151}\)) may facilitate this process.

The role for MRN in NHEJ is much less clear. In \(S.\) cerevisiae, where the MRN complex is called MRX, NHEJ requires MRX in a plasmid end-joining assay that monitors the efficiency of recircularization of restriction enzyme cut plasmids transformed into yeast\(^{156}\). In contrast, \(S.\) pombe MRN is not required in the same assay\(^{157}\). Similarly, the presence or absence of MRN shows only a minor effect in a mammalian NHEJ assay where chromatin is cleaved in vivo by a restriction enzyme\(^{158,159}\). Different requirements for MRN/MRX in these eukaryotes are surprising as the nuclease functions and subunit interactions that form the complex are well-conserved\(^{101,148,149,160}\). Consequently, the role of MRN in NHEJ is still poorly understood.

DNA repair has also been studied by recovery from exposure to ionizing radiation and transposon excision, both of which are distinct from the assays that monitor restriction enzyme cut DNA in that upstream nucleolytic processing is also assessed. While ionizing radiation produces random breaks, excision of an individual transposon produces a lesion at a defined site. Transposon excision therefore allows one to examine the sequences of the DNA substrates and the repaired products in detail. The \(hAT\) transposons are ideal tools for this purpose as this transposon-transposase system can integrate into DNA in vitro and into the genomes of many eukaryotes\(^{161-164}\). The transposase that causes integration
also excises the transposon to generate covalently closed DNA hairpins \(^{84,165}\) that lack the free 5’ phosphate and 3’ hydroxyl groups required by DNA ligase. Repair therefore requires processing by cellular nucleases. DNA hairpins can normally form in eukaryotic genomes at the end of structures formed by palindromes, inverted repeats, trinucleotide repeats and during vertebrate VDJ joining \(^{84,143,166-168}\). The correct processing of DNA hairpins is essential for genome stability, as these structures can lead to chromosome translocations or gene amplification in tumors \(^{169-173}\). Therefore, understanding how DNA hairpins are repaired has important consequences for our understanding of genome integrity and carcinogenesis.

Work in \textit{S. cerevisiae} has provided some insight into the genes required for processing of hairpins that arise from palindromic genomic sequences as well as hAT transposon excision events. The nuclease activities of MRX and the associated enzyme Sae2 are required for normal repair of these palindromes, leading to the conclusion that Mre11 or Sae2 nuclease activity cleaves the DNA hairpin \(^{146}\). The repair of hAT transposon excision events also requires the \textit{S. cerevisiae} MRX and Sae2, as well as the NHEJ protein Ku. NHEJ joins ends with a small loss of sequence (0-5 nt); however transposon excision was usually associated with a larger sequence loss (8 - 22 nt). The results are therefore surprising as the genetics indicate a Ku-dependent NHEJ process but the repaired excision sites (transposon footprints) indicate a Ku-independent MMEJ process. These \textit{S. cerevisiae} results also contrast with hAT transposon excision in \textit{Drosophila}, Maize and other organisms in which the majority of transposon
footprints have a small number of base deletions, insertions or changes \(^{174-177}\). These comparisons suggest that \(S.\ cerevisiae\) may repair structures such as DNA hairpins by a combination of evolutionarily conserved processes as well as unique mechanisms not used in other eukaryotes.

\(S.\ pombe\) provides an alternative model system with the potential to yield significant new insights into DNA repair that may be more broadly conserved, as its repair proteins (e.g. MRN) and ratio of HR to end-joining are more similar to mammalian cells than \(S.\ cerevisiae\) (reviewed in \(^{178-180}\)). Our results on the repair of hAT transposon excision events in \(S.\ pombe\) revealed a novel role for MRN and its associated nuclease Ctp1 in NHEJ. The majority of transposon footprints reveal an NHEJ process with small deletions or mutations, and efficient repair required the NHEJ factors Ku and DNA ligase 4 as well as MRN. Surprisingly, the nuclease activities of Mre11 were not required for efficient repair. These results indicate a role for MRN in NHEJ repair of DNA hairpins, and that such repair frequently occurs with little or no sequence loss.
Materials and Methods

**Hermes transposon insertion**

To produce cells with only one transposon insertion, we developed a “transient transfection” assay using two plasmids that separately introduce the transposon and a transposase expression cassette. Wild type *S. pombe* KRP1 or KRP201 cells (Table 2.1) bearing the *Hermes* transposase expressing plasmid pHL2578u were grown on EMM + adenine, leucine and histidine (EMM-ura) medium\(^{181}\) to induce transposase expression. Cells were then transformed with 0.3, 1.0 or 3.0 µg of the *Hermes* transposon donor plasmid pHL2577 and divided between non-selective plates with either inducing (EMM + adenine, uracil, leucine and histidine) or non-inducing (YES) media. Cells were then grown for 24 hr at 30°C, equivalent to two to five cell divisions. The resulting lawns of cells were then replica plated to YES medium containing supplements that both select for the transposon (200 µg/ml G418) and against both plasmids (1 mg/ml FOA). After three days of growth at 30°C, colonies were counted and a subset were analyzed for insertions.

**Mapping Hermes insertion sites**

Sequences of the insertion sites were determined by Inverse-PCR. Genomic DNA from the KRP1 and KRP201 insertion strains (5 µg) were digested with *Sau3A* I (20 units in a 50 µl reaction incubated at 37°C for 16 hr, followed by inactivation at 65°C for 20 min). Digested DNA (1 µg) was diluted into a 200 µl ligation reaction with 40 units of T4 DNA ligase (New England Biolabs) and incubated at 16°C overnight. The resulting material was ethanol precipitated,
resuspended in 10 µl 10 mM Tris – 1 mM EDTA pH 8.0, and 500 ng of DNA was amplified by the outward facing primers pHL2577-3633S (5’ CACAACCTAACAACACAGTTGTTTG 3’) and pHL2577-3527AS (5’ TTGTGCTTATCTATGTGGCTTAC 3’) using Expand polymerase. The inverse PCR products were purified and sequenced with the primer Hermes_Sau3A I (5’ CACAACCTAACAACACAGTTGTTTG 3’). The genomic DNA adjacent to Hermes was compared to the S. pombe genome using the BLAST utility at Pombase (www. Pombase.org) to map the Hermes insertion site. The transposon insertions in KRP201 were also mapped using the enzymes Mse I, Apo I and Mfe I, and only single insertions were found. The insertion strains, along with all other strains used in these studies, are listed in Table 2.1.

Table 2.1 S. pombe strains used in this work.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild type and transposon insertion strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRP1</td>
<td>h- ade6-M216 leu1-32 ura4-D18 his7-366</td>
<td>This study</td>
</tr>
<tr>
<td>KRP3-4</td>
<td>h- ade6-M216 leu1-32 ura4-D18 his7-366 chr I 1,279,108::Hermes</td>
<td>This study</td>
</tr>
<tr>
<td>KRP3-3</td>
<td>h- ade6-M216 leu1-32 ura4-D18 his7-366 chr I 5,443,647::Hermes</td>
<td>This study</td>
</tr>
<tr>
<td>KRP201</td>
<td>h+ ade6-M216 leu1-32 ura4-D18</td>
<td>This study</td>
</tr>
<tr>
<td>P9G2</td>
<td>h+ ade6-M216 leu1-32 ura4-D18 chr I 633,603::Hermes</td>
<td>This study</td>
</tr>
<tr>
<td>P10G2</td>
<td>h+ ade6-M216 leu1-32 ura4-D18 chr II 3,487,136::Hermes</td>
<td>This study</td>
</tr>
<tr>
<td>P12G2</td>
<td>h+ ade6-M216 leu1-32 ura4-D18 chr I 1,057,536::Hermes</td>
<td>This study</td>
</tr>
<tr>
<td>P13B2</td>
<td>h+ ade6-M216 leu1-32 ura4-D18 chr I 2,684,662::Hermes</td>
<td>This study</td>
</tr>
<tr>
<td>P15B11</td>
<td>h+ ade6-M216 leu1-32 ura4-D18 chr II 147,681::Hermes</td>
<td>This study</td>
</tr>
</tbody>
</table>
Strains used to construct mutants

TN1257  h+ ade6-M216 leu1-32 ura4-D18 his3-D1 pku70Δ::KanMX  S Sanders
TN2099  h+ ade6-M210 leu1-32 ura4-D18 his3-D1 lig4Δ::KanMX  S Sanders
NR2840  h+ ade6-M216 leu1-32 ura4-D18 his3-D1 rad50Δ::KanMX  T Nakamura
TN2389  h- ade6-M216 leu1-32 ura4-D18 his3-D1 nbs1Δ::KanMX  T Nakamura
JW4166  h+ leu1-32 ura4-D18 mre11-13myc:kanMX6  P Russell
JW4167  h+ leu1-32 ura4-D18 mre11-L77K-13myc:kanMX6  P Russell
JW4169  h+ leu1-32 ura4-D18 mre11-L77K L154D-13myc:kanMX6  P Russell
JW4170  h+ leu1-32 ura4-D18 mre11-H68S-13myc:kanMX6  P Russell
JW4171  h+ leu1-32 ura4-D18 mre11-H134S-13myc:kanMX6  P Russell
P34G06  h+ ade6-M216 leu1-32 ura4-D18 mre11Δ::KanMX4  Bioneer Version 2
P20C05  h+ ade6-M216 leu1-32 ura4-D18 ctp1Δ::KanMX4  Bioneer Version 2

Deletion strains used in Figures. 2.8, 2.9, and 2.10.

KRP180  h- ade6-M216 leu1-32 ura4-D18 chr I 1,279,108::Hermes  This study
KRP175  h+ ade6-M216 leu1-32 ura4-D18 chr I 1,279,108::Hermes  This study
KRP176  h+ ade6-M216 leu1-32 ura4-D18 chr I 1,279,108::Hermes ctp1Δ::KanMX4  This study
KRP177  h+ ade6-M216 leu1-32 ura4-D18 chr I 1,279,108::Hermes mre11Δ::KanMX4  This study
KRP178  h+ ade6-M216 leu1-32 ura4-D18 chr I 1,279,108::Hermes rad50Δ::KanMX4  This study
**mre11 mutant strains used in Figures. 2.11, 2.12, 2.13.**

KRP240  
- h- ura4-D18 leu1-32 chr I 1,279,108::Hermes mre11-13myc:kanMX6
- This study

KRP241  
- h- ura4-D18 leu1-32 chr I 1,279,108::Hermes mre11-L77K-13myc:kanMX6
- This study

KRP243  
- h- ura4-D18 leu1-32 chr I 1,279,108::Hermes mre11-L77K-L154D-13myc:kanMX6
- This study

KRP245  
- h- ura4-D18 leu1-32 chr I 1,279,108::Hermes mre11-H134S-13myc:kanMX6
- This study

KRP275  
- h- ura4-D18 leu1-32 chr I 1,279,108::Hermes mre11-H134S-13myc:kanMX6 ctp1Δ::KanMX4
- This study

KRP263  
- h- ade6-M216 ura4-D18 leu1-32 chr I 1,279,108::Hermes mre11-13myc:kanMX6
- This study

KRP258  
- h- ade6-M216 ura4-D18 leu1-32 chr I 1,279,108::Hermes mre11-H68S-13myc:kanMX6
- This study

**Hermes transposon excision induction and DSB repair assays**

Transposon excision was induced, and the excision products examined by PCR. KRP3-3 and KRP3-4 cells bearing a *Hermes* insertion were transformed with 1 µg of the pHL2578u transposase expressing plasmid, and excision was induced by growth on EMM-ura. A small colony (1 mm diameter) was inoculated into 1 ml of EMM liquid medium with 2% glucose and adenine, leucine and
histidine medium and grown for 48 hr at 30°C to 3 – 5 x 10⁷ cells/ml. Cells were harvested, washed with sterile, milli-Q filtered water, resuspended in 100 µl zymolyase buffer (1.2 M sorbitol, 0.1 M Na phosphate pH 7.4, 2.5 mg/ml Zymolyase 20T (MP biomedicals LLC)) and incubated at 37°C for 3 hours. The cells were pelleted and incubated with 150 µl lysis buffer (0.1 M Tris, 50 mM EDTA, 1% SDS, pH 8.0) at 65°C 20 min, followed by the addition of 50 µl of 7.5 M NH₄OAc and incubation for 30 min on ice. The reaction was extracted with an equal volume of 25:24:1 phenol:chloroform:isoamyl-alcohol, precipitated with isopropanol, washed with 70% ethanol and resuspended in 60 µl of H₂O.

Genomic DNA (250 ng) of KRP 3-4 cells was used as template in a 30 µl “first round” PCR reaction with primers 3-4_S (5’ TCCCCGTATGTGAAATCAAA 3’) and 3-4_AS (5’ TCGCGCTTGCCATTCTTTCTAAT 3’) with 5 PRIME MasterMix (5 PRIME). The template DNA was denatured (94°C, 2 min), followed by 45 cycles of PCR (94°C 30s, 52°C 30s, 65°C 1min), followed by 5 min at 65°C. The first round PCR product (0.5 µl) was used as the template in a 30 µl reaction for the “second round” PCR using the nested primers 3-4_2S (5’ CCCCGTATGTGAAATCAAAGAAAGC 3’) and 3-4_2AS (5’ AATTGCCTCTCACGAGCAGCAGAT 3’). The DNA was denatured (at 94°C 2 min), followed by 30 cycles of PCR (94°C 30s, 56°C 30s, 65°C 1min). The PCR products (5 µl) were analyzed on 1% agarose gels stained with ethidium bromide. Gel images were captured on a Bio-Rad gel scanner with Quantity One software (v. 4.6.2.70, 2006), which inverts the grayscale images to dark bands on a white background. These raw images were used to determine the frequency of
*Hermes* excision as described in the Results. All mutant strains were analyzed in parallel with a wild type strain, which showed median excision frequencies from $10^{-4}$ to $3 \times 10^{-4}$ in all experiments. Excision events of KPR 3-3 cells were amplified in a similar manner using the primers 3-3S (5’ CATGCAATAACCCATTGAAAAA 3’) and 3-3AS (5’ GGGGTATGGAGTGAGAAGGTT3’).

To determine the sequence of the repaired excision events, PCR products were cloned into pCR2.1-TOPO using the TOPO TA kit (Life Technologies) and sequenced using the M13 forward (5’ TGTAAAACGACGGCCAGT 3’) or reverse (5’ CAGGAAACAGCTATGAC 3’) primers.

**Reconstruction assay**

To quantify the frequency of *Hermes* excision, we mixed wild type cells and cells with a transposon to mimic different proportions of transposon excision, and used these mixtures to prepare DNA to produce a standard curve for excision frequency. KRP3-4 cells ($5 \times 10^7$) bearing a *Hermes* insertion were mixed with $10^5$, $10^4$, $10^3$, $10^2$ or 10 wild type KRP1 cells that lack *Hermes*. Genomic DNA was prepared as described for the excisions assays and then analyzed in the first and second PCR assays. Each experiment was performed in triplicate to reveal the variation in amplification of rare templates.
Results

A transient transfection assay in *S. pombe* to maximize single transposon insertions

To use transposon excision to investigate DNA repair requires that cells bear only a single transposon, and we therefore developed an approach to efficiently generate single genomic insertions. We adapted a system Evertts *et al.* developed for *S. pombe*, in which one plasmid expresses the transposase while another contains a transposon with a selectable marker for G418 resistance 83. Induction of the transposase allows the transposon to insert into the genome, and growing cells on medium with G418 can select for cells with transposon integrants. This system was efficient at generating transposon insertions, often several per cell, and has been successfully used to map *S. pombe* genes required for growth 182.

To maximize the fraction of cells with single insertions, the transposon plasmid was only transiently maintained in cells expressing transposase prior to selection for genomic insertions (Figure 2.1A). The plasmids containing the transposon and the transposase expression cassettes were altered so that both contained markers (*URA3*+, *ura4*+) that could be selected against with the same drug (FOA). *S. pombe* cells bearing the transposase expression plasmid were grown up under inducing conditions to produce a population of cells with transposase enzyme, and then transformed with transposon plasmid. Cells were plated onto non-selective medium, allowed to grow for 24 hr and then replica plated to medium that selects for the transposon and against both plasmids. This
24 hrs of growth and cell division allowed cells bearing genomic transposon insertions to lose the mitotically unstable transposon and transposase plasmids. Consequently, selection for the transposon and against the plasmids identifies those cells that allowed transposition and quickly lost both plasmids. This approach limits the opportunity for multiple transposition events.

The growth on non-selective medium following transformation was performed using conditions where transposase was, or was not, induced (Figure 2.1B). This approach was used because the production of cells with single transposon insertions might require continual exposure to transposase, or, alternatively, continual transposase exposure might produce multiple genomic insertions. Analysis of the number of potential transposon insertions revealed that growth for 24 hr on medium that induced the transposase produced about 10 times more insertions than growth on non-inducing medium (Table 2.2). The effect of different amounts of transposon plasmid was also tested, which showed that the number of transformants per microgram decreased as the amount of transposon DNA increased (Table 2.2). Thus, the transient transfection approach could produce significant numbers of cells with transposon insertions.
Figure 2.1 A transient transfection assay to generate mutants with single transposon insertions. A. Cells bearing the pH2578u plasmid express the *Hermes* transposase from the *nmt1* promoter under inducing conditions (in EMM-based medium). The pH2577 transposon donor plasmid contains the ends of the natural transposon flanking a kanamycin selectable marker (KanMX) that allows resistance to G418. Introducing the transposon plasmid into cells expressing the transposase allows insertion of the transposon into the genome, generating a target site duplication of 8 bp (indicated by thin arrows). B. Cells expressing transposase were transformed with
different amounts of the transposon plasmid. To monitor whether continued expression of transposase impacted transposition, each transformation was split to non-selective plates with non-inducing (YES) or inducing (EMM) media. The EMM plates contained supplements to complement cellular auxotrophies (where aluh signify the addition of adenine, leucine, uracil and histidine)\(^1\). After 24 hr of growth during which cells can lose both plasmids, the lawn of cells was replica plated onto medium that selects for the transposon (G418-resistance) and against the \textit{ura4} and \textit{URA3} genes (FOA) on the two plasmids. The yield of colonies bearing transposon insertions is shown in Table 2.2.

<table>
<thead>
<tr>
<th>Medium</th>
<th>Transforming DNA ((\mu)g)</th>
<th>G418(^R) FOA(^R) Transformants(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>YES</td>
<td>0.3</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>95</td>
</tr>
<tr>
<td>EMM+alu(^h)</td>
<td>0.3</td>
<td>325</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>690</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>804</td>
</tr>
</tbody>
</table>

\(^1\) Number of colonies on the final YES+G418+FOA plates following growth on non-selective medium in which transposase is (EMM) or is not (YES) induced.

\(^2\) Adenine, leucine, uracil and histidine (alu\(h\)) were added to EMM medium to complement all auxotrophies and allow continued transcription of transposase.

To determine if these cells contained one or more transposon insertions, the transposon locations in the genome of 12 strains bearing insertions from two different transformations were mapped by inverse PCR (Materials and Methods)\(^1\). Ten of these strains had single insertions (Figure 2.2). The remaining two strains retained the transposon plasmid with no genomic insertions, possibly by mutating the \textit{URA3} negative selectable marker on the plasmid. All 10 genomic insertions showed the 8 bp target site duplication bordering the insertion site, as expected. These data indicate that the transient transfection approach can efficiently produce large numbers of strains with single
transposon insertions. Two independent strains (KRP3-3 and KRP3-4) were subsequently used to examine *Hermes* excision and repair of the excision site.

Figure 2.2  The majority of *Hermes* transposon insertions from the transient transfection protocol are single events. Transposase producing cells were transformed with the transposon plasmid (as in Figure 2.1) and the site of *Hermes* genomic integration was determined by inverse circle PCR. The results from two independent transformations are shown. The top two integrants (KRP3-3 and KRP3-4) are in the strain KRP1 while the remaining integrants are in KRP201. All of the randomly chosen colonies shown contain single insertions. The distance from intergenic *Hermes* insertions to the ATG or stop codon of the nearest ORF is shown. Two colonies from the screen did not have a genomic *Hermes* insertion, but retained the transposon plasmid (pHL2577, Figure 2.1).
Transposon excision footprints indicate repair by NHEJ

While Hermes transposase can catalyze transposon insertion and excision by itself\textsuperscript{84,165}, repair of the DSB left by transposon excision relies on cellular factors. Transposon excision has not been tested in S. pombe, so it was unknown whether repair of hAT transposon excision leaves footprints that create new mutations or restore the wild type sequence as occurs, for example, in Drosophila\textsuperscript{184}. We therefore tested whether expressing transposase caused the removal of an inserted Hermes transposon. An important consideration in S. pombe is that haploid cells spend most of the cell cycle in G2, and are therefore functionally diploid and can repair DSBs by HR\textsuperscript{185}. This repair will recreate the original insertion if Hermes excises from only one of the two chromosomes, because repair of the hairpin capped DSB will copy the other chromosome that retains the Hermes transposon (Figure 2.3). Consequently, permanent removal of the transposon is expected to result only from NHEJ and MMEJ events. Such end-joining events are usually associated with the G1 phase of the cell cycle, and would be predicted to occur at lower frequency than HR events.
Figure 2.3 Removal of a *Hermes* transposon requires end-joining repair.

Transposon excision produces hairpin-capped DSB ends that must be repaired, or cells will lose essential genetic information during mitosis and die. Repair by end-joining (NHEJ or MMEJ, left panel) involves nuclease activities and ligation of the broken ends, which can be identified by PCR. Rare events can be detected by a second round of PCR with a set of nested primers. Repair by HR in G2 cells that retain an unexcised transposon (right panel) will regenerate the *Hermes* insertion. This product will not be detected in this assay due to the use of short extension times that do not allow amplification of the *Hermes* insertion. Grey and black arrows indicate primers for the first and second PCR reactions, respectively.

We designed a serial PCR assay to detect excision events over a wide range of frequencies. Primers were chosen at a sufficient distance from the insertion site based on the known size range of deletions generated by transposon excision in other organisms (i.e. *S. cerevisiae*, *Drosophila* and Maize). PCR extension times were then chosen that would amplify the excision
site but not amplify the larger Hermes insertion. The products of the first PCR were then used in a second PCR with a set of nested primers to quantitate rare events (Figure 2.3). This assay was semi-quantitative as determined by a reconstruction test. Different amounts of wild type cells that lack a Hermes insertion were mixed with cells that have an insertion, and DNA preparations from these mixtures were tested by PCR. The results showed that the fraction of cells where the transposon had excised, i.e. excision frequency, could be monitored over a $10^4$-fold range (Figure 2.4A). By comparing the levels of both the first and second PCR products, frequencies that differed by 10-fold were distinguishable in the range of $10^{-4}$ to $10^{-6}$ events/cell.
Figure 2.4  Quantitation and sequence analysis of Hermes transposon excision events.
A. A reconstruction test for estimating the frequency of *Hermes* excision. The number of wild type cells (WT) indicated, which lack a *Hermes* insertion, were mixed with $5 \times 10^7$ KRP 3-4 cells bearing a transposon and used to prepare DNA. PCR to detect only the excision products (as in Figure 2.3) was performed in triplicate for each sample. The approximate frequencies of the excision-mimicking wild type cells are shown. B. *Hermes* excision events occur at low frequency. Single KRP 3-4 cells bearing a single transposon insertion and the transposase plasmid were grown to $5 \times 10^7$ cells under conditions that induced transposase expression. DNA was prepared and used in 2 rounds of PCR. Each lane shows the results from a single culture. Lane “C” is the control PCR product DNA from wild type cells to show the size of the expected fragment. C. Sequences of *Hermes* excision events indicate NHEJ. Products from the first round PCR in panel B were cloned and sequenced. The underlined sequence is the 8 bp duplication generated during transposon insertion. Base changes are indicated in red, and deletions by a colon. A bracket indicates the sequences used to model the mechanism of repair (described below). Similar results were obtained with another strain containing a different transposon insertion (KRP 3-3 cells, Figure 2.6).
The frequency of transposon excision has not been tested in *S. pombe*, and we therefore measured this value and whether it could be determined quantitatively. KRP 3-4 cells containing a single transposon insertion (Figure 2.2) were transformed with the transposase expression plasmid to generate colonies from single cells. After growth of individual colonies, DNA was prepared and transposon excision frequency assayed in 10 independent colonies. Excision frequency varied over an approximately 100-fold range, from $1 \times 10^3$ to $1 \times 10^5$ cells. This wide range of frequencies suggests a low frequency of excision that occurs at different times during growth of the culture, with events that occur early producing higher frequencies and events that occur late producing lower frequencies. We therefore followed previously established methods of quantitating low frequency events with high variance by determining the median excision frequency of the 10 different cultures to obtain an excision frequency of a given strain $^{186-190}$. KRP 3-4 cells showed a median excision frequency that ranged from $10^{-4}$ events/cell (Figure 2.4B) to $3 \times 10^{-4}$ events/cell in a total of 4 independent assays (not shown), indicating that the median excision frequency was reproducible. These events depended upon transposase, as the frequency of these events in the absence of this enzyme was more than 1000-fold less (Figure 2.5A).
Figure 2.5 Efficient Hermes transposon excision requires transposase expression. A. KRP 3-4 cells lacking the transposase expression plasmid were tested in the transposon excision assay. The low level of correct second round PCR products indicates a median excision frequency of less than $10^{-7}$ per cell, or 1000-fold less than the excision frequency in transposase-expressing cells (Figure 2.4). Thus, transposase is required to produce the excision products shown in Figure 2.4. The assay in part A was performed in cells bearing the pku70Δ (B) or rad50Δ (C) mutations and lacking the transposase expression plasmid. The aberrant excision events seen in these mutants in Figure 2.8 were not detected in the absence of transposase.

To examine the repair of the DSB induced by Hermes excision, products from the first PCR from the colonies in Figure 2.4B were cloned and sequenced. The majority of recovered excision events (42/58) had sequences that strongly indicated NHEJ: 0 to 5 bp of the flanking sequences were deleted, and the mutations were confined to sequences near the point of end-joining (Figure 2.4C, Figure 2.6). The same types of events were observed for the excision of two independent transposons (in KRP3-3 and KRP3-4 cells), indicating that this repair was independent of the neighboring DNA sequences. These results are inconsistent with MMEJ, which generates larger deletions $^{136}$, and are quite different from excision in S. cerevisiae, where the majority of hAT transposon excision sites showed larger deletions (8-22 bp). The remaining S. pombe excision events included 2 revertants where one of the 8 bp target site
duplications had been removed, indicating that *Hermes* excision can restore the original wild type allele in *S. pombe*. Interestingly, a small number of excision events retained a portion of the *Hermes* transposon accompanied by a deletion of chromosomal sequences. While the origin of these events is not due to NHEJ and is unknown, these types of products were not obtained in cells that did not express transposase (Figure 2.5A), indicating that the presence of the transposase is required to generate these unusual products.

![Figure 2.6 Hermes excision footprint sequences from KRP 3-3 cells indicate repair by NHEJ.](image)

**Figure 2.6** *Hermes* excision footprint sequences from KRP 3-3 cells indicate repair by NHEJ. *Hermes* excision events were monitored by cloning the first round PCR products from several colonies in a KRP 3-3 cell transposon excision assay performed as in Figure 2.4. The underlined sequence is the 8 bp duplication generated during transposon insertion. Base changes are shown in red, and deletions are indicated by colons. Fourteen of the sixteen events show small deletions (0 - 5 nt) and mutations that implicate NHEJ.\textsuperscript{136,138}

**NHEJ repair of excision events requires Ku, DNA ligase 4 and MRN**

The NHEJ pathway depends upon Ku and DNA ligase 4 activities, and we therefore tested whether these factors are required for repair of *Hermes* excision
events. As the hairpin ends generated by excision must be converted to ligatable DNA ends prior to repair, we also tested a candidate nuclease. Hairpin cleavage by MRN appeared to be most likely because it is broadly implicated in DNA repair and biochemical and genetic evidence indicate that both human MRN as well as *S. cerevisiae* MRX and its associated protein Sae2 can cleave DNA hairpins \(^{142,191,192}\). Structural analysis indicates that Mre11 dimers should be able to facilitate subsequent base pairing of 5′ single-stranded overhangs for end-joining \(^{151}\). Our theoretical model of hairpin cleavage using MRN-mediated 5′ base pairing and repair (Figure 2.7) could explain the three most frequently isolated KRP 3-4 transposon footprints (bracketed in Figure 2.4C). MRN was therefore tested.
Figure 2.7  Theoretical mechanism of transposon footprint formation suggests a role for the MRN complex in S. pombe NHEJ. Hermes excision produces hairpin ends and such structures can be cleaved by MRN/MRX in vitro. The structure of Mre11 indicates that it can mediate base pairing of the 5’ overhangs. Cleavage of the hairpin (red arrows) and subsequent removal of unpaired bases and/or strand extension similar to the normal processes associated with NHEJ could account for the 3 most frequently observed KRP 3-4 excision events (bracketed in Figure 2.4C). Therefore, a role for MRN in NHEJ was investigated.

Hermes excision frequency in cells lacking a subunit of Ku (pku70Δ) or DNA ligase 4 (lig4Δ) was greatly reduced, as no excision events were detected in the first round of PCR, and few correct excision events were detected in the second PCR (Figure 2.8). The second round of PCR revealed aberrant products.
not due to normal NHEJ. These rare products were of the incorrect size, and sequencing revealed that many of these contained portions of the *Hermes* transposon and were accompanied by large deletions (Figure 2.9). These products were not observed in the absence of transposase (Figure 2.5B), indicating that their generation requires transposase activity. The results from the 2 PCRs from all *pku70Δ* and *lig4Δ* cultures indicated that the frequency of excision was $10^{-7}$ events/cell or less (Figure 2.10). Thus, efficient repair of *Hermes* excision events occurs by the Ku-dependent NHEJ pathway.

**Figure 2.8 Efficient NHEJ requires Ku70, DNA ligase 4 and MRN.** KRP 3-4 cells bearing no mutation (WT) or the indicated mutations were tested in the excision assay. Ten individual colonies were tested for each strain. Lane “C” is the control PCR product from wild type DNA to show the size of the expected NHEJ fragment. Footprint sequences and the median excision frequencies are shown in Figure 2.9 and 2.10, respectively.
Figure 2.9 Hermes excision footprints from different NHEJ mutants. Sequencing of the KRP 3-4 excision events from mutant cells revealed that some (mre11Δ, nbs1Δ, ctp1Δ) gave footprints similar to wild type cells. Other mutants (pku70Δ, rad50Δ) retained a portion of the Hermes transposon (shown by a hollow box) and showed larger deletions. These results suggest that other repair processes can heal the hairpin capped DSBs, but at a much lower efficiency.
Excision levels in cells lacking each of the MRN subunits were very similar to those in cells lacking Ku (Figure 2.8, Figure 2.10). No products were detected in the first PCR, and the second PCR showed either no product or products of aberrant size. Sequencing of products from the second PCR of the rad50Δ and mre11Δ mutants revealed aberrant products, similar to those identified in pku70Δ cells (Figure 2.9), which also required transposase for their production (e.g. Figure 2.5C). This requirement of MRN for efficient NHEJ repair of Hermes excision events contrasts with the results from end-joining assays with cut plasmids where the MRN complex is dispensable. A double mutant lacking DNA ligase 4 and Mre11 gave results very similar to the single mutant lacking
DNA ligase 4 (Figure 2.8, Figure 2.10), consistent with Mre11 and DNA ligase 4 acting in the same pathway.

**The Mre11 dimerization domain is required for efficient NHEJ**

The MRN complex has been implicated in holding two DNA ends in close proximity in HR, and mutations in the dimerization domain of Mre11 impair this function \(^{151}\). Structural analysis of the archael Mre11 revealed a key pair of leucine residues at the dimer interface. Conversion of either of these hydrophobic residues to charged residues disrupted dimerization and, when placed into *S. pombe* Mre11, inhibited DNA repair, with simultaneous mutation of both residues causing even greater inhibition \(^{151}\). We therefore tested whether these Mre11 dimerization domain mutants (Leu77Lys and the Leu77Lys Leu154Asp double mutant in *S. pombe* Mre11) affected NHEJ.

The *Hermes* insertion from KRP 3-4 cells was introduced into the wild type, *mre11-L77K* and *mre11-L77K L154D* strains from Williams *et al.* \(^{151}\) and excision frequency was monitored. The wild type cells showed a median excision frequency similar to the original KRP 3-4 strain (i.e. between $10^{-3}$ and $10^{-4}$ (Figure 2.11, Figure 2.12)). In contrast, both the *mre11-L77K* single mutant and *mre11-L77K L154D* double mutant showed reduced excision frequencies. Fewer products were observed in the first PCR and several cultures had no excision products or aberrant products in the second PCR (Figure 2.11). The median excision frequency for the *mre11-L77K* mutant was ~30-fold lower than the wild type strain, while the frequency for the double mutant was ~300-fold lower (Figure 2.11, Figure 2.12). The altered frequency in these mutants indicates that
an intact Mre11 dimerization domain is required for efficient NHEJ.

**Figure 2.11** Mutations in the Mre11 dimerization domain, but not the nuclease domain, greatly reduce NHEJ. Strains bearing the *mre11* mutations indicated were tested in the excision assay as in Figure 2.4. The indicated median excision frequencies and excision footprint sequences are shown in Supplemental Figures 2.12 and 2.13, respectively.
Figure 2.12 Graphical summary of Hermes excision frequencies of different mre11 mutants from Figure 2.11. Each data point represents the excision frequency of a culture derived from a single colony, and the red bar shows the median frequency.

Mre11 and Ctp1 nuclease functions are dispensable for efficient NHEJ.

Mre11 in the MRN complex has both single-strand DNA endonuclease and 3’ to 5’ exonuclease in vitro activities, which has led to the hypothesis that this complex plays a role in the repair of DNA hairpins by HR. The structure of the archael Mre11-DNA complex revealed key residues required for exonuclease activity or both nuclease activities. Mutation of the residues equivalent to *S. pombe* Mre11 H68S have ablated the in vitro exonuclease activity of archael enzyme, while the mutation equivalent to *S. pombe* H134S abolished both exo- and endonuclease activities. *S. pombe* cells bearing the exonuclease deficient *mre11-H68S* allele have the same DNA damage
phenotype as wild type cells, while the nuclease deficient \textit{mre11-H134S} cells show the same increased DNA damage sensitivity as cells lacking Mre11 \textsuperscript{151}. We therefore tested the effect of both nuclease deficient alleles on NHEJ.

Surprisingly, excision frequency in the \textit{mre11-H68S} and \textit{mre11-H134S} cells was substantially higher than that of cells lacking Mre11 or the Mre11 dimerization mutants, and only slightly less than that of wild type cells. Excision frequencies were approximately 10\% or more of the wild type strain level (Figure 2.11, Figure 2.12), which strongly contrasts with the absolute requirement for these activities in HR \textsuperscript{151}. The sequences of the transposon footprints from these nuclease mutants were very similar to those from wild type cells (Figure 2.13, Figure 2.4C). The frequency and type of excision events in these mutants indicate that the nuclease activities of Mre11 are largely dispensable for NHEJ.
Figure 2.13 Sequences of mre11 nuclease mutant excision events. Hermes excision events from sequencing the first round PCRs (Figure 2.11) show transposon footprints similar to wild type cells but also contain mutations outside of the 8 bp direct repeats (underlined). Base changes are shown in red, deletions indicated by a colon and portions of the Hermes transposon are shown by a hollow box. The bottom sequence from ctp1Δ mre11-H134S cells also contains a “C” insertion that is shown below the sequence.

These results indicated that another nuclease besides Mre11 processes the Hermes excision site to produce a ligatable DNA end for NHEJ. The only other nuclease implicated to date in hairpin cleavage is Ctp1, which is MRN associated. Ctp1 is a homolog of mammalian CtIP and *S. cerevisiae* Sae2, both of which have nuclease activity in vitro\(^\text{160,191,194,195}\). Excision in cells lacking Ctp1 was approximately 30-fold less than wild type cells, and at a higher frequency than the MRN mutants (Figure 2.8, Figure 2.10). Sequencing of products from the first PCR of DNA from ctp1Δ cells revealed transposon footprints similar to those seen in wild type cells (Figure 2.9). This small reduction in repair of the
excision site in cells lacking Ctp1 compared to the other mutants tested raised the question as to whether Ctp1 plays a redundant role in hairpin processing. We therefore examined *Hermes* excision in a *ctp1Δ mre11-H134S* mutant to test if cells lacking both nuclease activities would greatly reduce NHEJ.

Surprisingly, the median excision frequency in the *mre11* nuclease-deficient mutant lacking Ctp1 was nearly the same as in wild type cells (Figure 2.12), with a clearly detectable product in the first PCR in 9 of the 10 cultures (Figure 2.11). Sequencing of the products from independent *ctp1Δ mre11-H134S* cultures revealed two classes of excision footprints. Ones similar to wild type cells were detected that include small deletions and mutations, consistent with repair by NHEJ (Figure 2.13). The second class of footprints had mutations in sequences flanking the 8 bp direct repeats, which was only observed in Mre11 nuclease mutants and not in any other *Hermes* excision events in the different strains we examined. These results indicate that loss of the Mre11 and Ctp1 nuclease activities still allow normal NHEJ repair of *Hermes* excision events, and reveal a more error-prone mechanism that also occurs in these cells.
Discussion

Transposon excision that monitors both breakage and end-joining is a valuable approach for studying the processes of DNA repair, and it was used to gain insight into this process in *S. pombe*. We developed an efficient method to produce single *Hermes* transposon insertions in the *S. pombe* genome and showed that excision can be induced to remove the transposon. The sequence of the excision sites combined with the requirement for Ku and DNA ligase 4 showed that these excision events occur by NHEJ. However, in marked contrast to how *S. pombe* repairs DNA ends produced by restriction enzyme digestion, this NHEJ process also required MRN and, to a lesser extent, Ctp1\(^{157,195,196}\). Loss of each MRN component reduced NHEJ to the same extent, and produced similar transposon footprints, as loss of Ku and DNA ligase 4 (Figures 2.8, 2.9, 2.10). Surprisingly, and in contrast to HR\(^{151}\), Mre11 and Ctp1 nuclease functions were largely dispensable for NHEJ (Figure 2.11, 2.12). A primary role for MRN is to hold ends together\(^{151-155}\), and an intact Mre11 dimerization domain was required for repair (Figure 2.11), indicating a role for this domain *S. pombe* NHEJ. While the MRN-Ctp1 complex was the most likely candidate for hairpin cleavage based on its broad impact on repair and known properties\(^{146}\), our data suggest that an alternative activity opens this structure. Possible candidates are DNA topoisomerase II\(^{197,198}\) and the xeroderma pigmentosum G enzyme\(^{199,200}\), whose functions have been implicated in more specialized types of repair. Our results show that *Hermes* transposon excision in *S. pombe* provides an excellent model system to understand the processing of similar DNA structures that can
arise from palindromes, inverted repeats and trinucleotide repeats in mammalian genomes to threaten genomic stability and drive tumorigenesis 168,172,201.

The requirement of MRN for repair of hairpins but not restriction cut DNA 157 indicates that MRN is required for events upstream of end-joining, and is of interest regarding the known functions of Ku. The Ku heterodimer has the ability to bind and join the ends of restriction enzyme cut DNA 202,203, which could allow end-joining when restriction enzyme digested plasmids are transformed into cells 157,204,205. Ku can also bind DNA hairpins in vitro 206,207, and could similarly tether the hairpin-capped ends of the DSB created by Hermes excision. However, the requirement for MRN and the Mre11 dimerization domain in the NHEJ repair of Hermes excision events shows that this Ku activity is insufficient in vivo. MRN-Ctp1 association with DNA breaks has been proposed to remove Ku to allow the recruitment of HR factors such as the DNA damage checkpoint kinase ATR 108. We propose that MRN-Ctp1 also allows recruitment of factors that open DNA hairpins to allow NHEJ (Figure 2.14). Future work with the Hermes system may help reveal such factors.
Figure 2.14 Model for the role of MRN-Ctp1 in NHEJ repair of Hermes excision sites. The initial hairpin capped ends are bound by Ku, which has end protection functions that may prevent further processing \(^{203,206,207}\). MRN-Ctp1 is proposed to bind and replace Ku and allow synapsis to bridge the two DNA ends (as in \(^{208}\)) and to recruit factors that open the hairpin. Subsequent base pairing at microhomologies (yellow box) may be potentiated by MRN or other factors prior to ligation. While MRN has a known role in strand resection \(^{136,145,146}\), the most frequently observed end-joining events (Figure 2.4, Figure 2.9) indicated that little or no strand resection occurred to expose homologies.

Such physical analysis of DNA hairpin processing will require improvements to increase the levels of DNA intermediates in vivo, as the level of transposase-mediated excision was relatively low (about 1 in \(10^3\) to \(10^4\) cells, Figures 2.4, 2.11). Potential approaches would be to increase the expression and
nuclear localization of the transposase to allow the production of hairpins at much higher frequencies. We note that the present levels of excision are sufficient to use *Hermes* as a genetic tool to produce novel mutants, as these strains could be isolated by combining our PCR assay to detect excision with the methods used to isolate individual lambda phage clones from a genomic DNA library. Therefore, the *Hermes* excision system has significant potential for both the physical analysis of DNA repair and the generation of novel mutants.

MRN has been implicated in two other end-joining events besides transposon excision in *S. pombe*, where the substrates of the reaction are more poorly defined. Decottignes described an assay where a *ura4* fragment transformed into cells spontaneously circularized to capture DNA, either from mitochondria or fragments co-transformed with the *ura4* DNA. Rad50 was not required for circularization but was required for the capture of exogenous DNA. Analysis of circularized plasmids revealed both large and small sequence loss, but Ku dependence was not tested. These events suggest a potential role for the MRN complex in the capture phenomenon, which could occur by NHEJ or MMEJ.

Reis *et al.* studied the effect of MRN by extending the Decottignes assay, and also by studying the end-to-end joining of defective telomeres in G1-arrested cells that also lack the major double-stranded telomere binding protein Taz1. Both events required Ku, indicating repair by NHEJ, and a role for Ku, MRN and DNA ligase 4, but not Ctp1, was shown in both assays. Mutants in the Mre11 dimerization domain also blocked end-joining in both assays, similar to
what we observed with transposon excision. However, end-joining activity in both assays was defective in the \textit{mre11-H68S} and \textit{mre11-H134S} mutants, which strongly contrasts with our entire results showing that these two mutants were both clearly proficient in NHEJ (Figure 2.11). These data indicate that the processing of DNA hairpins in vivo has distinct requirements compared to the events that occur during DNA transformation or the fusion of defective telomeres. Thus, the MRN complex performs related but distinct functions in these different NHEJ reactions, which can be differentiated by the \textit{mre11} nuclease-deficient alleles.

There are significant differences between the products of end-joining in \textit{S. pombe} and \textit{S. cerevisiae} arising from the excision of \textit{hAT} transposons that leave hairpin ends. The majority of the transposon footprints we observed in \textit{S. pombe} showed deletion of small numbers of bases typical for an NHEJ reaction (Figure 2.4C), while similar events in \textit{S. cerevisiae} produced larger deletions more typical of MMEJ. This increased level of deletion suggests that the amount of strand resection prior to end-joining is greater in \textit{S. cerevisiae} compared to \textit{S. pombe}. We note that in mammalian cells, the majority of NHEJ events involving the ligation of DSBs created by in vivo restriction enzyme cutting also show only small amounts of DNA sequence loss\textsuperscript{158}, similar to our results in \textit{S. pombe}. These considerations suggest that results on NHEJ processing in \textit{S. pombe} may have broader implications for mammalian cells. For example, the repair of DNA adducts resulting from etoposide treatment of arrested human cells occurs by NHEJ in a process that requires MRN\textsuperscript{144}, but the extent of sequence removal is
unknown. An extension of our S. pombe results suggests that these NHEJ events may occur with only small amounts of DNA sequence loss or mutation at each DSB. It may therefore be possible to enhance etoposide toxicity if one can devise a combination therapy that increases the level of strand resection during repair.
Chapter 3

A *Schizosaccharomyces pombe* transposon insertion library for high-throughput genome-wide studies

Yanhui Li$^{1,2}$, Neil Molyneaux$^2$, Gang Zhou$^3$, Mark Adams$^2$, Kathleen L. Berkner$^4$ and Kurt W. Runge$^{1,2,*}$

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

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

Author contributions:
Conceived and designed the experiments: KWR, YL, MA. Performed the experiments: YL, GZ. Analyzed the data: YL, NM, KWR. Wrote the manuscript: YL, KWR, KLB.

Note: This chapter is a manuscript in preparation.
Abstract

Arrayed libraries of mutants constructed for phenotypic screening have been used in elucidating gene function in the post-genomic era. Yeast haploid gene deletion libraries have pioneered this effort. However, these deletion mutations do not reveal many of the potential phenotypic impacts of a given gene that may occur with altered expression or partial function, and essential genes whose deletions are lethal are absent from these collections. We have therefore constructed a defined library of insertion mutants that will provide a wider range of phenotypes for use in phenotypic screening. Our library contains 4,095 sequenced DNA barcode-tagged transposon insertion mutants in the S. pombe genome. While the construction and sequencing of similar transposon insertion libraries in other organisms were labor-intensive endeavors, we developed a novel three-dimensional pooling strategy and multiplexed high-throughput analysis pipeline to sequence the transposon insertion sites and DNA barcodes from thousands of samples simultaneously. The library contains insertions in non-essential genes, essential genes whose complete deletion is lethal, and genes for non-coding RNAs. Insertions occur in the open reading frames, 5’ and 3’ regulatory regions of genes.

As a test of the utility of the library, we examined the phenotypes of selected mutants for growth on a medium containing a non-fermentable carbon source or in the presence of the topoisomerase I inhibitor CPT. This analysis revealed that, while some insertion mutants have the same phenotypes as the cognate deletion mutants, others have novel phenotypes. This library therefore
represents an important resource for the international S. pombe community and a valuable approach for the construction and analysis of insertion mutant libraries in a wide variety of model systems.
**Introduction**

The fission yeast *S. pombe* has become an increasingly popular model for studying eukaryotic biological processes such as aging and gene-drug interactions\(^{128,212}\). *S. pombe* provides powerful molecular genetics combined with evolutionarily conserved processes that are similar to mammals including cell-cycle control, RNA splicing, RNAi-mediated gene silencing, telomere function and chromosomes with large repetitive centromeres\(^{80,213}\). Comparative sequence analysis revealed 50 *S. pombe* genes orthologous to human disease genes associated with cancer and metabolic, neurological and cardiac diseases\(^80\).

Mutant libraries in model organisms have been extensively used in large-scale screenings to systematically define gene functions\(^9,11,12,14,15\). In *S. pombe*, a haploid gene deletion collection has been created by removal of predicted coding sequences\(^{14}\). This collection (Version II) deleted 3,308 nonessential genes, ~68% of all annotated protein-coding genes\(^{14}\). A major issue with this library is that many of the deletions only removed part of the coding sequences. It is not known whether they are true null mutants\(^{14}\). Another issue is that essential genes and large numbers of functionally unknown non-coding RNA genes are not included in the haploid collection.

An alternative to gene deletion is random insertion mutagenesis. In generating genome-wide mutant collections, insertion mutagenesis has several advantages over targeted deletion technologies, i.e. homologous recombination, ZFN, TALEN or CRISPR\(^{47}\). First, random insertion by genetic elements such as
transposons or retroviruses can disrupt essential, non-essential and non-coding genes as well as regulatory elements such as promoters, enhancers and silencers. Second, random insertion can generate a wider variety of phenotypes such as null and hypomorphic mutants and the mis-expression of proteins, which are more valuable when modeling human disease genes. Third, while similar gene deletions or mutations can be made with ZFN, TALEN or CRISPR, these systems require significant effort to generate a mutation in specific gene compared to a random insertion. Fourth, with the development of high-throughput sequencing, the challenge of characterizing a library of random insertion sites can be overcome by sequencing the DNA flanking the insertion sites and mapping these sites to the genome.

Currently, there are no insertion mutant libraries for *S. pombe*. The *Hermes* transposon can efficiently insert into the *S. pombe* genome, indicating its potential as a tool for large-scale mutagenesis. *Hermes* originated in the housefly *Musca domestica* and uses its transposase enzyme to bind the terminal inverted repeats (TIRs) at both ends of the transposon to excise an integrated copy, which can then integrate elsewhere in the genome. This system has been adapted to a two-plasmid system where the transposase is separately expressed to allow the direct selection of cells with insertions of a transposon which is present on a second plasmid encoding drug resistance. Insertion of the modified *Hermes* 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 efficient construction and characterization of a genome-wide insertion library required a method to fully utilize high-throughput sequencing and connect those results to individual mutants in a large array of mutant cultures. We developed a novel three-dimensional cell pooling strategy to sequence and map individual Hermes transposon insertion sites to the genome. Our strategy allows processing and sequencing of thousands of mutants at one time. The pipeline is applicable to any insertion element generating defined boundaries when inserted into the genome. The strategy we developed makes the construction and sequencing of a mutant library by random insertion a very rapid and attractive way to study gene functions systematically in various model organisms.

Here we report the construction of a barcode-tagged Hermes transposon insertion mutant library for S. pombe. We generated 4,095 individual haploid mutants carrying 4,381 transposon insertions. Individual mutants with sequenced transposon integration sites and DNA barcodes were arrayed in 96-well plates. Over 90% of the strains carried a single insertion. The collection disrupted ∼20% of S. pombe annotated essential genes, ∼40% of the non-essential genes and ∼30% of the non-coding RNA genes. The Hermes transposon insertion library complements the gene deletion collection and will serve as a valuable resource for studying S. pombe gene functions.
Results

Construction of a barcode-tagged *Hermes* transposon insertion mutagenesis library

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 have been successfully employed in the gene deletion collections of *S. cerevisiae* \(^\text{15}\) and *S. pombe* \(^\text{14}\) to enable phenotypic analysis of the whole collection in pooled competitive growth assays \(^\text{14,212,215-217}\). To generate a *Hermes* transposon collection amenable to such pooled assays, we introduced DNA barcodes into the *Hermes* transposon such that each mutant gene would be tagged by 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 cloned into the *Hermes* transposon vector and transformed into *E. coli DH5α* to produce about 1-2 x10^5 bacterial colonies. Ten transformations were performed to generate 10 barcoded-*Hermes* transposon plasmid libraries. Each library was used to generate \(~1,000\) *S. pombe* insertion mutants (Figure 3.1) with a >99% probability that each barcode was unique.
Figure 3.1 Overview of Hermes library construction. (A) Transposase is expressed from the nmt1 promoter, which is active on EMM (minimal medium) but not on YES (rich medium). Transposon plasmids were transformed into S. pombe cells grown on EMM and preloaded with transposase. The cells 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 transposon carried a KanMX marker for G418 resistance. 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 and no transposon or transposase plasmid retained could grow on the YES+G418+5-FOA plate. A total of 96 plates (9,024 mutants) were picked into 96-well plates. (B) The architecture of the *Hermes* transposon. TIR, terminal inverted repeats.

We previously established a method of efficiently generating single *Hermes* transposon insertions in *S. pombe* (Figure 3.1) (Chapter 2). 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. Loss of both plasmids was then selected on a medium containing 5-FOA. *S. pombe* cells bearing the transposase plasmid were grown under inducible conditions, then transformed with barcode-tagged transposon plasmids. Cells were allowed only 2-5 divisions for transposition, then transferred to YES+G418+5-FOA to select for transposon integration 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 (Figure 3.1).

**Development of a novel three-dimensional pooling and high-throughput multiplexed sequencing strategy to map transposon integrations and DNA barcodes**

To sequence the transposon insertion sites and the associated DNA barcodes in a cost- and labor-effective manner, we developed a 3D pooling
strategy. Combined with next-generation sequencing, we processed cells from twenty-four 96-well plates (2,256 \textit{S. pombe} mutants) for sequencing on the Illumina GAIIx 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. Each layer contained 12 rows and 16 columns (Figure 3.2A). Cells were collected as 16 row pools, 12 column pools and 12 layer pools. These 40 pools contained 3 copies of each mutant in a different row, column and layer pool. Determining the location of a mutant in each row, column and layer pool, provides three-dimensional coordinates that identified each mutant in the 24 plate sub-library.
A

Stack  Row Pool  Column Pool  Layer Pool

B

16 Row pools, 12 column pools and 12 layer pools

Prepare genomic DNA from each pool

Barcodes

Transposon ends

Amplify barcodes with Index tag primers

Digest with Mse I, Apo I and Mfe I

Ligate on linkers and amplify with index tag primers

Amplify to add adaptors for Illumina sequencing

Combine products from 40 pools for sequencing
Figure 3.2 Overview of 3D pooling strategy and multiplexed high-throughput sequencing to map transposon integrations and DNA barcodes. (A) 3D pooling. The collection of 96 plates was divided into four groups of 24 plates. Each group (2,256 mutants) contained plates stacked 2 plates per layer with a total of 12 layers. Cells were cultured to saturation and collected in the format of a Row Pool (144 strains per pool), a Column Pool (196 strains per pool) and a layer Pool (192 strains per pool). A total of 40 pools of cells were collected, including 16 Row Pools, 12 Column Pools and 12 Layer Pools. (B) Multiplexed high-throughput sequencing. Genomic DNAs were prepared from each pool and then arrayed into 96-well plates and digested with three restriction enzymes. Double-strand linkers were ligated to the genome fragments. Genome fragments from the same pools were mixed, followed by amplification of transposon-genome junctions from both ends of insertion. PCR products of each pool were further amplified by indexed primers. DNA barcodes were directly amplified from each pool of genomic DNA using primers to common sequences that flank all of the barcodes and add index tags specific to each pool. After the addition of Illumina-sequencing adaptors by PCR, transposon-genome flanking fragments and DNA barcodes products from 24 plates of mutants were pooled and sequenced in a single lane of Illumina GAIIx or HiScan for a 79 bp single-end read. (C) Integration site mapping for 3D pooling.
**strategy.** After sequencing, flanking genomic sequences at both ends of the transposon were aligned to the *S. pombe* reference genome to identify the insertion points using the Bowtie software. 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. The independent right end and left end reads served as an internal validation of the decoded transposon insertion sites.

The 3D pooling strategy utilized three important principles. First, each pool was constructed with a unique sequence or index tag that identified which pool contained the associated sequence. Second, each pool was amplified in three ways with primers specific to the *Hermes* left or right ends and to the barcode. The combination of the index tag and specific primer sequences identified the pool and the type of sequence (barcode or *Hermes*-genome junctions). We therefore used 79 nt 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 triangulation software approach was designed that 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. 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 (Figure 3.2A) identified the individual well.
containing this barcode sequence. 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 sequences 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. The 24 plates contained 2,256 mutants, and each barcode and transposon end was sequenced 3 times, once each in a row, column and layer pool. About 20,000 distinct products (9 x 2,256) were expected. The 40 row, column and layer pools were processed in 40 PCR reactions to create DNA libraries for sequencing on the Illumina GAIIx platform, which can output ~10 million reads. This sequencing depth allowed each product to be read about 500 times. Ligation-mediated PCR (LM-PCR) was employed to amplify transposon flanking DNA sequences (Figure 3.2B, Figure 3.3)\textsuperscript{219,220}. Genomic DNA from each pool was digested by different restriction enzymes, MseI and Apol/MfeI, to increase the chance of capturing appropriately sized flanking genomic DNA fragments for sequencing libraries. The fragment size of Hermes DNA library for sequencing was around 150-1000 bp. The Illumina sequencer can analyze fragments ranging from 200 to 500 bp, excluding adapters. Larger
fragments up to 1000 bp can be sequenced, but with increasingly lower efficiency and yield \(^{221}\). 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.
Figure 3.3 Amplification of Hermes Insertion sites. An example for the Hermes left side is shown. Genomic DNA from a row, column or layer pool was fragmented by restriction enzyme digestion. The predicted average fragment size in the S. pombe genome was 66 bp for Mse I, 266 bp for Apo I and 2,962 bp for Mfe I. 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. In 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 (TIR). 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 (xxxxxxxx). To adapt the Illumina high-throughput sequencing platform, Illumina adaptors and sequencing primers were added to the final PCR products. The same approach was adapted to sequence the Hermes right end using different specific primers and index tags.

The DNA barcodes were directly amplified from pools of genomic DNA (Figure 3.2B, Figure 3.4). 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 GAIIx machine. The
whole insertion mutant collection of 9,024 mutants assembled in 96 plates was sequenced in a total of four lanes to yield 79 bp single-end reads.

**Figure 3.4 Amplification of barcodes.** Genomic DNAs extracted from pooled 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.

**Sequence identification and assignment to individual strains**

We developed a customized bioinformatics pipeline to decode transposon insertions and DNA barcodes using the 3D pooling strategy (Figure 3.2C). 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 triangulation of row, column,
and layer pool reads assigned the barcodes and integration sites to strains. We manually examined the barcodes and integration sites from 20 randomly chosen strains, and found that all assignments were consistent with the high-throughput mapping.

The pipeline successfully decoded insertion sites in unique regions of the genome. However, if the insertion site was within repetitive sequences such as the centromere, sub-telomere, rDNA repeats and mating type region, the triangulation program could not assign the integration site to a location in 96-well plates. Strains with insertions in the repeated DNA were present in the original set of transposon insertion mutants. The genomic sequences from the row, column and layer pools were compared to the \textit{S. pombe} genome by BLAST, and repeated DNA sequences were detected. Because we could not map these sequences to microtiter plate wells in the library, the number of insertions in repetitive DNA is unknown.

\textbf{\textit{S. pombe} mutations generated from random insertions of the Hermes transposon}

We successfully mapped \textit{Hermes} transposon integrations sites and DNA barcodes in 4,095 \textit{S. pombe} mutants out of the 9,024 mutants sequenced. A total of 4,391 distinct insertion sites were recovered. 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 (Figure 3.5A), as defined by the \textit{S. pombe} genome database as of May, 2013\textsuperscript{222}. The frequency
of insertion in each chromosome was proportional to chromosome size (Figure 3.5B), consistent with nearly random integration of \textit{Hermes}.

The \textit{Hermes} transposon shows 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 \textit{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 or UTRs ($p<0.01$) (Figure 3.5C).
**A**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of strains sequenced</td>
<td>9,024 (96 plates)</td>
</tr>
<tr>
<td>No. of strains mapped in unique region</td>
<td>4,095</td>
</tr>
<tr>
<td>No. of insertion sites</td>
<td>4,391</td>
</tr>
<tr>
<td>within genes (protein coding, non-coding genes)</td>
<td>3,022 (69%)</td>
</tr>
<tr>
<td>in intergenic regions</td>
<td>1,369 (31%)</td>
</tr>
</tbody>
</table>

**B**

- **Chr I**: 5.58 Mb
- **Chr II**: 4.54 Mb
- **Chr III**: 2.45 Mb

The diagrams illustrate the distribution of genetic material across three chromosomes (Chr I, Chr II, Chr III) along the forward strand, with specific markers indicating gene locations.
Figure 3.5 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, rDNA blocks. (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.
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 protein-coding regions. The insertions in essential genes were more enriched in UTRs (87%). Only 13% were in the protein-coding regions (Figure 3.6A). We further analyzed the insertions within the first and last 150 bp of the protein-coding 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 coding region of non-essential genes than that of essential genes (Figure 3.6B). All of the *Hermes* insertions in the coding regions of essential genes are shown in Figure 3.6C.

The heterochromatic centromere and telomere regions contained unique regions that could be mapped, but also these transcriptionally silenced regions had fewer transposon insertions identified. We recovered only 4 insertions within centromeres and 5 from chromosome I and II telomere regions (Figure 3.7A). This result most likely reflects silencing of the kanMX gene, which impairs the selection for G418 resistance. Consistent with this idea, ~500 insertions were identified in the Chromosome III ends containing rDNA tandem gene arrays, where gene silencing is much weaker \(^{45}\) (Figure 3.7B,C).
(A) Hermes insertions in *S. pombe*

<table>
<thead>
<tr>
<th></th>
<th>Essential genes</th>
<th>Non-essential genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NO. of insertions</strong></td>
<td>363</td>
<td>2,470</td>
</tr>
<tr>
<td>5’ UTR</td>
<td>201</td>
<td>967</td>
</tr>
<tr>
<td>3’ UTR</td>
<td>115</td>
<td>479</td>
</tr>
<tr>
<td>Coding exons</td>
<td>47</td>
<td>947</td>
</tr>
<tr>
<td>Introns</td>
<td>0</td>
<td>77</td>
</tr>
</tbody>
</table>

(B) The comparison of *Hermes* insertion distributions in essential and non-essential genes
Figure 3.6 The distribution of Hermes insertions in essential genes and non-essential genes. (A) In essential genes, Hermes insertions were enriched in the UTRs. The table shows 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 exons). 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. All 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.
Figure 3.7 The distribution of *Hermes* mutants on 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 I.
and II. (C) **Hermes Distributions in *S. pombe* rDNA.** The *Hermes* collection contained ~480 insertions in the rDNA blocks at the ends of chromosome III. In chromosome III, telomere sequences were not available because the unsequenced rDNA repeats prevent sequencing.

Our current *Hermes* insertion collection contains 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 3.1). The collection contains individual mutants with *Hermes* insertions in the same gene, but at different sites. There were 363 essential gene insertions, 2,470 non-essential gene insertions and 1,159 non-coding gene insertions in the collection.

**Table 3.1 Comparison of the *Hermes* library with the Bioneer library.**

<table>
<thead>
<tr>
<th></th>
<th>Essential genes</th>
<th>Nonessential genes</th>
<th>Non-coding genes</th>
<th>Intergenic insertions</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. pombe</em></td>
<td>1,260</td>
<td>3,576</td>
<td>1,876</td>
<td></td>
</tr>
<tr>
<td>Hermes Library</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Total Mutants)</td>
<td>363</td>
<td>2,470</td>
<td>1,159</td>
<td>1,369</td>
</tr>
<tr>
<td>Hermes Library</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(unique genes)</td>
<td>268</td>
<td>1,472</td>
<td>589</td>
<td>1,369</td>
</tr>
<tr>
<td>Haploid ORF</td>
<td>0</td>
<td>3,308</td>
<td>N/A</td>
<td>0</td>
</tr>
<tr>
<td>Deletion Library*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


**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 large number of mutants in essential genes that we isolated. To determine whether additional classes of genes 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 \(^{16}\). Thus, mutants in respiratory chain complex genes are expected to have impaired growth on non-fermentable carbon sources. There were 16 mutants in the library categorized under the GO term ‘respiration chain complexes I, II, III, IV, V and assembly proteins’ (Table 3.2). All 3 mutants with insertions in coding exon, 2 mutants with insertions in introns and 3 with insertions in the 5’ UTR showed defective growth (Figure 3.8A). In contrast, 6 UTR mutants (3 each in the 5’ and 3’ UTRs) showed normal growth (Figure 3.8B, Table 3.2). Thus, our insertion library can identify regions of the UTRs critical for gene function.
Table 3.2 **Hermes** insertion mutant strains used in Figure 3.8.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Encoded Protein</th>
<th><em>Hermes</em> Insertion</th>
<th>Growth on YEEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>atp14</td>
<td>F1-ATPase subunit H (predicted)</td>
<td>5' UTR Defective</td>
<td></td>
</tr>
<tr>
<td>atp16</td>
<td>F1-ATPase delta subunit (predicted)</td>
<td>5' UTR Defective</td>
<td></td>
</tr>
<tr>
<td>atp2</td>
<td>F1-ATPase beta subunit Atp2</td>
<td>5' UTR Defective</td>
<td></td>
</tr>
<tr>
<td>atp3</td>
<td>F1-ATPase gamma subunit (predicted)</td>
<td>5' UTR Normal</td>
<td></td>
</tr>
<tr>
<td>atp5</td>
<td>F0-ATPase delta subunit (predicted)</td>
<td>Intron Defective</td>
<td></td>
</tr>
<tr>
<td>cox13</td>
<td>cytochrome c oxidase subunit Vla (predicted)</td>
<td>3' UTR Normal</td>
<td></td>
</tr>
<tr>
<td>cox14</td>
<td>cytochrome c oxidase assembly protein Cox14 (predicted)</td>
<td>3' UTR Normal</td>
<td></td>
</tr>
<tr>
<td>cox17</td>
<td>metallochaperone Cox17 (respiratory chain complex IV assembly)</td>
<td>5' UTR Normal</td>
<td></td>
</tr>
<tr>
<td>cox5</td>
<td>cytochrome c oxidase subunit V</td>
<td>3' UTR Normal</td>
<td></td>
</tr>
<tr>
<td>cox6</td>
<td>cytochrome c oxidase subunit Vi (predicted)</td>
<td>5' UTR Normal</td>
<td></td>
</tr>
<tr>
<td>cox9</td>
<td>cytochrome c oxidase subunit VIIa (mitochondrial respiratory chain complex IV)</td>
<td>5' UTR Normal</td>
<td></td>
</tr>
<tr>
<td>mzm1</td>
<td>mitochondrial respiratory chain complex III assembly protein Mzm1</td>
<td>Intron Defective</td>
<td></td>
</tr>
<tr>
<td>sdh1</td>
<td>succinate dehydrogenase Sdh1 (mitochondrial respiratory chain complex II)</td>
<td>Coding Exon Defective</td>
<td></td>
</tr>
<tr>
<td>sdh2</td>
<td>succinate dehydrogenase (ubiquinone) iron-sulfur protein subunit</td>
<td>Coding Exon Defective</td>
<td></td>
</tr>
<tr>
<td>tim11</td>
<td>F0-ATPase subunit E</td>
<td>5' UTR Normal</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.8 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. So 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. Of the 16 insertion mutations in genes for respiration chain complexes examined, all insertions in coding exons and introns and some in the 5’ UTR showed defective growth on the medium that requires full mitochondrial function.
Camptothecin (CPT) is a topoisomerase inhibitor causing replication fork breakage when the replisome encounters the topoisomerase-CPT-DNA adduct. 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. We searched our insertion library for mutation in the CPT-sensitivity gene set and found 54 mutants with insertions in 37 genes (Table 3.3). We tested these mutants for growth on different concentrations of CPT.

Many of the insertion mutants showed sensitivity to either low or high concentrations of CPT, including 17 mutants with insertions in coding exons, 1 with an insertion in an intron and 14 with insertions in the 5’ or 3’ UTR. One coding exon mutant and 20 UTR insertion mutants showed no change in CPT sensitivity, thus identifying regions of the 20 genes that are dispensable for CPT resistance (Figure 3.9, Figure 3.10, Table 3.3). We observed 19 insertion mutants that were more sensitive to CPT than wild type cells were less sensitive than the corresponding deletion mutant, showing that our insertion mutants had distinct phenotypes from the deletion library.
Table 3.3 *Hermes* insertion strains used in the CPT-sensitivity assay.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Insertion</th>
<th>Insertion point</th>
<th>Sensitive to CPT(µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tdp1</td>
<td>ORFΔ</td>
<td>Coding Exon III:530334-530335</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coding Exon III:529408-529409</td>
<td>5</td>
</tr>
<tr>
<td>alp14</td>
<td>ORFΔ</td>
<td>Coding Exon III:65899-65900</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' UTR III:1011566-1011567</td>
<td>10</td>
</tr>
<tr>
<td>rnc1</td>
<td>ORFΔ</td>
<td>Coding Exon III:66958-66959</td>
<td>10</td>
</tr>
<tr>
<td>mug24</td>
<td>ORFΔ</td>
<td>5' UTR III:1918372-1918373</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>mhf2</td>
<td>ORFΔ</td>
<td>Coding Exon III:2101723-2101724</td>
<td>10</td>
</tr>
<tr>
<td>dad5</td>
<td>ORFΔ</td>
<td>3' UTR III:1670295-1670296</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3' UTR III:1669975-1669976</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3' UTR III:1670148-1670149</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3' UTR III:1670121-1670122</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>git3</td>
<td>ORFΔ</td>
<td>5' UTR III:1545380-1545381</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' UTR III:1545742-1545743</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' UTR III:1546043-1546044</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>SPCC1393.11</td>
<td>ORFΔ</td>
<td>5' UTR III:821273-821274</td>
<td>10</td>
</tr>
<tr>
<td>sgf73</td>
<td>ORFΔ</td>
<td>5' UTR III:2123943-2123944</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>SPBC725.10</td>
<td>ORFΔ</td>
<td>Coding Exon II:1225168-1225169</td>
<td>5</td>
</tr>
<tr>
<td>kin1</td>
<td>ORFΔ</td>
<td>5' UTR II:2694295-2694296</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>trt1</td>
<td>ORFΔ</td>
<td>Coding Exon II:2065448-2065449</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>sat1</td>
<td>ORFΔ</td>
<td>3' UTR II:3858892-3858893</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3' UTR II:3858892-3858893</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coding Exon II:3856366-3856367</td>
<td>5</td>
</tr>
<tr>
<td>Gene</td>
<td>ORF Δ</td>
<td>Region</td>
<td>Sense</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>-------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>bdc1</td>
<td>ORFA</td>
<td>5' UTR II:2425546-2425547 10</td>
<td></td>
</tr>
<tr>
<td>rps402</td>
<td>ORFA</td>
<td>5' UTR II:1651866-1651867 10</td>
<td></td>
</tr>
<tr>
<td>csn1</td>
<td>ORFA</td>
<td>3' UTR II:4027332-4027333 10</td>
<td></td>
</tr>
<tr>
<td>sgf29</td>
<td>ORFA</td>
<td>Coding Exon II:2419576-2419577 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5' UTR II:2420034-2420035 Not Sensitive</td>
<td></td>
</tr>
<tr>
<td>SPBC16A3.17c</td>
<td>ORFA</td>
<td>Coding Exon II:4265968-4265969 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coding Exon II:4266736-4266737 5</td>
<td></td>
</tr>
<tr>
<td>cbp1</td>
<td>ORFA</td>
<td>3' UTR II:3510683-3510684 Not Sensitive</td>
<td></td>
</tr>
<tr>
<td>rad26</td>
<td>ORFA</td>
<td>Coding Exon I:4453049-4453050 10</td>
<td></td>
</tr>
<tr>
<td>SPAC9.02c</td>
<td>ORFA</td>
<td>3' UTR I:1457123-1457124 Not Sensitive</td>
<td></td>
</tr>
<tr>
<td>bst1</td>
<td>ORFA</td>
<td>Coding Exon I:1686804-1686805 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coding Exon I:1688372-1688373 5</td>
<td></td>
</tr>
<tr>
<td>mrc1</td>
<td>ORFA</td>
<td>Coding Exon I:4209641-4209642 10</td>
<td></td>
</tr>
<tr>
<td>ryh1</td>
<td>ORFA</td>
<td>3' UTR I:1193477-1193478 Not Sensitive</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3' UTR I:1192931-1192932 15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3' UTR I:1193455-1193456 Not Sensitive</td>
<td></td>
</tr>
<tr>
<td>pab1</td>
<td>ORFA</td>
<td>5' UTR I:509884-509885 5</td>
<td></td>
</tr>
<tr>
<td>rad57</td>
<td>ORFA</td>
<td>5' UTR I:2122831-2122832 10</td>
<td></td>
</tr>
<tr>
<td>mkh1</td>
<td>ORFA</td>
<td>Coding Exon I:615881-615882 5</td>
<td></td>
</tr>
<tr>
<td>rad1</td>
<td>ORFA</td>
<td>Coding Exon I:4979380-4979381 10</td>
<td></td>
</tr>
<tr>
<td>rho2</td>
<td>ORFA</td>
<td>Coding Exon I:4423192-4423193 5</td>
<td></td>
</tr>
<tr>
<td>mre11</td>
<td>ORFA</td>
<td>Coding Exon variant</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>UTR Type</td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>SPAC12G12.12</td>
<td>5' UTR</td>
<td>I:436049-436050</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>3' UTR</td>
<td>I:439737-439738</td>
<td>10</td>
</tr>
<tr>
<td>rps3001</td>
<td>ORFΔ</td>
<td>I:325051-325052</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>5' UTR</td>
<td>I:325051-325052</td>
<td>10</td>
</tr>
<tr>
<td>arp42</td>
<td>ORFΔ</td>
<td>I:4352502-4352503</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td>5' UTR</td>
<td>I:4352502-4352503</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>gpa2</td>
<td>ORFΔ</td>
<td>I:2519102-2519103</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td>5' UTR</td>
<td>I:2519102-2519103</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>rad9</td>
<td>ORFΔ</td>
<td>I:1714404-1714405</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Coding Exon</td>
<td>I:1714404-1714405</td>
<td>5</td>
</tr>
<tr>
<td>SPBC20F10.07</td>
<td>ORFΔ</td>
<td>I:3295617-3295618</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td>5' UTR</td>
<td>I:3295617-3295618</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>mug24</td>
<td>ORFΔ</td>
<td>II:3295617-3295618</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td></td>
<td>5' UTR</td>
<td>III:1918455-1918456</td>
<td>Not Sensitive</td>
</tr>
<tr>
<td>Gene</td>
<td>YES</td>
<td>CPT</td>
<td>Insertion location</td>
</tr>
<tr>
<td>--------------</td>
<td>-----</td>
<td>-----</td>
<td>--------------------</td>
</tr>
<tr>
<td>SPAC1556.01c (rad50)</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>SPAC16.01 (rho2)</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>SPBC1921.07c (sgf29)</td>
<td></td>
<td></td>
<td>Intron</td>
</tr>
<tr>
<td>SPCC757.09c (no1)</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>SPAC227.07c (pab1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPCC757.09c (no1)</td>
<td></td>
<td></td>
<td>5'UTR</td>
</tr>
<tr>
<td>SPCC757.09c (no1)</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>SPCC757.09c (no1)</td>
<td></td>
<td></td>
<td>5'UTR</td>
</tr>
<tr>
<td>SPCP31B10.05 (tdp1)</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>SPCP31B10.05 (tdp1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPCP31B10.05 (tdp1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPCP31B10.05 (tdp1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPCP31B10.05 (tdp1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPCP31B10.05 (tdp1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPAC824.02 (bpl1)</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>SPAC824.02 (bpl1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPAC824.02 (bpl1)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>Gene</td>
<td>YES</td>
<td>CPT</td>
<td>Insertion location</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----</td>
<td>-----</td>
<td>--------------------</td>
</tr>
<tr>
<td>WT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(leu3::Hermes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPBC16A3.170</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>(SPBC16A3.170)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPAC1F3.02c</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>(mkh1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAC664.07c</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>(rad9)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPBC725.10</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>(SPBC725.10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPCC1393.11</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>(SPCC1393.11)</td>
<td></td>
<td></td>
<td>CDS</td>
</tr>
<tr>
<td>SPCC417.02</td>
<td></td>
<td></td>
<td>3UTR</td>
</tr>
<tr>
<td>(dad5)</td>
<td></td>
<td></td>
<td>5UTR</td>
</tr>
<tr>
<td>SPAC694.06c</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>(mro1)</td>
<td></td>
<td></td>
<td>3UTR</td>
</tr>
<tr>
<td>SPCC895.07</td>
<td></td>
<td></td>
<td>3UTR</td>
</tr>
<tr>
<td>(alp14)</td>
<td></td>
<td></td>
<td>3UTR</td>
</tr>
<tr>
<td>SPAC20H4.07</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>(rad57)</td>
<td></td>
<td></td>
<td>5UTR</td>
</tr>
<tr>
<td>SPBC215.03c</td>
<td></td>
<td></td>
<td>ORFA</td>
</tr>
<tr>
<td>(csn1)</td>
<td></td>
<td></td>
<td>3UTR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure 3.9** 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. Figures demonstrate that the lowest CPT concentration for a 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.
Figure 3.10 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 (Figure 3.11 and Table 3.3). The different insertion phenotypes identify gene regions required and dispensable for CPT resistance.

Importantly, one 5’ UTR mutant displayed a CPT resistant phenotype (Figure 3.11). 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 mutation has a novel phenotype. To exclude the possibility that the CPT resistant
phenotype was generated from unrelated genomic mutations, we reintroduced the 5’ UTR transposon insertion into a wild type background. The resulting new mutants were still resistant to CPT (Figure 3.11, 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. The S. cerevisiae azr1Δ deletion mutant decreased the resistance to DNA damaging reagents and anti-fungal drugs. Overexpression of AZR1 increased resistance, suggesting that the S. pombe 5’ UTR mutant might increase transporter expression and CPT export. Future experiments will directly test whether the 5’ UTR insertion causes an increase in SPBC16A3.17c protein levels.

**Figure 3.11 A 5’ UTR 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 was more CPT resistant than wild type cells. The wild type strain carries a Hermes insertion in the leu2 gene. Please note that all strains are leu- due to a background leu1-32 mutation.
Material and Methods

Construction of barcoded-\textit{Hermes} transposon plasmids

\textit{Hermes} transposon donor plasmid (pHL2577) and transposase plasmid (pHL2578) were from Dr. Henry Levin. We replaced the \textit{LEU2} gene with a \textit{ura4} marker in the backbone of pHL2578 (described in chapter 2).

The 78 bp barcode oligo (5'- /5Phos/TG GCC ACC CGG GCC ANN NAN ANN NANN ANN ANN ANN ANN ANN ANN NAN ANN ANN ANN ANN ANN ANN ANN ANN ANN ANN NAG GGC CAC CCG GGC CGG CGC GCC -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; 1min 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 barcodes were cloned within the \textit{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-\textit{Hermes} transposon plasmids (pHL2577-barcode) were isolated by Plasmid Midi Kit (Qiagen).

Generation of a library of \textit{Hermes} insertion mutants

\textit{S. pombe} KRP201 (\textit{h+}, \textit{ade6-m216}, \textit{leu1-32}, \textit{ura4-D18}) cells were transformed with 1 µg of PHL2578 plasmid and grown on an EMM-ura plate. Frozen competent cells were made as described. pHL2577-barcode plasmids (1µg) were transformed sequentially and plated on EMM+ ade, leu, his, and ura plates for 24 hours then replica plated on YES+G418 (200 µg/ml) +FOA (1g/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: \( P = 1 - (1-f)^N \), where:

\[ f = \frac{1}{\text{Number of barcoded-}\text{Hermes transposon plasmids}} \]

\[ N = \text{number of } S. \text{pombe clones sampled} \]

\[ P = \text{the probability of getting a barcode} \]

\[ 1-P = \text{the probability of not getting a barcode} \]

This calculation predicts a >99% probability of not getting the same barcode if picking 1,000 *S. pombe* colonies from each barcoded-\text{Hermes transposon plasmid} transformation.

**Pooling cells, genomic DNA preparation, fragmentation**

Frozen cells 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. Genomic DNA was extracted from 16 row-pooled, 12 column-pooled and 12 layer-pooled cells. A total of 40 pools of *S. pombe* cells were collected and cultured to saturation in YES+G418+5-FOA liquid. For each pool, \(10^9\) cells from each pool were resuspended 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 down by 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 Msel, Apol or MfeI (NEB) digestions in parallel. The digestion was done at 37°C for 8 h for Msel and MfeI, at 50°C for 8 h for Apol. The reactions were heat inactivated for 10 min at 80°C. The digestions were cleaned up by a PCR Purification Kit (Qiagen). Apol and MfeI digestions were mixed and cleaned up together.

**Ligation-mediated PCR**

The Msel ds linkers were generated by annealing the upper strand oligo (5Phos/TAGTCCCTTAAGCGGAG/3AmM/-amino) to the lower strand oligo (5GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC). Apol and MfeI ds linkers were generated by annealing the upper strand oligo (5Phos/AATTGTCCCTTAAGCGGAG/3AmM/-amino modified) to the lower strand oligo (5GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC). 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 (Figure 3.3). To amplify the *Hermes* transposon right end (HR) insertion sites, the first step was done by the HR outside primer (5GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC) specific to the transposon right end using the linker primer
(5GTAATACGACTCACTATAGGGCTC) 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 (CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGTAATACGACTCACTATAGGGCT) and an 8-bp indexed HR-nested primer (5ACACTCTTTTCCCTACACGACGCTCTTCCGATCTXXXXXXXXTATGTGGCTTACGTTTGCCTGTGG), 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 (5AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) 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, 1 min at 72°C and a final incubation for 5 min at 72°C.

The *Hermes* transposon left end insertion sites were amplified using 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

$5'\text{ACACTCTTTCCCTACACGACGCTCTTCCGATCT}xxxxxxxTATCCCGGGAT$

TTTGGCCAC) and barcode reverse primer

$\text{CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCTGCAGCGAGGAGCCGTAAT}$ 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 (Figure 3.4). Equal molars of products were mixed.

Mapping of Integration sites

Single end sequencing of multiplexed samples was performed on multiple lanes of the Illumina GAIIX or HiScan. Sequence reads were extracted from FASTQ files from the sequencers (Figure 3.12). The raw sequence data was parsed into row, column or layer pools and read by the 8-bp index tags. This was followed by trimming the adaptor sequences. Then the data was further sorted and trimmed by the reads preceding the barcode

(TATCCCGGGATTTTGCCACCCCGGCC), transposon right end

(TATGTGGCTTACGTTTGCCTGTGGCTTGTTGAAGTTCTCTG) or left end
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, then intersected row, column and layer pool reads were assigned to barcodes and integration sites to strains (Figure. 3.12).

**Figure 3.12 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).
Verification of high-throughput sequencing results

Random strains were picked from the Hermes library. To verify Hermes insertion sites, Inverse-PCR (described in chapter 2) was performed on individual mutants and the insertion points were compared to the high-throughput results. 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 if the insertion existed.

To verify the barcode sequences, primer 3829s (CAAGACTAGGAAAAGAGCATAAG) and 4171as (GACTGTCAAGGAGGTATTC) 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 the GO term were spot tested on YES and non-fermentable YEEG (0.5% yeast extract, 2% glycerol, 2% ethanol, 2g/L cosamino acids, amino acids mix) plates. Photos were taken after 5 days. The first spot contained $2 \times 10^4$ cells. The rest were 5-fold dilutions. Growth from defective strains was inoculated in liquid YEEG at $OD_{600} 0.2$ and cultured 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 $3 \times 10^6$ cells. The rest were 5-fold serial dilutions. Photos were taken on the third day or until phenotypes were observed.
Discussion

We have constructed a novel collection of 4,095 sequenced and uniquely barcoded *S. pombe* insertion mutants that are amenable to high-throughput phenotypic analysis for investigation of gene functions. These insertions distributed across 368 essential protein-coding genes, 2,470 non-essential protein-coding genes and 1,159 non-coding RNA genes.

**The Insertion library mutants display a wide range of phenotypes.**

Some *Hermes* mutants revealed different phenotypes compared to the haploid 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 a wider range of phenotypes in our insertion library compared to the deletion library (Figure 3.9, Table 3.3).

Third, one 5' UTR insertion mutant showed the opposite phenotype of CPT resistance. Thus, our new library provides a useful new tool for the analysis of gene function and can reveal new phenotypes.

Construction and analysis of this insertion mutant library demonstrated a method for constructing similar sequenced libraries in other organisms, and also revealed several key points for designing future approaches. 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. How the Hermes insertion library met these criteria is discussed below.

The Hermes transposon has minor target site preferences. We chose the Hermes transposon as a tool to mutagenize the S. pombe genome because Hermes efficiently targets ORF and regions upstream and downstream of ORF, according to the Hermes insertion pattern from 26 samples. Three years later, the Hermes transposon was reported to preferentially insert into nucleosome-free regions in S. cerevisiae. More recently, an intensive study of the Hermes transposon target sites in S. pombe was performed, suggesting in vivo Hermes insertions are influenced by chromatin structures. A map of Hermes transposon in vivo insertions in S. pombe revealed that 33% occurred within ORFs and 67% occurred in the intergenic regions. The insertion levels were higher in the intergenic regions upstream of the ORFs. In contrast, in vitro Hermes insertion reactions using naked DNA and purified transposase revealed a different pattern. Of 1.36 million independent integration events into naked DNA, 63% occurred in ORFs, correlating well with 60.2% of genome is coding
content. The fraction of insertions in the upstream of ORFs was similar to downstream.

*Hermes* transposon prefers to insert nucleosome-free regions in *S. pombe* as in *S. cerevisiae* Lewin et al. mapped nucleosome occupancy of the *S. pombe* chromosome III and overlapped the nucleosome map with in vivo *Hermes* transposon insertion sites. *Hermes* transposon insertions are located at the lowest nucleosome occupancy and at highest frequencies between nucleosomes.

*S. pombe* nucleosome occupancy on approximately 150 nt upstream of transcription start sites (TSSs) is very low. From the start of ORFs, nucleosomes are positioned in phased arrays, which explained the lower in vivo ORF insertion rate (33%) than the in vitro naked DNA ORF insertion rate (63%) As described in Chapter 3, our current *Hermes* transposon collection contains 1,054 (24%) ORF insertions and 3,337 (76%) intergenic region insertions, which are similar to Levin’s in vivo *Hermes* transposon insertion profile.

The current *Hermes* transposon mutant collection is not saturated, with insertions in about 20% of *S. pombe* annotated essential genes, 40% of non-essential genes and 30% of non-coding genes. To make a complete mutant library where every gene is targeted, one could continue to generate and sequence mutants with the same approach to increase the coverage of genes. However, an alternative approach that may be more efficient is to utilize *Hermes* transposon in combination with other elements that could target nucleosome
occupancy regions, similar to the multi-transposon approach used in *Drosophila* library constructions.

Unfortunately, transposons that target nucleosome-bound DNA in *S. pombe* have not been discovered yet. Tf1, a long terminal repeat retrotransposon in *S. pombe* also prefers to insert nucleosome-free regions. In a high-throughput target profiling study, 76% of the Tf1 insertions were distributed in intergenic regions where 31% of the promoters were located. The *S. cerevisiae* retrotransposon Ty1 prefers to target nucleosome-bound DNA. A detailed mapping of Ty1 sites of insertion revealed that Ty1 targets a specific surface of the nucleosome at the H2A/H2B interface. The majority (~90%) of insertions were within the predicted 5' region of Pol III-transcribed genes. It is not known whether Ty1 is active in *S. pombe*, but the element is well understood. Consequently, it may be possible to adapt Ty1 to construct barcode-tagged insertions in the same way we adapted the *Hermes* system. Recently, a rice miniature inverted repeat transposable elements (MITEs) was applied to *S. cerevisiae*. About 65% of the insertions were in genes. MITE is also a potential complement for the *Hermes* transposon in large-scale mutagenesis.

3D cell pooling combined with a deep-sequencing strategy greatly speeds sequencing of a random-insertion mutant library. The rapid advancement of deep-sequencing technology provides a new path to define genome-wide collections of mutated genes at reduced cost and effort, as we and others have shown. Insertion elements such as transposons, retrovirus or retroviral vectors that generate defined boundaries upon integration into the
genome are very suitable to the deep-sequencing system as the individual insertion sites’ sequences can be readily obtained.

The key to efficient sequencing is multiplexing, meaning that every sample is indexed and can be identified after sequencing. For example, the large-scale zebrafish mutant project utilized 6-mer DNA tags arrayed in 96-well format to index each mutant sample. After sequencing, genome sequences flanking the integration could be assigned to each sample by the index tag \(^{13}\). Compared to traditional capillary sequencing of individual samples, this strategy greatly reduced the cost of sequencing by pooling all samples into a single lane of the Illumina sequencing platform.

However, this strategy still required producing thousands of individual samples for genomic DNA preparation, which was very labor intensive. In addition, thousands of pre-synthesized index tags were required. Our strategy of sequencing the *Hermes* transposon library greatly improved this approach by 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. As described in Chapter 3, We designed the strategy based on the Illumina GAIIx platform, which was able to generate 10-20 million reads per lane, which allowed us to process ~2,300 strains (24 plates) at one time. Each product could be sequenced several hundred times to generate a reliable sequence. With continuing improvements in technology, our strategy is becoming more powerful. Our entire mutant collection of ~9,200 strains (96 plates) can be
processed at one time on a newer, higher-throughput platform, such as Illumina Hiseq 2000, which generates ~300 million per lane (Figure 3.13).

Figure 3.13 Increased read depth of the deep-sequencing platform allows a 3D pooling strategy to scale up and process thousands of samples in parallel. (A) On the Illumina GAIIx platform, 24 plates of cells were collected in 16 row pools, 12 column pools and 12 layer pools. A total of 40 indexed primers were required. (B) Using a newer, higher-throughput platform such as Illumina Hiseq 2000, four times as many samples, or the whole *Hermes* transposon library (96 plates), can be processed all at once. Only an additional 24 index primers would be required.

The 3D pooling strategy we developed required the synthesis of multiple indexed primers to tag each row pool, column pool and layer pool to allow the sequences to be analyzed. The strategy also required the development of a customized data analysis pipeline to triangulate the three dimensional row-column-layer data and assign the sequence to each mutant strain, which can be
easily adapted to a large set of sequences. Expanding our approach to a higher-throughput platform will only require a few more index primers (24) to sequence the whole set of the collection (Figure 3.13).

Our pooling and multiplexed sequencing strategy will have important utility in a variety of settings. The main advantages are greatly reduced cost and simple sample preparation that is amenable to scaling. The technique can be readily modified to map any DNA element being inserted into any sequenced genome. In multicellular eukaryotes, a genome-wide mutagenesis method is lacking. Random mutagenesis is a fast and convenient way to create genome-wide mutant collections. Sequencing such libraries by adapting our strategy will only take several months.

**Engineering insertion elements to facilitate future use of the mutants**

Creation of a sequenced mutant library will provide the research community with a resource to perform genome-wide gene function studies. How the insertion element is designed can also greatly facilitate the future utilization of individual mutant strains and the library.

One previous example is the mTn3 transposon library in *S. cerevisiae*. It was designed to generate various mutant alleles from a single insertion. mTn3 insertion generates a *lacZ* gene fusion. Screening β-galactosidase activity identified about 300 new yeast protein-coding genes. Moreover, mTn-insertion alleles can be converted to HA-tagged alleles when introducing Cre recombinase. The epitope-tagged genes will facilitate downstream analysis of the mutant alleles, such as examining protein expression, localization within cell or levels of
proteins by Western blots. Another example is the DNA tags. The DNA barcodes of the *S. pombe* and *S. cerevisiae* gene deletion collections allowed high-throughput analysis of the whole set of mutants in pooled competitive growth assays and phenotype screens\(^{14,232}\).

Our *Hermes* transposon insertion mutants are each tagged with a unique DNA barcode, suitable for high-throughput phenotype screens. Our *Hermes* transposon mutants can be easily modified for more diverse uses. 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 (Figure 3.14), as I did with the CPT resistant mutant (Figure 3.11). Second, the *Hermes*-KanMX insertions can be converted to ORF deletions using a NatMX gene with flanking homology to the genome. NatMX integration will replace the *Hermes*-KanMX marker, so one can screen for Nat resistant and G418 sensitive strains to rapidly identify the deletions (Figure 3.15). Third, mutated genes of interest can be epitope-tagged for analyzing truncated protein functions generated from transposon insertion in coding exons. *Hermes* transposon can be replaced by universal “epitope tag-selective marker” fragments, resulting in a C-terminal in-frame fusion of epitope tag (Figure 3.16). Fourth, conditional alleles can be generated from mutants bearing the *Hermes* transposon within the promoter or 5’ UTR regions. The *Hermes* transposon can be replaced by a universal “selective marker-conditional promoter” fragment. Gene expression can then be “turned on” or “turned off” using inducible promoters, temperature sensitive (*ts*) promoters\(^{233}\) or over-expressed by strong
promoters (Figure 3.17). The conversion allows examination of immediate effects of gene silencing, which may be different from the long-term effects of a null mutation.

**Figure 3.14** *Hermes* transposon insertion mutants can be easily transplanted to any desired 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.

**Figure 3.15** A *Hermes* transposon insertion can be easily converted 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.
Figure 3.16 Epitope tagging of disrupted genes. When the Hermes transposon inserts coding-exons, universal tagging elements containing Hermes transposon homologous arms, a selective marker and an epitope tag can be easily transformed into cells to make a C-terminal tagged strain. The epitope tag can be designed to be in-frame fusion with the protein-coding gene.

Figure 3.17 Conditional alleles generated from Hermes transposon insertions. When the Hermes transposon inserts 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.
Our defined, barcode-tagged mutant library can be applied in high-throughput phenotypic screens in *S. pombe* studies. For example, the collection allows systematic phenotypic screens utilizing the DNA barcodes. The whole set of mutants can be assayed in pooled competitive growth experiments to monitor the survival of individual mutants in response to stresses such as drug treatment. The abundance of an individual mutant in the pooled culture can be represented by its barcode abundance. Following barcode amplification, the abundance of each barcode can be detected via microarray or high-throughput sequencing, enabling quantitative phenotyping of individual strains to be analyzed 14,212,216.

Our library complements the existing Bioneer *S. pombe* ORF deletion library for advancing basic research. Compared to the Bioneer *S. pombe* haploid gene deletion library (Version II), which created null alleles of 3,308 non-essential genes 14, the transposon collection provides a wider variety of mutants and covers essential, and functionally unknown, non-coding RNA genes in addition to the non-essential genes (Table. 3.1). The Bioneer library is very expensive (~$15,000) and limited in its use by extensive legal restrictions. Our library is being transferred to (TransOMIC) for open distribution to the fission yeast research community for the lowest possible price (~$3,500) with no “reach through” restriction on the strains. Therefore, our library will be available to a wider range of *S. pombe* labs world-wide to facilitate biological research.
Chapter 4

General discussion and future directions
4.1 What nucleases open the hairpin in Mre11-mediated NHEJ?

Specific roles of Mre11 in NHEJ of DSBs that require DNA end processing are not clear. We observed that Mre11 nuclease activity is not required for repair of hairpin-capped DNA in order to heal the DSBs in *S. pombe*. These observations raise the question as to what cellular factors open chromosome hairpins for NHEJ repair. One candidate, Ctp1, is recruited by MRN upon DNA damage \(^{113}\). Its ortholog Sae2 is an endonuclease that can open hairpins in vitro \(^{191}\). However, *S. pombe* cells lacking both Ctp1 and Mre11 nuclease activities can still repair *Hermes* transposon excision efficiently. Thus, hairpin opening does not absolutely require ctp1. In mammalian cells, Artemis opens hairpin intermediates formed during V(D)J recombination \(^{122}\), but *S. pombe* lacks a clear Artemis ortholog. The hairpin-resolving factors are also required to process genomic DNA structures such as palindromes and cruciform \(^{234}\). Deficiency in these factors will result in genome instability and cancer development.

4.2 A proposed screen for hairpin-opening nucleases

There are 76 genes under the *S. pombe* Gene Ontology term “endonuclease activity.” Excluding 37 ribonucleases or gene products in mitochondria, the rest, 28 endonuclease genes and 11 retrotransposable elements bearing endonuclease activities, are the most likely candidates for hairpin-opening genes (Table 1). Here I propose a screen for *S. pombe* endonucleases using the *Hermes* transposon excision induced hairpin structure system to search candidate factors (Figure 4.1).
To identify hairpin-opening factors in *S. pombe*, a transposon excision PCR assay can be performed. Following induction of the *Hermes* transposon excision, non-excision events or HR repair will restore the *Hermes* transposon insertion. NHEJ repair will be detected by excision PCR. Cells without hairpin-resolving enzymes or NHEJ proteins are expected to have decreased NHEJ repair events. An initial screen can be started with endonuclease genes listed in Table 1. The system allows a whole genome screen utilizing the gene deletion library.

When the *Hermes* transposon is mobilized from the original insertion site in the genome, a cell’s lack of hairpin-resolving proteins or NHEJ proteins will have decreased NHEJ efficiency, which can be measured using our transposon excision PCR assay. To screen these endonuclease candidates, a universal fragment containing homologous arms and the *Hermes* transposon will be transformed into a *S. pombe* strain lacking an endonuclease candidate gene.
The *Hermes* transposon can be introduced into the endonuclease deletion strains in a high-throughput manner in 96-well format.

**Table 4.1 S. pombe endonuclease candidates for hairpin-resolving screen assay.**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Name</th>
<th>Gene description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPBC13G1.04c</td>
<td>abh1</td>
<td>alkB homolog/2-OG-Fe(II) oxygenase family Abh1 (predicted)</td>
</tr>
<tr>
<td>SPCC622.17</td>
<td>apn1</td>
<td>AP endonuclease Apn1</td>
</tr>
<tr>
<td>SPBC3D6.10</td>
<td>apn2</td>
<td>AP-endonuclease Apn2</td>
</tr>
<tr>
<td>SPCC736.11</td>
<td>ago1</td>
<td>argonaute</td>
</tr>
<tr>
<td>SPCC338.08</td>
<td>ctp1</td>
<td>CtIP-related endonuclease</td>
</tr>
<tr>
<td>SPCC188.13c</td>
<td>dcr1</td>
<td>dicer</td>
</tr>
<tr>
<td>SPAC30D11.07</td>
<td>nth1</td>
<td>DNA endonuclease III</td>
</tr>
<tr>
<td>SPBC4F6.15c</td>
<td>swi10</td>
<td>DNA repair endonuclease Swi10</td>
</tr>
<tr>
<td>SPCC970.01</td>
<td>rad16</td>
<td>DNA repair endonuclease XPF</td>
</tr>
<tr>
<td>SPBC3E7.08c</td>
<td>rad13</td>
<td>DNA repair nuclease Rad13</td>
</tr>
<tr>
<td>SPBC16D10.04c</td>
<td>dna2</td>
<td>DNA replication endonuclease-helicase Dna2</td>
</tr>
<tr>
<td>SPAC17A5.11</td>
<td>rec12</td>
<td>endonuclease Rec12</td>
</tr>
<tr>
<td>SPBC19C7.09c</td>
<td>uve1</td>
<td>endonuclease Uve1</td>
</tr>
<tr>
<td>SPAC3G6.06c</td>
<td>rad2</td>
<td>FEN-1 endonuclease Rad2</td>
</tr>
<tr>
<td>SPAC12G12.16c</td>
<td>SPAC12G12.16c</td>
<td>Fen1 family nuclease, XP-G family (predicted)</td>
</tr>
<tr>
<td>SPAPB1E7.06c</td>
<td>eme1</td>
<td>Holliday junction resolvase subunit Eme1</td>
</tr>
<tr>
<td>SPCC4G3.05c</td>
<td>mus81</td>
<td>Holliday junction resolvase subunit Mus81</td>
</tr>
<tr>
<td>SPAC25G10.02</td>
<td>cce1</td>
<td>mitochondrial cruciform cutting endonuclease Cce1</td>
</tr>
<tr>
<td>SPAC13C5.07</td>
<td>mre11</td>
<td>Mre11 nuclease</td>
</tr>
<tr>
<td>SPBC19F8.04c</td>
<td>SPBC19F8.04c</td>
<td>nuclease</td>
</tr>
<tr>
<td>SPAC139.01c</td>
<td>SPAC139.01c</td>
<td>nuclease, XP-G family (predicted)</td>
</tr>
<tr>
<td>SPCC18B5.06</td>
<td>SPCC18B5.06</td>
<td>peloto ortholog (predicted)</td>
</tr>
<tr>
<td>SPAC167.01</td>
<td>ire1</td>
<td>serine/threonine protein kinase, sensor for unfolded proteins in the ER Ire1</td>
</tr>
<tr>
<td>SPAC11H11.03c</td>
<td>SPAC11H11.03c</td>
<td>SMR and DUF1771 domain protein (predicted)</td>
</tr>
<tr>
<td>SPAP27G11.15</td>
<td>slx1</td>
<td>structure-specific endonuclease catalytic subunit</td>
</tr>
<tr>
<td>SPAC688.06c</td>
<td>slx4</td>
<td>structure-specific endonuclease subunit Slx4</td>
</tr>
<tr>
<td>SPAC688.13</td>
<td>scn1</td>
<td>TatD DNase family Scn1</td>
</tr>
<tr>
<td>SPBC17A3.08</td>
<td>SPBC17A3.08</td>
<td>TatD homolog (predicted)</td>
</tr>
<tr>
<td>SPBC9B6.02c</td>
<td>Tf2-9</td>
<td>retrotransposable element/transposon Tf2-type</td>
</tr>
<tr>
<td>SPAPB15E9.03c</td>
<td>Tf2-5</td>
<td>retrotransposable element/transposon Tf2-type</td>
</tr>
<tr>
<td>SPBC1289.17</td>
<td>Tf2-11</td>
<td>retrotransposable element/transposon Tf2-type</td>
</tr>
</tbody>
</table>
An alternative approach would be a genome-wide genetic screen for hairpin-opening and NHEJ proteins (Figure 4.2). This approach would connect an easily-scored phenotype to a *Hermes* excision event. The *Hermes* transposon is placed between a constitutive expressed promoter (CaMV) and a selectable marker (*his5*<sup>+</sup>)<sup>235</sup>. The *Hermes* insertion will inhibit *his5*<sup>+</sup> expression. If *his5*<sup>+</sup> is weekly expressed in this context, growth can be inhibited by addition of 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of *his5*<sup>+</sup><sup>236</sup> Typical NHEJ will place the CaMV promoter next to *his5*<sup>+</sup> and allow its expression. A mutation that reduces hairpin processing would produce inaccurate end-joining and reduce the number of cells that can express the marker. Our *Hermes* insertion and the gene deletion libraries could be screened by introducing this construct and the transposase plasmid into every strain. NHEJ frequencies would then be assayed qualitatively by inducing the transposase expression and then patching cells onto a selective medium (Figure 4.2). Such a screen would identify the NHEJ components we previously tested (Chapter 2) as well as potential modifiers of hairpin-opening proteins.
Figure 4.2 Genome-wide screen for hairpin-resolving nucleases. The transposase expression is controlled by the \textit{nmt1} promoter, which is repressed by thiamine. The \textit{his5}\textsuperscript{+} expression is controlled by a CaMV1 promoter, which is constitutively expressed. The construct containing the CaMV promoter, \textit{Hermes} transposon (NatMX) and \textit{his5}\textsuperscript{+} ORF is used to monitor the NHEJ efficiency. The \textit{Hermes} transposon is bordered by 10 bp linker sequences (red box). The \textit{his5}\textsuperscript{+} gene cannot be expressed in this construct. When the transposon is excised, the \textit{his5}\textsuperscript{+} ORF is placed right next to the CaMV promoter, enabling expression. To determine NHEJ frequency, transposase plasmid is transformed into wt or mutant cells and plated on EMM+ his medium. A single colony is picked and patched on the same medium, then replicated plate on EMM-his+ thiamine, where transposase expression is turned off and \textit{his5}\textsuperscript{+} expression is turned on. The \textit{his5}\textsuperscript{+} expression represents the NHEJ efficiency, which can be monitored by counting the number of papillae. Cells lacking the NHEJ or hairpin-opening proteins will have a decreased number of papillae.

4.3 Advantages of generating a sequenced library by insertion mutagenesis for a reverse genetic screen.

Although gene deletion is the basic way to study gene functions, the construction of a gene knockout library is a very difficult and demanding effort. Low efficiency of homologous recombination in multicellular organism limits the
application of the gene knockout strategy in a systematic way \(^{237}\). Strategies like RNAi require production of large libraries of double-stranded RNAs. Genome-wide mutagenesis utilizing ZFN, TALEN or CRISPR requires a library of protein components or guide RNAs for each gene locus and is also very challenging. Scaling the RNAi, ZFN, TALEN or CRISPR method to mutagenize thousands of genes in different organisms is difficult. In contrast, highly efficient DNA transposons can be engineered to be active in a broad range of eukaryotes \(^{85,238-240}\) (Table 4.2). With the advancements in deep-sequencing technology and the novel three-dimensional pooling strategy we developed, the construction of a mutant library by random insertion mutagenesis and sequencing each mutant in a high-throughput manner is becoming a very efficient and attractive way to study gene functions systematically in various model organisms.

**Table 4.2 Comparison of tools for genome-wide mutagenesis**

<table>
<thead>
<tr>
<th></th>
<th>Transposon</th>
<th>Homologous Recombination</th>
<th>Viral</th>
<th>TALEN/ZFN</th>
<th>CRISPR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mutation Efficiency</strong></td>
<td>High</td>
<td>Very low</td>
<td>High</td>
<td>Good</td>
<td>High</td>
</tr>
<tr>
<td><strong>Scaling to whole genome</strong></td>
<td>Fast</td>
<td>Very difficult</td>
<td>Fast</td>
<td>Difficult</td>
<td>Difficult</td>
</tr>
<tr>
<td><strong>Site Specific</strong></td>
<td>Random insertion</td>
<td>Yes</td>
<td>Random insertion</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

In addition, random insertion libraries can contain multiple alleles for most protein-coding genes, providing allelic series that could be valuable for functional genome studies and generating models for human diseases. Insertions by various elements such as transposons, proviruses or T-DNA may generate knockout alleles either from coding exons or splice junction insertions. Hypomorphic alleles may be generated from truncated proteins maintaining
partial functions or decreased protein expression when disrupting UTRs and affecting transcription or translation. Hypomorphic mutants are valuable. For example, hypomorphic daf-2 mutants of C. elegans exhibit a long-lived phenotype $^{129,241}$. In contrast, severe mutations of daf-2 cause larvae to arrest at the dauer larval stage $^{242}$. Moreover, many human genetic diseases are hypomorphic mutations $^{243-246}$. Extensive literature in C. elegans $^{247}$, Drosophila $^{248,249}$ and mice $^{244,250,251}$ demonstrate that hypomorphic alleles are often more revealing than a null allele when analyzing gene function.

Random insertion libraries have the additional advantage of disrupting intergenic regions that include a large number of non-coding genes, regulatory elements such as promoters, enhancers, etc… that are not included in the gene deletion collection, RNAi, TALEN or CRISPR libraries. Non-coding RNAs have important roles in gene expression and genome maintenance, which are becoming an intense area of investigation. The ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) have central roles in protein synthesis and chromatin organization. The small nuclear RNAs (snRNA) are components in splicosome. Micro RNAs (miRNA) are important in post-transcriptional regulation of gene expression. Long non-coding RNAs (lncRNA) assist chromosome looping and modulation of histone modifications $^{252}$. Insertions in these non-coding genes will extend the understanding of their biological roles. The novel methods developed in this work will facilitate the extension of these benefits to organisms beyond yeast.
4.4 Future application of barcode-tagged Hermes transposon insertion mutant library in genome-wide screens

Application of the Hermes transposon insertion library in genome-wide aging studies The identification of new and uncharacterized genes affecting a phenotype is one of the major uses of the collection. We plan to utilize the Hermes transposon collection to perform a genome-wide identification of aging regulators.

Aging is under genetic control. Some genes are critical to maintain a normal lifespan. LMNA gene mutation causes Hutchinson-Gilford progeria syndrome, which shows rapid aging. Werner syndrome, another progeria syndrome, is caused by autosomal recessive homozygous or compound heterozygous mutations in the RECQL2 gene. Genes that increase lifespan also exist. The heterozygous mutation in the IGF1-receptor gene is associated with human longevity. Mice are long-lived when the growth hormone receptor gene is knocked out. Yeast cells live longer when the sch9 gene (AKT ortholog) is deleted. However, the actual aging mechanisms are still not clear. A genome-wide screening of pathways and regulators that control aging will provide important mechanistic information and identify drug targets to increase lifespan or lower the prevalence of age-related diseases.

Yeast has been used as a tool to study aging mechanisms because of its powerful molecular genetics and relatively short lifespan. In S. cerevisiae, the mechanisms of aging are investigated by using two distinct kinds of studies. Replicative life span (RLS) refers to the number of times a single cell can divide.
Chronological life span (CLS) measures the time a population of cells can survive in stationary phase. RLS and CLS aging are controlled independently. *S. pombe* is an emerging model for CLS aging studies because *S. pombe* has many features more similar to multicellular eukaryotes than *S. cerevisiae*.

Our lab has previously developed a robust CLS assay in *S. pombe* that recapitulates the evolutionarily conserved features of aging (Figure 4.3). In a single cell culture, lifespan of a mutant strain is identified directly by CLS assay. Long-lived or short-lived mutants can be classified by their survival curves. However, in a population of random mutant cells, it is very hard to tell which mutants survived in the aged cultures and are long-lived.

The DNA barcodes in our *Hermes* mutants will greatly facilitate the identification of longevity genes. In a population of random mutants, long-lived mutants are only in a small portion. When the culture ages, normal lifespan mutants and short-lived mutants gradually die, and the frequency or proportion of their barcodes will decrease. In contrast, the extended survival of the long-lived mutants means that the frequency of their barcodes will gradually increase (Figure 4.4). Because the relative abundance of mutants is represented by their barcode abundance, the frequency changes will be reflected by the barcode of each mutant. Enrichment of long-lived mutants will be reflected by the enrichment of their barcodes.
Figure 4.3 *S. pombe* CLS assay. To measure the chronological lifespan of a fission yeast strain, the cells are cultured to saturation. Cells are then sampled everyday to monitor their viability by plating them on a fresh, rich medium until the number of viable cells/ml drops to less than 10 cells per ml culture. Lifespan curves are generated by plotting the viable cells/ml for each day. 128,259
Time 0- Day 0 of CLS assay. Determine starting barcode proportion of each strain.
Time 1- Around day 12.
Time 2- Around day 16.
Time 3- Viable cell is less than 1000/ml (0.01% of total number of cells).

B

Seed culture of Hermes transposon mutant library and grow to saturation

Sample cells at regular intervals to monitor cell viability by ability to grow on fresh medium

Prepare libraries for high-throughput sequencing of barcodes at different time points

C

Normal lifespan mutants
Short-lived mutants
Long-lived mutants
Figure 4.4 Identification of longevity genes by screening the *Hermes* transposon library using the CLS assay and high-throughput barcode sequencing. (A,B) To screen the *Hermes* library for long-lived mutants, CLS assay is performed using the whole set of mutants. To monitor the frequency changes of the three types of mutants, the cultures are sampled at multiple time points. DNA barcodes will be amplified from an individual time point and sequenced on the deep-sequencing platform. Barcode counts from sequencing will reflect the enrichment of the corresponding mutant at that time point. The day 0 sample will determine the starting frequency of all mutants. The normal lifespan of a *S. pombe* cell is about 15 days. Sampling on day 15 is expected to capture long-lived mutants. Sampling when survival is less than 0.01% of the day 0 culture is expected to capture extremely long-lived ones. (C) By calculating the barcode counts of each mutant log₂ (Timeₙ/Time₀), the frequency of short-lived and normal-lifespan mutants is expected to decrease overtime (blue and black lines), while the long-lifespan mutants increase (green line).

Our lab previously conducted such a screen with an unsequenced barcoded insertion library where the insertion sites and barcode sequences were unknown. This approach was successful and identified a new pathway that regulates lifespan and autophagy (a process that recycles cellular components). Therefore, this screening approach has been shown to be feasible with our *S. pombe* aging assay. However, use of this library was labor-intensive and would have to be repeated with every screen. This new barcoded *Hermes* insertion library will transform these projects into high-throughput experiments where multiple time points and conditions can be easily assayed by a next-generation sequencing platform and the mutations can be quickly identified.

**Application of the *Hermes* transposon insertion library in a genome-wide screen** The barcode-tagged mutant collection is also useful in screening for
gene-environment interaction functions such as the identification of drug targets. To screen on a larger scale, competitive growth in pools of multiple strains could be used. The abundance of each strain in the pool would be assayed by monitoring the barcodes using deep-sequencing, providing quantitative phenotyping via barcode counts of individual mutants\textsuperscript{216}. Deep sequencing allows multiplexed samples to be mixed and sequenced in a single lane. Samples from different treatments and different time points can be indexed and sequenced together to further reduce cost\textsuperscript{216}.

An alternative to competitive fitness in a pooled culture is to use the Hermes transposon collection to direct assay the phenotype of individual mutants on plates. For example, one could screen the collection for sensitivity to a DNA-damaging agent, and strains with growth deficiency can be easily identified on the drug plates. Because the collection is sequenced and each strain has a known location, the DNA-damaging agent targeted genes can be rapidly identified. While this method does not require bioinformatic support, it may be more challenging to quantitify the growth of treated and untreated cells. The barcode sequencing approach has the potential to be more quantitative.

4.5 Future application of Hermes transposon insertion mutagenesis in disease-gene discovery

Genome-wide transposon insertion mutagenesis has been utilized for cancer gene discovery\textsuperscript{240,260}. In higher organisms, genetic screens have been hampered by lack of efficient insertion mutagenesis tools. DNA transposons are usually inactivated in vertebrate genomes. Sleeping Beauty (SB), a TC1/mariner
transposon, and *PiggyBac* (PB), from the cabbage looper *moth Trichoplusia ni*, have been engineered to be highly active in mammalian cells, including mouse somatic and embryonic stem cells \(^{240,260}\) (Figure 4.5). An engineered transposon integrated into a mouse genome can stimulate the expression of an oncogene or inactivate the expression of a tumor suppressor gene. Application of these transposons in mouse somatic cells to induce cancer has led to the discovery of a large collection of known and new candidate cancer genes and signaling pathways that can drive tumor formation \(^{240,260}\).

**Figure 4.5 Engineered sleeping beauty transposon is active in mammalian cells.** This transposon contains splice acceptors (SA), splice acceptors (SD) and a bi-directional poly (A) and can terminate transcriptions. It also contains a murine stem cell virus (MSCV) which can promote gene expression.\(^{240}\)

A SB transposon, T2/Onc2 is introduced into the mouse germ line by microinjection to establish transgenic mouse lines. To mobilize T2/Onc2 and increase mutagenesis, a genetically enhanced transposase - SB11 has been knocked in the mouse Rosa26 locus (Rosa26–SB11) - allows transposase to widely express during development and in adult tissues. RosaSB mice are then crossed to each T2/Onc2 transgenic line to induce transposition and generate a cohort of mice bearing various transposition events. The double-transgenic mice are tumor-prone. An analysis of transposon integrations identified 7 previously
validated genes in human cancer, 7 previously validated genes in mouse leukemia and 6 novel cancer genes. The transposition and integration into the mouse genome of modified PB by PB transposase introduces random mutations in mouse somatic cells. Analysis of 14 mouse lines and nearly 900 progeny uncovered many cancer genes not identified in previous Sleeping Beauty screens. Thus, transposon mutagenesis allows rapid identification of cancer genes and provided new models to understand tumor development in mice.

The identification of different cancer genes by different transposon element reflects that PiggyBac and Sleeping Beauty have different integration preferences. It will be interesting to adapt the Hermes transposon for use in mammals. The Hermes transposon is likely to have a different target site preference from PiggyBac and Sleeping Beauty. It may also increase the types of mutations that can be isolated and novel phenotypes that can be identified in this genetic approach.
Bibliography


(75) large-scale screening for targeted knockouts in the Caenorhabditis elegans genome. *G3 (Bethesda)* 2012, 2, 1415-25.


(175) Baran, G.; Echt, C.; Bureau, T.; Wessler, S.: Molecular analysis of the maize wx-B3 allele indicates that precise excision of the transposable Ac element is rare. Genetics 1992, 130, 377-84.


181


(230) Bridier-Nahmias, A.; Lesage, P.: Two large-scale analyses of Ty1 LTR-retrotransposon de novo insertion events indicate that Ty1 targets nucleosomal DNA near the H2A/H2B interface. Mobile DNA 2012, 3, 22.


