Integration Profiling of Gene Function With Dense Maps of Transposon Integration

Yabin Guo,*,[1] Jung Min Park,*,[1] Bowen Cui,[†] Elizabeth Humes,* Sunil Gangadharan,[‡] Stevephen Hung,* Peter C. FitzGerald,[§] Kwang-Lae Hoe, Shiv I. S. Grewal,[†] Nancy L. Craig,[‡] and Henry L. Levin*,[2]**

*Section on Eukaryotic Transposable Elements, Program in Cellular Regulation and Metabolism, Eunice Kennedy Shriver National Institute of Child Health and Human Development, [†]Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, and [§]Genome Analysis Unit, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, [‡]Howard Hughes Medical Institute and Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, and **Department of New Drug Discovery and Development, Chungnam National University, Yusong, Daejeon 305–764, Republic of Korea

ABSTRACT Understanding how complex networks of genes integrate to produce dividing cells is an important goal that is limited by the difficulty in defining the function of individual genes. Current resources for the systematic identification of gene function such as siRNA libraries and collections of deletion strains are costly and organism specific. We describe here integration profiling, a novel approach to identify the function of eukaryotic genes based upon dense maps of transposon integration. As a proof of concept, we used the transposon Hermes to generate a library of 360,513 insertions in the genome of *Schizosaccharomyces pombe*. On average, we obtained one insertion for every 29 bp of the genome. Hermes integrated more often into nucleosome free sites and 33% of the insertions occurred in ORFs. We found that ORFs with low integration densities successfully identified the genes that are essential for cell division. Importantly, the nonessential ORFs with intermediate levels of insertion correlated with the nonessential genes that have functions required for colonies to reach full size. This finding indicates that integration profiles can measure the contribution of nonessential genes to cell division. While integration profiling succeeded in identifying genes necessary for propagation, it also has the potential to identify genes important for many other functions such as DNA repair, stress response, and meiosis.

THE accelerated rate of gene discovery in an increasing number of species has challenged the existing methods for determining the functions of genes. Traditional approaches for characterizing the function of genes rely on obtaining mutant alleles and testing them in individual experiments for phenotypes. Direct and systematic methods for evaluating gene function have been developed. Genome-wide RNAi screens of cultured cells require the synthesis, validation, and refinement of large libraries of double stranded RNAs or vectors that express double stranded RNAs. While RNAi screens have successfully identified many genes that may contribute to key functions such as the replication of human immunodeficiency virus (Brass *et al.* 2008; Konig *et al.* 2008; Zhou *et al.* 2008), the production of RNAi libraries is resource intensive and substantial complications exist, such as off-target effects and incomplete mRNA knockdown.

An alternate approach for characterizing gene function in haploid cells relies on targeted gene deletions. For *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, collections of strains have been created that contain systematic deletions of the predicted coding sequences (Winzeler *et al.* 1999; Kim *et al.* 2010). These comprehensive collections of strains can be readily screened under a variety of conditions to probe the contributions of individual genes to specific processes. However, considerable effort and resources are required to generate deletion sets and once generated, it is difficult to study the deletions in combination with other mutations. Another limitation with deletion collections is the rate of erroneous deletions associated with generating strains *en masse*. A significant proportion of strains can

doi: 10.1534/genetics.113.152744
Manuscript received April 30, 2013; accepted for publication July 20, 2013
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152744/-/DC1.
The raw sequences for the *in vitro* and *in vivo* inserts were deposited in the SRA database of GenBank under accession no. SRA043841.1.
[1]These authors contributed equally to this work.
[2]Corresponding author: NIH, 18 Library Dr., Room 106, Bethesda, MD 20892. E-mail: henry_levin@nih.gov

retain copies of genes that were targeted for deletion (Hughes *et al.* 2000). Finally, information from deletion collections cannot address the function of the ORFs and noncoding RNAs that have yet to be discovered.

Here we describe integration profiling, a simple transposon-based method capable of directly probing the function of single copy sequences throughout the genome of haploid eukaryotes. Integration profiling recently became feasible with the availability of deep sequencing and the discovery of highly efficient DNA transposons that are active in a broad spectrum of eukaryotic organisms. With transposons that readily disrupt ORFs, and sequencing technology that can position high numbers of insertions, the analysis of a single culture can measure the contribution of each gene to cell division. As a proof of principle we describe a study with 360,513 independent insertions of the Hermes transposon in the genome of *S. pombe*. As a control for the properties of the Hermes transposase, we analyzed 1,362,743 insertions generated *in vitro* with purified Hermes transposase and naked *S. pombe* DNA. The *in vivo* integration achieved an average of one insertion for every 29 nucleotides of nonrepetitive genome and these insertions favored nucleosome free regions. However, 33% of the insertions disrupted ORFs and those with lower levels of integration correlated well with the genes reported to be essential while the ORFs with higher integration densities corresponded with the genes reported to be nonessential (Kim *et al.* 2010). Importantly, ORFs with intermediate densities of insertion correlated with genes that, while not essential, do contribute to growth. In addition, discrepancies between the genes reported to be nonessential by the deletion consortium and our integration data revealed that ~10% of the deletion strains retain a copy of the ORFs targeted for deletion.

Materials and Methods

Yeast strains

The *S. pombe* strains used for this study are listed in Table S5. The *S. pombe* strain YHL912 (*h-, leu1-32, ura4-294*) was transformed with the donor (pHL2577) and expression plasmids (pHL2578) separately by lithium acetate method. The donor plasmid was transformed into YHL912 first to create the strain YHL9451. This strain was then transformed with the expression plasmid to create YHL9609, which was used for transposition in liquid cultures. An empty expression vector (Rep81x) was introduced in place of the expression plasmid to create YHL9176 as a negative control for transposition.

Heterozygous diploid deletion *S. pombe* strains were purchased from the Bioneer Corporation and Korea Research Institute of Biotechnology and Bioscience (http://pombe.bioneer.co.kr/). Strains carrying a deletion in *mmf1*, *mrpl16*, *mrpl19*, or *SPBC2d10.08c* were transformed with a plasmid copy of the deleted gene (see Table S6) by lithium acetate method, followed by a second transformation with a plasmid that allows for sporulation (pHL2806). The haploid deletion strains used in this study were from the Bioneer collection version 2.0.

Plasmid constructs

The donor plasmid (pHL2577) and expression plasmid (pHL2578) have been previously described and are listed in Table S6. The *Ura3*-marked donor plasmid consists of the *KanMX6* gene cloned between the Hermes terminal inverted repeats (TIRs). The Leu-marked expression plasmid contains the Hermes transposase gene driven by the Rep81x *nmt1* promoter. A detailed description of plasmids used in this work is included in Supporting Information. The oligonucleotides used in this study are listed in Table S7.

Media

Strains were grown in liquid Edinburgh minimal media (EMM) supplemented with 2 g/liter dropout powder (all amino acids in equal weights, leaving out leucine and uracil, plus adenine in 2.5 times more weight) (Forsburg and Rhind 2006). To eliminate the Hermes donor plasmid, the final cultures were diluted in EMM supplemented with uracil, leucine, 5-fluoroorotic acid (FOA), and B1 (at 50 μg/ml, 225 μg/ml, 1 mg/ml, and 10 μM, respectively). This culture was then diluted in YES media (YE plus complete dropout mix consisting of all the amino acids) supplemented with FOA and G418 (at 1 mg/ml and 500 μg/ml, respectively) to isolate cells with insertion events.

Transposition frequency was measured by plating cells on 2% agar plates containing EMM and dropout mix supplemented with leucine (250 μg/ml), uracil (50 μg/ml), 5-FOA (1 mg/ml), and 10 μM thiamine, and also plating on YES plates supplemented with FOA and G418 at the concentrations noted above.

Pombe Minimal (PM) media used for drop assays is identical to EMM, except with 3.75 g/liter glutamic acid substituted for 5 g/liter NH_4Cl.

Generating a library of Hermes insertions

A culture of YHL9609 was grown to $OD_{600} = 5$, and was used to inoculate a series of cultures at $OD_{600} = 0.05$. The serial passaged cultures were continued until the percentage of cells with integration reached 13%, or ~80 generations of cell division. In all, it took a series of 12 sequential cultures for the strain to reach this point.

The final 50-ml culture was grown in FOA to select against the donor plasmid and then in G418 to select cells with Hermes insertions.

The protocols for preparing Hermes insertion libraries for high-throughput Illumina sequencing are extensive and are included in Supporting Information. The process includes the extraction of genomic DNA, the digestion of the DNA with *Mse*I, the ligation of linkers to the *Mse*I fragments, the PCR amplification of insertion sites, the gel purification of the PCR products, and the Illumina sequencing of the amplified

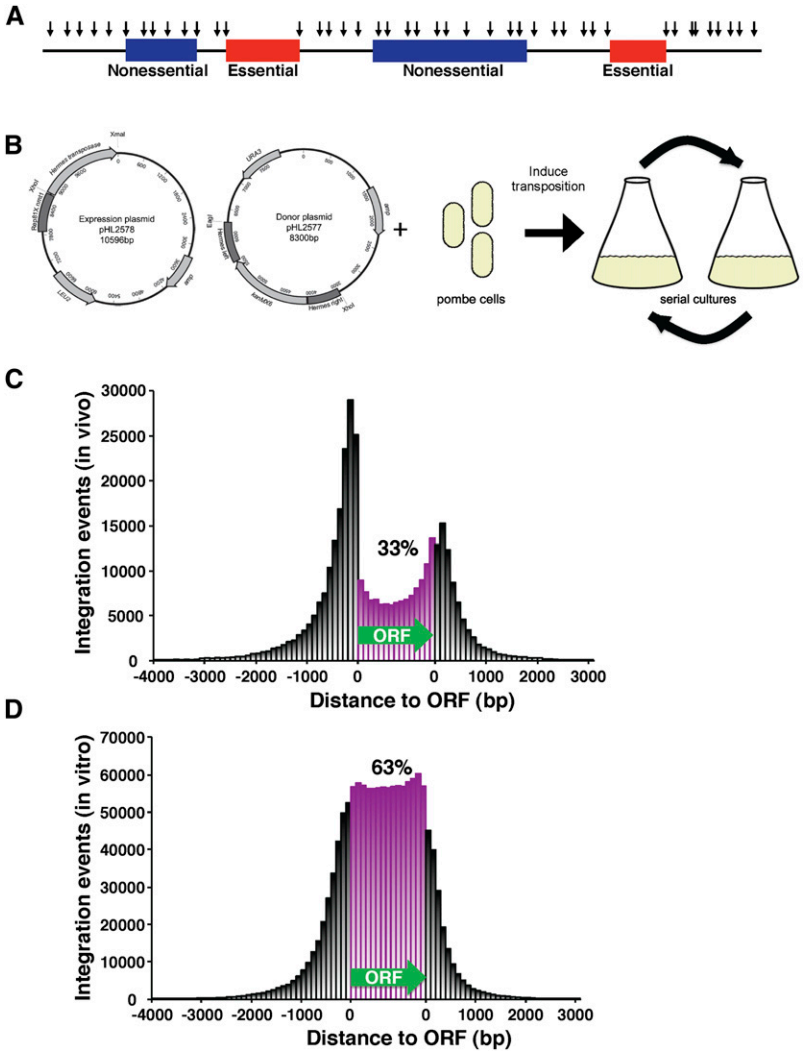

Figure 1 Integration profiling of *S. pombe* resulted in 360,513 independent insertions. (A) With integration profiling, genes required for cell division are predicted to have many fewer insertions than genes that are not required for propagation. (B) Transposition was induced in cells that contained one plasmid expressing the Hermes transposase and a second plasmid that contained the Hermes TIRs flanking *kanMX6*. Cells induced for transposition were propagated for many generations in serial cultures. (C) The positions of the integrations generated *in vivo* were plotted relative to ORFs. Each ORF was divided into 15 equal-sized segments and the number of insertions in each segment was displayed. Integration sites in intergenic regions closer to the 5′ end of an ORF were plotted upstream of the ORF in 100-bp intervals. The insertions closer to the 3′ end of an ORF were plotted downstream of the ORF. (D) The position of the insertions generated *in vitro* were plotted relative to ORFs as in C.

DNA. The raw sequences for the *in vitro* and *in vivo* inserts were deposited in the Sequence Read Archive (SRA) database of GenBank with accession no. SRA043841.1.

Mapping Hermes integration sites on the genome of S. pombe

Sequence reads from Illumina were screened for those containing Hermes left end. Then the Hermes sequences were trimmed. The trimmed sequences were aligned to the *S. pombe* genome using the National Center for Biotechnology Information (NCBI) BLAST software (blastall) on a local computer. The BLAST results were filtered to collect matches with genomic sequence that started from the first nucleotide after the Hermes end and with identities ≥95% and expect (*e*) values ≤0.05. Then, of the matches that met these criteria, the one with the highest bit score was used to obtain the coordinates for the unique insertion sites (Table S4). Sequences that were found to have the same insertion coordinate and the same orientation were considered to be duplicate reads and were counted as only one independent integration event.

Matched random control

For each Hermes insertion site, the distance between the integration site and the responsible *Mse*I site (d) was calculated. Then another *Mse*I site coordinate (m) was randomly chosen from the *S. pombe* genome. Then m + d or m − d was taken as a matched random control (MRC) site. To "add" or "subtract" was also randomly determined. Thus, the MRC dataset has the same size as the experimental integration dataset and matches the distances to *Mse*I sites.

General bioinformatic analysis and programming

The scripts for screening the raw sequences, filtering the BLAST outputs, extracting features from .embl files, determining the locations of Hermes insertion according to chromosomal features, generating MRC and other analyses were written in Perl or Ruby programming languages.

Nucleosome DNA preparation and sequencing

Nucleosomal DNA was prepared as described previously (Yamane *et al.* 2011), and DNA samples were sequenced by

Table 1 Distribution of Hermes insertions in *Schizosaccharomyces pombe*

	Total insertions	Upstream of ORFs (%)	In ORFs (%)	Downstream of ORFs (%)
In vivo	360,513	165,402 (45.9)	117,587 (32.6)	77,524 (21.5)
In vitro	1,362,743	313,326 (23.0)	858,763 (63.0)	190,654 (14.0)

Insertions were designated as upstream of an ORF if the insert is closer to the 5′ end of an ORF than to the 3′ end. They are designated as downstream of an ORF if the insert is closest to the 3′ end of an ORF.

using the Illumina sequencing protocol. Bowtie (Langmead *et al.* 2009) was used to map the Illumina sequencing reads, trimmed to 25 bp of high quality reads, against the reference genome allowing for two mismatches. The mapped data were filtered to remove all sequences that mapped to more than a single location. The end positions of the reads were aligned relative to the center of the nucleosome by shifting the plus strand reads by +73 bp and the minus strand reads by −73 bp. The final nucleosome maps were produced by applying gaussian smoothing to the raw data to reduce noise. More detailed methodology will be described in a forthcoming manuscript on the genome-wide mapping of nucleosomes in *S. pombe*.

Essentiality test

To test the viability of four representative Bioneer strains heterozygous for deletions in nuclear-encoded mitochondrial genes designated as essential, we performed a plasmid shuffle assay. The details of this process are included in Supporting Information.

Results

Integration profiling is a method developed to determine which genes are essential for cell division. With this method, populations of cells with transposon insertions are grown for many generations. The culture becomes depleted of cells that have insertions in genes important for division. Insertion sites in genomic DNA from cells in the culture are amplified by ligation-mediated PCR and the location of the insertions are determined by deep sequencing. The integration profiles of the cultures were expected to contain high densities of insertions throughout the genome except in genes that are required for propagation. Figure 1A illustrates a model of the integration pattern expected in an integration profile.

The Hermes transposon of the housefly, *Musca domestica*, has high levels of integration activity in *S. pombe* and as much as 50% of the insertions occur in ORFs (Evertts *et al.* 2007; Park *et al.* 2009). We therefore applied the Hermes system as previously described to create an integration profile of the *S. pombe* genome. Briefly, the Hermes transposase was expressed from a plasmid in cells that also contained a plasmid copy of the Hermes TIRs flanking a *kanMX6* gene, which confers resistance to G418 (Figure 1B) (Evertts *et al.* 2007; Park *et al.* 2009). The transposase excises *kanMX6* with the TIRs and inserts this DNA into chromosomal sequences of *S. pombe*. Cells with insertions are selected using media containing G418 and the majority of G418-resistant cells have a single insertion (Evertts *et al.* 2007). To create sufficient numbers of insertions and to allow for selection of the fittest cells, the cultures expressing the transposase were grown with repeated dilutions for a total of 74 generations (Figure 1B). Quantitative measures of transposition revealed that 13.4% of the cells in the final culture contained an integrated copy of Hermes (Supporting Information, Figure S1). In the absence of transposase, no insertions were detected. The positions of the insertions were determined by ligation-mediate PCR followed by deep sequencing (Guo and Levin 2010) (and Materials and Methods section in File S1).

The data from sequencing generated 46 million high-quality sequence reads that were 50 nt in length. The first 10 nt of the reads contained the end of Hermes and the remaining 40 nt were derived from the insertion sites in chromosomal DNA. Using the homology algorithm BLAST, 27.5 million reads were matched to a unique sequence in the *S. pombe* genome. Because PCR amplification or cell division can create multiple copies of the same integration event, duplicate reads were omitted. The remaining data identified 360,513 independent insertions and these are the sites that we analyzed (Table 1). On average, this amount of integration constituted one insertion for every 29 nt of nonrepetitive sequence in the genome.

The insertion sites favor nucleosome-depleted positions

The integration activity of Hermes in *S. cerevisiae* was recently described (Gangadharan *et al.* 2010). In *S. cerevisiae*, 41% of the insertions were in ORFs and of the 59% that occurred in the intergenic regions, sites upstream of the ORFs were favored. This pattern of integration correlated with nucleosome-free regions and was attributed to greater DNA accessibility at these positions (Gangadharan *et al.* 2010).

To determine which properties of integration are intrinsic to the Hermes transposon, we compared the distribution of Hermes integration in *S. pombe* to that reported in *S. cerevisiae*. A map of the insertions in *S. pombe* revealed that 33% occurred within ORFs (Table 1 and Figure 1C). The insertion levels were higher in the intergenic regions than in the ORFs and the insertions in intergenic regions of *S. pombe* exhibited the same preference for sequences upstream of the ORFs that occurred in *S. cerevisiae* (Table 1).

To test whether the higher levels of integration in the intergenic sequences were caused solely by sequence preferences of the transposase, we generated a library of insertions

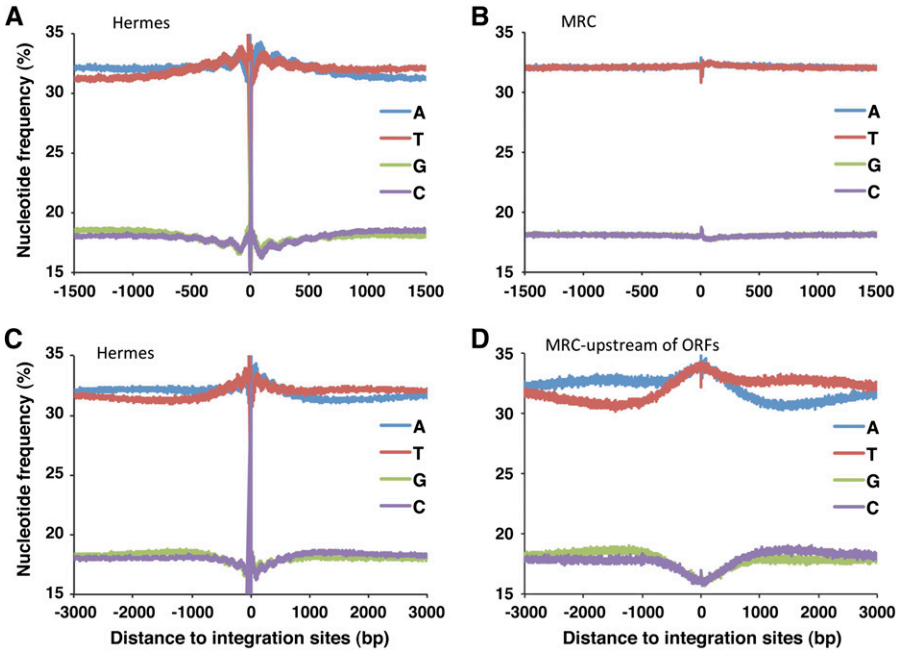

Figure 2 The sites of Hermes integration have a two-component pattern of nucleotide preferences that extends 3 kb from the insertion sites. (A) The *in vivo* insertion sites were aligned and the average nucleotide frequencies were calculated for a window that extended 1.5 kb on either side of the integration sites. (B) The average nucleotide frequencies were calculated for the MRC, the matched set of random insertions. (C) The average nucleotide frequencies were calculated for Hermes insertions in a window that extended 3.0 kb on either side of the insertion sites. (D) The average nucleotide frequencies were plotted for just the MRC sites that were upstream of ORFs.

in naked DNA using purified Hermes transposase in an *in vitro* reaction. Deep sequencing of the *in vitro* insertions identified 1.36 million independent integration events in *S. pombe* DNA (Table 1). Sixty-three percent of these inserts occurred in ORFs and the fraction of insertions in the intergenic sequences upstream of ORFs was similar to the insertions downstream (Figure 1D and Table 1). This distribution correlated well with the coding content of the *S. pombe* genome, which is 60.2% (Wood *et al.* 2002) and suggests in the *in vivo* experiment the integration preference for sequences upstream of ORFs was due to features of chromatin structure.

To characterize the sequence preferences of Hermes integration in *S. pombe*, we analyzed the nucleotide frequencies at the insertion sites. The Hermes transposase recognizes the consensus sequence nTnnnnAn (Zhou *et al.* 2004) and as expected, the vast majority of the insertions in *S. pombe* had this sequence (Figure S2A). The same nTnnnnAn was found at the sites of Hermes integration generated *in vitro* (Figure S2B).

Many transposons and retroviruses exhibit nucleotide preferences in the form of palindromes thought to result from contacts between the integrases and nucleotides that extend as far as 50 nt to 100 nt from the sites of insertion (Holman and Coffin 2005; Wang *et al.* 2007; Gangadharan *et al.* 2010; Guo and Levin 2010; Maertens *et al.* 2010). The nucleotide frequencies surrounding the insertions we generated possessed an unusual palindromic pattern of nucleotide frequencies that extended >1.5 kb from the insertion sites (Figure 2A). The sequences had a surprising oscillation of nucleotide frequencies with a wavelength of 150 bp that extended ~500 bp from the insertion sites. These oscillating patterns of nucleotide frequencies were not observed in a set of random insertions called a MRC generated *in silico* that matches the distances to *Mse*I sites of the authentic insertions (Figure 2B). Nor were they observed in the *in vitro* experiment (Figure S3). As a result, these oscillating preferences for A's, T's, G's, and C's reflect the *in vivo* context of the insertion sites and not a bias generated by the procedure used to cut the genomic DNA or contacts the transposase makes with the target DNA. The 150-bp repeat patterns suggested that the Hermes transposase was influenced by nucleosomes during integration.

Extending our analysis of nucleotide frequencies further from the insertion sites revealed a second bias consisting of a single palindromic depression in A/T content that continued a full 3 kb from the insertion sites (Figure 2C). This pattern was also absent in the MRC (Figure 2B) and in the *in vitro* integration data (Figure S3). Interestingly, the 3-kb nucleotide bias was stronger when just the insertions upstream of ORFs were analyzed (Figure S4). To test whether the 3-kb depression in A/T frequency resulted from a property of sequence upstream of ORFs, we analyzed the nucleotide frequencies of the subset of random insertions (MRC) that occurred upstream of ORFs (Figure 2D). The random insertions upstream of ORFs did possess a strong A/T depression that was very similar in magnitude and extent to the upstream insertions of Hermes (Figure S4). Consistent with the finding that the A/T depression was an intrinsic property of sequence upstream of ORFs, we found that the average nucleotide frequencies upstream and downstream of the *S. pombe* ORFs do have a lower frequency of A relative to T (Figure S5, A and B). This is an unusual genome structure that to our knowledge has not been previously reported for *S. pombe* and is not as extensive in *S. cerevisiae* (Figure S6, A and B).

The oscillating pattern of A's, T's, G's, and C's surrounding the Hermes insertions was not observed with the set of random insertions positioned upstream of ORFs, suggesting

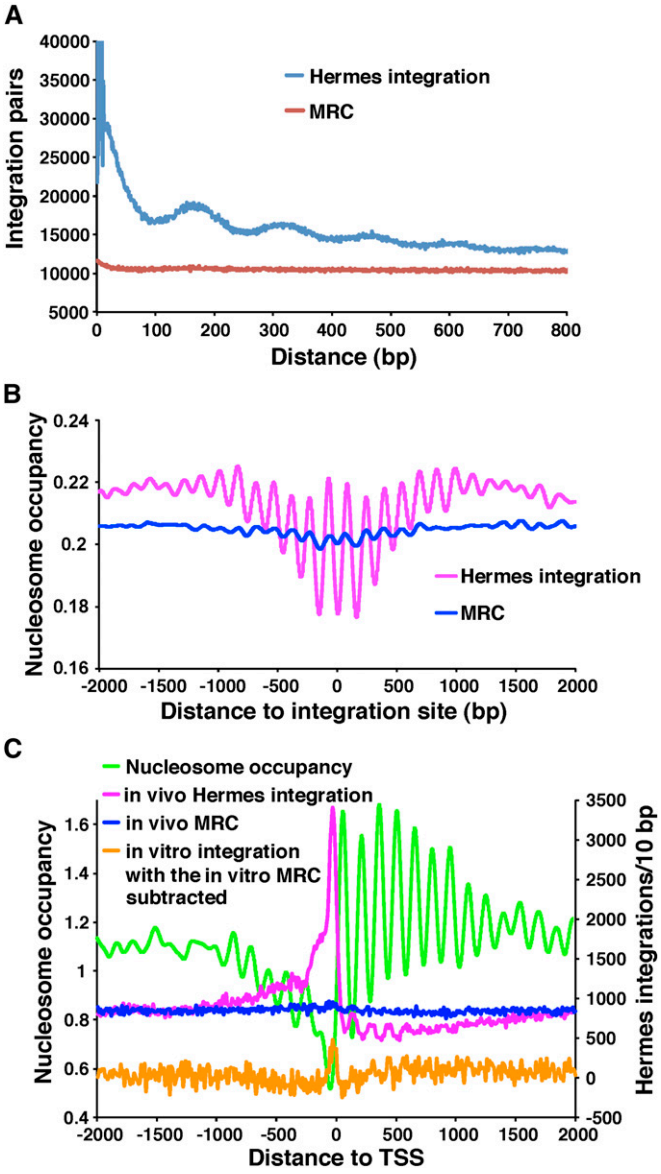

Figure 3 Hermes integration occurs at sequences with low nucleosome occupancy. (A) The distances between all pairs of Hermes insertions were calculated and the distribution of these distances was plotted for the pairs that were <800 bp apart. Also shown are the distances between insertions for the MRC sites created to match the *in vivo* insertions. (B) The insertion sites in chromosome 3 were aligned and the average nucleosome occupancies were plotted for each position in a window that extended 2 kb on both sides. (C) The pol II transcribed genes on chromosome 3 were aligned at the start sites of transcription. The numbers of *in vivo* insertions (pink) in 10-bp intervals were plotted for a 4-kb window. The MRC insertions matched to the *in vivo* inserts are shown in blue. Shown in orange are the *in vitro* insertion numbers minus the MRC values calculated for the *in vitro* data. The average nucleosome occupancies are shown in green.

that this 150-nucleotide periodicity was mediated by regularly spaced proteins such as nucleosomes. If this insertion bias were due to integration into arrays of phased nucleosomes then the closest distances between insertions should be multiples of ~150 bp. An analysis of all the pairs of insertions closer than 800 bp revealed that the most common distances were indeed multiples of 150 bp (Figure 3A). Such a pattern was not seen with the random controls. To test directly whether integration was influenced by nucleosomes, we generated a map of nucleosome occupancy for chromosome 3 (Methods section in File S1). An analysis of the average nucleosome occupancy extending 2 kb on either side of the insertions created *in vivo* revealed a highly regular oscillation with a wavelength of 150 bp (Figure 3B). The insertion sites were located at the lowest nucleosome occupancy, indicating the highest frequencies of integration were between nucleosomes. The MRC sites did exhibit oscillations but the amplitudes were substantially smaller than the *in vivo* insertions, indicating that the positions of *Mse*I sites did not contribute to the strong integration bias that favored sites between nucleosomes.

The correlation between the positions of integration and oscillating nucleosome occupancy suggested the pattern was due to the position of nucleosomes. To test this possibility we examined the insertion levels at the transcription start sites (TSSs) (annotated in the Sanger Center, February 2011 chromosome contigs based on Dutrow *et al.* (2008) and Lantermann *et al.* (2010) (Figure 3C). As expected, nucleosome occupancy was very low in the regions ~150 nt upstream of TSSs and from the start of ORFs, nucleosomes were positioned in phased arrays. Integration generated *in vivo* was very high in the nucleosome-free region upstream of the TSSs and from the start of ORFs the integration was reduced but oscillated in sync with the spaces between nucleosomes. This analysis supported the correlation seen in Figure 3B, that integration generated *in vivo* favored nucleosome-free sites (Figure 3C). This oscillating pattern was not observed in the MRC set matched to the *in vivo* insertions nor in the *in vitro* inserts.

Integration density of ORFs identify essential genes

While the nucleosomes within ORFs appeared to inhibit integration, a full 33% of all integrations did occur in coding sequence. To investigate whether the densities of the integration generated in cells could be used to discriminate between essential and nonessential genes, we graphed the insertions surrounding three known essential genes, *cdc2*, *cdc19*, and *cdc25* (Figure 4, A–C). In each case, little or no integration was detected in these three essential ORFs while high levels of insertions were seen in the adjacent intergenic sequences and nonessential ORFs. This initial evaluation of integration profiling suggested that integration density could be used to identify essential genes.

We next tested whether Hermes integration could identify essential ORFs throughout the *S. pombe* genome by analyzing the integration densities (inserts per kilobase per million integrations, units normalized to millions of insertions per dataset so that densities from the *in vitro* experiment could be compared) for all annotated ORFs (Figure 5A). Two dominant groups of ORFs were observed with peak densities of 8.3 inserts/kb/million and 50 inserts/kb/

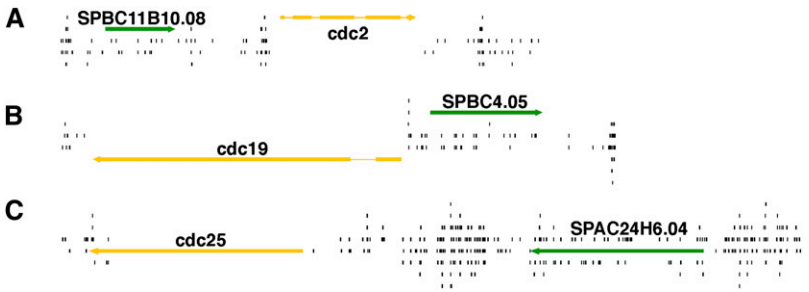

Figure 4 Essential genes accumulated far fewer insertions than nonessential genes. The Hermes insertions (black) are shown in regions encoding the essential genes (yellow) *cdc2* (A), *cdc19* (B), and *cdc25* (C). The adjacent genes are nonessential (green).

million. A consortium has systematically deleted individual ORFs in *S. pombe* to create a set of heterozygous diploids. By sporulating the strains and monitoring the germination of spores carrying the deletions, they designated which ORFs were essential for growth (Kim *et al.* 2010). Using these designations, we plotted the integration densities for the nonessential and essential ORFs separately (Figure 5B). The integration densities in the essential ORFs were clearly lower than in the nonessential ORFs, indicating that the integration profiles did discriminate between essential and nonessential genes. The integration density at the peak with the maximum ORFs was 5.5 inserts/kb/million for the essential ORFs and 50 inserts/kb/million for the nonessential ORFs. These two peaks corresponded with the two maximum peaks of integration in the total group of ORFs shown in Figure 5A. We also analyzed the integration densities of a subclass of nonessential genes that, when deleted, resulted in small colonies. Importantly, these nonessential ORFs had intermediate densities of integration (Figure 5B). This indicates that the intermediate levels of integration were detected because these nonessential genes made important contributions to growth. The low levels of integration in the essential ORFs and the intermediate levels of integration in the nonessential ORFs that contribute to colony growth were not due to properties of the sequences or the transposase, since an analysis of the *in vitro* integration showed the essential and nonessential ORFs contained equivalent densities of integration (Figure 5C).

While the integration densities generally reflected whether an ORF was essential, there was some overlap in the distributions of the two groups (Figure 5B). For example, there were ~50 ORFs designated by the deletion consortium as essential that had integration densities of 59.4 insertions/kb/million up to 161 insertions/kb/million, levels that suggested the ORFs were nonessential (Table S1). One possible explanation for this discordance is that these genes were actually not required for cell division but instead played a key role in spore formation or germination. Since the deletion consortium relied on lack of spore germination to identify essential genes, proteins required for spore function would be incorrectly designated as essential for growth. Interestingly, over half of these 50 disparate ORFs play an important role in mitochondrial function. In contrast to *S. cerevisiae*, it is thought that *S. pombe* requires mitochondrial DNA and many nuclear-encoded mitochondrial genes for viability even when fermentable carbon sources are available (Haffter and Fox 1992; Kim *et al.* 2010). However, the possibility existed that these nuclear-encoded mitochondrial genes of *S. pombe* are required for spore germination, not cell division.

We tested directly whether four representative nuclear-encoded mitochondrial genes from Table S1 were required for cell division with a "plasmid shuffle" method that uses a haploid strain containing a plasmid-encoded candidate gene but lacking the chromosomal copy of the same gene. If the strain grows on medium containing 5-FOA, a compound that selects for plasmid loss, then the gene is nonessential (Sikorski and Boeke 1991; Ben-Aroya *et al.* 2008). The genes that we tested encoded mitochondrial ribosomal proteins (*mrpl16*, *mrpl19*, and *SPBC2D10.08c*) and a factor involved in mitochondrial DNA maintenance (*mmf1*). Each of these four ORFs had high numbers of insertions (79.4 to 128 insertions/kb/million) but were designated as essential by the consortium. In each case, when the strains carrying plasmid copies of the genes were diluted onto medium containing 5-FOA, no growth occurred, indicating that these genes were truly essential (Figure 6A, right panel). In comparison, when a strain lacking *SPBC21C3.09c*, a nonessential gene, was diluted on the same plate, cells grew readily, indicating they did not require a plasmid encoding *SPBC21C3.09c*. Surprisingly, when patches of the same deletion strains were replica printed from rich medium directly onto medium containing 5-FOA, strains lacking the mitochondrial genes showed strong levels of growth while equivalent strains lacking a functional Cdc19p, an essential protein, did not grow (Figure 6B). Taken together with the high numbers of insertions in these ORFs, these data indicate that while we found the nuclear-encoded mitochondrial genes are truly essential for division, the pools of the proteins and the number of mitochondria that segregate during mitosis are sufficient to support many cycles of cell division after their genes were disrupted. This result suggests that the 50 ORFs identified in Table S1 represent a baseline of genes that produce pools of protein that were too large to be depleted during the growth of our cultures.

Ten percent of the strains in the deletion collection retain a copy of the genes that were targeted for deletion

Another aspect of the Hermes integration data that is discordant with the designations of the deletion consortium is that among the genes reported by the consortium to be

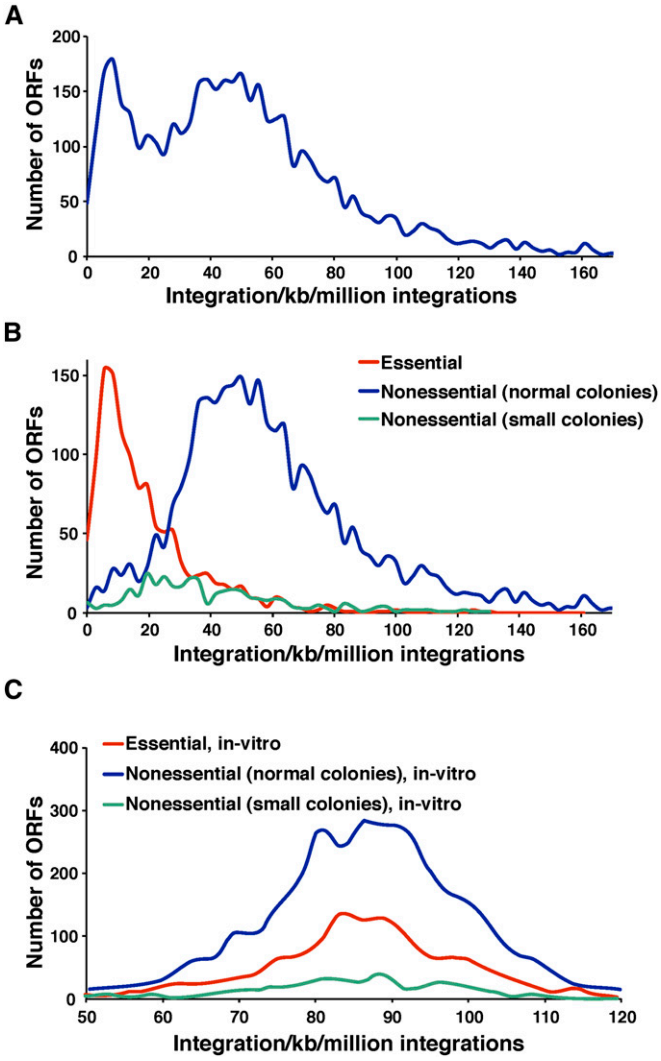

Figure 5 Essential ORFs exhibited low integration densities. (A) The *in vivo* integration densities (inserts/kb/million integrations) for each ORF were determined and the distributions were plotted. (B) The distributions of *in vivo* integration densities were plotted for ORFs designated by the deletion consortium as essential (red) and nonessential (blue). The distribution of densities was also plotted for ORFs that when deleted, result in small colonies (green). (C) The *in vitro* integration densities (inserts/kb/million integrations) for each ORF were determined and the distributions were plotted as above in B.

nonessential, there were 220 ORFs that had integration densities below the average for the essential ORFs (19.7 inserts/kb/million). This low amount of Hermes integration suggests these genes are actually essential for viability (Table S2).

Some of these ORFs encode highly conserved proteins such as eIF2 gamma, eIF6, and Pub1p that are known to be essential in other eukaryotes. One explanation for the low integration in these ORFs is that they are essential for cell division and in the process of deleting the genes or in the subsequent step of sporulating the heterozygous diploids, chromosomal rearrangements occurred that produced haploid strains with the desired deletion and an ectopic copy of the ORF that had been deleted. Strong selection for these events would exist if the ORFs were essential. Just such a set of processes occurred with the *S. cerevisiae* deletion set where a full 8% of 300 strains tested were found to retain an intact copy of the deleted ORFs either in aneuploid or rearranged chromosomes (Hughes *et al.* 2000).

To test whether the strains in the *S. pombe* deletion set retained copies of ORFs thought to be deleted, we assayed for the presence of these ORF sequences with PCR. For these experiments, we chose a representative set of 77 ORFs reported by the consortium to be nonessential and that had increasing integration densities from 0 to 42 inserts/kb/million (Table S3). We generated two pairs of primers to amplify different segments of each of these nonessential ORFs (Figure 7A). We found that 31 (40%, counting deletions producing the same number of PCR bands as wild type, wt) of the haploid strains tested retained ORF sequences reported to be deleted (Table S3). In addition, when the ORFs tested were grouped by the amounts of Hermes integration they had, a clear trend was observed. The ORFs with lower Hermes integration densities were much more likely to be retained in the deletion strains (Figure 7B). Of the 19 representative ORFs tested that had <5.5 inserts/kb/million integrations, 11 (58%) were still present in their respective deletion strains. The representative ORFs with greater levels of integration were more likely to be absent in their deletion strains. For example, the 10 representative ORFs with 41–42 insertions/kb/million insertions were all deleted successfully. To check whether our PCR reactions might have spuriously detected ORF sequences due to contaminating DNA, we used DNA blots and tested the presence of a sampling of ORFs. Six ORFs that had varying densities of Hermes integration and that we found by PCR to be present in the deletion strains were probed on DNA blots. All six ORFs were clearly present in the deletion strains in which they had been reported to be deleted (Table S3). Given that DNA contamination was not a significant problem with our PCR assays, we used the data in Figure 7B and the levels of Hermes integration to estimate the number of strains in the deletion collection that have been incorrectly reported to lack the specified ORFs. Based on a linear regression of integration densities of ORFs designated to be nonessential with 0–42 inserts/kb/million integrations, we estimate that ~300 (10%) of the deletion strains designated nonessential are actually essential and contain the ORFs said to be deleted.

Discussion

We described here integration profiling, a transposon-based technique that relies on integration densities to discriminate between the essential and nonessential genes of *S. pombe*. The 360,513 independent sites of *in vivo* integration resulted in an average of one insertion for every 29 nucleotides of nonrepetitive genome. Essential genes accumulated many fewer insertions than nonessential genes and importantly, nonessential

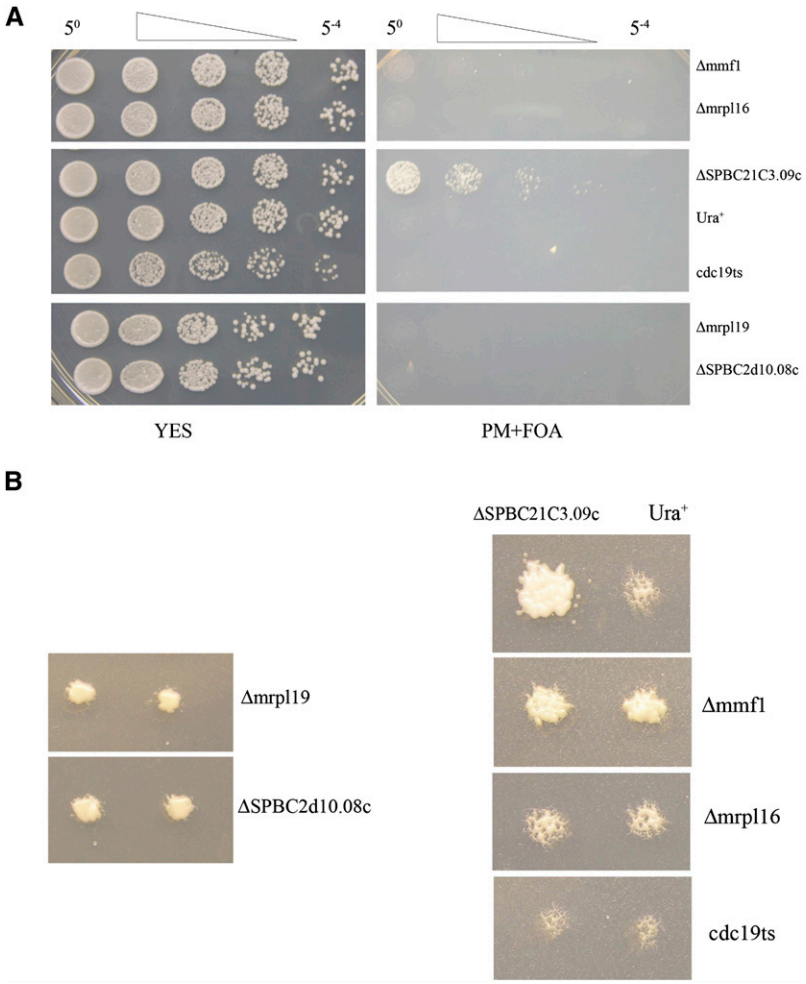

Figure 6 Essential genes important for mitochondrial function tolerated high numbers of insertions. Strains lacking chromosomal-encoded genes for mitochondrial ribosomal proteins mrpl16, mrpl19, and SPBC2D10.08c or for the mitochondrial DNA maintenance factor mmf1 each contained plasmids encoding the absent genes. Each strain was tested for viability on media (5-FOA) that selected against the plasmids. A strain lacking the nonessential gene SPBS21C3.09c was included as a control. Also shown is a control strain with a chromosomal copy of *ura4* (Ura+). (A) Serial fivefold dilutions of cultures were spotted onto rich medium (left panel) and onto medium containing 5-FOA (right panel). (B) The strains described in A were replica printed directly onto medium containing 5-FOA.

genes that contribute to colony growth accumulated intermediated densities of insertions.

A total of 67% of the *in vivo* insertions occurred in intergenic sequences. This fraction represents a high level of enrichment relative to 39.8%, the intergenic proportion of the nonrepetitive genome (Wood *et al.* 2002). In comparison, 37.0% of the insertions generated *in vitro* occurred in intergenic sequence, a fraction close to the intergenic portion of the nonrepetitive genome. This indicates the high level of intergenic integration that occurred *in vivo* was not due to the selection of these sequences by the transposase itself. It is formally possible that the *in vivo* insertions were detected in the intergenic regions because integration in ORFs was more likely to be detrimental. However, our data indicate that Hermes integrated more often in intergenic regions because these sequences lack nucleosomes, and sequences with low nucleosome occupancy were favored targets of integration. The nucleotide frequencies of sequences flanking insertion sites in *S. pombe* exhibited an unusual oscillation of A/T content with a wavelength of 150 bp (Figure 2A). Insertion numbers peaked in spaces between positioned nucleosomes and in the nucleosome-free regions at the TSSs upstream of ORFs (Figure 3, B and C). We suggest this pattern arises because nucleosomes occlude Hermes integration. A similar proposal was made for Hermes integration in *S. cerevisiae* where insertions were strongly associated with nucleosome-free regions (Gangadharan *et al.* 2010).

The nucleotide content downstream of the insertions generated *in vivo* exhibited an unusual bias favoring T that extended as far as 3000 bp (Figure 2C). The same 3000-bp bias was observed when we analyzed the MRC set that occurred in regions upstream of ORFs. This caused us to evaluate the A/T content of sequences upstream of ORFs and led to the surprising finding that for >500 bp upstream of ORFs the average nucleotide frequency of T is higher than A (Figure S5A). This appears to be an unusual structure for a genome as it did not exist in *S. cerevisiae* or in the genomes of other organisms we examined (Figure S6A and data not shown).

The high activity of Hermes in *S. pombe* and its ability to disrupt ORFs resulted in integration densities that were sufficient in most cases to distinguish between essential and nonessential ORFs. Recently published experiments with prokaryotic transposons demonstrated that dense integration maps can identify essential genes of bacteria (Gawronski *et al.* 2009; Langridge *et al.* 2009; van Opijnen *et al.* 2009;

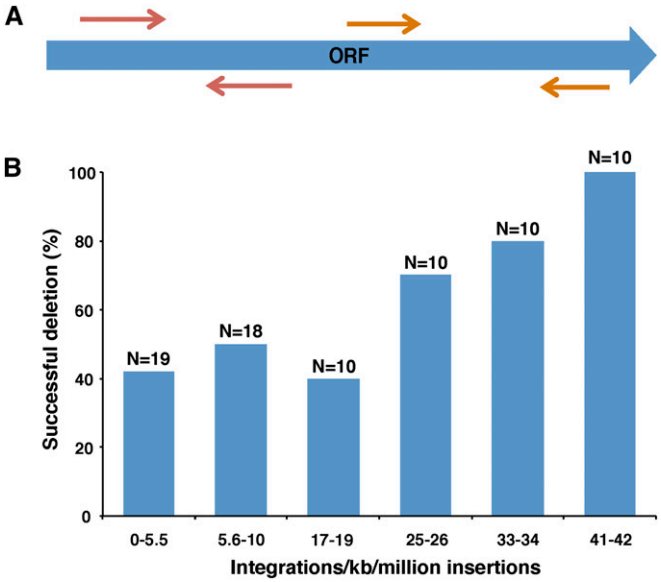

Figure 7 PCR assays of deletion strains revealed some genes had not been successfully removed. (A) A collection of 77 representative strains each said by the consortium (Kim *et al.* 2010) to carry a deletion of a nonessential ORF was assayed by PCR with two pairs of primers to determine whether the ORF sequences were present. (B) The ORFs defined by the consortium to be nonessential were divided into six bins based on their densities of integration. The number of ORFs tested in each bin was designated N. To be conservative the deletions were classified as successful if neither of the PCR products produced by the wild-type strain were generated by the deletion strain. The percentage of ORFs in each group that was successfully deleted is shown.

Christen *et al.* 2011). It is our application of the Hermes transposon that now makes this approach possible in a eukaryotic system.

After 80 generations of cell division, the pool of *S. pombe* cells with a Hermes insertion contained few disruptions of essential ORFs. However, there was a set of 50 ORFs with high densities of integration that were designated essential. In a sample of four of these genes, we confirmed their essential status (Figure 6A). We noted that many genes in this set were important for mitochondrial function (Table S1). The high amount of mitochondrial protein in cells suggested the possibility that these essential ORFs had many insertions because after these genes were disrupted, cells grew for many generations before their gene product was depleted. This idea was supported by our study of four of these deletions where we found that strains grew for many more generations after these essential genes were removed (Figure 6B). While large pools of proteins make it difficult to detect their essential function using integration profiling, the number of genes expressing such large pools is a small percentage of the entire gene set of *S. pombe*.

Integration profiling provided a means of determining which genes are essential, independent of the deletion method used by the consortium (Kim *et al.* 2010). Of the ORFs that had low densities of integration, we found a surprising number had been designated nonessential by the consortium. PCR and/or DNA blot assays of 77 ORFs revealed that at least 40% of these strains retained a copy of the ORF that was targeted for deletion. This result and the finding that many of these genes are known to be essential in other organisms led us to conclude that these ORFs with low integration densities are essential. We found that ORFs targeted for deletion were more likely to be retained if the ORF had low amounts of Hermes integration. Using a linear regression of the integration densities we project that ~300 ORFs thought to be nonessential were not successfully deleted in the deletion collection. The consortium determined which genes are essential by deleting one copy of an ORF in a diploid and following sporulation, testing whether haploids with the deletion are viable. Deletion of an essential ORF creates strong selection for suppressor mutations or chromosomal rearrangements that produce an ectopic copy of the gene. We believe this has happened with many of the deletion strains either in the haploid, the diploid, or during meiosis. An analysis of the *S. cerevisiae* collection of deletions revealed that ~8% of the genes deleted were nevertheless present in aneuploid or rearranged chromosomes (Hughes *et al.* 2000). Since these types of genetic alterations led to mistaken predictions of essentiality in both the *S. cerevisisae* and *S. pombe* deletion sets it is clear that there is a need for an independent method for establishing which genes are essential. Integration profiling is such an independent method for reliably identifying essential genes. One significant advantage of integration profiling is that it also can identify which nonessential genes make important contributions to growth. The intermediate integration levels of ORFs that, while not essential, contributed significantly to colony growth indicates that integration densities can be a measure of function even for nonessential genes. As a result, the capacity of integration profiling to document intermediate contributions to growth allows for a more accurate estimate of gene function than the designations of essential and nonessential.

Although we have validated this application of integration profiling under ideal growth conditions, it is clear that the method can be readily adapted to measure gene contribution to a wide variety of processes such as responses to environmental stress, repair of DNA damage, or viability during long periods of quiescence. Integration profiling can also be applied to identify gene interactions in a "synthetic lethal" approach by conducting integration in strains that have mutations of interest. As sequencing technology continues to improve, integration profiling will also have applications in organisms with more complex genomes, in the identification of genes that can be haploinsufficient or dominant negative. Such approaches in cultured cells will have important applications in the identification of disease pathways and physiological systems.

Acknowledgments

We thank Jacqueline Hayles for kindly sharing data on the colony size of the strains with gene deletions. This research

was supported by the Intramural Research Program of the National Institutes of Health (NIH) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and by the Intramural Program of the National Cancer Institute. The work by S.G. and N.L.C. was supported by NIH grant RO1GM076425.

Literature Cited

Ben-Aroya, S., C. Coombes, T. Kwok, K. A. O'Donnell, J. D. Boeke *et al.*, 2008 Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. Mol. Cell 30: 248–258.

Brass, A. L., D. M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman *et al.*, 2008 Identification of host proteins required for HIV infection through a functional genomic screen. Science 319: 921–926.

Christen, B., E. Abeliuk, J. M. Collier, V. S. Kalogeraki, B. Passarelli *et al.*, 2011 The essential genome of a bacterium. Mol. Syst. Biol. 7: 528.

Dutrow, N., D. A. Nix, D. Holt, B. Milash, B. Dalley *et al.*, 2008 Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat. Genet. 40: 977–986.

Evertts, A. G., C. Plymire, N. L. Craig, and H. L. Levin, 2007 The hermes transposon of Musca domestica is an efficient tool for the mutagenesis of Schizosaccharomyces pombe. Genetics 177: 2519–2523.

Forsburg, S. L., and N. Rhind, 2006 Basic methods for fission yeast. Yeast 23: 173–183.

Gangadharan, S., L. Mularoni, J. Fain-Thornton, S. J. Wheelan, and N. L. Craig, 2010 Inaugural Article: DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. Proc. Natl. Acad. Sci. USA 107: 21966–21972.

Gawronski, J. D., S. M. Wong, G. Giannoukos, D. V. Ward, and B. J. Akerley, 2009 Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. Proc. Natl. Acad. Sci. USA 106: 16422–16427.

Guo, Y., and H. L. Levin, 2010 High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in Schizosaccharomyces pombe. Genome Res. 20: 239–248.

Haffter, P., and T. D. Fox, 1992 Nuclear mutations in the petite-negative yeast Schizosaccharomyces pombe allow growth of cells lacking mitochondrial DNA. Genetics 131: 255–260.

Holman, A. G., and J. M. Coffin, 2005 Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proc. Natl. Acad. Sci. USA 102: 6103–6107.

Hughes, T. R., C. J. Roberts, H. Dai, A. R. Jones, M. R. Meyer *et al.*, 2000 Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet. 25: 333–337.

Kim, D. U., J. Hayles, D. Kim, V. Wood, H. O. Park *et al.*, 2010 Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28: 617–623.

Konig, R., Y. Zhou, D. Elleder, T. L. Diamond, G. M. Bonamy *et al.*, 2008 Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell 135: 49–60.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Langridge, G. C., M. D. Phan, D. J. Turner, T. T. Perkins, L. Parts *et al.*, 2009 Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 19: 2308–2316.

Lantermann, A. B., T. Straub, A. Stralfors, G. C. Yuan, K. Ekwall *et al.*, 2010 Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat. Struct. Mol. Biol. 17: 251–257.

Maertens, G. N., S. Hare, and P. Cherepanov, 2010 The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature 468: 326–329.

Park, J. M., A. G. Evertts, and H. L. Levin, 2009 The Hermes transposon of Musca domestica and its use as a mutagen of Schizosaccharomyces pombe. Methods 49: 243–247.

Sikorski, R. S., and J. D. Boeke, 1991 In vitro mutagenesis and plasmid shuffling: from cloned gene to mutant yeast. Methods Enzymol. 194: 302–318.

van Opijnen, T., K. L. Bodi, and A. Camilli, 2009 Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat. Methods 6: 767–772.

Wang, G. P., A. Ciuffi, J. Leipzig, C. C. Berry, and F. D. Bushman, 2007 HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res. 17: 1186–1194.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson *et al.*, 1999 Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285: 901–906.

Wood, V., R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne *et al.*, 2002 The genome sequence of Schizosaccharomyces pombe. Nature 415: 871–880.

Yamane, K., T. Mizuguchi, B. Cui, M. Zofall, K. Noma *et al.*, 2011 Asf1/HIRA facilitate global histone deacetylation and associate with HP1 to promote nucleosome occupancy at heterochromatic loci. Mol. Cell 41: 56–66.

Zhou, H., M. Xu, Q. Huang, A. T. Gates, X. D. Zhang *et al.*, 2008 Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe 4: 495–504.

Zhou, L. Q., R. Mitra, P. W. Atkinson, A. B. Hickman, F. Dyda *et al.*, 2004 Transposition of hAT elements links transposable elements and V(D)J recombination. Nature 432: 995–1001.

Communicating editor: D. Voytas

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152744/-/DC1

Integration Profiling of Gene Function With Dense Maps of Transposon Integration

Yabin Guo, Jung Min Park, Bowen Cui, Elizabeth Humes, Sunil Gangadharan, Stevephen Hung, Peter C. FitzGerald, Kwang-Lae Hoe, Shiv I. S. Grewal, Nancy L. Craig, and Henry L. Levin

DOI: 10.1534/genetics.113.152744

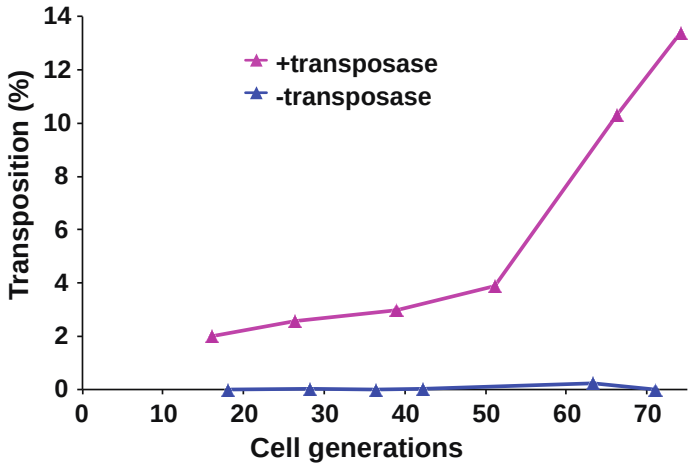

Figure S1 The percentage of cells with an insertion was measured during prolonged growth in liquid culture. After 74 generations 13.4% of the cells expressing the transposase contained an insertion (Pink). Cells in a control culture did not express the transposase (Blue).

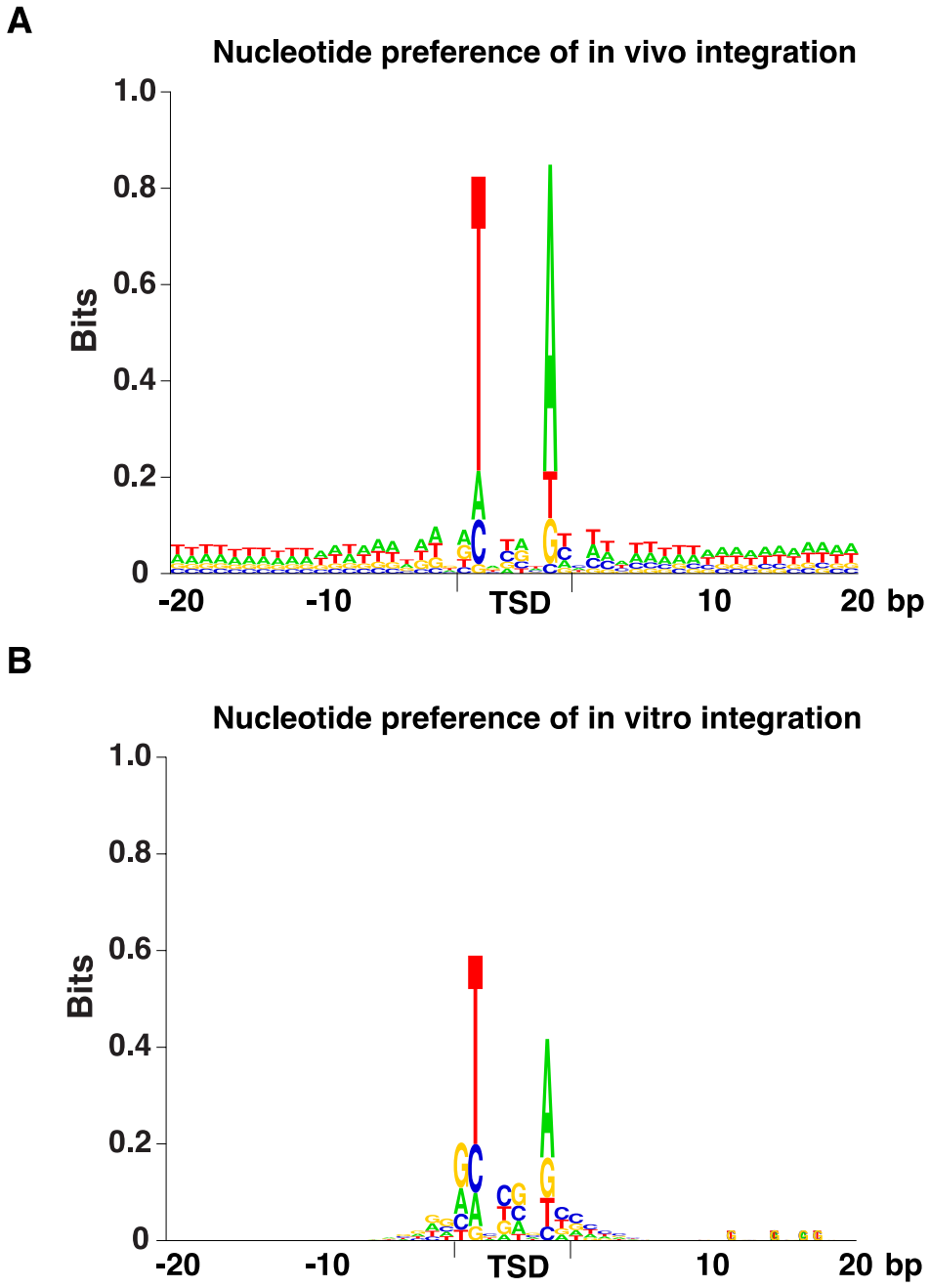

Figure S2 The average nucleotide frequencies flanking the insertion sites were determined. All Hermes insertions were aligned and the nucleotide patterns were analyzed by sequence logo created by a program written in Visual Basic based on an existing algorithm (Schneider and Stephens, 1990). The positions of the eight nucleotides at the target sites that are duplicated during integration are indicated by TSD. The nucleotide preferences for the *in vivo* integration sites are shown in (A) and the *in vitro* integration sites are shown in (B).

Schneider, T. D., and Stephens, R. M., 1990 Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18:6097-6100.

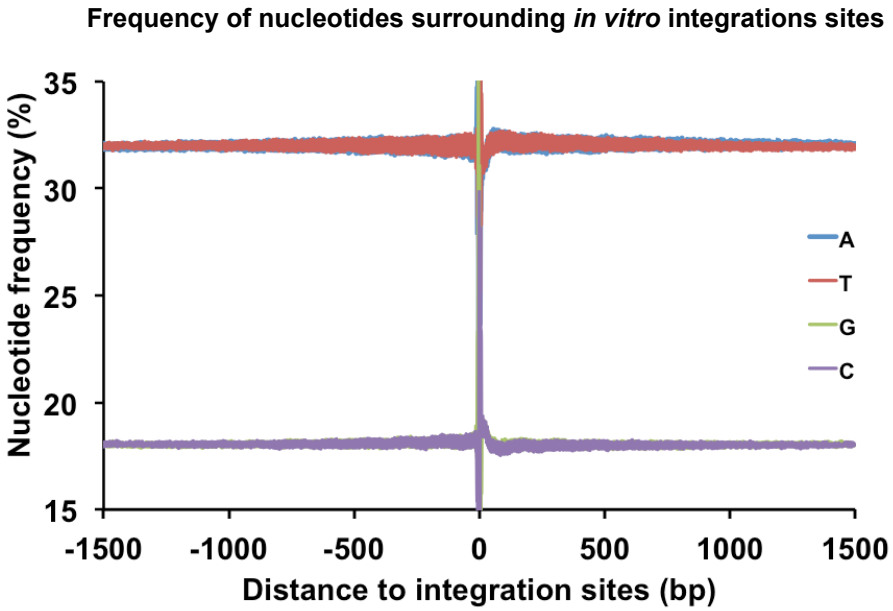

Figure S3 The average nucleotide frequencies were determined for the positions on either side of the *in vitro* insertion sites. The insertion sites were aligned and the nucleotide frequencies were calculated for a window that extended 1.5 kb on both sides.

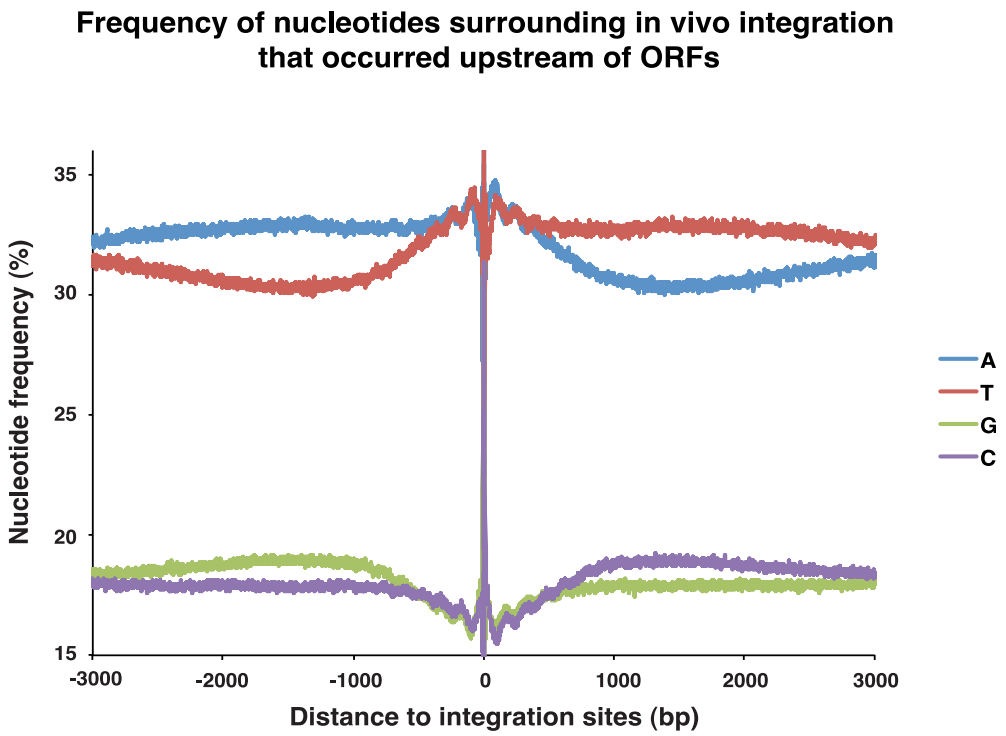

Figure S4 The average nucleotide frequencies were determined for the positions on either side of the *in vivo* insertion sites that occurred upstream of ORFs. The insertion sites were aligned and the nucleotide frequencies were calculated for a window that extended 3.0 kb on both sides.

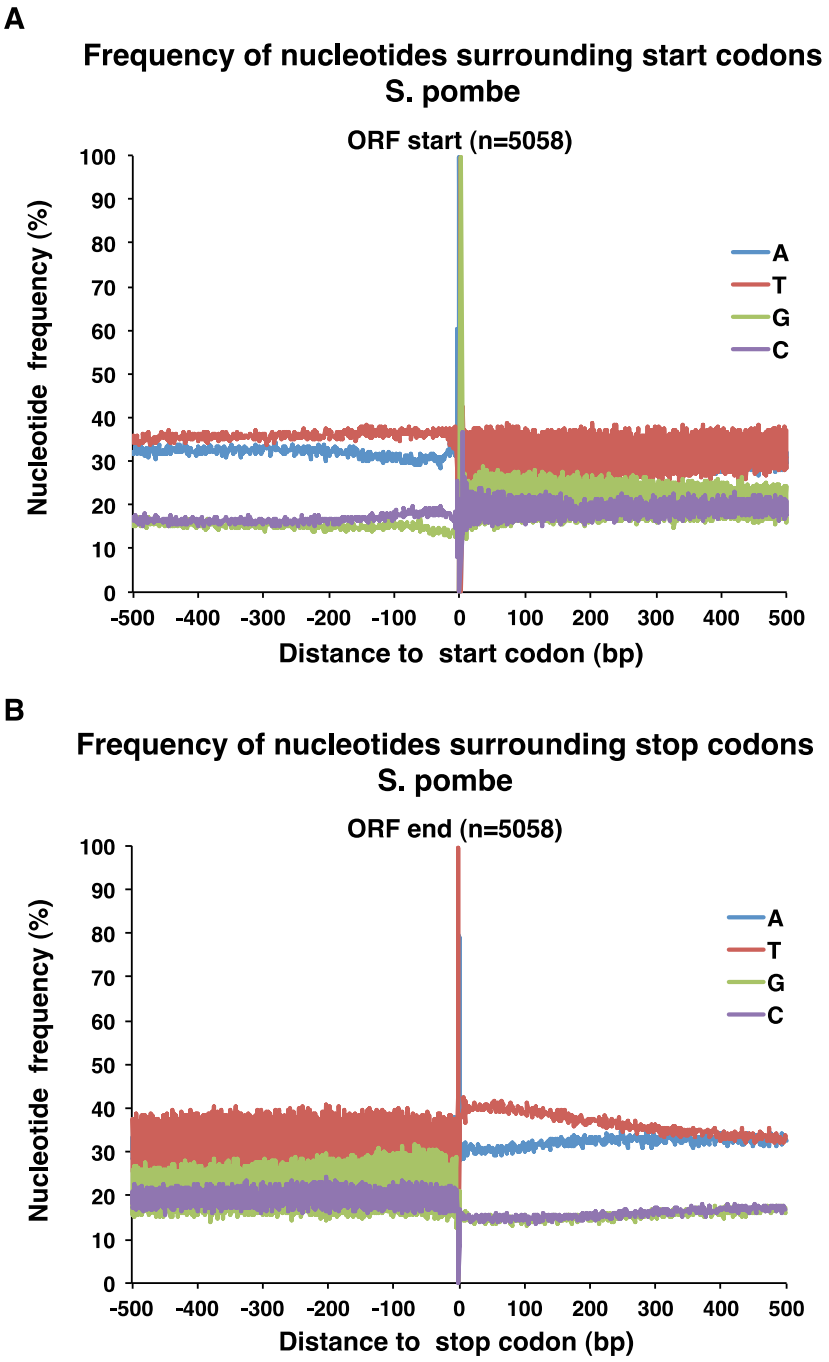

Figure S5 The average nucleotide frequencies upstream and downstream of ORFs were determined. A. The ATGs of the *S. pombe* ORFs were aligned and the nucleotide frequencies upstream of the ORFs were plotted. B. The stop codons of the *S. pombe* ORFs were aligned and the nucleotides frequencies downstream were plotted.

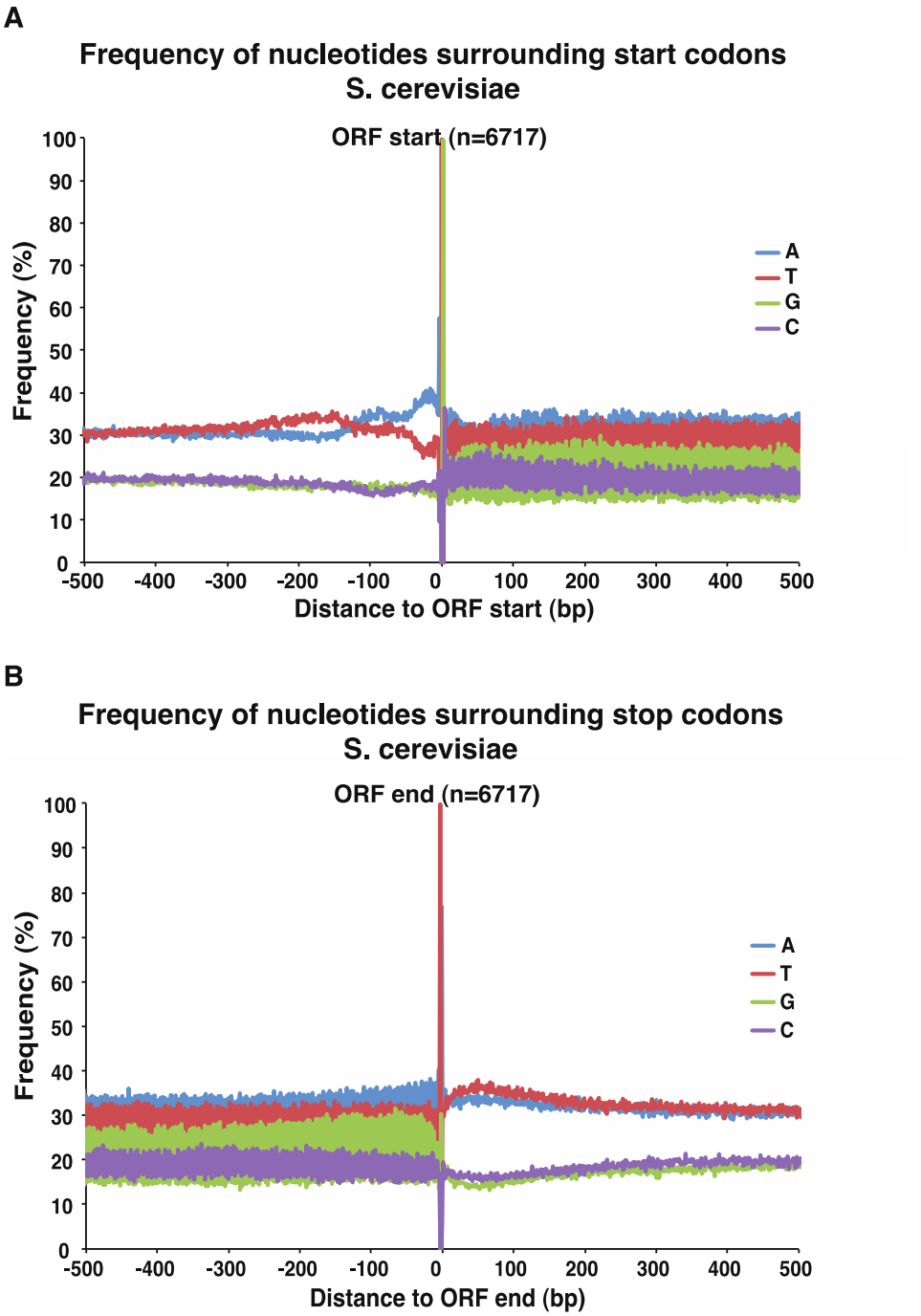

Figure S6 The average nucleotide frequencies upstream and downstream of ORFs in *Saccharomyces cerevisiae* were determined. A. The ATGs of the *S. cerevisiae* ORFs were aligned and the nucleotide frequencies upstream of the ORFs were plotted. B. The stop codons of the *S. cerevisiae* ORFs were aligned and the nucleotides frequencies downstream were plotted.

File S1

Materials and Methods

Plasmid constructs

The donor plasmid (pHL2577) and expression plasmid (pHL2578) have been previously described and are listed in Table S6. The Ura-marked donor plasmid consists of the KanMX6 gene cloned between the Hermes TIRs. The Leu-marked expression plasmid contains the Hermes transposase gene driven by the Rep81x nmt1 promoter.

The plasmid pHL2806 was created by combining the HindIII fragment of pHL2578 containing LEU2 with the HindIII fragment of pON177 in order to replace the URA3 marker of pON177 with LEU2.

To generate the plasmids for the essentiality test, the DNA fragments encoding mmf1, mrpl16, mrpl19, and SPBC2d10.08c were amplified by PCR using genomic DNA from YHL912 as a template and the primers in Table S7. Each PCR reaction consisted of 5 µL 10X Pfx amplification buffer, 1.5 µL 10 mM dNTPs, 1 µL 50 mM MgSO4, 1.5 µL primer mix (10 µM each), 100 ng template DNA, 0.4 µL Platinum Pfx DNA polymerase (2.5 U/µL), and water to 50 µL. The thermocycling conditions were as follows: 94ºC 2 min, 25-35x [94ºC 15 sec, 55ºC 30 sec, 68ºC for 1 minute per kb], 72ºC 10 min, hold 4ºC. The products were purified on Qiagen PCR purification columns and cloned into pCR□2.1-TOPO (Invitrogen) following manufacturer's instructions. The constructs were then digested with SalI and XhoI (NEB) (SPBC2d10.08c) or with XhoI and SacI (NEB) (mmf1, mrpl16, and mrpl19) to isolate the PCR gene inserts. The gene inserts were then cloned into pHL2612 to generate the deletion plasmids listed in Table S6.

Generating a library of Hermes insertions

A single colony from YHL9609 was used to inoculate a 5 mL starter culture of EMM-Leu-Ura+B1 that was grown for 16 hours at 32ºC and ~200 rpm. The culture was then pelleted at 3,000 rpm for 5 minutes and washed three times with 50 mL EMM-Leu-Ura-B1 to remove any traces of thiamine. The pellet was resuspended in 2 ml of the same media and the OD600 of the culture was measured. These measurements were used to calculate the volume of cells needed to inoculate 50 ml EMM-Leu-Ura-B1 at OD600=0.05. The absence of thiamine from this culture induced transcription of the Hermes transposase enzyme and initiated transposition. This culture was grown to OD600 =5, and was used to inoculate a new series of cultures at OD600=0.05. The serial passaged cultures were continued until the percent of cells with integration reached 13%, or about 80 generations of cell division. In all, it took a series of 12 sequential cultures for the strain to reach this point.

The final 50 ml culture was used to inoculate 500 ml of EMM+Leu+Ura+B1+FOA at OD600=0.25 to select against the donor plasmid. The cultures were grown for 4 days until an OD600 of 3 was reached. The EMM+FOA culture was pelleted, washed, and used to inoculate 500 mL YES+FOA+G418 at OD600=0.5 to select cells with Hermes insertions. The YES+FOA+G418 culture was grown for 48 hours to an OD600 of 10.

To measure transposition frequencies over the course of serial passaging, cells from cultures that reached OD600=5 were used to make a series of five 10-fold dilutions, starting at 10^8 cells/ml and ending with 10^4 cells/ml. From the three most dilute cultures, 100 µl of cells were spread on EMM+Leu+Ura+FOA+B1 plates. The same volume from the three least dilute cultures was spread on YES+FOA+G418 plates. Transposition frequency was calculated by dividing the number of colonies on EMM+Leu+Ura+FOA+B1 by the number of colonies on YES+FOA+G418.

To confirm that the majority of the cells lacked the donor plasmid and contained a Hermes integration event following growth in YES+FOA+G418 cultures, cells from the final YES+FOA+G418 cultures were plated as above, but on EMM complete and EMM-Ura. The percentage of cells retaining the donor plasmid was calculated by dividing the number of colonies on EMM -Ura by the number of colonies on EMM complete.

The protocol that follows below describes methods to prepare Hermes insertion libraries for high-throughput Illumina sequencing.

Genomic DNA extraction

The final 500 ml YES+FOA+G418 liquid culture was spun down at 3,000 rpm for 5 minutes. The pellet of cells was resuspended in 35 ml of Sp1 (1.2 M sorbitol, 50 mM citric acid monohydrate, 50 mM Na2HPO4*7H20, and 40 mM EDTA, pH 5.6) containing 105 mg Zymolyase 100T (Seikagaku) previously dissolved. Cells were incubated at 37ºC for 1-2 hours with gentle shaking. The cells were pelleted and resuspended in 105 mL of 5X TE and 1% SDS. Cells were incubated at 25ºC for 1 hour and then at 65ºC for five minutes. After this point, 35 ml of 5M KOAc was added to the mixture, and the cells were incubated on ice for 30 minutes. The cells were centrifuged for 15 min at 4200 rpm in an SS34 rotor, and an equal volume of ice-cold isopropanol was added to the supernatant. The mixture was placed at -20ºC overnight. The mixture was then centrifuged at 8000 rpm in an SS34 rotor for 10 minutes. The DNA pellet was resuspended in 21 ml of 5X TE, and RNaseA (Qiagen) was added to a final concentration of 100 µg/ml. The solution was incubated at 37ºC for 1-2 hours. Three phenol extractions and one phenol/chloroform/isoamyl alcohol extraction were performed to remove proteins. The nucleic acids were precipitated using 1/10 volume of 5 M NaCl and 2.5 volumes of 100% ethanol. The pellet was washed with 14 ml 70% ethanol, air dried, and resuspended in 300 µL 1X TE. The yield was approximately 100 µg.

Restriction endonuclease digestion

Genomic DNA was digested with MseI in a series of 6 duplicate digests. Each digest consisted of 1-2 µg DNA, 10 µl 10X NEB Buffer #2, 1 µL 100X BSA (NEB), 3 µl MseI (10U/µl, NEB), and water to 100 µl. The digests were incubated at 37ºC for 16 hours, followed by purification of each digest on a Qiagen PCR purification column. DNA was serially eluted from each column with a total of 150 µl EB.

Linker ligation

This step ligates MseI compatible ends of digested genomic DNA to linkers containing MseI restriction-site overhangs. Linkers were prepared by mixing equal amounts of HL1870 and HL1871 at a concentration of 10 µM in PCR buffer. For linker annealing, the mixture was denatured for 1 minute at 95ºC in a PCR machine, and the temperature was decreased to 80ºC for 7 minutes, and continued decrease by 10ºC every 7 minutes until 20ºC. The mixture was stored at -20ºC when not in use.

The linkers were ligated to MseI-digested DNA in a reaction consisting of 125 µl of purified MseI digest product, 40 µl 5X Ligase buffer, 35 µl annealed linker oligonucleotides (10 µM), and 5 µl T4 DNA Ligase (1U/µl). The reaction mixture was divided in 20 µl aliquots among 10 tubes, which were incubated at 18ºC for 16 hours, and stored at -20ºC.

PCR amplification of Hermes insertion sites

This step uses the linker-ligated genomic DNA fragments as templates for PCR amplification of genomic sequences flanking Hermes integration sites. A series of 40 PCR reactions were carried out in a 96-well plate. Each reaction consisted of 8 µl linker ligation template, 5 µl 10X PCR buffer, 1 µl dNTPs (10 mM), 1 µl LTR primer (10 µM, HL2215), 1 µl linker primer (10 µM, HL2216), 1 µl 50X Titanium Taq polymerase (Clontech), and water to 50 µl. Amplification was performed with the following cycling conditions: 94ºC 1 min, 6x [94ºC 15 sec, 65ºC 30 sec, 72ºC 30 sec], 24x [94ºC 15 sec, 60ºC 30 sec, 72ºC 30 sec], 72ºC 10 min, hold 4ºC.

Gel extraction and purification

Following amplification, PCR reactions were pooled and purified on 6 Qiagen PCR purification columns. The DNA was eluted in 400 µl EB and the entire volume of eluate was loaded in 1 lane (tape together multiple wells to create one large lane) of a 2% low melting point agarose, 1X TAE gel and run a few cm at 80V, next to a 100 bp DNA size standard for size reference. The DNA from 150-500 bp was excised with a clean razor blade and extracted using the Qiagen Gel purification kit following manufacturer's instructions. The DNA was eluted with 500 µL EB and extracted with one phenol extraction, one phenol:chloroform:isoamyl alcohol extraction, and one final chloroform extraction, using a 1:1 volume for each extraction. The products were purified on one Qiagen PCR purification column and eluted in 70 µl TE. DNA was quantified with a Quant-iT PicoGreen Fluorescence kit for dsDNA (Invitrogen).

Illumina sequencing

The DNA samples were sequenced by the Illumina Fast Track Sequencing Service in Hayward, California. The sequence reads were 50 nt in length. The raw sequences for the *in vitro* and *in vivo* inserts were deposited in the SRA database of Genbank with accession number SRA043841.1.

Genomic resources

All genomic resource data of *S. pombe* used in this study was downloaded from the Wellcome Trust Sanger Institute (ftp://ftp.sanger.ac.uk/pub/yeast/pombe/). The chromosome contigs are the Aug. 2007 version. The CDS coordinates are the Feb. 2009 version. The UTR data was extracted from the .embl files of 20110204 version.

Mapping Hermes integration sites on the genome of S. pombe.

Sequence reads from Illumina were screened for those containing Hermes left end. Then the Hermes sequences were trimmed. The trimmed sequences were aligned to the S. pombe genome using the NCBI BLAST software (blastall) on a local computer. The BLAST results were filtered to collect matches with genomic sequence that started from the first nucleotide after the Hermes end and with identities greater than or equal to 95% and expect (e) values less than or equal to 0.05. Then, of the matches that met these criteria, the one with the highest bit score was used to obtain the coordinates for the unique insertion sites. Sequences that were found to have the same insertion coordinate and the same orientation, were considered to be duplicate reads, and were counted as only one independent integration event.

Matched random control (MRC)

For each Hermes insertion site, the distance between the integration site and the responsible MseI site (d) was calculated. Then another MseI site coordinate (m) was randomly chosen from the S. pombe genome. Then m+d or m-d was taken as an MRC site. To "add" or "subtract" was also randomly determined. Thus, the MRC dataset has the same size as the experimental integration dataset and matches the distances to MseI sites.

General bioinformatic analysis and programming

The scripts for screening the raw sequences, filtering the BLAST outputs, extracting features from .embl files, determining the locations of Hermes insertion according to chromosomal features, generating MRC and other analyses were written in Perl or Ruby programming languages.

Nucleosome DNA preparation and sequencing

Nucleosomal DNA was prepared as described previously (YAMANE *et al.* 2011), and DNA samples were sequenced by using the Illumina sequencing protocol. Bowtie (LANGMEAD *et al.* 2009) was used to map the Illumina sequencing reads, trimmed to 25 bp of high quality reads, against the reference genome allowing for 2 mismatches. The mapped data was filtered to remove all sequences that mapped to more than a single location. The end positions of the reads were aligned relative to the center of the nucleosome by shifting the plus strands reads by +73 bp and the minus strand reads by -73 bp. The final nucleosome maps were produced by applying gaussian smoothing to the raw data in order to reduce noise. More detailed methodology will be described in a forthcoming manuscript on the genome wide mapping of nucleosomes in S. pombe.

Essentiality Test

To test the viability of four representative Bioneer strains heterozygous for deletions in nuclear-encoded mitochondrial genes designated as essential, we performed a plasmid shuffle assay. Following transformation with both the plasmids encoding the deleted genes and pHL2806 (needed to allow sporulation), the diploid strains carrying the plasmid with the deleted genes were sporulated by inoculating them at OD600=0.05 in 200 µl of low nitrogen media, and incubating them for 3 days at 25ºC in a rotating wheel. Low nitrogen media is identical to EMM media, except with 1g/l glutamic acid in place of NH4Cl. The cells were then pelleted and resuspended in water to OD600=0.1, and 20 µl of a 10-fold diluted stock of glusulase (PerkinElmer) was added to the mixture. The mixtures were allowed to incubate overnight at 32ºC in a rotating wheel. The cells were then pelleted, resuspended in 300 µl of 30% ethanol, and incubated for 30 minutes at 25ºC. Subsequently, the cells were pelleted, resuspended in 100 µl of water, plated on YES+G418 to select for haploid cells containing the KanMX6 deletion cassette, and grown for 3 days at 32ºC.

Gene essentiality was tested on solid media by patch assays and drop assays. For each patch assay, 15 cells from the spore germination plates were colony purified on YES+G418 plates and then patched to YES master plates that were grown at 32ºC for 2 days. These master plates were replica printed to YES+FOA plates (to select against the deletion plasmid) and grown at either 32ºC or 37ºC for 46 hours.

For each drop assay, a 1 ml stock of cells at OD= 0.5 was used to create a series of 5-fold dilutions. From each dilution, 10 µl of cells was plated on YES and PM+FOA plates and grown at 32ºC or 37ºC for 2-3 days.

Table S1

ORFs designated by the consortium to be essential but had high integration densities.

Gene	Gene function	integration/kb/million insertions
SPBC21C3.10c	5-amino-6-(5-phosphoribosylamino) uracil reductase	161.5
tim10 SPAC222.03c	Tim9-Tim10 complex subunit Tim10 (predicted)	130.2
mrpl16 SPBC1105.03c	mitochondrial ribosomal protein subunit L16	128.4
SPBP8B7.05c	carbonic anhydrase (predicted)	124.5
SPBC2D10.08c	mitochondrial ribosomal protein subunit Yml6	123.5
mrpl31 SPCC16A11.11	mitochondrial ribosomal protein subunit L31	111.6
ste11 aff1 stex SPBC32C12.0	transcription factor Ste11	108.4
SPAC1486.07c mrpl19	mitochondrial ribosomal protein subunit L19	102.0
SPCC1672.01	histidinol-phosphatase (predicted)	100.8
SPCC2H8.04	sequence orphan	99.4
met9 met5 SPAC56F8.10	methylenetetrahydrofolate reductase Met9	94.9
SPAC17H9.07	signal recognition particle subunit Srp21 (predicted)	94.6
tpx1 SPCC576.03c	thioredoxin peroxidase Tpx1	86.2
vph2 SPCC757.10	endoplasmic reticulum membrane involved in assembly of the V-ATPase	84.5
ups1 ups SPAC31G5.08	uroporphyrinogen-III synthase Ups1	80.7
mmf1 pmf1 SPBC2G2.04c	YjgF family protein Mmf1	79.4
grx4 SPBC26H8.06	glutaredoxin Grx4	79.3
SPAC30C2.03	sequence orphan	78.9
SPBC2A9.10	Bin3 family	78.4
SPAC12B10.02c	sequence orphan	78.4
mrp10 SPAC24C9.13c	mitochondrial ribosomal protein subunit Mrp10	78.0
SPBC3B9.14c mrpl3	mitochondrial ribosomal protein subunit L3	77.3
SPCC1682.09c	guanine nucleotide transporter	74.1
SPAC1002.16c	nicotinic acid plasma membrane transporter (predicted)	74.0
SPAC688.09	mitochondrial pyrimidine nucletide transporter (predicted)	70.9
SPBC577.09	ERCC-8 DNA repair homolog (predicted)	70.3
SPAC17G8.02	uridine ribohydrolase (predicted)	69.8
SPAC31G5.06	mitochondrial protein	69.8
dea2 SPBC1198.02	adenine deaminase Dea2	67.8
SPCC63.10c	dolichol kinase (predicted)	67.4
SPAC24C9.06c	aconitate hydratase	66.5
SPBC354.06 mrps16	mitochondrial ribosomal protein subunit S16	65.5
SPBP4H10.15	aconitate hydratase/mitochondrial ribosomal protein subunit L49, fusion protein	64.9
bet5 SPAC688.15 SPAC3G9.1	TRAPP complex subunit Bet5 (predicted)	64.6
usp103 yhc1 SPBP35G2.09	U1 snRNP-associated protein Usp103 (predicted)	63.6
crk1 mop1 mcs6 SPBC19F8.(	cyclin-dependent kinase activating kinase Crk1	63.3
SPAPB17E12.09	sequence orphan	63.2
mrp20 SPAC31A2.08	mitochondrial ribosomal protein subunit L23 (predicted)	62.8
fta3 SPBP8B7.12c sma3	Sim4 and Mal2 associated (4 and 2 associated) protein 3	62.8
SPBC887.07 mrpl38	mitochondrial ribosomal protein subunit L38	62.7
SPBC2F12.10	mitochondrial ribosomal protein subunit L35	62.3
SPCC4B3.09c	mitochondrial ribosomal protein subunit L12	62.1
SPAPB24D3.06c	DUF1749 family protein	61.3
SPBC119.18	mitochondrial distribution and morphology protein Mdm35 (predicted)	61.2
SPBC17G9.13c	sequence orphan	61.1
SPAC2F7.15 rsm24	mitochondrial ribosomal protein subunit S24 (predicted)	60.9
SPBC16H5.15	conserved fungal protein	60.9
SPCC1393.11	mitochondrial ribosomal protein subunit L20	60.7
sum2 SPBC800.09	G2/M transition checkpoint protein Sum2	60.6
alr2 SPBC359.02	alanine racemase Alr2	59.8
SPAC3A12.19	mitochondrial ribosomal protein subunit L27	59.4

Table S2 The levels of integration for each ORF of the genome

Table S2 is available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152744/-/DC1.

Table S3

Test of deletion strains version 2.0 for presence of ORFs designated by the consortium to be nonessential.

chromosome	gene	start	end	strand	Number of integrations	Density (integration/kb /million insertions)	Gene description	bands in colony PCR	bands with the same primers in wt.	ORF detected on Southern blot
chr1	rpl36a SPAC15E1.03 rpl42	3722509	3723019	+	0	0.00	60S ribosomal protein L36/L42	2	2	√
chr1	ubi1 SPAC11G7.04	3969009	3969395	+	0	0.00	ribosomal-ubiquitin fusion protein Ubi1	1	2	
chr1	SPAC186.06	5542419	5542973	+	0	0.00	human MAWBP homolog	0	2	
chr2	smd3 SPBC19C2.14	1702182	1702628	+	0	0.00	Sm snRNP core protein Smd3	2	2	
chr2	SPBC16A3.16	4267784	4268090	-	0	0.00	mitochondrial inner membrane protein	0	2	
chr3	SPCC14G10.03c ump1	748037	748530	-	0	0.00	proteasome maturation factor Ump1	1	2	
chr3	SPCC162.01c	1587869	1588603	+	0	0.00	RNA-binding protein	2	2	
chr1	rhp54 rad54 SPAC15A10.03c	3680587	3683145	-	1	1.08	Rad54 homolog Rhp54	0	2	
chr2	SPBC577.06c	758408	764229	-	4	1.91	phosphatidylinositol kinase	2	2	
chr2	tif213 SPBC17G9.09	2187712	2189052	+	1	2.07	translation initiation factor eIF2 gamma subunit	2	2	√
chr2	arp5 SPBC365.10	2512667	2514887	+	2	2.50	actin-like protein Arp5	0	2	
chr1	fas1 SPAC926.09c	3900367	3906588	-	6	2.67	fatty acid synthase beta subunit Fas1	2	2	
chr1	SPAC186.03	5535198	5536280	+	1	2.56	L-asparaginase	1	1	
chr1	nup184 SPAP27G11.10c	1625076	1629841	-	6	3.49	nucleoporin Nup184	1	1	
chr1	spt6 SPAC694.07c SPAC1F7.01c	4212706	4216803	-	6	4.06	transcription elongation factor Spt6	2	2	√
chr3	bgs4 orb11 cwg1 SPCC1840.02c	2249279	2255146	-	9	4.25	1,3-beta-glucan synthase subunit Bgs4	2	2	
chr2	gpd1 SPBC215.05	4035814	4036971	+	2	4.79	glycerol-3-phosphate dehydrogenase Gpd1	0	1	
chr3	tif6 SPCC1919.09	2223154	2224257	+	2	5.03	translation initiation factor eIF6	2	2	√
chr1	pub1 SPAC11G7.02	3963323	3965870	+	5	5.44	ubiquitin-protein ligase E3	1	2	
chr2	atp16 SPBC13E7.04	3046873	3047355	+	1	5.74	F1-ATPase delta subunit	2	2	
chr2	SPBC1539.08 arf6	4375857	4376810	+	2	5.82	ADP-ribosylation factor, Arf family	1	2	
chr2	SPBC27B12.08	1335724	1341821	+	13	5.91	AP-1 accessory protein	0	1	
chr1	arp42 arp4 SPAC23D3.09	4353826	4355118	-	3	6.44	SWI/SNF and RSC complex subunit Arp42	0	2	
chr1	SPAC869.04	5511137	5512369	-	3	6.75	formamidase-like protein	1	2	
chr2	ypt7 SPBC405.04c	3147478	3148230	-	2	7.37	GTPase Ypt7	1	1	
chr1	vps28 SPAC1B3.07c	4940690	4941436	-	2	7.43	ESCRT I complex subunit Vps28	0	2	
chr3	SPCC5E4.10c	651622	651994	-	1	7.44	sequence orphan	0	2	
chr2	wis1 spc2 smf2 SPBC409.07c	1143967	1145784	-	5	7.63	MAP kinase kinase Wis1	2	2	
chr1	ies6 SPAC222.04c	950298	950651	-	1	7.84	chromatin remodeling complex subunit Ies6	2	2	
chr1	SPAC664.13	1729505	1729858	+	1	7.84	sequence orphan	2	2	
chr2	SPBPB10D8.01	81295	82953	+	5	8.36	cysteine transporter	0	2	
chr3	rpt4 SPCC1682.16	405612	407251	+	5	8.46	19S proteasome regulatory subunit Rpt4	2	2	√
chr2	sec63 SPBC36B7.03	2387093	2389038	+	6	8.55	ER protein translocation subcomplex subunit Sec63	2	2	√
chr1	SPAC20G8.09c	1415542	1418750	-	10	8.64	N-acetyltransferase Nat10	2	2	
chr2	mcs4 SPBC887.10	3557274	3558842	+	5	8.84	two-component response regulator	0	2	
chr2	SPBC9B6.07	1824562	1825497	+	3	8.89	nucleolar protein Nop52 family	0	2	
chr1	plc1 SPAC22F8.11	4807786	4810485	+	9	9.25	phosphoinositide phospholipase C Plc1	2	2	
chr3	ekc1 SPCC777.16c SPCC663.01c	1626547	1629120	-	16	17.24	protein phosphatase regulatory subunit Ekc1	2	2	
chr1	mcl1 slr3 SPAPB1E7.02c	3283195	3285745	-	16	17.40	DNA polymerase alpha accessory factor Mcl1	0	2	
chr2	swi5 SPBC409.03	1138056	1138530	+	3	17.52	Swi5 protein	1	1	
chr3	pof3 SPCC338.16	1345394	1347127	-	11	17.60	F-box protein Pof3	0	2	
chr1	SPAC6F6.03c	2734109	2736126	-	13	17.87	ribosome export GTPase	2	2	
chr3	SPCC1919.04	2211103	2211873	+	5	17.99	sequence orphan	1	2	
chr1	SPAC16C9.05 cph1	799550	800764	+	8	18.26	PHD finger containing protein	2	2	
chr2	SPBC119.12	740275	741480	+	8	18.40	Golgi matrix protein	0	2	
chr2	lsd1 SPBC146.09c swm1 saf110	1016000	1019002	-	20	18.47	histone demethylase SWIRM1	2	2	
chr2	nfs1 SPBC21D10.11c	2422769	2424265	+	10	18.53	iron-sulfur cluster assembly protein Nfs1	2	2	
chr1	pzh1 SPAC57A7.08	1527964	1529511	-	14	25.09	serine/threonine protein phosphatase Pzh1	0	2	
chr3	set1 SPCC306.04c	412765	415527	-	25	25.10	histone lysine methyltransferase Set1	0	1	
chr3	set9 SPCC4B3.12	1154281	1155606	-	12	25.10	histone lysine methyltransferase Set9	2	2	
chr2	rad3 SPBC216.05	904365	911525	+	65	25.18	ATR checkpoint kinase	0	2	
chr2	pdb1 SPBC30D10.13c	3070543	3071643	+	10	25.19	pyruvate dehydrogenase e1 component beta subunit Pdb1	2	2	
chr1	pds5 SPAC110.02	1912042	1915659	+	33	25.30	cohesin-associated protein Pds5	0	2	
chr1	cit1 SPAC6C3.04	2328675	2330096	+	13	25.36	citrate synthase	2	2	
chr3	ccr4 SPCC31H12.08c SPCC5E4.02c	637461	639533	-	19	25.42	CCR4-Not complex subunit Ccr4	0	2	
chr2	hip3 SPBC31F10.14c hir3	3777778	3782670	-	45	25.51	HIRA interacting protein Hip3	0	2	
chr2	SPBC1685.04	502873	503850	+	9	25.53	sequence orphan	0	1	
chr2	SPBC30D10.14	3067996	3068745	-	9	33.29	dienelactone hydrolase family	0	2	
chr3	SPCC645.13	1258634	1260799	+	26	33.30	transcription elongation regulator	0	2	
chr3	SPCC569.07	2415649	2417061	-	17	33.37	aromatic aminotransferase	0	2	
chr1	atp2 SPAC222.12c	968783	970360	-	19	33.40	F1-ATPase beta subunit	2	2	
chr2	SPBC1A4.05	1984516	1986671	+	26	33.45	sequence orphan	1	2	
chr1	spo15 SPAC1F3.06c	625430	631303	-	71	33.53	sporulation protein Spo15	0	2	
chr1	fhl1 SPAC1142.08 SPAC8C9.01	3639988	3642219	+	27	33.55	fork head transcription factor Fhl1	0	2	
chr2	amo1 SPBC15D4.10c	3028705	3030192	-	18	33.55	nuclear rim protein Amo1	1	2	
chr1	hem2 SPAC1805.06c	2783941	2784930	-	12	33.62	porphobilinogen synthase Hem2	2	2	
chr2	vps13a SPBC31F10.18c SPBC21C3.01c	3789386	3798703	-	113	33.64	chorein homolog	0	2	
chr1	gmh3 SPAC22E12.06c	5029000	5029998	-	15	41.65	alpha-1,2-galactosyltransferase Gmh3	0	2	
chr3	SPCC1840.07c	2270637	2271635	-	15	41.65	phosphoprotein phosphatase	0	2	
chr3	SPCC622.19 jmj4 mug149	1438375	1439839	+	22	41.65	Jmj4 protein	0	2	
chr1	SPAC24C9.05c mug70	3045121	3047313	-	33	41.74	conserved protein (fungal and plant)	0	2	
chr1	SPAC16E8.17c	3532418	3533413	-	15	41.77	succinate-CoA ligase	0	1	
chr1	SPAC31A2.12	412084	413874	+	27	41.82	arrestin/PY protein 1	0	2	
chr1	msh3 swi4 SPAC8F11.03	2855931	2858980	+	46	41.83	MutS protein homolog 3	0	2	
chr2	SPBC1773.16c	318693	320480	-	27	41.89	transcription factor	0	2	
chr3	SPCC1620.12c	2167799	2169586	-	27	41.89	GTPase activating protein	0	2	
chr1	idp1 SPAC6G10.08	3231599	3232855	+	19	41.93	isocitrate dehydrogenase Idp1	1	2	

Table S4 Chromosome coordinates of Hermes insertions

Table S4 is available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152744/-/DC1.

Table S5 Strains of *S. pombe* used in this study

Yeast strain	Genotype	Reference
YHL#10046	FY243, h-cdc19-P1 ura4-D18 leu1-32 ade6-M210	Gift of Susan Forsburg
YHL240	h+ ura4-294::neo-ura4+	This study
YHL912	h- leu1-32 ura4-294	(*1*)
YHL9176	YHL912 with pHL423 and pHL2577	
YHL9451	YHL912 with pHL2577	This study
YHL9609	YHL 9451 with pHL2578	This study
YHL9799	yBN146 with pHL2808	This study
YHL9800	yBN147 with pHL 2809	This study
YHL9801	yBN148 with pHL2810	This study
YHL10044	yBN149 with pHL2825	This study
YHL10045	YHL10046 with pSLF124	
yBN24	h- ade6-M210/ade6-M216 ura4-D18 leu1-32 SPBC21C3.09c::KanMX4/ SPBC21C3.09c	Bioneer Corporation collection BG_3108
yBN146	h+/h+ ade6-M210/ade6-M216 ura4-D18/ura4-D18 leu1-32/leu1-32 mmf1::KanMX4/mmf1	Bioneer Corporation collection BG_3366
yBN147	h+/h+ ade6-M210/ade6-M216 ura4-D18/ura4-D18 leu1-32/leu1-32 mrpl16::KanMX4/mrpl16	Bioneer Corporation collection BG_2355
yBN148	h+/h+ ade6-M210/ade6-M216 ura4-D18/ura4-D18 leu1-32/leu1-32 mrpl19::KanMX4/mrpl19	Bioneer Corporation collection BG_0301
yBN149	h+/h+ ade6-M210/ade6-M216 ura4-D18/ura4-D18 leu1-32/leu1-32 SPBC2d10.08c::KanMX4/SPBC2d10.08c	Bioneer Corporation collection BG_3334

1. H. L. Levin, *Mol Cell Biol* **15**, 3310 (1995).

Table S6 Plasmids.

Plasmid	Markers	Description	Reference
pHL423	amp, LEU2	Rep3 *nmt1* promoter, control plasmid lacking Hermes transposase	(*1*)
pHL2577	amp, URA3	Hermes donor with TIR's flanking *kanMX6*	(*2*)
pHL2578	amp, LEU2	rep81X *nmt1* promoter expressing wt Hermes transposase	(*2*)
pHL2612	amp, Ura4	URA4 cloned into pRO322 vector	(*3*)
pHL2806	amp, LEU2	A 2.2 kb Hind III fragment with LEU2 from REP81X was inserted into the 9.0 kb Hind III fragment of PON177{Kim, 2010 #4464}. This plasmid allows sporulation of Bioneer diploids.	this study
pHL2808	amp, Ura4	mmf1 gene PCR cloned into pHL2612	this study
pHL2809	amp, Ura4	mrpl16 gene PCR cloned into pHL2612	this study
pHL2810	amp, Ura4	mrpl19 gene PCR cloned into pHL2612	this study
pHL2825	amp, Ura4	SPBC2d10.08c gene PCR cloned into pHL2612	this study
pSLF124	amp, Ura4	carries genomic clone of cdc19 to complement disruption of cdc19 in FY243	gift of Susan Forsbrug

1. K. Maundrell, *Gene* **123**, 127 (1993).
2. A. G. Evertts, C. Plymire, N. L. Craig, H. L. Levin, *Genetics* **177**, 2519 (Dec, 2007).
3. C. Adams, D. Haldar, R. T. Kamakaka, *Yeast* **22**, 1307 (Dec, 2005).

Table S7 Oligonucleotides.

Name	Sequence; function
HL1870	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC; linker primer
HL1871	[phos]TAGTCCCTTAAGCGGAG[AmC7~Q]; linker primer
HL2215	AATGATACGGCGACCACCGAGATCT-CTATGTGGCTTACGTTTGCCTG; PCR primer with P5 tag followed by Hermes seq
HL2216	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT-GTAATAACGACTCACTATAGGGC' PCR primer with P7 tag followed by linker seq
HL2434	CTATGTGGCTTACGTTTGCCTGTGGCTTGTTG; custom Illumina sequencing primer
HL2550	ATCTCGAGTCCAGGAGCAGATGCAGCAC; PCR primer to clone mmf1
HL2551	ATGAGCTCGGCGCACGAAAATGAAAAGC; PCR primer to clone mmf1
HL2552	ATCTCGAGCAACGGAGCAGAGAGCAGCA; PCR primer to clone mrpl16
HL2553	ATGAGCTCTGAGGCAAGGGGTGGTAACG; PCR primer to clone mrpl16
HL2554	ATCTCGAGTCGTCGTCGTCGTGATCCTC; PCR primer to clone mrpl19
HL2555	AACTCTGCATCCCGGGACAA; PCR primer to clone mrpl19
HL2753	GTCGACAGGTTTCCATGCATTAGCTGGGCA; PCR primer to clone SPBC2d10.08c
HL2754	CTCGAGTGCAGCAATCTGAACGACGGCA; GTCGACAGGTTTCCATGCATTAGCTGGGCA PCR primer to clone SPBC2d10.08c