SAP1 DEPENDENT REPLICATION FORK BARRIERS GUIDE INTEGRATION OF LTR RETROTRANSPOSONS IN *S. POMBE*

By

JAKE ZACHARY JACOBS

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Biochemistry

Written under the direction of

Dr. Mikel Zaratiegui, Ph.D.

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2016

ABSTRACT OF THE DISSERTATION

SAP1 DEPENDENT REPLICATION FORK BARRIERS GUIDE INTEGRATION OF LTR RETROTRANSPOSONS IN *S. POMBE*

By JAKE ZACHARY JACOBS

Dissertation Director: Dr. Mikel Zaratiegui

Long Terminal Repeat retrotransposable elements (LTR-TEs) are a large group of eukaryotic Transposable Elements characterized by flanking repeats in tandem orientation—the LTRs. The LTRs of these elements contain sequences that recruit proteins involved in their expression, replication, silencing, organization, and stability. A successful transposable element must maximize its reproductive amplification without jeopardizing its host, and several characterized LTR-TEs appear to accomplish this through the selection of integration sites away from protein coding sequences. However, despite their high relatedness, a universal mechanism that explains how these parasitic elements avoid coding sequences has not been established. Through sequencing of *de novo* integration sites of the LTR-TE Tf1 from the fission yeast *Schizosaccharomyces pombe*, we found a strong integration preference for locations near the binding site of Sap1. Sap1 has been previously shown to be a DNA-binding protein that controls the directionality of DNA replication by causing polar fork arrest. Sap1 mutations that mildly affect binding but strongly affect fork barrier activity decrease Tf1 retrotransposon efficiency ten-fold, indicating that Sap1 replication fork barrier activity is a stronger predictor of Tf1 integration than DNA binding. Further, synthetic Sap1 binding sites placed near DNA origins are only competent at Tf1 recruitment when placed in blocking orientation. Interestingly, the fork arresting activity of an independent factor provided *in*

cis can increase the integration efficiency of a barrier-incompetent Sap1 binding site. Thus, both Sap1 binding and replication fork arrest are necessary for Tf1 integration. Together, these data suggest that Sap1 guides insertion of Tf1 by tethering the intasome and blocking the progression of the replication fork, and that the Tf1 transposon uses features of arrested forks to insert into the host genome. Since fork arrest is detectable in many genomic features that recruit LTR-RT integration, such as type III promoters and heterochromatic sequences, these observations point to a universal mechanism for determination of LTR-TE tropism.

The questions surrounding the molecular mechanism of Tf1 transposition led to the examination of the CRISPR/Cas9 system as a tool for tethering Tf1 to stalled forks *in vivo*. However, the CRISPR/Cas9 toolkit had not been developed for *S. pombe*. Using a novel processed RNA Pol II promoter and the Hammerhead ribozyme we developed a highly efficient CRISPR/Cas9 expression system, leading to >95% modification efficiencies without selection.

Dedication

To all the scientists who came before me, and the millions of hours of thinking and research they performed. Without them, the questions examined in this thesis would not have even been possible to ask.

Acknowledgements

My thesis would not have been nearly as successful or interesting if it were not for the strong support network that helped me through the work presented here. First, I'd like to thank Dr. Zaratiegui for being an absolutely incredible mentor. His experimental support, daily input, and passion for science radiates throughout his laboratory, creating a wonderful atmosphere for scientific discovery.

I also would like to thank the other lab members who were critical in the work in this thesis. Vincent, for teaching me the art of the experiment, the importance of never believing your own results, and the value of persistence. Keith, for stepping up and learning a ton of biochemistry and molecular biology when the time was needed, and for assisting with CRISPR/Cas9 experiments. Susanne, for valuable experimental support during the review process and other great discussions. Jesus, for tremendous experimental support, helpful discussions, and importantly, performing, troubleshooting, and analyzing the 3C experiments (Figure 13).

Importantly, I also thank my incredibly supportive family. First, my parents who constantly supported my childhood curiosity, and never let it fade. Also, my wife Julie, who remained supportive through the whole process, and never asked too much of me when I was burdened. Also, my dog Wilbur and newborn baby Jordan, who greeted me with much needed excitement and joy every time I came home.

As a final note, it should be noted that the work contained within Chapter II and V are mostly reproduced works from scientific publications completed as part of this thesis (Jacobs, Ciccaglione, Tournier, & Zaratiegui, 2014; Jacobs et al., 2015).

Table Of Contents

List Of Tables

List Of Illustrations

Chapter I - An Introduction to Transposons and CRISPR/Cas9

1. Transposons

1-1. The Biological Impact of Transposon Activity

The success of biological life on earth is a direct result of the astounding ability of molecules to self-organize and replicate with just the right amounts of fidelity and error. This error creates diversity, and diversity allows organisms to happen-upon genetic changes that allow them to conquer new environments, outcompete other species, restrict parasites, improve their reproductive rate, and explore other phenotypes that improve their fitness in their ecological niche. These errors are the substrate upon which natural selection acts; without them, there would be no biological diversity.

With the knowledge that error drives diversity, it is difficult to imagine that biologists ever believed that the Eukaryotic genome was largely static. During this time, the recent synthesis of genetics and cytology had allowed biologists to directly observe genome dynamics like recombination, deletion, and inversion, but it was generally believed that a genes relative location in the genome itself was fixed (Benzer, 1956). However, this dogma was dramatically challenged in the middle of the 20th century when Barbara McClintock's discovered transposons, genes capable of movement around the genome. Her discovery arose from the study of mosaicism in kernels of corn, where she realized that variegated pigmentation was entirely explained by the mobility of a gene called Dissociation (Ds). Ds mobilized in a subset of cells within the kernel, pigmentation alleles were lost, resulting in mosaicism (Barahona, 1997).

Transposable elements (TE), like Ds discovered by McClintock, are defined as

any segment of DNA with the capacity to change its location in the genome. TEs that move by excising themselves and reinserting elsewhere (cut and paste) are classified as class II TE, while those that use a transcribed RNA intermediate (copy and paste) are class I TE (Wicker et al., 2007). TEs are found in all known Eukaryotic organisms, and in several cases, make up the majority of the DNA sequence of the genome—approximately half of the sequence in the human genome is TE derived, while approximately 85% of the maize genome is TE (Feschotte & Pritham, 2007; Schnable et al., 2009).

Since the discovery that they dominate the majority of sequence in many eukaryotic genomes, debates have arisen over what role they play in our evolution and genome biology. McClintock's work had already directly shown that transposon mobilization could influence the phenotype of an organism by rearranging nearby genes. Other researchers speculated that the novel insertion sites could lead to novel patterns of gene regulation, particularly during cell development and differentiation (Britten & Davidson, 1969). However, despite the recent observation that TE had been shown to be capable of mutagenesis in the *E. coli* genome (Shapiro, 1969), most biologists at the time believed that transposon mobility was too rare to have a significant impact on the genetic diversity of species (Biémont, 2010).

These concerns dissipated in the late 1970's upon the realization that hyperactivation of DNA transposons called P elements explained how wild male strains of *Drosophila melanogaster* produced sterile offspring when mated with closely related lab strains (Rubin, Kidwell, & Bingham, 1982). Since the lab strains were collected in the early 1900's and mated with wild files collected in the 1970's, it was hard to argue that transposons dynamics are too slow to directly impact the evolutionary trajectory of a

species. Though P-M hybrid dysgenesis cannot be considered biological speciation, as sterility fades with age (Khurana et al., 2011), a similar process between geographically separated species and activation of TE has been implicated in the speciation of hybrid species of sunflower (Ungerer, Strakosh, & Zhen, 2006).

P elements were only the beginning. The great awakening to the of impact TEs on all biological life would come sometime later, when DNA sequencers got to work cataloging the genetic makeup of life on earth. Equipped with thousands of detailed maps of the genome of closely related species, the magnitude of their impact became undeniable. These detailed genomic maps have painted a complex picture of how transposon mobility impacts genomes. Analysis of transposon sequence has revealed a diverse group of transcription factor, enhancer, and chromatin remodeler binding sites. The mobilization of these binding sites by TE can have direct and permanent effects on the expression and regulation of nearby genes (Feng, Leem, & Levin, 2012; Slotkin & Martienssen, 2007). Mobilization of these binding sites also allows for rapid rewiring of regulatory networks. As a few examples of this in action, remnants of degenerate transposon sequence exist within a third of the p53 binding sites in the human genome (T. Wang et al., 2007), SINE element mobilization of CTCF binding sites has wired the developmental network of several mammalian species (Schmidt et al., 2012), and a full 25% of human promoters have sequence related to known TE (Jordan, Rogozin, Glazko, & Koonin, 2003).

In addition to mobilizing regulatory motifs, transposon proteins involved in their movement have also been retooled to serve critical functions in Eukaryotic genomes. Domesticated transposases have been repurposed to serve critical roles in chromatin

silencing, V(D)J recombination, and cell development (Agrawal, Eastman, & Schatz, 1998; Cam, Noma, Ebina, Levin, & Grewal, 2007; Liang et al., 2015). Undeniably, the sequence contribution and mobilization of regulatory elements by TE has dramatically shaped our evolutionary history.

However beneficial transposons have been throughout evolution in enabling the rapid sampling of genetic traits, transposons also have long been known to be capable of causing damage to their hosts. TEs can precipitate disease through several types of genome modifications. First, TE cause mutagenesis by inserting DNA in areas where DNA spacing matters— coding sequences, intron splicing, regulatory motifs, and so on. Second, the regulatory motifs that TE bring, or the host machinery poised to silence it, can also influence nearby gene expression. Lastly, because TE are repetitive sequences, recombination between them can lead to chromosomal translocations, deletions, or inversions. These types of transposon-induced changes have been linked to over 120 distinct human diseases (Hancks & Kazazian, 2016; Solyom & Kazazian, 2012).

Because of the deleterious effects associated with transposon hyperactivation, host genomes have evolved diverse and complex mechanisms to transcriptionally silence transposons through epigenetic processes like DNA methylation, heterochromatin formation, and RNAi (Levin & Moran, 2011). More rarely, some organisms entirely eliminate TE from their transcriptionally active somatic nuclei through a piRNA-mediated excision process (Fang, Wang, Bracht, Nowacki, & Landweber, 2012).

TE walk a thin line between benefitting and damaging their hosts, as examples of both are readily available in nature. Is their prevalence and diversity mostly a function of

their ability to successfully evade a host's best efforts to constrain them, or do hosts tolerate TE as a repository for genetic diversity? The complex evolutionary forces at play that that explains TE diversity and persistence were first addressed through population genetic models. Early models argued that, after an initial colonization phase through either horizontal transfer or reactivation of a dormant TE, the persistence of TE in genomes is best explained through their ability to replicate faster than selection and mutation can act upon them (Charlesworth, Sniegowski, & Stephan, 1994). These 'Red Queen' based models assume that, after initial colonization, TE families persist in genomes in an equilibrium governed by replication-selection dynamics.

However, modern investigations based on simulations that assume random mating and account for copy number variation, mutation, the effect of copy number on host fitness, the variability of the selective impact of an insertion, and variability in transposon activity, reveal that transposon-host dynamics are likely in a non-equilibrium state (Le Rouzic, Boutin, & Capy, 2007). These dynamics result in several invasion and decline cycles, occasionally leading to a repurposing of the transposon sequence to a beneficial function (domestication) before the family is mostly lost to genetic drift. While these models are difficult to experimentally test on any realistic time-scale, they do seem to agree with the snapshot biological diversity that exists in our vast genome sequencing databases.

These models provide an important insight into some of the evolutionary forces TE are subject to when invading new genomes, but all make the assumption that the beneficial impact of transposition relies on random insertion. In the 1980's, McClintock theorized that transposon activity could reshape genomes in the face of environmental

stress (McClintock, 1984), and since then, several families of transposon (mostly class I) have been shown to mobilize in response to environmental stressors (Capy, Gasperi, Biémont, & Bazin, 2000). Of course, the activation of transposons in response to stress could be due to a generalized genome derepression or a hijacking of the stress-response pathway for activation, but studies suggest that this activation could have co-evolved. For example, in the fission yeast *S. pombe*, transposon insertion nearby heat shock genes enhances their transcriptional activation upon heat, but does not enhance the transcriptional activation of genes not involved in the heat shock response (Feng et al., 2012). Analysis of 161 natural isolates of *S. pombe* further suggests that this pathway is active in nature—intra-population transposon content differences were enriched nearby stress response or highly expressed genes (Jeffares et al., 2015). These results suggest that the transposon is cooperating with transcriptional activation pathways, and that transposon integration could lead to adaptation to stress. If a genome is able to better adapt to stress through the activation of transposons, it would challenge long-established Luria-Delbruck models of selection which posit that mutations not as a response to selection, but in the absence of it (Luria & Delbrück, 1943). However, if transposons really can shape genomes in response to selection, it would not be the first time transposons broke with paradigms.

1-2. Transposon Mobilization: From Colonization to Target Site Selection

TE are locked in their hosts and can only spread to new genomes through horizontal transfer. This horizontal transfer was directly observed in the wild when the P element spread within related *Drosophila* species like wildfire (Clark & Kidwell, 1997), but sex between closely related species is not the only means of transfer. TE have long been described to exist in large viral genomes, opening the possibility that they could be horizontally transferred between diverse species (D. W. Miller & Miller, 1982). TE also spread across kingdoms through host-parasite interactions. Interactions like these have been implicated in the horizontal transfer of at least 4 transposon families between vertebrates and invertebrates (C. Gilbert, Schaack, Pace, Brindley, & Feschotte, 2010). For example, a retrotransposon family was found to have spread between nematodes and birds in at least two bursts, demonstrating the power of TE to colonize to effectively colonize the genomes of distantly related species (Suh et al., 2016). TE can also horizontally spread in bacteria through the use of their own conjugative plasmids. This massive family of transposable element has been shown to disseminate antibiotic resistance within bacterial populations (Salyers, Shoemaker, Stevens, & Li, 1995). Indeed, every studied gene transfer mechanism has transposons hitchhikers, invading new genomes to improve their odds of survival through proliferation and expansion.

TE do not just amplify in genomes through horizontal transfer, but also, through reactivation of old or non-autonomous elements within their host genome. For example, the mobilization of a large family of hominid specific non-autonomous TE, the SVAs (SINE/VNTR/*Alu*), are entirely dependent on the expression of LINE-1 (L1) element transposition machinery, and actively compete with L1 elements for these transposition

factors (Doucet, Wilusz, Miyoshi, Liu, & Moran, 2015; Raiz et al., 2012).

Once a TE enters a new genome, spread of the TE within the genome is dependent on its ability to mobilize to new areas that do not negatively affect host fitness. This is especially true for small population sizes in asexual populations, where any deleterious effect on fitness leads to rapid loss of the TE (Dolgin & Charlesworth, 2006). To this end, most TE favor integration into regions of the genome that have minimal effects on fitness. Some target gene-dense regions of the genome like promoters, but others target regions like tRNAs, rDNA, heterochromatic regions, or telomeric repeats (Levin & Moran, 2011). Transposons that integrate into gene-dense regions tend to exhibit strong preferences for regions that do not disturb ORFs. For example, the LTR retrotransposon Tf1 from *S. pombe* is primarily targeted to RNA Pol II transcribed gene promoters (Behrens, Hayles, & Nurse, 2000). *Drosophila* P elements were once thought to target gene promoters, but recent reports have revealed that these DNA transposons likely preferentially target DNA replication origins, which often overlap with regions of open chromatin and promoters (Bellen et al., 2011; Spradling, Bellen, & Hoskins, 2011). Since P elements are DNA transposons that cannot directly copy themselves through an RNA intermediate, this preference for origins of replication allows them to increase their copy number by jumping ahead of moving replication forks. It is not currently understood how transposons can target gene promoters, but this targeting strategy is widely used within the mobile element kingdom, as retroviral elements like Human Immunodeficiency Virus (HIV) and Murine Leukemia Virus (MLV) also exhibit similar preferences for Pol II transcribed genes (Mitchell et al., 2004).

Transposon target site selection has been extensively studied in some members of

the Ty family of TE found in *S. cerevisiae*. This is particularly true for the Ty1/*copia* and Ty3/*gypsy* families, both of which exhibit a strong preference for regions upstream of RNA Pol III promoters, like tRNAs (Lesage & Todeschini, 2005). Ty1 integrates within nucleosome-occupied regions ~700 bp region upstream of tRNAs, and accomplishes this specific targeting through an interaction between its transposase and the AC40 subunit of RNA Pol III (Baller, Gao, Stamenova, Curcio, & Voytas, 2012; Bridier-Nahmias et al., 2015). Ty3 insertion is similarly directed to tRNA genes, but through distinct mechanisms. Ty3 inserts just 2-3 base pairs upstream of tRNA transcription start sites through direct interaction with the RNA Pol III transcription factors TFIIIB and TFIIIC (Kirchner, Connolly, & Sandmeyer, 1995). Distantly related families evolving distinct and independent mechanisms for targeting the same genomic 'safe harbor' is an elegant example of convergent evolution, and a recapitulation of the Hippocratic oath of TE mobilization: first, do no harm.

TE need not just jump into the promoters of genes for success. Another well-studied family of transposon in *S. cerevisiae,* the LTR retrotransposon Ty5/*gypsy* exhibits strong preferences for heterochromatin through a direct interaction between its integrase and Sir4p, a protein critical for the formation of heterochromatin (Zou, Ke, Kim, & Voytas, 1996). Poorly characterized transposons in organisms other than yeast also exhibit preferences for gene-poor regions. For example, the non-LTR HeT-A and TART transposons in *Drosophila melanogaster* exhibit a strong preference for telomeres, and their repeated integration into chromosomal ends replaces the role of telomerase plays in most Eukaryotic genomes (Pardue, Danilevskaya, Lowenhaupt, Slot, & Traverse, 1996).

While the vast majority of TE show particular insertion site preferences, not all

appear to specifically utilize an interaction between host and transposon binding factors. Some TE have more general preferences for integration sites, and are largely influenced by nucleosome occupancy, CpG islands, or particular conformations of DNA (Gangadharan, Mularoni, Fain-Thornton, Wheelan, & Craig, 2010; Pryciak & Varmus, 1992; Yant et al., 2005). Other TE, like the L1, SINE, and *Alu* elements found in hominid genomes, have integration profiles and mechanisms that are poorly understood. The studies that have either mapped relatively low number of novel L1 insertions, however, have generally found that L1 insertions are random, and do not avoid ORFs (Beck et al., 2010; Ovchinnikov, Troxel, & Swergold, 2001).

Retroviruses, which are closely related to LTR-TE, also seem to use specific host factor-integrase interactions to target insertions toward specific areas of the genome. However, because these elements are not locked into their host genomes, insertions only need to occur in regions that ensure productive expression, not in areas that minimize their impact on host fitness. As two examples of this targeting, MLV uses an interaction with BET proteins to direct integration into enhancer sequences of genes (Sharma et al., 2013), while HIV uses an interaction with LEDGF to direct its integration to the bodies of highly expressed RNA Pol II genes (Ciuffi et al., 2005). These integration preferences appear to be entirely mediated through their integrases, as experiments swapping HIV and MLV integrase also swap their corresponding integration preferences (Lewinski et al., 2006).

Perhaps the large differences between the integration profiles of TE within genomes are a function of the ecology of the genomes from which they arise. In gene dense genomes like yeast and bacteria, the vast majority of studied native transposons

exhibit tight integration profiles. As an extreme example, the Tn7 transposon is largely contained to a single location in the *E. coli* genome through direct tethering by its DNA binding protein TnsD to *attTn7* (Kuduvalli, Rao, & Craig, 2001). High-throughput insertion analysis of TEs from yeast generally reveal ~90% specificity to their preferential regions, whether it be RNA Pol III genes, heterochromatin, or gene promoters (Baller et al., 2012; Gai & Voytas, 1998; Yabin Guo & Levin, 2010; Lesage & Todeschini, 2005). However, in mammalian or plant genomes with low gene density, this high level of specificity seems to be largely lost, and most TE in these genomic contexts appear to integrate much more liberally, occurring either randomly or in highly-dispersed regions, like those of increased DNA flexibility (Beck et al., 2010; Vrljicak et al., 2016). It is possible that these differences reflect the selective pressure TE are under to avoid coding sequences. Alternatively, these dynamics could be a result of an evolutionary arms race between host silencing machinery and TE targeting factors, like is seen with viruses and their respective host restriction factors (Compton, Malik, & Emerman, 2013). Either way, like molecular versions of Darwin's finches, TE are dramatically shaped by their ecosystem.

In light of the highly specific targeting mechanisms some TE use to avoid ORFs in yeast and bacterial genomes, how can these TE be successful when they spread to new genomes? One recent study revealed that these specific targeting mechanisms might not be necessary for successful colonization and ORF avoidance. In this study, abolishing the interaction between Ty1 and its targeting factor does not lead to full random integration, instead, the transposon is redirected to subtelomeric regions (Bridier-Nahmias et al., 2015). Further, expression of the LTR retrotransposon Tj1 from the fission yeast

Schizosaccharomyces japonicus to the naïve environment of the related *S. pombe* genome shows similar avoidance of coding sequences, and integrates in close proximity RNA Pol III genes, similar to the Ty3/*gypsy* family from budding yeast (Yabin Guo, Singh, & Levin, 2015b). Further, loss of the tethering factor Sir4p redirects Ty5 integrations to rDNA and Ty elements (Zhu, Zou, Wright, & Voytas, 1999). These striking studies raise the possibility that some TE may have secondary mechanisms for ORF avoidance.

It is possible that this secondary targeting ability is what allows TE to readily colonize new genomes. If other TE display secondary targeting preferences, what role do these primary targeting pathways play in host or TE fitness? On one hand, it is possible that these primary pathways are a result of TE domestication or co-evolution, and the presence of TE at these loci benefit the host under certain conditions. Alternatively, these primary preferences could be a direct result of parasite-host dynamics, like TE avoiding silencing or eviction by its host genome.

1-3. LTR Retrotransposons: Structure and Function

LTR retrotransposons are a massive family of class I TE that is characterized by two long terminal repeats (LTRs) that flank their internal coding sequence. LTR-TE are both structurally and functionally related to the well-studied retroviruses like HIV, MLV, and Prototype Foamy Virus (PFV), but lack extracellular infectivity pathways encoded by *env* genes (Peterson-Burch & Voytas, 2002). It is not known whether TE gave rise to retroviruses through gain of *env*, or retroviral integrations into the genome became trapped by loss of *env*.

Like all retrotransposons, LTR-TE propagate in their host genomes through reverse-transcription of their RNA into a cDNA molecule that can insert elsewhere in the genome (Figure 1). This process has been examined in detail for the two large families of LTR-TE that exist in yeast, the *Metaviridae* (Ty3/*gypsy*) and *Pseudoviridae* (Ty1/*copia*) transposon families (Havecker, Gao, & Voytas, 2004). First, the transposon RNA is transcribed by the host's RNA Pol II from a promoter that lies within the 5'LTR. Second, after mRNA export and translation, the long polyprotein is cleaved into *gag* and *pol* encoded components parts by its own protease. The *gag* gene encodes for a single protein, Gag, an important structural protein (Atwood, Lin, & Levin, 1996). The *pol* gene encodes three critical functional components: protease (PR), reverse transcriptase (RT), and integrase (IN). Third, Gag self-assembles into a virus-like particle (VLP), a small capsule that organizes and coordinates the subsequent steps (Atwood et al., 1996). Most importantly, this involves holding in place the transposon RNA, and several reaction intermediates, during conversion of the RNA to cDNA by RT. Lastly, IN binds to the ends of the cDNA, forming the preintegration complex (PIC or intasome) and this

complex is imported into the nucleus through a poorly understood Gag-dependent process (Teysset, Dang, Kim, & Levin, 2003).

For VLP formation to occur, Gag needs to be in significant molar excess to Pol, and LTR-TE vary in their strategies to achieve this necessary ratio from a single RNA and promoter. This aforementioned separation of LTR-TE genes into *gag* and *pol* is done by the LTR-TE to regulate the relative ratios of these components. Approximately half of LTR-TE, regardless of family, encode the *gag* and *pol* genes in a single reading frame, and use a targeted degradation process to ensure a high ratio of Gag to Pol (Atwood et al., 1996; X. Gao, Havecker, Baranov, Atkins, & Voytas, 2003). Members of the *Metaviridae* generally use a slippery ribosomal site that occasionally promotes a -1 ribosomal frameshift to produce the *pol* genes (X. Gao et al., 2003). Interestingly, the method used to regulate the method of *pol* expression is more of a function of the host than the transposon family (Havecker et al., 2004). This relationship may be the result of silencing mechanisms that target these priming events.

After the ratio of Gag to Pol is achieved, the VLP is formed through the self-assembly of the Gag protein. In most LTR-TE, the Gag protein is processed into smaller nucleocapsid proteins that restrain the transposon RNA to the VLP through RNA binding motifs. These motifs are widely conserved among LTR-TE, and usually take the form $Cx_2Cx_4Hx_4C$ (Peterson-Burch & Voytas, 2002). With the Gag nucleocapsid formed around the transposon RNA, RT begins the complex process of reverse transcription. RT is a highly conserved enzyme within the all LTR-TE, and is usually functionally separated into its RNA-dependent DNA polymerase activity and its RNase H activities (Xiong & Eickbush, 1988). Because of this high level of conservation, its sequence forms

the basis of all LTR-TE classification (Peterson-Burch & Voytas, 2002).

Since transcription both begins and ends inside the transposon LTR sequences, reverse transcription of the RNA into a full-length cDNA with intact LTRs is a multi-step, complex, and tightly regulated process. This process classically begins with a cellular tRNA priming to a sequence located close to the 5' LTR called the primer binding site (PBS), and reverse transcription to the 5' end of the 5' LTR fragment. However, some LTR-TE, like the *S. pombe* Tf family, carry out their own priming. In this process, the 5' end of the LTR loops back and binds to a complementary sequence, where the RNase H activity of RT reveals a 3'OH to begin priming (Levin, 1995; 1996). Through either this self-primed or tRNA primed reverse transcription, the first reaction intermediate is called the (-) strong-stop cDNA. This short ssDNA is then transferred to the 3' end of the transposon RNA transcript, where it anneals to the 3'LTR and forms the majority of the first strand of the cDNA, only lacking sequence 5' of the PBS. A second priming event initiated from a polypurine tract (PPT) near the 3'LTR then synthesizes the double-stranded 3'LTR and continues synthesis back into the PBS, creating the (+) strong-stop cDNA. This plus strand strong stop DNA then initiates a second transfer event back to the 5' end of the nascent cDNA, and filling in of the 3' ends results in the full length cDNA (Lauermann & Boeke, 1997; M. Wilhelm, Heyman, Friant, & Wilhelm, 1997).

Once the VLP completes synthesis of the cDNA, molecules of IN are bound to the conserved 5'CA cDNA ends and Gag mediates nuclear import of the cDNA/IN complex, where it is free to integrate into the host genome (Teysset et al., 2003). Integration in the genome is generally mediated through the IN, but cDNA can also

recombine into the genome with considerable frequency, especially if existing homologous elements are present in high copy number in the host genome {Hoff:1998wv}. LTR-RT INs are members of the $DDX_{35}E$ superfamily of IN. Integrase in both *Pseudoviridae* and *Metaviridae* contain three domains—the N-terminal zinc finger binding motif (called the HHCC or NTD), the catalytic core domain (CCD) containing the characteristic $DDX_{35}E$ catalytic core, and the C terminal domain, which usually adopts an SH3-like fold (Hare, Gupta, Valkov, Engelman, & Cherepanov, 2010; Peterson-Burch & Voytas, 2002). While the HHCC and CCD domains are highly conserved among LTR-RT, the C-terminal domain differs considerably between the two families, with *Pseudoviridae* harboring a GKGY motif, and the *Metaviridae* harboing a GPF/Y motif (Peterson-Burch & Voytas, 2002). The residues within the catalytic core domain (CCD) coordinate Mg^{2+} or Mn^{2+} to catalyze the strand transfer of the 3'OH ends of the cDNA to two staggered cuts in the genome, usually separated by 4-5 base pairs (Hare et al., 2010; Levin, Weaver, & Boeke, 1990). These staggered cuts and their resulting repair by host DNA repair machinery result in the characteristic target site duplication (TSD) that is associated with *bona fide* LTR-TE insertions. Crystal structures of closely related retroviral IN, like that of the prototype foamy virus (PFV) suggest that LTR-TE INs bind their cDNA as a tetramer and all three conserved domains (NTD, CCD and CTD) participate in critical protein-protein or protein-DNA interactions to create the functional intasome (Hare et al., 2010). Importantly, these crystal structures also strongly suggest that all three conserved domains can participate in target DNA binding.

Additionally, some *Metaviridiae* LTR-TE contain a C-terminal chromodomain (CHD), a histone binding motif known to bind methylated histones (Malik & Eickbush,

1999). These chromodomains are not always critical for TE targeting (Chatterjee et al., 2014) but can sometimes guide insertion site selection (X. Gao, Hou, Ebina, Levin, & Voytas, 2008). Currently, due to their low conservation and lack of structural data, very little is known about the role of CHD in TE targeting.

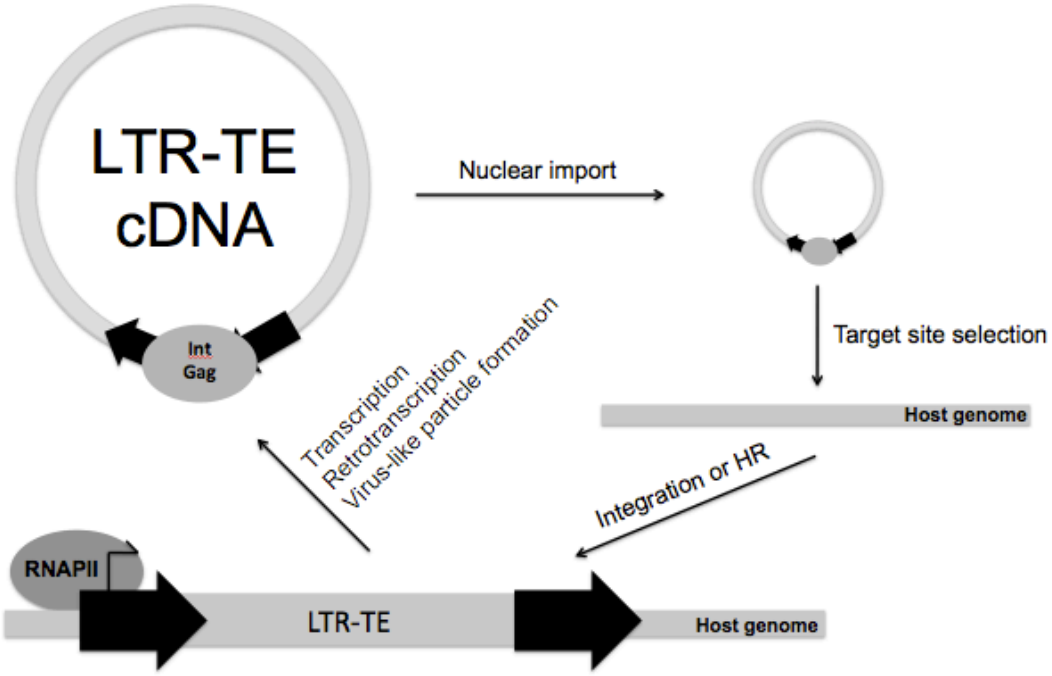


Figure 1 - The LTR-TE lifecycle. First, RNA polymerase II transcribes the LTR-TE mRNA from genomic copies of LTR-TE from a promoter located within the 5'LTR. Translation of this mRNA in the cytoplasm of the cell results in a long polyprotein, and encodes for components that carry out the subsequent steps of reverse transcription, nuclear import, and integration.

1-4. LTR Retrotransposon Target Site Selection in *Schizosaccharomyces pombe*

The fission yeast *Schizosaccharomyces pombe* contains only a single family of LTR retrotransposons within the *Metaviridae (*Ty3/*gypsy*) family, Tf, consisting of Tf1 and Tf2. In laboratory strains, this family is present as 13 full-length Tf2 elements, ~250 solo-LTR, and 5 Tf fragments, covering approximately 1.1% of the genome (Bowen, 2003; Levin et al., 1990). Biochemical analysis reveals that these full length Tf2 are capable of retrotransposition, and RNA-seq experiments show low levels of expression, indicating Tf2 is presently active in these lab strains (Hoff, Levin, & Boeke, 1998; Mourier & Willerslev, 2010). In addition to full-length elements, the *S. pombe* genome also contains ~250 solo LTR fragments, remnants of past insertions that were evicted due to processes like intra-LTR recombination. These relics of past transposon activity reveal that the retrotransposon Tf1 was also once present and active in full-length form in the annotated strain, with 28/250 bearing close resemblance to LTRs sequenced from strains of *S. pombe* where Tf1 is still active (Bowen, 2003). The remainder of the solo LTR bear either significant homology to active Tf2 (60/250), are solo LTR from lineages of Tf that are now extinct, or are the result of solo-LTR duplications in the subtelomeric repeats (Bowen, 2003).

The sequence of full-length Tf1 is known from a wild strain of *S. pombe,* NCYC 132 (Levin et al., 1990; Levin & Boeke, 1992). Tf1 is virtually identical to Tf2 in its IN and RT regions, but is divergent in its *gag,* PR, 5'UTR, and in the U3 region of its LTRs (Hoff et al., 1998). Early studies examining the mobilization of Tf2 by overexpression in laboratory strains revealed that Tf2 inserts ~10-20 times less efficiently than Tf1. Moreover, 70% of the mobilization events are a result of homologous recombination

(HR) with existing Tf2 elements (Hoff et al., 1998). By comparison, >95% of Tf1 mobility is dependent on the presence of IN (Levin, 1995). These large mobility differences are partially explained by the increased efficiency of Tf1 at carrying out the steps involved in retrotransposition. Despite similar RNA levels, Tf1 produced 4 times more cDNA and IN than Tf2, likely due to large differences in the PR activity between the strains (Hoff et al., 1998). However, these differences in protein levels do not explain why most integration events occur independent of IN. Native genomic copies of Tf2 mostly contain TSDs (12/13), indicating Tf2's spread was primarily mediated through IN-dependent pathways. One simple explanation is the increased HR seen in experiments is due to the presence of hundreds of homologous sequences within the genome. However, a more intriguing possibility is that the increased HR rate of Tf2 cDNA is an attempt by the transposon to homogenize Tf2 sequences within the genome, protecting them from mutation.

Whatever the reason for Tf2's low frequency of IN-mediated integration, the increased retrotransposition activity of Tf1 has made it an attractive target of studies examining transposon site selection in *S. pombe*. In these studies, an antibiotic-resistance tagged Tf1 (Tf1-*neo*) is expressed from an inducible promoter, antibiotic-resistant clones are selected, and inverse PCR, Southern blot, or high-throughput sequencing is performed to determine the location individual insertions (Behrens et al., 2000; Yabin Guo & Levin, 2010; Levin, Weaver, & Boeke, 1993).

Early low-throughput characterization of a relatively large number of Tf1 insertions by inverse PCR revealed a strong preference for regions 100-420 nucleotides upstream of RNA Pol II transcription start sites (TSS) (Behrens et al., 2000; Singleton &

Levin, 2002). The sequence requirements of this preference were first directly examined using target sites located on autonomously replicating plasmids. In these important studies, Tf1 was found to target the nucleosome-free regions within the promoters of 5 tested class II genes (Leem et al., 2008). Deletion analysis of these promoter constructs, and later studies that ramped up expression with the transcriptional activator LexA, reveal that RNA Pol II transcription is not important for plasmid targeting (Leem et al., 2008; Majumdar, Chatterjee, Ripmaster, & Levin, 2011). In one tested gene, *fbp1*, the insertion sites clustered around an upstream activating sequence (UAS1), which was known to bind the activating stress-response transcription factor Atf1. Abrogation of Atf1 or mutation of UAS1 resulted in loss of targeting to the plasmid, and biochemical analysis reveals that the Tf1 integrase and Atf1 co-immunoprecipitate (Leem et al., 2008). However, deletion of Atf1 does not change transposition efficiency genome-wide, indicating that factors other than Atf1 can mediate target site selection (Majumdar et al., 2011).

Later high-throughput analysis confirmed this preference for Pol II promoters. Mapping of ~73k unique insertion sites confirmed a strong preference for nucleosome-free regions upstream of ORFs, with >90% of all mapped integrations occurring in these regions (Yabin Guo & Levin, 2010). This analysis also confirmed that the RNA Pol II transcription machinery was likely not responsible for Tf1 integration, as there was no correlation between transcript abundance and insertion number. However, not all promoters were equally targeted—76% of all mapped Tf1 integration events occurred in just 20% of the available intergenic regions, indicating that some promoters can guide integration better than others (Yabin Guo & Levin, 2010). Interestingly, these highly-

targeted regions are enriched for genes that are activated upon exposure to a variety of environmental stressors (Yabin Guo & Levin, 2010). A more recent analysis of Tf1 insertion site preferences using serial-tagged Tf1 identified ~1.1 million independent insertion sites, and recapitulated the strong preference for Pol II promoters (Chatterjee et al., 2014). This study also further revealed that the most targeted regions of the genome have a stronger sequence signature than all genomic insertions, indicating that particular DNA sequences guide target site selection of Tf1 (Chatterjee et al., 2014).

Studies with plasmid traps have also examined the Tf1 IN domains responsible for plasmid targeting, and have shown that the chromodomain (CHD) of the integrase is also essential for efficient targeting to all tested promoters (Chatterjee, Leem, Kelly, & Levin, 2009). Tf1 expression plasmids lacking the IN CHD transpose 14 times less frequently than wild-type IN, indicating that CHD is important for either target site selection or integration efficiency (Chatterjee et al., 2009). This loss of targeting is unlikely due to a change in catalytic activity, as *in vitro* studies of recombinant IN lacking CHD show dramatically increased catalytic activity (Hizi & Levin, 2005). Importantly, while IN and cDNA levels are virtually identical in both strains, IN binding to cDNA ends is reduced 3 fold without CHD (Chatterjee et al., 2009). This drop of binding could partially explain some, if not all, of the loss of Tf1 transposition frequency. However, when IN-lacking CHD strains were subject to high-throughput analysis, insertion profiles were virtually identical to wild-type Tf1 (Chatterjee et al., 2014). This shows that plasmid based assays are limited in their ability to accurately reflect what occurs in the genome.

Recently, examinations of factors involved in the silencing of Tf family members revealed another factor associated with Tf insertions, Sap1 (Zaratiegui et al., 2010). Sap1 is an essential DNA-binding protein that controls replication polarity by causing replication fork arrest in one orientation (Krings & Bastia, 2006). Sap1 has been previously implicated in checkpoint signaling (C. Noguchi & Noguchi, 2007), chromosome segregation (de Lahondes, Ribes, & Arcangioli, 2003), mating-type switching (Arcangioli & Klar, 1991), fork arrest at rDNA repeats (Krings, 2005), and is the main determinant of nucleosome free-regions in fission yeast (Tsankov, Yanagisawa, Rhind, Regev, & Rando, 2011). Sap1 ChIP-seq profiles show that it is primarily bound to the promoters of genes, Tf1 and Tf2 LTRs, rDNA, and at the mating-type locus (Zaratiegui et al., 2010). These binding regions are similar to previously described integration preferences of Tf1 (Yabin Guo & Levin, 2010).

Sap1 binds as a homopolymer to 5-bp repeats at various locations in the genome, and this mode of this binding determines its activity as a fork barrier (Krings & Bastia, 2006). Low-resolution structures of Sap1 obtained from small-angle X-ray scattering suggest that it binds to DNA as a T-shaped dimer, and previous *in vitro* gel mobility shift assays shows that these dimers can tetramerize (Bada, Walther, Arcangioli, Doniach, & Delarue, 2000; Ghazvini, Ribes, & Arcangioli, 1995). This tetramerization activity may function to mediate long range interactions in the genome. *In vitro* double-stranded oligo pull-down studies with purified Sap1 have identified the consensus sequence of Sap1 is a 5-bp direct repeat separated by 5 nucleotides, called DR2 (Ghazvini et al., 1995). Well-characterized Sap1 binding sites within the genome include TER1 (RFB1) at the rDNA, SAS1 at the mating-type locus, and within Tf1/2 LTRs (Arcangioli & Klar, 1991; Krings,

2005; Zaratiegui et al., 2010). At TER1, Sap1 binds a sequence that is similar to DR2 and causes strong fork arrest (Mejia-Ramirez, Sanchez-Gorostiaga, Krimer, Schvartzman, & Hernandez, 2005). At SAS1, Sap1 binds 5-bp inverted repeats separated by 12 nucleotides; this binding does not cause fork arrest (Dalgaard & Klar, 2000). The Sap1 binding site at the LTR is likely a direct repeat, and has been shown to cause fork arrest on plasmids (Zaratiegui et al., 2010).

Plots of Sap1 enrichment around solo Tf LTR's show additional peaks outside of the LTR, indicating that the Sap1 binding site predated the transposon insertion (Zaratiegui et al., 2010). Reinforcing this link between Sap1 binding and Tf1 tropism, Tf1 insertion points are also significantly enriched for Sap1 binding sites (Zaratiegui et al., 2010). Interestingly, a full ~66% of Sap1 peaks identified from the ChIP-seq data contain no detectable Tf1 insertion, and a mere 23% of Sap1 binding sites explain 99% of all insertion into Sap1 enriched regions (Hickey et al., 2015). These observations suggest additional requirements of Tf1 recruitment outside of Sap1 binding.

As mentioned previously, not all Sap1 binding sites are created equal—only some are capable of fork arrest. Tf1 insertion into the mating-type locus provides an opportunity to look at the effects of Sap1 binding and fork barrier activity in isolation and in combination. This region contains a strong Sap1 binding site that does not block the fork (SAS1), and fork barriers RTS1, which harbors a nearby Sap1 binding site, and MPS1, which does not (Arcangioli & Klar, 1991; Krings & Bastia, 2006). Tf1 insertions are enriched at RTS1, a region that both binds Sap1 and blocks the replication fork, and virtually absent at the other two sites that either block the fork without Sap1 (MPS1) or bind Sap1 without blocking the fork (SAS1). Since RTS1 does not depend on Sap1 for its

barrier activity (Eydmann et al., 2008), these results suggest that the fork barrier that determines a transposition hotspot can be furnished *in cis* by an independent fork barrier protein, provided that Sap1 binds in the vicinity. This secondary requirement for efficient insertion could explain why ~77% of the Sap1 binding sites in the genome do not significantly recruit Tf1 insertions, and why only a minority of Sap1 binding sites in the genome explain almost all Tf1 insertion (Hickey et al., 2015).

Other described LTR-TE target sites are associated with fork barriers. For example, budding yeast transposons Ty1 and Ty3 are targeted to RNA Pol III genes, which have a known fork barrier activity (Deshpande & Newlon, 1996), and Ty5 is guided to heterochromatin by Sir4, which is also recruited to stalled forks (Dubarry, Loïodice, Chen, Thermes, & Taddei, 2011). Additionally, several factors linked to Okazaki fragment processing and stalled fork repair have been linked to Ty1 hypermobility phenotypes, including *SGS1, RAD6, ELG1, FEN1, CDC9, RAD52,* and *RAD3* (Lesage & Todeschini, 2005). Further, deletion of a helicase that involved in the progression of replication forks through DNA-protein barriers changes the integration profile of Ty1 (Mularoni et al., 2012). The targeting to replication forks could also function in the avoidance of ORFs, as fork stalling generally occurs within intergenic regions (Dardalhon, de Massy, Nicolas, & Averbeck, 1998). RFB targeting would be especially useful for TE pioneers in new genomes, where specific DNA-protein interactions for ORF avoidance may not exist. In light of these studies, it is tempting to speculate that replication fork stalling may generally enhance the tropism and efficiency of all LTR retrotransposons.

2. CRISPR/Cas9 Genome Editing in *S. Pombe*

2-1. Genome Editing History and Principles

Site-specific modification of the DNA of a living organism to fix genetic disorders or create better disease models has been a dream of scientists since the dawn of molecular biology. Until recently, site-specific modification has only been feasible in simple organisms like bacteria and fungi. However, even in these organisms, modification occurs at low frequency and requires the use of selectable markers. These requirements have limited the utility of genome modification in complex genomes. However, recent developments in engineered nuclease technology such as zinc-finger nucleases (ZFNs), TAL effector nucleases (TALENs), and the CRISPR/Cas9 systems have made site-specific genome modification a reality in virtually all cell types (Sander & Joung, 2014).

Techniques for genome modification can be separated into three categories: addition, subtraction, and replacement/modification. Additive technologies aim to introduce a functional copy of a gene to a cell where the native gene is defective. To avoid mutating the genome, these technologies usually do not make changes to the host's genome. Rather, they rely on episomal vectors or transient transgene expression from linear dsDNA to complement a phenotype (Gaj, Mercer, Sirk, Smith, & Barbas, 2013). Subtractive and replacement technologies are much more technically difficult since they both require modifying a specific area of the genome. Subtractive gene therapy can be used to remove key genes from cancer cells, genes that viruses use to proliferate, or delete a disease-causing dominant mutation. However, by far the most exciting type of gene therapy is through replacement—*in vivo* site-specific mutagenesis. Replacing or site-specifically altering a specific gene avoids many of the problems associated with

additive gene therapy—mainly, gene replacement is permanent, not limited by the size of the target gene, and does necessitate exogenous promoters or regulatory sequences.

Most gene therapies currently on the market or in clinical trials currently use additive techniques. These additive techniques were spearheaded by the use of viral delivery vectors like Retrovirus and Adeno-Associated Virus (AAV), which has shown has shown some promise in the clinic for the treatment of genetic disorders. Still, these additive technologies like AAV have significant drawbacks. First, because AAV vectors do not replace the native copy of the disease gene, these vectors can only be used for complementing a recessive gene. Second, because the therapeutic gene is usually not integrated into a patient's genome, it is lost over time and thus requires continual administration, increasing the odds of anti-vector immunity. Lastly, these vectors can only be used to deliver genes of relatively small size, ~4.2kb, greatly limiting the scope of the conditions they can treat through complementation alone (Gaj et al., 2013).

An attractive alternative to addictive techniques is to introduce a site-specific changes in genome the through stimulation of a host's native homologous recombination (HR) pathway. In the cell, the HR pathway uses homologous sequences to repair single and double strand DNA (dsDNA) breaks that naturally occur in the genome. The system's reliance on homology can be leveraged to integrate desired sequences in the genome. While the process of using HR to introduce changes in fungal and bacterial genomes is as simple as transforming a gene flanked with sequences that specify the desired integration point, early attempts to efficiently do this in higher eukaryotic cells were precluded by the low rate of spontaneous targeting (Malkova & Haber, 2012; Sedivy & Sharp, 1989). This went unchanged until the late 90's, when two independent

groups showed that introducing dsDNA breaks in the genome with the budding yeast homing endonuclease I-SceI enhanced integration of a transgene three to five orders of magnitude, and enabled vector independent and marker-less insertion of arbitrary sequences into mammalian genomes (Jasin, 1996; Sargent, Brenneman, & Wilson, 1997). However, because these I-SceI sites were still being introduced through randomly integrating vectors, this early work was only a conceptual triumph that could not alone deliver on the promise of site-specific genome modification. To accomplish that, researchers still needed to find a way to introduce dsDNA breaks at specific sequences in the genome. But the important discovery that began the revolution had already been made: DSBs enable site-specific modification of genomes.

To understand why DSBs are so useful for genome editing, it is useful to understand what almost all cells do when faced with a DSB. The how and why cells decide to go into one of three different repair pathways is poorly understood, but we do know that it varies greatly between organisms, cell types, cell developmental stages, and also, in the genomic region the break is occurring. The non-homologous end joining pathway (NHEJ) and microhomology end-joining pathways (MMEJ) are both error-prone pathways, and can leave mutations after repairing the genome. The homologous recombination (HR) pathway is the preferred pathway of DSB repair, as it fixes the break error-free, using a homologous sequence as a template. The HR pathway essentially trims back the 5' end of the two ends of the break and uses the 3' overhang to scan the genome for homology, hoping to find another copy of the broken gene on the sister chromatid or the homologous chromosome. Once homology is detected, it is used as a template to copy over the gap introduced by the DSB, cuts are made to untangle the strands, and the DSB

is fixed. This process acts to mediate meiotic recombination between homologous chromosomes, but can also repair DSB caused by genotoxic insults in mitotic cells. Mitotic HR can be co-opted into introducing specific changes by providing an exogenous template for repair.

It is through these pathways that we can decide the nature of the changes we want to make. If we want to inactivate a gene, we can cause a DSB in the protein coding sequence, and hope that error-prone pathways of DNA repair inactivate the gene. If we want to make more targeted changes, we can cause a DSB near where we want the changes, but favor HR by simultaneously introducing a repair template containing the desired changes. With this theoretical framework for site-specific modification firmly in place, the only challenge now was to actually figure out how to create designer endonucleases.

Early proof-of-principle experiments fused naturally occurring non-specific restriction enzyme domains to highly specific DNA binding domains, and were successful at creating dsDNA breaks at targeted sequences (Y. G. Kim, Cha, & Chandrasegaran, 1996; Porteus & Baltimore, 2003). C_2H_2 zinc finger domains were an early attractive target for providing this DNA specificity, spawning what are now known as zinc finger nucleases (ZFNs) (Pabo, Peisach, & Grant, 2001). ZFNs have the disadvantage of not being modular, but their design assumes modularity—C_2H_2 domains are mixed and matched to create target site specificity, and fusion of this array to a nuclease makes the cut. A decade later, another nuclease platform called TAL effector nucleases (TALENs) was developed. TALs are bacterial plant pathogen transcription factors that have simple repeated domains that individually recognize each of the four

DNA nucleotides. By cracking the protein sequences which specify each base, researchers were quickly able to engineer TALENs for virtually any sequence (Moscou & Bogdanove, 2009). Both designer endonuclease technologies have the capability to make highly specific dsDNA breaks in the genome, but researchers currently need to spend months designing, validating, and employing non-traditional cloning techniques to build these systems. Additionally, some domains in the ZFN toolkit have context-dependent interactions with neighboring domains and have to be computationally designed (Maeder et al., 2008). This is not a problem for TALENs, but the highly-repetitive nature of their DNA sequence makes them unsuitable for certain types of viral delivery and in some cases, unstable on bacterial plasmids due to their ability to recombine and rearrange during plasmid replication (Holkers et al., 2013). Further, certain chromatin environments inhibit both ZFN and TALEN cleavage activity, making design of nucleases for methylated regions of the genome virtually impossible in some cases (Christian et al., 2010). Nonetheless, TALEN/ZFN technologies have been used in laboratory settings to correct disease-causing mutations in both transformed and primary cells for X-linked severe combined immune deficiency, hemophilia B, sickle-cell disease, and a handful of others (Gaj et al., 2013).

In addition to ZFN and TALEN technologies, the bacterial CRISPR/Cas9 system has recently emerged as an attractive alternative, mainly due to its simplicity and broader range of applications. In bacteria, the CRISPR/Cas9 system protects cells from transforming DNA and RNA like bacteriophage and conjugative plasmids through the use of an RNA-guided nuclease called Cas9. The Cas9 enzyme complexes with two short RNAs termed the crRNA and trans-activating crRNA (tracrRNA), the former of which

specifies the target DNA to be cleaved. Cas9 will form a DSB at these crRNA-designated loci through simple Watson-Crick base pairing, provided they are also adjacent to a three nucleotide sequence known as a protospacer-adjacent motif (PAM) (Sander & Joung, 2014). The most commonly used version of this system uses the Cas9 from *S. pyogenes* (SpCas9) and a fusion of the crRNA and fixed tracrRNA (guide RNA, sgRNA, or gRNA) to minimize the number of RNAs that must be expressed. For this version, the PAM must be 5'-NGG. Thus, any target DNA sequence with the structure 5' N_{20}-NGG can be targeted for cleavage (Jinek, Chylinski, Fonfara, Hauer, & Doudna, 2012). Demonstrating the portability of this system, only three short years after its initial description in 2012, this CRISPR/Cas9 system has been successfully used in bacteria, mice, rats, monkeys, humans, worms, frogs, yeast, tobacco, zebrafish, fruit flies, and a handful of other model organisms (Hwang, Fu, Reyon, Maeder, & Tsai, 2013; Niu et al., 2014).

The CRISPR/Cas9 system has tremendous advantages over ZFN and TALEN technologies. First, Cas9 is capable of introducing several DSBs at a single time, allowing researchers to create large deletions and inversions, or delete and modify five or more genes in a single shot by transfecting multiple gRNAs (Jao, Wente, & Chen, 2013; J. F. Li, Norville, Aach, McCormack, & Zhang, 2013). This ability is therapeutically important too, and crucial in cases where a genetic disorder is a result of multiple distant mutations, or in cases other cases where several loci need to be changed to achieve a desired end-goal, like simultaneous deletion of a provirus integration site and the non-essential receptor it used to enter. Second, the simplicity of designing gRNAs allows researchers to design large gRNA arrays that will aid in target discovery (Shalem et al., 2014). Third, through there are a few exceptions, Cas9 cleavage in the human genome is

mostly independent of the chromatin landscape (Hsu et al., 2013). However, cleavage is influenced by nucleosome positioning (Isaac et al., 2016).

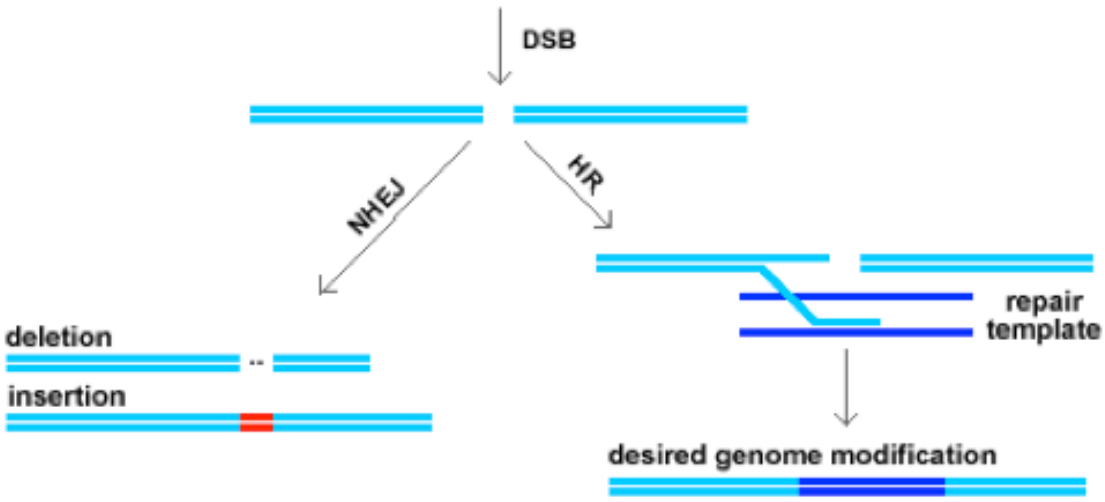


Figure 2 – DSBs enable genome engineering. After a DSB is created through a site-specific nuclease technology, the cell attempts to fix the break via HR. If HR continuously fa ils to fix the cut, as it does in the presence of nuclease repeatedly cutting a desired sequence, NHEJ pathways are invoked, which often result in small deletions or insertions.

2-2. Genome Modification in Fission Yeast

Schizosaccharomyces pombe has long been recognized as a powerful model organism for the study of basic cell and molecular biology pathways, mostly because of their high level of similarity to metazoan cells (Hoffman, Wood, & Fantes, 2015). At the core of this power is their rapid rate at which mutants can be screened, isolated, and characterized.

Early genetic analysis of *S. pombe* was limited to the study genes involved in the ability to grow in the absence of a key nutrient. These auxotrophic markers served as important early tools in complementation studies, and were later used in the maintenance of autonomously replicating plasmids and selection of vectors capable of integration into the genome. Identified genes of interest could be targeted with cassettes containing functional copies of the mutant gene, or complemented with episomal vectors for further analysis. These strategies were great for synthetic recovery screens, copy-number suppression screens, and the identification of other synthetic interactions. For example, the most popular auxotrophic marker, *ura4*, allows for both positive selection via complementation with *S. pombe* Ura4 or *S. cerevisiae* URA3, or negative selection with 5 fluoroorotic acid (5-FOA) (Grimm, Kohli, Murray, & Maundrell, 1988). These markers allow genetic features, or strains, to be tracked relatively easily in experiments with *S. pombe*. However, while these markers have been useful genetic tools, deletions of genes involved in biosynthetic pathways occasionally has unintended consequences for the biology of the yeast, and can confound the interpretation of the phenotype being studied (Hartmuth & Petersen, 2009; Nishino, Kushima, Matsuo, Matsuo, & Kawamukai, 2015). These concerns can be obviated by the use of exogenous antibiotic resistance genes

(Hentges, Van Driessche, Tafforeau, Vandenhaute, & Carr, 2005). However, these antibiotics can be extremely costly, and resistance requires high expression of these genes that may alter the genomic environment of the targeted sequences. This problem can be obviated by clever *Delitto perfetto* methods that remove the marker after the modification has been made (Storici, Lewis, & Resnick, 2001). However, these methods require several expensive sets of primers, multiple transformation steps, and the screening of a large number of colonies.

Until recently, *in situ* modifications to the fission yeast genome were performed using polymerase chain reaction (PCR) gene targeting methods. These methods utilize cassettes flanked by targeting regions homologous, introduced by PCR, to the area of the desired mutation to introduce either large deletions or insertions into the genome. Various laboratories have developed vector based cassettes that can be used for both N and C-terminal gene tagging, deletion analysis, and other desired modifications, using a range of tags and selectable markers (Bähler et al., 1998; Van Driessche, Tafforeau, Hentges, Carr, & Vandenhaute, 2005). These methods have been successful over the years, but for unknown reasons, gene targeting greatly varies (6-63%) in efficiency between different targets, making the creation of some genetic constructs extremely time-consuming (Bähler et al., 1998). Nonetheless, there are methods that researchers have discovered over the years to improve these targeting efficiencies. These methods can include increasing the length of the region of homology (Grallert, Nurse, & Patterson, 1993), starving transformed cells in media lacking nitrogen, or deletion of critical non-homologous end-joining (NHEJ) proteins to prevent aberrant integration events (Fennessy et al., 2014). Whatever the method used, certain types of modification still

remain difficult, require large numbers of oligos, screening large numbers of colonies, and most importantly, are limited by the number of selectable markers available in the strain to be modified.

The recently developed CRISPR/Cas9 system has the potential to circumvent these various problems in gene targeting efficiencies, the difficulties involved in reagent design and availability, and the number of selectable markers. However, the implementation of CRISPR/Cas9 in *S. pombe* faces a hurdle in the difficulty of expressing the sequences of arbitrary RNA with defined 5' and 3' ends, a feature needed for optimal gRNA expression, Cas9 binding, and activity. In human cells, this is achieved through use of a H1 or U6 RNA Pol III promoters, which begins transcription with a defined G (or any purine in the case of H1) and ends with a poly-T tract (Ranganathan, Wahlin, Maruotti, & Zack, 2014). Unfortunately, in both fission and budding yeasts, these analogous RNA Pol III promoters have *cis*-acting internal dependencies that would severely limit the suite of gRNAs to be expressed. This requirement was recently obviated in *S. cerevisiae* with the use of the RNA Pol III snoRNA promoter (SNR52) and terminator (from SUP4) (DiCarlo et al., 2013). Other labs have avoided the problems with RNA Pol III promoters entirely, and have used RNA Pol II promoters to produce gRNA. Because they must not be capped or poly-adenylylated, these expression systems require that the RNA be post-transcriptionally cleaved, and make use of RNA triple helices, micro RNAs, and introns to produce gRNAs with desired 5' and 3' ends (Nissim, Perli, Fridkin, Perez-Pinera, & Lu, 2014). However, these strategies have not yet been applied to fission yeast.

The unlocking of the CRISPR/Cas9 toolkit for *S. pombe* could allow for the rapid

introduction of point mutations and tagged proteins without the need for selectable markers. This implementation will likely have a positive influence on the rate at which *S. pombe* can be genetically manipulated and studied. CRISPR/Cas9 will allow for more rapid and targeted genetic screens, faster mutational analysis. Further, expression of arbitrary gRNA will allow use of the catalytic-dead versions of Cas9 (dCas9), which are particularly powerful ways of anchoring proteins of interest to regions of interest.

Chapter II - Arrested Replication Forks Guide Tf1 Integration

1. Introduction

Retrotransposons are mobile genetic elements that replicate through an RNA intermediate that is reverse transcribed into a cDNA capable of insertion elsewhere in the genome. By virtue of this amplifying mechanism, retrotransposons comprise large portions of many eukaryotic genomes, and have a critical influence on their evolution (Burns & Boeke, 2012). To prevent affecting the fitness of their host species, fungal LTR retrotransposons minimize their mutagenic potential by carefully selecting integration sites away from protein coding sequences (Bushman, 2003). The different families of LTR retrotransposons employ a variety of strategies for this selective target site selection, but current models presuppose tethering interactions between retrotransposon proteins and host DNA binding factors (Bushman, 2003). However, despite the high relatedness of the studied LTR retrotransposons, a universal mechanism guiding their target site selection hasn't been established.

The fission yeast genome shows signs of ancient and persistent colonization by the LTR retrotransposons Tf1 and Tf2, members of the Metaviridae/Ty3-gypsy like group of transposable elements (Bowen, 2003). Both Tf1 and Tf2 exhibit a preference for insertion into promoters of RNA polymerase II transcribed genes (Bowen, 2003; Yabin Guo & Levin, 2010), coinciding with the nucleosome free region (NFR) that is usually present preceding the transcription start site. The main determinant of NFR presence in fission yeast promoters is Sap1 (Tsankov et al., 2011), which binds DNA as homopolymers to clusters of a 5-bp sequence motif (Arcangioli, Ghazvini, & Ribes, 1994; Krings & Bastia, 2006). Because of this high association between Sap1 binding

profiles and Tf1 integration preferences, we sought out to directly investigate whether Sap1 could be the targeting factor for Tf1, and if so, what activities of Sap1 are responsible for targeting.

2. Materials and Methods

Transposon Trapping Assays

Plasmid based transposon trap assays were performed as previously described (Leem et al., 2008). Briefly, independent transformants of *S. pombe* strains transformed with a nmt1 driven Tf1-*neo* expressing plasmid were patched onto EMM plates with 2g/L of appropriate drop-out mix lacking thiamine and grown for 4 days at 32°C. For the plasmid-based transposition assays, cells were then replica plated onto 5-FOA-leu twice to remove the expression plasmid and finally onto EMM –leu + G418 (500µg/mL) + FOA (1mg/mL) plates to select for transposition events while maintaining selection for the target plasmids. For the genomic intron trap assays, cells were only replica plated a single time onto 5-FOA (1mg/mL) plates to select for transposition events. For the plasmid-based assay, DNA was purified from the patches (see DNA extraction protocol), electroporated into bacteria, and grown on plates containing Carbenicillin (100µg/mL) to quantify total extracted plasmid or Carbenicillin (100µg/mL) and Kanamycin (50µg/mL) to quantify the subpopulation of total plasmid containing insertion events. For the genomic intron traps, individual colonies were picked and subjected to colony PCR to map insertion sites. To determine whether FOA-resistant colonies from the genomic intron traps experiment arose from Tf1-*neo* insertions or mutations in *ura4(bpb1intron)+*, two PCRs were performed—one with oMZ25/26 to amplify *ura4(bpb1intron)+*, and the second with oM25/26/130 to detect Tf1-*neo* at the locus. Colonies negative for the former PCR and positive for the latter PCR were considered to have Tf1-*neo* insertions and were subjected to a 3rd PCR with oM25/26/146 to confirm the TSD. Colonies positive for both PCRs were considered to have mutations in *ura4(bpb1intron)+*.

DNA extraction

Pellets were resuspended in in 1mg/mL zymolyase 100T in CPS buffer (50mM citrate phosphate buffer, pH=5.6, 1.2M sorbitol), treated for 1 hr at 37°C (or until >90% have lost their refringence under brightfield microscopy), then spun down for 3' at 3,000g and resuspended in 5xTE (50mM Tris-HCl, 5mM EDTA, pH=8). 1/10th volume of 10% SDS was added and cells were lysed by incubation at 65°C for 5'. Next, KAc was added to 1.25M final and incubated 30' on ice. Samples were then spun to remove cell debris and ethanol precipitated. DNA pellets were then brought up in TE + RNase A (100µg/mL) and treated for 1 hour at 42°C. Supernatant was extracted with phenol/chloroform/iso-amyl alcohol (pH=8) twice, then concentrated by ethanol precipitated a final time. Pellets were then washed with 70% ethanol once before being resuspended in a small volume of TE or EB (10mM Tris-HCl, pH=8).

Quantitative Transposition Assays

Quantitative measurements of Tf1-*neo* transposition frequencies were measured as previously described (Atwood, Choi, & Levin, 1998). Four independent colonies of cells transformed with either a wild-type Tf1-*neo* expression vector, or a Tf1-*neo* expression vector containing mutations in the catalytic site of the Tf1 integrase, were patched onto EMM plates lacking thiamine and containing 2g/L drop-out mix lacking uracil and grown for 4 days at 32°C. After, cells were patched onto 5-FOA to remove the Tf1-*neo* expression vector, and dilutions were plated onto 5-FOA or YES + G418 (500µg/mL) + FOA (1mg/mL) + 2g/L drop-out minus uracil mix plates to measure transposition

frequencies. The proportion of G418 and 5-FOA resistant colonies to 5-FOA resistant colonies represents the transposition frequency.

Cloning, transformations, and PCR

Target plasmids used for the transposon trap assay were created by digesting pArt1(*38*) with BamHI and SphI and cloning the appropriate phosphorylated and annealed oligos (ter1F = oZ15/16, ter1R, = oZ17/18, dr2F = oZ7/8, dr2R = oZ9/10, dr2dF = oZ11/12, dr2dR = oZ13/114, scrF = oZ19/20, scrR = oZ21/22) to create the target site directly upstream of ARS1. Plasmid traps containing the Reb1 binding site *ter2* were cloned by digesting the appropriate pArt1 derivative with SmaI, then ligating in the annealed and phosphorylated oligo pair oM683/oM684. The catalytic dead Tf1-*neo* expression vector was constructed by creating a Tf1 integrase fragment containing the E73A and D38A mutations with oligo pairs oZ05/oZ02, oZ06/oZ03, and oZ04/oZ01, then performing cross-over PCR all three fragments and oZ05/oZ06. This fragment was then digested with NarI and BsrGI and cloned into the backbone of pHL414 cut with the same enzymes to create pMZ209. The LEU2 marked Tf1-*neo* expression vector used for the genomic intron traps was constructed by amplifying the LEU2 gene from pArt1 with oM750 and oM751 and ligating it into the BamHI/SphI backbone of pHL414, pMZ209, pHL449, and pHL476 creating pMZ325, pMZ326, pMZ416, and pMZ417 respectively. The frameshifted Tf1-*neo* expression vector used as a control for cDNA and Tf1 integrase quantification was created by cutting pHL414 with AvrII (NEB), filling in the 5' overhangs with Klenow DNA polymerase, and religating the blunt ends with T4 DNA ligase, creating pMZ246. This introduces a stop codon 198 amino acids into the Tf1

ORF. The intron from *S. pombe bpb1* was reconstructed with a CspCI site in the middle by mixing oligos oM744 and oM745 and incubation with Klenow (NEB). This fragment was then Gibson assembled into the SmaI backbone of pUC19 with two URA4 gene fragments created by colony PCR of a *ura4+* strain with oligo pairs oM738/739 and oM741/743 (*39*). This created pMZ303, which was then cut with CspCI and used as a backbone to clone in appropriate phosphorylated and annealed oligos pairs (ter1F = oM734/735, ter1R = oM735/736, scrF = oM752/753, scrR = oM754/755, ter2F = oM756/757, ter2R = oM758/759) to create the intron trap plasmids. These plasmids were then linearized and transformed into a strain of *S. pombe* containing the *ura4-DS/E* allele (ZB1465). The constructs fix the deletion in *ura4-DS/E* and give rise to URA+ colonies if the intron is functional. To create the *E. coli* expression vector the Tf1 integrase for antibody production, Tf1 integrase was amplified by PCR (oM228/229) and cloned into pET28(a)+ at the NdeI/XhoI restriction sites in the multiple cloning site, creating pMZ193 (6xHis-Tf1 integrase). To create the *neo* fragment generating plasmid used for cDNA blotting, pHL414 was digested with NarI/XhoI, blunt ended with Klenow, and the ~800 bp *neo* fragment was cloned into pUC19 at SmaI, generating pMZ243. All *E. coli* transformations were performed by electroporation with the Bio-Rad Gene Pulser (1.6kV, 200Ω, 25μF), and all *S. pombe* transformations were performed using the Lithium Acetate/PEG transformation protocol. Unless otherwise indicated, all PCR was performed with high fidelity Phusion DNA polymerase (NEB) in the supplied 1xHF buffer.

Chromatin Immunoprecipitation

ChIP experiments were roughly performed as previously described (Pidoux, Mellone, & Allshire, 2004). *S. pombe* strains were either growth in drop-out media lacking leucine (for plasmid ChIP), YEA (for intron ChIP), or EMM-ura (for cDNA ChIP) to OD_{600} 0.8-1.5. Cells for cDNA ChIP were induced for Tf1 expression as described in the "cDNA and Integrase Level Profiling" section of the methods. Cells were fixed in 1% formaldehyde for 30' at room temperature, then quenched with 125mM glycine for 5' before harvesting by centrifugation. To prepare chromatin extracts, cells were treated with 5 mL 0.4mg/mL zymolyase 100T (USBiological) in PEMS (100mM PIPES, 1.2M sorbitol, 1mM $MgCl_2$, 1mM EDTA, pH = 7.5) for 30' at 37°C, or until cells lost their refringence under brightfield microscropy. Cells were then brought up in 1mL 1x ChIP lysis buffer (50mM HEPES-KOH, 140mM NaCl, 1mM EDTA, 1% TritonX-100, 0.1% sodium deoxycholate, pH=8) with freshly added protease inhibitors (Merck/Millipore 539136) and 2mM PMSF, and sonicated for 10 cycles on high, 30s on, 30s off, in a bath sonicator at 4°C (Diagenode) to obtain dsDNA fragments of approximately 200bp. Insoluble debris was removed with two 15' spins at max speed (16,300g) at 4°C. Chromatin was then quantified by Bradford and equal amounts (approximately 1mg per IP) were immunoprecipitated with 10µL Protein A Dynabeads (Life Technologies) conjugated to polyclonal Sap1 antibody. A 1/10 IP volume aliquot was brought up to 250 microliters with 1xTES (50mM Tris-HCl, 10mM EDTA, 1% SDS, pH = 8), and saved at -80°C as the whole cell extract. Complexes were allowed to form overnight, then beads were washed once with 100 vol. 1x ChIP lysis buffer, once with 1x ChIP lysis buffer with 0.5M NaCl final, once with 1x ChIP wash buffer (10mM Tris-HCl, 250mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1mM EDTA, pH=8) and once with 1xTE (10 mM

Tris-HCl, 1 mM EDTA, pH=8). Beads were then resuspended in 250µL 1xTES and placed at 65°C overnight (~12hr) to reverse the crosslinking. The immunoprecipitated DNA and WCE DNA from the freezer in was then diluted in half to reduce the SDS concentration to 0.5% and Proteinase K was added to 1.2 mg/mL. Protein was digested for 2-3 hours at 37°C before being phenol/chloroform extracted once, concentrated by ethanol precipitation and resuspended in 100µL EB (10mM Tris-HCl, pH=8) plus 100 µg/mL RNase A. RNA was digested with 30' treatment at 37°C. Before qPCR, all IP and WCE DNA was PCR column purified (Qiagen) to ensure it was free of PCR inhibitors. qPCR was done on an Eppendorf Realplex Mastercycler with 1xKAPA SYBR FAST qPCR mix appropriate oligo pairs (Kapa Biosystems). ChIP enrichment was calculated with the $\Delta\Delta C_t$ method.

Yeast Two-Hybrid

Yeast two-hybrid analysis was done, with modifications to allow for the testing of interactions by mating, with components from the DupLEX-A yeast two-hybrid system (OriGene Technologies, Rockville, MD). To create the LexA expression vectors used for Y2H analysis, Tf1 integrase was amplified from pHL414 using oligos oY1/oY2. Sap1 was amplified in fragments from *S. pombe* genomic DNA with oY3/oY4 (full length), oY3/oY5 (domain V), oY6/oY7 (domain I to III), oY8/oY4 (domain IV), oY3/oY7 (domains V, I, II, III), and oY6/oY4 (domains I to IV). All these fragments were then digested with AscI and NotI and cloned into the backbone of both pEG202 (containing the LexA DNA binding domain, or BD) and pJG4-5 (containing the LexA activation domain, or AD) cut with the same enzymes to create in-frame N-terminal LexA fusions.

All LexA-AD fusions were transformed into SB1035 and maintained on synthetic dropout media lacking tryptophan, and all LexA-BD fusions were transformed into EGY48+pSH18-34 and maintained on standard synthetic dropout media lacking uracil and histidine. Transformation of *S. cerevisiae* was performed with a standard lithium acetate protocol. To test for interaction between two proteins, equal amounts of appropriate strains were mixed, spotted on YEPD plates, and allowed to mate overnight at 30°C. After, cells were lifted and diploids were selected on synthetic dropout (SD) plates lacking histidine, uracil, and tryptophan, and allowed to grow for 2 days at 30°C. Interaction was then assayed by spotting selected diploids on synthetic dropout lacking leucine, histidine, tryptophan, and uracil and containing 2% galactose instead of 2% glucose. To test the robustness of the interaction, spots were also performed on SD(gal) lacking histidine, tryptophan, and uracil, and containing X-Gal (40 μg/mL). Empty LexA vectors were used as a negative control, while the RAB-10 / CNT-1 interaction was used as a positive control (Shi et al., 2012). Internal controls were performed for each construct by mating each with an empty vector to rule out any auto-activation effects. Cells were imaged after 2 days of growth at 30°C.

2D Gel Electrophoresis

Visualization of DNA replication intermediates by 2D gel electrophoresis was done as previously described with a few minor modifications (Zaratiegui et al., 2010). Briefly, approximately 500mL of cells growing in drop-out lacking leucine were grown to an OD_{600} of 1-1.5 at 32°C before killing them in 0.1% sodium azide and harvesting them by centrifugation. Pellets were then brought up in 5mL nuclear isolation buffer (17%

glycerol, 50mM MOPS-NaOH, 150mM Kac, 2mM $MgCl_2$, 500µM spermine, 150µM spermidine, 2mg/mL zymolyase 100T, pH=7.2) and incubated for approximately 45' or until the cells lost their refringence under brightfield microscopy. A crude nuclei pellet was then obtained by adding 4 volumes of deionized H_2O and spinning 10' at 7.5k RPM in a JA-17 rotor at 4°C. Nuclei were then very gently resuspended in 5mL TEN (50mM Tris-HCl, 50mM EDTA, 100mM NaCl, pH=8), then after through resuspension, SDS was added to 0.1% to lyse the nuclei, and RNase A was added to 0.5mg/mL before incubation at 37°C for 30 minutes. Next, proteinase K was added to 0.3mg/mL, incubated 1 hour at 37°C. After 1 hour, the SDS concentration was increased to 1% and the mixture was incubated an additional hour at 37°C. Samples were then cooled on ice for 30' before KAc was added to 1.1M final, incubated an additional hour on ice, then spun at 10k RPM in a JA-17 rotor at 4°C for 10 minutes. The clarified supernatant was then transferred to an Oakridge tube and 1 volume of 2-propanol was added. This mixture was then spun in the JA-17 for 30' at 4°C at 10k RPM to precipitate the DNA, and pellets were then resuspended overnight in 5mL 5xTE (50mM Tris-HCl, 5mM EDTA, pH=8) with gentle rocking at 4°C overnight. The next day, the DNA was phenol/chloroform extracted 5 times and chloroform extracted once before being transferred to a clean Oakridge tube for ethanol precipitation with 1/10 vol. 3M NaAc and 2.5 vol 100% EtOH. The pellet was resuspended with gentle rocking at 4°C overnight. The next day, the DNA was transferred to a 1.5mL tube and digested with ClaI and PvuII simultaneously for 6 hours at 37°C. After the digestion, the DNA was ethanol precipitated, washed twice with 500µL 70% EtOH, and then resuspended in 30µL 1xTBE prepared in 3x purple loading dye (NEB) with a cut tip. To ensure even running, DNA was quantified before loading and no

more than 15μg of digested DNA was loaded. This DNA was then loaded into a 0.7% agarose gel prepared in 1xTBE and run at 0.4V/cm for 2 days at room temperature with flanking 2 log ladders (NEB) as size markers. The gel was covered in aluminum foil to protect the DNA from exposure to ambient UV light. After the run, the gel was stained in 1xTBE with 500μg/L EtBr, and vertical strips were cut between the 2kb and 8kb bands and placed into the wells of a 1% agarose gel in 1xTBE prepared with 500μg/L EtBr. The strips were then sealed in place with 1% agarose dissolved in 1xTBE and cooled to 50°C. This gel was then run at 3.5V/cm at 4°C for approximately 6 hours before being equilibrated in 0.4M NaOH, 1M NaCl, nicked with 0.4J of UV (Stratagen 1800 Crosslinker), and transferred to charged nylon overnight through vertical capillary transfer. This membrane was then probed with ~100ng of the ~600 bp EcoRV/ClaI LEU2 fragment from pArt1 labeled with [α-32P]-dCTP (Perkin-Elmer, 3000Ci/mmole, 10mCi/mL) with a random primed labeling kit (Roche) and purified on a homemade G50 Sephadex (Sigma) column. Hybridization occurred overnight at 40°C before the membrane was washed twice with 2xSSC (0.3M NaCl, 30mM trisodium citrate, pH=7) with 0.1% SDS, then twice with 0.1xSSC with 0.1% SDS. All washes took place at 40°C for 15' each. The membrane was then exposed to a phosphorscreen for 1-3 days before being imaged on a phosphorimager.

cDNA and Integrase Level Profiling

To measure cDNA and integrase levels, *S. pombe* cells containing a Tf1-*neo* expression plasmid (or indicated controls) were grown in 5mL EMM-leu+10μM B1 to stationary phase O/N at 32°C, washed 3 times in sterile 50mL dH_2O to remove the B1, then seeded

in 50mL fresh EMM-leu lacking thiamine to induce expression of Tf1-*neo*. Cells were grown to OD_{600} = 1.0-1.5 for one day, then diluted again in EMM-leu lacking thiamine so that they would reach OD_{600} = 1.0-1.5 the next day. After the 2 days of growth without thiamine, cells were harvested by centrifugation and split in half. Half the cells were subjected to a standard genomic DNA isolation (see DNA extraction protocol), and the other half of cells was resuspended in Buffer A (50mM HEPES-KOH, 150mM NaCl, 1% TritonX-100, 0.5% sodium deoxycholate, 10% glycerol) with 1/100 vol. freshly added protease inhibitors (Merck/Millipore 539136) and glass beads up to the meniscus. Protein extracts were prepared by bead beating, then clarified with two 15' centrifugations at 4°C. Extracts were then quantified by Bradford, diluted to equal concentrations, and then diluted in half with 2x Laemmli loading dye. Samples were run on a 8% SDS-PAGE gel, transferred to nitrocellulose in 1x Bjerrum transfer buffer with 20% MeOH (1.4V/cm^2 of gel for 1 hour), then blocked in 5% non-fat milk for 1 hour at room temperature before being incubated with the appropriate antibodies overnight in 1xTBST (50mM Tris-HCl, 150mM NaCl, 0.1% Tween-20, pH=7.4) with 2.5% non-fat milk. Membranes were then washed, incubated with the appropriate secondary HRP-conjugated antibody, and developed with ECL (GE). cDNA was visualized by roughly following previously described protocols (Atwood et al., 1996). First, 1µg DNA extract was digested with 30U of BstXI overnight in 1xCutSmart (NEB), then loaded onto a 1% agarose gel prepared in 0.5xTAE. This gel was run approximately 13 cm, then transferred to charged nylon and probed with a random primed labeled [α-32P]-dCTP (Perkin-Elmer, 3000Ci/mmole, 10mCi/mL) 861bp *neo* fragment produced by digesting pMZ243 with EcoRI/XhoI. The transfer, washing, and probe generation were done as described in the 2D gel

electrophoresis protocol.

High throughput mapping of Tf1 insertion points

High-throughput mapping of Tf1 insertion points was done as previously described but with slight modifications for the Illumina miSeq platform (Yabin Guo & Levin, 2010). Twenty plates containing 16 independent patches were used for experiments in both wild-type and *sap1-c* backgrounds. After DNA purification, digestion, and linker ligation, libraries were prepared for high-throughput sequencing by PCR with oM538/544 (for WT), and oM538/543 (for *sap1-c*). These primers put the p7 and p5 tags necessary for paired-end Illumina sequencing on the ends. PCR was done as described previously with Titanium Taq (Clontech) on a 96 well plate (Yabin Guo & Levin, 2010). Library DNA was column purified, and DNA between 130bp and 500bp was isolated for flowcell cluster formation using a Pippin prep (Sage Biosciences).

Bioinformatic Analysis

The insertion points obtained in the high throughput transposition experiments were fitted to a logistic regression model, with three fold random matched sites as controls, obtained sampling positions between 150 and 500bp of MseI sites present in the genome, using the fold enrichment signal for Sap1 ChIP, Abp1 ChIP{Zaratiegui:2011gi}, Cdc20 ChIP(Sabouri, Capra, & Zakian, 2014) and g-H2A ChIP(Rozenzhak et al., 2010) as predictors, performing best subset selection with 10-fold cross-validation, using the R package *bestglm*. The resulting best model contained only Sap1 as a predictor. The Receiver Operating Characteristic (ROC) curves were obtained with the *pROC* package.

The estimated error for the Area Under the Curve (AUC) was obtained by 50-fold Bootstrapping using the *BOOT* package.

Chromosome Conformation Capture (3C)

DNA interactions between transposon-derived cDNA and genomic DNA were analyzed by a modified version of 3C described previously (B. N. Singh, Ansari, & Hampsey, 2009). *S. pombe* cells containing the *ura4(bpb1-target site)* construct and a LEU2 marked Tf1-*neoAI* expression plasmid were grown to induce Tf1 expression in EMM-leu for 2 days, as described previously for cDNA and INT profiling. After the 2 days of growth without thiamine, cells were crosslinked in 1% formaldehyde for 20', then quenched in 0.125M glycine for 5'. The cell pellet obtained was washed once with 5mL of 1x TBS buffer (50mM Tris-HCl, 150mM NaCl, pH=7.5) containing 1% Triton X-100 and once with 2mL of FA-lysis buffer (50mM Tris-HCl , 150mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 1mM PMSF, pH=7.5). Approximately 1.2mL of acid-washed glass beads was added, and cells were lysed by beat beating for a total of 15' at 4°C in a mini bead beater. Lysates were then collected and clarified by a 15' spin at 13.3k RPM spin at 4°C in a mini-centrifuge. The chromatin pellet was washed once with 500µL of FA-lysis buffer and resuspended in 3mL of 10mM Tris-HCl (pH 7.5) and aliquots of 100µL were stored at -80°C. Prior to endonuclease cleavage, chromatin was denatured in 1% SDS for 20' at 65°C. SDS was then sequestered by incubation with 750µL of 1% Triton X-100. Samples were centrifuged, and the pellets were dissolved in 79µL of HPLC grade water. 10µL of 10X CutSmart Buffer and 6µl of the enzyme XmaI (50,000 units/mL, NEB) were added to the chromatin. The restriction digestion was

carried out for ~12hr with gentle mixing. The reaction was then stopped by adding 10µL of 10% SDS and incubating the samples for 20' at 65°C. SDS was sequestered again as done previously. Samples were centrifuged, and the pellet was dissolved in 345µL of HPLC grade water. Next, 350µL of 2X Quick Ligase Buffer and 5µl of Quick T4 Ligase were added to chromatin (NEB). Ligation reactions were performed for 1hr at 25°C. To ensure complete removal of RNA, 10µg of DNase-free RNase A (AMRESCO) was added to the reaction mixture and incubated 1hr at 37°C. The crosslinks were reversed for 12hr at 65°C in the presence of 1% SDS and 100µg Proteinase K (Thermo Scientific). The samples were extracted with phenol-chloroform and precipitated by ethanol in the presence of glycogen. PCR reactions were performed using the primer pairs indicated below. Because Tf1-*neo* insertion into the genome at the *ura4(bpb1)* locus has the potential to confound the results, samples were first screened with oligos oM25/oM130, oM25/oM146, oM26/oM130, and oM26/oM146 to detect potential insertions at the locus. Importantly, the samples did not contain any detectable Tf1-*neo* insertion, indicating the 3C signal is due to strand ligation by T4 DNA ligase and not the integrase (not shown). Ligation products between target and cDNA were amplified with a primer pair that anneals to the *ura4(bpb1)* target region and to the cDNA of Tf1-*neo* (oJR1/2), producing a band of 251bp. Input PCR products were generated using a convergent primer pair that amplifies a region of the donor plasmid (oJR3/4). PCR products were separated on a 1.5% agarose gel and visualized by ethidium bromide staining using a Gel Logic 1000 Imaging System (Kodak).

Sap1/Tf1 Integrase Antibody generation

To produce recombinant integrase for antibody production, BL21 was transformed with pMZ193 (N-terminal 6xHis tagged Tf1 integrase) or pMZ150 (for N-terminal tagged 6xHis tagged Sap1) (Arcangioli et al., 1994), diluted from a saturated O/N culture of LB with appropriate antibiotics to OD_{600} = 0.1, then grown until OD_{600} = 0.5 in 50mL LB+antibiotics. Cells were then induced for Tf1 integrase expression by addition of IPTG (Thermo Scientific) to 1mM final and grown at 16°C for 16 hours. Lysates were prepared by lysozyme treatment and sonication (0.5s pulses for 20 seconds) in NETN buffer (20mM Tris-HCl, 100mM NaCl, 10mM imidazole, 1mM PMSF, 0.1% TritonX-100, pH=7.4). 6xHis-Tf1-INT was bound to Ni-NTA agarose beads (Genscript) for 1 hour at 4°C, then washed extensively (500 bed volumes) with NETN buffer with 200mM NaCl and 20mM imidazole. The integrase was then eluted from the column with the addition of 100mM imidazole. For Sap1 antibody, IPTG induction was done with 0.5mM IPTG at 37°C for 5 hours and purified on Ni-NTA beads in 1xNi lysis buffer (10mM Tris-HCl, 10% glycerol, 500mM NaCl, 0.1% NP-40, 10mM imidazole, 8M urea, pH=8), then washed extensively (500 bed volumes) with 1xNi wash buffer (20mM Tris-HCl, 20% glycerol, 100mM KCl, 20mM imidazole, 8M urea, pH=8) eluted with 1xNi wash buffer with the imidazole concentration brought up to 100mM final. 5.0 mg of purified Tf1 integrase and Sap1 was used for polyclonal antibody production in rabbits (Yenzym Antibodies LLC).

Growth of sap1-c Containing Strains

Our previous studies on *sap1-c* led us to the conclusion that this allele was lethal in *S. pombe* in an otherwise WT background (Zaratiegui et al., 2010). However, further

analysis revealed that this phenotype was due to death in saturation in YEA prepared with the standard 3% glucose (not shown). We have now found that *sap1-c* is viable in stationary phase in both EMM and synthetic drop-out (SD) media, and interestingly, in YEA prepared with 2% or less glucose (not shown). All experiments performed with *sap1-c* strains in this study were performed in EMM or SD medias. In these media, *sap1-c* doubles at approximately same rate as WT strains.

Table 1 - Oligonucleotides used in this chapter

name	5'->3' sequence	purpose	common use name
oM130	ATTATTTGTAGATCACAAGAGTTCAG	used to seq. 5' LTR insertion point	tf388R
oM146	TCCATGTTGGAATTTAATCG	used to seq. 3' LTR insertion point	neo139R
oM744	gtaagtgttttaagagattattatgggcttaactCAAttcttGTGGgtta	bpb1 intron cloning	bpb1 half 1 (CspCI)
oM745	ctaaaaatattaataatttgcacagagagataacCCACaagaaTTGagtt	bpb1 intron cloning	bpb1 half 2 (CspCI)
oM738	gccagtgaattcgagctcggtacccCAGCTAGAGCTGAGGGGATG	bpb1 intron cloning	ura4I USF
oM739	taataatctcttaaaacacttacCTTGTATAATACCCTCGCCTGG	bpb1 intron cloning	ura4I USR - bpb1
oM741	caaattattaatattttagGCCTCAAAGAAGTTGGTTTACCTT	bpb1 intron cloning	ura4I DSF - bpb1
oM743	tgcaggtcgactctagaggatccccTTAATGCTGAGAAAGTCTTTGCTG	bpb1 intron cloning	ura4I DSR
oM752	GTGTTGGTTTAAAGACGTATGACATTTTGTgt	bpb1 intron cloning	scrF bpb1 +
oM753	ACAAAATGTCATACGTCTTTAAACCAACACta	bpb1 intron cloning	scrF bpb1 -
oM754	ACAAAATGTCATACGTCTTTAAACCAACACgt	bpb1 intron cloning	scrR bpb1 +
oM755	GTGTTGGTTTAAAGACGTATGACATTTTGTta	bpb1 intron cloning	scrR bpb1 -
oM734	AGGGATTTAACGCAGTGCAAGGAGCTATgt	bpb1 intron cloning	ter1F bpb1 +
oM735	ATAGCTCCTTGCACTGCGTTAAATCCCTta	bpb1 intron cloning	ter1F bpb1 -
oM736	ATAGCTCCTTGCACTGCGTTAAATCCCTgt	bpb1 intron cloning	ter1R bpb1 +
oM737	AGGGATTTAACGCAGTGCAAGGAGCTATta	bpb1 intron cloning	ter1R bpb1 -
oM756	GTGCATTACCCTTACCTgt	bpb1 intron cloning	ter2F bpb1 +
oM757	AGGTAAGGGTAATGCACta	bpb1 intron cloning	ter2F bpb1 +
oM758	AGGTAAGGGTAATGCACgt	bpb1 intron cloning	ter2R bpb1 +
oM759	GTGCATTACCCTTACCTta	bpb1 intron cloning	ter2R bpb1 -
oM750	GATCgcatgcTCGACTACGTCGTTAAGG	making pMZ325/326	SphILEU2R
oM751	ATGCggatccTCGAGGAGAACTTCTAGTATATCTA	making pMZ325/326	BamHILEU2F
oM693	tgacctgaccatttgatgga	Sap1 ChIP (plasmid traps)	qTS-F1
oM694	tacgtcgttaaggccgtttc	Sap1 ChIP (plasmid traps)	qTS-R1
oM532	ttgtgctcttcatcctgtgc	Sap1 ChIP normalization	q6F6-F
oM533	gaatccgagatttcgtccaa	Sap1 ChIP normalization	q6F6-R
oM857	gcaggatttcgaccaggata	Sap1 ChIP (intron traps)	qURA-F
oM858	tgagcccaagaagcaatttt	Sap1 ChIP (intron traps)	qURA-R
oM684	gtgcattacccttacct	ter2 cloning into DR2 R (@SmaI)	reb1BS
oM683	aggtaagggtaatgcac	ter2 cloning into DR2 R (@SmaI)	rcreb1BS
oZ01	ATCCAAAAGAAATCATTGCAGCAAATGATCATATTTTTACTTCTCAAACA	catalytic dead pHL414 (pMZ209)	D1038A F
oZ02	TGTTTGAGAAGTAAAAATATGATCATTTGCTGCAATGATTTCTTTTGGAT	catalytic dead pHL414 (pMZ209)	D1038A R
oZ03	CACAAACTGATGGACAAACTGCACGTACAAACCAAACTGTGGAGAAATTA	catalytic dead pHL414 (pMZ209)	E1073A F
oZ04	TAATTTCTCCACAGTTTGGTTTGTACGTGCAGTTTGTCCATCAGTTTGTG	catalytic dead pHL414 (pMZ209)	E1073A R
oZ05	CATCCAGGCATTGAACTTCTT	catalytic dead pHL414 (pMZ209)	tfMutF2
oZ06	GAGCATTACGCTGACTTGAC	catalytic dead pHL414 (pMZ209)	tfMutR3
oM175	TTATGCCTCTTCCGACCATC	cDNA southern blot	neoProbe F
oM176	GCCTGAGCGAGACGAAATAC	cDNA southern blot	neoProbe R
oM536	gtaatacgactcactatagggctccgcttaagggac	Tf1 high-throughput mapping	HL1870
oM537	5Phos/tagtcccttaaccggag/3AmMO	Tf1 high-throughput mapping	HL1871
oM538	caagcagaagacggcatacgagatgtaataacgactcactatagggc	Tf1 high-throughput mapping	p7 linker
oM543	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcagatcgatgcataggaatttagtttatgg	Tf1 high-throughput mapping	sap1-c barcode
oM544	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgatgtgatgcataggaatttagtttatgg	Tf1 high-throughput mapping	WT barcode
oM228	GGGGCATATGACAGCTGATTTTAAAAACCAAG	Tf1 integrase cloning	NdeI overhang
oM229	ATGCCTCGAGGATATTTAAGATTATTGTTTTTAATATAATC	Tf1 integrase cloning	XhoI overhang
oZ07	AGCGGAATAGCGTGGTGTAGCGAGCCCG	plasmid traps pART1	DR2F+
oZ08	gatcCGGGCTCGCTACACCACGCTATTCCGCTcatg	plasmid traps pART1	DR2F-
oZ09	CGGGCTCGCTACACCACGCTATTCCGCT	plasmid traps pART1	DR2R+
oZ10	gatcAGCGGAATAGCGTGGTGTAGCGAGCCCGcatg	plasmid traps pART1	DR2R-
oZ11	AGCGGAATAGCGTGGTGCAAGGAGCTAT	plasmid traps pART1	DR2DF+
oZ12	gatcATAGCTCCTTGCACCACGCTATTCCGCTcatg	plasmid traps pART1	DR2DF-
oZ13	ATAGCTCCTTGCACCACGCTATTCCGCT	plasmid traps pART1	DR2DR+
oZ14	gatcAGCGGAATAGCGTGGTGCAAGGAGCTATcatg	plasmid traps pART1	DR2DR-
oZ15	AGGGATTTAACGCAGTGCAAGGAGCTAT	plasmid traps pART1	TER1F+
oZ16	gatcATAGCTCCTTGCACTGCGTTAAATCCCTcatg	plasmid traps pART1	TER1F-
oZ17	ATAGCTCCTTGCACTGCGTTAAATCCCT	plasmid traps pART1	TER1R+
oZ18	gatcAGGGATTTAACGCAGTGCAAGGAGCTATcatg	plasmid traps pART1	TER1R-
oZ19	ATAGCTCCTTGCACTGCGTTAAATCCCT	plasmid traps pART1	SCRF+
oZ20	gatcACAAAATGTCATACGTCTTTAAACCAACACcatg	plasmid traps pART1	SCRF-
oZ21	ACAAAATGTCATACGTCTTTAAACCAACAC	plasmid traps pART1	SCRR+
oZ22	gatcGTGTTGGTTTAAAGACGTATGACATTTTGTcatg	plasmid traps pART1	SCRR-
oM683	GTGCATTACCCTTACCT	plasmid traps pART1-ter2	TER2+
oM684	AGGTAAGGGTAATGCAC	plasmid traps pART1-ter2	TER2-
oM684	AGGTAAGGGTAATGCAC	plasmid traps pART1-ter2	TER2-
oM25	CCCACTGGCTATATGTATGCATTTGT	intron trap	ura4-F

oM26	GCTTGTGATATTGACGAAACTTTTTGA	intron trap	ura4-R
oY1	atgcGGCGCGCCtacagatgattttaaaaaccaag	Tf1 integrase LexA fusion cloning	AscI-Tf1IN-F
oY2	atgcGCGGCCGCTTAGATATTTAGATTATTGTTTTTAATATAATC	Tf1 integrase LexA fusion cloning	NotI-Tf1IN-R
oY3	atgcGGCGCGCCtATGGAAGCTCCCAAGATGGA	Sap1 LexA fusion cloning	Sap1AscI-F
oY4	atgcGCGGCCGCTTAATGGTCACCAAGATTAGG	Sap1 LexA fusion cloning	Sap1NotI-FL-R
oY5	atgcGCGGCCGCttaGGAACTGGGAGATAACGAAG	Sap1 LexA fusion cloning	Sap1NotI-domV-R
oY6	atgcGGCGCGCCtTCTCCCGCCAAGGCACAGC	Sap1 LexA fusion cloning	Sap1NotI-ItoIII-F
oY7	atgcGCGGCCGCttaGCCAACCAAGAATTTTTCGC	Sap1 LexA fusion cloning	Sap1NotI-ItoIII-R
oY8	atgcGGCGCGCCtAAGCGCAAGTGTGATTGCAA	Sap1 LexA fusion cloning	Sap1AscI-domIV-F
HL1576	CCTGATTGCCCGACATTATCGCG	Sap1 cDNA ChIP	HL1576
oM478	Aagatccgggttacattgc	Sap1 cDNA ChIP	qNeoTf1R
oJR1	AATGGGCTCGCGATAATGTC	3C analysis, cDNA intron spanner	3C-cDNA-neoAI-R
oJR2	GAATCGTTGCCATCGATTCG	3C analysis, target site forward	3C-ura4-F
oJR3	GTTAGCAGGACACTGAATCG	3C analysis, input forward	3C-pHL-F
oJR4	CAAGGACCCTCAATTCTTCC	3C analysis, input reverse	3C-pHL-R

Table 2 - Strains used in this chapter

name	mat	genotype
PB1	h90	ura4-D18, leu1-32
ZB1069	h+	sap1-c, ura4-D18
ZB1465	h-	ura4-DS/E, leu1-32, ade6-M210
ZB1518	h-	ura4(bpb1intron-TER1F), leu1-32, ade6-M210
ZB1519	h-	ura4(bpb1intron-TER1R), leu1-32, ade6-M210
ZB1585	h-	ura4(bpb1intron-SCRF), leu1-32, ade6-M210
ZB1586	h-	ura4(bpb1intron-SCRR), leu1-32, ade6-M210
ZB1587	h-	ura4(bpb1intron-TER2F), leu1-32, ade6-M210
ZB1588	h-	ura4(bpb1intron-TER2R), leu1-32, ade6-M210
SB1035	MATa	ade2-1, ade3::hisG, ura3-1, his3-11,15, trp1-1, leu2-3,112, LYS2, can1-100, RAD5
EGY48	MATα	his3, trp1, ura3, $LexA_{op(x6)}$-LEU2

Table 3 - Plasmids used in this chapter

name	description	cloning strategy	parent plasmid
pART1		{McLeod:1987vr}	
pMZ14	TER F	BamHI/SphI - recombinant oligos	pART1
pMZ15	TER R	BamHI/SphI - recombinant oligos	pART1
pMZ16	DR2 F	BamHI/SphI - recombinant oligos	pART1
pMZ17	DR2 R	BamHI/SphI - recombinant oligos	pART1
pMZ18	DR2D F	BamHI/SphI - recombinant oligos	pART1
pMZ19	DR2D R	BamHI/SphI - recombinant oligos	pART1
pMZ203	SCR F	BamHI/SphI - recombinant oligos	pART1
pMZ204	SCR R	BamHI/SphI - recombinant oligos	pART1
pMZ183	DR2R reb1(NB)	reb1 cloned @ SmaI	pMZ17
pMZ184	DR2R reb1 (B)	reb1 cloned @ SmaI	pMZ17
pMZ205	SCRF reb1 (NB)	reb1 cloned @ SmaI	pMZ203
pMZ206	SCRF reb1 (B)	reb1 cloned @ SmaI	pMZ203
pMZ209	Tf1, catalytic dead integrase	mutated PCR product placed into pHL414 NarI/BsrGI backbone	pHL414-D38A,E73A
pHL414	Tf1 IN+	{Levin:1993wd}	
pMZ246	Tf1, frame shifted protease	pHL414 cut with AvrII, Klenow filled, then religated	pHL414
pHL2673	Tf1 IN+ for HTS	{Guo:2010bs}	
pMZ325	pHL414LEU2	LEU2 marked pHL414	pHL414
pMZ326	pMZ209LEU2	LEU2 marked pMZ209	pMZ209
pMZ416	pHL449LEU2	LEU2 marked pHL449 {Levin:1995uh}	pHL449
pMZ417	pHL476LEU2	LEU2 marked pHL476 {Levin:1995uh}	pHL476
pMZ193	antibody production	NdeI/XhoI Tf1IN	pET28(a)+
pMZ150	antibody production	from Benoit Arcangoli	
pMZ303	intron traps template	cross-over PCR and Gibson	pUC19
pMZ306	TER F	CspCI cloning - recombinant oligos	pURA4Ibpb1(CspCI)
pMZ307	TER R	CspCI cloning - recombinant oligos	pURA4Ibpb1(CspCI)
pMZ344	SCR F	CspCI cloning - recombinant oligos	pURA4Ibpb1(CspCI)
pMZ345	SCR R	CspCI cloning - recombinant oligos	pURA4Ibpb1(CspCI)
pMZ243	pUCneo	NarI/XhoI from pHL414, Klenow, then ligation @ SmaI	pUC19
pSH18-34	$LexA_{op(x6)}$-LacZ reporter	from OriGene Technologies	
pEG202	adh1-LexA-BD	from OriGene Technologies	
pJG4-5	adh1-LexA-AD	from OriGene Technologies	

3. Results

Sap1 is Associated with Tf1 Insertion Sites Genome-Wide

To test whether Sap1 binding coincided with transposition hotspots we performed high-throughput sequencing of transposon-host genome junctions in cultures overexpressing a genetically marked Tf1 transposon (Yabin Guo & Levin, 2010). Genome-wide correlation analysis shows a strong association of Sap1 enrichment (Zaratiegui et al., 2010) with insertion sites (Figure 3 A,B, Figure 4a). Sap1 is strongly enriched at the previously described Tf1 hotspots, like in the promoters of class II genes (Figure 3A, 4b). Peaks of significant Sap1 enrichment (MACS(Y. Zhang et al., 2008)) account for 63.1% of transposition points, while covering only 5.1% of the host genome, and contained more efficient insertion points than the rest of the genome (Figure 4c). Logistic regression analysis revealed that Sap1 binding is a strong predictor of insertion position (AUC-0.5^{WT}=0.217, Figure 5a). However, correlation between Sap1 fold enrichment and number of insertion points, while significant (Spearman's rho=0.70, p=1e-10), shows a wide variability beyond the threshold of significant enrichment (Figure 4 a,b), suggesting that Sap1 binding is not the only factor affecting target site competence. Insertion points coincide precisely with a maximum of Sap1 enrichment(Zaratiegui et al., 2010), strongly indicating that Sap1 determines Tf1 target site selection (Figure 3C). To test the involvement of Sap1 in Tf1 transposition we performed high-throughput insertion analysis in a *sap1* mutant with a lower affinity for DNA (*sap1-c*){Zaratiegui:2011gi}. *sap1-c* mutants exhibited a drastically reduced transposition frequency (Student's t-test, $p<0.001$, n=21, Figure 3D). Additionally, the strong association of insertion points with Sap1 was decreased (Figure 3C), the portion of

insertions in Sap1-enriched regions fell to 49.9%, and the accuracy of Sap1 binding as a predictor of insertion dropped (AUC-$0.5^{sap1\text{-}c}$=0.097, Figure 5a), indicating that transpositions are dispersed away from Sap1 binding peaks. Importantly, the *sap1-c* background showed no defects in cDNA processing or significantly altered levels of integrase, suggesting that the transposition defect is due to impaired integration (Figure 6). Together, these data show that Sap1 is a major determinant of Tf1 insertion target site selection.

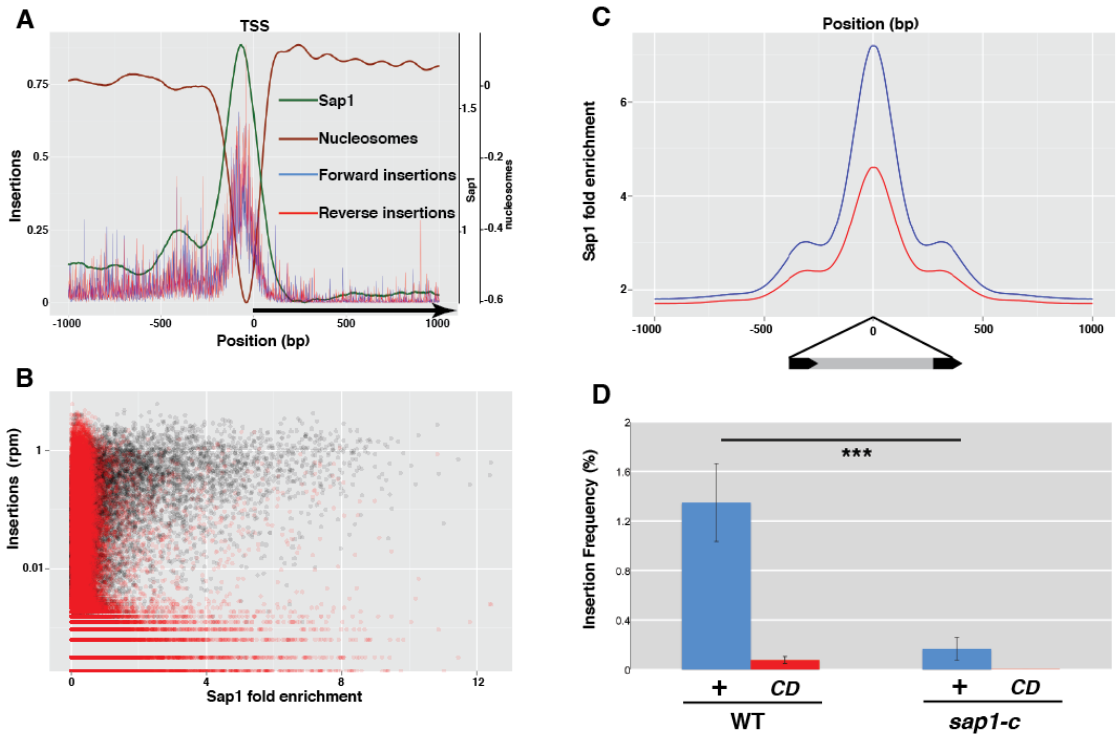


Figure 3 - Tf1 transposition into Sap1 binding regions. (A) Sap1, nucleosome positioning and average insertion number in reads per million (rpm) at type II genes aligned at the Transcription Start Site (TSS). (B) Genome-wide correlation between transposition (insertion number, rpm) and Sap1 binding in 500bp windows. Black: genomic windows; Red: randomized value pairs. (C) WT Sap1 enrichment around WT insertions (blue) and *sap1-c* insertions (red). (D) Transposition frequency in WT and *sap1-c* mutant of Integrase + (+) and Catalytic Dead (CD) Tf1. Error bars depict s.d. and asterisks depict statistically significant differences.

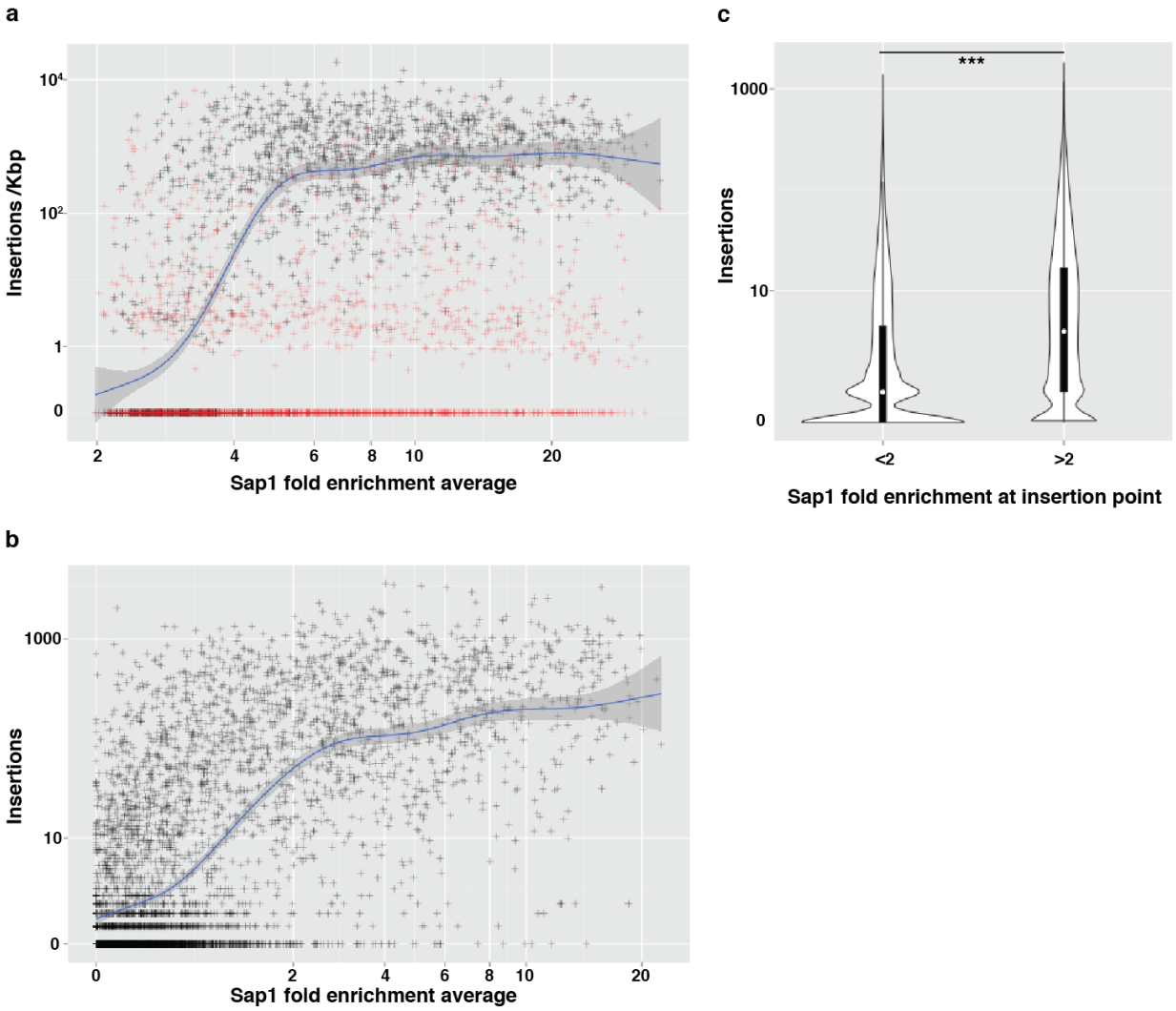


Figure 4 - Correlation between Sap1 enriched regions and Tf1 transposition. (a) Scatterplot of transposition density in Sap1 significantly enriched regions (MACS fold enrichment>2). Black: genomic windows. To show that the association is not due to uniform distribution of Sap1 and insertions, the value pairs are randomized and plotted in red. (b) Scatterplot of transposition number in the nucleosome free region of protein coding gene promoters, defined as 200bp upstream of the transcription start site. A spline smoothed trend line with shaded 95% confidence intervals is overlaid in blue. (c) Violin plot of average normalized insertion count at insertion points located within (Sap1 fold enrichment >2) and outside (Sap1 fold enrichment <2) Sap1 significantly enriched regions. A box-and-whiskers plot with the median as a white point is overlaid. Insertion points within Sap1 enriched regions present a higher number of insertions (Median of 3.96 vs. 0.98, Mann-Whitney U, $p<2.2x10^{-16}$)

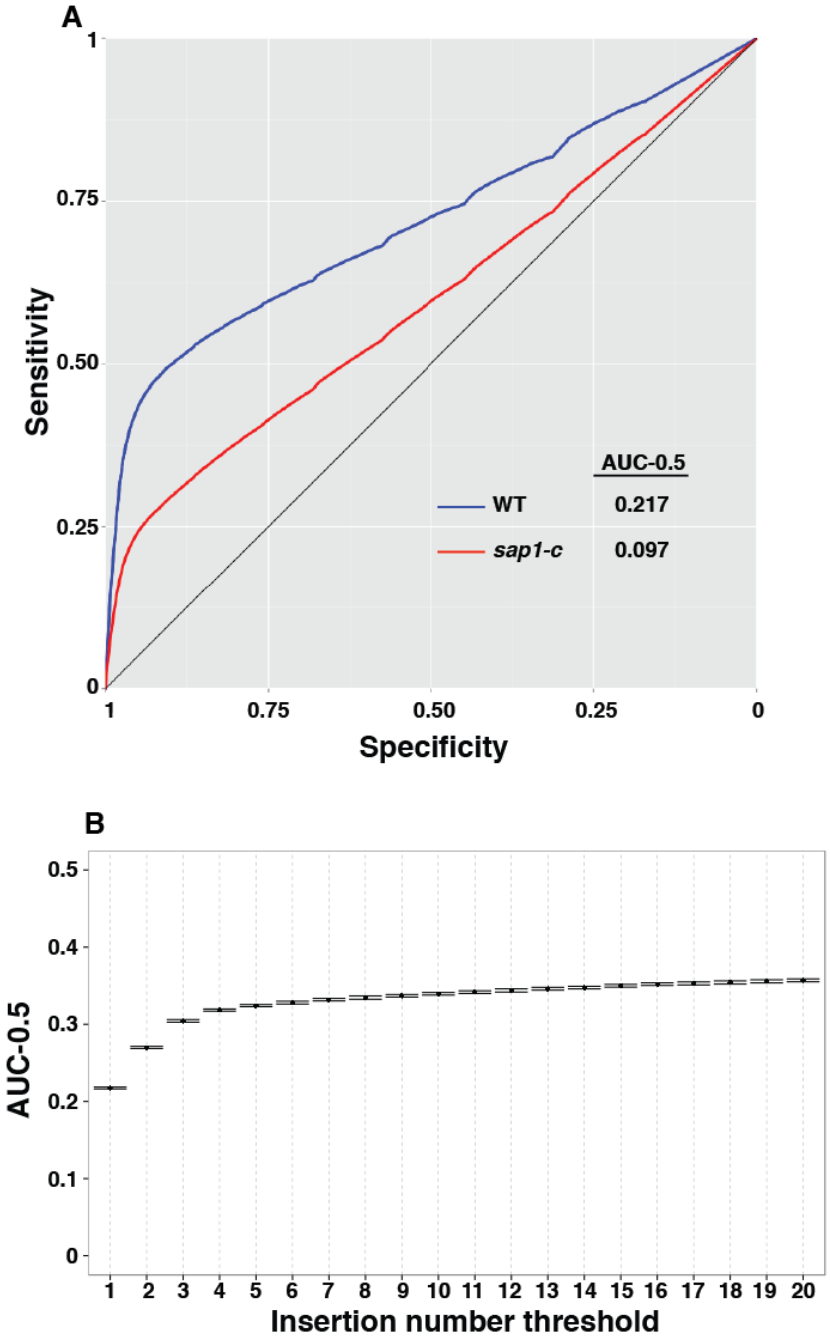


Figure 5 – Sap1 binding strength is highly predictive of Tf1 insertion. (A) Receiver Operating Characteristic (ROC) curves for the logistic regression model using WT Sap1 binding as a predictor of insertion positions in WT (blue) and *sap1-c* (red). The Area Under the Curve (AUC, after subtracting 0.5, representing pure random prediction) represents the predictive power of the model. (B) AUC of logistic regressions using insertion points with increasing number of insertions as threshold to be included in the model. Error bars depict standard deviation obtained from 50-fold bootstrapping.

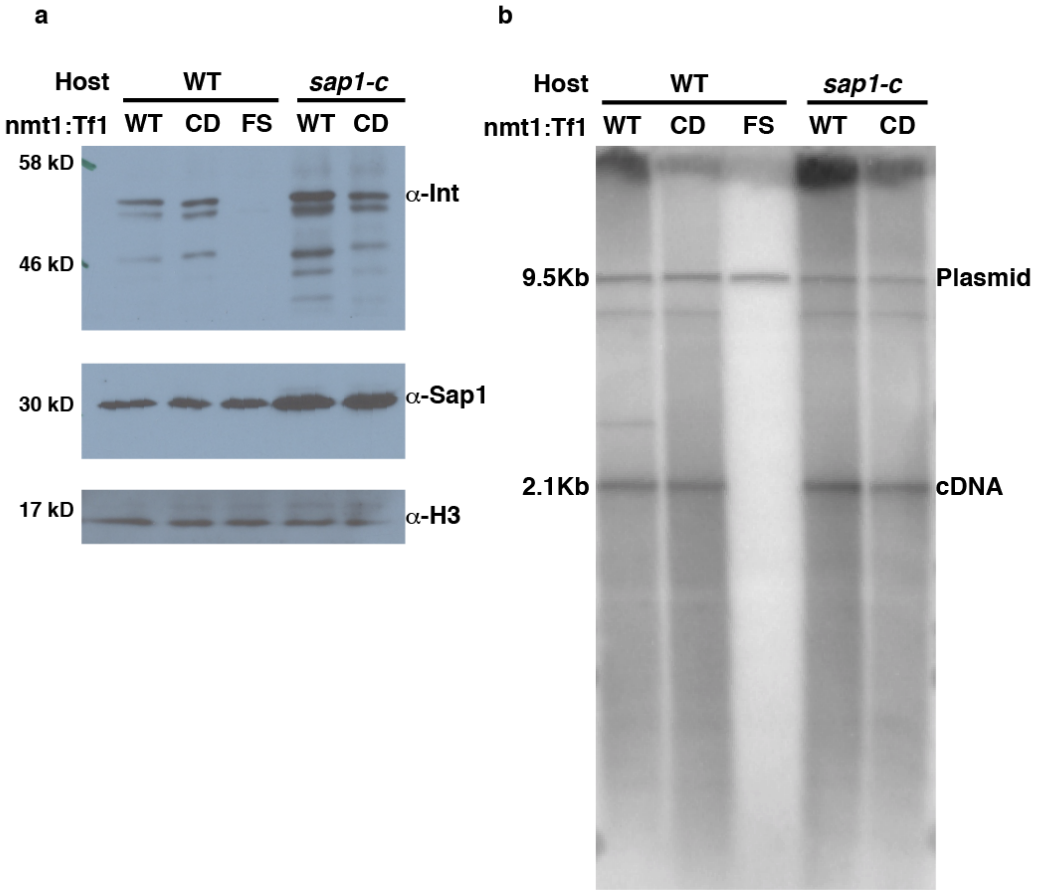


Figure 6 - Transposase expression and cDNA production are normal in WT and *sap1-c* mutants. WT: wild type, CD: catalytic dead, FS: frameshift mutation. (a) Integrase and Sap1 western blot. H3: Histone 3. (b) cDNA southern blot

Tf1 insertions are asymmetric and periodic around Sap1 binding sites

Sap1 is an essential factor with functions affecting genome integrity during DNA replication(de Lahondes et al., 2003). It has a demonstrated role in forming directional, i.e. orientation dependent, replication fork barriers (RFB) (Krings, 2005; Mejia-Ramirez et al., 2005) that arrest replication fork progression in one direction but allow passage in the opposite direction.. We plotted Tf1 insertion density around Sap1 binding motifs taking into account their orientation (Fig. 2A). Sap1 binding motifs exhibit enrichment of insertions around them (Figure 7B) indicating that Sap1 binding directs transposition but protects its footprint. Strikingly, most insertion events occurred 3' of the Sap1 binding motif (Wilcoxon signed rank test [5,99<mu<7,99] 95%CI, p<2e-16, n=888), displaying a prominent periodicity of peaks (Figure 7B) that was also observable in autocorrelation analysis of insertion sites genomewide (Figure 8). This is the side of the motif where confirmed instances of Sap1 dependent RFB cause fork arrest(Krings & Bastia, 2006; Mejia-Ramirez et al., 2005; Zaratiegui et al., 2010). Moreover, in both confirmed Sap1 dependent RFB, the replication terminator Ter1 located at rDNA(Krings, 2005; Mejia-Ramirez et al., 2005) and the solo LTR interspersed in the genome(Zaratiegui et al., 2010), most insertions occurred on the blocking side of the Sap1 barrier, suggesting that the RFB influences site selection (Figure 7, C and D).

Consistently, Tf1 insertion hotspots and Sap1 binding regions coincide with domains of γ−H2A deposition and with DNA Pol ε (Cdc20) maxima in undisturbed S-phase, both markers of replication fork arrest(Rozenzhak et al., 2010; Sabouri et al., 2014) (Figure 9). Since Sap1 fork barrier activity is not a function of binding affinity but of binding site structure(Krings & Bastia, 2006), this observation could potentially

explain the variability in transposition competence of Sap1 binding sites (Figure 4a,b), and why the *sap1-c* allele, which only modestly lowers DNA binding but severely affects RFB activity(Zaratiegui et al., 2010), so dramatically decreases Tf1 transposition (Figure 3D).

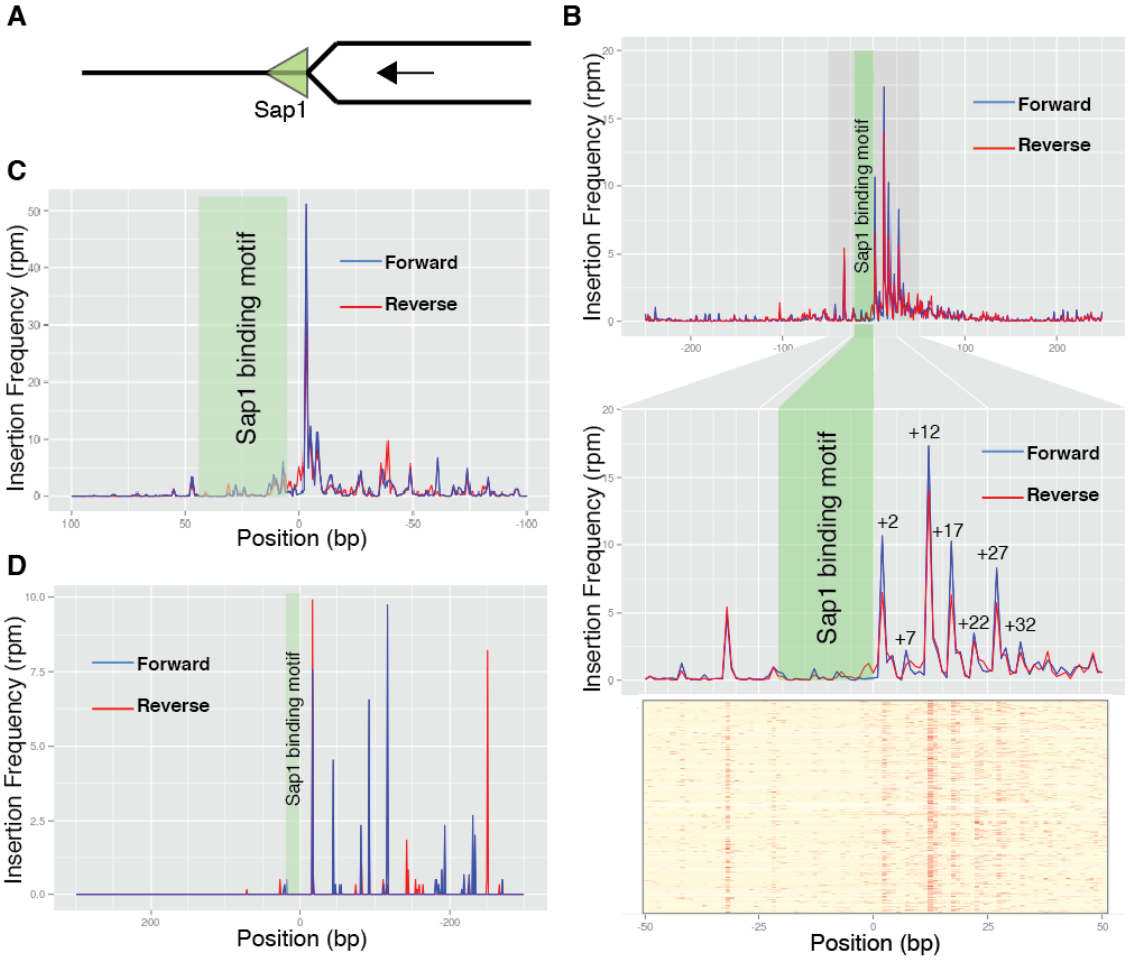


Figure 7 – Tf1 insertion profiles around Sap1 binding motifs. (A) Sap1 binding motifs are oriented as blocking replication forks advancing in the right-to-left orientation. Forward insertions (blue) place their coding strand in the top strand, Reverse (red) in the bottom strand. (B) Averaged transposition frequency around Sap1 binding motifs (n=888). Upper panel: 500bp window; middle panel: 100bp zoom-in window; lower panel: 100bp heat-map of individual motifs. (C) Averaged transposition frequency around Tf2 LTR (n=152). (D) Averaged transposition frequency around Ter1 (n=3)

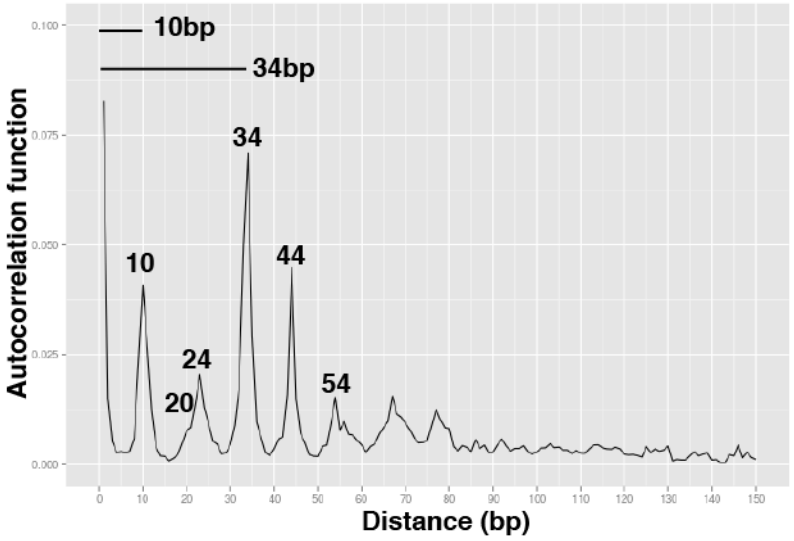


Figure 8 - Autocorrelation plot of insertion points, showing two periods: one of 10bp, corresponding to consecutive insertion hotspots, and one of 34bp, corresponding to the distance between the two hotspots on either side of the Sap1 binding motif.

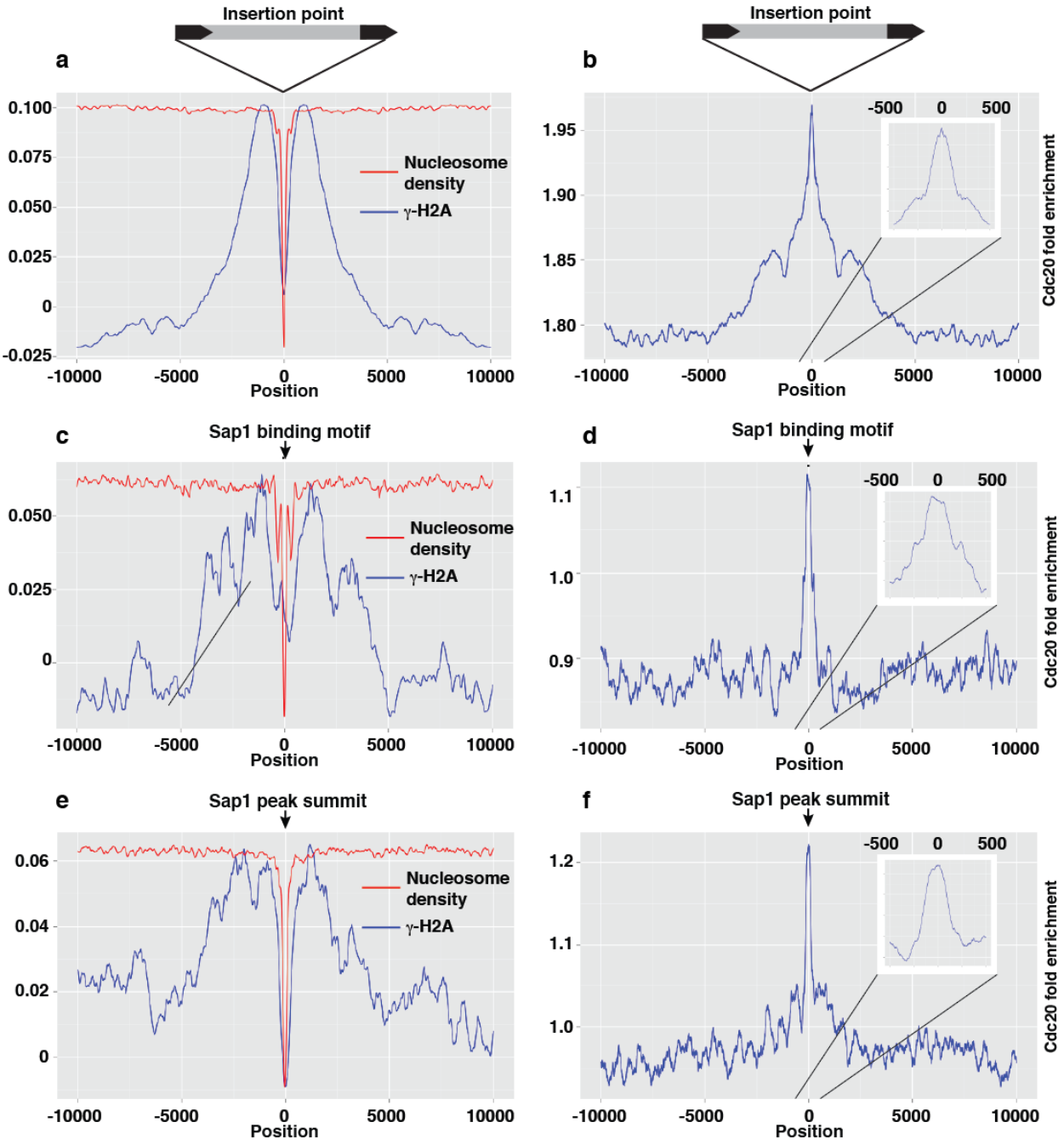


Figure 9 - Tf1 insertions colocalize with marks of endogenous replication fork arrest in undisturbed S phase. Insertions (a), Sap1 binding motifs (c) and Sap1 peaks (e) are placed at the center of a ~10Kb region of g-H2A deposition, signaling replicative DNA stress. The peak is notched in the center because of the absence of nucleosomes in the Sap1-determined NFR (Tsankov et al., 2011). Accordingly, Insertions (b), Sap1 binding motifs (d) and Sap1 peaks (f) exhibit a sharp peak of Cdc20/DNA polymerase ε catalytic subunit. Cdc20 data from (Sabouri et al., 2014), γ-H2A data from (Rozenzhak et al., 2010) and nucleosome data from (Tsankov et al., 2011).

Sap1 barrier activity and not binding strength influences Tf1 transposition

To test the hypothesis that Sap1-dependent fork arrest guides Tf1 insertion, we assessed whether the transposition competence of Sap1 binding sites correlates with the intensity of Sap1 binding or with their RFB activity. We tested the influence of Sap1 binding site orientation with respect to fork progression on transposition efficiency in wild type cells, using three well characterized Sap1 binding sites: (i) the rDNA replication terminator Ter1(Krings, 2005; Mejia-Ramirez et al., 2005), a very efficient RFB; (ii) the synthetic sequence DR2, derived from in vitro Sap1 binding selection(Ghazvini et al., 1995) but an inefficient RFB; and (iii) DR2D, a mutation of DR2 that restores its RFB activity (Krings & Bastia, 2006). We introduced these Sap1 binding sites in one of the two orientations (Blocking: B; or Non-Blocking: NB) into autonomously replicating plasmids, in close proximity to a replication origin so as to control the predominant direction of fork progression over the motif. We then used these plasmids as transposition acceptors in a targeting assay (Leem et al., 2008) (Figure 10A). The results are summarized in Figure 10B. 2D native-native gel electrophoresis of replication intermediates confirmed that Ter1 and DR2D, but not DR2, are efficient RFB in their blocking orientation (Figure 10B). ChIP analysis showed little difference in Sap1 enrichment between the two orientations of each motif but revealed that DR2 is the strongest Sap1 binder, with DR2D as the weakest and Ter1 showing intermediate enrichment (Figure 11A). The results of the transposition trap experiment show that RFB competency (Ter1 and DR2D, but not DR2) as well as blocking ability (B orientation) determined higher transposition frequency into the target site (n=3 biological replicates per condition, Tukey Range test, p=0.0006). Importantly, all insertions displayed 5bp

target site duplications (TSDs, not shown), indicating that they were integrase-mediated transpositions and not the result of arrested-fork induced recombination {Hoff:1998wv}. These results indicate that transposition into Sap1 binding regions depends not on their Sap1 binding affinity but on their efficiency as RFB.

We next tested if the effect of target site orientation extended to genomic positions. We set up a transposon trap system in which the target site is placed inside an artificial intron in the reporter gene *ura4*, allowing selection of insertions by treatment with the counterselection drug 5-Fluoroorotic acid (5-FOA). *ura4* is passively replicated by forks approaching from two nearby replication origins on its centromeric side (Lambert, Watson, Sheedy, Martin, & Carr, 2005) allowing us to correlate the target site efficiency with its competence as a RFB (Fig. 3C). Blocking (B) or non-blocking (NB) orientations of Ter1 showed equal binding of Sap1 (Figure 11B). However, insertion frequency was 10 fold larger in the Ter1 motif placed in the blocking orientation (t-test, $p<0.001$, $n=4$ biological replicates per condition). Once again, all insertions exhibited TSDs (not shown). We conclude that the efficiency of insertion near a Sap1 binding motif depends on its ability to cause fork arrest.

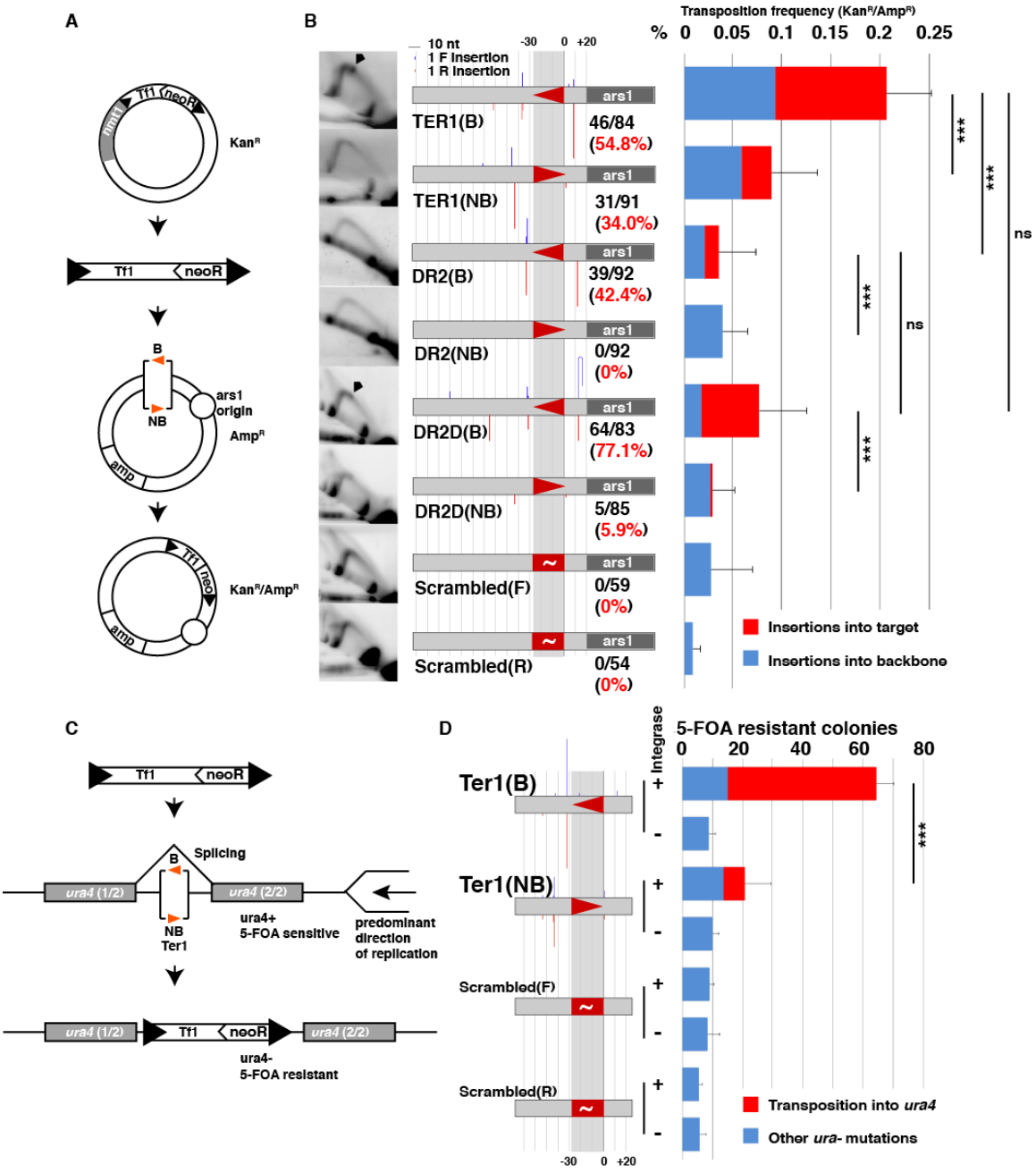


Figure 10 - Transposition competence of Sap1 binding sites depends on RFB activity. (A) Plasmid transposition trap strategy. (B) Transposition into Ter1, DR2, DR2D and Scrambled binding motifs in plasmid transposition trap assay. Left column: 2D gel electrophoresis; RFB signals are marked with an arrowhead. Middle column: diagram of target site with insertion sites depicted as columns (blue in forward, red in reverse orientation) at the insertion position of height proportional to number of insertions. Orientation of the Sap1 binding site is depicted by triangles, in blocking (B, pointing left) and non-blocking (NB, pointing right) orientations. The fraction of insertions into the target site (defined as a 150bp window around the Sap1 binding motif) over the total number of sequenced plasmid insertions is depicted as a fraction and percentage in red

numbers. Right column: Frequency of transposition into the plasmid (Kan^R/Amp^R plasmids over Amp^R total plasmids), with insertions into the target site as defined as a 150bp window around the binding motif in red, and insertions into the plasmid backbone in blue. Error bars depict s.d. (C) Intron transposition trap strategy. (D) Insertion into Ter1 and Scrambled binding motifs in intron transposition trap assay. Diagram of motif arrangement and insertions as in (B). Proportion of 5-FOA resistant colonies due to transposition into *ura4* in red, due to other mutations in blue.

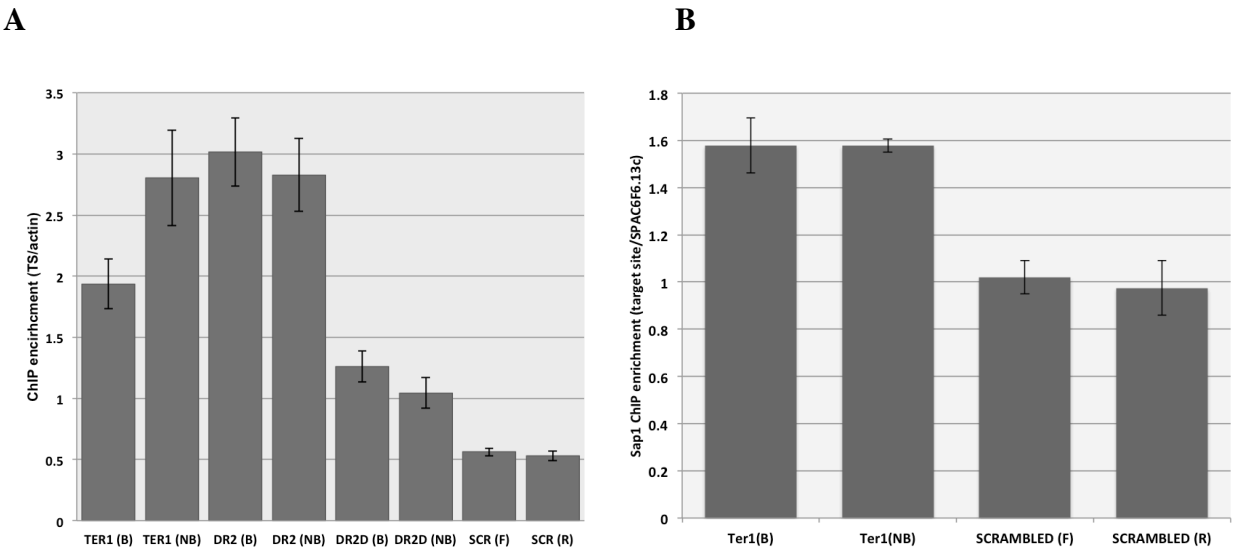


Figure 11 - Sap1 ChIP performed on transposon trapping targets. (A) Plasmid trap target sites. (B) Intron trap target sites.

Sap1 binding and barrier activity are necessary and separate requirements for Tf1 recruitment

These observations prompted us to examine how Sap1, the Tf1 intasome (integration complex), and the replication fork interact. Sap1 could be influencing Tf1 target site selection through direct interaction with the integrase, or through interactions between Sap1 binding sites within the LTR and genomic bound Sap1. We split Sap1 into functional domains and tested for interactions with the full length Tf1 integrase with a yeast two-hybrid assay, which revealed that the Tf1 integrase interacts directly with the dimerization domain of Sap1(Ghazvini et al., 1995) (Figure 12). To evaluate the role of this interaction and the arrested fork in Tf1 transposition, we turned to chromosome conformation capture (3C) to measure tethering of mature cDNA at Sap1 dependent and independent RFB (Fig 4A). Sap1 bound to Ter1 in the blocking orientation led to prominent recruitment of Tf1 cDNA, while the non-blocking orientation was unable to recruit (Figure 13B, Ter1 B/NB panels). Tethering to Ter1 was also dependent on WT Sap1 and the presence of Tf1 integrase in the intasome (Figure 13B, *sap1-c*, Δ*int* panels). This suggests that the direct interaction of Sap1 with Integrase (Figure 12) participates in intasome recruitment, and that Sap1 bound to cDNA (Figure 14) through its cognate binding sequences in the LTR(Zaratiegui et al., 2010) is not sufficient to localize the intasome by multimerization with genome-bound Sap1. A Sap1-independent RFB (Ter2, dependent on the DNA binding factor Reb1 (Sánchez-Gorostiaga, López-Estraño, Krimer, Schvartzman, & Hernández, 2004)) did not tether the cDNA (Figure 13, Ter2B/ Ter2NB panels). This is consistent with our genome-wide observations that class III genes and other Sap1 independent RFB are not hotspots for Tf1 transposition despite

causing polar fork arrest (not shown) (Sabouri et al., 2014; Sánchez-Gorostiaga et al., 2004). Combined, these results suggest that integrase, Sap1, and fork barrier activity must be present to tether Tf1 cDNA to the target site and guide insertion.

We next tested if the RFB and Sap1 binding requirements are separable. If so, we could rescue insertion into a non-RFB Sap1 binding site by providing an independent RFB in *cis*. We cloned Ter2 next to the DR2 binding site placed in the non-blocking orientation (Figure 13C). The presence of Ter2 in either orientation did not change the binding of Sap1 to DR2 (Figure 15). Ter2 rescued the targeting efficiency of DR2 only when the former was placed in the blocking orientation (Mann-Whitney U test, $p<0.001$, n=3 biological replicates per condition, Fig. 4B). Part of the increase in targeting to DR2 could be caused by replication forks converging onto the Ter2 blocked fork to complete S phase, approaching DR2 in the blocking orientation. Accordingly, insertions are detectable on the blocking side of DR2 (Figure 13D). However, transposition also occurred near Ter2 into the side of the motif where Reb1 stops the fork, suggesting that features of the arrested fork, and not the location of binding sites, are the major determinants of target site choice. Together, these results reveal that Tf1 transposition targeting requires two separable conditions, both necessary but neither sufficient: (1) Sap1 binding and (2) an active RFB.

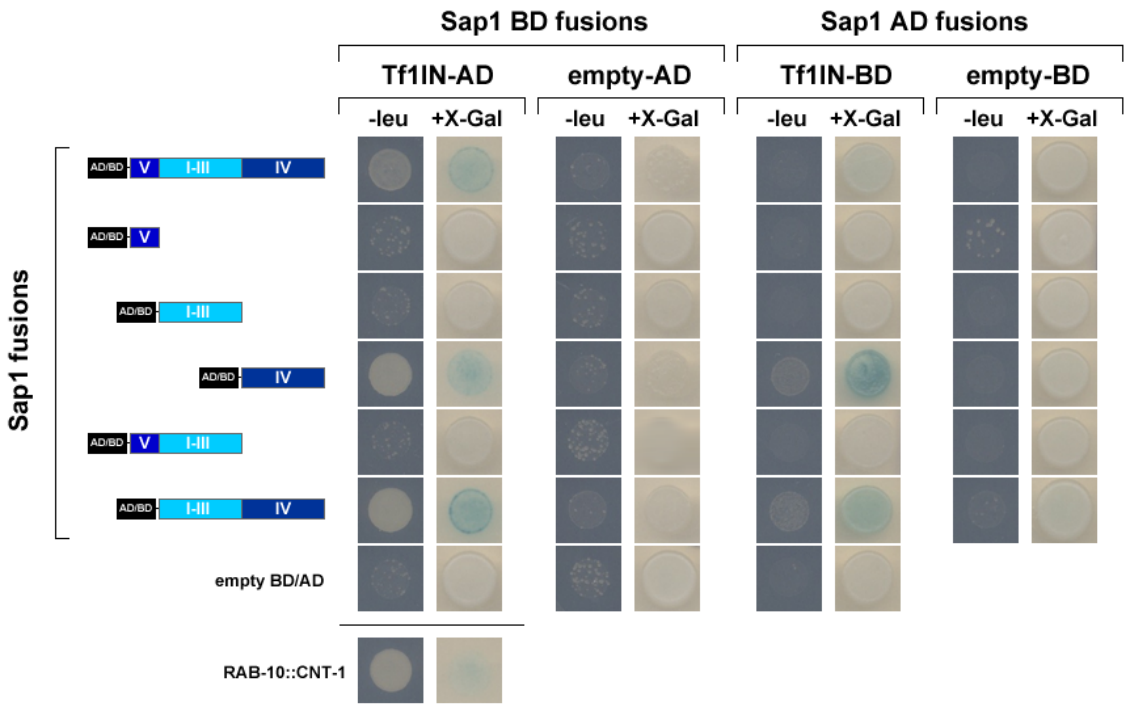


Figure 12 - Tf1 integrase (Tf1IN) directly interacts with domain IV (dimerization domain) from Sap1 in a yeast 2-hybrid assay. LexA activation domain (AD), LexA DNA binding domain (BD) fusion proteins assayed are indicated above. The RAB-10::CNT-1 interaction was used as a positive control (Shi et al., 2012).

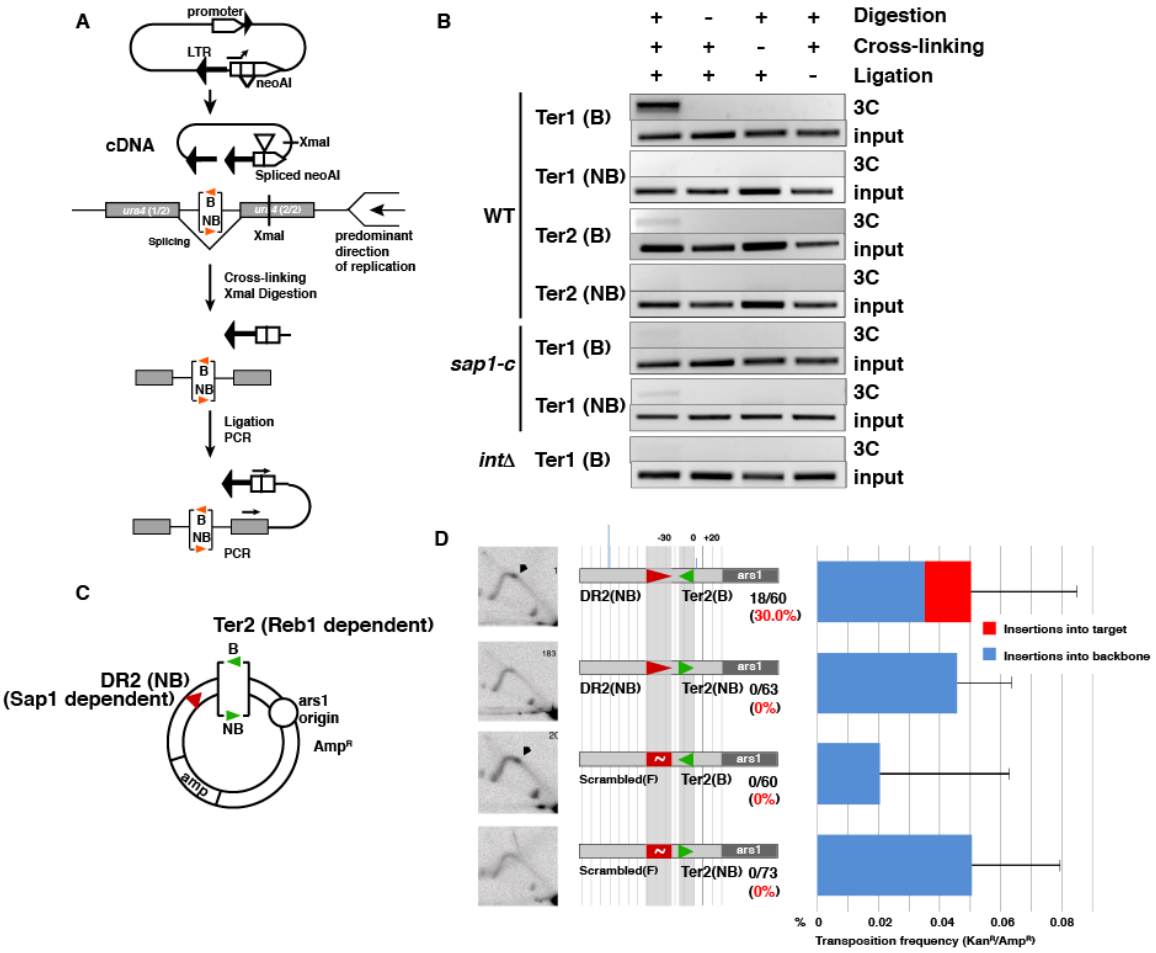


Figure 13 - Sap1 binding and RFB activity collaborate to tether the intasome. (A) Strategy for 3C analysis of cDNA tethering. (B) cDNA tethering at Sap1 dependent (Ter1) and independent (Ter2) RFB in WT, *sap1-c* mutant and integrase frameshift Tf1 mutant. (C) Strategy for separation of Sap1 binding and RFB activities. (D) Results of plasmid trap assay. Left column: 2D gel electrophoresis. Middle column: Diagram of arrangement of Sap1 (DR2 and Scrambled) and Reb1 (Ter2) binding motifs, insertion points and frequency depicted as in Fig. 10B. Right panel: Transposition frequency into target plasmid as in Fig. 10B.

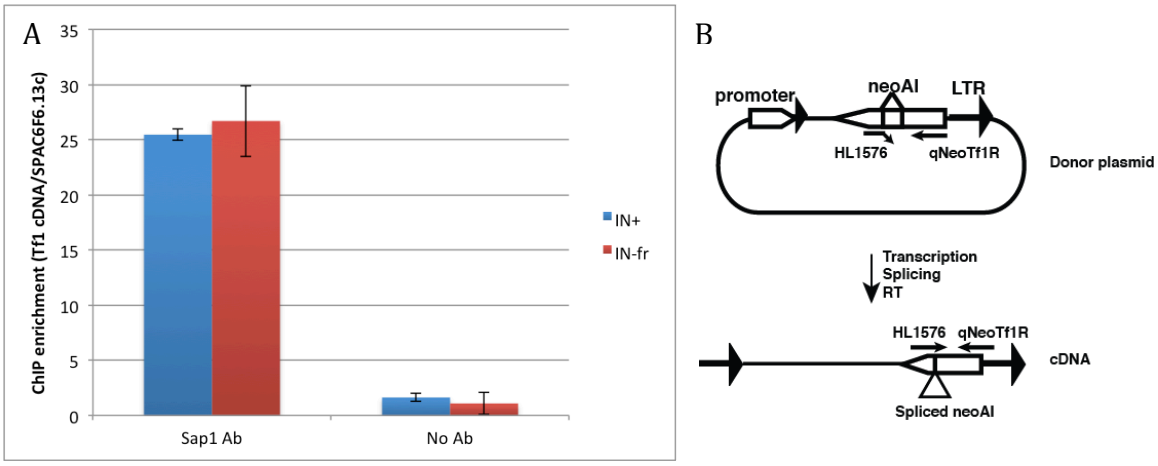


Figure 14 - Sap1 is bound to cDNA and its binding is not affected by the presence of integrase. (A) ChIP enrichment of Sap1 in a strain containing WT Tf1 (IN+) or a version of Tf1 that contains a frameshifted integrase (IN-fr). Both versions of Tf1 produce comparable amount of cDNA (Atwood et al., 1998). (B) Schematic demonstrating how the qPCR distinguishes between plasmid and cDNA sequences (Chatterjee et al., 2009)

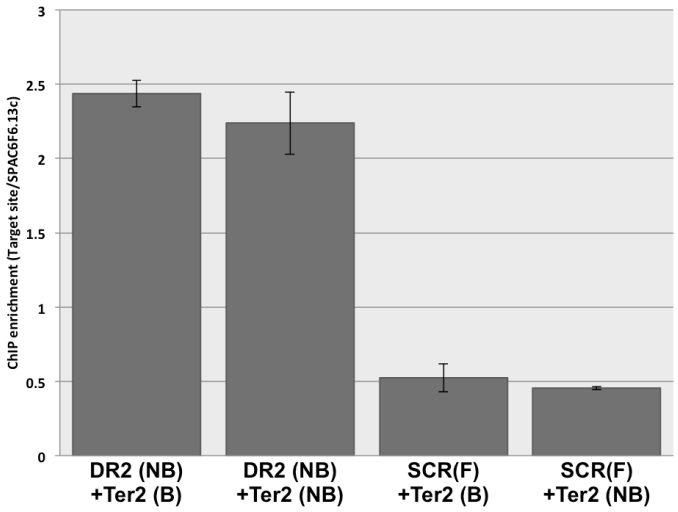


Figure 15 - Sap1 binding is not affected by a nearby independent RFB.

4. Discussion

These results strongly suggest that Sap1 and replication fork arrest are the main determinants of Tf1 tropism in *S. pombe*. Sap1 is correlated with Tf1 insertions genome wide, but mutations in Sap1 that severely affect RFB, not binding, dramatically affect integration efficiency and genome-wide insertion profiles. Beyond correlation, direct studies of Sap1 binding on Tf1 transposition using plasmid and genomic trapping assays reveal that maximal Tf1 targeting is achieved when Sap1 is placed in an orientation capable of RFB. Further, we show that the Sap1 C-terminal domain directly interacts with the IN, and that this interaction is likely responsible for cDNA tethering to the genome.

Other LTR retrotransposons may display similar preference for arrested replication forks. The *S cerevisiae* LTR retrotransposons Ty1 (Copia group) and Ty3 (Gypsy group) insert upstream of RNA Pol III transcribed genes like tRNA and 5S(Devine & Boeke, 1996; Kirchner et al., 1995), which are confirmed RFB(Deshpande & Newlon, 1996), and several regulators of fork progression suppress Ty element retrotransposition(Bairwa, Mohanty, Stamenova, Curcio, & Bastia, 2011; Baller et al., 2012). Similarly, other LTR retrotransposons with insertion preference for heterochromatin(Tsukahara et al., 2012) might use fork stalling at satellite repeats in pericentromeric DNA(Zaratiegui et al., 2011). Like Tf1, insertion hotspots for *S cerevisiae* LTR retrotransposons Ty1 and Ty3 (Baller et al., 2012; Mularoni et al., 2012; X. Qi et al., 2012) also coincide with accumulation of γ-H2A and DNA polymerase ε (Szilard et al., 2010) (Figure 16), indicating that the association between insertion and

replication fork arrest is conserved in LTR retrotransposons of the Gypsy and Copia groups.

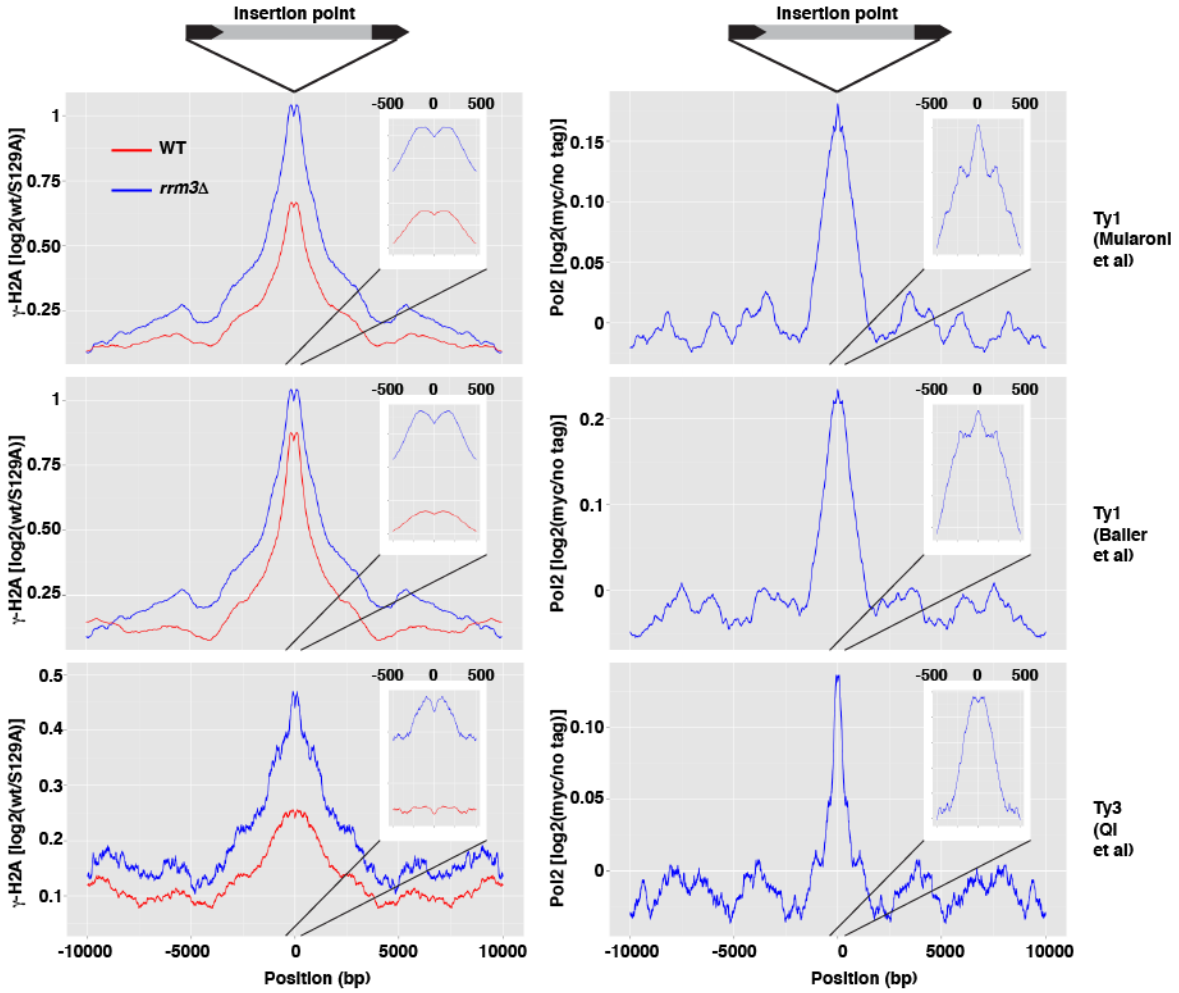


Figure 16 - Ty1 and Ty3 colocalize with marks of endogenous replication fork arrest in undisturbed S phase in *S. cerevisiae*. Left, γ-H2A in WT/*rrm3*Δ; Right, DNA Pol2 enrichment in WT. Insertion data from (Baller et al., 2012; Mularoni et al., 2012; X. Qi & Sandmeyer, 2012); γ-H2A in WT, *rrm3*Δ, and DNA Pol2 ChIP data from (Szilard et al., 2010).

Chapter III – Genetic Examination of the Role of RFB on Tf1 Transposition

1. Introduction

Despite large differences in DNA origin structure, genome size, and complexity, the core DNA replication machinery is highly conserved among all Eukaryotes (Bell & Dutta, 2002). Upon each cell division, thousands of DNA replication origins are forced to coordinate to faithfully copy the genome exactly once, and with as few errors as possible. This process is fraught with challenges, as demonstrated the ubiquity of diverse pathways of DNA checkpoint and repair pathways that exist to protect our genomes from mitotic catastrophe (Canman, 2001). Part of the difficulty involved in the replication of genomes is the existence of replication fork barriers (RFB). These barriers exist naturally at several loci in the genome, like rDNA (Krings, 2005), but can also form from head-to-head collisions between the transcription and DNA replication machinery, forming structures called R-loops (Aguilera & García-Muse, 2012). Consistently, some highly expressed genes have been shown to cause fork arrest (Azvolinsky, Giresi, Lieb, & Zakian, 2009). RFB can also occur from G-quadruplex formation (Sabouri et al., 2014), depletion of nucleotide precursors, or DNA adducts (Barlow et al., 2013).

RFB have the ability to both stabilize and destabilize the genome. The influence of natural RFB on processes of genome stability was first studied in the *E. coli Ter*-Tus system, a natural RFB that ensures replication termination occurs at the correct locus (Mulcair et al., 2006). Early studies on this complex revealed that these *Ter* sites were highly recombinogenic, but only when the Tus protein could facilitate formation of RFB (Horiuchi, Fujimura, Nishitani, Kobayashi, & Hidaka, 1994). RFB can also act to suppress recombination. RFB activity at rDNA prevents repeats from being highly

recombinogenic (Takeuchi, Horiuchi, & Kobayashi, 2003), and RFB at transposon LTR direct specific DNA repair activities around them (Zaratiegui et al., 2010). However, the link between RFB and recombination is not entirely always clear, and RFB intensity has been unlinked to recombinogenic potential at some RFB (Pryce, Ramayah, Jaendling, & McFarlane, 2009). Until the various pathways controlling RFB establishment, bypass, and resolution can be reconciled, the outcome of RFB on a particular cellular process therefore needs to be individually examined.

Cells have evolved a variety of pathways designed to minimize the impact of RFB. Studies in yeast have found that replisomes are stabilized at RFB by cellular checkpoint factors, presumably to keep the replisome anchored for subsequent DNA repair events, and to protect the reactive replication intermediates (Katou et al., 2003). A fork stalled by this pathway can then either wait for the obstacle to be cleared, or remain stable long enough for a converging fork to complete synthesis of the region. In locations of the genome where DNA synthesis is largely unidirectional, however, this may never happen, and cells lean on fork restart pathways, like homologous recombination (HR), for viability. HR has been observed to be necessary for viability in a number of scenarios where natural RFB occur—including the centromere (Zaratiegui et al., 2011), the *S. pombe* mating-type locus (Roseaulin et al., 2008), and even artificial RFB created in regions of the genome where replication is unidirectional (Lambert et al., 2005). HR-dependent fork restart is highly error-prone and can lead to gross-chromosomal rearrangements, particularly at repetitive DNA where fork arrest is common (Iraqui et al., 2012; Mizuno, Miyabe, Schalbetter, Carr, & Murray, 2013). Because of this, these pathways likely represent a last resort in the face of an impenetrable RFB.

Several proteins coordinate to slow down the progression of S-phase in the face of replication stress. In fission yeast, this intra-S checkpoint is mediated through activation of Cds1, and activation of Cds1 depends on Mrc1, a member of the fork protection complex (FPC). Mrc1 preferentially binds branched DNA structures (H. Zhao & Russell, 2004), suggesting that it recognizes structural components of stalled forks to activate Cds1. *In vitro* studies show that Mrc1 binding to DNA is enhanced by other members of the FPC, the Swi1 and Swi3 heterodimer (Tanaka et al., 2010). Swi1 and Swi3 were originally identified as proteins responsible for mating-type switching, a process that requires a series of programmed RFB and subsequent directed gene conversion events to successfully complete (Dalgaard & Klar, 2000). The Swi1-Swi3 heterodimer is also important for establishment of the important RFB at the TER1, TER2, and TER3 barriers at rDNA (Krings & Bastia, 2004), and the mating-type locus (Dalgaard & Klar, 2000; E. Noguchi, Noguchi, McDonald, Yates, & Russell, 2004). Mrc1 is not responsible for RFB (Calzada, Hodgson, Kanemaki, Bueno, & Labib, 2005).

The FPC is essential for genome stability, and mutants are hypersensitive to DNA damaging agents (C. Noguchi & Noguchi, 2007). In FPC mutants, the loss of RFB at rDNA results in high levels of recombination and loss of rDNA repeats (Sommariva et al., 2005). While Swi1-Swi3 independent RFB, like tRNA, have been shown to exist in the genome, Swi1-Swi3 still play a role in mediating genome stability at these RFB, as fork collapse and HR is more common in FPC mutants at these loci (Pryce et al., 2009). Swi1-Swi3 thus probably play a global role in protecting the replication fork, even outside of RFB they are not responsible for forming. Consistent with this theory, Swi1-Swi3 track with the replication fork and directly interact with several replisome

components (McFarlane, Mian, & Dalgaard, 2010; E. Noguchi et al., 2004). These interactions are thought to play important roles in fork protection through the coordination of both helicase progression and leading and lagging strand DNA synthesis (McFarlane et al., 2010). This link between helicases and DNA synthesis is partially potentiated by the fourth member of the FPC, Mcl1. Mcl1 has been poorly studied in *S. pombe*, but its homologue in *Xenopus laevis* links helicases to DNA polymerase alpha progression (Errico et al., 2009). In line with these roles, fork progression is delayed in FPC mutants. As either a cause of consequence of this delay, replisome components are degraded by a ubiquitin-dependent pathway in Swi1-Swi3 mutants (Roseaulin et al., 2013).

The helicase Pfh1 also has been implicated in the resolution of RFB in *S. pombe*. Pfh1 is an ATP-dependent 5'→3' DNA helicase in the Pif1 family of helicases, a family that exists in almost all Eukaryotes (Bochman, Sabouri, & Zakian, 2010). Pfh1 has been shown play critical roles in the bypass of difficult-to-replicate regions of the genome, and its depletion leads to accumulation of X-shaped termination structures and increases in fork arrest at a variety of protein-DNA barriers (Sabouri, McDonald, Webb, Cristea, & Zakian, 2012; Steinacher, Osman, Dalgaard, Lorenz, & Whitby, 2012). Loss of Pfh1 has been linked to fork stalling at some highly expressed genes, class III genes (Sabouri et al., 2012), converged replication forks (Steinacher et al., 2012), and recently, G-quadruplexes (Sabouri et al., 2014). In *pfh1* mutants, stalled fork intermediates accumulate at these sites, and in turn, these sites become hyper recombinogenic and contribute to genome instability (Steinacher et al., 2012; Zhou et al., 2013). Aside from mediating RFB bypass, genetic and biochemical analysis shows that Pfh1 also plays a role in Okazaki fragment

maturation, probably by promoting the dissociation of endonuclease-resistant flaps from the lagging strand (Ryu, 2004).

Pfh1 and the FPC complex members Swi1 and Swi3 thus have similar effects on suppressing recombination at RFB, but opposite effects on RFB establishment. These opposite effects make it somewhat surprising that Swi1 and Pfh1 are synthetically lethal, as many RFB that are enhanced by Pfh1 depletion are also dependent on the FPC (Sabouri et al., 2012). Thus, this synthetic lethality likely results from the increased severity of Swi1-independent fork barriers in Pfh1 mutants. As previously mentioned, Swi1-independent fork barriers still depend on Swi1 for suppressing recombination at these loci, suggesting that the FPC plays a global role in fork management (Pryce et al., 2009). This global role is likely even more important in Pfh1 depleted cells, where RFB activity is likely enhanced genome-wide.

The opposite effects of Pfh1 and Swi1-Swi3 prompted us to look at how these proteins influenced Tf1 transposition. Tf1 transposition has been recently linked to Sap1-dependent RFB in *S. pombe,* indicating that features of the RFB are important for targeting (Jacobs et al., 2015). If the Tf1 transposon is recognizing substrates of the stalled fork, then it is possible that enhancement or depletion of these RFB through deletion of the FPC or Pfh1 would influence target site selection or efficiency. To this end, we mapped Tf1 insertions genome-wide in deletion strains of for Swi1, Swi3, or strains depleted of Pfh1, and compared Tf1 insertion patterns, autocorrelation, and transposition frequency to previously generated datasets from wild-type and *sap1-c* strains.

2. Materials and Methods

Sap1 ChIP-seq

ChIP-seq was done according previously published protocols adapted for *S. pombe* (Rosado-Lugo & Hampsey, 2014). Briefly, cells were grown to OD600=0.8, then fixed in 1% formaldehyde for 20' before quenching in 125mM glycine for 5'. Pellets were then washed in 1xTBS (50mM Tris-HCl, 150mM NaCl, pH=7.5) + 1% TritonX-100, then were washed in 1xFA lysis (50mM Tris-HCl, 200mM NaCl, 1mM EDTA, 0.1% sodium deoxycholate, 1% TritonX-100) with 1mM freshly prepared PMSF. Pellets were then resuspended in 500µL 1xFA lysis, filled with acid-washed glass beads to the meniscus, and broken by bead-beating at 3,000 rpm with 3x cycles of 5' at 4°C. Lysate was then removed and clarified by centrifugation. IPs were rotated overnight at 4°C with 2.5µL Sap1 antibody (Jacobs et al., 2015) per 10µL of Protein A Dynabeads (Sigma). Beads were then washed once with 1mL each of 1xFA lysis, 1xFA lysis with 500mM NaCl, 1xWash buffer (10mM Tris-HCl, 1mM EDTA, 250mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, pH=7.5), and 1xTE (10mM Tris-HCl, 1mM EDTA, pH=7.5). Washes were carried out at RT for 10' each. DNA was then released from the protein-bead complexes by reverse crosslinking overnight at 65°C 1xTES (1xTE + 1%SDS), treated with RNase A and Proteinase K, extracted with phenol/chloroform (pH=8), and ethanol precipitated. Libraries were prepared with 5ng of IPed DNA with the NEBNext Library Kit (New England Biolabs).

ChIP-qPCR

ChIP-qPCR was done as described previously (Chapter 2, Materials and Methods), using

the $\Delta\Delta C_t$ method. For enrichment at TER1 and LTR, qTER-F/R, and qLTR-F/R were used, and compared to enrichment in input samples using a region with low Sap1 binding as measured by ChIP-seq (q6F6-F/R). The qLTR-F/R pair sequences were obtained from (Cam et al., 2007). See Table 5 for the sequences of these oligos.

Quantitative Transposition Assay

Measurements of Tf1-*neo* transposition were performed as previously described (Levin, 1995). Briefly, 4 independent transformants of pHL414 into the appropriate strain (Table 4) were patched into ~2.5cm^2 patches, induced for 4 days at 32°C on EMM-ura, then patched onto 5-FOA to remove pHL414. After two days, colonies are lifted and 10 fold dilutions are prepared in water. The three highest dilutions are plated on YES + G418 (500μg/mL) + FOA (1mg/mL) + 2g/L DO-ura, and the three lowest dilutions are plated on 5-FOA. Frequency is measured by comparing the number of G418/FOA-resistant colonies to FOA-resistant colonies after 3-5 days of growth. Statistical analysis was done in R using the *multcomp* package, and plots were generated with *ggplot2*.

High-throughput sequencing of Tf1 insertion points

High-throughput mapping of Tf1 insertion points was done as previously described but with slight modifications for the Illumina miSeq platform (Yabin Guo & Levin, 2010). Twenty plates containing 16 independent patches were used for all RFB mutants measured. After DNA purification, digestion, and linker ligation, libraries were prepared for high-throughput sequencing by PCR with custom barcoded primers. These primers put the p7 and p5 tags necessary for paired-end Illumina sequencing on the ends. PCR

was done as described previously with Titanium Taq (Clontech) on a 96 well plate (Yabin Guo & Levin, 2010). Library DNA was column purified, and DNA between 130bp and 500bp was isolated by gel purification on 2% agarose.

Bioinformatic Analysis

Insertion points were mapped in *bowtie2* and converted into fwig format before being imported into in R for analysis. Data was analyzed with custom scripts and compared to Sap1 ChIP-seq data from (Zaratiegui et al., 2010), and insertion data from (Jacobs et al., 2015). Insertion density box-and-whisker plots were generated by dividing the normalized average number of insertions in each region of significant Sap1 enrichment (MACS >2 (Y. Zhang et al., 2008)) by the number of base pairs that each region spanned. Autocorrelation was performed on the normalized insertions (per million base pairs) data sets with the acf() function in R. Plots were generated with the base R plot package or the *ggplot2* package.

Table 4 – A list of strains used in this chapter

name	mat	genotype
PB1	h90	ura4-D18, leu1-32
ZB1069	h+	sap1-c, ura4-D18
ZB1038	h-	Δswi1::natMX, leu1-32, ura4-D18
ZB1039	h-	Δswi3::natMX, leu1-32, ura4-D18
ZB1173	h+	ade6-M216, leu1-32::pJK148-pfh1-m21,170,265,320,-NES-GFP, Δpfh1::natMX, his3-D1, ura4-D18

Table 5 – Oligonucleotides used in this chapter

name	**5'->3' sequence**
qLTR-F	TGATAGGTAACATTATAACCCAGT
qLTR-R	ACGCAGTTTGGTATCTGATT
qTER-F	GGTAAGGTAGGTCGTGAATCG
qTER-R	ATTTGAAAAGGGGGGAACCAC
q6F6-F	TTGTGCTCTTCATCCTGTGC
q6F6-R	GAATCCGAGATTTCGTCCAA

3. Results

Sap1 binding is largely unchanged in Swi1, Swi3, and sap1-c

Because a full 63% of Tf1 insertion sites occur within Sap1 enriched regions of the genome, we first assayed Sap1 enrichment in these RFB mutants to be sure that differences in binding could not explain differences in the Tf1 integration profile or transposition frequency. We first assessed Sap1 binding in RFB mutants in Δswi1, Δswi3, and *sap1-c* by ChIP-qPCR at two well-characterized Sap1 binding sites that cause RFB, the rDNA and LTR. RFB at TER1 (rDNA) is dependent on Swi1 and Swi3 (Krings & Bastia, 2004), but not *sap1-c* (not shown). The fork pausing signal in the Tf2 LTR is dependent on the FPC (E. Noguchi, personal communication), and is lost in *sap1-c* strains (Zaratiegui et al., 2010). Sap1 binding to TER1 is reduced 10-20% in Δ*swi1* and Δ*swi3*, but unaffected at Tf2 LTR (Figure 17A). This pattern is reversed in *sap1-c*, with binding reduced 20% at Tf2 LTR (Zaratiegui et al., 2010), but unaffected at TER1 (Figure 17A).

We next asked if the large changes in Tf1 transposition frequency and targeting observed in *sap1-c* could be explained by large changes in genome-wide Sap1 enrichment. Tf1 integrations in *sap1-c* occur 10 times less frequently than WT, and integration events move away from Sap1-enriched regions (Jacobs et al., 2015). ChIP-qPCR had revealed only modest decreases in Sap1 binding at LTR, but not TER1 (Figure 17A). However, because the vast majority of insertions occur away from LTR and TER1, we wondered if *sap1-c* affects binding at these other sites. To address these concerns, we performed a Sap1 ChIP-seq study on WT and *sap1-c*. Sap1 binding in WT and *sap1-c* were strongly correlated (Pearson R square = 0.95), with linear regression coefficient of

0.96, indicating that average Sap1 binding decreases ~4% in the *sap1-c* mutant. Sites of significant Sap1 enrichment (MACS fold enrichment >2) showed a significant but mild decrease of Sap1 binding (fold *sap1-c*/WT mean = 0.90, 95% CI [0.86, 0.93], p<2e-9, Wilcoxon signed-rank test, Figure 17B). These observations suggest that the major loss in Tf1 transposition frequency in the *sap1-c* mutant cannot be explained through Sap1 binding alone.

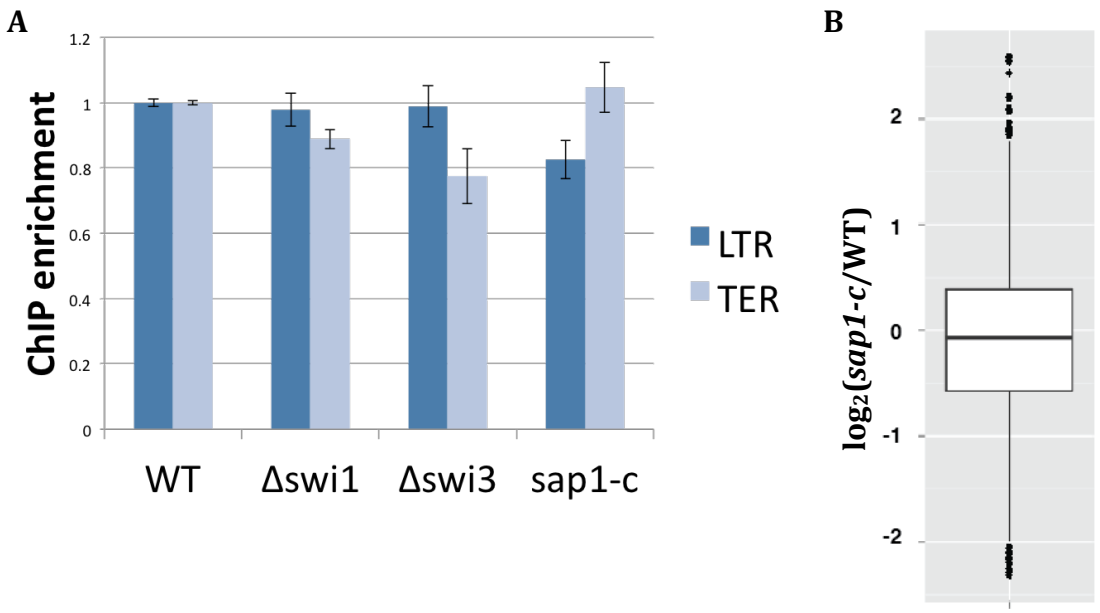


Figure 17 – Sap1 binding is largely unchanged in Δ*swi1*, Δ*swi3*, and *sap1-c*. (A) ChIP-qPCR of Sap1 binding at two well-characterized Sap1 binding sites. (B) Sap1 ChIP-seq box and whisker plot comparison of sites of significant Sap1 enrichment (MACS fold enrichment >2) in WT and *sap1-c* strains.

Tf1 transposition frequency in RFB mutants

We next aimed to measure Tf1 transposition frequency in wild-type, Δ*swi1*, Δ*swi3, pfh1mt**, and *sap1-c* strains. *pfh1mt** is a mutant of Pfh1 that localizes exclusively to the mitochondria, hereby referred to as Δ*pfh1* due to its exclusion in the nucleus (Pinter, Aubert, & Zakian, 2008). If Tf1 is recognizing features of stalled replication forks, then FPC mutants are expected to lose transposition efficiency or targeting, and Pfh1 mutants should gain it. We measured Tf1 transposition frequency in a strain of *S. pombe* overexpressing an antibiotic marked Tf1 (Tf1-*neo*), according to previously published protocols (Levin, 1995). With the exception of *sap1-c*, RFB mutants revealed very minor changes in overall frequencies (Figure 18). Tf1 transposition in Δ*swi3* but not Δ*swi1* was slightly but significantly lower than WT, retaining ~70% of its transposition frequency. In Δ*pfh1,* average transposition was slightly higher, but was very variable between replicates and not significantly different than WT. These data suggest that genome-wide effects on Swi1-Swi3, and Pfh1 RFB do not lead to large changes in Tf1 transposition efficiencies. Because >90% of Tf1 transposition into Sap1 binding sites occurs in the uncharacterized Sap1 binding sites within intergenic regions, not LTR and rDNA, it is possible that RFB at these loci are largely unaffected by mutations in Swi1, Swi3, and Pfh1, but majorly affected by the *sap1-c* allele. These observations warrant further investigation.

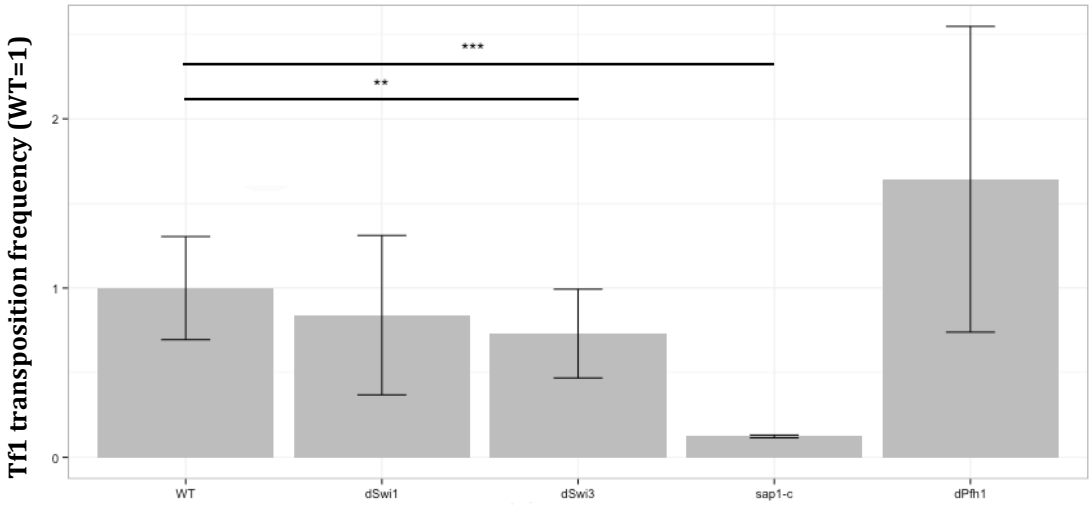


Figure 18 – Tf1 transposition frequencies in WT and RFB mutants. Error bars represent s.d. and asterisks indicate different levels of significance (**, $p<0.05$; ***, $p<0.01$). Significance calculated with ANOVA followed by a *post hoc* analysis with Dunnett's procedure (to WT), N=16 measurements for WT, dSwi1, dSwi3, and *sap1-c*. N=4 for dPfh1.

High-throughput analysis of Tf1 frequency in RFB mutants

We next analyzed the genome-wide insertion profile in these RFB mutants. Loss of the Ty1 targeting factor at the tRNA results in a large relocalization to the subtelomeric repeats without loss in overall transposition frequency (Bridier-Nahmias et al., 2015). Thus, it remained a possibility that Tf1 transposition frequency was unchanged, but targeting was affected. We performed high-throughput Tf1 insertion analysis in WT, Δ*swi1,* Δ*swi3, sap1-c,* and Δ*pfh1* (*pfh1mt**) according to previously published protocols (Yabin Guo & Levin, 2010; Jacobs et al., 2015). A summary of the total mapped insertions per strain is summarized in Table 6.

We next analyzed these insertion profiles in the context of previously known Tf1 targeting factors. Previous analysis revealed the proportion of Tf1 insertions within Sap1 enriched peaks drops from 63.1% in WT to 49.9% in *sap1-c* (Jacobs et al., 2015). This is due to a large decrease in the number of insertions in regions of significant Sap1 enrichment (Figure 19, compare WT to s*ap1-c*). In Δ*swi1,* Δ*swi3*, and Δ*pfh1*, the proportion of insertions in Sap1 enriched regions is unchanged (Table 6). However, the average density of insertions in regions of significant Sap1 enrichment significantly changes between WT and Δ*pfh1,* but not in Δ*swi1* and Δ*swi3* (Figure 19). Unlike in *sap1*-c, this change arises from decreased Tf1 insertion density within some but not all Sap1 enriched regions. This indicates that Pfh1 decreases the efficiency of Tf1 insertion at specific Sap1 sites.

Another hallmark of Tf1 insertions is their asymmetry and periodicity around Sap1 binding sites. We plotted Tf1 insertions around Sap1 binding motifs (n=888), taking into account their orientation, for WT and the RFB mutants we mapped. Insertion around

Sap1 binding sites was unchanged, and the asymmetric profile was preserved, with the majority of insertions occurring on the 3' end of the Sap1 binding motif in all strains (Figure 20). Importantly, Tf1 insertions genome-wide retained their average periodicity of 10 and 34 bp in all tested mutants, indicating that average spacing of insertions is also unchanged (Figure 21).

We next asked whether Tf1 association with Sap1 binding sites changes between WT and RFB mutants. To this end, average Sap1 ChIP enrichment around all Tf1 insertion sites was plotted, and the predictive value of Sap1 binding strength to insertion was analyzed by logistic regression, as done previously (Jacobs et al., 2015) (Figure 22). Relative to WT, Δ*pfh1* and *sap1-c* associated with weaker Sap1 binding sites (Figure 22A), and Sap1 binding strength was less predictive of Tf1 insertion (Figure 22B). In the FPC mutants Δ*swi1* and Δ*swi3*, a more complicated picture arose. Average Sap1 binding around Tf1 insertions increased in Δ*swi1,* but not Δ*swi3*, indicating that insertions in Δ*swi1* associate with stronger Sap1 binding sites *in vivo*. However, despite stronger association with Sap1 enriched regions in the genome, Sap1 binding strength was less predictive of overall insertion in Δ*swi1*, and was unchanged in Δ*swi3*.

Table 6 – Summary of high-throughput Tf1 insertion profiling in RFB mutants

Strain	**Number of insertion sites**	**Number of mapped insertions**	**Percentage of insertions within Sap1 enriched regions (MACS>2)**
WT	67,783	1,985,891	63.1%
Δ*swi1*	44,032	1,795,213	65.5%
Δ*swi3*	62,853	1,875,595	65.7%
sap1-c	180,371	6,418,228	49.9%
Δ*pfh1* (*pfh1mt**)	111,402	4,721,258	64.0%

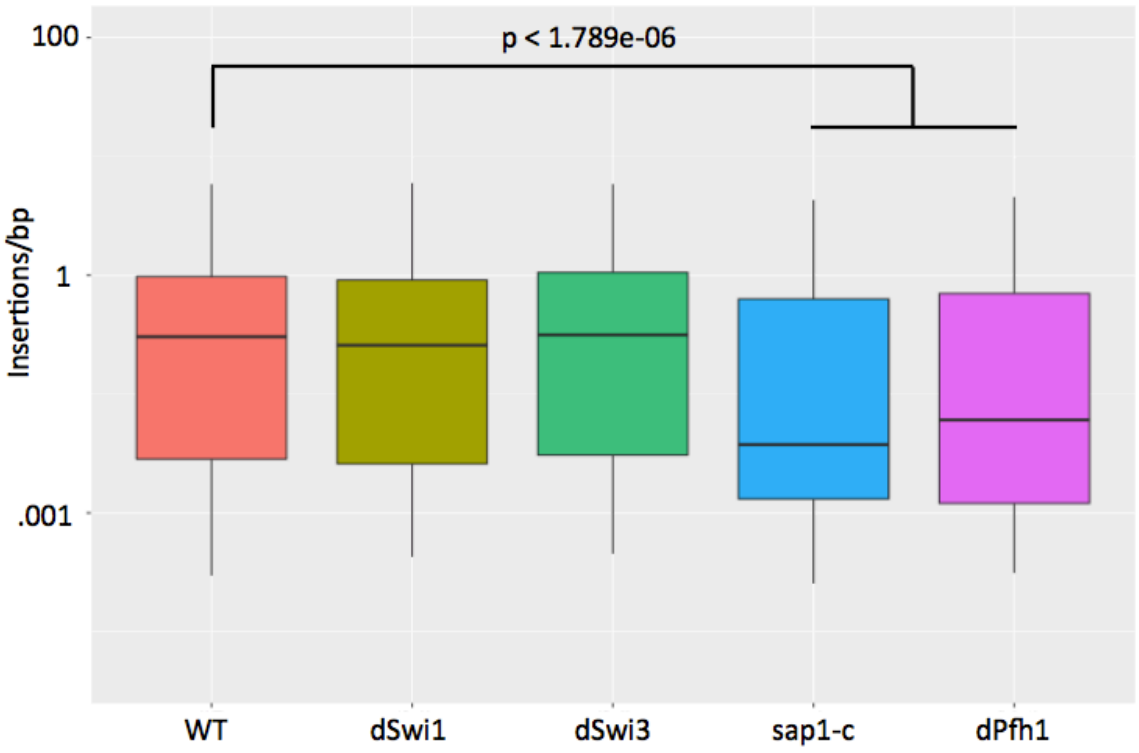


Figure 19 – Averaged insertions/bp in regions of significant Sap1 enrichment (MACS>2). Only dPfh1 and *sap1-c* are significantly different than WT (Mann-Whitney test, p<1.789e-06).

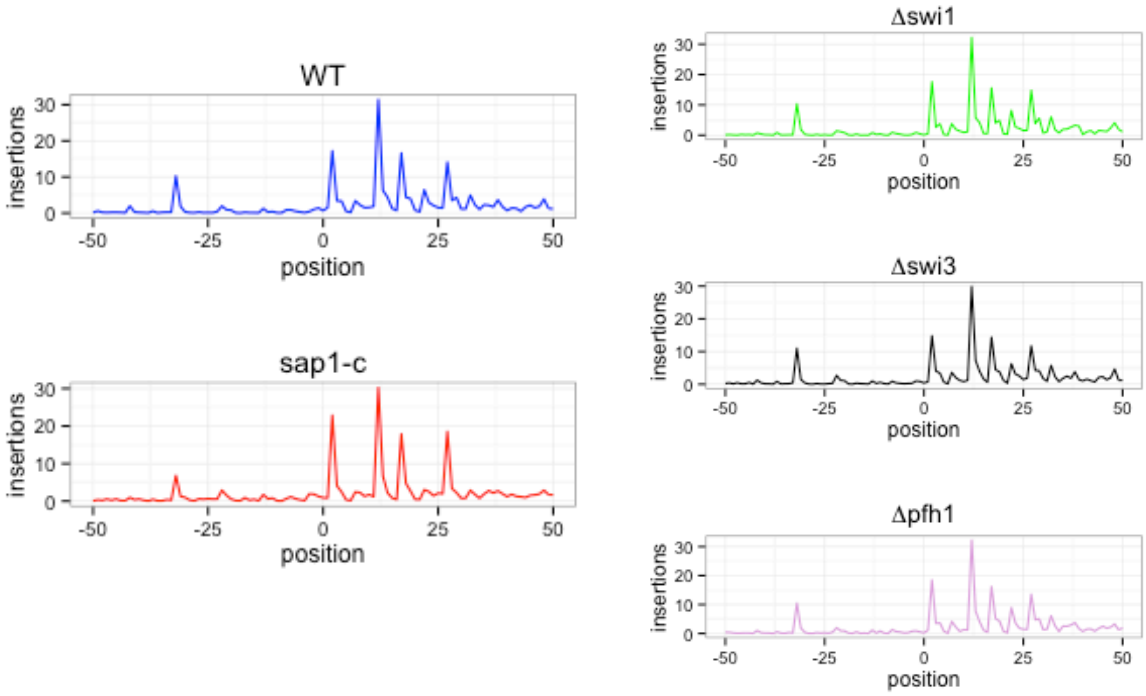


Figure 20 – Averaged Tf1 insertion profile around Sap1 binding sites (n=888) in a 100 bp window. The Sap1 motif is oriented such that forks approaching from the right hand side would be blocked.

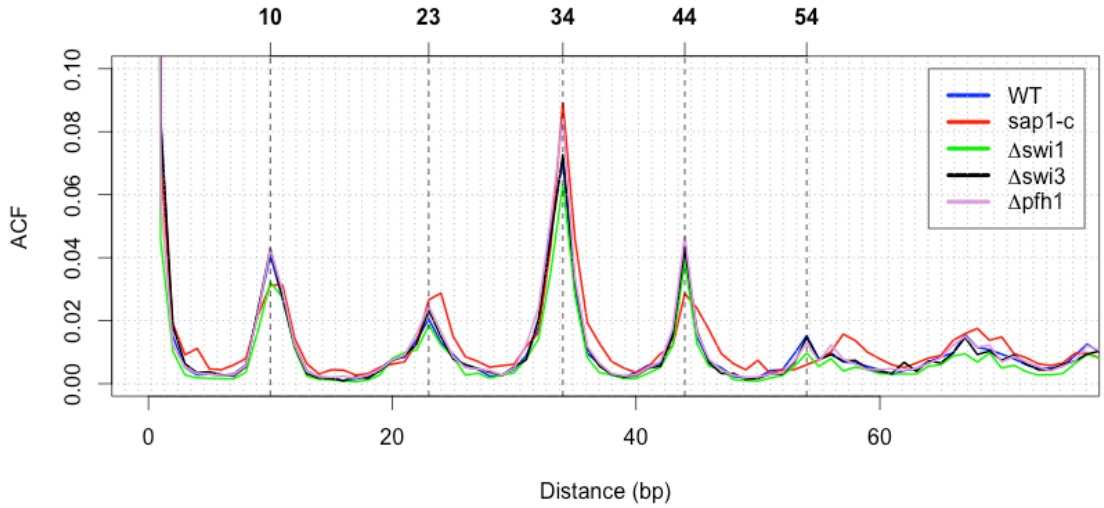
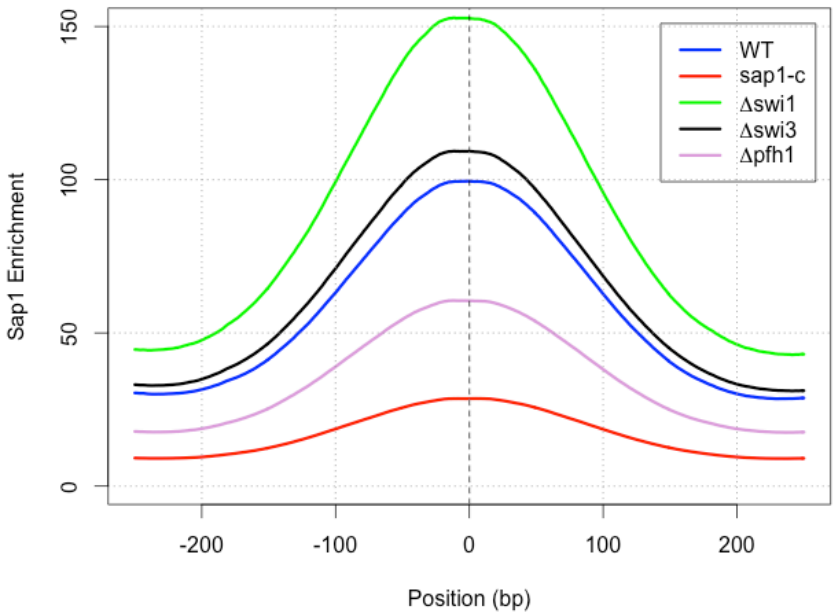


Figure 21 – Autocorrelation of Tf1 insertions in the indicated RFB mutants. Horizontal lines demonstrate two overlapping periods of 10 and 34 base pairs.

A



B

Strain	AUC-0.5 (for insertions >=1)
WT	.1966
Δswi1	**.1479**
Δswi3	.2191
Δpfh1	**.1437**
sap1-c	**.0982**

Figure 22 – Tf1 association with Sap1 changes in RFB mutants. (A) Average Sap1 enrichment around Tf1 insertions in the indicated strains; Sap1 ChIP data from (Zaratiegui et al., 2010). (B) ROC analysis of Tf1 association with Sap1 as done in (Jacobs et al., 2015); strains demonstrating a lower predictive value are bolded.

4. Discussion

The previously characterized link between Sap1 RFB and Tf1 insertion competence prompted us to study Tf1 transposition in mutants known to influence this barrier. We expected that FPC mutants that establish these barriers, presumably by stabilizing protein-DNA interactions (Sabouri et al., 2012), would lead to dramatic changes in Tf1 transposition frequency or targeting. Similarly, we expected that mutations in a helicase known to be critical for fork bypass of RFB to enhance Tf1 transposition at Sap1 binding sites. Instead, we discovered a complicated relationship between RFB and Tf1 transposition that is hard to summarize succinctly. For all these data, it is important to keep in mind the overall Tf1 transposition frequency in these mutants. To account for differences in the number of insertions obtains from each sample, these genome-wide data are normalized to reads per million. Thus, the effects of RFB on transposition frequencies are best understood in the context of their transposition frequencies.

Despite being normally thought of as members of heterodimeric complex that cannot function without the other, it is clear that Δ*swi1* and Δ*swi3* have different effects on Tf1 transposition. Transposition rates drop in both mutants, but are only significantly reduced in Δ*swi3* (Figure 18). Insertions in Δ*swi3* are similar to WT, but Δ*swi1* insertions are more strongly associated with stronger Sap1 binding sites. Unexpectedly, Sap1 enrichment is less predictive of Tf1 insertion in Δ*swi1*. These differences are interesting considering that ~66% of insertions in both Δ*swi1* and Δ*swi3* occur in Sap1 enriched peaks (Table 6), and insertion density within these regions is unchanged (Figure 19).

Interestingly, both Δ*pfh1* (*pfh1mt**) and *sap1-c* show similar patterns of Tf1 redistribution, but for different reasons. Both mutants show significant drops in insertion density within Sap1 enriched regions (Figure 19), and insertions associated with weaker Sap1 binding sites genome-wide (Figure 22). However, as compared to WT, the same percentage of insertions in Δ*pfh1* are associated with Sap1 enriched regions. This is in contrast to *sap1-c,* where a larger proportion of insertions occur outside of these regions (Table 6). This could mean that Pfh1 activity normally directs Tf1 transposition to a subset of Sap1 enriched regions, and depletion of Pfh1 results in redistribution of Tf1 insertions, mostly to weaker Sap1 binding sites. This may explain why Tf1 insertion density drops within more regions in Δ*pfh1* (Figure 19), and why Tf1 insertion points in Δ*pfh1* mutants are associated with weaker Sap1 binding sites on average (Figure 22).

It is important to realize that these data comparing Sap1 binding strength to Tf1 insertion profiles are performed using Sap1 ChIP enrichment values obtained from WT strains. Our own genome-wide measurements of Sap1 binding in *sap1-c* has revealed Sap1 binding to be largely the same, and a reanalysis of the *sap1-c* insertion data with Sap1 binding profile from that mutant does not explain the redistribution of insertions away from Sap1 binding sites (not shown). However, it remains to be seen what Sap1 enrichment profiles look like genome-wide in Δ*swi1*, Δ*swi3,* and Δ*pfh1*. It may be the case that the differences we see in these mutants are fully explained by differences in the genome-wide profile of Sap1 binding. While our low-throughput characterization of Sap1 binding at LTR and TER1 (Figure 17A) suggests that binding changes will likely be minor, it still warrants further investigation.

Regardless, this analysis of insertion profiles around Sap1 binding sites reveals that it is unlikely that stalled forks contribute to the asymmetry or periodicity of insertions around Sap1 binding motifs (Figure 20 and 21). We do not currently know whether these sites are affected by the FPC, Pfh1, or *sap1-c*. However, we do know that Swi1-Swi3 affect Sap1 RFB at rDNA (Krings & Bastia, 2004), and that Pfh1 depletion enhances RFB at this locus (Sabouri et al., 2012). Notwithstanding different effects on RFB establishment, insertion profiles and frequency (as a percentage of total insertions) at rDNA (TER1) and LTR are identical in all tested strains (not shown). This indicates that the presence of a paused fork signal is not predictive of Tf1 insertion profiles. Our previous analysis of RFB competence and Sap1 binding revealed that the RFB enhanced integration of Tf1 in both plasmid and genomic contexts (Jacobs et al., 2015), so this result was unexpected. This could indicate that the efficiency of Tf1 insertion is governed mostly by the orientation of the approaching fork, and not whether it is stalled or not. Clearly, other currently unknown factors are responsible for this orientation-dependent targeting. One possibility is that the binding of other factors depends on the orientation of Sap1, despite Sap1 binding itself not being influenced by orientation (Figure 11). A similar phenomenon was observed in a recent analysis of CTCF binding orientations in the human genome which revealed that inversion of the site does not change CTCF binding, but affects cohesin binding (Ya Guo et al., 2015a). Interestingly, cohesin binding has recently been shown to inhibit Ty1 transposition (Ho et al., 2015). It remains to be seen whether cohesin influences Tf1 transposition; this subject is a current area of active research in our laboratory.

Chapter IV - Efficient Retrotransposon Targeting and Periodicity Is Accomplished Through The Tf1 Chromodomain

1. Introduction

Fungal LTR-TE integration preferences are mediated by specific interactions between their integrase and host factors. These interactions are critical for the LTR-TE lifecycle, as they enable the TE to select targets away from regions of the genome critical to host fitness. However, even within the same LTR-TE family, ORF avoidance is accomplished through interaction with different host factors. Within the Metaviridae family, for example, Ty5, Ty3, and Tf1 target heterochromatin, class III genes, and class II genes, respectively (Levin & Moran, 2011). Further, within the Pseudoviridae family, Ty1 is targeted to class III genes, similar to Ty3 (Lesage & Todeschini, 2005). The specific host factor integrase interactions that guide targeting have been well characterized for Ty1 and Ty5. Both factors utilize interactions between their integrase C-terminal domains and host factors to target integration sites (Bridier-Nahmias et al., 2015; Xie et al., 2001).

Among LTR-TE, this C-terminal region is poorly conserved, suggesting that it evolves rapidly (Peterson-Burch & Voytas, 2002), and possibly plays a general role in mediating target site selection of diverse groups of LTR-TE. However, a notable exception is the cooption of a C-terminal chromodomain (CHD) within a group of LTR-TE within the Metaviridae family (Malik & Eickbush, 1999). The CHD is a ~40-50 amino acid motif that folds into three beta strands packed against a C-terminal alpha helix (Nielsen et al., 2002), and is known to interact with protein, DNA, and RNA (Brehm, Tufteland, Aasland, & Becker, 2004). Famously, CHDs are active in the recognition of methylated residues on proteins (Nielsen et al., 2002).

While studies investigating the role of LTR-TE CHD's are lacking, experiments have revealed that some LTR-TE CHD are functional and can recognize methylated lysines (X. Gao et al., 2008). Some transposons with CHD are often located within gene-poor and transposon-rich areas of the genome, which are often regions coated in repressive histone marks decorated by methylated histones (H3K9me, H3K27me), suggesting that CHD are directly involved in TE targeting (X. Gao et al., 2008).

The Gypsy family LTR-TE Tf1 from *S. pombe* also contains a C-terminal CHD. *In vitro,* recombinant Tf1 IN proteins lacking CHD (INΔCHD) show increased strand transfer and disintegration activities (Hizi & Levin, 2005), showing that CHD can modulate IN enzymatic activity. In contrast, *in vivo* studies with overexpressed Tf1-*neo*(ΔCHD) show 14 fold reduced transposition activity, reduced IN binding to cDNA ends, and a loss of plasmid targeting to class II genes (Chatterjee et al., 2009). This loss of transposition activity is likely due to a changed activity of the IN, as levels of Gag, RT, IN, and cDNA are similar between the strains (Chatterjee et al., 2009). Despite loss of plasmid targeting, high-throughput analysis of insertion points show virtually identical targeting when compared to wild-type profiles (Chatterjee et al., 2014). These studies suggest that CHD modulates transposition efficiency, not targeting, and that these changes in efficiency are the result of a decrease in IN catalytic activity.

The essential DNA binding domain Sap1 mediates Tf1 targeting to class II genes, and its C-terminal domain has been shown to directly interact with the Tf1 IN (Hickey et al., 2015; Jacobs et al., 2015). Here, we take a complementary approach and ask whether the Tf1 IN CHD interacts with the Sap1 C-terminal domain. Further, we probe the biological significance of the previously described Sap1-IN interaction.

2. Materials and Methods

Bioinformatic Analysis

Tf1 IN domains were predicted with InterPro (https://www.ebi.ac.uk/interpro). Tf1 insertion data was extracted from (Chatterjee et al., 2014) and converted into fwig files for analysis in R. ROC analysis was done as described previously (Jacobs et al., 2015).

Yeast 2-Hybrid

Yeast 2-Hybrid was done as described previously (Jacobs et al., 2015), with minor modifications for the Y2H fragment screen. For the targeted integrase fragment analysis, a mating approach was taken to assess interaction between fragments. For the Sap1 C-terminal mutagenesis fragment screen, screening was done in haploid cell yZB05.

Sap1 C-terminal fragment interaction screen

Mutagenic PCR was done in a total 100µl volume, with 1µl 1M Tris-HCl (pH=8), 2µl 2.5M KCl, 0.7µl 1M $MgCl_2$, 1µl 100mM pyrimidine dNTPs, 1µl 20mM purine dNTPs, 20µl of 10µM each primer (sap1FL-F, sap1-ISPNL-R, Table 7), 10pg of template DNA, and water to 97µl. Template DNA was a column purified Sap1 C-terminal fragment generated from PCR amplification of Sap1 from *S. pombe* genomic DNA using oligos Sap1FL-F/Sap1-ISPNL-R and Phusion DNA polymerase (New England Biolabs). Right before the PCR was started, 2µl of 1M MnCl2 was added to the reaction, and the reaction was cycled at 94°C 1', 60°C 1', 72°C 3' for 12-15 cycles. After the first annealing step, 1µl of 5U/µl Taq polymerase (GenScript) was added. PCR reaction products purified from 12 and 15 cycles were digested with AscI/NotI and ligated into pEG202 to generate

mutagenic libraries. 10 independent ligations were then transformed into chemically competent cells (Lucigen). Colonies from ~20 plates were then pooled and 1µg of purified library was used per transformation into yZB05 + pMZ436 (Tf1CHD-AD fragment) with a standard lithium acetate transformation protocol. Ten independent lithium acetate transformation were performed, and each transformation was spread onto two SD-trp-his-ura plates, for a total of 20 plates. To test for interaction, colonies were then replica plated onto SD(gal)-trp-his-ura+X-Gal (40µg/mL) and SD(gal)-trp-his-ura-leu. SD(gal) are SD plates prepared with 2% galactose and no glucose. Mutations were then split into 3 classes based on their ability to turn blue and/or grow on –leu plates. Class I mutants were leu+ and blue. Class II mutants were leu+ and white. Class III mutants were leu- and white. 24 colonies of each class were then amplified with BDseqF and BDseqR (Table 7) by colony PCR and Sanger sequenced by Macrogen USA. Colony PCR sequences that were not empty plasmid, early nonsense mutations, or rearranged plasmid were aligned with Clustal Omega, and the alignment was visualized with BOXSHADE.

Table 7 – Oligos used in this chapter

name	**5'->3' sequence**
Tf1IN-NR	GATCGCGGCCGCTTAATACTTTTTAATAATTGTCCTAGTC
Tf1IN-tra5F	GATCGGCGCGCCtCATGAAGAAGGTAAATTGATACATC
Tf1IN-tra5R	GATCGCGGCCGCTTAAGCTGGTGAATAGCGATGTA
Tf1IN-GP(Y/F)F	GATCGGCGCGCCtTTATCACCTTTAGAGTTACCTAGCT
Tf1IN-GP(Y/F)R	GATCGCGGCCGCTTATTCTGAATTATGTCGATACTTTTCT
Tf1IN-chR-F	GATCGGCGCGCCtCTCAATTACACTACCATTGATGATT
Tf1IN-FL-F	atgcGCGGCCGCTTAGATATTTAGATTATTGTTTTTAATATAATC
Tf1IN-FL-R	GGCGCGCCtacagatgattttaaaaaccaag
Sap1-FL-F	atgcGGCGCGCCtATGGAAGCTCCCAAGATGGA
Sap1-ISPNL-R	atgcGCGGCCGCTTAGCTAGAATGGAGACCGCCAC

Table 8 – Plasmids used in this chapter

Name	**Description**	**Source**	**Parent Plasmid**
pSH18-34	$LexA_{op(x6)}$-LacZ reporter	from OriGene Technologies	
pEG202	adh1-LexA-BD	from OriGene Technologies	
pJG4-5	adh1-LexA-AD	from OriGene Technologies	
pAD-Tf1IN	Full Length Tf1 IN (AD)		pJ4-5
pBD-Tf1IN	Full Length Tf1 IN (BD)		pEG202
pMZ509-516	Int Fragment AD fusions	Cloned into AscI/NotI	pJ4-5
pMZ517-524	Int Fragment BD fusions	Cloned into AscI/NotI	pEG202
pSap1-C-term-ISPNL AD	Sap1 C-terminus truncation	Cloned into AscI/NotI	pJ-4-5

Table 9 – Strains used in this chapter

Strain ID	Mat	Parent Strain	Description (or transformed plasmid)
yZB01	MATalpha	EGY48	positive control {Shi:2012kv}
yZB02	MATalpha	EGY48	negative control (empty vectors)
yZB05	MATalpha	EGY48	pSH18-34
yZB10	MATa	SB1035	pSap1-Cterm-AD
yZB17	MATalpha	EGY48	pSap1-Cterm-BD
yZB21	MATa	SB1035	pAD-Tf1INT
yZB22	MATalpha	EGY54	pBD-Tf1INT, pSH18-34
yZB36	MATalpha	EGY48	pSap1-cH2-ISPNL-BD, pSH18-34
yZB37	MATa	SB1035	pMZ509
yZB38	MATa	SB1035	pMZ510
yZB39	MATa	SB1035	pMZ511
yZB40	MATa	SB1035	pMZ512
yZB41	MATa	SB1035	pMZ513
yZB42	MATa	SB1035	pMZ514
yZB43	MATa	SB1035	pMZ515
yZB44	MATa	SB1035	pMZ516
yZB45	MATalpha	EGY48	pMZ517, pSH18-34
yZB46	MATalpha	EGY48	pMZ518, pSH18-34
yZB47	MATalpha	EGY48	pMZ519, pSH18-34
yZB48	MATalpha	EGY48	pMZ520, pSH18-34
yZB49	MATalpha	EGY48	pMZ521, pSH18-34
yZB50	MATalpha	EGY48	pMZ522, pSH18-34
yZB51	MATalpha	EGY48	pMZ523, pSH18-34

3. Results

Sap1 C-terminus interacts with the Tf1 CHD

Prior studies had described an interaction between Sap1 and the Tf1 IN (Hickey et al., 2015; Jacobs et al., 2015), which was more precisely mapped to the C-terminal coiled-coil dimerization domain of Sap1 (Jacobs et al., 2015). We asked what domains of the IN are responsible for this targeting. First, we used InterPro to split the integrase into four functional domains—the HHCC (or N) terminal domain, the Tra5 domain containing the catalytic activity, the GPY/F domain, and the chromodomain (CHD). These predicted domains were then used to delineate functional domains for Y2H. All possible fragments and truncations of these domains were fused to the LexA AD or BD, and tested for interaction with the Sap1 C-terminal domain. In both AD and BD configurations, all fragments containing the Tf1 CHD interacted with the Sap1 C-terminal fragment (Figure 23). This indicates that the IN CHD domain is responsible for the interaction between Sap1 and the Tf1 IN.

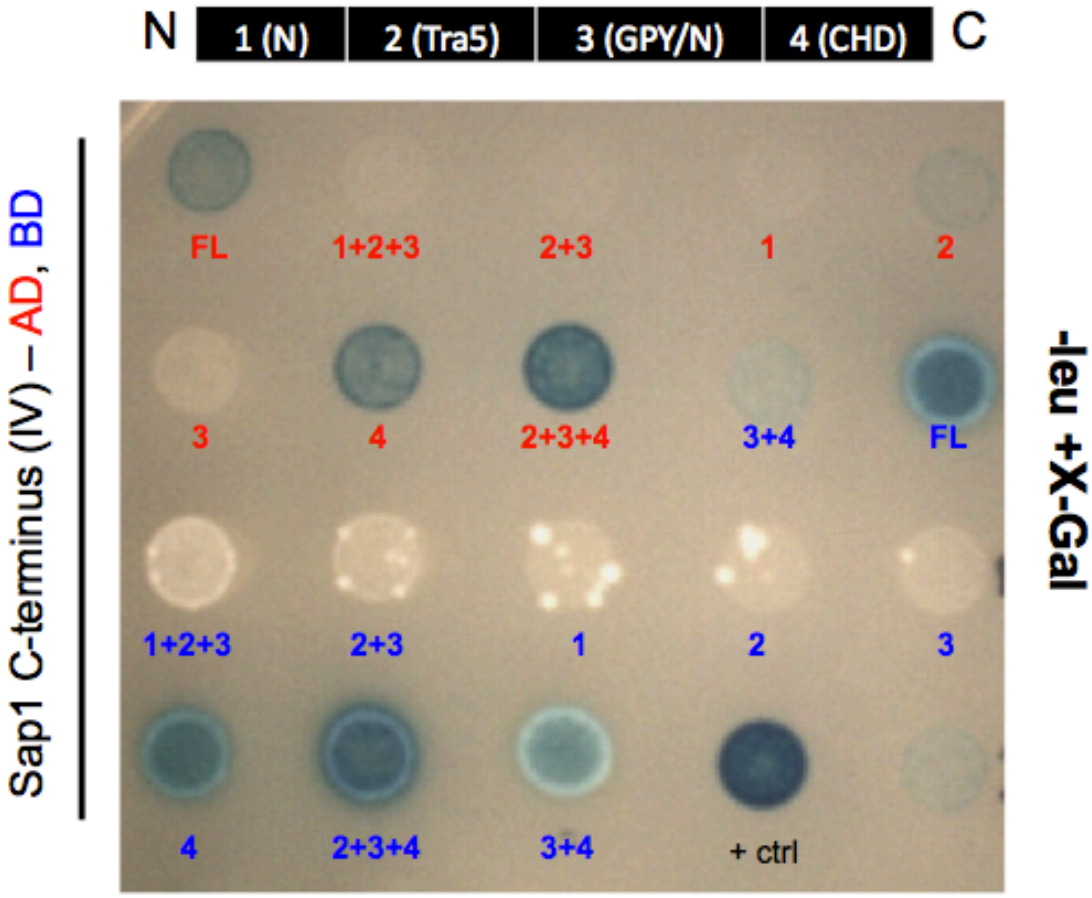


Figure 23 – The Tf1 IN CHD interacts with the Sap1 C-terminus. The relevant integrase domain fragment is indicated under the colony. Red indicates this fragment is fused to the LexA activation domain (AD), while blue indicates the fragment is fused to the LexA DNA-binding domain (BD). Growth and blue color indicate interaction. The RAB-10::CNT-1 interaction was used as a positive control (Shi et al., 2012).

Tf1 insertions in ΔCHD are still associated with Sap1 binding sites

The direct interaction between Sap1 and the IN CHD prompted us to investigate how the CHD affects Tf1 integration genome-wide. Previously generated genome-wide profiles of Tf1 insertion performed with wild-type IN and INΔCHD were reevaluated in the context of Sap1 binding (Chatterjee et al., 2014). In strains with WT IN, 53.9% of all Tf1 insertions occur within Sap1 enriched (MACS>2) regions, but drops to 42.7% in ΔCHD. Insertions are similarly associated with rDNA and LTR (not shown), indicating that the drop is explained by dispersion elsewhere in the genome. Despite this drop in overall association, Sap1 enrichment around insertion sites in WT and ΔCHD is similar (Figure 24A). Further, the ability of Sap1 binding strength to predict Tf1 insertion is only slightly lower, indicating that Tf1 is still strongly associated with Sap1 binding sites (Figure 24B). Thus, Sap1 is still strongly correlated with Tf1 insertion in ΔCHD mutants.

Another characteristic of insertions around Sap1 binding sites is their strong asymmetry and periodicity (Figure 7 and 8). We asked whether these patterns are influenced by the CHD. Interestingly, the autocorrelation of insertion positions is dramatically decreased in ΔCHD, indicating that the single-nucleotide preferences of Tf1 are changed (Figure 25). Binding patterns around Sap1 binding motifs are the largest contributor to this periodicity, so we next examined whether these patterns changed in ΔCHD. Patterns of Tf1 insertion around these Sap1 binding motifs (n=888) slightly changed between WT and ΔCHD, with some peaks appearing offset compared to WT (Figure 26). This suggests that the CHD is partially responsible for the periodicity of insertions around Sap1 binding sites.

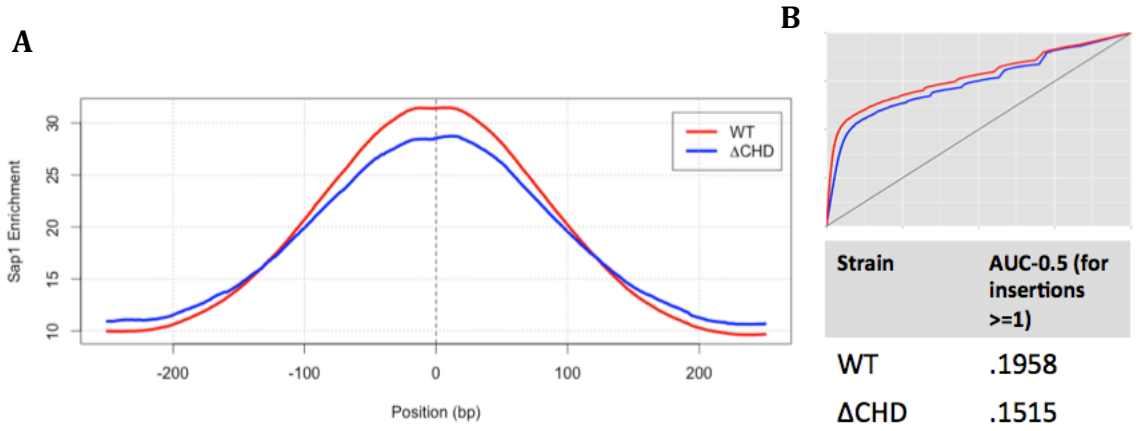


Figure 24 – Tf1 insertions profiles around Sap1 binding sites in WT and ΔCHD. (A) Averaged Sap1 enrichment around Tf1 insertions from WT and ΔCHD. (B) ROC curves from WT (red) and ΔCHD (blue), with corresponding AUC-0.5 values below. High-throughput insertion data from (Chatterjee et al., 2014) and Sap1 binding data from (Zaratiegui et al., 2010).

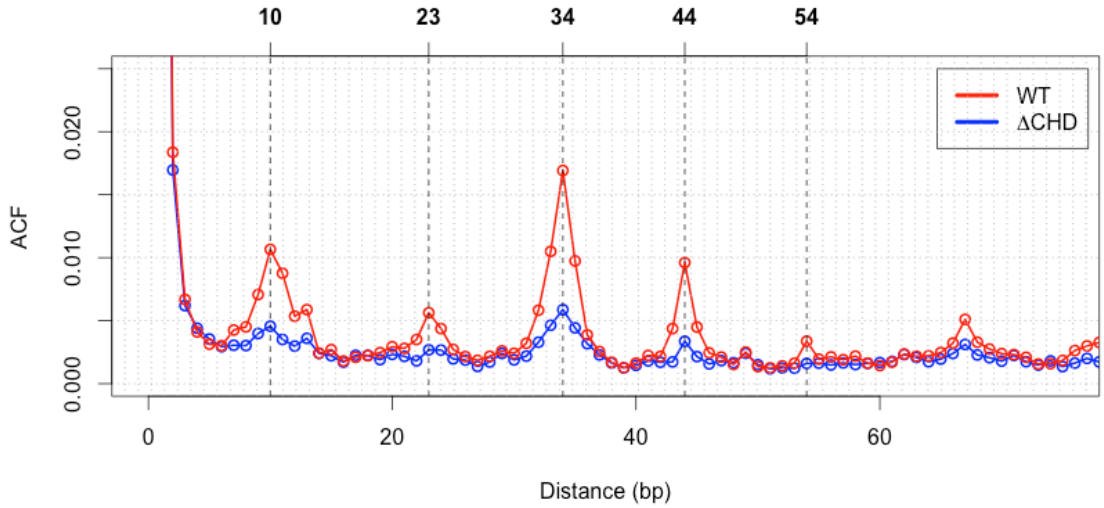


Figure 25 – Autocorrelation of Tf1 insertions in WT and ΔCHD.

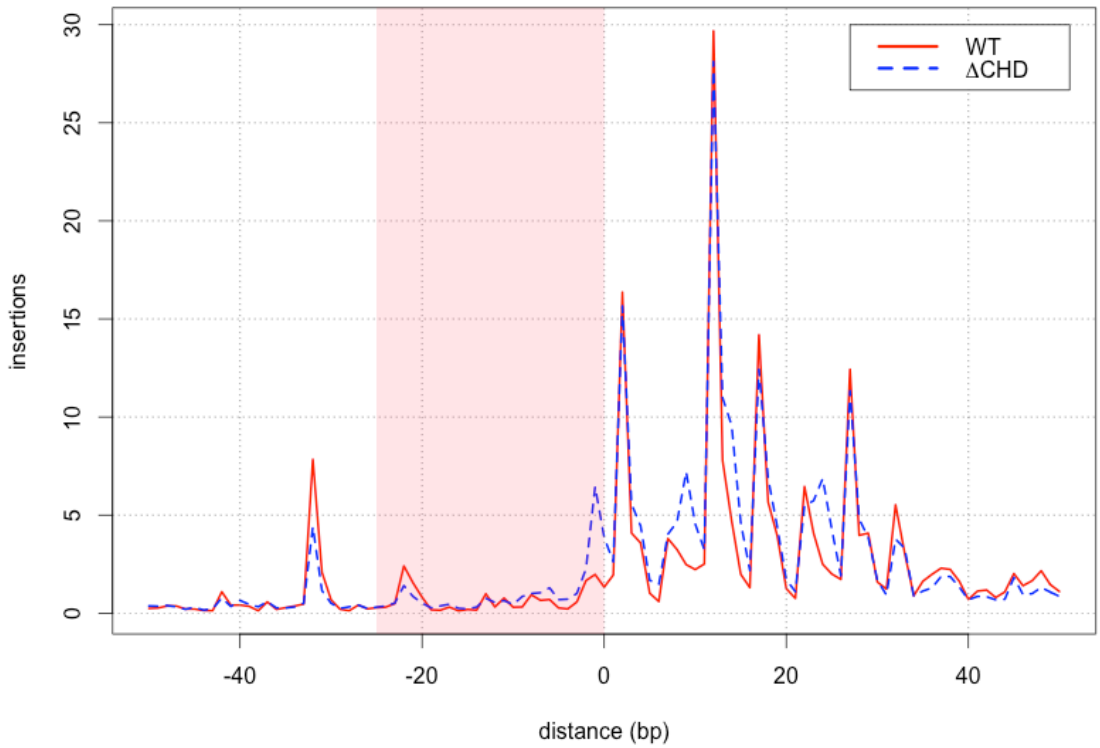


Figure 26 – Tf1 insertion profiles around Sap1 binding motifs (n=888) in WT and ΔCHD. The Sap1 binding motif location used to generate the anchor plot is indicated by the red rectangle.

Separating Sap1 binding from IN catalytic activities

We next asked whether these differences in Tf1 insertion around Sap1 binding sites are the result of a loss of Sap1 interaction or a result of changed IN enzymatic activity. To address this question, we attempted to generate mutations in the Sap1 C-terminal fragment that lose interaction with the Tf1 IN CHD but retain WT levels of DNA binding, dimerization, and fork barrier activity. The creation of such a mutation would allow for the functional separation of Tf1 IN enzymatic activity and its ability to bind Sap1 *in vivo*.

To create such mutants, we screened a Sap1 C-terminal fragment library arising from a mutagenic PCR for interaction with the Tf1 IN with Y2H. Colonies arising from this screen were separated into three classes based on their ability to activate Y2H reporters, and in turn, interact with the Tf1 IN (Figure 27A). All colonies (15/15) that lost their ability to grow on leucine were rearranged plasmids, early nonsense mutations, or empty plasmids resulting from the library preparation (not shown). Colonies that retained their ability to grow on leucine but were no longer able to turn blue were mostly colonies that contained 1-2 amino acid changes (Figure 27B, class II mutants). These mutants were then aligned against mutants that did not result in a loss of color change (Figure 27B, class I mutants), and the protein sequence of Sap1 from a sister species, *S. japonicus*, to show conservation. *S. japonicus* Sap1 was also used for alignment because it retains interaction with Tf1 IN in Y2H assays, despite having large changes in its C-terminus relative to *S. pombe* Sap1 (not shown).

This screen turned out 4 sequences that both result in a loss of blue color and contain only one amino acid change relative to WT. These mutations were then compared

to the BLOSUM62 matrix to assess their relative severity (S. Henikoff & Henikoff, 1992) (Table 10). The BLOSUM62 matrix is a position-scoring matrix that calculates the severity of amino acid substitutions on protein function by determining how underrepresented those changes are in nature, relative to what is expected. All of the recovered mutations were unfavorable mutations, but the one least likely to affect Sap1 folding and function is II-6 (M162T). Future experiments will need to be conducted to determine how these mutations affect Sap1 function, and more importantly, how they affect Tf1 transposition. Until that point, will remain unclear whether the effects the CHD has on Tf1 periodicity are a result of changed catalytic activity or a loss of interaction with Sap1.

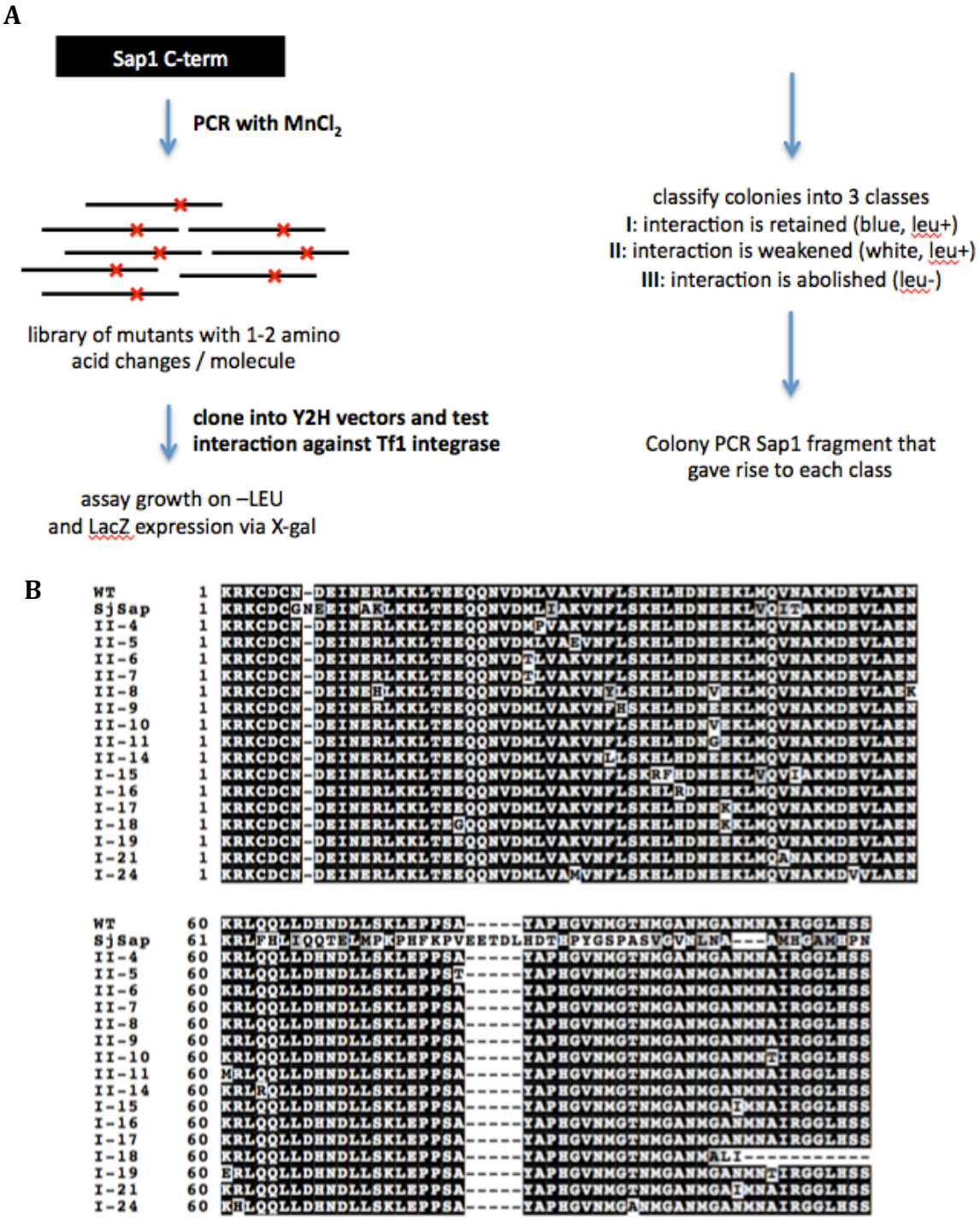


Figure 27 – Identifying point mutations that abolish interaction with the Tf1 IN CHD. (A) Schematic of the strategy, and (B) Clustal Omega alignment of colony PCR results, generated with BOXSHADE. Blue arrows indicate clones that both influence binding of the Tf1 IN CHD to Sap1, and only contain one mutation. WT= *S. pombe* Sap1. SjSap1 = *S. japonicus* Sap1.

Table 10 – BLOSUM62 penalty scores of recovered mutations

Name	**Substitution**	**BLOSUM62 Score**
II-4	L163P	-3
II-6	M162T	-1
II-9	L170H	-3
II-11	E178G	-2

4. Discussion

Little is known about the specific host factors that guide insertion of LTR-TE. The rapid evolution within the C-terminal domain of the Metaviridae, and acquisition of chromodomains, however, may be important determinants of targeting. For the Tf1 IN, the C-terminal CHD appears to be responsible for interaction with Sap1. Unexpectedly, Tf1 expressed without the CHD retains its ability to localize to Sap1 binding sites. In contrast, disruption of other IN-host factor targeting redistributes LTR-TE insertions (Bridier-Nahmias et al., 2015).

Several possibilities could explain this discrepancy. Most simply, the CHD may not be important for targeting, or may be a false positive result from Y2H. Supporting this model, genome-wide insertion profiles are largely unchanged (Chatterjee et al., 2014), and retain association with Sap1 (Figure 24). The 14-fold reduction in transposition efficiency could be explained entirely by a destabilized pre-integration complex, or altered enzymatic activity of the IN. However, *in vitro* studies show increased IN activity in CHD mutants (Hizi & Levin, 2005). IN binding to cDNA is reduced 3-fold in ΔCHD, but nuclear localization of the remaining cDNA, and overall cDNA levels, are virtually identical (Chatterjee et al., 2009). These changes seem out of proportion with the 14-fold reduction in Tf1 transposition efficiency, but it is possible that the described 3-fold reduction is an underestimate due to the crosslinking steps involved in ChIP. Altered enzymatic activity of the CHD is also another possible reason for these observations, but *in vitro* data suggests that this is unlikely to be the case (Hizi & Levin, 2005).

It is also possible that another factor targets Tf1 to Sap1 binding sites, but that the CHD-Sap1 interaction is responsible for the efficiency and periodicity of insertions at these sites. In support of this model, loss of CHD leads to slightly alerted Tf1 insertion profiles around Sap1 motifs, and the genome-wide periodicity of insertions is largely gone (Figure 25 and 26).

These two possibilities cannot currently be reconciled, but we are taking steps to separate this tethering interaction from the possible effects on the Tf1 transposition machinery (Figure 27). Namely, a targeted Y2H screen found 4 potential mutants that reduce or abolish binding to the CHD (Table 10). If CHD really is important for single-nucleotide targeting, then these mutants should result in similar Tf1 insertion profiles around Sap1 binding sites, and large reductions in transposition efficiency. However, if the interaction is a false positive Y2H result, or not important for targeting, then these mutants will likely have no impact on Tf1 transposition. Experiments are underway to introduce these mutations into *S. pombe*, and their effects on Tf1 transposition and Sap1 function will be assayed in detail.

Chapter V - Implementation of CRISPR/Cas9 in Fission Yeast

1. Introduction

The "Cas9" Type II CRISPR-Cas system for genome editing has garnered considerable attention because of its high efficiency and flexibility. Originally discovered in prokaryotes, where it participates in genome defense against invading DNA, a minimal set of components has recently been defined and put to use in the specific editing of eukaryotic genomes (Sorek, Lawrence, & Wiedenheft, 2013). This minimal system combines the RNase III-like endonuclease Cas9 and a single guide RNA (sgRNA) containing both the trans-activating RNA and CRISPR RNA (Jinek et al., 2012). These components direct the cleavage of cognate DNA sequences determined by the sequence of the sgRNA, creating a double strand break (DSB). The DSB will persist as long as the repair process that ensues is error-free resulting in the same target sequence, re-eliciting cleavage. Survivors of this cycle of cleavage and repair have acquired mutations in the target sequence that abrogates sgRNA/Cas9 recognition. The repair process can be directed to introduce specific changes by providing a DNA editing template, containing the desired mutations, that engages in Homologous Recombination (HR) with the cleaved region. Since unmutated sequences are repeatedly cleaved, leading to cell death, no additional markers are needed to select for mutation. Alternatively this system can be used to tether protein factors to specific regions of the genome by their expression as chimeric fusion proteins with cleavage-deficient Cas9 mutants (L. A. Gilbert et al., 2013), which retain their binding ability to sgRNA-programmable cognate sequences but don't cause mutagenic DNA damage. The CRISPR/Cas9 system has been adapted to virtually every model organism from baker's yeast to mammals (reviewed in (R. M. Terns & Terns, 2014)).

Expression of the sgRNA has unusual requirements because both its 5' and 3' ends need to be precisely defined in order to produce a functional Cas9/sgRNA ribonucleoprotein. This is usually achieved by means of the RNA Pol III promoter for the spliceosomal U6 snRNA that starts transcription in a defined G and terminates in a poly-T stretch. In yeast the transcribed region contains *cis*-acting sequences integral to the promoter, preventing the use of RNA Pol III promoters for expression of arbitrary sequences. This was circumvented in *Saccharomyces cerevisiae* through the use of a snoRNA promoter (DiCarlo et al., 2013). In this system, a precursor with a leader RNA containing the *cis*-acting sequences that is subsequently cleaved off at a precise location yields a mature sgRNA 5' end, with the 3' end defined by the RNA Pol III terminator.

The fission yeast *Schizosaccharomyces pombe* has proven to be a useful model organism because of its higher degree of similarity to genomes of higher eukaryotes than the classic yeast model *S. cerevisiae*. *S. pombe* remains less well studied than *S. cerevisiae*, and lags behind in availability of molecular tools. In particular, lack of a portable RNA Pol III promoter to express sgRNA has prevented the implementation of the CRISPR/Cas9 system. Here, we aim to use a RNA Pol II promoter to express gRNA with defined 5' and 3' ends, and test these gRNA in combination with Cas9 to impart single-nucleotide changes to the *S. pombe* genome.

2. Materials and Methods

Creation of constructs

All plasmids (Table 12) are available from the Addgene repository (www.addgene.org). Cas9 was built by Gibson assembly(Gibson et al., 2009) of two Cas9 fragments amplified from plasmid p414-TEF1p-Cas9-CYC1t (Addgene) (Oligos OM446/OM449, and OM447/OM448), introducing a silent mutation in the CspCI site, into plasmid pART1 digested with PstI, resulting in plasmid pMZ222. The gRNA expression construct was built by Gibson assembly of a synthetic dsDNA containing the *rrk1* promoter and leader RNA, the CspCI placeholder, and the gRNA (Figure S1), into plasmid pUR19 (OM473/OM474), yielding pMZ252. We inserted the Hammerhead Ribozyme from the satellite RNA of Tobacco Ringspot Virus (sTRSV) (OM552/OM553) immediately downstream of the gRNA by Gibson assembly with plasmid pMZ252 (PCR-amplified with OM550/OM551), yielding pMZ283. The *ade6* targeting gRNA constructs were built by digestion of pMZ283 with CspCI and ligation with double stranded oligonucleotides with the desired sequence (oligonucleotides OM450/OM456 [WT]; OM452/OM457 [M210]; OM454/OM458 [L469]) yielding pMZ284 (wt), pMZ285 (M210) and pMZ286 (L469). The combination Cas9/sgRNA plasmid pMZ374 was cloned by amplifying the adh1:Cas9 expression cassette (oligonucleotides OM554/OM555) and inserting it into the ZraI site of pMZ283 by Gibson assembly. Targeting sgRNA were cloned into the CspCI placeholder as with pMZ284. The nmt1:Cas9 construct was made by Gibson assembly of a Cas9 fragment (OM777/OM778), the nmt1 promoter (OM801/OM802) and the BamHI/SphI digested pART1 vector, resulting in vector pMZ373.

The DNA provided as an HR donor template were produced by PCR of the targeted region in the case of WT and M210 (OM12/OM13), an unrelated PCR product contained within the *reb1* gene (OM100/OM101). PCR products were purified with silica membrane columns (Epoch Life Science) and used directly for transformation.
The cdc6-Nterm targeting gRNA was made by cloning oligos OM817/817 into pMZ283. To create the HA tagging repair template, HA was amplified from pFA6-3xHA (OM821/OM822). The 5' and 3' regions flanking the *cdc6* start codon were amplified with OM820/OM823 and oM825/oM824 respectively. These 3 fragments were then gel purified and joined by cross-over PCR with (OM820/OM825).

Transformation and mutagenesis experiments

All strains used contained the *ura4-D18* and *leu1-32* alleles, and one of 4 *ade6* alleles (WT, M210, L469 or M216). Strains were grown in 5 ml of MB media (Sunrise Science Products #2016) with the appropriate supplements at 32°C to mid-log phase (OD=0.5) and transformed by the Lithium Acetate/PEG/Heat shock method (Bähler et al., 1998). Transformation of pMZ222 was selected on Dropout media without Leucine. All mutagenesis experiments were carried out by cotransformation of 1 μg of the sgRNA (for Cas9 expressing strains) or Cas9/sgRNA construct (for single-step mutagenesis), and 1mg of the PCR product used as HR template. 5mL of mid-log phase cultures were grown in MB media with supplements, as above, transformed as above, spread onto EMM (US Biological #E2205) plates supplemented with 10 mg/L adenine immediately after heat shock and grown at 32°C for 4 days. For colony PCR individual colonies were picked, restreaked to single colonies, boiled in 0.02M NaOH for 10 minutes, and the *ade6*

target region was amplified and sequenced.

5' RACE

WT strains expressing Cas9 via vector pMZ222, transformed with a sgRNA directed against the *ade6-M210* allele were picked form single colonies and grown in EMM media to mid-log phase and harvested and total RNA was extracted with the hot phenol method. 5µg of total RNA was used with the 5´ RACE System Version 2.0 kit (Life technologies) using OM546 for RT and OM547 for nested PCR, cloned by TopoTA and sequenced.

Northern blot

WT strains expressing Cas9 via vector pMZ222, either without sgRNA vector or transformed with a sgRNA directed against the *ade6-M210* allele (to avoid any selection effects from the targeting and cleavage of the endogenous allele) were picked from single colonies, grown in EMM media to mid-log phase, harvested and total RNA was extracted with the hot phenol method. 5µg of total RNA was run alongside a radioactively labeled Decade Marker (Life Technologies) on a 8% PAGE with 7M Urea and blotted onto charged nylon membrane by semi-dry transfer. The blot was hybridized to labeled oligonucleotides OM599 (U1) or OM 546 (sgRNA).

HA-tag characterization

An *ura4-D18, leu1-32* strain was mutagenized as described above, offering the HA-tagged cdc6 N-terminus fragment as HR donor template, and plated onto media selective for the Cas9/sgRNA vector. Five colonies were picked at random and colony PCR

performed as above. The same colonies were grown in EMM with the appropriate supplements to mid-log phase, pelleted and boiled in Laemmli loading buffer, run in an 8% SDS-PAGE gel and transferred by semi-dry transfer onto 0.2 mm pore Nitrocellulose. The membrane was blocked and incubated with anti-HA antibody (Abgent AP1012a) or with anti-Sap1 antibody (Jacobs et al., 2015), then developed with secondary anti-rabbit-HRP conjugate and ECL (GE).

Table 11 - Oligonucleotides used in this chapter

OM12	CAAAGATCCTGTCGAATCACCTG
OM13	CAGTTATGTCTATGGTCGCCTATGC
OM100	CTGCGGAACATTGGGACTAT
OM101	TTCTTTGCTCCACACACAGC
OM446	TTAAGCAAGAGAATTGCTGCAGGTCGACTCTAGAGATGGACAAGA AGTACTCCATTG
OM447	GGAAGGTTGGAATTCGAGCTCGGTACCCGGGGATCGCAAATTAAA GCCTTCGAGCGTC
OM448	TGGGATCCGAGAtAAGCAGtcTGGAAAGACAATCCT
OM449	AGGATTGTCTTTCCAgaCTGCTTaTCTCGGATCCCA
OM450	TCTATTGTTCAGATGCCTCGgt
OM452	TCTATTGTTCAGATGCtTCGgt
OM454	TCTATTGTTCAGATGCCTtGgt
OM456	CGAGGCATCTGAACAATAGAtt
OM457	CGAaGCATCTGAACAATAGAtt
OM458	CaAGGCATCTGAACAATAGAtt
OM473	cgagtcggtggtgcTTTTTTTCggtaccgtggggatcctctagagtcgac
OM474	ACTACCACCAACATAAGCAAAAggtacccggggtaccgagctcgaattc
OM550	gaaagcacatccggtgacagggcaccaccgactcggtgccactc
OM551	gagtccgtgaggacgaaacaggggtaccgtgggGATCCTCTAGAG
OM552	cctgtcaccggatgtgctttccggtctgatgagtccgtgaggacgaaacagg
OM553	cctgtttcgtcctcacggactcatcagaccggaaagcacatccggtgacagg
OM546	ACTCGGTGCCACTCACTTTT
OM547	ACGGACTAGCCTTATTTTAACTT
OM599	AGCTGACCTTAGCCAGTCCA
OM554	cccgaaaagtgccacctgacGCCCTACAACAACTAAGAAAATG
OM555	gataataatggtttcttagacGGGGATCGCAAATTAAAGCC
OM777	tgcaggtcgactctagagatggaca
OM778	attgtctttccagactgcttatctc
OM801	tgtccatctctagagtcgacctgcagaGgatatgccaggattcctcttcc
OM802	aacgacgtagtcgacaagcttGCATGCcgccataaaagacagaataagtc
OM817	TAATTACTAGAATGACAGATgt
OM818	ATCTGTCATTCTAGTAATTAtt
OM820	TCAGCAGTTGACACTCATACCA
OM821	AAATAAAAAATTAATTACTAGAATGtacccatacgatgttcct
OM822	CCTCATTTGAAGACCTATCTGTCATgcactgagcagcgtaatc
OM823	TCTAGTAATTAATTTTTTATTT
OM824	ATGACAGATAGGTCTTCAAATGAGG
OM825	TTGCCACTGCTGATCTTTTG

Table 12 - Plasmids used in this chapter

	expression cassette	**sgRNA Target**	**Marker /replication origin**	**backbone**	**Addgene number**
pMZ222	adh1:Cas9	N/A	*LEU2/ars1*	pART1[20]	52223
pMZ283	rrk1:sgRNA	empty	*ura4/ars1*	pUR19[21]	52224
pMZ284	rrk1:sgRNA	*ade6+*	*ura4/ars1*	pUR19	52225
pMZ285	rrk1:sgRNA	*ade6-M210*	*ura4/ars1*	pUR19	52226
pMZ286	rrk1:sgRNA	*ade6-L469*	*ura4/ars1*	pUR19	52227
pMZ374	adh1:Cas9/rrk1:sgRNA	empty	*ura4/ars1*	pUR19	59896
pMZ288	adh1:Cas9/rrk1:sgRNA	*ade6+*	*ura4/ars1*	pUR19	59897
pMZ289	adh1:Cas9/rrk1:sgRNA	*ade6-M210*	*ura4/ars1*	pUR19	59898
pMZ381	adh1:Cas9/rrk1:sgRNA	*ade6-L469*	*ura4/ars1*	pUR19	59899

3. Results

Creation and validation of a novel RNA Pol II driven sgRNA expression platform

We sought to establish an sgRNA expression system in *S pombe*. Since, similarly to those of *S cerevisiae*, *S pombe* RNA Pol III promoters include promoter elements in the transcribed region we explored the possibility of using a transcript with a cleavable leader RNA. One such example is the *rrk1* gene that codes for K RNA, a component of the RNAse P ribonucleoprotein (Krupp, Cherayil, Frendewey, Nishikawa, & Söll, 1986). While it has been reported that *S pombe rrk1* lacks a leader RNA, careful inspection of available high-throughput RNA sequencing data reveals that in mutants of *rrp6*, a component of the exosome that degrades by-products of RNA maturation, *rrk1* is preceded by a leader RNA of around 250 nucleotides(B. T. Wilhelm et al., 2008).

We constructed an expression cassette by joining positions -1 to -358 of the *rrk1* gene with the sgRNA sequence (DiCarlo et al., 2013; Mali, Esvelt, & Church, 2013), preceded by a CspCI restriction target site that serves as a placeholder for cloning of the targeting sequence (Figure 28). Since *rrk1* is synthesized by RNA Pol II (Krupp et al., 1986) which has complex transcription terminators and yields polyadenylated transcripts, we cloned a Hammerhead Ribozyme (Dower, Kuperwasser, Merrikh, & Rosbash, 2004; Y. Gao & Zhao, 2014) immediately downstream of the construct to precisely determine the 3' end of the mature sgRNA (Figure 28 and 29A). Northern blotting and 5' RACE analysis revealed the accumulation of sgRNA with correct cleavage of the 5' and 3' ends, as well as the presence of a larger precursor (Figure 29B,C).

5’
TTTTGCTTATGTTGGTGGTAGTTGGCATGCGTAGACTGATGACTAGTCAGCAA
GGAGCGTAGAACAGTCACACTCGTTATATATGTGCTTCCAAGAAAACTCAAG
AATTTACCATTAGCAAACACTTTTTTGAAATGTTAGACATTTAAATGACGAAG
GCATATAGAAGCTTTGAATAGGTGTTGTAAAGTGTTGATTTATGTGACGCTGA
GGGTGCGCATGAAAGGAATGTTGGGTCACGATTATTAAACAGTTTGCTAGCT
TGGACACTTGAGTATTGGAAGTTGTTGAATTCTAAAAAACTTTCAGTTGATTT
GAATAGTTGCTGTTGCCAAAAAACATAACCTGTACCGAAGAAtgggcttaactCAAtt
cttGTGGgttatctctctgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaaagtgagtggc
accgagtcggtggtgcCCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGG
ACGAAACAGG
3’

Figure 28 – DNA sequence of the gRNA cassette. *rrk1* promoter and leader RNA: Capital letters; CspCI placeholder: straight underlined; sgRNA: small caps, not underlined; Hammerhead Ribozyme: wavy underlined capital letters.

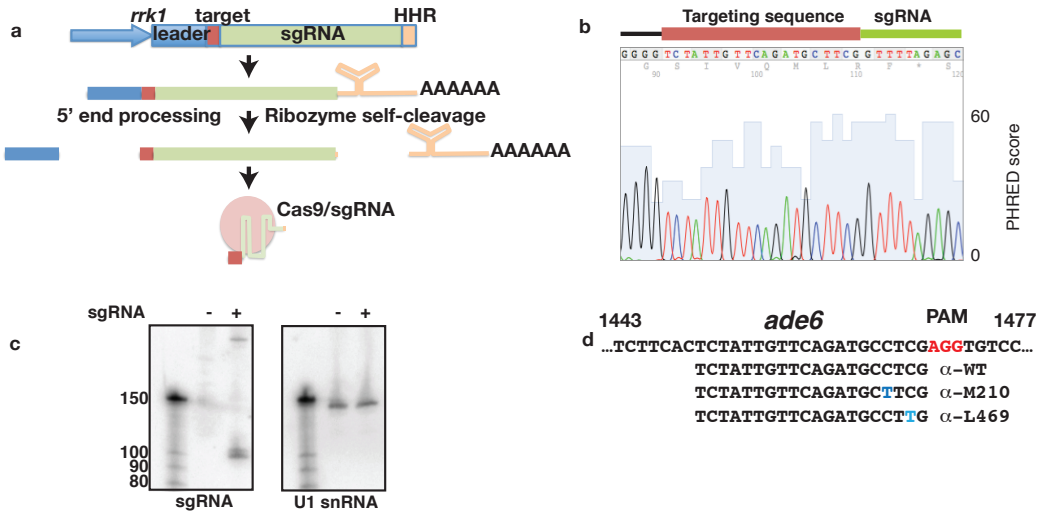


Figure 29 - sgRNA expression system. (A) Schema of sgRNA expression construct and processing. (B) 5' RACE sequence. Only 4bp of the oligodG attached to the 5' end, resulting from the method, are shown. (C) Northern analysis of Cas9 expressing , ade6+ strains with either no sgRNA vector or a sgRNA vector targeted against ade6-M210. Marker sizes are shown to the left and hybridization probe below. (D) *ade6* targeting sequences.

Initial testing of sgRNA:Cas9 to potentiate genome modification

As a proof of principle of the Cas9/CRISPR mutagenesis system, we attempted the editing of the *ade6* gene, which when mutated causes accumulation of a red colored precursor in media with low adenine. We generated sgRNA constructs targeting three different alleles of *ade6*: *ade6+* (*WT*), *ade6-M210* and *ade6-L469*. The *M210* and *L469* mutations are located within a 3 bp window near a Protospacer Adjacent Motif (PAM) necessary for Cas9/sgRNA recognition (Figure 29C), and are easily characterized because they yield deep red colonies and cause an XhoI restriction site polymorphism. As a DNA template for directed HR we generated PCR products containing the *WT* and *M210* alleles, as well as a non-homologous amplicon of the same size, called Control. We transformed strains containing WT, *ade6-M210*, *ade6-L469* or *ade6-M216* (an allele mutated outside of the targeted region yielding light pink colonies) alleles with an *adh1*:Cas9 constitutive expression construct and rechecked the *ade6* phenotype of the transformants to ensure that Cas9 induces no changes without the sgRNA. We then picked individual transformants from each strain, transformed them with all combinations of one sgRNA expression construct and one mutation donor DNA, and spread them onto Edimburgh Minimal Media supplemented with a low concentration of adenine. We scored the phenotype of the colonies generated, and picked individual colonies for sequencing of the *ade6* gene.

Transformation of sgRNA targeting alleles different from the one present in the transformed strain yielded no changes in phenotype, indicating that the single nucleotide polymorphisms were sufficient to confer complete specificity to the mutagenesis. When transformed with a sgRNA targeting the allele present in the strain, the Cas9 expressing

transformants exhibited a wide variation of mutagenesis efficiencies. Some of them (8 out of 12) changed color (white to red, red to white and pink to red) with efficiencies between 50% and 98%, with the highest efficiencies corresponding to those co-transformed with an HR template donor that changes the targeted sequence (85-98%). Targeting of the *WT* sequence present in the *WT* and *M216* strains (yielding white and pink colonies respectively) resulted in the generation of dark red colonies, indicating that the targeted region acquired new mutations that phenocopied the *M210* and *L469* mutations. Conversely, targeting of the *ade6-M210* and *ade6-L469* alleles resulted in the appearance of white colonies indicating reversion to the *WT* allele. However, a portion of the Cas9 expressing transformants (4 out of 12) were resistant to CRISPR mediated mutagenesis, with mutation efficiencies between 0 and 5% even when the allele present was targeted by the sgRNA.

We sequenced the *ade6* gene from unmutated and mutated colonies (according to their color) in experiments from mutagenesis competent transformants (Figure 30). The sequencing results show that when the allele present in the strain is targeted by the sgRNA, mutagenesis occurs through two different mechanisms depending on the HR donor template co-transformed. If the donor template is homologous to the targeted region and it provides a mutation that would confer resistance to cleavage by the sgRNA/Cas9, almost all the mutated colonies had acquired the template mutation (22/23), indicating that repair of the cleavage had occurred through HR with the donor template. However, when the donor template encodes the same allele that is being targeted in the strain or is non-homologous to the targeted region (Control), the resulting mutations were small deletions and insertions at the cleavage point, typical of repair

through Non-Homologous End Joining (NHEJ). Colonies that retained the original phenotype also showed the original genotype in their sequence, indicating that they had somehow become resistant to Cas9/sgRNA cleavage without mutating the targeted sequence. If the sgRNA transformed targeted a different *ade6* allele from that present in the strain, no mutations were detected, indicating that the single nucleotide polymorphisms were sufficient to confer complete specificity to the mutagenesis.

The transformation efficiencies observed in the mutagenesis-competent clones differed depending on the sgRNA used: they were very high (10^4 cfu/μg) if it targeted an allele different from the one present in the strain, but dropped precipitously (between 10 and 10^2 cfu/μg) when the sgRNA targeted the allele present in the strain. This suggests that the combination of Cas9 with a targeting sgRNA challenges the survival of the transformant. In contrast, Cas9 expressing transformants resistant to CRIPSR mutagenesis showed consistently high transformation efficiencies, independently of the sgRNA used, indicating that cleavage of the target was impaired in these clones.

We sought to investigate the source of the drastic differences in CRISPR competency of the different Cas9-expressing clones. We hypothesized that mutations or rearrangements in the Cas9 and/or sgRNA expressing plasmids could render them inactive. To test this hypothesis we recovered the transformed plasmids from CRISPR competent and incompetent clones, and characterized them by restriction digest and sequencing. Both sgRNA and Cas9 expressing plasmids were intact in CRISPR competent clones (not shown). sgRNA plasmids from CRISPR incompetent clones were also intact. In contrast, Cas9 expressing plasmids from CRISPR incompetent clones exhibited rearrangements and mutations that would render them inactive (Figure 31).

The relatively high frequency of these events suggested that Cas9 inactivation underwent positive selection, implying that Cas9 overexpression is toxic to fission yeast, even without a sgRNA. Stochastic mutations of Cas9 would overcome the chosen transformant resulting in sporadic loss of CRISPR competency. We tested this hypothesis generating a construct that expresses Cas9 driven by the inducible nmt1 promoter, which is active in the absence of thiamine. The size of the colonies transformed with this construct depended on the induction conditions: plating onto media without thiamine (nmt1 promoter active) resulted in colonies much smaller than those plated in media with thiamine (nmt1 repressed) indicating impaired growth upon Cas9 expression (Figure 32). Accordingly, clones obtained from the nmt1-repressed plate showed a drastically reduced growth rate when grown in liquid media without thiamine (Figure 32). We conclude that Cas9 overexpression is toxic to fission yeast, and that maintenance of Cas9 expressing clones carries the risk of incapacitating mutations that impede CRISPR genome editing.

original genotype	Template	Phenotype		target PAM
				CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC WT
				CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC M210
				CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC L469
WT	+	White	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
		Red	1442>	CTCTTCACTCTATTGTTCAGATGCCCCTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-C-CGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC--TCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC--TCGAGGTGTCCCTGTCGCCACTGTTGC
	210	White	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
		Red	1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
	Control	White	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
		Red	1442>	CTCTTCACTCTATTGTTCAGATGCCCCTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC--TCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
M216	+	Pink	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
		Red	1442>	CTCTTCACTCTATTGTTCAGATGC-C-C-AGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCA------CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGCCCCTCGAGGTGTCCCTGTCGCCACTGTTGC
	M210	Pink	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
		Red	1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
	Control	Pink	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
		Red	1442>	CTCTTCACTCTATTGTTCAGAT---TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATG---TCGAGGTGTCCCTGTCGCCACTGTTGC

Figure 30 (1/2) – Sequencing results of the targeted *ade6* region in survivor colonies.

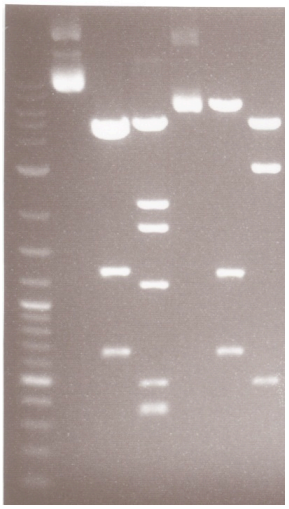
M210	+	Red	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
		White	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
	M210	Red	1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGCTTTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC--------TGTCCCTGTCGCCACTGTTGC
		White	1442>	CTCTTCACTCTATTGTTCAGATGCTTTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC--TCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGA-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATG---TCGAGGTGTCCCTGTCGCCACTGTTGC
	Control	Red	1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGAT----TCGAGGTGTCCCTGTCGCCACTGTTGC
		White	1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGCTTTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
L469	+	White	1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTCGAGGTGTCCCTGTCGCCACTGTTGC
	M210	Red	1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-TTCGAGGTGTCCCTGTCGCCACTGTTGC
	Control	Red	1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGAT---CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC
			1442>	CTCTTCACTCTATTGTTCAGATGC-CTTGAGGTGTCCCTGTCGCCACTGTTGC

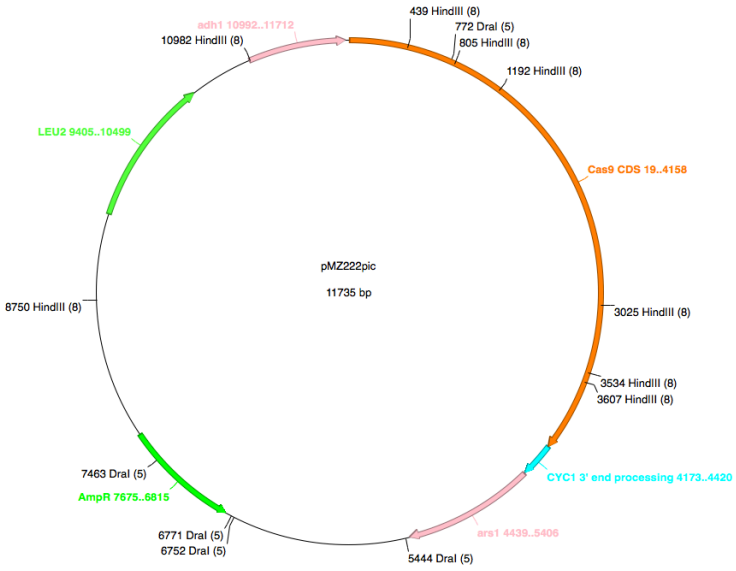
Figure 30 (2/2) - Sequencing results of the targeted *ade6* region in survivor colonies.

```
                R  K  N  R  I  C  Y  L  Q  E  I  F  S  N  E  M  A  K  V    93
Cas9+           AGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTG   278
Cas9mut1        AGAAAGAATCGG--CTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTG
                R  K  N  R     L  L  P  A  G  D  L  *

                E  E  D   K  K  H  E  R  H  P  I  F  G      120
Cas9+           GAGGAGTCC-TTTTTGGTGGAGGAGGATAAAAAGCACGAG     369
Cas9mut2        GAGGAGTCCATTTTTGGTGGAGGAGGATAAAAAGCACGAG
                E  E  D  I  F  G  G  G  G  *
```

1 2 3 4 5 6 7





1 - DNA ladder (NEB #N3200)
2-4: pMZ222
- 2: undigested
- 3: DraI
- 4: HindIII

5-7: pCas9mut3
- 5: undigested
- 6: DraI
- 7: HindIII

Figure 31 - Mutations and rearrangements detected in Cas9 expression vectors isolated from CRISPR incompetent clones.

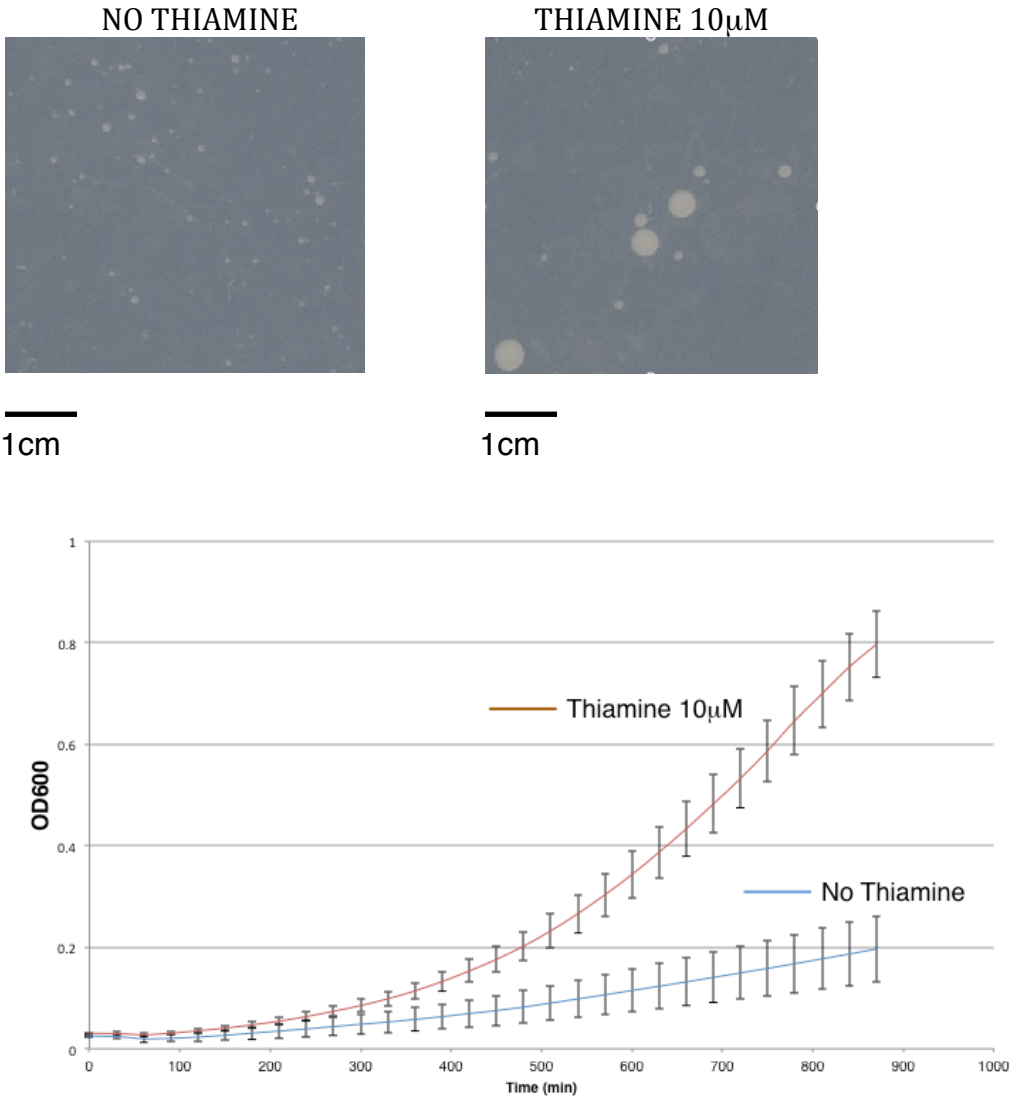


Figure 32 - Cas9 toxicity. Upper panel: colony size after transformation with nmt1:Cas9 construct and plating in EMM without thiamine (nmt1 promoter active) or with 10μM Thiamine (nmt1 promoter repressed). Lower panel: growth curves of clones picked from the repressed plate, grown in EMM with or without 10μM Thiamine (3 clones each). Error bars depict standard deviation.

Overcoming Cas9 toxicity with single-vector CRISPR/Cas9 modification

In order to prevent the selection of Cas9 mutations in the mutagenesis protocol, we evaluated the efficiency of mutagenesis induced by a combined vector expressing both Cas9 and the sgRNA. This approach has the added advantage of requiring a single transformation instead of two. We cloned the adh1 promoter driven Cas9 expression cassette into the sgRNA expression plasmid, and then cloned the *ade6+*, *ade6-M210* and *ade6-L469* targeting sequences into the sgRNA cassette. We then transformed them into the same *ade6+*, *ade6-M216*, *ade6-M210* and *ade6-L469* strains used in the two-plasmid system, along with the same *ade6+*, *ade6-M210* and Control HR donor templates, in all combinations. The results are summarized in Figure 33. Transformation of the single-plasmid vector with a sgRNA targeting the allele of *ade6* present in the strain caused a high frequency of mutation with robust reproducibility. Recapitulating the results observed in the CRISPR competent Cas9 expressing clones, the highest efficiencies (85-90%) were obtained by co-transformation with an HR template that mutates the targeting sequence.

genotype	gRNA	pcrDNA	mutagenesis efficiency	std.dev.	n
WT	α-WT	WT	69.70%	3.84%	3
		M210	88.85%	3.78%	3
		Control	59.95%	4.21%	3
	α-M210	WT	1.34%	N/A	1
		M210	0%	N/A	1
		Control	2.22%	N/A	1
	α-L469	WT	0.82%	N/A	1
		M210	0.47%	N/A	1
		Control	0%	N/A	1
M216	α-WT	WT	79.70%	1.95%	3
		M210	91.13%	3.96%	3
		Control	65.40%	4.76%	3
	α-M210	WT	4.35%	N/A	1
		M210	3.61%	N/A	1
		Control	4.52%	N/A	1
	α-L469	WT	0.43%	N/A	1
		M210	0.59%	N/A	1
		Control	1.20%	N/A	1
M210	α-WT	WT	0%	N/A	1
		M210	0%	N/A	1
		Control	0%	N/A	1
	α-M210	WT	86.21%	6.38%	3
		M210	6.80%	3.96%	3
		Control	0%	0%	3
	α-L469	WT	0.24%	N/A	1
		M210	0%	N/A	1
		Control	0%	N/A	1
L469	α-WT	WT	1.84%	N/A	1
		M210	1.58%	N/A	1
		Control	1.50%	N/A	1
	α-M210	WT	1.82%	N/A	1
		M210	1.23%	N/A	1
		Control	1.64%	N/A	1
	α-L469	WT	87.62%	4.82%	3
		M210	1.80%	4.44%	3
		Control	4.55%	2.30%	3

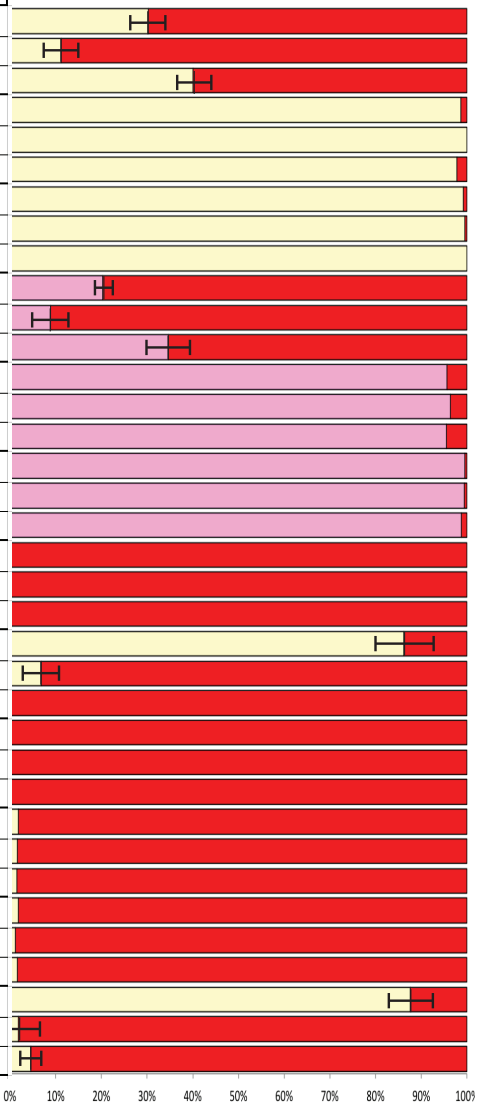
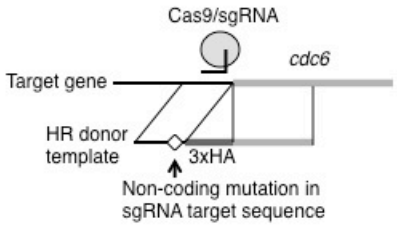


Figure 33 - *ade6* mutagenesis. The original genotype is shown in the first column, the sgRNA expression vector in the second, the PCR product used as HR template in the third, number of independent transformations in the fourth. The phenotype of the survivor colonies is shown in stacked bar format, with error bars depicting standard deviation.

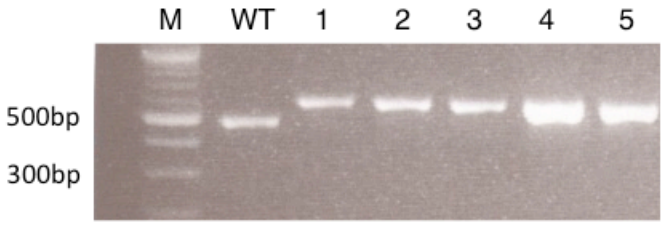
N-terminal HA-tagging of an essential gene with CRISPR/Cas9

We then attempted a real-life application of CRISPR mutagenesis by inserting an N-terminal epitope tag into *cdc6,* the catalytic subunit of DNA polymerase Delta. N-terminal tagging by classic means involves a multistep process with antibiotic markers and the generation of large transforming fragments. We created an sgRNA targeting the N-terminus of *cdc6*, and generated a PCR product containing a 3xHA tag in frame with the N-terminus of *cdc6*, destroying the targeted site, with 200bp homology regions on both sides of the tag (Figure 34). We co-transformed these constructs into a WT strain, and picked 5 clones at random for characterization of *cdc6*. All 5 clones showed a the expected size increase by PCR, and western blotting confirmed that a protein of size corresponding to that of Cdc6 tagged with 3xHA was expressed.

A



B



C

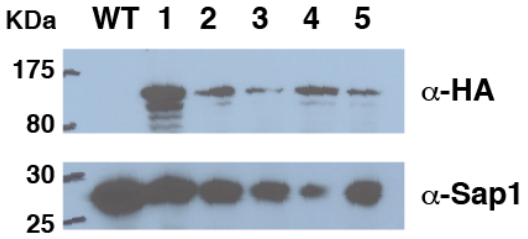


Figure 34 - CRISPR/Cas9 mediated tagging of the N-terminus of Cdc6. (A) Tagging strategy. (B) PCR of the original strain (WT) and 5 clones (1-5) transformed with the Cas9/sgRNA construct and tag HR donor template. (C) Western blot of the same strains as in B, with anti-HA antibody or anti-Sap1 antibody (as loading control).

4. Discussion

Schizosaccharomyces pombe is the model organism of choice for studies of chromosome biology because of its conservation of crucial aspects of chromosome function, such as complex heterochromatic centromeres, with higher organisms. However, modification of the *S. pombe* genome is often laborious and time consuming, and requires the use of marker genes that may interfere with the phenomenon being investigated. *Delitto perfetto* approaches that subsequently remove the marker genes require multiple selection steps and screening of large number of candidate colonies to obtain the desired modification (Storici et al., 2001). Furthermore, owing to the lower efficiency of HR with exogenous DNA observed in *S pombe* as compared to *S cerevisiae*, extended homology regions are required, necessitating additional cloning steps(Y. Gao & Zhao, 2014; Steinhauser et al., 2012). Since targeted cleavage of the unmutated sequence constitutes a built-in negative selection, CRISPR/Cas9 mutagenesis achieves near complete efficiencies and obviates the need for selectable markers. The single vector expressing Cas9 and the sgRNA is marked with *ura4*, allowing for plasmid removal by counterselection with Fluoroorotic Acid and enabling subsequent mutagenesis of additional targets.

We expect that the flexibility of the *rrk1* sgRNA expression system will allow for implementation of the full Cas9 toolset, from sgRNA/Cas9 systems with alternate PAMs (Esvelt et al., 2013) to nicking and catalytic dead mutants of Cas9 (L. A. Gilbert et al., 2013; L. S. Qi et al., 2013). The RNA Pol II-expressed *rrk1*/Hammerhead Ribozyme cassette may also prove useful in other situations where expression of RNA of defined arbitrary length and sequence such as short interfering RNA or long non-coding RNA is

needed, an advantage over RNA Pol III systems. The methods and reagents presented here will prove useful for genomics research in *S pombe* by enabling rapid and specific genome editing.

Chapter VI – Epilogue

The work described here strongly suggests that Sap1 is the tethering factor for Tf1. Short DNA oligos were competent at Tf1 recruitment (Figure 10), and genome-wide studies reveal that 63% of Tf1 insertions lie within Sap1 enriched areas of the genome (Hickey et al., 2015; Jacobs et al., 2015). Insertions around Sap1 sites displayed asymmetry and periodicity, with the majority of insertions occurring on the 3' end of the binding motif (Figure 7). Helping to explain this correlation, the Sap1 C-terminal domain directly interacts with the Tf1 IN chromodomain (Figure 23).

It's a familiar tale of transposon targeting, but with a new twist—we also found Sap1 replication fork barrier activity to be a requirement. Flipping short oligonucleotides containing minimal Sap1 binding sites changed Tf1 insertion profiles and efficiency, but not binding of Sap1 to these sites (Figure 10). Consistently, Tf1 insertions were correlated with DNA polymerase epsilon maxima and marks of DNA damage (Figure 9).

This correlation between Tf1 transposition and fork barriers turned our attention to mutants that globally modulate these activities. Unexpectedly, we found that deletion of the genetic requirements for barrier activity or bypass had mild effects on Tf1 transposition. In both FPC (Swi1-Swi3) and Pfh1 mutants, transposition rates were virtually identical (Figure 18), insertion points were still enriched in Sap1 regions (Table 6), and insertion patterns around Sap1 binding sites still displayed asymmetry and periodicity (Figure 20 and 21). Measurements of RFB activity are performed by directly measuring the accumulation of fork-shaped intermediates by 2D gel electrophoresis in actively growing cells. The lack of changes in the Tf1 insertion profile around RFB in these mutants, despite large changes in these RFB signals, indicates that fork stalling is

unlikely to be the secondary targeting requirement. In FPC mutants, there is no detectable barrier at rDNA, but Tf1 insertion profiles at this locus are indistinguishable from wild-type, and occur just as frequently. However, we also know that in FPC mutants, even FPC-independent barriers become highly recombinogenic (Pryce et al., 2009). It is possible that these observations point to the existence of additional features of protein-DNA barriers that cannot be seen by 2D gel electrophoresis. The localization of these features would, of course, have to dependent on the orientation of Sap1.

The only mutants that dramatically influence Tf1 transposition efficiency are mutations in the targeting factor Sap1—the *sap1-c* and *sap1-1* alleles, which both reduce transposition 10 fold. The *sap1-c* allele displays virtually identical genome-wide binding (Figure 17), but Tf1 insertion points have reduced association with Sap1 binding sites (63% in WT, 49% in *sap1-c*). On the other hand, the *sap1-1* allele has an identical Tf1 insertion profile. We suspect that the differences between *sap1-c* and *sap1-1* are due to their different effects on the Sap1 fork barrier. Only *sap1-c* has been shown to affect RFB in the genome. Interestingly, *sap1-c* maps to the DNA binding domain, while *sap1-1* (L181S) maps to the dimerization domain, only three residues away from one of our interaction mutations (II-11, E178G). It might be that *sap1-c* loses insertion targeting and efficiency because it loses RFB activity, while *sap1-1* only loses insertion efficiency because it loses interaction with IN but does not affect RFB activity.

It is unclear why these mutants have such severe effects on Tf1 transposition. It is possible that these alleles only affect the two Sap1 binding sites within the LTR. Sap1 binds to DNA as a dimer, and these dimers are capable of tetramerization *in vitro* (Ghazvini et al., 1995). Thus, Sap1 tetramerization between cDNA and Sap1 bound in the

genome may be what guides insertion. Unfortunately, our attempts at mutating the Sap1 binding sites within the LTR completely abolished cDNA synthesis (not shown), so we were unable to experimentally determine the significance of these sites. Future work will aim to abolish Sap1 binding sites at the LTR without affecting cDNA synthesis.

One important consequence of the involvement of the replication fork in transposon targeting is the prediction that integration occurs during S phase. In bacteria, transposons and integrative bacteriophage have been directly shown to target the template strands during DNA replication (Fricker & Peters, 2014). In Eukaryotes, the evidence for retrotransposition occurring during S-phase is more circumstantial. The DNA transposon Sleeping Beauty has been shown to slow down cell cycle progression by direct interaction with cyclins (Walisko et al., 2006), and in human cells, HIV enhances retrotransposition of the non-LTR transposon L1 in a manner dependent on its ability to cause cell cycle arrest (Jones et al., 2013). In yeast, factors linked to Okazaki fragment processing and stalled fork repair have been linked to LTR-TE hypermobility (Lesage & Todeschini, 2005). These studies suggest, but do not directly show, that transposition may occur in S-phase.

Our own attempts at addressing this question have come up empty handed, mainly due to the lack of tools available to limit the ability of Tf1 to integrate to one phase of the cell cycle. We do know that transposition does not occur during G_0, a metabolically active cell cycle phase where no DNA replication occurs, but unfortunately, neither cDNA, Sap1, or Tf1 IN are detectable during this phase of the cell cycle. We are currently developing degron-based tools to limit IN activity to one phase of the cell cycle to address this question in cycling cells.

Why might LTR-TE favor insertion into stalled forks or during S-phase cell cycle? One possibility is that it allows TE to colonize new genomes without necessitating the evolution of specific host factor-IN interactions. A transposon with the capability to target arrested replication forks would be capable of enhanced horizontal transfer, because the structure of the replication fork is universal. DNA replication could also be a period of vulnerability. S-phase could provide proteins or substrates for integration, or increase access to cellular DNA by unmasking targeting factors, or recruit DNA repair factors needed for insertion. Alternatively, integration could occur outside of S-phase, if fork arrest leaves epigenetic marks that guide insertion. Although the link between S-phase and TE is not air tight, the correlation is strong enough to warrant further study.

References

Agrawal, A., Eastman, Q. M., & Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*, *394*(6695), 744–751. http://doi.org/10.1038/29457

Aguilera, A., & García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. *Molecular Cell*, *46*(2), 115–124. http://doi.org/10.1016/j.molcel.2012.04.009

Arcangioli, B., & Klar, A. J. (1991). A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient mat1 switching in Schizosaccharomyces pombe. *The EMBO Journal*, *10*(10), 3025–3032.

Arcangioli, B., Ghazvini, M., & Ribes, V. (1994). Identification of the DNA-binding domains of the switch-activating-protein Sap1 from S.pombe by random point mutations screening in E.coli. *Nucleic Acids Res*, *22*(15), 2930–2937.

Atwood, A., Choi, J., & Levin, H. L. (1998). The application of a homologous recombination assay revealed amino acid residues in an LTR-retrotransposon that were critical for integration. *Journal of Virology*, *72*(2), 1324–1333.

Atwood, A., Lin, J. H., & Levin, H. L. (1996). The retrotransposon Tf1 assembles virus-like particles that contain excess Gag relative to integrase because of a regulated degradation process. *Molecular and Cellular Biology*, *16*(1), 338–346.

Azvolinsky, A., Giresi, P. G., Lieb, J. D., & Zakian, V. A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular Cell*, *34*(6), 722–734. http://doi.org/10.1016/j.molcel.2009.05.022

Bada, M., Walther, D., Arcangioli, B., Doniach, S., & Delarue, M. (2000). Solution structural studies and low-resolution model of the Schizosaccharomyces pombe sap1 protein. *Journal of Molecular Biology*, *300*(3), 563–574. http://doi.org/10.1006/jmbi.2000.3854

Bairwa, N. K., Mohanty, B. K., Stamenova, R., Curcio, M. J., & Bastia, D. (2011). The intra-S phase checkpoint protein Tof1 collaborates with the helicase Rrm3 and the F-box protein Dia2 to maintain genome stability in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *286*(4), 2445–2454. http://doi.org/10.1074/jbc.M110.189456

Baller, J. A., Gao, J., Stamenova, R., Curcio, M. J., & Voytas, D. F. (2012). A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. *Genome Research*, *22*(4), 704–713. http://doi.org/10.1101/gr.129585.111

Barahona, A. (1997). Barbara McClintock and the transposition concept. *Archives internationales d'histoire des sciences* (Vol. 46, pp. 309–329).

Barlow, J. H., Faryabi, R. B., Callén, E., Wong, N., Malhowski, A., Chen, H. T., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell*, *152*(3), 620–632. http://doi.org/10.1016/j.cell.2013.01.006

Bähler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., Steever, A. B., et al. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast*, *14*(10), 943–951. http://doi.org/10.1002/(SICI)1097-0061(199807)14:10<943::AID-

YEA292>3.0.CO;2-Y
Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., et al. (2010). LINE-1 retrotransposition activity in human genomes. *Cell*, *141*(7), 1159–1170. http://doi.org/10.1016/j.cell.2010.05.021
Behrens, R., Hayles, J., & Nurse, P. (2000). Fission yeast retrotransposon Tf1 integration is targeted to 5' ends of open reading frames. *Nucleic Acids Research*, *28*(23), 4709–4716.
Bell, S. P., & Dutta, A. (2002). DNA replication in eukaryotic cells. *Annual Review of Biochemistry*, *71*, 333–374. http://doi.org/10.1146/annurev.biochem.71.110601.135425
Bellen, H. J., Levis, R. W., He, Y., Carlson, J. W., Evans-Holm, M., Bae, E., et al. (2011). The Drosophila gene disruption project: progress using transposons with distinctive site specificities. *Genetics*, *188*(3), 731–743. http://doi.org/10.1534/genetics.111.126995
Benzer, S. (1956). Genetic fine structure and its relation to the DNA molecule. *Brookhaven Symposia in Biology*, (8), 3–5.
Biémont, C. (2010). A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics*, *186*(4), 1085–1093. http://doi.org/10.1534/genetics.110.124180
Bochman, M. L., Sabouri, N., & Zakian, V. A. (2010). Unwinding the functions of the Pif1 family helicases. *DNA Repair*, *9*(3), 237–249. http://doi.org/10.1016/j.dnarep.2010.01.008
Bowen, N. J. (2003). Retrotransposons and Their Recognition of pol II Promoters: A Comprehensive Survey of the Transposable Elements From the Complete Genome Sequence of Schizosaccharomyces pombe. *Genome Research*, *13*(9), 1984–1997. http://doi.org/10.1101/gr.1191603
Brehm, A., Tufteland, K. R., Aasland, R., & Becker, P. B. (2004). The many colours of chromodomains. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *26*(2), 133–140. http://doi.org/10.1002/bies.10392
Bridier-Nahmias, A., Tchalikian-Cosson, A., Baller, J. A., Menouni, R., Fayol, H., Flores, A., et al. (2015). Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposon integration. *Science*, *348*(6234), 585–588. http://doi.org/10.1126/science.1259114
Britten, R. J., & Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science*, *165*(3891), 349–357.
Burns, K. H., & Boeke, J. D. (2012). Human transposon tectonics. *Cell*, *149*(4), 740–752. http://doi.org/10.1016/j.cell.2012.04.019
Bushman, F. D. (2003). Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell*, *115*(2), 135–138.
Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., & Labib, K. (2005). Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes & Development*, *19*(16), 1905–1919. http://doi.org/10.1101/gad.337205
Cam, H. P., Noma, K.-I., Ebina, H., Levin, H. L., & Grewal, S. I. S. (2007). Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature*, *451*(7177), 431–436. http://doi.org/10.1038/nature06499
Canman, C. E. (2001). Replication checkpoint: preventing mitotic catastrophe. *Current*

Biology : CB, *11*(4), R121–4.
Capy, P., Gasperi, G., Biémont, C., & Bazin, C. (2000). Stress and transposable elements: co-evolution or useful parasites? *Heredity*, *85 (Pt 2)*, 101–106.
Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, *371*(6494), 215–220. http://doi.org/10.1038/371215a0
Chatterjee, A. G., Esnault, C., Guo, Y., Hung, S., McQueen, P. G., & Levin, H. L. (2014). Serial number tagging reveals a prominent sequence preference of retrotransposon integration. *Nucleic Acids Research*, *42*(13), 8449–8460. http://doi.org/10.1093/nar/gku534
Chatterjee, A. G., Leem, Y.-E., Kelly, F. D., & Levin, H. L. (2009). The chromodomain of Tf1 integrase promotes binding to cDNA and mediates target site selection. *Journal of Virology*, *83*(6), 2675–2685. http://doi.org/10.1128/JVI.01588-08
Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, *186*(2), 757–761. http://doi.org/10.1534/genetics.110.120717
Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., et al. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nature Medicine*, *11*(12), 1287–1289. http://doi.org/10.1038/nm1329
Clark, J. B., & Kidwell, M. G. (1997). A phylogenetic perspective on P transposable element evolution in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(21), 11428–11433.
Compton, A. A., Malik, H. S., & Emerman, M. (2013). Host gene evolution traces the evolutionary history of ancient primate lentiviruses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1626), 20120496. http://doi.org/10.1098/rstb.2012.0496
Dalgaard, J. Z., & Klar, A. J. (2000). swi1 and swi3 perform imprinting, pausing, and termination of DNA replication in S. pombe. *Cell*, *102*(6), 745–751.
Dardalhon, M., de Massy, B., Nicolas, A., & Averbeck, D. (1998). Mitotic recombination and localized DNA double-strand breaks are induced after 8-methoxypsoralen and UVA irradiation in Saccharomyces cerevisiae. *Current Genetics*, *34*(1), 30–42.
de Lahondes, R., Ribes, V., & Arcangioli, B. (2003). Fission Yeast Sap1 Protein Is Essential for Chromosome Stability. *Eukaryotic Cell*, *2*(5), 910–921. http://doi.org/10.1128/EC.2.5.910-921.2003
Deshpande, A. M., & Newlon, C. S. (1996). DNA replication fork pause sites dependent on transcription. *Science*, *272*(5264), 1030–1033.
Devine, S. E., & Boeke, J. D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes & Development*, *10*(5), 620–633. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=8598291&retmode=ref&cmd=prlinks
DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research*, *41*(7), 4336–4343. http://doi.org/10.1093/nar/gkt135
Dolgin, E. S., & Charlesworth, B. (2006). The fate of transposable elements in asexual populations. *Genetics*, *174*(2), 817–827. http://doi.org/10.1534/genetics.106.060434

Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y., & Moran, J. V. (2015). A 3' Poly(A) Tract Is Required for LINE-1 Retrotransposition. *Molecular Cell*, *60*(5), 728–741. http://doi.org/10.1016/j.molcel.2015.10.012

Dower, K., Kuperwasser, N., Merrikh, H., & Rosbash, M. (2004). A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. *RNA (New York, N.Y.)*, *10*(12), 1888–1899. http://doi.org/10.1261/rna.7166704

Dubarry, M., Loïodice, I., Chen, C. L., Thermes, C., & Taddei, A. (2011). Tight protein-DNA interactions favor gene silencing. *Genes & Development*, *25*(13), 1365–1370. http://doi.org/10.1101/gad.611011

Errico, A., Cosentino, C., Rivera, T., Losada, A., Schwob, E., Hunt, T., & Costanzo, V. (2009). Tipin/Tim1/And1 protein complex promotes Pol alpha chromatin binding and sister chromatid cohesion. *The EMBO Journal*, *28*(23), 3681–3692. http://doi.org/10.1038/emboj.2009.304

Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., & Church, G. M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods*, *10*(11), 1116–1121. http://doi.org/10.1038/nmeth.2681

Eydmann, T., Sommariva, E., Inagawa, T., Mian, S., Klar, A. J. S., & Dalgaard, J. Z. (2008). Rtf1-mediated eukaryotic site-specific replication termination. *Genetics*, *180*(1), 27–39. http://doi.org/10.1534/genetics.108.089243

Fang, W., Wang, X., Bracht, J. R., Nowacki, M., & Landweber, L. F. (2012). Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. *Cell*, *151*(6), 1243–1255. http://doi.org/10.1016/j.cell.2012.10.045

Feng, G., Leem, Y. E., & Levin, H. L. (2012). Transposon integration enhances expression of stress response genes. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gks1185

Fennessy, D., Grallert, A., Krapp, A., Cokoja, A., Bridge, A. J., Petersen, J., et al. (2014). Extending the Schizosaccharomyces pombe molecular genetic toolbox. *PLoS ONE*, *9*(5), e97683. http://doi.org/10.1371/journal.pone.0097683

Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, *41*, 331–368. http://doi.org/10.1146/annurev.genet.40.110405.090448

Fricker, A. D., & Peters, J. E. (2014). Vulnerabilities on the lagging-strand template: opportunities for mobile elements. *Annual Review of Genetics*, *48*, 167–186. http://doi.org/10.1146/annurev-genet-120213-092046

Gai, X., & Voytas, D. F. (1998). A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. *Molecular Cell*, *1*(7), 1051–1055.

Gaj, T., Mercer, A. C., Sirk, S. J., Smith, H. L., & Barbas, C. F. (2013). A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Research*, *41*(6), 3937–3946. http://doi.org/10.1093/nar/gkt071

Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J., & Craig, N. L. (2010). DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. *Proceedings of the National Academy of Sciences*, *107*(51), 21966–21972. http://doi.org/10.1073/pnas.1016382107

Gao, X., Havecker, E. R., Baranov, P. V., Atkins, J. F., & Voytas, D. F. (2003). Translational recoding signals between gag and pol in diverse LTR retrotransposons.

RNA (New York, N.Y.), *9*(12), 1422–1430.
Gao, X., Hou, Y., Ebina, H., Levin, H. L., & Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Research*, *18*(3), 359–369. http://doi.org/10.1101/gr.7146408
Gao, Y., & Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of Integrative Plant Biology*, *56*(4), 343–349. http://doi.org/10.1111/jipb.12152
Ghazvini, M., Ribes, V., & Arcangioli, B. (1995). The essential DNA-binding protein sap1 of Schizosaccharomyces pombe contains two independent oligomerization interfaces that dictate the relative orientation of the DNA-binding domain. *Molecular and Cellular Biology*, *15*(9), 4939–4946.
Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. http://doi.org/10.1038/nmeth.1318
Gilbert, C., Schaack, S., Pace, J. K., Brindley, P. J., & Feschotte, C. (2010). A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature*, *464*(7293), 1347–1350. http://doi.org/10.1038/nature08939
Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*(2), 442–451. http://doi.org/10.1016/j.cell.2013.06.044
Grallert, B., Nurse, P., & Patterson, T. E. (1993). A study of integrative transformation in Schizosaccharomyces pombe. *Molecular & General Genetics : MGG*, *238*(1-2), 26–32.
Grimm, C., Kohli, J., Murray, J., & Maundrell, K. (1988). Genetic engineering of Schizosaccharomyces pombe: a system for gene disruption and replacement using the ura4 gene as a selectable marker. *Molecular & General Genetics : MGG*, *215*(1), 81–86.
Guo, Ya, Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., et al. (2015a). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900–910. http://doi.org/10.1016/j.cell.2015.07.038
Guo, Yabin, & Levin, H. L. (2010). High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in Schizosaccharomyces pombe. *Genome Research*, *20*(2), 239–248. http://doi.org/10.1101/gr.099648.109
Guo, Yabin, Singh, P. K., & Levin, H. L. (2015b). A long terminal repeat retrotransposon of Schizosaccharomyces japonicus integrates upstream of RNA pol III transcribed genes. *Mobile DNA*, *6*, 19. http://doi.org/10.1186/s13100-015-0048-2
Hancks, D. C., & Kazazian, H. H. (2016). Roles for retrotransposon insertions in human disease. *Mobile DNA*, *7*, 9. http://doi.org/10.1186/s13100-016-0065-9
Hare, S., Gupta, S. S., Valkov, E., Engelman, A., & Cherepanov, P. (2010). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*, *464*(7286), 232–236. http://doi.org/10.1038/nature08784
Hartmuth, S., & Petersen, J. (2009). Fission yeast Tor1 functions as part of TORC1 to control mitotic entry through the stress MAPK pathway following nutrient stress. *Journal of Cell Science*, *122*(Pt 11), 1737–1746. http://doi.org/10.1242/jcs.049387
Havecker, E. R., Gao, X., & Voytas, D. F. (2004). The diversity of LTR retrotransposons. *Genome Biology*, *5*(6), 225. http://doi.org/10.1186/gb-2004-5-6-225

Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(22), 10915–10919.

Hentges, P., Van Driessche, B., Tafforeau, L., Vandenhaute, J., & Carr, A. M. (2005). Three novel antibiotic marker cassettes for gene disruption and marker switching in Schizosaccharomyces pombe. *Yeast*, *22*(13), 1013–1019. http://doi.org/10.1002/yea.1291

Hickey, A., Esnault, C., Majumdar, A., Chatterjee, A. G., Iben, J. R., McQueen, P. G., et al. (2015). Single-Nucleotide-Specific Targeting of the Tf1 Retrotransposon Promoted by the DNA-Binding Protein Sap1 of Schizosaccharomyces pombe. *Genetics*, *201*(3), 905–924. http://doi.org/10.1534/genetics.115.181602

Hizi, A., & Levin, H. L. (2005). The integrase of the long terminal repeat-retrotransposon tf1 has a chromodomain that modulates integrase activities. *The Journal of Biological Chemistry*, *280*(47), 39086–39094. http://doi.org/10.1074/jbc.M506363200

Ho, K. L., Ma, L., Cheung, S., Manhas, S., Fang, N., Wang, K., et al. (2015). A role for the budding yeast separase, Esp1, in Ty1 element retrotransposition. *PLoS Genetics*, *11*(3), e1005109. http://doi.org/10.1371/journal.pgen.1005109

Hoff, E. F., Levin, H. L., & Boeke, J. D. (1998). Schizosaccharomyces pombeRetrotransposon Tf2 Mobilizes Primarily through Homologous cDNA Recombination. *Molecular and Cellular Biology*, *18*(11), 6839–6852. http://doi.org/10.1128/MCB.18.11.6839

Hoffman, C. S., Wood, V., & Fantes, P. A. (2015). An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. *Genetics*, *201*(2), 403–423. http://doi.org/10.1534/genetics.115.181503

Holkers, M., Maggio, I., Liu, J., Janssen, J. M., Miselli, F., Mussolino, C., et al. (2013). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Research*, *41*(5), e63. http://doi.org/10.1093/nar/gks1446

Horiuchi, T., Fujimura, Y., Nishitani, H., Kobayashi, T., & Hidaka, M. (1994). The DNA replication fork blocked at the Ter site may be an entrance for the RecBCD enzyme into duplex DNA. *Journal of Bacteriology*, *176*(15), 4656–4663.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832. http://doi.org/10.1038/nbt.2647

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., & Tsai, S. Q. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature*.

Iraqui, I., Chekkal, Y., Jmari, N., Pietrobon, V., Fréon, K., Costes, A., & Lambert, S. A. E. (2012). Recovery of arrested replication forks by homologous recombination is error-prone. *PLoS Genetics*, *8*(10), e1002976. http://doi.org/10.1371/journal.pgen.1002976

Isaac, R. S., Jiang, F., Doudna, J. A., Lim, W. A., Narlikar, G. J., & Almeida, R. A. (2016). Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife*, *5*. http://doi.org/10.7554/eLife.13450

Jacobs, J. Z., Ciccaglione, K. M., Tournier, V., & Zaratiegui, M. (2014). Implementation of the CRISPR-Cas9 system in fission yeast. *Nature Communications*, *5*, 5344. http://doi.org/10.1038/ncomms6344

Jacobs, J. Z., Rosado-Lugo, J. D., Cranz-Mileva, S., Ciccaglione, K. M., Tournier, V., & Zaratiegui, M. (2015). Arrested replication forks guide retrotransposon integration. *Science*, *349*(6255), 1549–1553. http://doi.org/10.1126/science.aaa3810

Jao, L.-E., Wente, S. R., & Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences*, *110*(34), 13904–13909. http://doi.org/10.1073/pnas.1308335110

Jasin, M. (1996). Genetic manipulation of genomes with rare-cutting endonucleases. *Trends in Genetics*.

Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Převorovský, M., Mourier, T., et al. (2015). The genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nature Genetics*, *47*(3), 235–241. http://doi.org/10.1038/ng.3215

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., & Doudna, J. A. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*.

Jones, R. B., Song, H., Xu, Y., Garrison, K. E., Buzdin, A. A., Anwar, N., et al. (2013). LINE-1 retrotransposable element DNA accumulates in HIV-1-infected cells. *Journal of Virology*, *87*(24), 13307–13320. http://doi.org/10.1128/JVI.02257-13

Jordan, I. K., Rogozin, I. B., Glazko, G. V., & Koonin, E. V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics : TIG*, *19*(2), 68–72.

Katou, Y., Kanoh, Y., Bando, M., Noguchi, H., Tanaka, H., Ashikari, T., et al. (2003). S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature*, *424*(6952), 1078–1083. http://doi.org/10.1038/nature01900

Khurana, J. S., Wang, J., Xu, J., Koppetsch, B. S., Thomson, T. C., Nowosielska, A., et al. (2011). Adaptation to P element transposon invasion in Drosophila melanogaster. *Cell*, *147*(7), 1551–1563. http://doi.org/10.1016/j.cell.2011.11.042

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(3), 1156–1160.

Kirchner, J., Connolly, C. M., & Sandmeyer, S. B. (1995). Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science*, *267*(5203), 1488–1491.

Krings, G. (2005). Sap1p Binds to Ter1 at the Ribosomal DNA of Schizosaccharomyces pombe and Causes Polar Replication Fork Arrest. *Journal of Biological Chemistry*, *280*(47), 39135–39142. http://doi.org/10.1074/jbc.M508996200

Krings, G., & Bastia, D. (2004). swi1- and swi3-dependent and independent replication fork arrest at the ribosomal DNA of Schizosaccharomyces pombe. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(39), 14085–14090. http://doi.org/10.1073/pnas.0406037101

Krings, G., & Bastia, D. (2006). Molecular architecture of a eukaryotic DNA replication terminus-terminator protein complex. *Molecular and Cellular Biology*, *26*(21), 8061–8074. http://doi.org/10.1128/MCB.01102-06

Krupp, G., Cherayil, B., Frendewey, D., Nishikawa, S., & Söll, D. (1986). Two RNA species co-purify with RNase P from the fission yeast Schizosaccharomyces pombe. *The EMBO Journal*, *5*(7), 1697–1703.

Kuduvalli, P. N., Rao, J. E., & Craig, N. L. (2001). Target DNA structure plays a critical role in Tn7 transposition. *The EMBO Journal*, *20*(4), 924–932. http://doi.org/10.1093/emboj/20.4.924

Lambert, S., Watson, A., Sheedy, D. M., Martin, B., & Carr, A. M. (2005). Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell*, *121*(5), 689–702. http://doi.org/10.1016/j.cell.2005.03.022

Lauermann, V., & Boeke, J. D. (1997). Plus-strand strong-stop DNA transfer in yeast Ty retrotransposons. *The EMBO Journal*, *16*(21), 6603–6612. http://doi.org/10.1093/emboj/16.21.6603

Le Rouzic, A., Boutin, T. S., & Capy, P. (2007). Long-term evolution of transposable elements. *Proceedings of the National Academy of Sciences*, *104*(49), 19375–19380. http://doi.org/10.1073/pnas.0705238104

Leem, Y.-E., Ripmaster, T. L., Kelly, F. D., Ebina, H., Heincelman, M. E., Zhang, K., et al. (2008). Retrotransposon Tf1 Is Targeted to Pol II Promoters by Transcription Activators. *Molecular Cell*, *30*(1), 98–107. http://doi.org/10.1016/j.molcel.2008.02.016

Lesage, P., & Todeschini, A. L. (2005). Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenetic and Genome Research*, *110*(1-4), 70–90. http://doi.org/10.1159/000084940

Levin, H. L. (1995). A novel mechanism of self-primed reverse transcription defines a new family of retroelements. *Molecular and Cellular Biology*, *15*(6), 3310–3317.

Levin, H. L. (1996). An unusual mechanism of self-primed reverse transcription requires the RNase H domain of reverse transcriptase to cleave an RNA duplex. *Molecular and Cellular Biology*, *16*(10), 5645–5654.

Levin, H. L., & Boeke, J. D. (1992). Demonstration of retrotransposition of the Tf1 element in fission yeast. *The EMBO Journal*, *11*(3), 1145–1153.

Levin, H. L., & Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, *12*(9), 615–627. http://doi.org/10.1038/nrg3030

Levin, H. L., Weaver, D. C., & Boeke, J. D. (1990). Two related families of retrotransposons from Schizosaccharomyces pombe. *Molecular and Cellular Biology*, *10*(12), 6791–6798. http://doi.org/10.1128/MCB.10.12.6791

Levin, H. L., Weaver, D. C., & Boeke, J. D. (1993). Novel gene expression mechanism in a fission yeast retroelement: Tf1 proteins are derived from a single primary translation product. *The EMBO Journal*, *12*(12), 4885–4895.

Lewinski, M. K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathogens*, *2*(6), e60. http://doi.org/10.1371/journal.ppat.0020060

Li, J. F., Norville, J. E., Aach, J., McCormack, M., & Zhang, D. (2013). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature*.

Liang, S. C., Hartwig, B., Perera, P., Mora-García, S., de Leau, E., Thornton, H., et al. (2015). Kicking against the PRCs - A Domesticated Transposase Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory Component of Polycomb Repressive Complex 2. *PLoS Genetics*, *11*(12), e1005660.

http://doi.org/10.1371/journal.pgen.1005660
Luria, S. E., & Delbrück, M. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*, *28*(6), 491–511.
Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., et al. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular Cell*, *31*(2), 294–301. http://doi.org/10.1016/j.molcel.2008.06.016
Majumdar, A., Chatterjee, A. G., Ripmaster, T. L., & Levin, H. L. (2011). Determinants that specify the integration pattern of retrotransposon Tf1 in the fbp1 promoter of Schizosaccharomyces pombe. *Journal of Virology*, *85*(1), 519–529. http://doi.org/10.1128/JVI.01719-10
Mali, P., Esvelt, K. M., & Church, G. M. (2013). Cas9 as a versatile tool for engineering biology. *Nature Methods*, *10*(10), 957–963. http://doi.org/10.1038/nmeth.2649
Malik, H. S., & Eickbush, T. H. (1999). Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *Journal of Virology*, *73*(6), 5186–5190.
Malkova, A., & Haber, J. E. (2012). Mutations arising during repair of chromosome breaks. *Annual Review of Genetics*, *46*, 455–473. http://doi.org/10.1146/annurev-genet-110711-155547
McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, *226*(4676), 792–801.
McFarlane, R. J., Mian, S., & Dalgaard, J. Z. (2010). The many facets of the Tim-Tipin protein families' roles in chromosome biology. *Cell Cycle*, *9*(4), 700–705. http://doi.org/10.4161/cc.9.4.10676
Mejia-Ramirez, E., Sanchez-Gorostiaga, A., Krimer, D. B., Schvartzman, J. B., & Hernandez, P. (2005). The Mating Type Switch-Activating Protein Sap1 Is Required for Replication Fork Arrest at the rRNA Genes of Fission Yeast. *Molecular and Cellular Biology*, *25*(19), 8755–8761. http://doi.org/10.1128/MCB.25.19.8755-8761.2005
Miller, D. W., & Miller, L. K. (1982). A virus mutant with an insertion of a copia-like transposable element. *Nature*, *299*(5883), 562–564.
Mitchell, R. S., Beitzel, B. F., Schroder, A. R. W., Shinn, P., Chen, H., Berry, C. C., et al. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biology*, *2*(8), E234. http://doi.org/10.1371/journal.pbio.0020234
Mizuno, K., Miyabe, I., Schalbetter, S. A., Carr, A. M., & Murray, J. M. (2013). Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature*, *493*(7431), 246–249. http://doi.org/10.1038/nature11676
Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, *326*(5959), 1501. http://doi.org/10.1126/science.1178817
Mourier, T., & Willerslev, E. (2010). Large-scale transcriptome data reveals transcriptional activity of fission yeast LTR retrotransposons. *BMC Genomics*, *11*, 167. http://doi.org/10.1186/1471-2164-11-167
Mularoni, L., Zhou, Y., Bowen, T., Gangadharan, S., Wheelan, S. J., & Boeke, J. D. (2012). Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Research*, *22*(4), 693–703. http://doi.org/10.1101/gr.129460.111
Mulcair, M. D., Schaeffer, P. M., Oakley, A. J., Cross, H. F., Neylon, C., Hill, T. M., &

Dixon, N. E. (2006). A molecular mousetrap determines polarity of termination of DNA replication in E. coli. *Cell*, *125*(7), 1309–1319. http://doi.org/10.1016/j.cell.2006.04.040

Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., et al. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature*, *416*(6876), 103–107. http://doi.org/10.1038/nature722

Nishino, K., Kushima, M., Matsuo, Y., Matsuo, Y., & Kawamukai, M. (2015). Cell Lysis in S. pombe ura4 Mutants Is Suppressed by Loss of Functional Pub1, Which Regulates the Uracil Transporter Fur4. *PLoS ONE*, *10*(11), e0141796. http://doi.org/10.1371/journal.pone.0141796

Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P., & Lu, T. K. (2014). Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular Cell*, *54*(4), 698–710. http://doi.org/10.1016/j.molcel.2014.04.022

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., et al. (2014). Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*, *156*(4), 836–843. http://doi.org/10.1016/j.cell.2014.01.027

Noguchi, C., & Noguchi, E. (2007). Sap1 Promotes the Association of the Replication Fork Protection Complex With Chromatin and Is Involved in the Replication Checkpoint in Schizosaccharomyces pombe. *Genetics*, *175*(2), 553–566. http://doi.org/10.1534/genetics.106.065334

Noguchi, E., Noguchi, C., McDonald, W. H., Yates, J. R., & Russell, P. (2004). Swi1 and Swi3 are components of a replication fork protection complex in fission yeast. *Molecular and Cellular Biology*, *24*(19), 8342–8355. http://doi.org/10.1128/MCB.24.19.8342-8355.2004

Ovchinnikov, I., Troxel, A. B., & Swergold, G. D. (2001). Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. *Genome Research*, *11*(12), 2050–2058. http://doi.org/10.1101/gr.194701

Pabo, C. O., Peisach, E., & Grant, R. A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. *Annual Review of Biochemistry*, *70*, 313–340. http://doi.org/10.1146/annurev.biochem.70.1.313

Pardue, M. L., Danilevskaya, O. N., Lowenhaupt, K., Slot, F., & Traverse, K. L. (1996). Drosophila telomeres: new views on chromosome evolution. *Trends in Genetics : TIG*, *12*(2), 48–52.

Peterson-Burch, B. D., & Voytas, D. F. (2002). Genes of the Pseudoviridae (Ty1/copia retrotransposons). *Molecular Biology and Evolution*, *19*(11), 1832–1845.

Pidoux, A., Mellone, B., & Allshire, R. (2004). Analysis of chromatin in fission yeast. *Methods (San Diego, Calif.)*, *33*(3), 252–259. http://doi.org/10.1016/j.ymeth.2003.11.021

Pinter, S. F., Aubert, S. D., & Zakian, V. A. (2008). The Schizosaccharomyces pombe Pfh1p DNA helicase is essential for the maintenance of nuclear and mitochondrial DNA. *Molecular and Cellular Biology*, *28*(21), 6594–6608. http://doi.org/10.1128/MCB.00191-08

Porteus, M. H., & Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science*, *300*(5620), 763. http://doi.org/10.1126/science.1078395

Pryce, D. W., Ramayah, S., Jaendling, A., & McFarlane, R. J. (2009). Recombination at

DNA replication fork barriers is not universal and is differentially regulated by Swi1. *Proceedings of the National Academy of Sciences*, *106*(12), 4770–4775. http://doi.org/10.1073/pnas.0807739106

Pryciak, P. M., & Varmus, H. E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell*, *69*(5), 769–780.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183. http://doi.org/10.1016/j.cell.2013.02.022

Qi, X., & Sandmeyer, S. (2012). In Vitro Targeting of Strand Transfer by the Ty3 Retroelement Integrase. *Journal of Biological Chemistry*, *287*(22), 18589–18595. http://doi.org/10.1074/jbc.M111.326025

Qi, X., Daily, K., Nguyen, K., Wang, H., Mayhew, D., Rigor, P., et al. (2012). Retrotransposon profiling of RNA polymerase III initiation sites. *Genome Research*, *22*(4), 681–692. http://doi.org/10.1101/gr.131219.111

Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., et al. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Research*, *40*(4), 1666–1683. http://doi.org/10.1093/nar/gkr863

Ranganathan, V., Wahlin, K., Maruotti, J., & Zack, D. J. (2014). Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. *Nature Communications*, *5*, 4516. http://doi.org/10.1038/ncomms5516

Rosado-Lugo, J. D., & Hampsey, M. (2014). The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. *Journal of Biological Chemistry*, *289*(49), 33916–33926. http://doi.org/10.1074/jbc.M114.608695

Roseaulin, L. C., Noguchi, C., Martinez, E., Ziegler, M. A., Toda, T., & Noguchi, E. (2013). Coordinated degradation of replisome components ensures genome stability upon replication stress in the absence of the replication fork protection complex. *PLoS Genetics*, *9*(1), e1003213. http://doi.org/10.1371/journal.pgen.1003213

Roseaulin, L., Yamada, Y., Tsutsui, Y., Russell, P., Iwasaki, H., & Arcangioli, B. (2008). Mus81 is essential for sister chromatid recombination at broken replication forks. *The EMBO Journal*, *27*(9), 1378–1387. http://doi.org/10.1038/emboj.2008.65

Rozenzhak, S., Mejia-Ramirez, E., Williams, J. S., Schaffer, L., Hammond, J. A., Head, S. R., & Russell, P. (2010). Rad3 decorates critical chromosomal domains with gammaH2A to protect genome integrity during S-Phase in fission yeast. *PLoS Genetics*, *6*(7), e1001032. http://doi.org/10.1371/journal.pgen.1001032

Rubin, G. M., Kidwell, M. G., & Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. *Cell*, *29*(3), 987–994. http://doi.org/10.1016/0092-8674(82)90462-7

Ryu, G. H. (2004). Genetic and biochemical analyses of Pfh1 DNA helicase function in fission yeast. *Nucleic Acids Research*, *32*(14), 4205–4216. http://doi.org/10.1093/nar/gkh720

Sabouri, N., Capra, J. A., & Zakian, V. A. (2014). The essential Schizosaccharomyces pombe Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage. *BMC Biology*, *12*, 101. http://doi.org/10.1186/s12915-014-0101-5

Sabouri, N., McDonald, K. R., Webb, C. J., Cristea, I. M., & Zakian, V. A. (2012). DNA replication through hard-to-replicate sites, including both highly transcribed RNA Pol II and Pol III genes, requires the S. pombe Pfh1 helicase. *Genes & Development*, *26*(6), 581–593. http://doi.org/10.1101/gad.184697.111

Salyers, A. A., Shoemaker, N. B., Stevens, A. M., & Li, L. Y. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiological Reviews*, *59*(4), 579–590.

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*(4), 347–355. http://doi.org/10.1038/nbt.2842

Sargent, R. G., Brenneman, M. A., & Wilson, J. H. (1997). Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Molecular and Cellular Biology*, *17*(1), 267–277.

Sánchez-Gorostiaga, A., López-Estraño, C., Krimer, D. B., Schvartzman, J. B., & Hernández, P. (2004). Transcription termination factor reb1p causes two replication fork barriers at its cognate sites in fission yeast ribosomal DNA in vivo. *Molecular and Cellular Biology*, *24*(1), 398–406. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=14673172&retmode=ref&cmd=prlinks

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonçalves, Â., Kutter, C., et al. (2012). Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. *Cell*, *148*(1-2), 335–348. http://doi.org/10.1016/j.cell.2011.11.058

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science*, *326*(5956), 1112–1115. http://doi.org/10.1126/science.1178534

Sedivy, J. M., & Sharp, P. A. (1989). Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(1), 227–231.

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, *343*(6166), 84–87. http://doi.org/10.1126/science.1247005

Shapiro, J. A. (1969). Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. *Journal of Molecular Biology*, *40*(1), 93–105. http://doi.org/10.1016/0022-2836(69)90298-8

Sharma, A., Larue, R. C., Plumb, M. R., Malani, N., Male, F., Slaughter, A., et al. (2013). BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proceedings of the National Academy of Sciences*, *110*(29), 12036–12041. http://doi.org/10.1073/pnas.1307157110

Shi, A., Liu, O., Koenig, S., Banerjee, R., Chen, C. C.-H., Eimer, S., & Grant, B. D. (2012). RAB-10-GTPase-mediated regulation of endosomal phosphatidylinositol-4,5-bisphosphate. *Proceedings of the National Academy of Sciences*, *109*(35), E2306–15. http://doi.org/10.1073/pnas.1205278109

Singh, B. N., Ansari, A., & Hampsey, M. (2009). Detection of gene loops by 3C in yeast. *Methods (San Diego, Calif.)*, *48*(4), 361–367. http://doi.org/10.1016/j.ymeth.2009.02.018

Singleton, T. L., & Levin, H. L. (2002). A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. *Eukaryotic Cell*, *1*(1), 44–55.

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, *8*(4), 272–285. http://doi.org/10.1038/nrg2072

Solyom, S., & Kazazian, H. H. (2012). Mobile elements in the human genome: implications for disease. *Genome Medicine*, *4*(2), 12–12. http://doi.org/10.1186/gm311

Sommariva, E., Pellny, T. K., Karahan, N., Kumar, S., Huberman, J. A., & Dalgaard, J. Z. (2005). Schizosaccharomyces pombe Swi1, Swi3, and Hsk1 are components of a novel S-phase response pathway to alkylation damage. *Molecular and Cellular Biology*, *25*(7), 2770–2784. http://doi.org/10.1128/MCB.25.7.2770-2784.2005

Sorek, R., Lawrence, C. M., & Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual Review of Biochemistry*, *82*, 237–266. http://doi.org/10.1146/annurev-biochem-072911-172315

Spradling, A. C., Bellen, H. J., & Hoskins, R. A. (2011). Drosophila P elements preferentially transpose to replication origins. *Proceedings of the National Academy of Sciences*, *108*(38), 15948–15953. http://doi.org/10.1073/pnas.1112960108

Steinacher, R., Osman, F., Dalgaard, J. Z., Lorenz, A., & Whitby, M. C. (2012). The DNA helicase Pfh1 promotes fork merging at replication termination sites to ensure genome stability. *Genes & Development*, *26*(6), 594–602. http://doi.org/10.1101/gad.184663.111

Steinhauser, M. L., Bailey, A. P., Senyo, S. E., Guillermier, C., Perlstein, T. S., Gould, A. P., et al. (2012). Multi-isotope imaging mass spectrometry quantifies stem cell division and metabolism. *Nature*, *481*(7382), 516–519. http://doi.org/10.1038/nature10734

Storici, F., Lewis, L. K., & Resnick, M. A. (2001). In vivo site-directed mutagenesis using oligonucleotides. *Nature Biotechnology*, *19*(8), 773–776. http://doi.org/10.1038/90837

Suh, A., Witt, C. C., Menger, J., Sadanandan, K. R., Podsiadlowski, L., Gerth, M., et al. (2016). Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. *Nature Communications*, *7*, 11396. http://doi.org/10.1038/ncomms11396

Szilard, R. K., Jacques, P.-E., Laramée, L., Cheng, B., Galicia, S., Bataille, A. R., et al. (2010). Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. *Nat Struct Mol Biol*, *17*(3), 299–305. http://doi.org/10.1038/nsmb.1754

Takeuchi, Y., Horiuchi, T., & Kobayashi, T. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes & Development*, *17*(12), 1497–1506. http://doi.org/10.1101/gad.1085403

Tanaka, T., Yokoyama, M., Matsumoto, S., Fukatsu, R., You, Z., & Masai, H. (2010). Fission Yeast Swi1-Swi3 Complex Facilitates DNA Binding of Mrc1. *Journal of Biological Chemistry*, *285*(51), 39609–39622. http://doi.org/10.1074/jbc.M110.173344

Terns, R. M., & Terns, M. P. (2014). CRISPR-based technologies: prokaryotic defense

weapons repurposed. *Trends in Genetics : TIG*, *30*(3), 111–118. http://doi.org/10.1016/j.tig.2014.01.003

Teysset, L., Dang, V.-D., Kim, M. K., & Levin, H. L. (2003). A long terminal repeat-containing retrotransposon of Schizosaccharomyces pombe expresses a Gag-like protein that assembles into virus-like particles which mediate reverse transcription. *Journal of Virology*, *77*(9), 5451–5463.

Tsankov, A., Yanagisawa, Y., Rhind, N., Regev, A., & Rando, O. J. (2011). Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. *Genome Research*, *21*(11), 1851–1862. http://doi.org/10.1101/gr.122267.111

Tsukahara, S., Kawabe, A., Kobayashi, A., Ito, T., Aizu, T., Shin-I, T., et al. (2012). Centromere-targeted de novo integrations of an LTR retrotransposon of Arabidopsis lyrata. *Genes & Development*, *26*(7), 705–713. http://doi.org/10.1101/gad.183871.111

Ungerer, M. C., Strakosh, S. C., & Zhen, Y. (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Current Biology : CB*, *16*(20), R872–3. http://doi.org/10.1016/j.cub.2006.09.020

Van Driessche, B., Tafforeau, L., Hentges, P., Carr, A. M., & Vandenhaute, J. (2005). Additional vectors for PCR-based gene tagging in Saccharomyces cerevisiae and Schizosaccharomyces pombe using nourseothricin resistance. *Yeast*, *22*(13), 1061–1068. http://doi.org/10.1002/yea.1293

Vrljicak, P., Tao, S., Varshney, G. K., Quach, H. N. B., Joshi, A., LaFave, M. C., et al. (2016). Genome-Wide Analysis of Transposon and Retroviral Insertions Reveals Preferential Integrations in Regions of DNA Flexibility. *G3 (Bethesda, Md.)*. http://doi.org/10.1534/g3.115.026849

Walisko, O., Izsvák, Z., Szabó, K., Kaufman, C. D., Herold, S., & Ivics, Z. (2006). Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(11), 4062–4067. http://doi.org/10.1073/pnas.0507683103

Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., et al. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proceedings of the National Academy of Sciences*, *104*(47), 18613–18618. http://doi.org/10.1073/pnas.0703637104

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. http://doi.org/10.1038/nrg2165

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, *453*(7199), 1239–1243. http://doi.org/10.1038/nature07002

Wilhelm, M., Heyman, T., Friant, S., & Wilhelm, F. X. (1997). Heterogeneous terminal structure of Ty1 and Ty3 reverse transcripts. *Nucleic Acids Research*, *25*(11), 2161–2166.

Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R., & Voytas, D. F. (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Molecular and Cellular Biology*, *21*(19),

6606–6614.

Xiong, Y., & Eickbush, T. H. (1988). Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Molecular Biology and Evolution*, *5*(6), 675–690.

Yant, S. R., Wu, X., Huang, Y., Garrison, B., Burgess, S. M., & Kay, M. A. (2005). High-resolution genome-wide mapping of transposon integration in mammals. *Molecular and Cellular Biology*, *25*(6), 2085–2094. http://doi.org/10.1128/MCB.25.6.2085-2094.2005

Zaratiegui, M., Castel, S. E., Irvine, D. V., Kloc, A., Ren, J., Li, F., et al. (2011). RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. *Nature*, *479*(7371), 135–138. http://doi.org/10.1038/nature10501

Zaratiegui, M., Vaughn, M. W., Irvine, D. V., Goto, D., Watt, S., Bähler, J., et al. (2010). CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. *Nature*, *469*(7328), 112–115. http://doi.org/10.1038/nature09608

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. http://doi.org/10.1186/gb-2008-9-9-r137

Zhao, H., & Russell, P. (2004). DNA binding domain in the replication checkpoint protein Mrc1 of Schizosaccharomyces pombe. *The Journal of Biological Chemistry*, *279*(51), 53023–53027. http://doi.org/10.1074/jbc.M410449200

Zhou, Z.-X., Zhang, M.-J., Peng, X., Takayama, Y., Xu, X.-Y., Huang, L.-Z., & Du, L.-L. (2013). Mapping genomic hotspots of DNA damage by a single-strand-DNA-compatible and strand-specific ChIP-seq method. *Genome Research*, *23*(4), 705–715. http://doi.org/10.1101/gr.146357.112

Zhu, Y., Zou, S., Wright, D. A., & Voytas, D. F. (1999). Tagging chromatin with retrotransposons: target specificity of the Saccharomyces Ty5 retrotransposon changes with the chromosomal localization of Sir3p and Sir4p. *Genes & Development*, *13*(20), 2738–2749.

Zou, S., Ke, N., Kim, J. M., & Voytas, D. F. (1996). The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes & Development*, *10*(5), 634–645.