DNA Dynamics in Drosophila: Detecting Chromosomal Translocations Using High-Throughput Genome-Wide Translocation Sequencing

Vicki Do

Senior Honors Thesis
Department of Biology
Tufts University
Under the Advisement of Professor Mitch McVey and Nick Woodward
May 2nd, 2023

Table of Contents

Abstract

Chromosomal translocations are structural genome variants (SGVs) that arise when two non-homologous chromosomes exchange genetic material through the repair of two separate DNA double-strand breaks (DSBs). The repair of these breaks can lead to the formation of fusion genes or alterations in gene expression, which has been implicated in the development of many malignancies and diseases, particularly cancer. The repair process for these breaks involves either classical non-homologous end joining (c-NHEJ) or alternative end-joining (alt-EJ), with the latter being more mutagenic and utilizing microhomology (MH) sequences aligned by polymerase theta (POLQ). The literature on translocation formation has presented conflicting data regarding the underlying mechanisms and factors involved in this process. In humans, translocation formation appears to be driven by c-NHEJ whereas in mice, alt-EJ plays a more dominant role. High-throughput genome-wide translocation sequencing (HTGTS) is a technique that has been used to study translocations in mammalian cells and allows for the identification of translocation junctions between two different genomic loci termed “bait” and “prey” sites. This project attempts to adopt the HTGTS workflow for *Drosophila melanogaster*, utilizing Cas9 and single-guide RNAs (sgRNAs) with low specificity to generate translocations at various sites throughout the fly genome. Validation of efficient sgRNAs has been achieved through *in vitro* cutting assays, and these guide constructs have been successfully cloned into the pU6-BbsI-chiRNA plasmid expression vector. The next steps will be to continue transfecting these plasmids containing the sgRNAs in Cas9-expressing Drosophila S2 cells for Cas9 cleavage, and the translocation junctions will be sequenced using high-throughput amplicon sequencing and analyzed using a published HTGTS bioinformatic pipeline. This work will be done in different DNA repair backgrounds to elucidate the contributions of each in translocation formation as well as identify the role of POLQ and alt-EJ in mediating these events. The same sgRNA expression plasmids were also injected into early Drosophila embryos endogenously expressing Cas9 to investigate the potential for inducing breaks and translocations in their genomes, in parallel with the HTGTS project. In an additional but related study, a previously established plasmid-based system was adapted as a proxy to investigate the preferences for types of chromosomal rearrangements in Drosophila.

Introduction

Chromosomal translocations are structural genome variants (SGVs) that arise when two separate DNA double-strand breaks (DSBs) are repaired, resulting in the exchange of genetic material between two non-homologous chromosomes. Thus, repair of the incorrect break ends must be repaired *in trans* as opposed to *in cis*. It is worth noting that because correct ends are located proximally to one another following a DSB, this means that joining *in cis* is expected to be favored over joining *in trans*, and this assumption has been confirmed by experimental evidence (Zhang et al., 2012). Furthermore, as chromosomes are arranged in distinct domains within the nucleus, translocation formation can also be influenced by the relative positioning of the chromosomal breaks and the mobility of the break ends, both of which may be constrained by the structure of the chromatin fiber (Clouaire & Legube, 2019).

There are two main classes of chromosomal translocations: reciprocal translocations and Robertsonian translocations (Figure 1). Reciprocal translocations occur when two non-homologous chromosomes exchange portions with each other, resulting in two new chromosomes that each contain a combination of genetic material from the original chromosomes. Robertsonian translocations arise when two acrocentric chromosomes (chromosomes with the centromere near one end) fuse together, producing a single chromosome containing genetic material from both the original chromosomes as well as a smaller fragment that is lost.

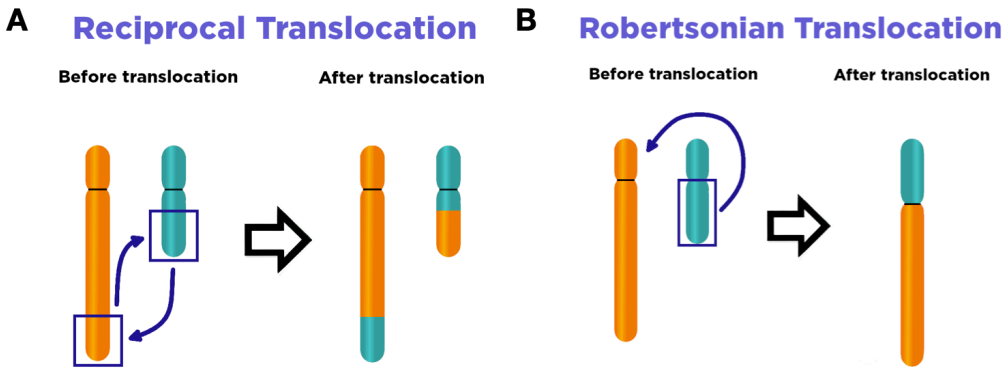

Figure 1. (A) Reciprocal translocations occur when two non-homologous chromosomes exchange genetic material. (B) Robertsonian translocations occur when two acrocentric chromosomes join at their centromeres to form a single, larger chromosome. Figure adapted from Mario Janakis.

In addition, translocations can be either balanced, where there is an equal exchange of genetic material between two chromosomes, resulting in no net loss or gain, or they can be unbalanced where genetic material is lost or gained in one or both chromosomes involved. Reciprocal and Robertsonian translocations can occur in both balanced and unbalanced states.

Chromosomal translocations can affect gene expression in several ways, one of which is the formation of a fusion gene. These arise when two different chromosomes break and the resulting broken ends are erroneously rejoined with each other, leading to the fusion of two previously separate genes. Alternatively, translocations can place a gene under the control of a different promoter or enhancer that increases or decreases gene expression. Translocations leading to fusion genes and altered gene expression have been implicated in various types of malignancies, and it has been estimated that known translocations drive approximately 20 percent of cancer cases (Mitelman et al., 2007). In many cases, these SGVs create oncogenic proteins that facilitate cell growth and proliferation. For example, the BCR-ABL fusion gene resulting from the reciprocal translocation between chromosomes 9 and 22 is associated with chronic myelogenous leukemia and produces a constitutively active tyrosine kinase. When these kinases are aberrantly activated, they can disturb downstream signaling pathways that can result in increased cell proliferation and resistance to apoptosis (Kang et al., 2016). This can lead to uncontrolled cell growth and ultimately promote cancer progression.

The formation of translocations has been linked to two different DNA repair pathways, namely classical non-homologous end joining (c-NHEJ) and alternative end-joining (alt-EJ), as shown in Figure 2. In somatic mammalian cells, c-NHEJ is the predominant pathway for repairing DSBs and operates in all phases of the cell cycle. It involves the synapsis of the broken DNA ends mediated by the Ku70/Ku80 heterodimer, followed by the recruitment of DNA-PKcs in association with Artemis for processing DSBs. The resulting gaps are then filled with the assistance of polymerases mu and lambda. The final step in c-NHEJ is mediated by the XRCC4-Ligase IV complex which catalyzes the ligation of the DNA ends, ultimately repairing the break. However, if DSBs are not immediately repaired, they can persist and form translocations via the end-joining of two distinct DSBs. While this pathway is generally considered error-free, it can generate insertions and deletions (indels) flanking the break site, typically in the range of 1-4

base pairs (bp). Compared to c-NHEJ, alt-EJ is a less well-characterized set of repair mechanisms and occurs as a more mutagenic form of repair when c-NHEJ is unavailable (Deriano & Roth, 2013). There is still a debate on whether alt-EJ involves a single mechanism or multiple overlapping ones. Previous studies have shown that one form of alt-EJ called microhomology-mediated end-joining (MMEJ) utilizes exposed 1-6 bp microhomology (MH) sequences that flank the break site prior to the ligation of the broken ends to drive repair. These MH sequences are typically aligned with polymerase theta (POLQ), an error-prone and highly conserved A-family DNA polymerase. Because of POLQ's significant role in facilitating this repair process, this form of alt-EJ is known as theta-mediated end-joining (TMEJ).

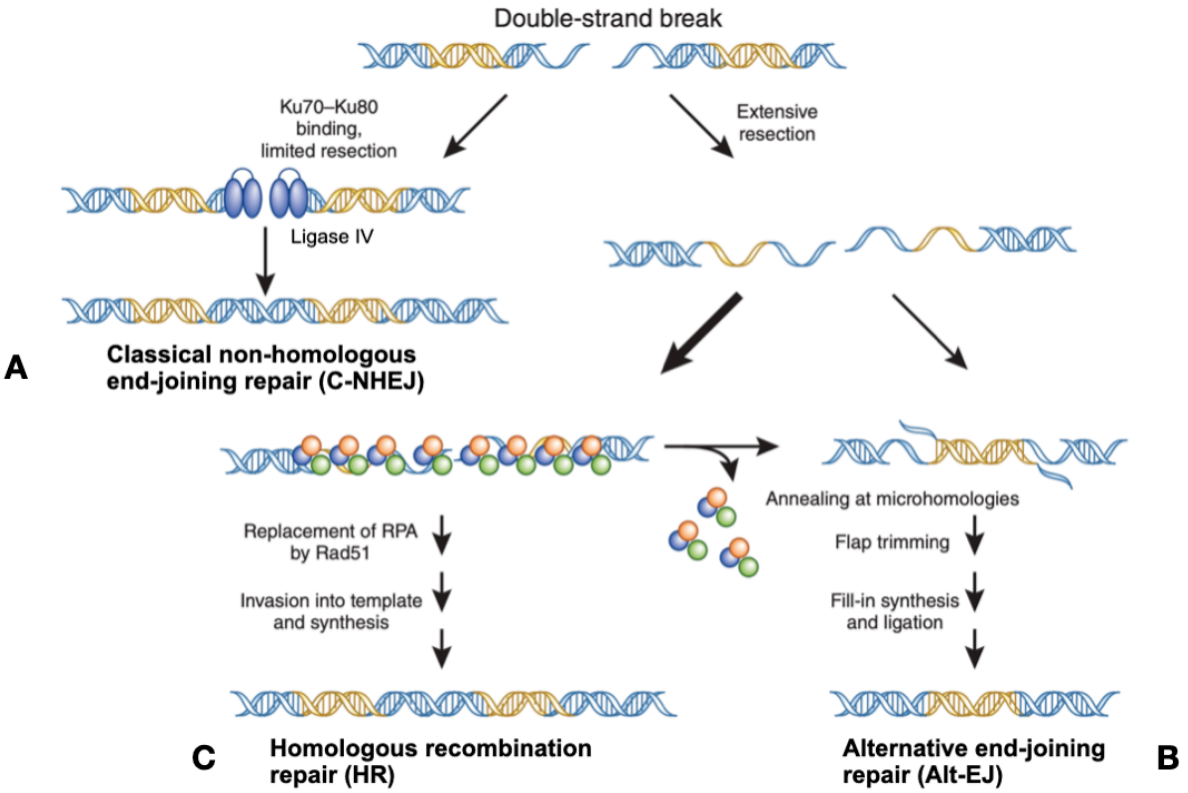

Figure 2. DNA repair pathways for resolving double-strand breaks: (A) C-NHEJ is mediated by a complex of proteins, including Ku70/80, DNA-PKcs, XRCC4, and Ligase IV, that work together to ligate the broken DNA ends without the need for a homologous template. (B) Alt-EJ involves short-range resection of DNA ends exposing microhomology sequences on opposite sides of the break that can be aligned, typically by POLQ, to facilitate repair. (C) Homologous recombination (HR) is a DNA repair mechanism that uses a homologous sequence as a template to accurately repair DNA double-strand breaks, resulting in error-free repair. Figure adapted from McVey, 2014.

Previous studies have demonstrated the central role of c-NHEJ in preserving genomic stability, as observed by the increase in the frequency of spontaneous chromosome and chromatid breaks in cells lacking functional c-NHEJ factors (Difilippantonio et al, 2000; Ferguson et al, 2000). This suggests that translocations should be suppressed by c-NHEJ, as increased breaks should lead to more translocations. In contrast, cellular depletion of alt-EJ factors such as DNA Ligase III and CtIP showed a reduction in translocation frequency in mouse cells (Simsek et al, 2011; Zhang & Jasin, 2011). The reduction in translocations indicates that canonical NHEJ does not play a dominant role in translocation formation in mice. However, published findings on the underlying mechanism of translocation formation appear to vary depending on the model organism studied. For instance, in mouse cells, it was shown that loss of c-NHEJ factors Ku and Ligase IV increased translocation frequency (Boboila et al., 2010). This supports the idea that when c-NHEJ is absent, alt-EJ is likely to become more dominant in repairing DSBs and contribute to the formation of translocations, as discussed earlier. In human cell lines, findings implicate the opposite: the frequency of translocations decreased in a c-NHEJ-deficient background (Ghezraoui et al., 2014).

Intriguingly, there have been inconsistencies in experimental outcomes even within the same model organism. For example, in one study, alt-EJ was implicated in translocation formation in which the absence of POLQ suppressed the overall frequency of translocations in mouse embryonic fibroblasts (Mateos-Gomez et al., 2015). However, another study demonstrated that there was a 4-fold increase in translocation frequency (between IgH and myc loci) in mouse B lymphocyte cells in the absence of POLQ, implying that POLQ functions to suppress genomic instability and translocations (Yousefzadeh et al., 2014). A third study investigated the influence of POLQ deficiency on translocations between Cas9-induced breaks in transformed mouse embryonic fibroblasts. It suggested that the loss of POLQ alone did not affect the frequency of translocations, but when combined with Ku/c-NHEJ deficiency, it resulted in increased frequency, implying that POLQ serves as a secondary pathway for suppressing translocations (Wyatt et al., 2016). Despite controlling for organisms, there are still significant variations in experimental models and outcomes among these published findings such as the techniques employed to introduce chromosome breaks and the types of cells being analyzed.

Taken together, these disparate results indicate that there is no consensus on the exact mechanisms involved in translocation formation.

Establishing a connection between translocations and alt-EJ mechanisms is challenging due to the intricate and incompletely understood nature of the pathway. That being said, one consistent factor in the literature on translocations is that translocation junctions contain short stretches of MH typically in the range of 1-6 bp (Zhu et al., 2002). Repair events that create these MH signatures frequently implicate POLQ (Mateos-Gomez et al., 2015). As described earlier, POLQ is an error-prone, translesion DNA polymerase belonging to the family of class A polymerases. It has a characteristic N-terminal helicase-like domain that is connected to a C-terminal polymerase domain by a disordered region called the linker domain (Zahn et al., 2015). POLQ is highly involved in alt-EJ mechanisms and functions to mediate the ligation of two resected 3' ends of DNA strands that share short regions of MH (Feng et al., 2019). In normal cells, homologous recombination (HR) is one of the repair pathways for accurately resolving DSBs (Figure 2C). However, when the HR pathway is compromised, as is the case with mutations in the *BRCA1* and *BRCA2* genes, cells must rely on error-prone alt-EJ mechanisms to repair DSBs. POLQ has been identified as a potential target for HR-deficient cancers because it demonstrates synthetic lethality interactions with HR, e.g. those involving *BRCA1/2* defects. Analysis of *BRCA1*- and *BRCA2*-mutated tumors revealed that these tumors have distinct mutational profiles characterized by higher numbers of point mutations and deletions, distinguishing them from sporadic tumors, and that there was a significant increase in interchromosomal translocations in *BRCA1*-mutated tumors than other classes of tumors. These translocations also displayed a tendency to have MH signatures ranging from 1-10 bp at the breakpoints (Lal et al., 2019). Furthermore, high levels of POLQ expression have been observed in certain subtypes of breast and ovarian tumors with HR deficiencies (Ceccaldi et al, 2015).

In 2016, Beagan and McVey reviewed the link between POLQ and the formation of chromosomal translocations, and presently, this link continues to remain unclear. As discussed previously, POLQ has contrasting effects on translocation formation depending on the genomic context. For example, POLQ was shown to be necessary for both alt-EJ-mediated translocations and telomere fusion events at de-protected telomeres in mouse cells, where it utilized pre-

existing telomeric MHs (Mateos-Gomez et al., 2015). In another study, POLQ inhibited spontaneous translocations between the Myc and IgH loci in a POLQ-defective mouse model, implying that it takes on a role in suppressing translocations and other translocation-prone end-joining pathways (Yousefzadeh et al, 2014; reviewed in Beagan & McVey, 2016). Again, the reasons for these discrepant effects of POLQ on translocation formation are unresolved, but it might include differences in the nature of the DSBs or the presence of other alt-EJ repair proteins that could modulate the activity of POLQ in different systems.

At the time of writing this senior thesis, there has been no published work to study the mechanisms of translocation formation in Drosophila. However, Drosophila may be an ideal model organism to investigate translocation formation due to its preference for using alt-EJ as its primary DNA repair pathway. This is supported by evidence from plasmid-based assays, which demonstrate that alt-EJ repair events account for the majority of inaccurate repairs (Hanscom et al., 2022). Moreover, in both chromosomal assays and plasmid-based systems that used DNA recovered from Drosophila embryo injections, more than 75% of repair products exhibited features consistent with TMEJ (Hanscom et al., 2022; Yu & McVey, 2010).

It is worth mentioning that in the work of a previous undergraduate in the McVey laboratory, the genetic requirements for translocation formation in Drosophila were investigated using a novel inter-plasmid system as a proxy (Zuckerman, 2022, p. 40) (Figure 3). This involved using two different plasmids, Iw7 and R421, that both contain single-guide RNA (sgRNA) recognition sites for the Cas9 endonuclease. The plasmids were co-injected into flies with varying repair-deficient genetic backgrounds such as Ligase IV null (c-NHEJ-deficient) as well as POLQ null, and translocation events were isolated following repair (Zuckerman, 2022, p. 40). Translocation events were categorized as apparent blunt joins (ABJ), microhomology joins (MHJ), or indels. ABJs and MHJs can arise from both c-NHEJ and alt-EJ repair pathways, but alt-EJ is more commonly associated with the formation of MHJs. Indels are caused by a specific type of alt-EJ known as synthesis-dependent microhomology-mediated end-joining (SD-MMEJ). Indels can sometimes lead to the creation of a repeat motif at the repair junction, which contains an insertion between the primer and the MH. MHJ and ABJ repeat motifs consist solely of the primer and MH, with no insertion. It was found that in a Ligase IV null background, there was an

increase in indel events during translocation formation which was reflected in a 10% decrease in ABJ events as well as a 10-15% decrease in MHJ events (Zuckerman, 2022, p. 43). In the POLQ null background, however, there was an increase in the percentage of ABJ events but a decrease in MHJs (Zuckerman, 2022, p. 43). Collectively, these results suggest that alt-EJ mechanisms are utilized more frequently in the formation of translocations.

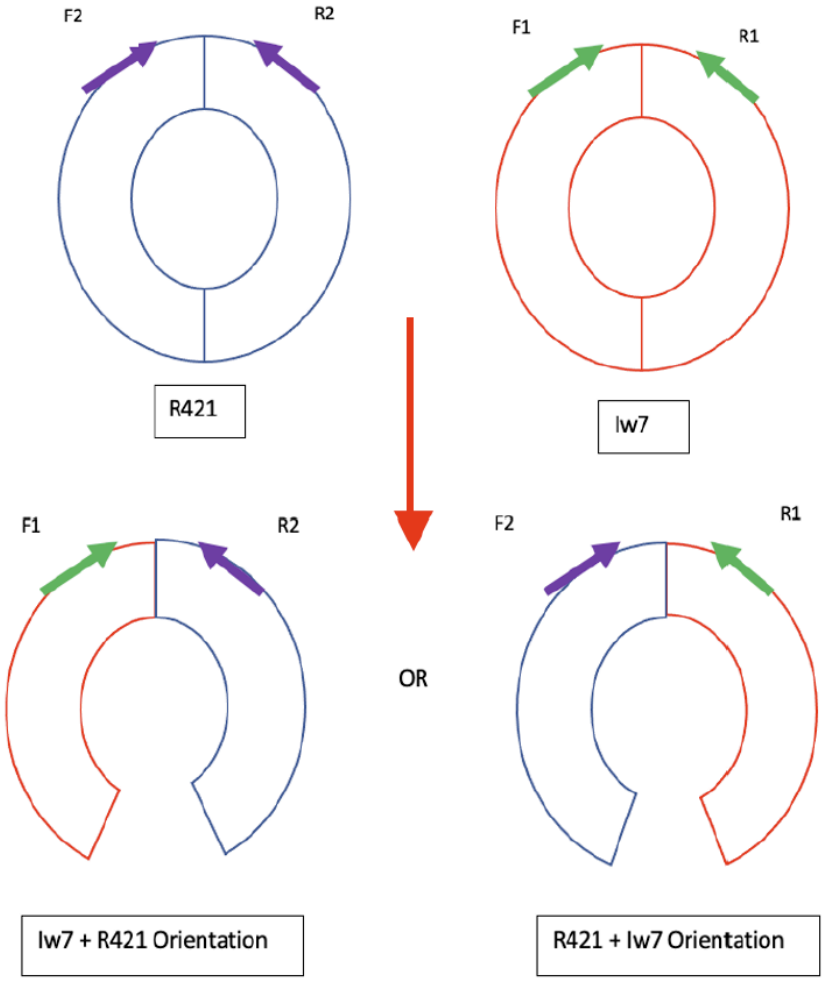

Figure 3. Schematic of the inter-plasmid system as a proxy to study translocations. The inter-plasmid translocation products of R421 (blue) and Iw7 (red) were isolated using PCR with the (A) F1 and R2 primers for the Iw7 + R421 orientation and the (B) F2 and R1 primers for the R421 + Iw7 orientation. Figure adapted from Zuckerman, 2022.

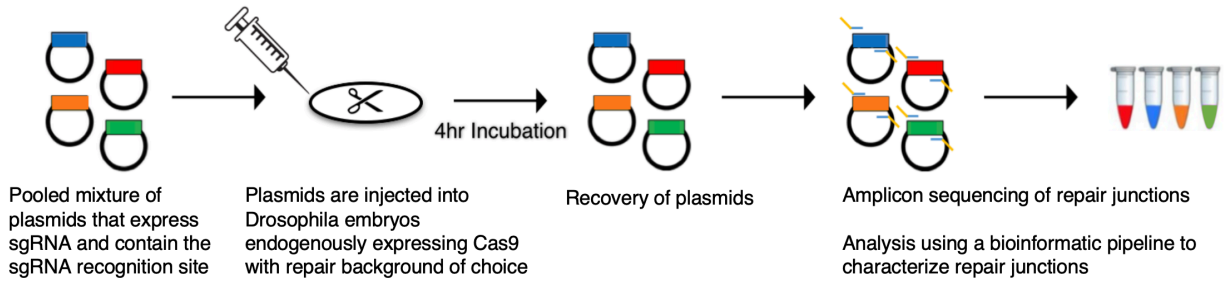

Figure 4. Workflow of the inter-plasmid system. A pooled mixture of plasmid constructs expressing the sgRNA as well as ones containing the Cas9 recognition site was injected into Cas9-expressing Drosophila embryos, followed by incubation to allow cutting and repair. The repaired DNA was then recovered, enriched for inaccurate repair, and subjected to high-throughput amplicon sequencing for the characterization of repair junctions.

While inter-plasmid repair is not entirely equivalent to chromosomal translocations, it may serve as an adequate proxy. Advantages of a plasmid-based system over a chromosome-based system include a significantly reduced time in isolating plasmid products and enriching them for inaccurate repair. It can also generate hundreds of thousands of reads from high-throughput amplicon sequencing that allows for efficient characterization of these repair junctions (Figure 4). Still, there are limitations to the inter-plasmid system. For instance, inconsistent results were noted concerning the primer distance from the break site for the insertional events. While the Iw7 + R421 orientation revealed an increase in mean primer distance in the POLQ null background, the R421 + Iw7 orientation had the opposite effect (Zuckerman, 2022, p. 43). This suggests that it is possible that the sequences flanking the Cas9 cut site might unintentionally favor one repair pathway over another, based on their ability to form DNA structures that drive alt-EJ repair.

Here, I report on the progress of using high-throughput genome-wide translocation sequencing (HTGTS) to detect chromosomal translocations in Drosophila S2 cells, effectively switching from a plasmid-based system to a chromosome-based one. HTGTS is a well-documented means of identifying DSBs and translocations across the genome in a high-throughput manner and specifically, allows for the identification of "prey" DSB formation by translocation to a "bait" DSB site. If a “bait-prey” junction arises from different chromosomes, then it is considered a translocation. This technique has been used to determine the off-target cleavage activity of programmable nucleases such as Cas9 as well as to identify spontaneous

translocations caused by exogenous damaging agents or endogenous processes such as replication and transcription (Frock et al., 2015). To generate translocation events at various locations across the Drosophila genome, Cas9 and sgRNAs with low specificity, i.e. with many predicted off-targets, were utilized.

Linear amplification-mediated HTGTS (LAM-HTGTS) begins with an incubation period following the transfection of Cas9-expressing S2 cells with sgRNAs to allow for DSB formation; after incubation, the genomic DNA is recovered (Figure 5). Following this, an initial PCR is performed using one biotinylated primer located on each end of the "bait" DSB site. This generates single-stranded DNA (ssDNA) products with a biotin molecule attached upstream of the break site, which mostly represents the "bait" locus but also includes any translocation formation that occurred. The biotinylated ssDNA is then purified using streptavidin beads. Double-stranded DNA (dsDNA) adapters are ligated onto the 3' end of the ssDNA, and a nested PCR is performed using a primer that anneals to the adapter to generate dsDNA. Using a rare, known restriction enzyme site or a second Cas9 sgRNA that cleaves downstream of the "bait" DSB site, the amplified perfect rejoins, uncut sequences, or sequences containing indels are removed, enriching for translocation products. An additional PCR is performed, now on the collection of translocation samples only, for Illumina sequencing preparation (Figure 6).

The Alt laboratory has developed a bioinformatic pipeline that can process the resulting raw reads from high-throughput sequencing, classify repair junctions for MH, and identify where in the genome the "prey" sequence originated and maps them across the genome (Hu et al., 2016). The pipeline uses raw sequencing reads and a reference genome file to identify the "prey" locus and sort the sequencing reads using a specific nested primer as an identifier. We have modified this approach for investigating the genetic requirements and repair mechanisms of translocation formation in Drosophila.

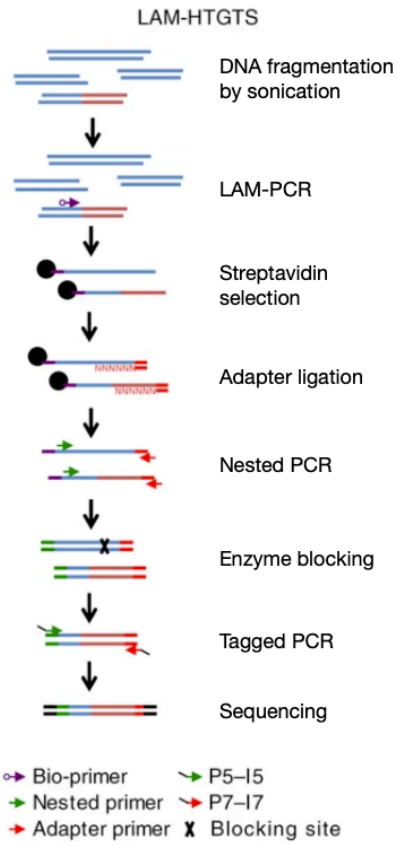

Figure 5. LAM-HTGTS is a method that enables direct amplification of genomic DNA junctions via LAM-PCR, followed by enrichment and bridge adapter ligation. This approach allows for the exponential amplification of enriched products, which can then be labeled for Miseq sequencing. Figure adapted from Hu et al., 2016.

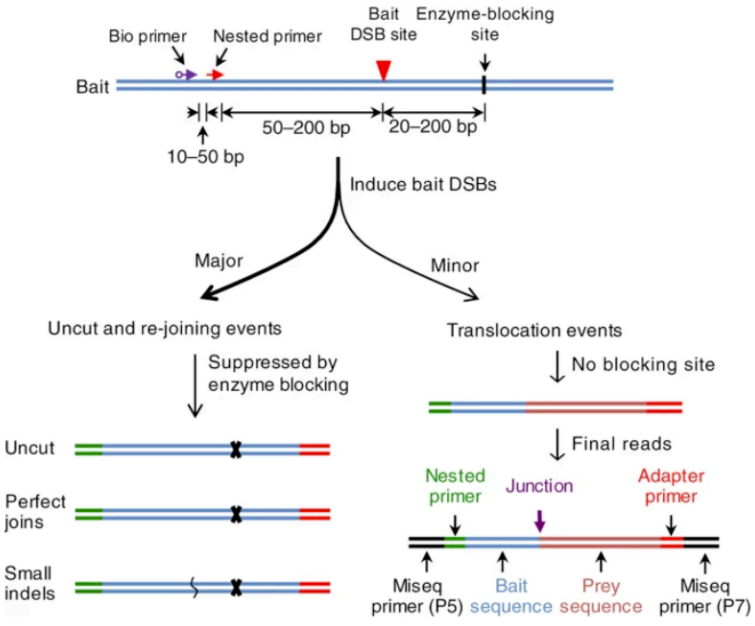

Figure 6. The LAM-HTGTS method uses biotinylated and nested primers to amplify junctions from sheared genomic DNA. Enzyme-blocking sites are chosen to suppress small insertions and deletions around the “bait” DSB sites. The method allows for the identification of translocations between bait DSBs and genome-wide DSBs that can be amplified and sequenced using Miseq. The amplified products contain the “bait” which is composed of the nested primer-binding sequence and the “prey” which refers to the specific DNA sequence from the genome that aligns with the bait sequence, and the junction represents the point at which the “bait” and “prey” sequences are joined together. Figure adapted from Hu et al., 2016.

CRISPR/Cas9 sgRNAs were designed to target highly transcribed, non-essential gene regions across the third fly chromosome. Highly transcribed regions were selected as they have been found to have better Cas9 binding in eukaryotic cells, owing to their euchromatic nature, while non-essential genes were chosen to minimize the lethality associated with Cas9 cleavage and translocation formation. These gene regions include Dci, Mtch, Cypl, Tpi, and Vps16B. One approach used to increase the likelihood of translocation events was the design of sgRNAs with predicted low specificity, thereby increasing the chance of off-target cleavage by Cas9 and the formation of potential "prey" DSBs. Prior to administration to the cells, the sgRNAs were validated *in vitro* for their cleavage efficiency. We have validated and selected nine sgRNAs to evaluate their potential in inducing translocations in one of the aforementioned gene regions. S2 cells expressing the Cas9 protein will be transfected with pU6-BbsI-chiRNA, a plasmid expression vector that has a specific sgRNA oligonucleotide cloned in to target a genomic locus. Successful cloning was verified through sequencing.

In addition to utilizing HTGTS for studying translocation formation in Drosophila S2 cells, a preliminary project was pursued to investigate the potential for inducing breaks and translocations in the genomes of early embryos through the use of the same sgRNA expression plasmids and endogenously expressed Cas9. This preliminary experiment was conducted by injecting a plasmid expressing the sgRNA into embryos younger than two hours old, similar to the inter-plasmid assay shown in Figure 4, and genomic DNA was recovered using phenol-chloroform extraction following overnight incubation. The repair junctions were amplified by polymerase chain reactions (PCRs), and the products were sent for Sanger sequencing to assess the extent of cleavage. This experiment was conducted in parallel with the HTGTS project and

allowed for the exploration of alternative approaches for studying translocation formation and repair mechanisms *in vivo*.

The HTGTS method not only isolates translocation junctions between a chromosomal "prey" DSB and other sequences genome-wide, but it is also highly sensitive in detecting various types of translocations, including intrachromosomal deletions and inversions as well as interchromosomal translocations (Figure 7). For example, in a 2012 paper by Chiarle et al., the researchers utilized HTGTS to identify tens of thousands of independent translocation junctions that were generated by inducing DSBs in the c-myc oncogene or IgH locus of mouse B lymphocytes through the use of I-SceI meganuclease. They were able to distinguish intrachromosomal and interchromosomal junctions and identify various potential outcomes such as deletions, inversions, extra-chromosomal circles, and acentrics.

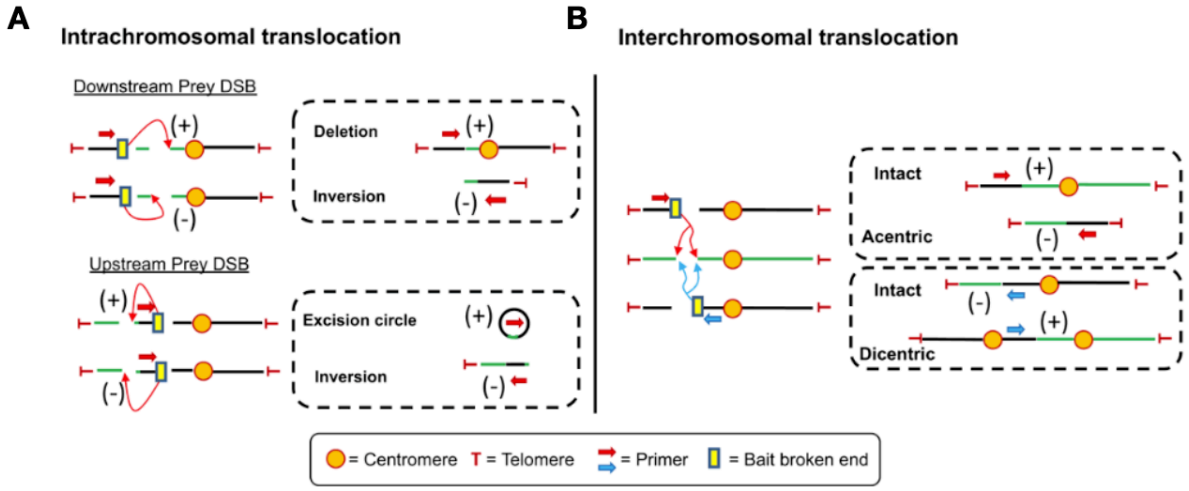

Figure 7. Types of chromosomal rearrangements. (A) Intrachromosomal translocation. Inversions occur when a fragment of DNA between two breakpoints on the same chromosome breaks off and reattaches in an inverted orientation. Deletions occur when the fragment between two breakpoints is not rejoined, resulting in the loss of that DNA. An excision circle can form when the two ends of the lost DNA fragment are rejoined, causing the fragment to circularize. (B) Interchromosomal translocation. Acentric translocations occur when two broken chromosomes lacking a centromere are rejoined. Dicentric translocations occur when two broken chromosomes with centromeres are rejoined, leading to the formation of a chromosome with two centromeres. Figure adapted from https://med.stanford.edu/frocklab/research.html.

To complement HTGTS and investigate the preferences for types of chromosomal rearrangements in Drosophila, the plasmid-based system as shown in Figure 3 was adapted as a

proxy, enabling the study of different types of plasmid rearrangements. A second sgRNA was designed approximately 100 bp downstream from the first sgRNA recognition site in both the Iw7 and R421 plasmids which would result in the creation of an interstitial fragment upon cleavage at the two Cas9 cut sites. In this preliminary experiment, multiple PCRs with unique primer pairs and subsequent gel electrophoresis to run the PCR products were used to identify inversion and deletion events in the Iw7 plasmid. This step was taken to detect these events in one plasmid initially, before moving on to isolate more complex rearrangements between two plasmids.

Materials and Methods

Determining Genomic Regions of Interest and Nomination of Single-Guide RNAs

To minimize the risk of lethal repair events, the ends of the third fly chromosome were scanned using Flybase's JBrowse feature, with a focus on regions with high expression levels and similar expression profiles between *Drosophila melanogaster* and Drosophila cell lines. The selection of chromosome ends as target regions aim to mitigate the possibility of an unbalanced translocation resulting in lethality prior to the time of genomic DNA collection. A filter was applied on Flybase to identify sgRNAs with a high number of off-targets, as off-target DSBs increase the likelihood of translocation repair outcomes.

Using a CRISPR/Cas9 target online predictor tool, CCTop, sgRNAs with a CRISPRater score of medium or high with multiple off-targets were noted as lead candidates for a specific locus. SgRNAs in which there were rare restriction enzyme sites located within 20-200 bp downstream from the cut-site were nominated as ideal candidates for this locus (Figure 6). Six genomic loci were identified using this method.

***In Vitro* Cutting Assay**

In order to evaluate the *in vitro* cutting efficiency of candidate sgRNAs, up to three sgRNAs per locus were assessed using an *in vitro* cutting assay. The guides in this assay were designed and *in vitro* transcribed using the EnGen® sgRNA Synthesis Kit, *S. pyogenes* Protocol (NEB #E3322), following the manufacturer's instructions. The *in vitro* cutting assay began with amplification of a region flanking the cut sites by PCR, approximately 500-600 bp in length, followed by incubation of the PCR product with pre-formed Cas9 ribonucleoprotein (RNP). Cas9 RNP was made by incubating 200 ng of the *in vitro* transcribed sgRNA, 5000 ng of Cas9, 2 µL of r3.1 buffer (NEB), and water until the reaction volume reached 20 µL. Following an incubation period of 10 minutes at 25°C, 300 ng of cutting substrate was added, and the mixture was incubated at 37°C for 1 hour. Then, 1 µL of proteinase K was added. The mixture was incubated for an additional 10 minutes at room temperature. The DNA was analyzed by gel electrophoresis to visualize any cleavage products. The efficiency of cutting was determined by comparing the amount of uncut DNA to the amount of DNA cut by Cas9.

Cloning of Oligonucleotides for sgRNA Constructs

1. Preparation of pU6-BbsI-chiRNA Plasmid Vector

The pU6-BbsI-chiRNA vector was linearized using a digestion protocol. To do so, a digestion mix was made that contained the following: 6 µg of pU6-BbsI-chiRNA plasmid vector (Addgene #45946), 6 µL of high-fidelity BbsI restriction enzyme (NEB), 5 µL of rCutSmart Buffer (NEB), and water until the mix reached 50 µL. The digestion mix was incubated at 37°C for 1 hour. The digested vector was run on a 1% agarose gel (85 volts for 1.5 hours), then extracted using a gel extraction protocol according to the manufacturer's specifications from the Monarch® DNA Gel Extraction Kit manual. The extracted vector was eluted in a final volume of 15 µL.

2. Designing and Annealing sgRNA Oligonucleotides Inserts

To clone a sgRNA into the pU6-BbsI-chiRNA vector, two single-stranded oligonucleotides (henceforth will be referred to as oligos) were designed containing the sgRNA target sequence. Oligos should be designed with the following configuration:

Forward oligo (top strand): 5'-CTTCG-sgRNA sequence-3'

Reverse oligo (bottom strand): 5'-AAAC-reverse complement sgRNA sequence-C-3'

The design of the sgRNA oligos creates overhangs during the annealing reaction that can be ligated to the pU6-BbsI-chiRNA vector. The oligos were ordered from Eton Bioscience. The ssDNA oligos were prepared by resuspending them in 1x STE Buffer—comprised of 0.1 M NaCl, 10 mM Tris pH 8.0, and 1 mM EDTA—to 100 µM DNA concentration, then assembled in a 10 µL reaction mix that contained the following: 1 µL oligo 1 (100 µM), 1 µL oligo 2 (100 µM), 1 µL 10x Polynucleotide Kinase (PNK) Buffer from New England Biolabs (NEB), 1 µL T4 PNK (NEB), and 6 µL of water. The reaction mix was incubated in a thermocycler at 37°C for 30 minutes to add 5'-phosphates to the ends of strands, followed by heating up the mix to 95°C for 5 minutes, and ramping down at 0.1°C/s from 95°C to 25°C to anneal the single-stranded oligos.

3. Ligation Reaction

To insert the sgRNA oligos into the pU6-BbsI-chiRNA vector, a ligation reaction was performed. A ligation mix of 50 ng digested pU6-BbsI-chiRNA, 1 µL annealed sgRNA oligo, (20 µM), 1 µL T4 DNA Ligase Reaction Buffer (NEB), 1 µL T4 DNA ligase, and water to 10 µL was made. The mix was incubated for at least 15 minutes at room temperature.

4. Bacterial Transformation

The ligation reaction was transformed into NEB 2987 DH5α chemically competent cells (henceforth referred to as "cells"). A tube containing the cells was thawed on ice for 10 minutes and approximately 22.5 µL was used for each transformation. 2.5 µL (12.5 ng) of the ligation reaction mix was added to the cells and mixed by flicking the tube 4-5 times. The cells were then incubated on ice for 30 minutes. After that, the cells were heat-shocked at 42°C for 45 seconds, then immediately placed on ice to incubate for an additional 5 minutes. 900 µL of room temperature LB broth was added to the mixture, and the tube was allowed to recover in a shaking incubator at 37°C for 1 hour. Following recovery, the cells were spun down at 6000 RCF to pellet, 700 µL of the medium was removed, and then the pellet was resuspended in the remaining 200 µL of the medium. The cells were plated onto an LB agar plate with 100 µg/mL ampicillin and were incubated at 37°C overnight.

5. Alkaline Lysis Mini-Preps of pU6-BbsI-chiRNA

A single colony from the transformed cells was inoculated in 2 mL of LB medium and ampicillin at a final antibiotic concentration of 100 µg/mL, and the culture was grown overnight in a shaking incubator at 37°C. The following day, 1.5 mL of the overnight culture was transferred into Eppendorf tubes and spun down at maximum speed (15,000 RPM) in a centrifuge for 2 minutes. The remaining 0.5 mL was saved as a 50% glycerol stock and stored at -80C°C. When the centrifugation was complete, the bacterial pellet was resuspended in 100 µL of P1 solution—comprised of 50 mM Tris-HCl pH 7.5 and 10 mM EDTA—by pipetting it up and down. 100 µL of P2 solution—comprised of 0.2 N NaOH and 1% SDS—was added to the bacterial suspension and mixed by pipetting up and down. The mixture was allowed to rest at

room temperature for up to 5 minutes. 100 µL of cold P3 solution—comprised of 1.32 M KOAc pH 4.8—was added and mixed by pipetting once again. The bacterial lysate was centrifuged at maximum speed for 3 minutes, and the supernatant was transferred to a new Eppendorf tube along with 750 µL of 95% ethanol and mixed by inverting the tube 3 times. The tube was then incubated on ice for 10 minutes and spun at maximum speed for 5 minutes. The pellet was washed with 500 µL 75% ethanol to remove excess salts, then centrifuged for an additional 2 minutes at maximum speed. The pellet was dried and resuspended in 30 µL of water. To prep products, RNase A was added to a final concentration of 0.5 mg/mL, and the tube was incubated at 37°C for 1 hour for full degradation of RNA. PEG precipitation was then performed with 2.5 M NaCl 20% PEG8000. An equal volume of PEG solution was added to the tube, and the mixture was allowed to incubate on ice for 15 minutes. Following that, the tube was spun at maximum speed for 5 minutes to pellet the DNA. The pellet was washed with 500 µL of 70% ethanol and spun again at maximum speed for 5 minutes. After centrifugation, the ethanol was poured off, and the pellet was allowed to dry. The pellet was resuspended in 30 µL of TE, incubated at room temperature for 20 minutes, and its concentration was then read on a Qubit fluorometer. The transformants were screened via Sanger sequencing.

6. Alkaline Lysis Midiprep

The procedure for an alkaline lysis midiprep follows a similar protocol as the alkaline lysis miniprep to purify plasmids but involves larger quantities. For the following protocol, the NucleoBond® Xtra Midi kit Takara Bio was used and modified. A starter culture of 3-5 mL containing LB medium and ampicillin at a final antibiotic concentration of 100 µg/mL was prepared and inoculated with 1 µL of bacteria known to contain the appropriate plasmid. It was allowed to grow in a shaking incubator at 37°C for 8 hours. An overnight culture in 100 mL LB medium and ampicillin (100 µg/mL) was inoculated with 100 µL of starter culture. The following day, the cells were pelleted in 50 mL conicals at 3000 g in a centrifuge for 10 minutes, and the cell pellet was resuspended in 8 mL resuspension buffer (RES) by pipetting. Then, 8 mL lysis buffer (LYS) was added to the cell suspension, and the mixture was inverted several times. The mixture was incubated at room temperature for up to 5 minutes. A NucleoBond® Xtra

column with an inserted column filter was equilibrated with equilibration buffer (EQU). The buffer was applied to the rim of the column filter and allowed to empty by gravity flow. Following incubation, 8 mL neutralization buffer (NEU) was added to the suspension mix by inverting until the mix became colorless. The lysate suspension was loaded on the column filter, and the column was again emptied by gravity flow. The column filter and column were washed with 5 mL EQU buffer, which was applied to the rim of the filter to ensure that the remaining lysate was washed. The filter in the column was removed, and the column was washed again with 8 mL buffer WASH. The plasmid sample was eluted with 5 mL elution buffer (ELU), and the elution was collected in a 50 mL conical (preheating the ELU buffer to 50°C may improve yield). The elution was split into Eppendorf tubes into 1 mL aliquots, and 400 µL of room-temperature isopropanol was added. The mixture was centrifuged at 15,000 g for 15 minutes at 4°C. Following centrifugation, the supernatant was discarded, and 500 µL of room-temperature 70% ethanol was added to the pellet. It was centrifuged at 15,000 g for 5 minutes at room temperature, and the supernatant was removed completely. The pellet was allowed to dry at room temperature, then reconstituted in 100 µL of TE. The TE from the first pellet was transferred to the next tube to reconstitute the second pellet, and the process was continued for all the remaining tubes. A sample was sent for Sanger sequencing to confirm that the vector contained sgRNA insert.

7. Diagnostic Restriction Enzyme Digestion

To confirm that the sgRNA was successfully cloned into the pU6-BbsI-chiRNA vector prior to Sanger sequencing, a diagnostic restriction enzyme digestion was performed. Controls, including an undigested vector and a digested vector, were run alongside individual clones on a gel to verify the expected plasmid vector size with the correct sgRNA insert. A successful clone should be resistant to subsequent BbsI digestion, as BbsI cleaves away from its recognition site and removes them in the process. Therefore, it should appear as a supercoiled plasmid following gel electrophoresis. In contrast, an unsuccessful clone should migrate like a linear plasmid because either the sgRNA insert was not cloned or the vector was not digested during the first

initial round of digestion. This assay provides an additional layer of confirmation before proceeding with Sanger sequencing to ensure that the correct construct had been obtained.

Transfections of Drosophila S2 Cells and Genomic DNA Extraction

1. Transfections

Plasmid DNA was transfected into Drosophila S2R+ derivative PT5 cells (NPT005), obtained from the Perrimon lab (10.7554/eLife.36333), using the Effectene Transfection Reagent (Qiagen) following the manufacturer's recommended protocol. Cells were grown until just confluent on T75 flasks in TNM-FH insect medium (Sigma-Aldrich) for 2-3 days, then plated at 3 × 10^6 per well into 12-well dishes. Transfections were performed using 1 μg of sgRNA expression plasmid per well, using a transfection mix containing 1 μg DNA, 8 μL Enhancer, and 4 μL Effectene Transfection Reagent in a final volume of 200 μL. The transfection mix was added to the wells in a dropwise fashion, and the plate was gently swirled. It was placed into a 27°C incubator for a 5-day incubation.

2. Precipitating Genomic DNA

The cells were resuspended in Eppendorf tubes, with each tube corresponding to one well plate. Approximately half of the media with the cells were removed from each tube. 500 μL of fresh media was added back to each tube. The cells were then pelleted at 3,000 RPM in a centrifuge for 3 minutes and resuspended in 500 μL of lysis buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 7.4, 2 mM EDTA pH 8.0, 0.2% w/v SDS, and 200 ng/mL proteinase K. Following that, 500 μL of isopropanol was added to each tube and mixed by inverting. The tubes were centrifuged at 14,000 RPM for 10 minutes at 4°C for 10 minutes, and the supernatant was poured off. The pellet was dried, resuspended in 50 μL of TE, and placed in a shaking incubator at 56°C for 20 minutes. The concentration was measured by NanoDrop (ng/μL). PCRs were performed in 25 μL reaction volumes per transfection. To purify, the Monarch® PCR & DNA Cleanup Kit was used, following the manufacturer’s instructions. The samples were eluted in 20 μL of elution buffer and sent for Sanger sequencing.

Injecting Plasmids into Drosophila Embryos

1. Preparing Injection Mix

To prepare, a 0.1-M sodium phosphate (pH 6.8) solution was made using 0.1 M NaH2PO4 (monobasic) and 0.1 M Na2HPO4 (dibasic). A 0.1-M potassium chloride solution was also made, and these two solutions were combined to create a 10-mL stock solution containing 1 mM sodium phosphate and 50 mM potassium chloride. The stock solution was filter sterilized using a 0.22-μm filter. The plasmids were diluted to a known concentration in the injection buffer to create the final injection mix, and several concentrations were tested. The injection mix was stored in 10 μL aliquots at -20°C and was thawed immediately before use.

2. Embryo Collection

Two days prior to injection, approximately 50-100 newly eclosed flies of both sexes were sorted into a bottle containing standard cornmeal agar supplemented with a thin layer of active yeast pellets to promote egg-laying. On the following day, fresh yeast paste was prepared and a grape plate was warmed to 25°C. An approximately 1 cm diameter dab of yeast paste was placed in the center of the grape plate, and the flies were transferred to an embryo collection cage along with the grape plate with yeast paste. The cage was wrapped tightly with a rubber band to secure the plate and placed on its side at 25°C overnight. On the third day, the old grape agar plate was replaced with a new plate containing fresh yeast paste, and the cage was again covered with the grape plate. The cage was wrapped with tin foil (to mimic darkness, as female flies generally lay the most during dawn and dusk) and placed on its side for 1 hour at 25°C. The plate-changing process was repeated as needed for additional collections, and the removed grape plate was saved for embryo collection and dechorionation in preparation for injections. Grape plates were rinsed and reused. Flies were typically collected for up to 5 days before being frozen overnight and disposed of.

3. Embryo Dechorionation

A mesh screen was positioned over the top of a 50 mL conical tube and a hole was cut in the middle of the conical cap. The cap was then screwed onto the conical tube to secure the mesh

screen. Using a spray bottle filled with ddH2O, water was gently dispensed onto the grape plate, and the bristles of a paintbrush were used to separate the embryos from the grape agar. The yeast paste was broken up to recover any embryos that were laid within. The embryos floating in the water were carefully poured into the conical tube and onto the screen. This process was repeated as necessary until all embryos were in the conical tube. DdH2O water from the faucet was used to gently wash the embryos in the conical tube, allowing all yeast paste and residual grape agar to flow through the screen until only embryos remained. Embryos were submerged and agitated for 1 minute in a weighing dish containing bleach diluted to 50% concentration in water. After 1 minute, the embryos were removed from the bleach and rinsed for at least 2 minutes using ddH2O from the faucet. The embryos were washed with embryo wash solution by submerging the bottom of the conical in a weighing dish containing 1x embryo wash. They were agitated for 4 minutes. Using the spray bottle of ddH2O, the walls of the conical tube were gently sprayed for 2 minutes to collect the embryos in the center of the screen and placed aside.

A rectangular piece of grape agar was cut out from the grate plate and placed on a microscope slide. Embryos were then picked up using the mesh screen and gently pressed onto the grape agar. Embryos older than 2 hours were not used. The embryos were aligned horizontally in a single file on the grape agar using a metal-tipped probe, then transferred to a glass slide using double-sided tape and placed on a coverslip. The coverslip with the embryos was then placed in a desiccating chamber for 1-2 minutes to dry the embryos before injection.

4. Preparing Injection Needles

To prepare for microinjection, Sutter Instrument Co. Model P-87 Flaming/Brown Micropipette Puller was used to prep the needles with the following settings of heat: 848; pull: 125; velocity: 70; time: 250 ms; pressure: 200. After pulling, the injection needles were cut with a razor blade such that it would have narrow bore wide enough to aspirate the injection mix but not so wide that embryos would rupture. To load the needles, 1-2 µL of injection mix was drawn into the needle using a 2-µL pipettor with a long pipette tip.

5. Embryo Injection

After drying in the desiccating chamber, the embryos were covered with a thin coat of halocarbon oil 700. The loaded needle was placed into the aspirator apparatus and the slide onto the microscope stage. The exterior edge of the embryo was aligned on the same plane as the injection needle. The tip of each embryo was punctured with the needle, the plasmid mix was injected using a Picospritzer, and the needle was immediately removed from the embryo. After completion of the injections, the coverslip containing the embryos was placed into a small covered container (e.g. empty Petri dish) and incubated overnight at 25°C to allow for cutting of the construct and subsequent repair. The needles could be saved if they were free of clogs and handled carefully by removing them from the injection apparatus and placing them in a protected container at -20°C. After incubation, DNA could be immediately extracted, or the embryos could be stored at -20°C for up to 2 weeks. The procedure for preparing and injecting the embryos, recovering plasmids and enriching them for inaccurate repair, and analyzing reads from amplicon sequencing follows a protocol that has been published by our laboratory (Hanscom *et al*., 2018).

6. DNA Extraction From Injected Embryos Using Phenol-Chloroform

After incubation, the embryos were scraped off from the microscope slide with a fresh razor, and a pestle was used to transfer them to an Eppendorf tube along with 200 µL of P1 solution which is comprised of 50 mM Tris-HCl pH 7.5 and 10 mM EDTA. It is recommended to use 100 µL of P1 solution first to mash the embryos, then another 100 µL to wash the remaining embryos off of the pestle. Next, 600 µL of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the tube, and the tube was inverted 10 times to mix after sealing with parafilm. The tube was then centrifuged for 1 minute at maximum speed (15,000 RPM), and the aqueous layer was collected and transferred to a new Eppendorf tube. An equal volume of phenol-chloroform:isoamyl alcohol (25:24:1) was added, sealed with parafilm, and inverted several times before centrifuging for 1 minute at maximum speed. The aqueous layer was collected again. Then, 1/5th volume of 8M potassium acetate (KAc) was added, and the tube was inverted to mix. 1 volume (including the KAc) of chloroform was added, sealed with parafilm, and inverted several times. The tube was centrifuged again for 1 minute at maximum speed, and the

aqueous layer was collected once more. To the collected aqueous layer, 0.7 volumes of isopropyl alcohol were added and inverted to mix, then incubated at -80°C for 30 minutes. Following incubation, the tube was centrifuged for 15 minutes at maximum speed, and the supernatant was poured off. The pellet was then washed with 200 µL of fresh 70% ethanol and centrifuged for 5 minutes at maximum speed before pouring off the supernatant. Finally, the pellet was dried for 5 minutes with the Eppendorf cap off, and 30 µL of TE was added as the final elution volume.

Results

Analysis of *In Vitro* Cutting Efficiency of sgRNAs for HTGTS

Highly transcribed gene regions in the fly genome were selected for Cas9 cutting, as they have been reported to provide better Cas9 binding and editing efficiencies in eukaryotic cells, owing to their euchromatic nature (Verkuijl & Rots, 2019) (Figure 8). To minimize the lethality resulting from Cas9 cleavage and translocation formation, only non-essential genes were chosen. A CRISPR/Cas9 target online predictor tool, CCTop, was used to evaluate the potential cutting efficiency of each candidate sgRNA. Only sgRNAs that were predicted to have medium or high cutting efficiency were chosen, based on the tool's algorithm. Initially, 18 potential sgRNAs were nominated for six genomic loci including Dci, Mtch, Cypl, CC2218, Tpi, and Vps16B (Table 1). For each genomic locus, there were three corresponding sgRNAs.

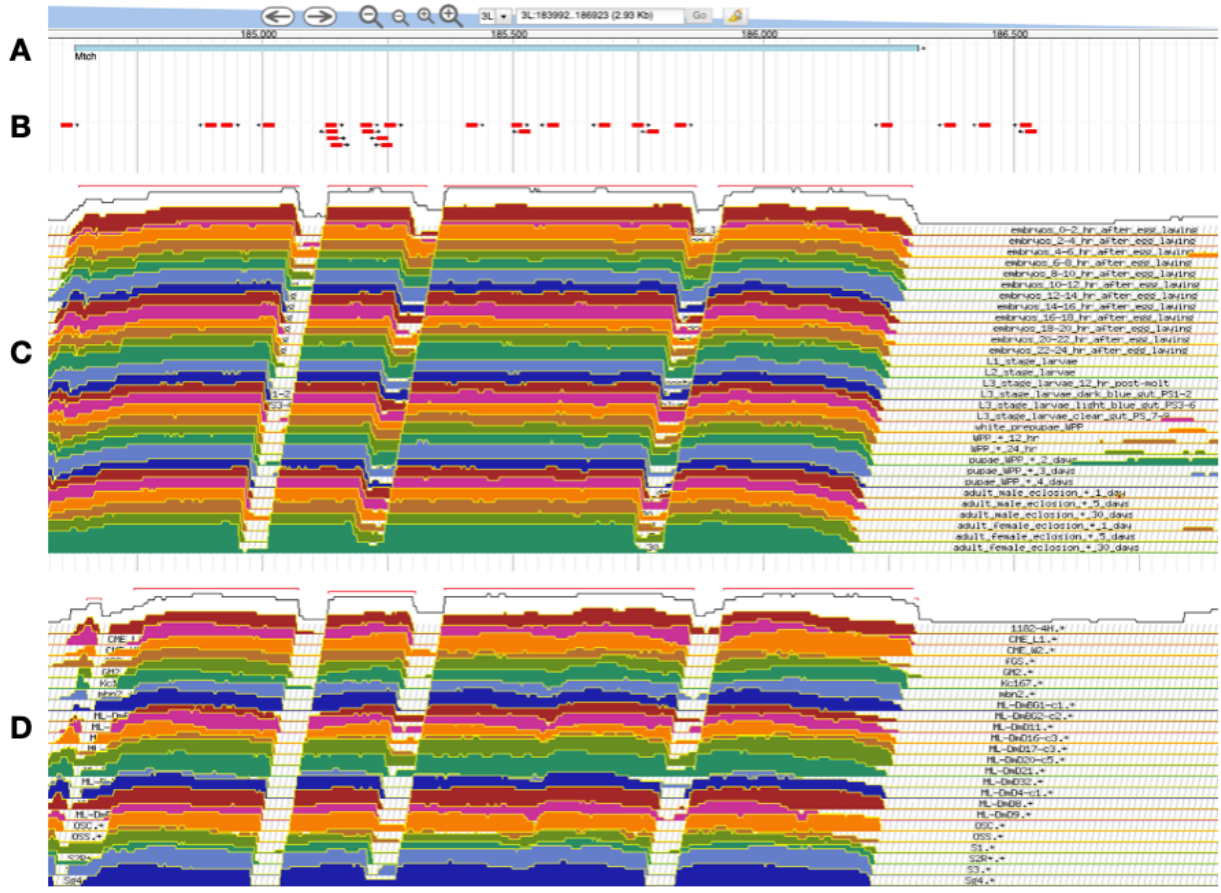

Figure 8. A genomic map on FlyBase of the (A) Mtch gene span. (B) Sequences predicted to be suitable as sgRNAs for genome editing which was filtered on Flybase to have many off-targets and low stringency. (C) RNA-Seq expression data at different stages of development in whole organism Drosophila. (D) RNA-Seq expression data for various Drosophila cell lines.

Before administering the sgRNAs to the S2 cells, they were validated for their cutting efficiency *in vitro* using an *in vitro* cutting assay. The efficiency of each sgRNA was evaluated by analyzing the relative amount of cleavage products generated from the cut genomic locus using gel electrophoresis (Figures 9 and 10). SgRNAs with lower cutting efficiency displayed lighter or absent bands such as in lane 7 of Figure 10 where the resulting cleavage products exhibited low visibility while those with higher cutting efficiency resulted in darker bands, with two distinct fragments being clearly visible. Guides with low efficiency were eliminated from further analysis. Following the *in vitro* cutting assay, we identified a total of nine sgRNAs (not shown) targeting five genomic loci of the third fly chromosome (Dci, Mtch, Cypl, Tpi, and Vps16B) that

exhibited high efficiency for inducing DSBs *in vitro*. CC2218 was eliminated from the list of targeted genomic loci due to the low cutting efficiency of its sgRNAs.

Name (guide number_gene region target)	**Sequence (5' → 3') PAM**	**Efficiency of sgRNA as predicted by CCTop**
sgRNA6_Dci	AAGAGATTAACCTATGGTTA TGG	0.65 medium
sgRNA13_Dci	GATATGAATGCTAATTCTTG TGG	0.78 high
sgRNA22_Dci	TATCTCTTGGGTTTTAAAAT AGG	0.71 medium
sgRNA27_Mtch	CAGCTCCCTTTTCTGGCGGT GGG	0.67 medium
sgRNA28_Mtch	AGTTTAGATCTACATTGAAG TGG	0.74 high
sgRNA33_Mtch	CTTAGGCAAGGGCGCGAATG AGG	0.75 high
sgRNA10_Cypl	GGGCGCGGGCATCCTCTCGA TGG	0.76 high
sgRNA12_Cypl	CGCGAACTCCGATCCGTAGA TGG	0.64 medium
sgRNA30_Cypl	GGCGGCGATCCCACGGGCAC TGG	0.75 high
sgRNA13_CC2218	GGGGAGAACCAAACTCATTG CGG	0.76 high
sgRNA35_CC2218	TTGCGGCAATCGACGGAAGG CGG	0.81 high
sgRNA44_CC2218	GCTGGCGGGCGGAGCACTTG AGG	0.85 high
sgRNA29_Tpi	CGATCTCGGCGATGGACTTC TGG	0.62 medium
sgRNA56_Tpi	GGATCGGGGGACAAAACAAG GGG	0.64 medium
sgRNA65_Tpi	CATGTTGTTGCTGTTGCTGT CGG	0.76 high
sgRNA12_Vps16B	GGTGGATGCCTATGCCACAGC GGG	0.80 high
sgRNA15_Vps16B	ACTGCACAGAAGCTGCTGCA AGG	0.69 medium
sgRNA20_Vps16B	GCATGAGAAGCAGCAGCTGC TGG	0.67 medium

Table 1. Predicted efficiencies of sgRNAs targeting various regions of Drosophila gene regions including Dci, Mtch, Cycle, CC2218, Tpi, and Vps16B, as determined by the CRISPR prediction tool CCTop. The predicted efficiencies of the sgRNAs range from 0 (poor) to 1 (high), with scores classified as either medium or high.

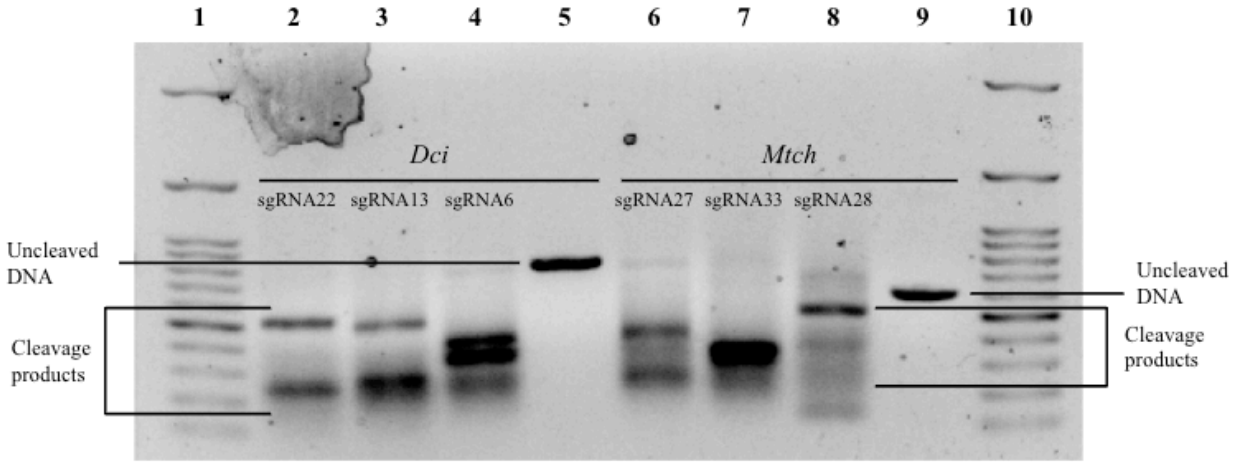

Figure 9. Gel electrophoresis (1% agarose with ethidium bromide) showing the results of an *in vitro* cutting assay to validate the cutting efficiency of sgRNAs in genomic loci Dci and Mtch. Lanes 1 and 10 are marker ladders (1+ kb DNA ladder). Lanes 2-4 show DNA fragments of the expected sizes of the Dci locus after incubation with sgRNA and Cas9. Lane 5 is uncut Dci template. Lanes 6-8 show DNA fragments of the expected sizes of Mtch locus after Cas9 RNP incubation. Lane 9 is uncut Mtch template.

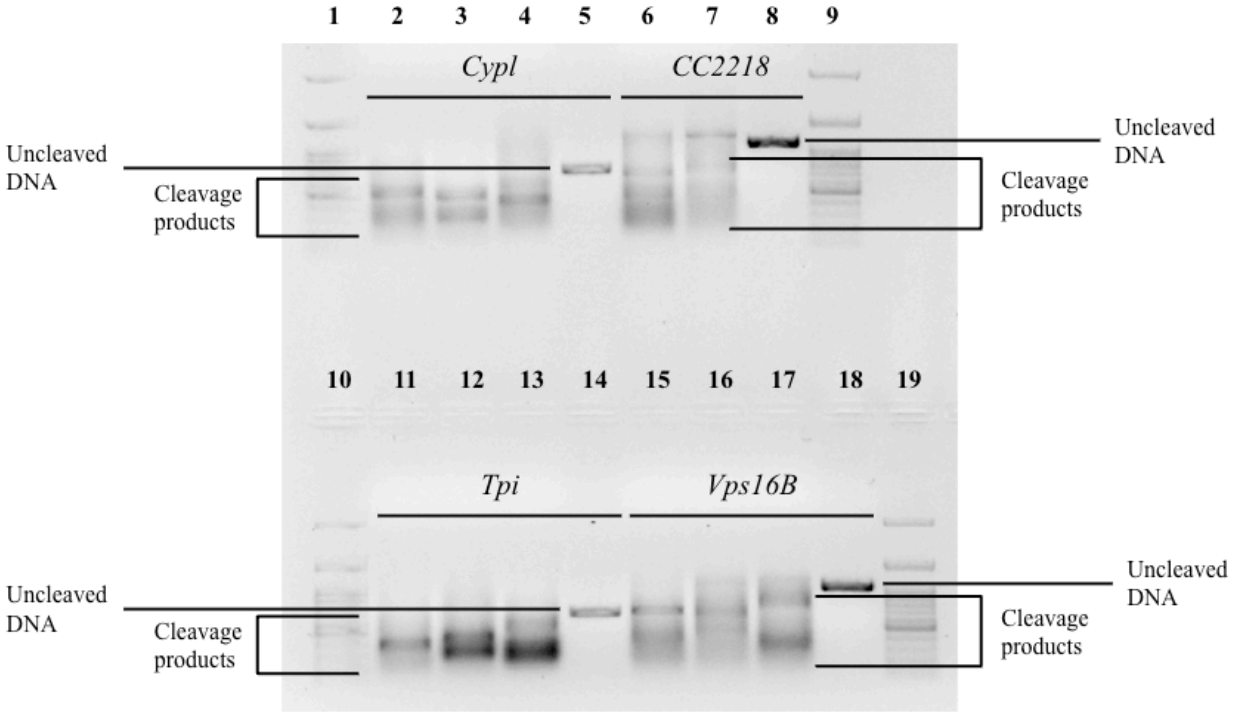

Figure 10. Gel electrophoresis (1% agarose with ethidium bromide) showing the results of an *in vitro* cutting assay to validate the cutting efficiency of sgRNAs in genomic loci Cypl, CC2218, Tpi, and Vps16B. Lanes 1, 9, 10, and 19 are marker ladders (1+ kb DNA ladder). Lanes 2-4 (sgRNA30, sgRNA12, and sgRNA10, respectively) show DNA fragments of the

expected sizes of the Cypl locus after incubation with sgRNA and Cas9. Lane 5 is uncut Cypl template. Lanes 6-7 (sgRNA35 and sgRNA13, respectively) show DNA fragments of the expected sizes of CC2218 locus after incubation with Cas9 RNP. Lane 8 is uncut CC2218 template. Lanes 11-13 (sgRNA56, sgRNA65, and sgRNA29, respectively) show DNA fragments of the expected sizes of Tpi locus after incubation with Cas9 RNP. Lane 14 is uncut Tpi template. Lanes 15-17 (sgRNA15, sgRNA20, and sgRNA12, respectively) show DNA fragments of the expected sizes of Vps16B locus after incubation with Cas9 RNP. Lane 18 is uncut Vps16B template.

Generation of High-Efficiency sgRNA Constructs in a Plasmid Vector

To construct the plasmid expression vector for the delivery of sgRNAs in S2 cells, the BbsI restriction enzyme was utilized to clone the nine selected sgRNAs that exhibited high-efficiency cutting into the pU6-BbsI-chiRNA vector (Figure 11). A unique feature of this cloning system is that BbsI generates 4-nucleotide non-complementary overhangs which facilitate unidirectional insertion of the sgRNA insert into the vector. In addition, the recognition site of BbsI is destroyed following initial cleavage, thereby preventing the plasmid from recircularizing without an insert containing the correct sticky ends (Figure 12).

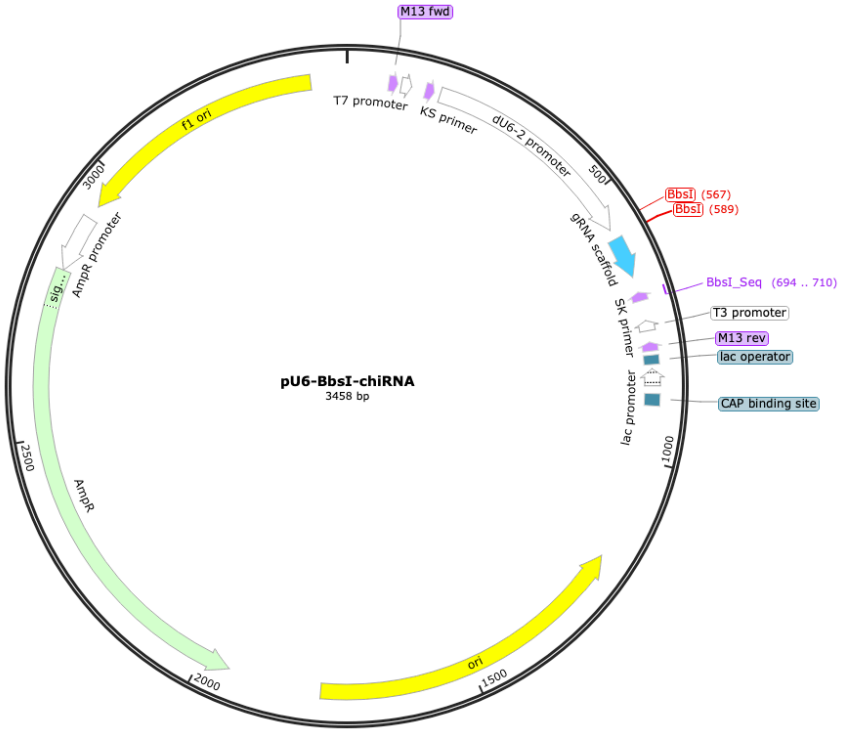

Figure 11. Plasmid map of the pU6-BbsI-chiRNA plasmid vector, showing the BbsI restriction enzyme site highlighted in red. The presence of the BbsI cut sites indicates that a sgRNA has not been cloned yet.

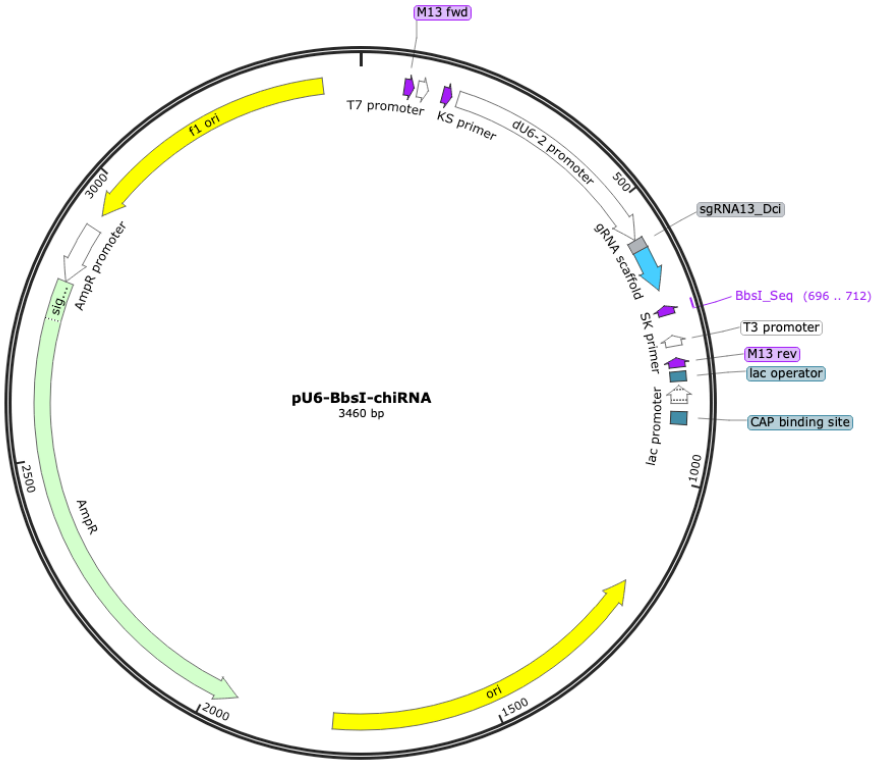

Figure 12. Plasmid map of the pU6-BbsI-chiRNA plasmid vector containing the sgRNA13_Dci insert, highlighted in grey. The absence of the BbsI recognition sites indicates that the sgRNA had been successfully cloned in.

On a few occasions, a diagnostic restriction enzyme digestion assay was performed to verify that the sgRNA was inserted correctly. After transforming *E. coli* with the pU6-BbsI-chiRNA vector containing the sgRNA insert, an alkaline miniprep procedure was used to isolate the plasmid. Following gel electrophoresis of the miniprepped samples, the digestion pattern of the plasmid with an insert and the undigested parent plasmid control should be identical, given that the BbsI recognition site is destroyed in the process of digestion. If, however, a plasmid that is suspected to have a clone is linearized during the digestion assay, this indicates that the insert did not ligate properly to the sticky ends, or that the initial preparation of the plasmid vector was unsuccessful. This is evident in lanes 6 and 7 in Figure 13. The resulting plasmids that had been confirmed via the digestion assay and/or Sanger sequencing were suitable for transfection into the S2 cells for our HTGTS studies.

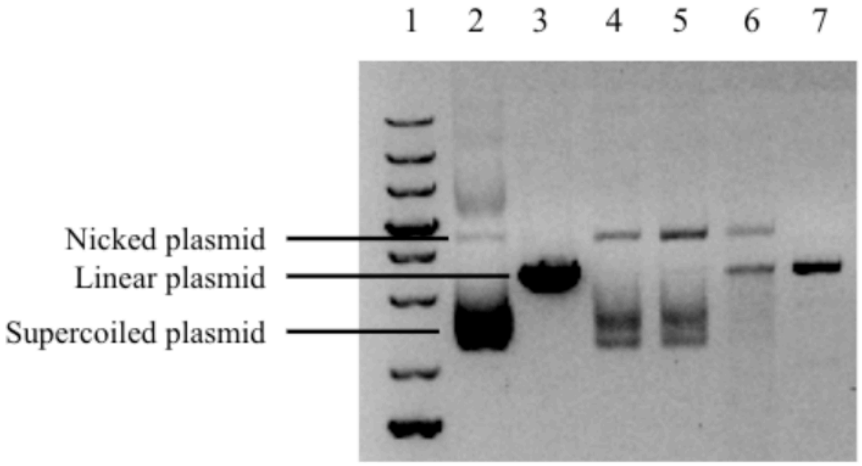

Figure 13. Gel electrophoresis (1% agarose with ethidium bromide) of diagnostic restriction enzyme digestion assay. Lane 1 represents the marker ladder (1+ kb DNA ladder). Lanes 2 and 3 are undigested and digested pU6-BbsI-chiRNA parent plasmid, respectively. Lanes 4 and 5 are miniprepped plasmids with the sgRNA13_Dci insert, incubated with and without BbsI, respectively. Because the products are running like supercoiled plasmids, cloning was successful. Lanes 6 and 7 are miniprepped plasmids with the sgRNA28_Mtch insert, incubated with and without BbsI, respectively. In this case, the linear migration pattern of the products indicates that the insert was not successfully cloned in.

Assessing Cutting Efficiencies of Dci sgRNA13 and Tpi sgRNA56 in S2 Cells Using Sanger Sequencing Traces

To test the cutting efficiencies of sgRNAs in S2 cells, two guides—Dci sgRNA13 and Tpi sgRNA56—that have previously demonstrated high efficacy for Cas9 cleavage *in vitro* were transfected into the cells. Following a 5-day incubation period for cutting and repair, genomic DNA was extracted via isopropanol precipitation. The cutting efficiency of a sgRNA can be inferred from Sanger sequence traces surrounding the Cas9-induced break from purified PCR products. In high-efficiency sgRNAs, the trace should be messy with no discernible peaks immediately after the cut site, indicating the presence of indels resulting from inaccurate repair. (Figure 14A). If sgRNAs do not exhibit these characteristics and instead have clear traces with identifiable peaks downstream, this suggests that cleavage most likely did not occur because the "bait" locus is still intact (Figure 14B).

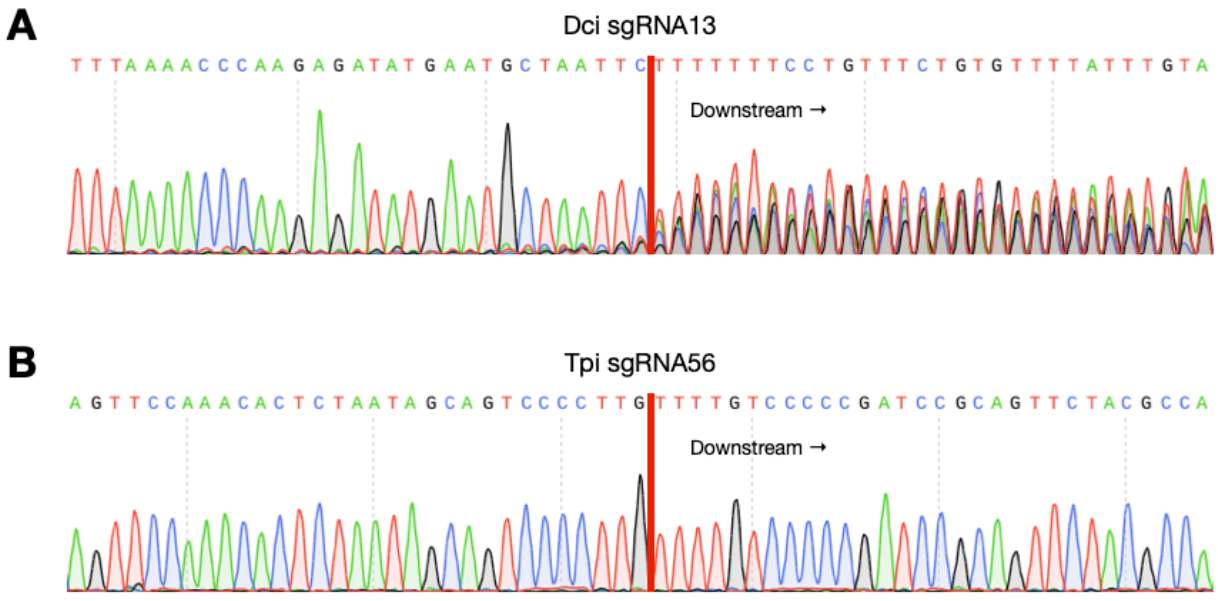

Figure 14. Sanger sequencing traces following transfections into S2 cells for (A) Dci sgRNA13 exhibits a high level of noise, indicating efficient cutting at the targeted site, whereas Tpi sgRNA56 exhibits a clean trace, suggesting little to no cleavage at the site. The vertical red line in each trace denotes the expected Cas9 cut site.

Assessing Efficiencies of Dci sgRNA13 and Mtch sgRNA27 in Early Drosophila Embryos

The potential of inducing DSBs and translocations in the genomes of early Drosophila embryos was explored. Compared to the 5-day timeline required for transfection and recovery of DNA products in S2 cells, the use of embryos for testing allows for a faster experimental timeline, usually within 24 hours. Two sgRNAs (Dci sgRNA13 and Mtch sgRNA27), which were previously confirmed to have high cutting efficiency *in vitro*, were utilized for embryo injections. To this end, a preliminary experiment was conducted by injecting sgRNA expression plasmids into early-stage embryos, followed by an overnight incubation for cutting and repair. Genomic DNA was recovered, and PCR was used to amplify repair junctions. Gel electrophoresis analysis (not shown) was performed on the PCR products to ensure that the genomic DNA extracted from the embryos was of good quality and had not undergone significant degradation during incubation. The presence of a single distinguishable band corresponding to the expected amplicon size was indicative of genomic DNA that had not been degraded. Both Dci sgRNA13 and Mtch sgRNA28 exhibited this. In contrast, a smear or absence of a band would indicate fewer intact templates. Unique primers specific to each genomic locus

were used for sequencing. The sequencing traces indicated low levels of cutting in both guides (Figure 15), suggesting that the embryo approach is unlikely to be successful. However, this experiment provided insight into the feasibility of applying HTGTS in future *in vivo* investigations, while also allowing for comparison with the 5-day transfection timeline.

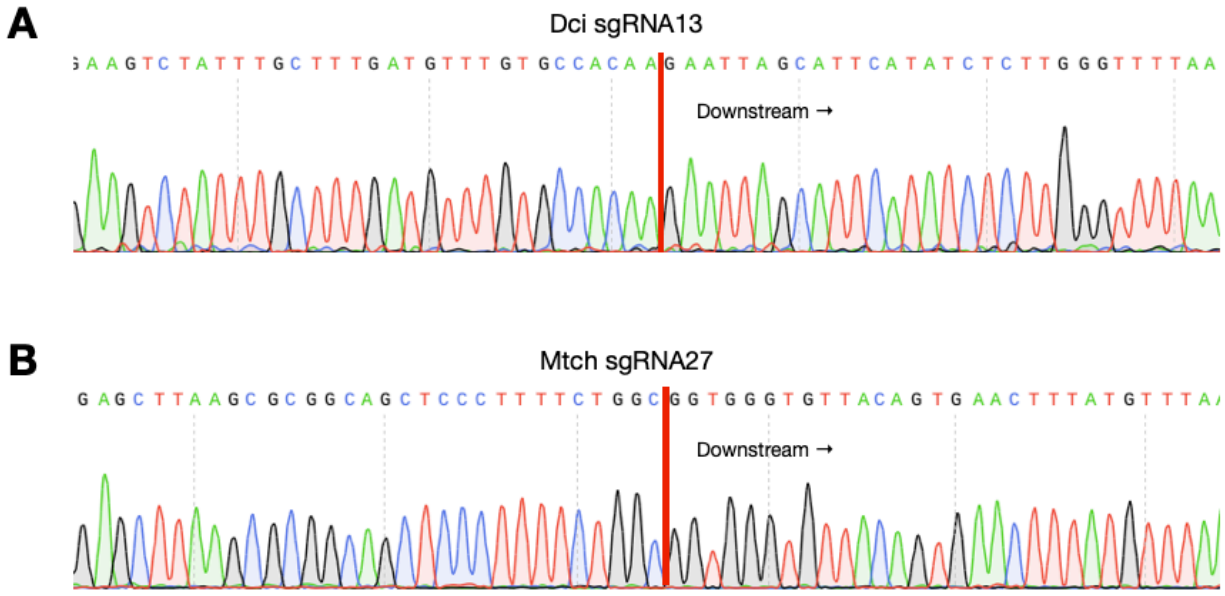

Figure 15. Sanger sequencing traces following embryo injections for (A) Dci sgRNA13 and (B) Mtch sgRNA27, both of which had no noise downstream from the expected cut site. This suggests that there was little to no cleavage.

Detection of Inversion and Deletion Events in Iw7 Plasmid Via PCR Assay

Due to delays in the S2 culture system, a PCR assay was developed as an alternative to HTGTS for investigating chromosomal rearrangements in a plasmid system. The study of such events in Drosophila, including inversions and deletions, involved the design and cloning of a second sgRNA (sgRNA2) into the pU6-BbsI-chiRNA plasmid expression vector, targeting a site approximately 100 bp downstream from the first sgRNA (sgRNA5) (Figure 16A). The cutting efficiency of sgRNA2 was validated using an *in vitro* cutting assay (not shown). It should be noted that these guides were intended to target and induce breaks in the Iw7 plasmid, which would subsequently undergo repair in embryos, following a similar workflow as depicted in Figure 4.

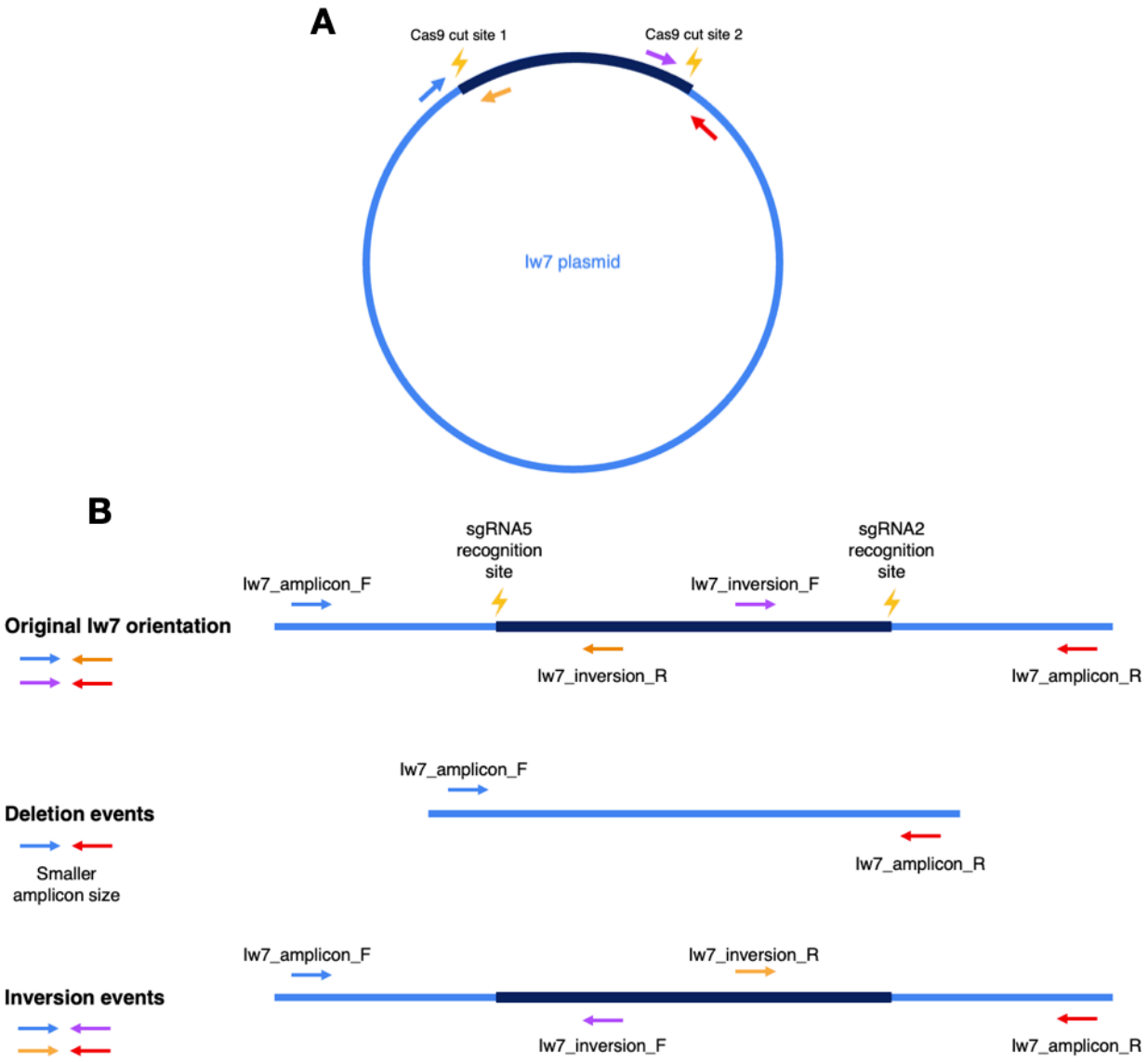

Figure 16. (A) Schematic of Iw7 plasmid with two Cas9 cut sites separated by approximately 100 bp (not drawn to scale). (B) Intra-plasmid events, such as deletion and inversion events, can occur between these cut sites. The amplicon size for Iw7 is 545 bp which can be generated using the Iw7_amplicon_F + Iw7_amplicon_R primers. Detection of deletion events can be achieved by amplifying the Iw7 amplicon, followed by gel electrophoresis to confirm a smaller amplicon size compared to the original Iw7 amplicon. To identify inversion events, two sets of primer pairs can be used: 1) Iw7_amplicon_F + Iw7_inversion_F (which replaces Iw7_inversion_R as the reverse primer), or 2) Iw7_amplicon_R + Iw7_inversion_R (which replaces Iw7_inversion_F as the forward primer).

In order to assess the occurrence of inversion and deletion events following Cas9 cleavage at the two cut sites in Iw7, PCR and gel electrophoresis analysis were conducted using

unique primer pairs to quantify the extent of such events (Figure 17). Deletion events should be easily distinguished as the interstitial fragment is lost, resulting in a PCR product that is approximately 100 bp shorter than the original Iw7 amplicon, which can be visualized on a gel. However, no such deletion events were observed in lane 2, as evidenced by the absence of a smaller product. As expected, the Iw7 amplicon band observed in lane 2, which served as the positive control, had high intensity. Similar high intensities were observed in lanes 3 and 4, which also served as positive controls in the assay. In contrast, potential inversion products in lanes 5 and 6, which were expected to occur at a much lower frequency than the controls, exhibited fainter band intensities on the gel. It was inferred that these were such events since the product ran at the expected size, approximately 200-250 bp. Gel band intensities were assessed using ImageLab software (Table 2) to provide a semi-quantitative estimation of the DNA content in each band, allowing for a comparison of potential inversion events with the control Iw7 amplicon. The band intensities observed on the gel were consistent with the results obtained from the ImageLab software analysis.

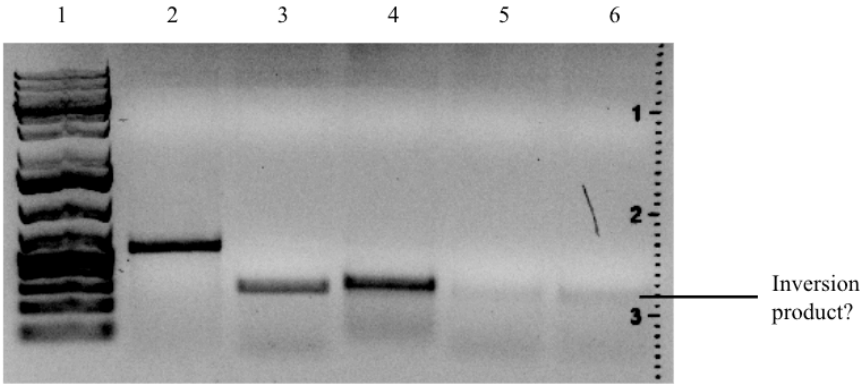

Figure 17. Gel electrophoresis (1% agarose with ethidium bromide) of PCR assay. Lane 1 represents the marker ladder (1+ kb DNA ladder). Lane 2 is the Iw7 amplicon as a positive control, and no deletion events were detected. Lanes 3 and 4 also are positive controls using Iw7_amplicon_F + Iw7_inversion_R and Iw7_amplicon_R + Iw7_inversion_F primers, respectively. Lanes 5 and 6 represent potential inversion products by using primer pairs Iw7_amplicon_F + Iw7_inversion_F and Iw7_amplicon_R + Iw7_inversion_R, respectively. The product size for inversion events is approximately 200-250 bp which is consistent with the gel.

Lane	Adjusted Volume (Intensity)	Software Relative Intensity
2	23553240	1

3	11568585	0.491167
4	16695701	0.708849
5	2160600	0.091733
6	2466751	0.104731

Table 2: Normalized gel band intensities from PCR assay using ImageLab, with amplicon control in Lane 2 as reference.

Discussion

Chromosomal translocations are hallmarks of many malignancies and diseases, particularly in cancer, and often lead to the activation of oncogenes or inactivation of tumor suppressor genes. These events require at least two simultaneous DSBs in separate chromosomes and can either be generated spontaneously or induced by exogenous exposure like ionizing radiation or chemical agents. While much is known about the products of translocations, the underlying mechanisms of their occurrence remain poorly understood. Therefore, a better understanding of these mechanisms is crucial for identifying potential therapeutic targets. Techniques such as HTGTS have allowed for the sensitive detection of DSBs across the genome, making it possible to investigate the genetic requirements and preferences for translocation formation. Researchers have identified several DNA repair pathways that are implicated in the translocation formation including c-NHEJ and alt-EJ, although the literature has yielded diverging findings. For instance, Ku deficiency decreases the frequency of translocations in human cell lines but has the opposite effect in mouse cell lines (Ghezraoui et al., 2014; Boboila et al., 2010). However, as noted earlier, a consistent finding observed in the literature on translocations is the presence of 1-6 bp MH signatures at the junctions. Interestingly, POLQ is often associated with these MH signatures left at the junctions during translocation events, as it uses MH that is exposed after DNA resection to join broken DNA ends. This type of repair is dubbed TMEJ. Additionally, POLQ mRNA overexpression has been observed in various human cancers (Lemée et al, 2010).

Here, I have reported my progress in adopting the HTGTS workflow to induce DSBs as an approach to studying translocation formation in Drosophila. Nine sgRNAs have been

designed to target five specific genomic loci across the third fly chromosomes, and these guide constructs have been successfully cloned into the pU6-BbsI-chiRNA plasmid expression vector. Cloning has been confirmed through diagnostic restriction enzyme digestion assays and/or Sanger sequencing. To increase the likelihood of DSBs and therefore translocation formation, low-specificity sgRNAs were chosen to induce off-target cleavage by Cas9 leading to "prey" DSBs. Also, highly transcribed gene regions were selected to maximize the efficiency of the CRISPR/Cas9 system in inducing DSBs. Euchromatin is a loosely packed and less condensed form of chromatin and is found in regions of the genome that contain genes that are actively being transcribed. We predict that by targeting euchromatic regions, Cas9 is more readily able to bind and cleave the DNA. Non-essential genes were chosen to minimize the possibility of any harmful or lethal effects resulting from Cas9 cleavage and translocation formation.

We initially performed transfections of two guides, Dci sgRNA13 and Tpi sgRNA56, in S2 cells to evaluate their cutting efficiencies. The success of the transfections was assessed by analyzing Sanger sequencing traces. However, due to various factors such as time constraints, cell culture contamination, and the need for optimizing incubation time for the sufficient recovery of “bait-prey" junctions, we were unable to perform transfections for all nine sgRNAs. Additionally, we attempted to inject expression plasmids of Dci sgRNA13 and Mtch sgRNA27 into early-stage embryos. However, the Sanger sequencing traces obtained showed no indication of cutting, suggesting that the guides may not have been effective in inducing the desired cuts at the target site in embryos. Delivery of editing reagents, such as the Cas9 endonuclease, into embryos through injection is critical for successful genome modification. However, if the sgRNAs are unable to diffuse throughout the embryos, this results in a large number of uncut chromosomes that may be preferentially amplified in the amplicon PCRs. To further investigate the cutting efficiency of the remaining sgRNAs, additional transfections in S2 cells and injections in embryos will be performed. Subsequent analysis of the Sanger sequence traces would allow for an initial assessment of which sgRNAs are effective in inducing the desired cuts at the targeted site, as has been done in these previous experiments.

The Cas9/HTGTS system has been adapted to study chromosomal translocations in both Drosophila S2 cells and embryos. S2 cells have the advantage of generating larger libraries of

sequencing reads of "bait-prey" junctions. Additionally, these cells can be easily transfected, allowing for quicker introduction of plasmid DNA. However, cutting and repair require an incubation period of at least five days following transfection, but there are ongoing efforts to optimize this incubation time. S2 cells are also polyploid, meaning these cells have multiple sets of chromosomes and thus may not accurately reflect the DNA content in the organism itself. On the other hand, embryos offer the advantage of a faster experimental timeline, with cutting and repair occurring within 24 hours after injections, in addition to being diploid. Compared to S2 cells which are cultured *in vitro*, experiments conducted in embryos provide an opportunity to perform *in vivo*-like studies. Moreover, it may be easier to study translocation formation in different genetic backgrounds, such as POLQ null or Ligase IV-deficient, in embryos as these fly stocks are readily available in the laboratory. However, introducing plasmids with the guide constructs into embryos via microinjection may take longer than transfection in S2 cells due to the labor-intensive steps involved. These include preparing the flies for egg-laying, identifying and sorting embryos that are suitable for injection (ones that are less than two hours old), and prepping the embryos for injection.

In the case where there is evidence of successful cleavage, as seen in the Sanger sequencing traces of Dci sgRNA13 where there was substantial background signal observed downstream from the cut site (Figure 14A), high-throughput sequencing can be a robust method for characterizing the sequences surrounding the break site. However, this approach can be time-consuming. An alternative approach to identifying indels that arise within a “bait” junction following introduction of DSBs is to use Tracking of Indels by DEcomposition (TIDE) (Brinkman & van Steensel, 2019). In this approach, two parallel PCRs can be performed: one of the junction following transfection and another of the junction with no transfection as a negative control to use as a baseline. The PCR product of both the Cas9-targeted cells and the uncut population are Sanger sequenced, and the resulting traces can be analyzed using an online TIDE tool. The TIDE algorithm reconstructs the spectrum of indels from the sequencing traces which reports the identity of the detected indels and their frequencies. In practice, TIDE can be employed on both genomic DNA extracted from S2 cell transactions as well as embryo injections.

The PCR assay shown in Figure 17 was developed as an alternative to HTGTS for studying chromosomal rearrangements in Drosophila, including inversions and deletions, by modifying the plasmid-based system. Although potential inversion products were detected in the assay, further investigation is required to establish their identity as true inversion events. This experiment's main limitation is that efficient cleavage is necessary at both of the Cas9 cut sites for interstitial fragment formation and subsequent inversion. The absence of deletion events, which should involve fewer steps than inversions, is a notable finding that warrants further scrutiny regarding the accuracy of the data and implies that the inversion products visualized on the gel may not be the result of a complete inversion. For instance, the interstitial fragment may have been inverted but not repaired at both cut sites (Figure 18). It was also speculated that deletion events may have been occurring but were below the level of detection by ethidium bromide gel electrophoresis. While the Iw7 amplicon primers were used to detect deletion events, unique primer pairs specifically designed to detect inversions were used, which were only expected to amplify the junction in the event of an inversion. Also, given that the Iw7 template may have been present in greater quantities than the smaller amplicon produced through deletion events, the Iw7 amplicon may have been preferentially amplified.

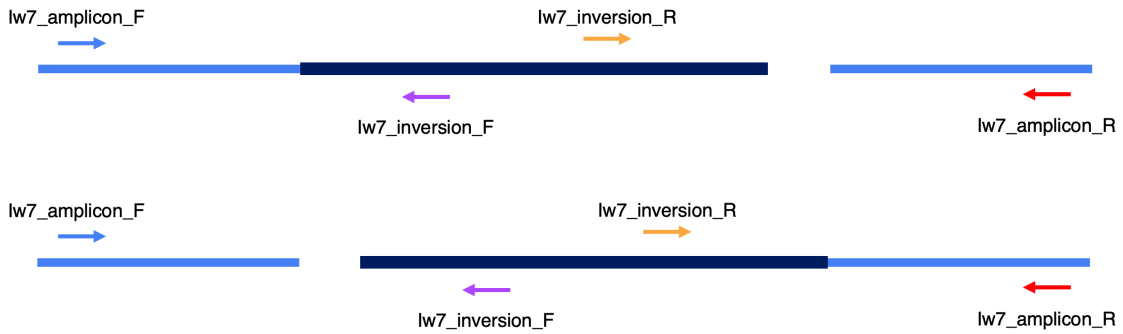

Figure 18. Schematic of partial inversion event. Inversions require Cas9 cleavage at both sites cute sites, followed by inversion of the resulting interstitial fragment, and repair at the original cut sites. It is possible that the product observed on the gel is the result of inversion and repair at one cut site only.

If the products visualized on the gel were indeed inversion events, restriction enzyme cloning can be utilized to estimate the number of unique junctions present. This technique involves amplifying these junctions using primers that have restriction sites added to the 5' end

which would create sticky ends generated by a restriction enzyme digest. Subsequently, the resulting fragment can be ligated into a cloning vector containing complementary sticky ends. The recombinant DNA can be transformed into bacteria and plated, with each colony representing a unique junction that can be isolated and sent for Sanger sequencing to identify repair events. However, it is important to reiterate that these unique repair junctions may not necessarily represent complete inversion events and could instead be partial inversions.

Future Directions

1. Are the translocation junctions consistent, if any, with the synthesis-dependent microhomology-mediated end-joining (SD-MMEJ) model?

In 2010, the McVey laboratory proposed a new type of alt-EJ, termed SD-MMEJ, which specifically forms secondary structures such as loops or hairpins via the annealing of complementary MHs (Figure 19). Following the formation of these secondary structures, nascent DNA is synthesized and new MHs may arise as a result. SD-MMEJ can generate three types of repair junctions including apparent blunt joins (ABJs), microhomology joins (MHJs), and insertions/deletions (indels). ABJs are formed by the blunt joining of two ends during repair, while MHJs are repair products that exhibit 1 or more bp of MHs at the junction. Indels refer to repair products that involve deletions coupled with insertions of 1-10 base pairs. Through bioinformatic analyses, this model can predict the repair mechanism depending on the neighboring DNA sequence, including how far the sequence is from the break site. The McVey laboratory has developed a custom pipeline for analyzing next-generation sequencing reads for SD-MMEJ-consistent repair products. The aim is to append this existing pipeline to the HTGTS pipeline to further examine translocation junctions for their consistency with SD-MMEJ.

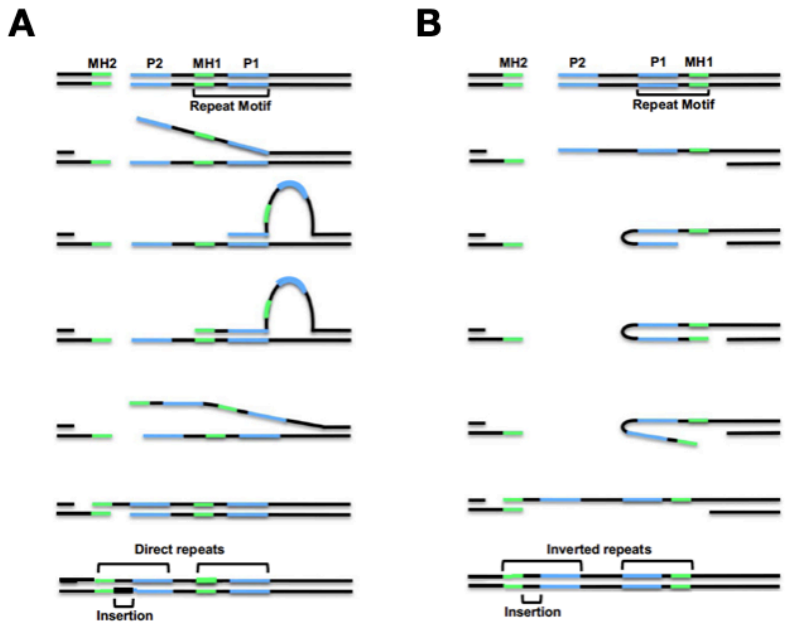

Figure 19. Model of SD-MMEJ. (A) Loop-out mechanism illustrates the unwinding of DNA prior to fill-in synthesis and ligation. (B) Snap-back mechanism illustrates the formation of a hairpin structure after extensive resection. The green represents MH repeats whereas the blue represents primer repeats, either direct or inverted. A unique characteristic of SD-MMEJ is that nascent DNA is synthesized following repair, generating templated insertions and deletions. Figure adapted from Khodaverdian et al., 2017.

In analyzing the different sequence contexts flanking a Cas9-induced break, the McVey laboratory has made the following observations: 1) there is a delicate balance between the length of primer repeats forming secondary structures and their distance, with shorter repeats being favored when the inter-repeat distance is small; 2) the impact of sequence flexibility/rigidity in the loops formed during secondary structure formation, with homopolymeric A sequences impeding hairpin formation and flexible AT loops promoting repair; and 3) the strong propensity toward SD-MMEJ repair by the presence of favorable MH sequences nearby the secondary-structure forming repeats (Hanscom et al., 2022). The current SD-MMEJ pipeline can account for certain types of alt-EJ repair events including blunt joins, simple deletions, and single-step insertions, although it is limited by the fact that it cannot model insertion events that occur in multiple steps. Apart from assessing the translocation junctions for their consistency with either NHEJ or alt-EJ, we can evaluate to what extent these junctions are consistent with SD-MMEJ repair. This involves examining factors such as the distance between primers and the break site,

the length of translocation insertions and/or deletions, as well as the relative percentages of ABJ, indel, and MHJ events. These factors have been extensively studied in the context of inter-plasmid repair. However, adapting the current SD-MMEJ and published HTGTS pipelines for studying translocation junctions in Drosophila S2 cells will be necessary to study SD-MMEJ consistency.

There is strong evidence that *Drosophila melanogaster* is an ideal model organism for investigating alt-EJ repair. For example, the absence of key mammalian c-NHEJ factors, such as DNA-PKcs and Artemis, may predispose Drosophila to resolve breaks through non-NHEJ pathways. I predict that translocation formation in Drosophila will proceed through alt-EJ mechanisms in a POLQ-mediated fashion, given its preference for this repair pathway, and that among these events, a subset will be consistent for SD-MMEJ. While there has been ongoing research to determine the genetic requirements for SD-MMEJ, it is a relatively complex model and most likely involves interactions between multiple alt-EJ proteins and other factors.

2. The Role of POLQ and Its Domains in Mediating Translocation Formation

The McVey laboratory has previously implicated POLQ in driving SD-MMEJ events through its involvement in alt-EJ mechanisms, particularly during insertion events where it aligns MH sequences during repair. POLQ comprises an N-terminal helicase-like domain, a polymerase domain, and a highly disordered central domain also termed the linker, as depicted in Figure 20.

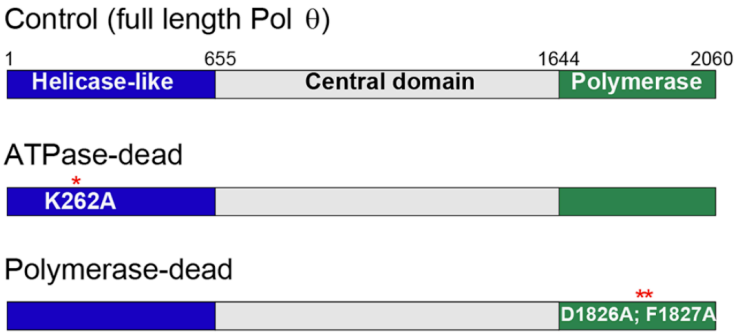

Figure 20. Schematic of *mus308* transgenic rescue stock with endogenous expression (top), ATPase-dead stock (middle), and polymerase-dead stock (bottom). The helicase-like domain (blue) is connected to the polymerase domain (green) by a central linker domain (grey). In the ATPase-dead stock, there is a single amino acid change from lysine to alanine at position 262. In

the polymerase-dead stock, there are two amino acid changes: aspartic acid to alanine at position 1826 and phenylalanine to alanine at position 1827.

To further investigate the role of POLQ in SD-MMEJ and whether it is driving translocation formation, we aim to explore how each of its domains affects repair events and the preferences of primer repeats and MHs that POLQ favors. The McVey laboratory has generated transgenic Drosophila stocks that lack the endogenous *mus308* gene encoding for POLQ, with accompanying mutations in either the Walker A box of the helicase-like domain (ATPase-dead) or in the catalytic region of the polymerase domain (polymerase-dead) via a transgene. These stocks, known as domain mutants, allow the disruption of the DNA-dependent ATPase activity of the helicase-like domain and DNA polymerase activity, potentially enabling us to pinpoint the function of each POLQ domain.

The helicase and polymerase domains for POLQ may have attributable roles in facilitating DSB repair and by extension, translocation formation. Although POLQ's polymerase activity has been well-studied, there is limited knowledge surrounding the functions of the helicase domain. However, it has been implicated that the helicase domain is involved with aligning DNA strands containing short MHs to facilitate MMEJ repair (Newman et al., 2015). Furthermore, although the ATPase function of the helicase domain has been demonstrated to be expendable in alt-EJ, it was noted that ATPase-dead mutants displayed a reduced frequency in insertion events during alt-EJ repair and utilize shorter MHs (2-3 bp) at repair junctions *in vivo* in Drosophila (Beagan et al., 2017). Remarkably, human POLQ has been shown to directly interact with RAD51, a recombinase that binds to DNA following a break, which results in the displacement of RAD51 from ssDNA in an ATPase-dependent manner (Ceccaldi et al., 2015). In studies with human cells, the POLQ ATPase activity has been observed to be crucial for survival in an HR-deficient background. Therefore, this makes the helicase domain an appealing target for cancer therapeutics, especially in HR-deficient tumors.

It has been demonstrated that the polymerase domain can carry out most of the enzymatic steps of MMEJ, joining DNA ends with limited microhomologies *in vitro* (Kent et al., 2015). If consistent with previous research indicating the indispensability of the polymerase domain, I anticipate a significant reduction in the frequency of alt-EJ events in Drosophila lacking the

polymerase activity of POLQ. By extension, the frequency of translocations in a polymerase domain mutant may also decrease as a result of the reduction in alt-EJ events, as this is the repair pathway that Drosophila typically prefers. For mutants lacking functional ATPase activity, I predict there will be simpler alt-EJ events with a reduced frequency of larger deletions as well as loss of indels. This is based on the model that the helicase domain plays a crucial role in DNA unwinding and scanning of MHs used in templated insertions and deletions.

3. Further *In Vivo* Experiments

As noted earlier, a limitation in using S2 cells for experiments is that they are polyploid and therefore may not accurately represent the DNA content found in the organism from which the cells were derived. As such, we hope to extend the findings of our experiments from the inter-plasmid system and eventually from S2 cells to whole organism Drosophila, as this would provide us with the most accurate representation and insight into translocation formation and how POLQ might be mediating these events. To validate this approach in whole organism, transgenic Drosophila stocks will be generated to express a sgRNA that has been validated to be highly efficient both *in vitro* and *in vivo* (embryos). A similar approach to identifying target genomic loci and potential "bait" sites can be applied, and *in vivo* translocations can be isolated by crossing these flies with flies that express Cas9. As mentioned in the previous section, transgenic stocks of POLQ domain mutants have been generated by the lab, which can be employed to further elucidate the exact role of POLQ in mediating translocation formation. Lastly, the integration of these guide constructs in the Drosophila genome would enable further investigation of SD-MMEJ.

Acknowledgments

First, I want to thank **Professor Mitch McVey** for his role as my major advisor and for providing me with the opportunity to be part of his research lab. His guidance has been instrumental to my academic and personal growth as well as my development as a young researcher. I would like to extend a special thank you to my graduate student mentor, **Nick Woodward**, for his exceptional mentorship over the past two years and for equipping me with the necessary tools and knowledge that I will carry forward in my future research endeavors. Thank you to **Dr. Justin Blanch**, **Dr. Mai Tran**, **Tokio Sano**, and **Lara Maggs** for being a constant source of positivity and encouragement and for being wonderful friends. Thank you to **Professor Lawrence Uricchio** for serving on my thesis committee and challenging me to think critically about this project. Thank you to **Dr. Sara Martin** for her valuable insights and feedback throughout this project and to **Dr. Alice Miller** for the laughs, fun, and support. Thank you to the rest of the McVey lab, members past and present, for their camaraderie and for making me feel like I belonged since day one.

References

Beagan, K., & McVey, M. (2016). Linking DNA polymerase theta structure and function in health and disease. *Cellular and Molecular Life Sciences*, CMLS, 73(3), 603–615. https://doi.org/10.1007/s00018-015-2078-9

Beagan, K., Armstrong, R. L., Witsell, A., Roy, U., Renedo, N., Baker, A. E., Schärer, O. D., & McVey, M. (2017). Drosophila DNA polymerase theta utilizes both helicase-like and polymerase domains during microhomology-mediated end joining and interstrand crosslink repair. *PLOS Genetics*, 13(5), e1006813. https://doi.org/10.1371/journal.pgen.1006813

Boboila, C., Jankovic, M., Yan, C. T., Wang, J. H., Wesemann, D. R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M., & Alt, F. W. (2010). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), 3034–3039. https://doi.org/10.1073/pnas.0915067107

Brinkman, E. K., & van Steensel, B. (2019). Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER. *Methods in Molecular Biology,* 1961, 29–44. https://doi.org/10.1007/978-1-4939-9170-9_3

Ceccaldi, R., Liu, J. C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M. I., O'Connor, K. W., Konstantinopoulos, P. A., Elledge, S. J., Boulton, S. J., Yusufzai, T., & D'Andrea, A. D. (2015). Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature*, 518(7538), 258–262. https://doi.org/10.1038/nature14184

Chiarle, R., Zhang, Y., Frock, R. L., Lewis, S. M., Molinie, B., Ho, Y. J., Myers, D. R., Choi, V. W., Compagno, M., Malkin, D. J., Neuberg, D., Monti, S., Giallourakis, C. C., Gostissa, M., & Alt, F. W. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell*, 147(1), 107–119. https://doi.org/10.1016/j.cell.2011.07.049

Clouaire, T., & Legube, G. (2019). A Snapshot on the Cis Chromatin Response to DNA Double-Strand Breaks. *Trends in Genetics*, 35(5), 330–345. https://doi.org/10.1016/j.tig.2019.02.003

Deriano, L., & Roth, D. B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annual Review of Genetics*, 47, 433–455. https://doi.org/10.1146/annurev-genet-110711-155540

Difilippantonio, M. J., Zhu, J., Chen, H. T., Meffre, E., Nussenzweig, M. C., Max, E. E., Ried, T., & Nussenzweig, A. (2000). DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature*, 404(6777), 510–514. https://doi.org/10.1038/35006670

Feng, W., Simpson, D. A., Carvajal-Garcia, J., Price, B. A., Kumar, R. J., Mose, L. E., Wood, R. D., Rashid, N., Purvis, J. E., Parker, J. S., Ramsden, D. A., & Gupta, G. P. (2019). Genetic determinants of cellular addiction to DNA polymerase theta. *Nature Communications*, 10(1), 4286. https://doi.org/10.1038/s41467-019-12234-1

Ferguson, D. O., Sekiguchi, J. M., Chang, S., Frank, K. M., Gao, Y., DePinho, R. A., & Alt, F. W. (2000). The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), 6630–6633. https://doi.org/10.1073/pnas.110152897

Frock, R. L., Hu, J., Meyers, R. M., Ho, Y. J., Kii, E., & Alt, F. W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnology*, 33(2), 179–186. https://doi.org/10.1038/nbt.3101

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J. B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A. E., Hendrickson, E. A., Giovannangeli, C., Jasin, M., & Brunet, E. (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Molecular Cell*, *55*(6), 829–842. https://doi.org/10.1016/j.molcel.2014.08.002

Hanscom, T., Khodaverdian, V. Y., & McVey, M. (2018). Recovery of Alternative End-Joining Repair Products From Drosophila Embryos. *Methods in Enzymology*, 601, 91–110. https://doi.org/10.1016/bs.mie.2017.11.027

Hanscom, T., Woodward, N., Batorsky, R., Brown, A. J., Roberts, S. A., & McVey, M. (2022). Characterization of sequence contexts that favor alternative end joining at Cas9-induced

double-strand breaks. *Nucleic Acids Research*, 50(13), 7465–7478. https://doi.org/10.1093/nar/gkac575

Hu, J., Meyers, R. M., Dong, J., Panchakshari, R. A., Alt, F. W., & Frock, R. L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nature Protocols*, 11(5), 853–871. https://doi.org/10.1038/nprot.2016.043

Kang, Z. J., Liu, Y. F., Xu, L. Z., Long, Z. J., Huang, D., Yang, Y., Liu, B., Feng, J. X., Pan, Y. J., Yan, J. S., & Liu, Q. (2016). The Philadelphia chromosome in leukemogenesis. *Chinese Journal of Cancer*, 35, 48. https://doi.org/10.1186/s40880-016-0108-0

Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y., & Pomerantz, R. T. (2015). Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nature Structural & Molecular Biology*, 22(3), 230–237. https://doi.org/10.1038/nsmb.2961

Lal, A., Ramazzotti, D., Weng, Z., Liu, K., Ford, J. M., & Sidow, A. (2019). Comprehensive genomic characterization of breast tumors with BRCA1 and BRCA2 mutations. *BMC Medical Genomics*, 12(1), 84. https://doi.org/10.1186/s12920-019-0545-0

Lemée, F., Bergoglio, V., Fernandez-Vidal, A., Machado-Silva, A., Pillaire, M. J., Bieth, A., Gentil, C., Baker, L., Martin, A. L., Leduc, C., Lam, E., Magdeleine, E., Filleron, T., Oumouhou, N., Kaina, B., Seki, M., Grimal, F., Lacroix-Triki, M., Thompson, A., Roché, H., … Cazaux, C. (2010). DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30), 13390–13395. https://doi.org/10.1073/pnas.0910759107

Mateos-Gomez, P. A., Gong, F., Nair, N., Miller, K. M., Lazzerini-Denchi, E., & Sfeir, A. (2015). Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature*, 518(7538), 254–257. https://doi.org/10.1038/nature14157

Mitelman, F., Johansson, B., & Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7(4), 233–245. https://doi.org/10.1038/nrc2091

Newman, J. A., Cooper, C. D. O., Aitkenhead, H., & Gileadi, O. (2015). Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway. *Structure*, 23(12), 2319–2330. https://doi.org/10.1016/j.str.2015.10.014

Simsek, D., Brunet, E., Wong, S. Y., Katyal, S., Gao, Y., McKinnon, P. J., Lou, J., Zhang, L., Li, J., Rebar, E. J., Gregory, P. D., Holmes, M. C., & Jasin, M. (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLOS Genetics*, 7(6), e1002080. https://doi.org/10.1371/journal.pgen.1002080

Verkuijl, S. A., & Rots, M. G. (2019). The influence of eukaryotic chromatin state on CRISPR-Cas9 editing efficiencies. *Current Opinion in Biotechnology*, 55, 68–73. https://doi.org/10.1016/j.copbio.2018.07.005

Wyatt, D. W., Feng, W., Conlin, M. P., Yousefzadeh, M. J., Roberts, S. A., Mieczkowski, P., Wood, R. D., Gupta, G. P., & Ramsden, D. A. (2016). Essential Roles for Polymerase θ-Mediated End Joining in the Repair of Chromosome Breaks. *Molecular Cell*, 63(4), 662–673. https://doi.org/10.1016/j.molcel.2016.06.020

Yousefzadeh, M. J., Wyatt, D. W., Takata, K., Mu, Y., Hensley, S. C., Tomida, J., Bylund, G. O., Doublié, S., Johansson, E., Ramsden, D. A., McBride, K. M., & Wood, R. D. (2014). Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. *PLOS Genetics*, 10(10), e1004654. https://doi.org/10.1371/journal.pgen.1004654

Yu, A. M., & McVey, M. (2010). Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Research*, 38(17), 5706–5717. https://doi.org/10.1093/nar/gkq379

Zahn, K. E., Averill, A. M., Aller, P., Wood, R. D., & Doublié, S. (2015). Human DNA polymerase θ grasps the primer terminus to mediate DNA repair. *Nature Structural & Molecular Biology*, 22(4), 304–311. https://doi.org/10.1038/nsmb.2993

Zhang, Y., & Jasin, M. (2011). An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nature Structural & Molecular Biology*, 18(1), 80–84. https://doi.org/10.1038/nsmb.1940

Zhang, Y., McCord, R. P., Ho, Y. J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., Becker, M. S., Alt, F. W., & Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, 148(5), 908–921. https://doi.org/10.1016/j.cell.2012.02.002

Zhu, C., Mills, K. D., Ferguson, D. O., Lee, C., Manis, J., Fleming, J., Gao, Y., Morton, C. C., & Alt, F. W. (2002). Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell*, 109(7), 811–821. https://doi.org/10.1016/s0092-8674(02)00770-5

Zuckerman, J. (2022). Understanding Synthesis-Dependent Microhomology-Mediated End Joining Repair of DNA Double Strand Breaks. (Unpublished senior thesis). Tufts University, Medford, Massachusetts.