Improving CRISPR Off-Target Detection Using Transposases

Ann Abraham, Connor Gazda, Alexis Malamas and Olivia Wood.
Drexel University School of Biomedical Engineering, Science, and Health Systems, and College of Medicine
Advisor: Dr. William Dampier

Abstract:

Innovations in CRISPR technology are imperative for researchers like Dr. William Dampier, seeking advanced solutions for genome editing. Addressing the limitations of current methods, we aim to enhance the Genome Wide Unbiased Estimation of Double-Stranded Breaks Enabled by Sequencing (GUIDE-seq) protocol, crucial for Cas9-based gene editing safety assessment. Our goal is to streamline the off-target screening process by integrating a novel enzyme that condenses the existing two-step protocol into a single "cut and paste" procedure. This enzyme, the EZ-TN5 Transposase, offers rapid and efficient DNA insertion capabilities. To ensure compatibility with the established GUIDE-seq framework, our design mandates adherence to specific DNA sequences during the "paste" phase while maintaining a processing time under two hours. Key design requirements include fragment lengths within the 100 to 2000 bp range, with an optimal mean of 300 bp, and sequencing results consistent with existing technology, surpassing noise levels at $p<0.01$. Tests confirmed that the EZ-TN5 Transposase effectively generates fragments meeting these criteria, validating its suitability for GUIDE-seq enhancement. Sequencing analyses demonstrated comparable results to current methods, affirming the enzyme's efficacy in achieving accurate off-target detection. The adoption of this streamlined GUIDE-seq protocol promises significant benefits for Dr. Dampier's HIV cure research, facilitating high-throughput screening and accelerating scientific advancements in genome editing. By simplifying and expediting off-target assessment, our innovative solution empowers researchers to navigate CRISPR technology with greater efficiency and precision, driving transformative impacts on both individual research endeavors and broader societal healthcare initiatives.

Need:

User Problem: Our users are Dr. William Dampier and other researchers interested in CRISPR innovation. We intend to improve the Genome Wide Unbiased Estimation of Double-Stranded Breaks Enabled by Sequencing (GUIDE-seq) current methods of CRISPR off-target detection, specifically with the procedure of GUIDE-seq used within Drexel's Immunology laboratory used to assess the safety of Cas9 based gene editing.

Cas9 based gene editing techniques are a novel technology for improving human health. It is important to measure the safety of these techniques using an unbiased approach. In this way, potentially harmful therapies can be screened-out early in the process, thus saving time and money. Current implementations of the GUIDE-seq technique require chemical or physical shearing genomic DNA to an appropriate size which are laborious, time-consuming, lossy, and difficult to control. HIV (Human Immunodeficiency Virus) affects 39 million people worldwide [1].

No treatments target the dormant viral DNA in some immune cells. Addressing this problem, Dr. Dampier's GUIDE-seq takes around 4 to 5 hours. This length of time mainly comes from 2 steps, the fragmentation and adapter ligation steps. There are prep and washing stages during these steps, in which some of the DNA is lost. We aim to improve upon these two steps in the GUIDE-seq technique protocol, yielding a simpler and faster protocol.

Existing Solutions Review: Although there are no existing products to improve GUIDE-seq there are alternative ways to improve the protocol. There are two main alternative improvements: sonication and enzymatic shearing. Sonication involves breaking up DNA with sound waves. The advantages to sonication are it has a quicker active time than enzymatic shearing, less washing steps, since it does not involve enzymes, and is generally cheaper. However, sonication does not provide for the exact cutting of DNA, some of the ends might be different between different fragments of DNA. Sonication also has a larger set-up time. Sonication also has no effect on the ligation step, which contributes a large amount of time to the protocol [2]. Enzymatic shearing involves using digestion enzymes to break up the DNA. Enzymatic shearing yields identical cuts, which are useful since it makes the addition of adaptors easier. However, these enzymes need to be incubated yielding for a longer active time and then washed off which reduces the overall DNA yield of the procedure. Again, this has no effect on the ligation steps of the procedure.

Both methods are available or already used within Dr. Dampier's lab, however, the main feature that both methods lack is they do not affect ligation. Simplifying/replacing ligation is a key feature that we will focus on. A solution that combines the speed of sonication, the precision and identical cutting of enzymatic shearing and can simplify or fully eliminate the need for ligation would be ideal.
Objective: Our objective is to improve CRISPR's off-target screening procedures by utilizing a new enzyme which boasts a single "cut and paste" step into the GUIDE-seq protocol. In other words, take a 2-step process and turn it into 1 step.

Design Inputs
Constraints Summary Table: The key constraints details are shown below, see appendix 2 for full details on constraints.

Title	Description	Values (units)
Compatible Illumina i5 & i7 adapters	Compatible DNA ends to adhere to flow cell in Illumina DNA Sequencer. Known as i5 and I7 sequences	Presence/Absence
Adds a unique molecular index to each molecule	Identifying PCR-duplicates molecules. Each cut must have a unique molecule.	Presence/Absence
Compatible with current CRISPR-tag	Protocol enhancements compatible with current systems	True/False

Requirements Summary Table: The key requirements details are shown below, see appendix 3 for full details on requirements.

Title	Description (Value, Units)	Justification
DNA Fragment Length	100-2000 bp with ideal mean of 300 bp	Downstream protocols require this DNA length
PCR Enrichment	Fragments resulting from the solution should result in an exponential increase in DNA concentration.	Exponential increase indicates correct alignment of CRSIPR tag and solution
DNA amount	Solution needs to accommodate with at least 400 ng of DNA	Upstream protocols yield this DNA amount
Incubation Time	Solution incubation time must be at most 2 hours	Current protocol takes 2 hours, improvement needs to be less

Solution:
Assembly & Use:

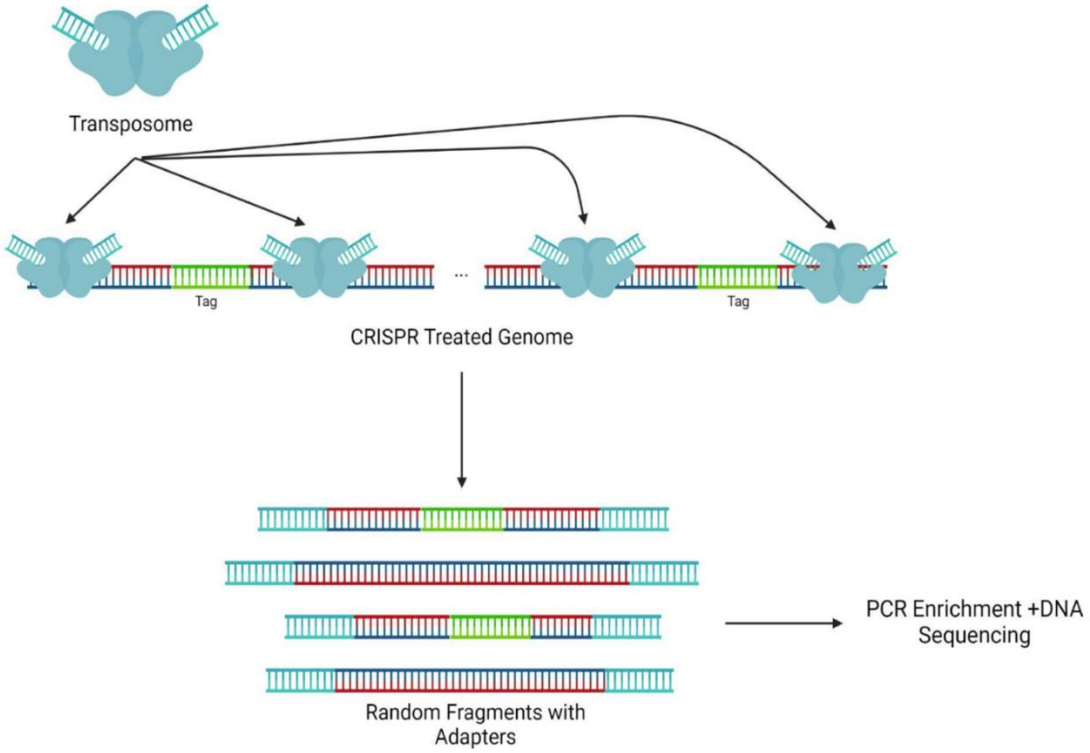


Figure 1 Transposase Reaction: The transposomes bind to random sites along the CRISPR treated genome, then shear at that site and insert the loaded adapter to the ends of each site. This results in random fragments all with adapters at the ends. These fragments will then be enriched by the PCR step then sequenced.

Built Solution Highlights: The transposase enzyme system has fast "cut and paste" and "copy and paste" functions. EZ-TN5 Transposase is a hyperactive, mutated form of TN5 transposase, a highly efficient enzyme for insertion of an EZ-TN5 Transposon into any target DNA, in vitro. All commercially available TN5 Transposases are EZ-TN5, some companies specify EZ-TN5, while others do not. For generalization, we will use TN5 moving forward in this report. We will be using TN5 transposase to insert a custom DNA sequence. The main design choice was deciding what TN5 transposase to use that fits our design requirements, the fragment cut length (R1), the DNA amount (R3), and the incubation time (R4). Our built solution will use the Varizymes TN5 Transposase kit, which contains TN5 enzymes, and buffer solutions for loading/TN5 Reaction. This enzyme fits our needs for the cut length and the timing, cutting 100ng in 30 minutes (R4) to fragments of 200 - 1000 bp in length (R1). Although the DNA amount recommended is 100 ng, it is sold in batches of 50 experiments, giving us plenty of enzymes to work with to find the ideal incubation time and enzyme amount to work with 400 ng of DNA (R3).

The second component is the adapter molecule, which will be loaded into the TN5. This component was designed based on our constraints (C5, C6, C7), and R2 as this adapter will allow for the same end product as the original GUIDE-seq protocol. The molecule build comprises 4 main parts: first, the Illumina Adapter which allows for the PCR enrichment and later Illumina dye sequencing (C5, C7), second the Unique Molecular Index which allows for the identification of each unique read (C6), third the synthetic sequence which acts as a space filler and last the TN5 recognition site which allows for the DNA to be loaded onto the TN5 enzyme. These sections are laid out in Figure 2. Our adapter was built with two types of adapters, literature involving TN5 enzymes regarding how it should be oriented, and we built both 3' and 5' orientations of the adapter. See our demo video of our solution for a more detailed diagram of how our solution works: link

Originally, we planned to use Beta Lifescience TN5, which had a $14.60 price per reaction with a reaction time of 1 hour, fitting our time constraint. We chose the Beta Lifescience TN5 enzyme due to it being $14.60 per reaction and fitting our constraint of 2 hours. This enzyme was sold out, however.
Build/Test Summary: Test 1 will test for the fragment length, mainly component 1, but component 2 will be present in testing. This test will verify that component 1 fulfills requirements R1, R3, and R4. Test 2 will verify that both component 1 and component 2 create fragments that will enrich for CRISPR tags through the PCR procedure of GUIDE-seq, as well as confirming which adapter orientation is correct.

Verification Testing:

Test 1:

Introduction: The first test will confirm the length of the DNA fragments resulting from various protocol setups (varying in the amount of enzyme and time of active shearing) (R1). Our method falls in line with the indications for use for the Aligent Bioanalyzer and has been used in scientific studies to measure the length of DNA and RNA fragments, [4, 5].

Resources:

- Aligent 2100 Bioanalyzer
- Immunology Lab equipment
    - Beakers, micropipettes, microcentrifuge tubes, etc
- Loaded TN5 enzymes (Components 1,2)
- Cell Line DNA

Method: Cell line DNA was treated with Loaded TN5 enzymes at various TN5 amounts and times, the fragments resulting from the reaction were purified using an Ampure Bead purification, then the fragment size was measured with an Aligent 2100 bioanalyzer. Fragment size data was averaged for each enzyme amount/time then a linear regression was performed on the data. Tests will be performed at increasing enzyme amounts and reaction times, until an amount/time results in 300 bp or resources to perform additional testing is exhausted. Should none be identified, use linear regression to predict enzyme amount/time that would result in 300 bp. Test will be considered a failure if no enzyme amounts/time yield fragments within 100-2000 bp range of R1.

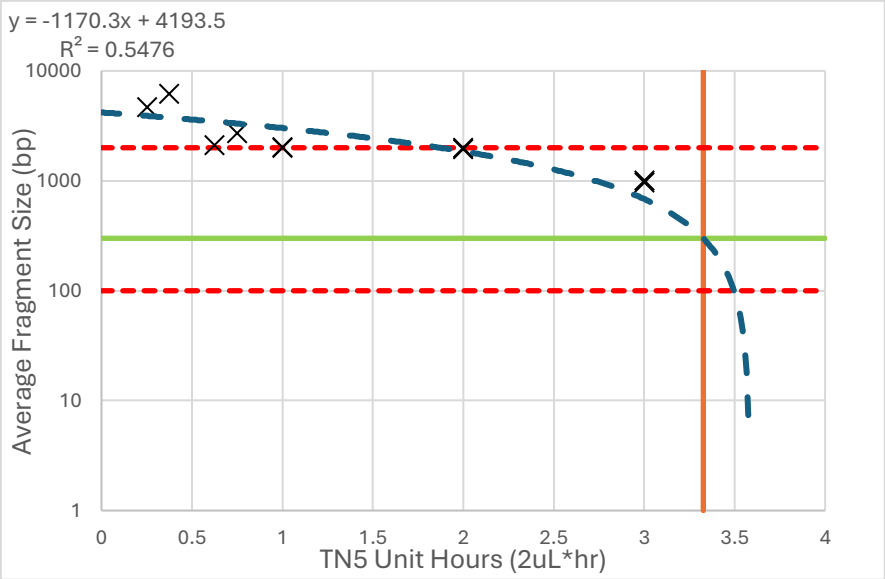


Figure 2 TN5 Cutting Test Data: Data is shown as black crosses, there are repeats of data at same TN5 unit hours points, as different enzyme amounts/times amounted to that point. Example: 2 TN5 units at 1 hour = 2 TN5 unit hours and 1 TN5 unit at 2 hours = 2 TN5 Unit hours. Equation and R squared of figure are shown in upper left corner.igure are shown in upper left corner.

Results: No enzyme amount/time directly resulting in the ideal 300 bp average, however higher enzyme amount/time test groups were within the 100-2000 bp range. Figure 3 shows the average DNA fragment size per TN5 unit hours, along with the regression of that data as the blue dashed line, and requirement bounds shown as. The regression identified ~3.32 TN5 unit hours as the experimental set up those results in the ideal mean of 300 bp, shown with the green and orange lines.

Discussion: The TN5 reaction *effectively shortened DNA* fragments to within 100-2000 bp, and the identified 3.32 TN5 unit hours, which moving forward we will use 2 TN5 Units with 100 min reaction time (~1.66hr). The main confidence issue with this test is the reliability of the regression data, the R^2 value = 0.5476, however since the lower TN5 unit hours groups were within the requirement bounds, and the new experiment setup is higher where the enzyme has more time to cut the DNA shorter, this will not affect the next test in a meaningful way.

Test 2:

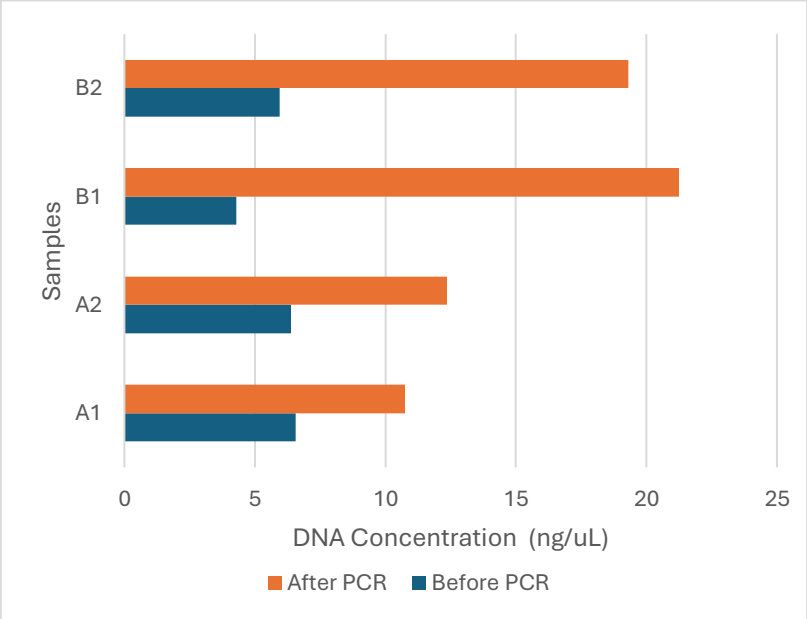
Introduction: The second test will confirm that TN5 treated DNA fragments are enriched by a PCR treatment with GUIDE-seq primers, DNA concentration should increase exponentially(R2). Our method is to follow PCR protocol of the existing GUIDE-seq procedure, giving us high confidence in our results.

Resources:

- Thermocycler
- Qubit Fluorometer/Qubit High Sensitivity kit
- Mixture of CRISPR treated DNA samples and GUIDE
- Loaded TN5 samples (Adapter Orientations A [3'], B [5'])
- Lab equipment
    - Beakers, micropipettes, microcentrifuge tubes, etc.

Method: DNA fragments will be treated with TN5 at the ideal experiment setup for 300 bp identified by the previous test and record DNA concentration with Qubit, then run the GUIDE-seq PCR protocol with Thermocycler and record new DNA concentration. Run test with 2 samples from each adapter orientation, then compare adapter orientations. Test will be considered a fail if both orientations increase the same amount or if there is no increase in DNA concentration.

Results: Data is shown in Figure 3, which shows the DNA concentrations before and after the PCR reaction. Both adapter orientations saw increases in the DNA concentration, however orientation B, especially B1 which increased from 4.29 ng/uL to 21.25 ng/uL had double the increase in DNA concentration as orientation A.



A1, A2. After PCR data is average of both PCR forward and reverse directions for each sample.

Discussion: Since our adapter orientation is a binary, 3' vs 5' orientations, orientation B is most likely to see an exponential increase in DNA concentration due to the larger increase, indicating a more successful PCR. Both orientations saw some increase as the polymerase reaction would add to the fragments but only the correct orientation would increase the most as the primers are aligned. Although the larger increase does not confirm an exponential increase, since our adapter orientations cover both possible orientations, we are confident that the B orientation is correct.

Conclusion

Dr. Dampier will possess a high-throughput GUIDE-seq, significantly expediting HIV cure research with our research of TN5. The transposases system, known for its "cut and paste" functions. Our need of compatible Illumina i5 and i7 adapters while also incorporating a unique molecular index to each molecule is a requirement for our solution. Maintaining compatibility with current oligo-tag and oligo length is imperative, alongside ensuring that the results align with existing technology standards.

Furthermore, the solution aims to streamline processes by condensing two steps into one, thereby enhancing efficiency. Additionally, it seeks to augment GUIDE-seq for Next-Gen Sequencing-compatible molecules, promising advancements in genomic research and analysis.
In summary with our verification tests:

Verification Test	1	2
Pass/Fail	Mixed	Pass
Why pass/fail?	Did not meet our 300bp fragment size average, however fragments are in 100-2000bp range	Found adapter orientation through PCR enrichment.

In our future revisions, we plan to explore the use of a TN5, which could potentially facilitate the cutting of larger amounts of DNA in a single process. Additionally, we aim to streamline our sequencing process by directly testing comparable samples, rather than relying on downstream DNA obtained from the laboratory. These enhancements hold significant potential implications, particularly in the realm of HIV therapy. By effectively locating dormant HIV DNA through our improved methodology and subsequently excising it, there exists the possibility of rendering current therapy for HIV patients obsolete. This advancement could represent a monumental shift in the treatment landscape for HIV, offering new avenues for targeted and potentially curative interventions.

The development of a simpler and quicker GUIDE-Seq protocol marks a significant advancement in genetic research methodologies. This innovation accelerates the pace of research and substantially reduces the protocol's operating costs. By enabling high throughput, researchers can now process large sample sets more rapidly, thereby enhancing overall efficiency. Automation plays a crucial role in this advancement, significantly reducing errors and ensuring greater consistency across experiments. Moreover, the scalability of the protocol allows for seamless adaptation to varying sample sizes and experimental needs. These improvements increase the likelihood of detecting rare genetic events, further expanding the utility and impact of GUIDE-Seq in genetic research endeavors.

A risk in our solution involves unintended off-target effects of the GUIDE-Seq, particularly with the TN5 cutting. This could result in cutting DNA in unintended areas, potentially inserting DNA into healthy regions rather than targeting infected DNA. These off-target effects could compromise genetic integrity, posing ethical and safety concerns. Mitigating this risk requires thorough protocol validation to minimize off-target effects and ensure precise targeting of pathogenic DNA. Ongoing monitoring is essential to assess efficacy and prevent unintended consequences.

References

[1] WHO. (2023). HIV. *Global Health Observatory*. Link.

[2] Sambrook J, Russell DW. "Fragmentation of DNA by sonication." CSH Protoc. 2006 Sep 1;2006(4):pdb.prot4538. doi: 10.1101/pdb.prot4538. PMID: 22485919.

[3] Apone et al. New England Biolabs. *Feature Articles*. Link.

[4] Krowczynska, A. "Analysis of DNA fragments using the Agilent 2100 Bioanalyzer." Covaris Inc. Woburn, Massachusets. Link

[5] Liang, B. et al. "Enrichment of the fetal fraction in non-invasive prenatal screening reduces maternal background interference" *Sci Rep* 8, 17675 (2018). https://doi.org/10.1038/s41598-018-35738-0

Appendix 1, Features Table:

#	Description
1	**Produces an Illumina compatible molecule.**
2	**Enriches for oligo-tag containing molecules which represent Cas9 cutting events.**
3	**Results match current technology.**
4	**Decreases handling steps**
5	**Reduce handling time.**

Appendix 2, Constraints Table:

C#	Title	Description	Values (units)	Feature(s) #
1	Compatible Illumina i5 & i7 adapters	Compatible DNA ends to adhere to flow cell in Illumina DNA Sequencer. Known as i5 and I7 sequences	Presence/Absence	1,3
2	Adds a unique molecular index to each molecule	Identifying PCR-duplicates molecules. Each cut must have a unique molecule.	Presence/Absence	3
3	Compatible with current CRISPR-tag	Protocol enhancements compatible with current systems	True/False	2,3

Appendix 3, Requirements Table:

R#	Title	Description	Definition (Value,Units)	Justification	Feature(s) #
1	DNA Fragment Length	100-2000 bp with ideal mean of 300 bp	Length, bp	Downstream protocols require this DNA length	1,2,3,4,5
2	PCR Enrichment	Fragments resulting from the solution should result in an exponential increase in DNA concentration.	DNA concentration, ng/uL	Exponential increase indicates correct alignment of CRSIPR tag and solution	1,2,3
3	DNA amount	Solution needs to accommodate with at least 400 ng of DNA	DNA mass, ng	Upstream protocols yield this DNA amount	1,2,3
4	Incubation Time	Solution incubation time must be at most 2 hours	Time, hr	Current protocol takes 2 hours, improvement needs to be less	4,5