Inverse Prime Editing Using IscB Protein and Circular Guide RNA

Marco Rendace

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science

Graduate Program In Chemistry
York University
Toronto, Ontario

July 2026

© Marco Rendace, 2026

Abstract

To improve editing efficiency from previous inverse prime editing (IPE) research, cutting activity from smaller sized Cas12 subtypes and IscB (Cas9 ancestor) were compared. IscB was later chosen to study IPE utilizing two separate guide RNA, splitting its cutting and editing purposes. The editing guide RNA was designed circularly to contain DNA-binding (PBS) and edit-containing regions (RTT) to conserve both regions from degradation. Successful IPEs were later shown to indicate that the IPE mechanism with IscB and circular RNA were possible, but the edits were not as consistent and frequent. Next generation sequencing (NGS) methods, specifically PCR primer quality, need to be troubleshot to fully observe its editing capabilities. If NGS is resolved but low editing frequency continues, assays for optimal plasmid composition within the same lipid nanoparticle (LNP) could be tested to improve plasmid delivery into the cell through LNP packing, transfection and nuclear transport.

Dedications

I dedicate this thesis and my entire life to my Father and to my Lord and Savior Jesus Christ. I would always read the same quote by Prof. Russell Cowburn at my desk; “A good day in the lab is a cause for worship, because you come out of it seeing God's creation just a little bit more clearly than when the day started.”

To my parents Giuseppe and Paola Rendace. Your sacrifice in ensuring that my future is filled with happiness and opportunities only drives me further into making you both proud. Thank you both for the loving support I could lean on as I continue to grow into the man I am today.

To my brother Luca Rendace, thank you for being a true friend in life as we support each other throughout every chapter. I know there are many great ones to come.

To my uncle (Zio) Oliver Rendace, thank you for your mentorship and constant support by treating me like one of your own. I already have a major advantage in knowing how to navigate my paths in life because of how much you have accomplished.

To my dog and childhood dream Brio (I’m aware he can’t read this). You teach me so many lessons and outlooks that no human could do as effectively by seeking joy in the smallest things I take for granted. Most importantly for your love that brightens my days.

To all of my family and friends. I thank each and every one of you for giving me your own form of love and support that I truly cherish. Although I’m blessed for the love I have from multiple people, I value and notice every single person that is in my corner.

Acknowledgements

Thank you to my supervisor Dr. S. Brian Kim for introducing me to CRISPR genome editing, a whole new form of research I did not realize I would be so passionate about in such a short amount of time. Most importantly, I thank his willingness in accepting me into his lab group.

Thank you to Dr. Ryan Hili for being a strong source of support and always being open in giving advice for academic, research and future career purposes. His quality of making me feel genuinely heard is something I don't take for granted.

Thank you to all of my colleagues. From those who added their ideas to the research, were a helping hand in the lab or to those who just say hi, thank you for making the workdays much easier.

Table of Contents

Abstract ii

Dedications iii

Acknowledgements iv

Table of Contents v

List of Tables vi

List of Figures vii

Chapter One: Introduction 1
- Genome Editing Background 1
- Inverse Prime Editing 2
- Proposals To Improve Inverse Prime Editing - Protein Size 4
- Proposals To Improve Inverse Prime Editing - Separate Guide RNA 7

Chapter Two: Experimental Methods 10

Chapter Three: Results and Discussion 13
- Testing Cas12 Subtypes: Transfection Efficiency Using Flow Cytometry 13
- Observing Cutting Activity Through Sanger Sequencing 14
- Observing Cutting Activity Through NGS Sequencing 16
- Inverse Prime Editing Experimental Design 17
- NGS Sample Preparation 18
- Inverse Prime Editing Results: Comparing Azenta and ISEQ Primers 18
- PCR Quality Using Azenta and ISEQ PCR-2 Primers 20

Chapter Four: Conclusions and Future Work 24

Bibliography 26

Appendices 30
- Appendix A: NGS Analysis of Precise Edits 30
- Appendix B: Tables of Expanded NGS Results 35
- Appendix C: List of Materials Used in Experimental Methods 40

List of Tables

Table 1: Gene sites tested through Sanger Sequencing 15

Table 2: Inverse Prime Edits (simplified NGS results using Azenta primers) 18

Table 3: Inverse Prime Edits (simplified NGS results using Azenta primers) 19

Table 4: IscB cutting activity (expanded NGS results) 34

Table 5: Inverse Prime Edits (expanded NGS results using Azenta primers) 35

Table 6: Inverse Prime Edits (expanded NGS results using ISEQ primers) 36

List of Figures

Figure 1: Inverse prime editing flowthrough using only crRNA 3

Figure 2: Difference between prime editing and inverse prime editing 4

Figure 3: Size comparison between Cas12 subtypes 5

Figure 4: Structure of enhanced IscB (enOgeuIscB) 7

Figure 5: Using crRNA and cRNA to incorporate an edit 8

Figure 6: RNA circularization via Tornado method 9

Figure 7: Average fluorescence of BFP and mCherry from transfected plasmids 13

Figure 8: Cutting activity of Cas12 subtypes through Sanger Sequencing 14

Figure 9: Cutting activity of Cas12f and IscB through NGS 16

Figure 10: Average cutting activity of IscB through NGS 17

Figure 11: Gel of PCR-2 samples from run NGS 11 using Azenta primers 20

Figure 12: PCR-1 and PCR-2 gels of NGS 17 (NGS 12 repeat) with ISEQ primers 21

Figure 13: NGS sequence analysis of NGS 17 (NGS 12 repeat) with ISEQ primers 23

Figure 14: NGS analysis of precise edits, run NGS 10 sample I-214 29

Figure 15: NGS analysis of precise edits, run NGS-11 sample I-211 30

Figure 16: NGS analysis of precise edits, run NGS 11 sample I-214 30

Figure 17: NGS analysis of precise edits, run NGS 12 sample I-210 31

Figure 18: NGS analysis of precise edits, run NGS 12 sample I-211 31

Figure 19: NGS analysis of precise edits, run NGS 17 (NGS 12 Repeat) sample I-210 32

Figure 20: NGS analysis of precise edits, run NGS 17 (NGS 12 Repeat) sample I-211 33

Introduction

Genome Editing Background

DNA is commonly termed as the code of life with each base in its long unique sequence being responsible for the vast complex systems that make up each living being. Genes act as parts that make up the entire genome and can be distinguished to express different characteristics. A mutated sequence within a gene can mostly lead to consequences such as disease, can benefit from what the gene is coding or not have any impact at all. Gene editing opens the possibility of controlling desired and undesired characteristics of genes so that the genome can be expressed in a customizable way. Gene editing has been used to alter gene expression in a variety of industries. For example, in the agriculture industry gene editing can improve the quality of plants in a variety of factors such as environmental stress tolerance and increasing fruit yield and disease resistance.[1] Gene therapy is a large field of medicine which uses gene editing to target mutated genes that prevents disease-causing genomic sequences.[2] Gene editing technologies have grown over time starting with Zinc Finger Nucleases (ZFN) in the 1990s which used a cutting enzyme (Fokl nuclease) and DNA binding proteins with zinc, specifically designed for each domain to bind onto particular DNA regions three bases at a time.[3] Transcription Activator-Like Effective Nucleases (TALENs), a larger protein with higher DNA-binding specificity compared to ZFN, was discovered twenty years later and reduced the number of recognized bases for each protein domain from three to one.[3] Both ZFN and TALENs still needed to go through the costly, time consuming and very complex protein design procedure until a breakthrough in Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) with Cas proteins made preparation for gene editing much simpler, cheaper and more feasible.[3, 4] CRISPR relies on using RNA as a guide to target and bind onto gene sites while using Cas as a DNA-cutting protein instead of designing the protein itself to bind onto specific regions.[3, 4] The flexibility of using RNA mediated binding allowed CRISPR technology to broaden its scope on different types of gene editing.[3] Variations of CRISPR had been discovered over time using different protein components for multiple methods such as gene editing, RNA-targeting or fluorescent DNA imaging.[4] However when using CRISPR for DNA-cutting purposes, Cas9 tends to be the popular choice due to it being a highly effective endonuclease by making a double-stranded blunt-end cut within the DNA strands (native SpCas9).[4] When Cas9 would cut

the DNA, the cell utilizes its non-homologous end joining (NHEJ) or homology directed repair (HDR) repair systems. Gene knockout was done to essentially turn off genes via NHEJ by cutting the DNA and allowing the cell to repair the cut by inserting or deleting nucleotides (indels) randomly.[4] HDR is a repair method in which the cell uses homologous DNA as a template derived from a sister chromatid to repair the cut DNA from Cas9.[5] Once Cas9 was able to work with other proteins tethered to it, CRISPR editing was much more precise. Base editing began with Cas9 engineered to only nick one of the strands to keep the DNA intact as an attached cytidine or adenine deaminase protein changes particular C to U or A to G bases.[6] Most recently, the most flexible and accessible form of CRISPR editing is prime editing which replaces whole sequences, instead of one base at a time, using reverse transcriptase (RT).[7] Regardless of how far CRISPR has come, the prime editing system could be further improved so that whole sequences can be placed anywhere within the genome. The focus of this thesis project is to utilize prime editing capabilities by expanding editing windows through inverse prime editing, in a way for controlled and desired edits to be incorporated with high precision and efficiency changing Cas protein and guide RNA characteristics.

Inverse Prime Editing

Prime editing is specifically done on the non-target DNA strand which contains the protospacer-adjacent-motif (PAM). Inverse prime editing is incorporating the edit on the target strand, the strand not containing PAM. The PAM is a series of nucleotides that the Cas protein recognizes. A guide RNA (crRNA) is attached to the Cas protein to guide it to the desired gene site for Cas to make a double-strand break. The crRNA selects the desired gene site because it contains the primer binding site (PBS) which is complementary to one of the DNA strands. The crRNA also consists of the reverse transcriptase template (RTT) which contains the complementary sequence of the desired edit. After PBS binds to the gene site and a double stranded break has occurred from the Cas protein, RT elongates the desired sequence complementary to RTT starting from the bound PBS to the unbound RTT region before the edit is fully incorporated, post DNA repair (Figure 1).[8] The need for two methods of prime editing to occur on each DNA strand is to increase access to editing windows. Since RT runs 5’ to 3’, the PAM location and cut site could dictate whether an edit could be possible. For example, if the

edit is desired upstream but the PAM is downstream, prime editing could not be used but inverse prime editing could be done instead (Figure 2).[8, 9]

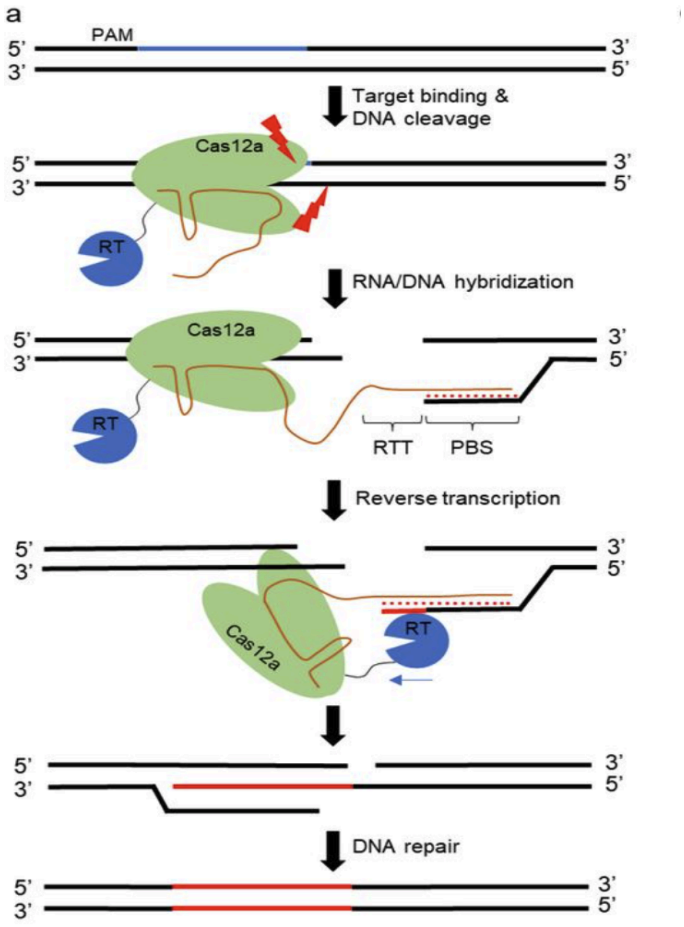


Figure 1: Inverse prime editing flowthrough using crRNA. Cas12a makes a staggered double stranded break while its crRNA binds to the target strand using its PBS region. The edit is completed on the strand not containing the PAM sequence (target strand) using RT elongating along the RTT region of the crRNA. Image taken from Bill Kim, Y; *et al*.[8]

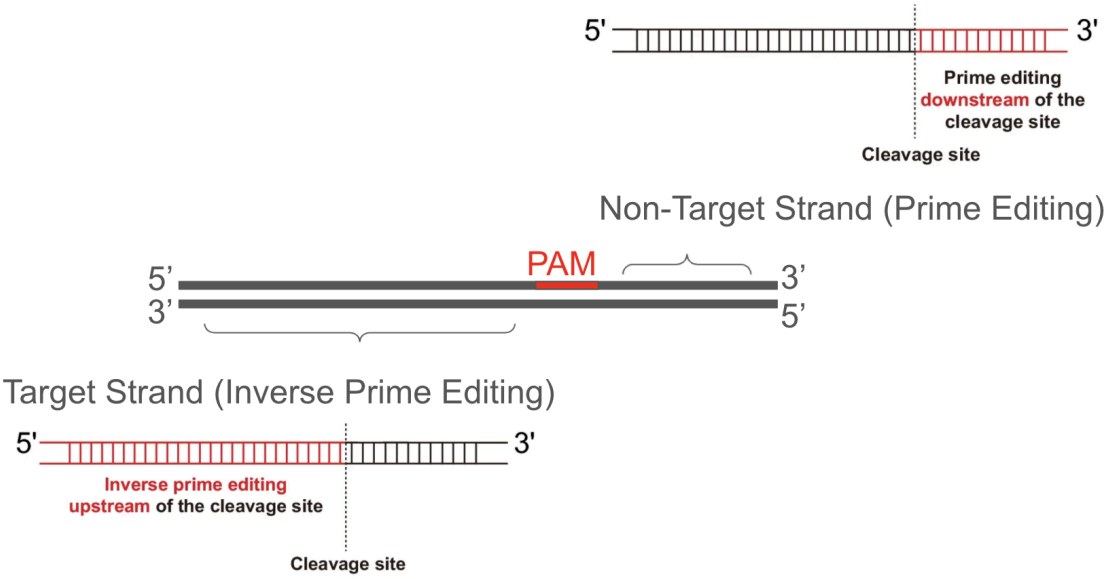


Figure 2: The difference between prime editing and inverse prime editing and when each are needed based on the PAM and cut site location. RT elongates from 5' to 3' which causes the need for inverse prime editing if the PAM is too downstream from a desired editing window. Images (top and bottom) were taken from Liang, R; *et al*.[9]

Proposals To Improve Inverse Prime Editing - Protein Size

Lipid nanoparticles (LNP) carrying plasmids that code for the inverse prime editing components are introduced to the cell via transfection. Previous research has observed why larger plasmids are less favorable by using chemically labelled fluorescent dyes that intercalate within the bases of the plasmid. When testing for cytosol diffusion rates, researchers observed that as plasmid size increased, it moved much slower within the cell.[10] Another study found that increasing plasmid size gradually decreases nucleus entry efficiency by comparing fluorescent intensity within the nucleus and cytoplasm over time.[11] Calculating a ratio between free and LNP packaged fluorescent plasmids indicated that increasing plasmid size reduces DNA packing efficiency in LNP.[12] Larger plasmids are also thought to reduce efficiency in escaping the endosome. Endosome escape efficiency can be inferred by observing fluorescent plasmids to a fluorescently labelled endosome, monitoring separation over time.[13] Therefore, a way to improve efficiency in plasmid delivery is by using smaller proteins with less amino acids (aa) which are introduced as

smaller plasmids with less kilobases (kb). Initially Cas12a (1300 aa/3.9 kb), CasMINI (Cas12f) (529 aa/1.587 kb) and Cas12j-8 (Cas12j) (757 aa/2.271 kb) were chosen as smaller Cas protein subtypes (Figure 3).[14, 15, 16]

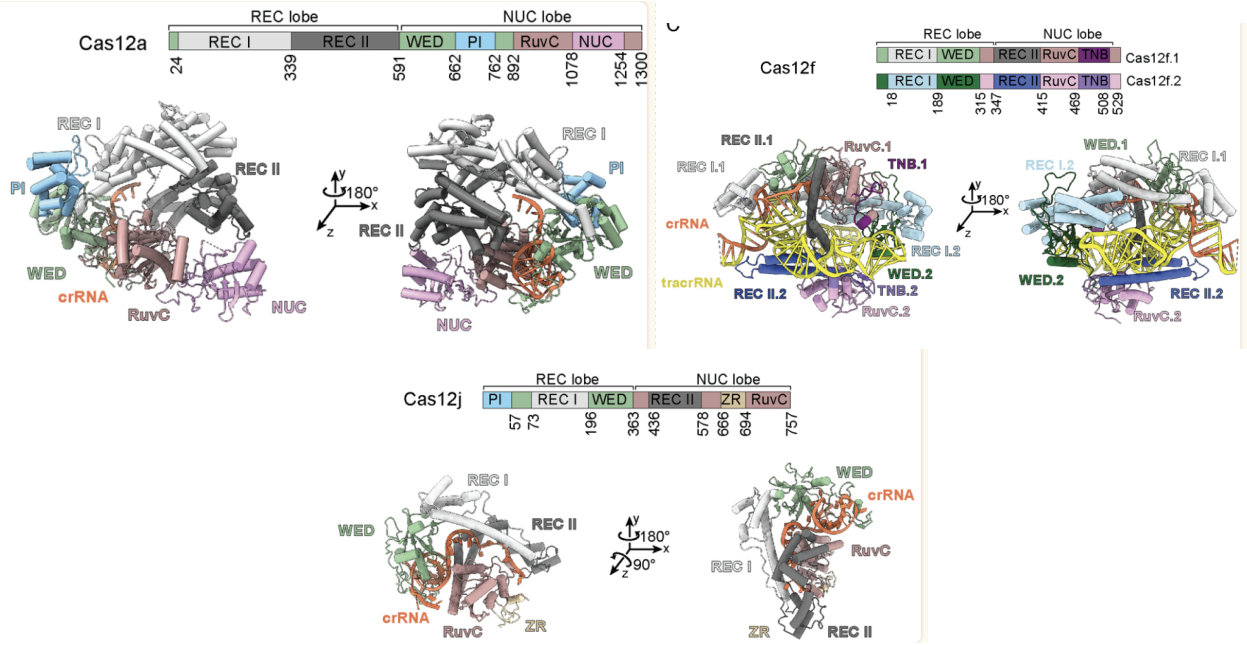


Figure 3: Size comparison between Cas12a, Cas12f and Cas12j proteins with their associated domains. Cas12f interacts with trans-activating crRNA (tracrRNA) and is dimerized to crRNA to form a larger guide RNA (sgRNA) as needed. Recognition lobe to crRNA (REC), nuclease lobes to cut the DNA strand (NUC and RuvC), wedge domain to recognize crRNA stem loops (WED), PAM interacting domain (PI), zinc finger domain (ZR) and target nucleic acid binding domains (TNB) are shown. Images taken from Nguyen, G.T; *et al*.[17]

What was desirable about Cas12j was its similar bilobed structure compared to Cas12a when carrying its crRNA, but at a smaller protein size.[17] Cas12j is a monomer like Cas12a but its reduction in size is due to lacking a NUC domain in the NUC lobe and uses its only RuvC domain to perform DNA cleavage instead.[17] Unlike Cas12f, Cas12j does not require sgRNA at 140 nucleotides in total compared to about 43 nucleotides for crRNA alone.[17] Specifically, Cas12j-8 was selected due to its simplistic and therefore highly targeted PAM TTM (M either being C or A) and was reported for its high cutting activity in HEK293T cells.[18] Previous research results indicated that within regions (loci) of the AAVS1 gene, its cutting activity

outperformed AsCas12a (5 out of 11 loci) and FnCas12a using the same PAM (8 out of 12 loci).[18] FnCas12a and AsCas12a are Cas12a orthologs which originate from different bacterial species. Cas12f is a dimer but its desired small size compared to Cas12a contains two RuvC domains, which cuts each DNA strand sequentially by cutting the non-target strand and then the target strand.[17] The chosen version of Cas12f was based on a study that tested an engineered Cas12f in mammalian cells called CasMINI.[19] The research consisted of a final engineered version of an inactive CasMINI (dCasMINI) fused to a transcriptional activator. Transcription activation was comparable with Cas12a but resulted in a 200-fold increase compared to the Cas12f wildtype, showing its improved DNA binding and specificity.[19] CasMINI also showed an increase in indel percentages compared to the Cas12f wildtype at VEGFA, HBB and IFNɣ gene sites.[19] CasMINI and Cas12j-8 were tested as versions of Cas12f and Cas12j to observe whether their small size and cutting activity could be implemented within inverse prime editing. Eventually the protein focus of this project became an endonuclease-enhanced IscB (enOgeuIscB) (496 aa/1.49 kb), a Cas9 ancestor (Figure 4).[20] Since wildtype IscB is inactive in human cells, enhancement of IscB was done with further protein engineering.[20] The small size of IscB was already attractive since its smaller than Cas12f and Cas12j, but the engineered enhancement increased DNA binding 4 fold in-vitro and cutting activity 30.4 fold ex-vivo compared to wild type in HEK293T cells.[20] More specifically, the enhancement of IscB outperformed its wildtype in cutting activity across 9 different gene sites within multiple loci (26 sites).[20] Within the study's indel percentage data, the cutting was most active at VEGFA, EMX1 and DMNT1 gene sites, which were later chosen as the gene sites to test IscB for cutting activity and inverse prime editing. All studied Cas proteins have been tested in comparison to SpCas9 (Cas9) (1368 aa/4.1 kb) as a positive control.

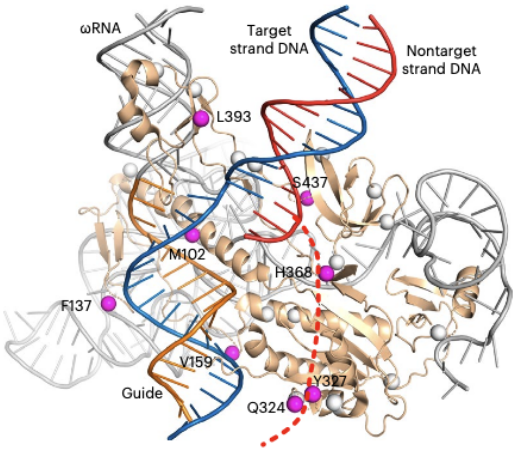


Figure 4: Cryo-EM structure of IscB with manipulated interacting residues to enhance endonuclease efficiency (highlighted in pink spheres). Other residue substitutions are also shown (grey spheres). IscB is encoded within a transposon region (IS200/IS605 family), which are genetic elements that move within the genome.[20,21] IscB uses a longer interactive guide RNA (ωRNA) (shown as a grey strand) and recognizes a TAM (target adjacent motif) as PAM but both still carry out the same functions. The guide in this figure (orange single strand) is the component of the ωRNA that specifically binds to the target strand of DNA (blue single strand). The red dashed line indicates a possible path where the non-target strand (red single strand) will go when entering the RuvC nuclease site.[18] Image taken from Hao, Y; *et al*.[20]

Proposals To Improve Inverse Prime Editing - Separate Guide RNA

The crRNA is used to assist in cutting and originally contained PBS and RTT. To improve the guide RNA design, separating the PBS and RTT regions from a linear crRNA conserves it from degradation as a separate circular RNA (cRNA).[9] Having separate guide RNA splits the cutting and editing roles into crRNA and cRNA respectively. The cRNA contains MS2 stem loops which is a binding site for MS2 coat protein (MCP), a fusion protein tethered with RT. The goal is to test whether cutting and editing can occur in separate events with separate guide RNA along with IscB, instead of Cas9 like previous studies (Figure 5).[9]

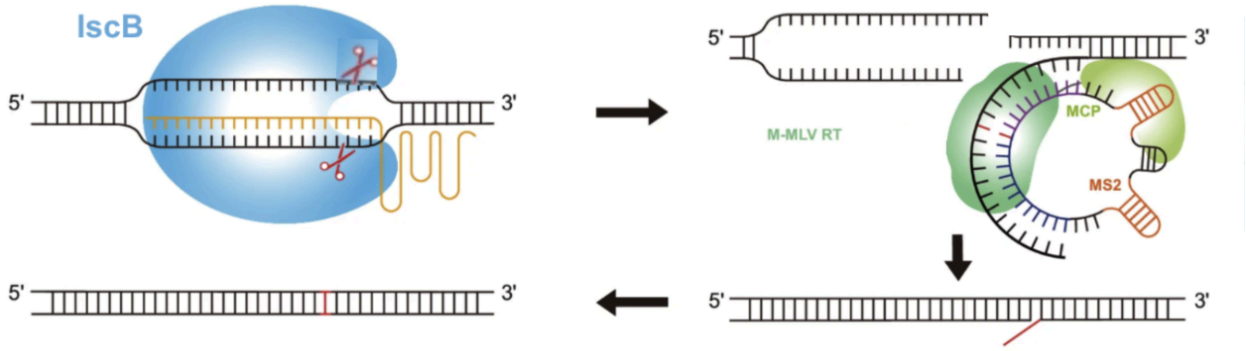


Figure 5: Overview of the proposed inverse prime editing design with IscB and cRNA. Two separate guide RNA occur at separate events to perform inverse prime editing. The cutting process requires crRNA to bind at a specific site complementary to the target strand so IscB could make a double-stranded break (top-left). What follows is incorporating the desired edit by using the RT with MCP and cRNA (top-right). The PBS (blue single strand) on the cRNA binds to the DNA target strand before MCP (with RT) binds to one of the MS2 stem loops. RTT (purple single strand) is within the cRNA and is used as a template for RT to elongate 5' to 3'on the target DNA strand and add the edit (red base) (top-right). The edit is completely incorporated after DNA repair (bottom). Image was taken and edited from Liang, R; *et al*.[9]

The cRNA is circularized using the Tornado system by having ribonuclease sites at both ends of the RNA sequence. The RNA polymerase III U6 promoter is used to transcribe the cRNA as a small non-coding RNA so that it therefore remains in the nucleus. Once transcription is complete, ribonucleases cleave the designated sites via folding and leave a 5' OH and 2', 3'-cyclic phosphate. Both ends come closer together with the help of the ends of the ligation sequences being complementary to each other. Finally an intracellular RNA ligase enzyme such as RTCB ligates both ends together resulting in a cRNA (Figure 6).[22]

Further testing needed to be done to observe whether the proposed adjustments to inverse prime editing were successful in sequential steps while using smaller Cas proteins. Separating the cutting (Cas/IscB protein and crRNA) and editing components (cRNA and RT + MCP) also avoids a complicated expression of both systems if they were tethered together.

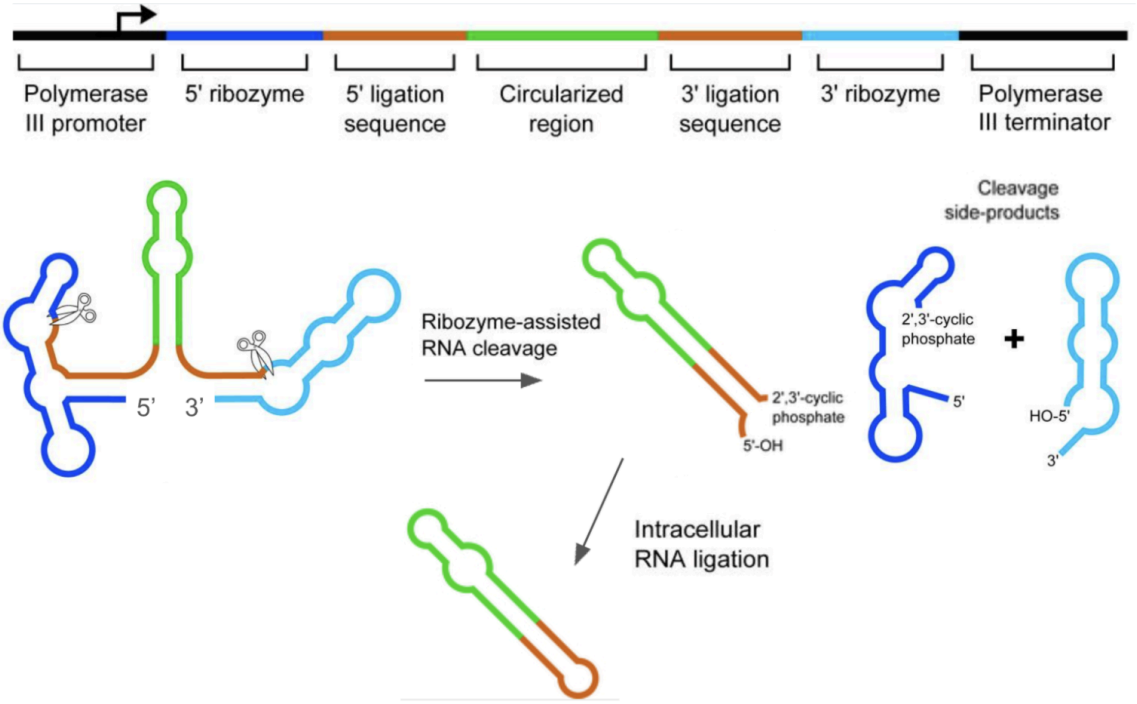


Figure 6: RNA circularization via Tornado system. Configuration of plasmid DNA for transcription (top) was used for the design of the cRNA to participate in the Tornado pathway. Image was taken and edited from Litke, J.L; *et al*.[22]

Experimental Methods

Agarose Gel Electrophoresis

Agarose gels were made with TAE buffer at 1% and stained with ethidium bromide. Electrophoresis conditions were at 140 V, 2.0 A for 30 minutes. A ratio of 4 µl sample to 2 µl of loading dye was used when loading samples.

Passaging HEK293T Cell Stocks and Preparing Cell Plates

HEK293T cell passaging was performed twice weekly to ensure that passaging occurred at around 75% confluency. Under a biosafety cabinet, old media was removed at the corner of the flask before adding PBS buffer, then TrpLe (trypsin solution) directly on the cells. The cell flask was rocked and then incubated for 15 minutes at 37℃. After incubation, the flask was rocked again before adding 8 ml of DMEM with pen-strep (DMEM-P/S) media. The suspended cell solution was pipetted up and down 10 times, dispensing at the bottom of the flask (on top of the cells). About 0.5 ml of the cell solution was left in the flask, while the rest was dispensed into a Falcon tube for cell plate preparation. 15 ml of DMEM-P/S was added to the flask before it was returned to the incubator.

10 µl of the cells from the Falcon tube with 10 µl of Tryptan Blue Dye were mixed so that 10 µl of the cell/dye mixture could be added to each side on a cell-counter slide. Both sides of the slide were inserted into a cell counter machine to record and calculate the average cell count of both reads. To ensure that there are 22,000 cells of 100 µl per well on a 96-well plate, totaling 12 ml of solution with DMEM (no P/S), the volume of cell stock added was calculated;

Volume of Cell Stock Added (ml) = $\frac{(220{,}000\ cells/ml\ ) \times (12\ ml)}{Average\ Cell\ Count/ml}$

Volume of DMEM (No P/S) Buffer (ml) = 12 ml - Volume of Cell Stock Added (ml)

The cell plate was left to incubate overnight before starting transfection.

Transfection

Transfection methods followed protocols from similar research in literature.[9] Each transfection requires prepared solutions consisting of a protein plasmid (Cas/IscB at 100 ng and/or MLV-RT + MCP at 50 ng), guide-RNA plasmids (50 ng of crRNA and/or 50 ng of cRNA) and Optimem, totaling 5 µl.[9] Approximate plasmid sizes are as follows: SpCas9 (4.1 kb), CasMINI (1.6 kb),

Cas12j-8 (2.3 kb), enOgeuIscB (1.5 kb), MLV-RT + MCP (5.9 kb), crRNA (2.7 kb) and cRNA (3.3 kb). LNP for cutting samples only contain the protein plasmid and crRNA plasmid. Inverse prime editing samples contain all four plasmids in a single LNP for simultaneous delivery. Based on the Invitrogen Lipofectamine 2000 protocol, an 8 µl solution was also prepared with a ratio of 88% Optimem (7.04 µl) and 12% lipofectamine (0.96 µl) at room temperature. Both solutions were mixed and incubated at room temperature for 15 minutes. 10 µl of the mixed solution was added in the well on the cell plate, ensuring not to hit the bottom of the well, before incubating for three days.

Genomic DNA Harvesting

Genomic DNA was harvested following New England Bio-Lab's and GeneBio's protocol.

PCR Amplification of Genomic DNA

PCR reaction mixtures total 25 µl and consist of 12.5 µl of PCR Master Mix, 10 µM of forward primer, 10 µM of reverse primer, 100 ng of genomic DNA and the remaining with Milli-Q water. Thermalcycler steps follow the GeneBio PCR Master Mix protocol. After amplification, clean-up was followed by the supplier's protocol before measuring concentration via Nanodrop. Cleaned samples and 1 kb Plus DNA ladder were run on an agarose gel and imaged to ensure the remaining band was the appropriate size.

Blunt-End Ligation Cloning

A reaction mixture was made consisting of 5 µl Milli-Q water, 1 µl of 10x T4 ligase buffer, 3 µl of PCR product and 1 µl each of T4PNK (10 kU/ml), DpnI (20 kU/ml) and T4 ligase enzyme (400 kU/ml). After incubation for 1 hour, transformation was performed with DH5-α competent cells following a protocol from New England Biolabs. The transformed cells were plated on 1x carbenicillin LB-agar plates and incubated overnight at 37℃.

Golden-Gate Cloning

Forward and reverse insert oligos were mixed at 100 µM each and were annealed by heating the mixture at 95℃ then cooled to room temperature. A reaction mixture was made consisting of 2 µl of 10x T4 ligase buffer, 50 ng of backbone plasmid, 1 µl of 1:25 diluted insert oligos, 2 µl of

T4PNK enzyme (10 kU/ml), 1 µl of BsmBI restriction enzyme (10 kU/ml), 2 µl of T4 ligase (400 kU/ml) enzyme and filled with Milli-Q water to total 20 µl for the total solution. PCR was performed using the following thermocycler process in its respective order; 37℃ for 5 minutes, 16℃ for 10 minutes (repeat to first step ten times), 37℃ for 15 minutes, 80℃ for 5 minutes. Transformation was performed with DH5-α competent cells following its protocol from New England Biolabs. The transformed cells were plated on 1x carbenicillin LB-agar plates and incubated overnight at 37℃ .

Mini-Prep

Preparing bacterial cultures for plasmids was done following the GeneBio protocol.

Flow Cytometry

Cell Preparation: Media was removed before cells were washed with 150 µl of 1x PBS buffer per well. PBS was removed after 5 minutes of incubation. For each well on a 92-well plate, 30 µl of TrypLE was added and incubated at 37℃ for 15 minutes. Cells were resuspended in 150 µl of Optimem and were pipetted up and down to break cell clumps before it was transferred to a U-bottom flow cytometer plate.

Measurement: Readings were analyzed following the Cytek Biosciences protocol.

Sanger Sequencing

Preparing solutions for Sanger sequencing shipment was done following the TCAG protocol to confirm plasmid sequences. Chromatograms from Sanger sequencing results were compared with wildtype (non-transfected) chromatogram samples through DECODR v3.0, a decoding computer software to obtain indel percentage results and determine whether the indels caused a frameshift mutation.

Next Generation Sequencing (NGS)

Preparing solutions for NGS was done following the Azenta Life Sciences and ISEQ protocol. Sequencing results were read using the CRISPResso2 NGS reading program. Indel percentages were calculated following the equation; $\text{Indel } \% = 100 \times \frac{Insertions + Deletions}{Reads\ Aligned}$

Results and Discussion

Testing Cas12 Subtypes: Transfection Efficiency Using Flow Cytometry

To confirm successful co-transfection of multiple plasmids to HEK293T cells, the following plasmids were transfected at once; blue fluorescent protein (BFP) plasmid, Cas12a protein plasmid with mCherry and its associated crRNA plasmid. Each crRNA is designed to guide the Cas protein to cut the BFP plasmid around its chromophore region (amino acid 66) to let NHEJ repair effectively turn-off blue fluorescence without solely relying on a frameshift.[23] The positive control consists of a BFP plasmid with a mCherry reporter plasmid (without Cas) and a non-targeting control plasmid to balance the amount of DNA transfected. Fluorescence from BFP and mCherry were measured from across multiple Cas and crRNA combinations indicating successful co-transfection. Also, a decrease in BFP along with Cas12a compared to the positive control could also indicate BFP plasmid being cut by Cas12a within the same assay. However, further testing needed to be done to look further into how effective the cutting activity really is between all Cas12 subtypes. High transfection efficiency was not a priority at this stage since the purpose of the experiment was to only ensure if transfection was successful (Figure 7).

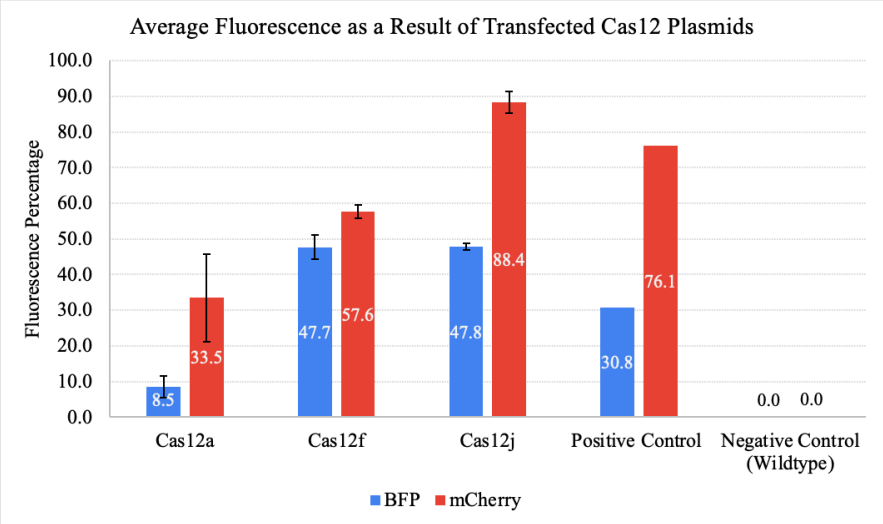


Figure 7: Average fluorescence percentages between BFP and mCherry-Cas12 subtypes post-transfection. Error bars represent the standard error across the triplicate dataset for each Cas12 protein, except Cas12f (duplicate). Each crRNA guided Cas12 within the BFP chromophore region. The positive control contains a shorter mCherry reporter plasmid without crRNA and Cas12 plasmids. The negative control (wildtype) is without any reporter plasmids to ensure non-contamination.

Observing Cutting Activity Through Sanger Sequencing

When DNA has a double strand break it undergoes NHEJ by introducing insertions and deletions (indels). Therefore, an increase in indel percentage indicates an increase in cutting activity.[24] Gene sites were chosen based on research from previous studies in the literature that reported high indel percentages. To ensure consistency, data was reproduced from previously tested sites.[18, 19, 25] Also, Cas12a targeted sites were tested using Cas12f and Cas12j to observe any differences.[25] Cutting experiments comprise of transfecting the Cas protein plasmid with the corresponding crRNA plasmid in HEK293T kidney cells. Once transfection is complete, DNA is extracted, the desired gene site is amplified via PCR and the purified sample is sent off for Sanger sequencing. Sanger sequencing quantifies the percentage of the DNA population that contains indels, indicating cutting activity. Each cutting sample is sent along with a wild-type sample without transfection as a negative control to ensure there is no contamination within the same cell plate. For most of the gene sites tested, cutting activity was not observed when using the smaller Cas12 subtypes (Table 1). However, where cutting activity was observed, Cas12f and Cas12j outperformed other Cas12a cut sites. Compared to Cas9 as a positive control with an indel percentage at 17.4%, the indel percentage from Cas12f and Cas12j were low ranging from 2.1% - 6.9% (Figure 8).

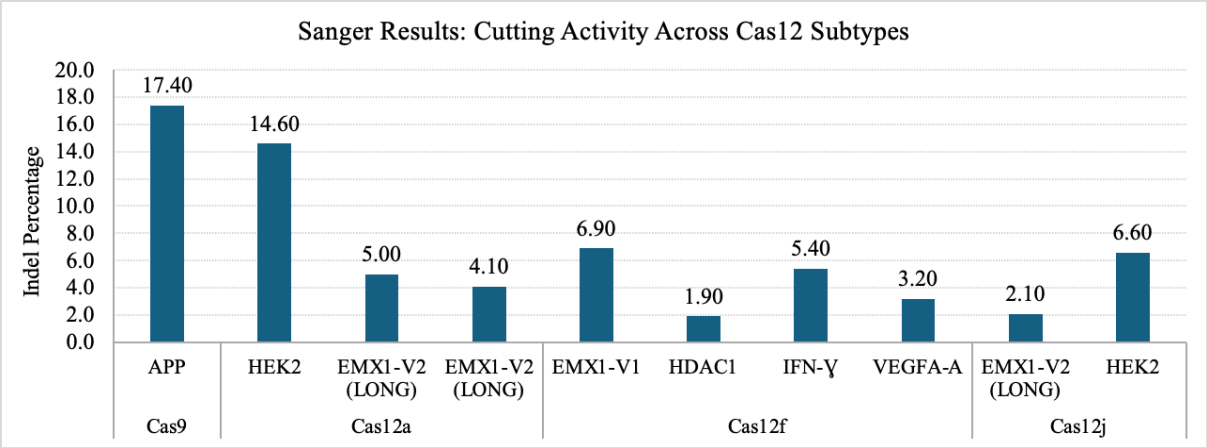


Figure 8: Cutting activity of Cas proteins shown as indel percentage data after Sanger sequencing across various gene sites. Cas12 subtypes were tested with Cas9 as a positive control. The results shown are the highest indel percentages after performing the experiments at least twice. All wildtype samples showed 0% indel indicating non-contamination.

Although previous studies had reported otherwise, there were gene sites where cutting activity was not observed.[18, 19] IscB protein was later introduced to observe if there would be any changes in cutting activity. Initially most of the no-cut results had a sample quality too low to measure for sequencing. Troubleshooting methods were done to promote sequencing reads such as increasing the sample concentration for lower quality DNA and using longer amplicons to ensure the entire gene site was sequenced without primers binding elsewhere. Also, DNA harvesting and PCR cleanup kits were replaced with new ones, which improved the yield of DNA but resulted in most samples initially having a low-quality non-read to a concluded no-cut read (Table 1).

Table 1: Gene sites tested with Cas12a (A), Cas12f (F), Cas12j (J) and IscB (I) with their corresponding crRNA number. Red boxes indicate no cut (no indels) after sequencing, and yellow boxes indicate low quality for measurement. Longer amplicons (LONG) were tested for some sites to assist amplification. Low-quality results were repeated at least twice to conclude a final result.

Gene	Gene Length (Base Pairs)	Cas Protein/crRNA
FANCF-B, FANCF (LONG)	300, 508	A89
		F96
		J104
DMNT1-A	300	A29
		F90
		J97
		J98
EMX1-V1, EMX1-V2 (LONG)	277, 320	F92
EMX1-B	301	J163
AAVS1	300	A82
		F94
AAVS1-B	342	J162
HDAC1	300	A84
		J103
VEGFA-B	307	I-168
EMX1-C	281	I-169
DMNT1-C	298	I-170

Observing Cutting Activity Through NGS Sequencing

Since troubleshooting for Sanger sequencing had not improved, next generation sequencing (NGS) was done instead, except only for Cas12f and IscB samples. Cas12f samples were tested since there was more consistent cutting activity shown within the Sanger sequencing results compared to the other Cas12 subtypes. IscB samples were also measured due to it being the shortest plasmid size compared to the other protein plasmids. The effect of small plasmid size on cell delivery efficiency was to be tested within a high throughput sequencing method like NGS. Cas12f NGS results had a significant decrease in indel percentage compared to Sanger results, even when using some of the same samples (about a 4 fold decrease at IFN-γ and 4.5 fold decrease at VEGFA) (Figure 8)(Figure 9). However, IscB samples showed a substantial indel percentage compared to the positive control using NGS (10.25% at DMNT1-C compared to 13.4% with Cas9 at APP) compared to having no indel reading at all when using Sanger (Figure 9) (Table 1). What stayed consistent between both sequencing methods was the indel percentage of Cas9 as the positive control (17.4% with Sanger and 13.4% with NGS).

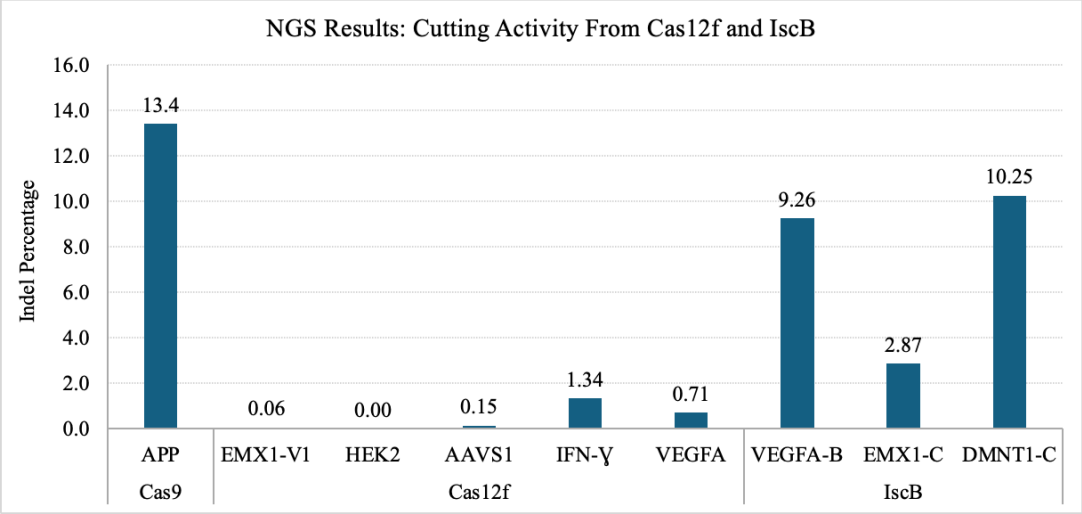


Figure 9: Cutting activity of Cas12f and IscB proteins shown as indel percentage data after NGS across various gene sites. Cas12f and IscB were tested with Cas9 as a positive control. All wildtype samples showed 0% indel indicating non-contamination.

With IscB outperforming Cas12f in cutting activity, IscB sites were tested twice more and analyzed via NGS to ensure if the results were consistent. Average indel percentages from NGS results showed that IscB had consistent cutting activity within the same gene sites (Figure 10).

IscB became the protein of interest tested against VEGFA-B, EMX1-C and DMNT1-C gene sites to replicate previous cutting activity observations and test inverse prime editing moving forward, analyzed via NGS.[20]

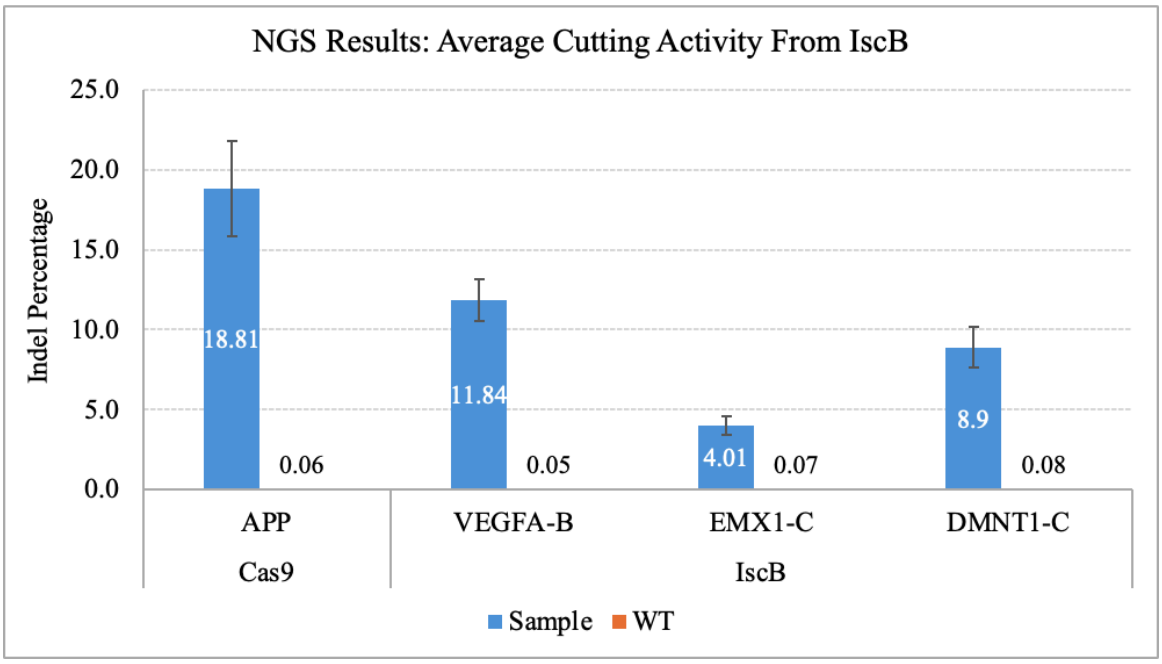


Figure 10: Average cutting activity of IscB shown as indel percentage within a triplicate dataset via NGS across various gene sites with Cas9 as a positive control. Average wildtype (WT) indel percentage being approximately zero indicates non-contamination within the samples. Error bars represent the standard error across the triplicate dataset. NGS data was obtained using Azenta Life Sciences (Azenta) primers and sequencing service.

Inverse Prime Editing Experimental Design

Consistent cutting activity from IscB and crRNA combinations were tested to see if it could contribute to inverse prime editing. RT + MCP and cRNA plasmids were additionally added within the same LNP to ensure simultaneous delivery of all components into the cell during transfection.[26] For each gene site there are two samples of inverse prime editing (edits of a codon transversion and a TAA stop codon insertion), one sample for prime editing (codon transversion), a positive control to test for cutting activity only (only IscB and crRNA plasmids are transfected) and a wildtype negative control to ensure non-contamination (no plasmids transfected).

NGS Sample Preparation

Before submitting NGS samples for sequencing, two PCR runs need to occur. The first PCR run amplifies the desired gene site on genomic DNA, post transfection and incubation, resulting in single stranded NGS tags at the ends of the amplicon. The second PCR run uses primers that bind to the NGS tags that give it a unique barcode so that multiple samples could be sequenced at once and could be distinguished within a pool of NGS samples. The samples will then be combined and gel extracted for complete preparation. Initially samples were sent to Azenta's NGS service and were prepared using Azenta's PCR primers. Once our lab acquired its own NGS sequencer, other PCR primers (ISEQ) were used to prepare the samples instead.

Inverse Prime Editing Results: Comparing Azenta and ISEQ Primers

Table 2: NGS results using Azenta primers across the three targeted gene sites with IscB. Samples only containing IscB and crRNA plasmids represent the positive control. "Approx. Total Reads" is the approximate total number of reads the NGS sequencer analyzed. "Percent Reads Aligned" is the percentage of reads having a sequence that matched the spacer sequence (editing region). "Precise Edit Reads" also shows the percentage of successful edited reads compared to the total number of reads. An expanded version could be seen on Table 5 in Appendix B.

NGS Run	Gene	Approx. Total Reads	Percent Reads Aligned	Percent Indel From Positive Control	Precise Edit Reads
NGS 10	VEGFA-B	5400 - 21,150	40% - 65%	2.43%	0
	EMX1-C	4000 - 16,000	93% - 97%	0.67%	0
	DMNT1-C	8000 - 12,200	96% - 98%	2.22%	65 (0.57%)
NGS 11	VEGFA-B	32,700 - 63,000	70% - 80%	2.64%	0
	EMX1-C	74,500 - 115,430	97.8% - 98.3%	0.10% - 4%	99 (0.43%)
	DMNT1-C	43,130 - 55,500	86.4% - 92%	1.4% - 6%	120 (0.29%)
NGS 12	VEGFA-B	11,000 - 14,000	2.2% - 10.3%	7.74%	28 (2.23%), 8 (0.61%)
	EMX1-C	5500 - 7500	58.9% - 81.2%	0%	0
	DMNT1-C	11,000 - 14,000	0.10%	0%	0

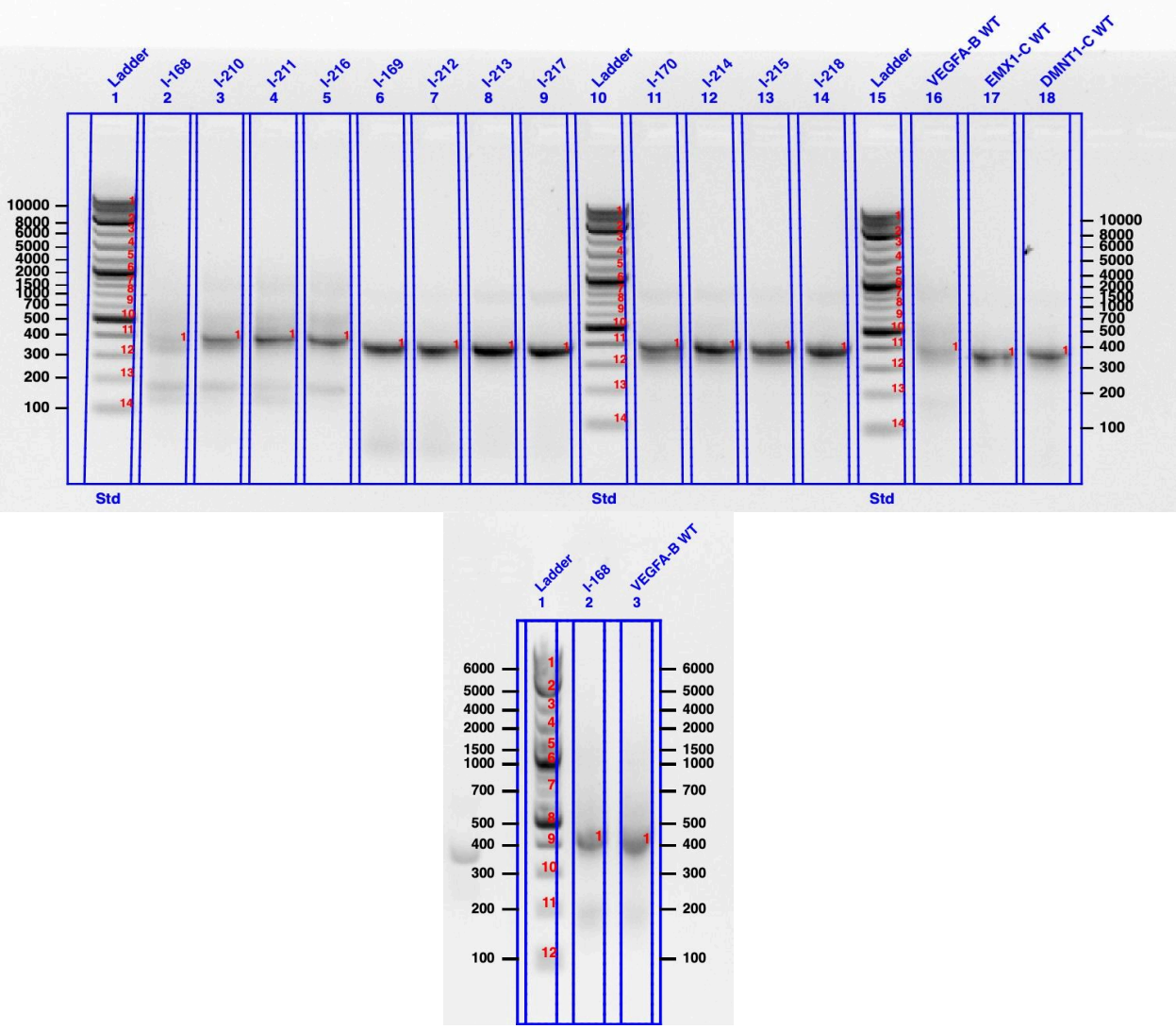
Table 3: NGS results using ISEQ primers across the three targeted gene sites with IscB. Samples only containing IscB and crRNA plasmids represent the positive control. "Approx. Total Reads" is the approximate total number of reads the NGS sequencer analyzed. "Percent Reads Aligned" is the percentage of reads having a sequence that matched the spacer sequence (editing region). "Precise Edit Reads" also shows the percentage of successful edited reads compared to the total number of reads. An expanded version could be seen on Table 6 in Appendix B.

NGS Run	Gene	Approx. Total Reads	Percent Reads Aligned	Percent Indel From Positive Control	Precise Edit Reads
NGS 14	VEGFA-B	60,000 - 88,000	12.3% - 22.4%	0%	0
	EMX1-C	21,000 -78,000	22.6% - 89.9%	1.18%	0
	DMNT1-C	69,000 - 77,000	73% - 88.4%	4.54%	0
NGS 15	VEGFA-B	103,000 - 134,000	3.8% - 9.1%	0.00%	0
	EMX1-C	78,000 - 118,000	9.4% - 18.4%	0.60%	0
	DMNT1-C	49,000 - 139,000	5.1% - 19.2%	2.75%	0
NGS 17 (NGS 12 Repeat)	VEGFA-B	23,000 - 26,000	0.14% - 1.22%	15.15%	4 (2.25%), 1 (0.36%)
	EMX1-C	22,000 - 29,000	0.14% - 4.98%	1.51%	0
	DMNT1-C	21,000 - 23,000	0.73% - 9.06%	5.76%	0
NGS 17 (2)	VEGFA-B	28,000 - 34,000	0.09% - 7.44%	0.00%	0
	EMX1-C	31,000 - 40,000	0.47% - 3.08%	0.79%	0
	DMNT1-C	27,000 - 35,000	0.13% - 3.43%	1.33%	0
NGS 17 (3)	VEGFA-B	34,000 - 45,000	1.83% - 10.61%	2.67%	0
	EMX1-C	43,000 - 49,000	0.10% - 8.72%	0%	0
	DMNT1-C	37,000 - 46,000	0.42% - 2.76%	0.97%	0

NGS runs using Azenta primers showed a lower number of reads but a higher percent reads aligned compared to using ISEQ primers. Even though the total reads using ISEQ were high, a low reads aligned percentage doesn't produce accurate data. The same genomic DNA for NGS 12 using Azenta primers was tested using ISEQ primers to observe whether the edit between both results would be present, without there being a variable in genomic DNA quality. The edit was still present using ISEQ primers indicating that genomic DNA quality was not a factor for the difference in NGS data between Azenta and ISEQ primers.

PCR Quality Using Azenta and ISEQ PCR-2 Primers

When observing PCR-2 bands on an agarose gel, the Azenta samples only contained the desired product band when the percent reads aligned was high (70% - 98.3%), while ISEQ samples all contained small intense byproducts with faint desired products (Figure 11) (Figure 12).



<u>Figure 11:</u> NGS 11 PCR-2 samples are shown due to the highest reads aligned percentage in Table 2. The desired product is about 400 base pairs for each gene site. Lanes are labelled according to guide RNA number and wildtype (WT). Samples are grouped by gene site in groups of 4 showing VEGFA-B (I168 - I216), EMX1-C (I169 - I217) and DMNT1-C (I170 - I218) respectively. PCR-2 was re-done for samples that had bands being too faint (bottom). Identification of each lane could be seen on Table 5 in Appendix B.

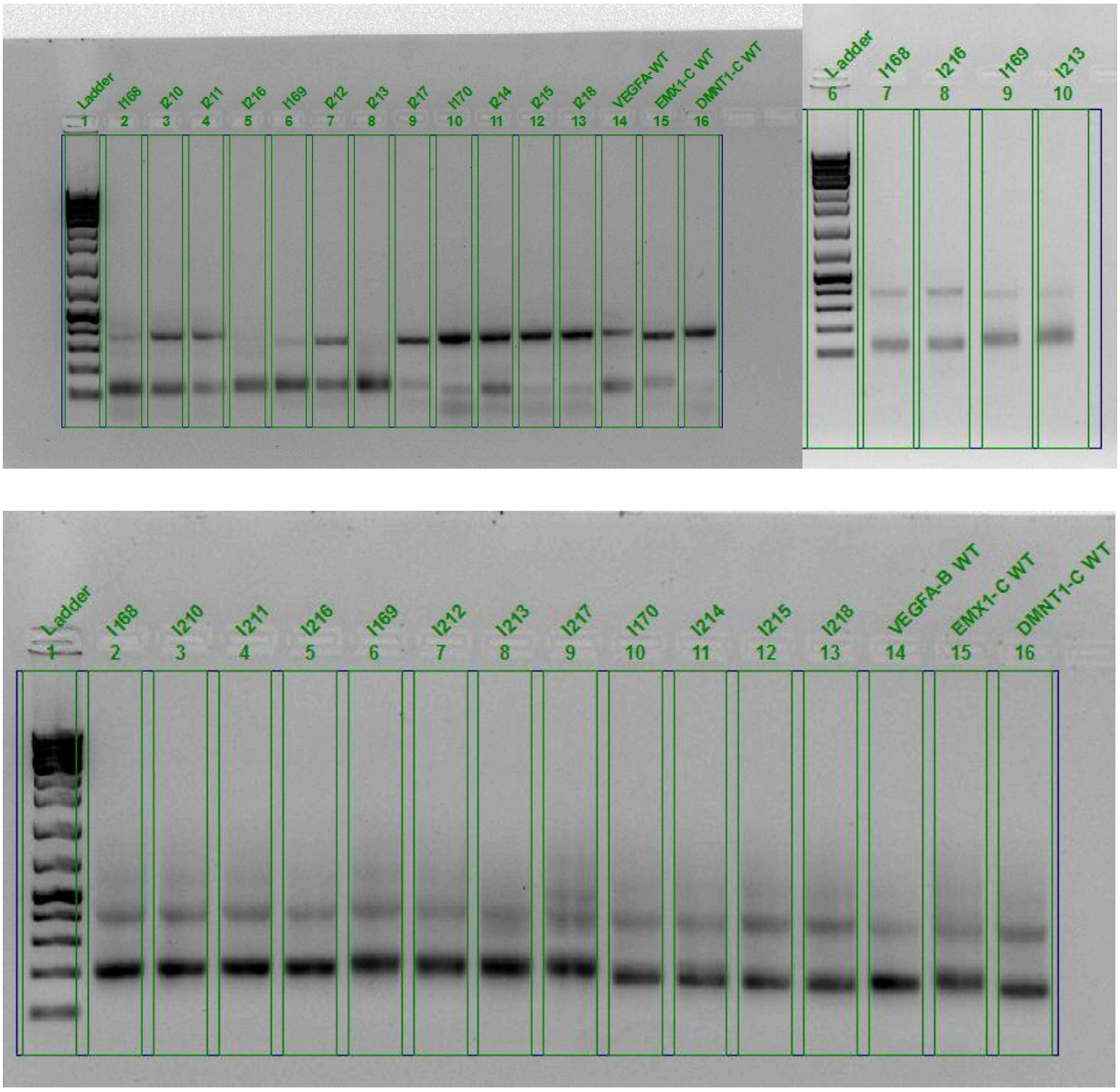


Figure 12: NGS 17 (NGS 12 Repeat) of PCR-1 (above) and PCR-2 (below) samples are shown due to the lowest reads aligned percentage in Table 3. The desired product is about 400 base pairs for each gene site (fourth band from the bottom on the ladder). Lanes are labelled according to guide RNA number and wildtype (WT). Samples are grouped by gene site in groups of 4 showing VEGFA-B (I168 - I216), EMX1-C (I169 - I217) and DMNT1-C (I170 - I218) respectively. Identification of each lane could be seen on Table 6 in Appendix B.

Troubleshooting steps were done to reduce intense off-target byproduct bands in PCR-2 by adjusting PCR conditions, such as using a gradient elongation temperature (55℃ - 72℃),

reducing the quantity of template DNA volume by half and cleaning PCR 1 samples, but the result of having intense short byproducts remained the same. It was thought that gel extracting the desired band at the 400 bp region would avoid issues with the intense byproducts below them. However, after viewing the NGS sequences of NGS 17 (NGS 12 Repeat) using a NGS sequence viewer program, all the sequences started with the correct sequence for about the first 25 nucleotides then became random downstream, even while the whole DNA library had a high base quality (Figure 13). A possibility for the cause of random sequences downstream is that the intense byproduct bands are affecting the desired product band during the PCR process. What could have happened was that the denaturing step first separated DNA strands for the short byproducts and the desired product. Once the annealing step occurs, one strand from each separate product anneals to each other forming an entirely different sequence, leading to very low reads aligned percentages.



Figure 13: A subset of NGS reads sequences for the VEGFA-B gene site from NGS 17 (NGS 12 Repeat) using NGS analysis software. The black box is the spacer sequence that should be unchanged (where crRNA binds). The red box is the edit region (can either be edited as a codon transversion to GCA or TAA insertion at the start of the codon). The percent reads aligned value is based on the presence of the sequence within the entire black box without the edit. Other sequences above and below the highlighted sequence of interest are assumed to be random sequences as a result of PCR byproducts.

Although at a low number of reads, multiple precise inverse prime edits were present showing that the inverse prime editing design with IscB and cRNA is possible (0.29% - 2.23% using Azenta and 0.36% - 2.25% using ISEQ). Using the same method did not show any precise edits for prime editing however, if the approximate total reads using Azenta were as high as ISEQ the true number of precise edit reads could possibly increase (Table 2) (Table 3).

Conclusions and Future Work

NGS results showed that inverse prime editing is possible using IscB with crRNA and cRNA in separate cutting and editing steps. The issue of either having lower reads with high reads aligned using Azenta or high reads with low reads aligned using ISEQ doesn't fully express the true number of edited reads. Future work needs to be done by improving the ISEQ PCR-2 primer design so that byproducts would not form, leading to cleaner desired products with a higher reads aligned percentage. Primers from the sequencer manufacturer could also be used to rule out any compatibility issues with the sequencer but could result in a lower amount of samples that could be sequenced at a time. Although decreasing the number of samples within the NGS library reduces crowding on the flow cell and therefore could improve the read accuracy. Overcrowding the flow cell causes an increase in mixed signals which reduces quality scores.[27] From there what needs to be focussed on is the consistency of indel percentages from the positive controls between NGS runs (Figure 10)(Table 5)(Table 6).

The cutting activity for editing and positive control samples need to be observed on whether cutting activity is affected according to having the extra RT + MCP and cRNA plasmids competing to pack within the same LNP. Further assays could be done to observe how the plasmids interact within LNP packaging, transfection and transport to the nucleus by using covalent fluorophores, similar to previous studies mentioned.[10,11,12,13] Once the activity between multiple plasmids is further understood the ratio between plasmids can be adjusted when packaging the LNP.

Other common molecular biology techniques could be done to visually observe how effective IscB transportation is to the nucleus. To preserve protein size, antibodies could be used. A primary antibody would bind to IscB while a second antibody containing a dye as a visual indicator would bind to the primary antibody. Another method would be to just encode green fluorescent protein (GFP) within the IscB plasmid but this results in increasing the size of an already desired smaller protein plasmid.

Long-term research could discover how well other small Cas subtypes are within this inverse prime editing design. As research progresses in literature, IscB itself could be further manipulated into a new and more improved version to increase cutting activity and promote the editing process. For example, a modified IscB linked to a separate DNA binding domain could be used to improve DNA recognition and increase cutting activity.[27] Further modifications to

IscB could be done by mutating one of its RuvC DNA-cutting domains, turning off its cutting capabilities on one of the DNA strands, effectively turning it into a nickase. Preventing double-stranded breaks can benefit further during in-vivo studies to reduce the likelihood of genetic toxicity that double stranded breaks may cause.[29]

Regardless of how inverse prime editing is improved next, what is most important is that the research continues. The evolution of genome editing leading to CRISPR already impacted the agriculture and therapeutics industry at its infancy, but improving inverse prime editing could unlock CRISPR's true potential of incorporating desired sequences that can benefit others.[1,2] Therapies would not have to rely on the specificity of drugs with the caveat of harmful side effects. Food could be more sustainable by increasing resistance in plants from environmental factors. Inverse prime editing could further contribute to such causes that all start from editing the genome with increased accessibility.

Bibliography

(1) Martín-Valmaseda, M.; Devin, S. R.; Ortuño-Hernández, G.; Pérez-Caselles, C.; Mahdavi, S. M. E.; Bujdoso, G.; Salazar, J. A.; Martínez-Gómez, P.; Alburquerque, N. CRISPR/Cas as a Genome-Editing Technique in Fruit Tree Breeding. *Int J Mol Sci* **2023**, *24* (23), 16656. https://doi.org/10.3390/ijms242316656.

(2) FDA Approves First Gene Therapies to Treat Patients with Sickle Cell Disease. *U.S Food and Drug Administration*. August 12, 2023. https://www.fda.gov/news-events/press-announcements/fda-approves-first-gene-therapies-treat-patients-sickle-cell-disease.

(3) Gaj, T.; Gersbach, C. A.; Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering. *Trends in Biotechnology* **2013**, *31* (7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004.

(4) Adli, M. The CRISPR Tool Kit for Genome Editing and Beyond. *Nat Commun* **2018**, *9* (1), 1911. https://doi.org/10.1038/s41467-018-04252-2.

(**5**) Liu, M.; Zhang, W.; Zhang, H. Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *Int. J. Mol. Sci.* **2020**, *21* (18), 6461. DOI: 10.3390/ijms21186461.

(6) Komor, A.; Kim, Y.; Packer, M.; Zuris, J.; Liu, D. Programmable Editing of a Target Base in Genomic DNA without Double-Stranded DNA Cleavage. *Nature* **2016**, *533*, 420–424. https://doi.org/doi:10.1038/nature17946.

(7) Anzalone, A.; Randolph, P.; Davis, J.; Sousa, A.; Koblan, L.; Levy, J.; Chen, P.; Wilson, C.; Newby, G.; Raguram, A.; Liu, D. Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA. *Nature* **2019**, *576*, 149–157. https://doi.org/10.1038/s41586-019-1711-4.

(8) Bill Kim, Y.; Pierce, E. B.; Brown, M.; Peterson, B. A.; Sanford, D.; Fear, J.; Nicholl, D.; San Pedro, E.; Reynolds, G. M.; Hunt, J. E.; Schwark, D. G.; Jali, S.; Graham, N.; Cesarz, Z.; Chapman, T. A. L.; Watts, J. M.; Hummel, A. W. A Novel Mechanistic Framework for Precise Sequence Replacement Using Reverse Transcriptase and Diverse CRISPR-Cas Systems. *bioRxiv* **2022**, 2022.12.13.520319. https://doi.org/10.1101/2022.12.13.520319.

(9) Liang, R.; Wang, S.; Cai, Y.; Li, Z.; Li, K. M.; Wei, J.; Sun, C.; Zhu, H.; Chen, K.; Gao, C. Circular RNA-Mediated Inverse Prime Editing in Human Cells. *Nat Commun* **2025**, *16* (1), 5057. https://doi.org/10.1038/s41467-025-59120-7.

(10) Lukacs, G. L.; Haggie, P.; Seksek, O.; Lechardeur, D.; Freedman, N.; Verkman, A. S. Size-Dependent DNA Mobility in Cytoplasm and Nucleus. *J Biol Chem* **2000**, *275* (3), 1625–1629. https://doi.org/10.1074/jbc.275.3.1625.

(11) Dauty, E.; Verkman, A. S. Actin Cytoskeleton as the Principal Determinant of Size-Dependent DNA Mobility in Cytoplasm: A NEW BARRIER FOR NON-VIRAL GENE DELIVERY *. *Journal of Biological Chemistry* **2005**, *280* (9), 7823–7828. https://doi.org/10.1074/jbc.M412374200.

(12) Kulkarni, J. A.; Cullis, P. R.; Van Der Meel, R. Lipid Nanoparticles Enabling Gene Therapies: From Concepts to Clinical Utility. *Nucleic Acid Therapeutics* **2018**, *28* (3), 146–157. https://doi.org/10.1089/nat.2018.0721.

(13) Varkouhi, A. K.; Scholte, M.; Storm, G.; Haisma, H. J. Endosomal Escape Pathways for Delivery of Biologicals. *J Control Release* **2011**, *151* (3), 220–228. https://doi.org/10.1016/j.jconrel.2010.11.004.

(14) Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; Koonin, E. V.; Zhang, F. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **2015**, *163* (3), 759–771. https://doi.org/10.1016/j.cell.2015.09.038.

(15) Harrington, L. B.; Burstein, D.; Chen, J. S.; Paez-Espino, D.; Ma, E.; Witte, I. P.; Cofsky, J. C.; Kyrpides, N. C.; Banfield, J. F.; Doudna, J. A. Programmed DNA Destruction by Miniature CRISPR-Cas14 Enzymes. *Science* **2018**, *362* (6416), 839–842. https://doi.org/10.1126/science.aav4294.

(16) Pausch, P.; Al-Shayeb, B.; Bisom-Rapp, E.; Tsuchida, C. A.; Li, Z.; Cress, B. F.; Knott, G. J.; Jacobsen, S. E.; Banfield, J. F.; Doudna, J. A. CRISPR-CasΦ from Huge Phages Is a Hypercompact Genome Editor. *Science* **2020**, *369* (6501), 333–337. https://doi.org/10.1126/science.abb1400.

(17) Nguyen, G.T.; Dhingra, Y.; Sashital, D. G. Miniature CRISPR-Cas12 Endonucleases – Programmed DNA Targeting in a Smaller Package. *Curr Opin Struct Biol* **2022**, *77*, 102466. https://doi.org/10.1016/j.sbi.2022.102466.

(18) Wang, Y.; Qi, T.; Liu, J.; Yang, Y.; Wang, Z.; Wang, Y.; Wang, T.; Li, M.; Li, M.; Lu, D.; Chang, A. C. Y.; Yang, L.; Gao, S.; Wang, Y.; Lan, F. A Highly Specific CRISPR-Cas12j Nuclease Enables Allele-Specific Genome Editing. *Science Advances* **2023**, *9* (6), eabo6405. https://doi.org/10.1126/sciadv.abo6405.

(19) Xu, X.; Chemparathy, A.; Zeng, L.; Kempton, H. R.; Shang, S.; Nakamura, M.; Qi, L. S. Engineered Miniature CRISPR-Cas System for Mammalian Genome Regulation and Editing. *Molecular Cell* **2021**, *81* (20), 4333-4345.e4. https://doi.org/10.1016/j.molcel.2021.08.008.

(20) Hao Yan; Xiaoqing Tan; Siyuan Zou; Yihong Sun; Ailong Ke; Weixin Tang. Assessing and Engineering the IscB–ωRNA System for Programmed Genome Editing. *Nat Chem Biol* **2024**, *20*, 1617–1628. https://doi.org/10.1038/s41589-024-01669-3.

(21) Siguier, P.; Gourbeyre, E.; Chandler, M. Bacterial Insertion Sequences: Their Genomic Impact and Diversity. *FEMS Microbiol Rev* **2014**, *38* (5), 865–891. https://doi.org/10.1111/1574-6976.12067.

(22) Litke, J. L.; Jaffrey, S. R. Highly Efficient Expression of Circular RNA Aptamers in Cells Using Autocatalytic Transcripts. *Nat Biotechnol* **2019**, *37* (6), 667–675. https://doi.org/10.1038/s41587-019-0090-6.

(23) Heim, R.; Prasher, D. C.; Tsien, R. Y. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proc Natl Acad Sci U S A* **1994**, *91* (26), 12501–12504. https://doi.org/10.1073/pnas.91.26.12501.

(24) Anzalone, A.; Koblan, L.; Liu, D. Genome Editing with CRISPR–Cas Nucleases, Base Editors, Transposases and Prime Editors. *Nature Biotechnology* **2020**, *38*, 824–844. https://doi.org/10.1038/s41587-020-0561-9.

(25) Liang, R.; He, Z.; Zhao, K.; Zhu, H.; Hu, J.; Liu, G.; Gao, Q.; Liu, M.; Zhang, R.; Qiu, J. L.; Gao, C. Prime Editing Using CRISPR-Cas12a and Circular RNAs in Human Cells. *Nat Biotechnol* **2024**. https://doi.org/10.1038/s41587-023-02095-x.

(26) Qiu, M.; Glass, Z.; Chen, J.; Haas, M.; Jin, X.; Zhao, X.; Rui, X.; Ye, Z.; Li, Y.; Zhang, F.; Xu, Q. Lipid Nanoparticle-Mediated Codelivery of Cas9 mRNA and Single-Guide RNA Achieves Liver-Specific in Vivo Genome Editing of Angptl3. *Proc Natl Acad Sci U S A* **2021**, *118* (10), e2020401118. https://doi.org/10.1073/pnas.2020401118.

(27) *How to achieve more consistent cluster density on Illumina sequencing platforms | Illumina Knowledge*. https://knowledge.illumina.com/instrumentation/general/instrumentation-general-reference_material-list/000001481 (accessed 2026-05-06).

(28) Xue, N.; Hong, D.; Zhang, D.; Wang, Q.; Zhang, S.; Yang, L.; Chen, X.; Li, Y.; Han, H.; Hu, C.; Liu, M.; Song, G.; Guan, Y.; Wang, L.; Zhu, Y.; Li, D. Engineering IscB to Develop Highly Efficient Miniature Editing Tools in Mammalian Cells and Embryos. *Molecular Cell* **2024**, *84* (16), 3128-3140.e4. https://doi.org/10.1016/j.molcel.2024.07.007.

(29) Tao, J.; Bauer, D. E.; Chiarle, R. Assessing and Advancing the Safety of CRISPR-Cas Tools: From DNA to RNA Editing. *Nat Commun* **2023**, *14* (1), 212.

Appendices

Appendix A: NGS Analysis of Precise Edits

NGS Analysis Using Azenta Primers (Table 5). Figures only show the non-target strand with transversion edits boxed in blue and insertions boxed in red pointed with a red arrow.

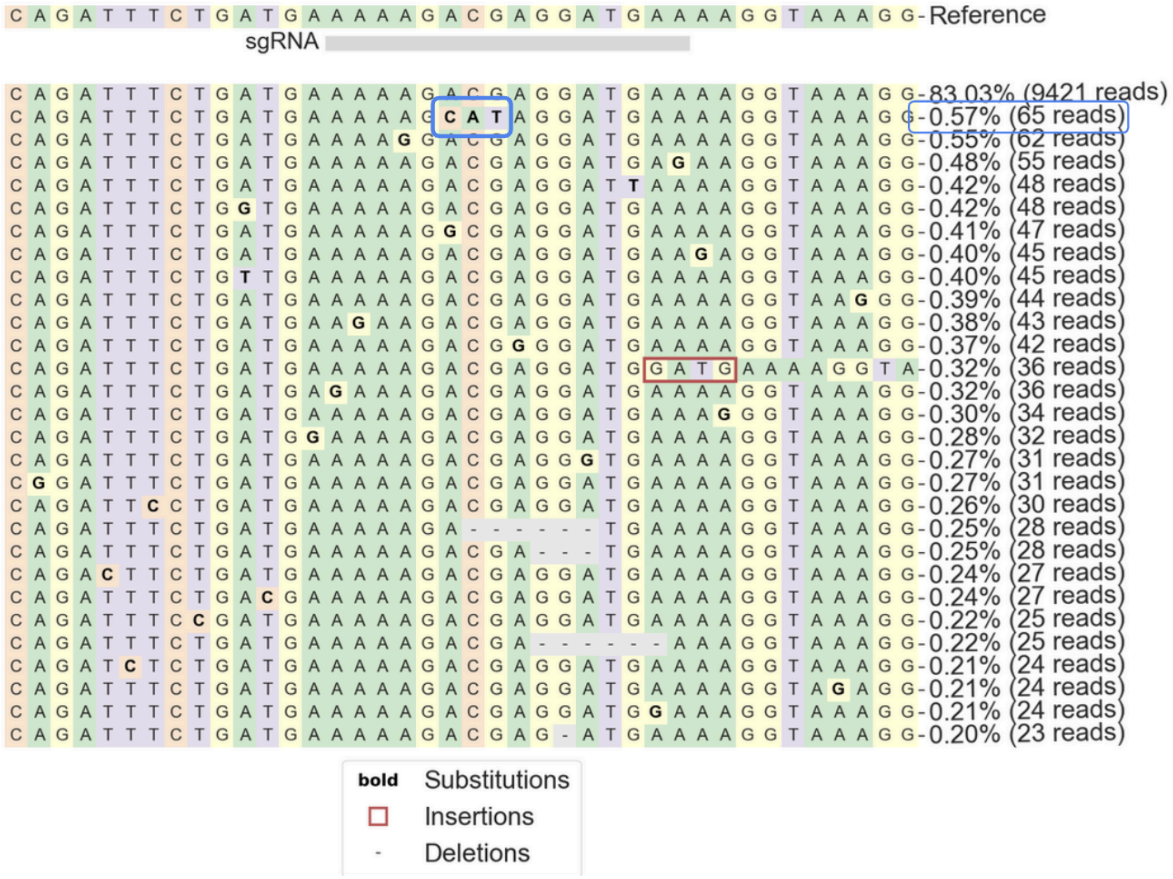


Figure 14: NGS 10 sample I-214 (DMNT1-C gene site): Inverse prime edit of CGT (reverse complement of ACG) to GTA (reverse complement of CAT) (65 reads at 0.57%).



Figure 15: From NGS-11 sample I-211 (VEGFA-B gene site), TAA insertion before AGT (99 reads at 0.43%).

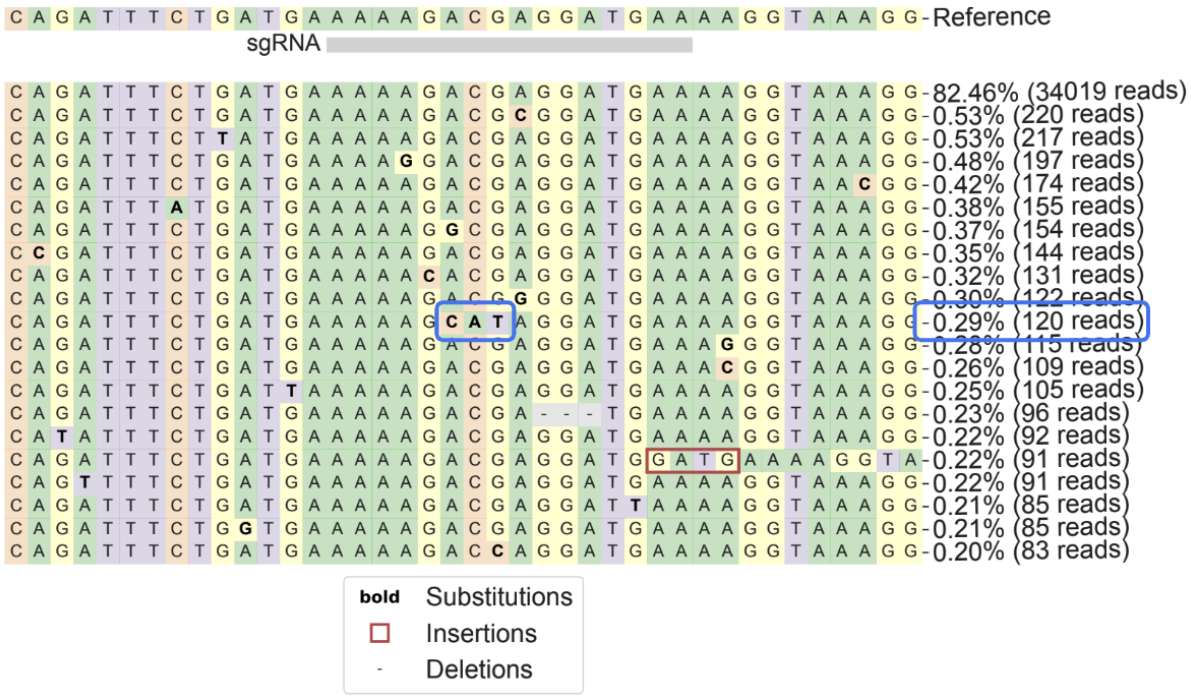


Figure 16: NGS 11 sample I-214 (DMNT1-C gene site): Inverse prime edit of CGT (reverse complement of ACG) to GTA (reverse complement of CAT) (120 reads at 0.29%).

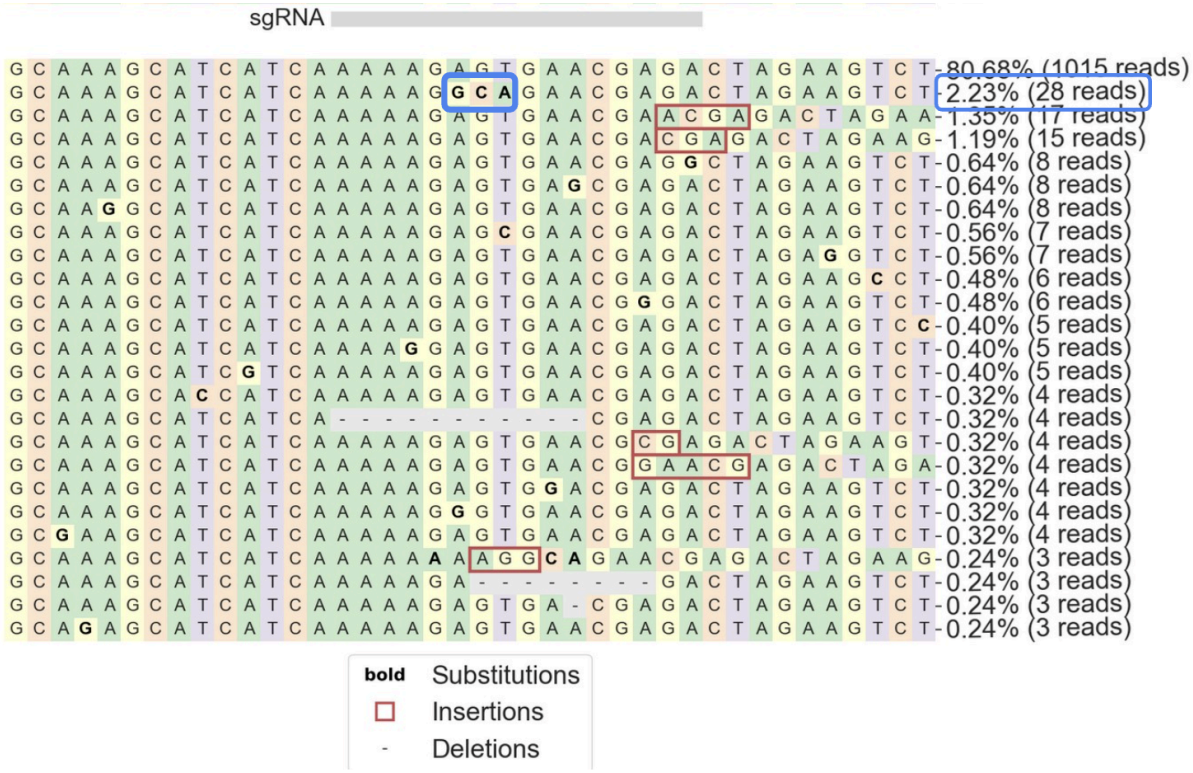


Figure 17: NGS 12 sample I-210 (VEGFA-B gene site): Inverse prime edit of ACT (reverse complement of AGT) to TGC (reverse complement of GCA) (28 reads at 2.23%)

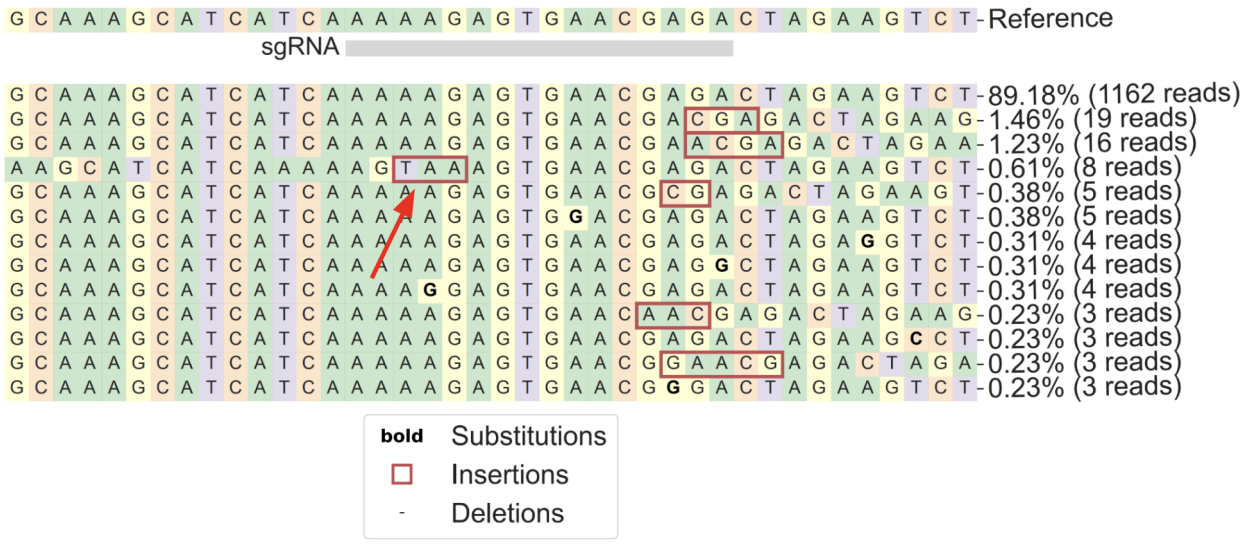


Figure 18: From NGS 12 sample I-211(VEGFA-B gene site), TAA insertion before AGT (8 reads at 0.61%).

NGS Analysis Using ISEQ Primers (Table 6). Figures only show the non-target strand with transversion edits boxed in blue and insertions boxed in red pointed with a red arrow.

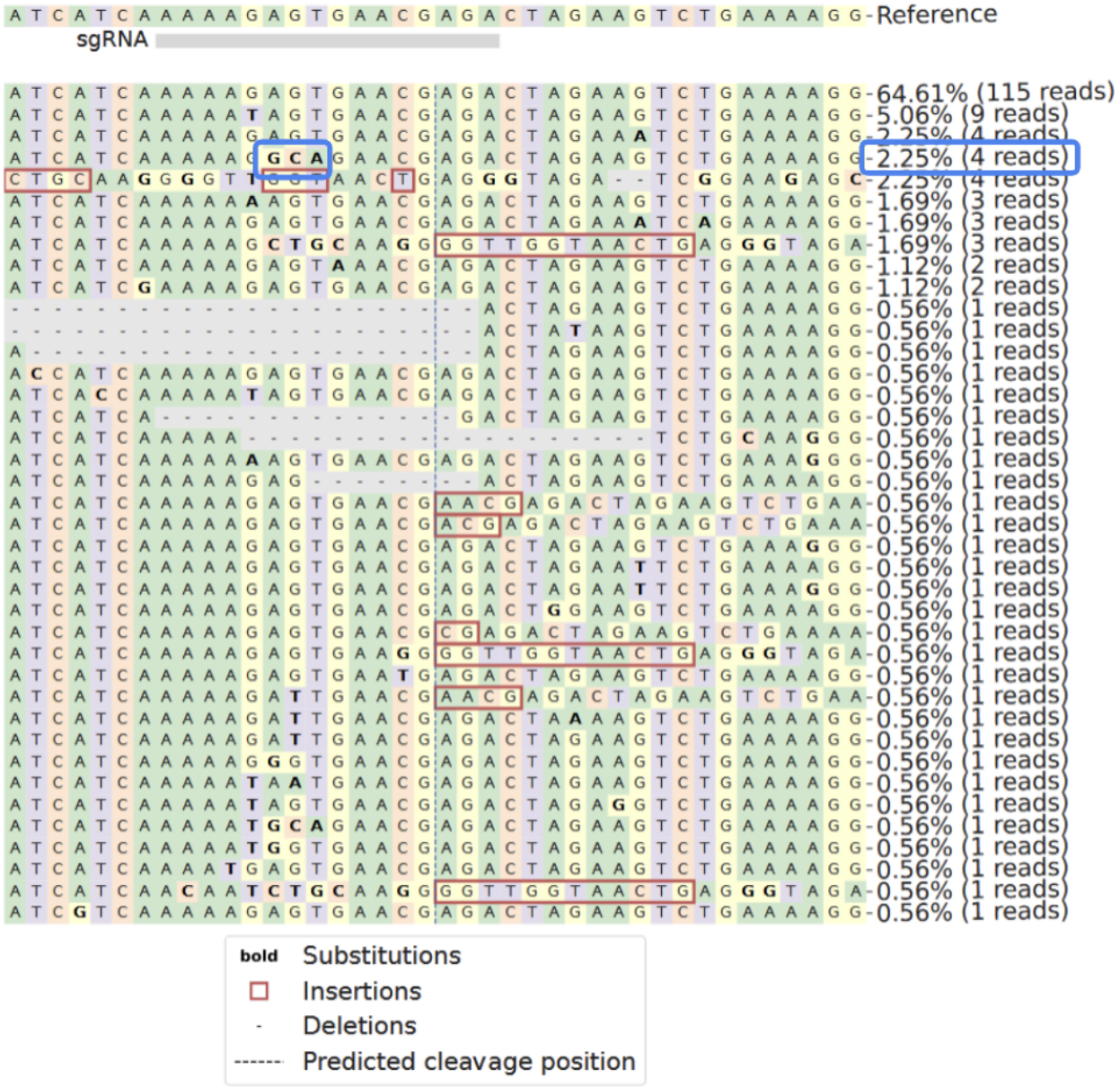


Figure 19: NGS 17 (NGS 12 Repeat) sample I-210 (VEGFA-B gene site): Inverse prime edit of ACT (reverse complement of AGT) to TGC (reverse complement of GCA) (4 reads at 2.25%).



Figure 20: From NGS 17 (NGS 12 Repeat) sample I-211(VEGFA-B gene site), TAA insertion before AGT (1 read at 0.36%).

Appendix B: Tables of Expanded NGS Results

Table 4: An expanded breakdown of NGS results using Azenta primers when only testing for IscB cutting activity. The average NGS run are the average results of all three NGS runs and correspond to the results from Figure 10.

NGS Run	Sample	Gene Site	Reads	Reads Aligned	Percent Reads Aligned	Percent Indel
1	I168	VEGFA-B	75,118	59,003	79	9.25
	WT		77,269	62,137	80	0
	I169	EMX1-C	87,174	84,691	97	2.87
	WT		63,446	60,503	95	0
	I170	DMNT1-C	69,084	67,213	97	10.25
	WT		47,884	45,562	95	0
2	I168	VEGFA-B	8,776	7,121	81	13.75
	WT		8,197	7,332	89	0
	I169	EMX1-C	8,394	8,018	96	4.4
	WT		7806	7504	96	0
	I170	DMNT1-C	6,716	6,299	94	10.13
	WT		7,468	7,075	95	0
3	I168	VEGFA-B	8,492	6,697	79	12.51
	WT		12,514	11,376	91	0
	I169	EMX1-C	6,004	5,718	95	4.77
	WT		8,943	8,636	97	0
	I170	DMNT1-C	6,360	5,963	94	6.32
	WT		9,105	8,658	95	0
Average	I168	VEGFA-B	30,795	24,274	80	11.84
	WT		32,660	26,948	87	0.00
	I169	EMX1-C	33,857	32,809	96	4.01
	WT		26,732	25,548	96	0.00
	I170	DMNT1-C	27,387	26,492	95	8.90
	WT		21,486	20,432	95	0.00

Table 5: An expanded breakdown of NGS results using Azenta primers for each designated sample with IscB (I) and its corresponding cRNA/crRNA number along with their function represented in the assay (PE: prime editing, IPE: inverse prime editing). Precise Edit Reads also shows the percentage of successful edited reads compared to the total number of reads.

NGS Run	Sample	Function	Gene Site	Reads	Reads Aligned	Percent Reads Aligned	Percent Indel	Precise Edit Reads
NGS 10	I-168	Control	VEGFA-B	7,275	2,883	39.63	2.43	
	I-210	Inverse PE		5,430	2,945	54.24	0.78	0
	I-211			6,257	2,748	43.92	0.36	0
	I-216	PE		6,436	3,320	51.58	0.27	0
	VEGFA-B WT	WT		21,142	13,730	64.94	0.12	
	I-169	Control	EMX1-C	3,997	3,718	93.02	0.67	
	I-212	Inverse PE		5,274	4,942	93.70	1.48	0
	I-213			5,953	5,642	94.78	0.37	0
	I-217	PE		4,757	4,474	94.05	0.38	0
	EMX1-C WT	WT		16,013	15,500	96.80	0.11	
	I-170	Control	DMNT1-C	7,969	7,701	96.64	2.22	
	I-214	Inverse PE		11,566	11,347	98.11	3.82	65 (0.57%)
	I-215			9,342	9,149	97.93	3.50	0
	I-218	PE		10,093	9,826	97.35	1.08	0
	DMNT1-C WT	WT		12,212	11,770	96.38	0.08	
NGS 11	I-168	Control	VEGFA-B	42,247	33,105	78.36	3.57	
	I-210	Inverse PE		37,944	28,549	75.24	0.75	0
	I-211			32,685	22,832	69.85	2.45	99 (0.43%)
	I-216	PE		42,691	33,488	78.44	0.61	0
	VEGFA-B WT	WT		63,012	50,324	79.86	0.00	
	I-169	Control	EMX1-C	83,141	81,352	97.85	2.64	
	I-212	Inverse PE		98,202	96,042	97.80	1.77	0
	I-213			82,211	80,385	97.78	4.03	0
	I-217	PE		74,526	72,984	97.93	0.94	0
	EMX1-C WT	WT		115,430	113,489	98.32	0.10	
	I-170	Control	DMNT1-C	55,478	51,044	92.01	4.32	
	I-214	Inverse PE		47,775	41,253	86.35	3.23	120 (0.29%)
	I-215			45,662	41,986	91.95	3.14	0
	I-218	PE		50,710	45,927	90.57	6.10	0
	DMNT1-C WT	WT		43,134	39,584	91.77	1.43	
NGS 12	I-168	Control	VEGFA-B	14,790	323	2.18	7.74	
	I-210	Inverse PE		12,269	1258	10.25	7.55	28 (2.23%)
	I-211			13,490	1303	9.66	6.45	8 (0.61%)
	I-216	PE		11,032	1117	10.13	2.06	0
	VEGFA-B WT	WT		5,554	36	0.65	0	
	I-169	Control	EMX1-C	5,554	4020	72.38	0	
	I-212	Inverse PE		7,554	4450	58.91	1	0
	I-213			6,163	4873	79.07	1.09	0
	I-217	PE		5,978	4856	81.23	0	0
	EMX1-C WT	WT		7,554	4450	58.91	1.2	
	I-170	Control	DMNT1-C	14,790	11	0.07	0	
	I-214	Inverse PE		12,269	14	0.11	0	0
	I-215			13,490	11	0.08	0	0
	I-218	PE		11,032	15	0.14	0	0
	DMNT1-C WT	WT		6,163	12	0.19	0	

Table 6: An expanded breakdown of NGS results using ISEQ primers for each designated sample with IscB (I) and its corresponding cRNA/crRNA number along with their function represented in the assay (PE: prime editing, IPE: inverse prime editing). Precise Edit Reads also shows the percentage of successful edited reads compared to the total number of reads.

NGS Run	Sample	Function	Gene Site	Reads	Reads Aligned	Percent Reads Aligned	Percent Indel	Precise Edit Reads
NGS 14	I-168	Control	VEGFA-B	91,805	20,544	22.38	0	
	I-210	Inverse PE		62,823	11,302	17.99	0	0
	I-211			62,901	7,755	12.33	4	55 (0.71%)
	I-216	PE		65,607	9,132	13.92	0	0
	VEGFA-B WT	WT		87,129	78,318	89.89	1.2	
	I-169	Control	EMX1-C	75,680	65,831	86.99	1	
	I-212	Inverse PE		78,096	58,357	74.72	1.57	0
	I-213			92,143	20,842	22.62	1	0
	I-217	PE		97,730	94,124	96.31	4.54	0
	EMX1-C WT	WT		69,654	58,408	83.85	0	
	I-170	Control	DMNT1-C	72,220	52,713	72.99	3.5	
	I-214	Inverse PE		77,717	68,730	88.44	2.07	0
	I-215			62,234	7,489	12.03	0	0
	I-218	PE		75,332	72,751	96.57	0	0
	DMNT1-C WT	WT		53,689	47,534	88.54	0	
NGS 15	I-168	Control	VEGFA-B	124,148	7,043	5.67	0	
	I-210	Inverse PE		107,514	9,800	9.12	0	0
	I-211			139,200	5,552	3.99	0	0
	I-216	PE		127,401	5,000	3.92	0	0
	VEGFA-B WT	WT		124,523	11,733	9.42	0	
	I-169	Control	EMX1-C	82,240	14,634	17.79	0	
	I-212	Inverse PE		110,301	13,403	12.15	0	0
	I-213			100,376	17,527	17.46	1	0
	I-217	PE		51,032	2,617	5.13	2.75	0
	EMX1-C WT	WT		61,435	9,918	16.14	1.48	
	I-170	Control	DMNT1-C	146,061	25,138	17.21	1.07	
	I-214	Inverse PE		127,512	24,466	19.19	1.34	0
	I-215			46,056	2,235	4.85	0	0
	I-218	PE		87,497	18,273	20.88	1	0
	DMNT1-C WT	WT		82,813	22,915	27.67	1.8	

NGS 17 (NGS 12 Repeat)	I-168	Control	VEGFA-B	23,481	33	0.14	15.15	
	I-210	Inverse PE		22,621	178	0.79	15.73	4 (2.25%)
	I-211			22,889	279	1.22	20.07	1 (0.36%)
	I-216	PE		23,129	91	0.39	24.18	0
	VEGFA-B WT	WT		25,941	85	0.33	0.00	
	I-169	Control	EMX1-C	24,242	265	1.09	1.51	
	I-212	Inverse PE		29,197	40	0.14	0.00	0
	I-213			29,509	1,469	4.98	1.23	0
	I-217	PE		22,296	814	3.65	6.02	0
	EMX1-C WT	WT		24,078	463	1.92	7.99	
	I-170	Control	DMNT1-C	21,449	1,944	9.06	5.76	
	I-214	Inverse PE		23,304	1,937	8.31	4.23	0
	I-215			22,570	164	0.73	12.20	0
	I-218	PE		21,355	1,439	6.74	0.90	0
	DMNT1-C WT	WT		22,734	1,640	7.21	2.07	
NGS 17 (2)	I-168	Control	VEGFA-B	34,071	31	0.09	0.00	
	I-210	Inverse PE		33,132	104	0.31	3.85	0
	I-211			32,140	44	0.14	2.27	0
	I-216	PE		31,996	38	0.12	13.16	0
	VEGFA-B WT	WT		28,567	2,126	7.44	0.09	
	I-169	Control	EMX1-C	31,107	756	2.43	0.79	
	I-212	Inverse PE		40,261	649	1.61	0.00	0
	I-213			40,244	1,241	3.08	0.08	0
	I-217	PE		33,005	156	0.47	1.92	0
	EMX1-C WT	WT		31,298	418	1.34	0.96	
	I-170	Control	DMNT1-C	31,345	376	1.20	1.33	
	I-214	Inverse PE		34,690	46	0.13	2.17	0
	I-215			27,869	819	2.94	1.47	0
	I-218	PE		26,705	917	3.43	0.55	0
	DMNT1-C WT	WT		32,328	65	0.20	1.54	

NGS 17 (3)	I-168	Control	VEGFA-B	41,559	1,724	4.15	2.67	
	I-210	Inverse PE		38,290	874	2.28	1.26	0
	I-211			45,261	827	1.83	4.11	0
	I-216	PE		34,638	1,084	3.13	1.85	0
	VEGFA-B WT	WT		34,029	3,612	10.61	0.28	
	I-169	Control	EMX1-C	46,069	48	0.10	0.00	
	I-212	Inverse PE		49,997	3,298	6.60	0.30	0
	I-213			48,562	4,234	8.72	0.05	0
	I-217	PE		43,795	513	1.17	0.19	0
	EMX1-C WT	WT		42,535	684	1.61	1.61	
	I-170	Control	DMNT1-C	46,463	723	1.56	0.97	
	I-214	Inverse PE		38,520	160	0.42	1.25	0
	I-215			37,881	1,047	2.76	2.01	0
	I-218	PE		36,885	360	0.98	0.00	0
	DMNT1-C WT	WT		38,409	825	2.15	0.61	

Appendix C: List of Materials Used in Experimental Methods

DMEM-P/S Media: 1x DMEM (Life Technologies, no. 11995-065), 10% fetal bovine serum (FBS), 100 µg/ml 1-penicillin and 100 µg ml 1-streptomycin (P/S), and GlutaMax (ThermoFisher, no. 35050061).
TrypLE Express: Trypsin solution with phenol red (Life Technologies, no. 12605036)
Optimem: OPTIMEM1 Reduced Serum Media (ThermoFisher, no. 31985).
Lipofectamine: Lipofectamine 2000 (Invitrogen, no. 11668027)
Cell Counter: Countess 3 Automated Cell Counter (Invitrogen, no. AMQAX2000)
Mini-Prep Kit: GB-Clean Plasmid Miniprep Kit (GeneBio Systems, no. P100103)
PCR Clean-up Kit: Gel-PCR 2in1 Kit-HiPure DNA Kit (GeneBio Systems, no. D21103)
PCR Master Mix: GB-Amp 2X PCR Master Mix (without loading dye) (GeneBio Systems, no. P4001)
Competent Cells: NEB 5-alpha Competent *E. coli* (New England Biolabs, no. C2987)
DNA Ladder: 1 kb Plus DNA Ladder (New England Biolabs, no. N3200S)
Genomic DNA Extraction Kit: E.Z.N.A. Tissue DNA Kit (Omega Bio-Tek, no. D3396-02)
Nanodrop: Nanodrop One (Thermo Scientific, no. ND-ONE-W)
Thermalcycler: C1000 Touch Thermalcycler (BioRad, no. 1851148)
Electrophoresis Chamber: Power Pac 200, Mini-Sub and Wide Mini-Sub horizontal electrophoresis chambers (BioRad, no. 1704466 and 1704468)
Imager: Molecular Imager Gel-Doc XR+ Imaging System (BioRad)
Flow Cytometer: Cytek Aurora (Cytek Biosciences)
Flow Cytometry Analysis Program: Spectroflo
Sanger Sequencing Service: The Centre for Applied Genomics (TCAG) DNA Sequencing Facility
Sanger Sequencing Analysis Program: DECODR v3.0
NGS Sequencing Service and Primers: Azenta Life Sciences
NGS Sequencing Primers: ISEQ Custom Primers
NGS Sequencer (In-lab): MiSeq i100 Series (Illumina, no. 20115694)
NGS Reading Program: CRISPResso2
NGS Sequence Viewer Program: Geneious Prime