UNIVERSITY OF CALIFORNIA SAN DIEGO

Studies of Genome Editors for Precision Therapeutics and Beyond

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Elizabeth Michelle Porto

Committee in charge:

Professor Alexis Komor, Chair
Professor Vicki Grassian
Professor Patricia Jennings
Professor Prashant Mali
Professor Alexandra Newton
Professor Susan Taylor

2023

The dissertation of Elizabeth Michelle Porto is approved, and it is acceptable in quality and form for publication on microfilm.

University of California San Diego

2023

In honor of my parents, Robert and Nancy – without their love and support I would never have gotten this far. This is for you, Mom and Dad.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AAV	Adeno-associated virus
ABE	Adenine base editor
BEs	Base editors
bpNLS	Bipartite nuclear localization signal
CBE	Cytosine base editor
CGSC	Chemistry Graduate Student Council
CRISPR	Clustered regularly interspaced short palindromic repeats
dCas9	Catalytically-dead or inactive Cas9
DSB	Double strand break
FACS	Fluorescence activated cell sorting
G1	Gap 1 phase of the cell cycle
G2	Gap 2 phase of the cell cycle
gDNA	Genomic DNA
GPSA	Graduate and professional student association
gRNA	Guide RNA or spacer
HDR	Homology-directed repair
HeLa	Henrietta Lacks immortalized cell line
HEK293T	Human embryonic kidney immortalized cell line
HTS	High-throughput sequencing
Indels	Insertions and deletions
M-phase	Mitosis phase of the cell cycle
MMR	Mismatch repair pathway

NHEJ	Non-homologous end joining
PAM	Protospacer adjacent motif
PBS	Primer binding sequence
PE	Prime editor
PE2	Prime editor 2
PE2max	Prime editor 2, max construct
PE3	Prime editor 3
pegRNA	Prime editing guide RNA
RT	Reverse transcriptase
RTT	Reverse transcription template
S-phase	Synthesis phase of the cell cycle
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SSB	Single strand break
ssDNA	Single-stranded DNA
SWIGS	Society for women in graduate studies
TALENs	Transcription-activator-like effector nucleases
UDG	Uracil DNA glycosylase
UGI	Uracil glycosylase inhibitor
ZFN	Zinc-finger nuclease

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I want to acknowledge my parents, Robert and Nancy Porto, without whom I would never have gotten this far. My Dad has been here every step of the way for my time in grad school, always providing support, words of encouragement, and just lending a sympathetic ear when I need to vent. Sadly my Mom has not been here to see my journey in the same way, as she passed away before my sophomore year in high school. Nevertheless, it was her who originally set me on my academic path and I know she would be thrilled with how far I've gone. Both of my parents got me here, and words will never be able to express how much I love and appreciate them.

With regards to my scientific endeavors, major thanks to my PI, Dr. Alexis Komor. It seems crazy that we signed the paperwork for me to be the first student in her new lab at UCSD six years ago. I appreciate Alexis giving me the opportunity to work in her lab, and entrusting me and the four other original Komor Lab members, Mallory, Brodie, Sifeng, and Cameron, with building the lab foundations and culture. I'm also extremely grateful Alexis supported my non-scientific work and my professional development in a space and time where most PIs steer their students toward academia. Many thanks to all past and present Komor Lab members as well, especially Dr. Zsolt Bodai and Dr. Kartik Rallapalli.

I also have to acknowledge the friends I came to grad school with, and those I made along the way. Coti, Forrest, Sam, Kristine, and Audrey are truly some of the best friends a gal could ask for and the countless FaceTime calls, Instagram memes, coffee dates, and encouraging texts truly helped get me through grad school. I love you all.

Chapter 1 is reproduced, in part, with permission, from: **Porto, E. M**., Komor, A. C., Slaymaker, I. M., and Yeo, G. W. (2020) Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov*., 19, 839-859. and **Porto, E. M**., and Komor, A. C. (2023). In the business of base editors: Evolution from bench to bedside. *PLOS Biology*, 21, e3002071. The dissertation

author was the primary author on all reprinted materials.

Chapter 2 is reproduced, in full, with permission, from: **Porto, E. M**., and Komor, A. C. (2023). Studies of the relationship between early prime editors and cell cycle dependence. *Nucleic Acids Research, submitted*. The dissertation author was the primary author on all reprinted materials.

Chapter 3 is reproduced, in full, with permission, from: **Porto, E. M**., and Komor, A. C. (2023). Studies of the relationship between early prime editors and cell cycle dependence. *Nucleic Acids Research, submitted*. The dissertation author was the primary author on all reprinted materials.

VITA

2017 Bachelor of Science, Chemistry, University of Missouri-Kansas City, Kansas City, MO

2017 Bachelor of Science, Biology, University of Missouri-Kansas City, Kansas City, MO

2019 Master of Science, Chemistry, University of California San Diego, La Jolla, CA

2023 Doctorate of Philosophy, Chemistry, University of California San Diego, La Jolla, CA

PUBLICATIONS

Porto, E. M., and Komor, A. C. (2023). Studies of the relationship between early prime editors and cell cycle dependence. *Nucleic Acids Research, submitted*.

Porto, **E. M.,** and Komor, A. C. (2023). In the business of base editors: Evolution from bench to bedside. *PLOS Biology,* 21, e3002071.

Porto, E. M., Komor, A. C., Slaymaker, I. M., and Yeo, G. W. (2020) Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.,* 19, 839-859.

FIELD OF STUDY

Major Field: Biochemistry

Studies in Biochemistry and Biophysics

Professor Alexis Komor

ABSTRACT OF THE DISSERTATION

Studies of Genome Editors for Precision Therapeutics and Beyond

by

Elizabeth Michelle Porto

Doctor of Philosophy in Chemistry

University of California San Diego, 2023

Professor Alexis Komor, Chair

With the advent of recombinant DNA technology in the 1970s, the idea of using gene therapies to treat human genetic diseases captured the interest and imagination of scientists around the world. Years later, enabled largely by the development of CRISPR-based genome editing tools, the field has exploded, with academic labs, startup biotechnology companies, and large pharmaceutical corporations working in concert to develop life-changing therapeutics.

Prime editing is a new genome editing methodology that utilizes novel intermediates. How the cell processes these intermediates into desired genome editing outcomes has yet to be fully

characterized. In Chapter 2 and 3, we detail our findings which aim to provide insights into the mechanisms governing prime editing and reveal this new technology to be more ubiquitous than traditional genome editing tools that rely on S- and G2-phase dependent DNA repair pathways.

Base editing, first reported in 2016, is capable of installing C•G to T•A and A•T to G•C point mutations, while largely circumventing some of the pitfalls of traditional CRISPR/Cas9 gene editing. Despite their youth, these technologies have been widely used by both academic labs and therapeutics-based companies. In Chapter 4, we detail work aimed at addressing two of the main pitfalls of base editors: limited point mutation installation capability and incompatibility with optimal viral therapeutic delivery vehicles.

Finally, in Chapter 5, non-scientific development garnered throughout the primary author and researcher's tenure is briefly discussed as this work has also led to both personal and professional advances.

Chapter 1

Introduction

1.1 Genome Editing

As our understanding of how the primary sequence of genomic DNA impacts human health has expanded, the therapeutic potential of genome editing has emerged. This revolution in how we think about human health and disease has in large part been driven by rapid advances in genome sequencing technologies, which have revealed causative mutations of genetic diseases. Furthermore, we are brought ever closer to the realization of precision medicine: the development of disease prevention and treatment strategies based on a patient's individual characteristics (such as their genomic sequence). It is therefore an exciting time for researchers in these fields as we

tackle some of the most noteworthy barriers to the utilization of genome editing for the treatment and cure of genetic diseases. Approximately half of the known pathogenic genetic variants are due to single nucleotide variants (SNVs), highlighting the need for the development of methods and tools capable of correcting SNVs with high efficiency.(1) Over 96% of observed human genetic variants are SNVs, with over 99% currently lacking a clinical interpretation.(2) Therefore, tools to introduce SNVs will also prove indispensable for improving our understanding of how human genetic variation impacts health.(3–7)

To be used as a therapeutic, a genome editing tool must demonstrate high on-target efficiency, minimal harmful or undesired off-target edits, and be deliverable to the organ(s) of interest. It is important to note that the disease target will dictate the exact degree to which these criteria must be met. Early efforts in the field used platforms such as zinc-finger nucleases (ZFNs) and transcription-activator-like effector nucleases (TALENs), but these methods were hampered by the requirement of designing and validating a new ZFN or TALEN protein for each new target editing site.(8) However, these extensive protein re-engineering requirements were alleviated with the discovery, mechanistic elucidation, and adaption for genome editing of clustered regularly interspaced short palindromic repeat (CRISPR) platforms.

CRISPR systems are a naturally occurring bacterial and archaeal defense against invading viruses that have been harnessed for genome and transcriptome editing by inducing double stranded DNA breaks (DSBs) or RNA cleavage at user-defined loci in living cells.(9–11) Re-programming these systems to perform genome editing at different genomic loci simply requires changing the sequence of a piece of RNA (called a spacer, guide RNA (gRNA), or single guide RNA (sgRNA)) via Watson-Crick-Franklin base pairing rules. DSBs are introduced at pre-programmed loci by the CRISPR-associated proteins, Cas9 and Cas12, and are typically repaired

through one of two competing endogenous repair pathways in mammalian cells: non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ-resolved DSBs result in nonspecific insertions or deletions (indels) at the site of the DSB, often resulting in frame-shifts and gene knock-out. Researchers can co-opt the HDR pathway to introduce desired and precise sequence edits into the genome by using an exogenous DNA repair template. Unfortunately, HDR efficiencies vary amongst mammalian cell type (i.e. unmodified cells exhibit low HDR efficiency), the HDR pathway is only active during certain phases of the cell cycle, and HDR is in constant competition with and usually outcompeted by NHEJ for repair of DSBs. Therefore, the development of new techniques and tools to improve HDR yields and/or suppress NHEJ rates have been primary areas of study in the field, have yielded many improvements, and have been reviewed elsewhere.(12–19)

In comparison with Cas9 and Cas12, Cas13 proteins function similarly as DNA targeting CRISPR systems to bind and cleave target RNA transcripts in a programmable manner. However, upon target RNA binding and cleavage, Cas13 will also non-specifically cleave nearby ssRNAs *in vitro*, which can potentially pose complications and thus limit the therapeutic potential of wtCas13s.(20) It is important to note that Cas9, Cas12, and Cas13 enzymes can be catalytically inactivated to produce dCas9, dCas12, and dCas13, which maintain programmable DNA or RNA binding capabilities but do not cleave their target.(21) Nucleic acid backbone-cleaving technologies are not within the scope of this review, but have been reviewed extensively elsewhere.(22–24)

One technology developed to address the challenge of creating targeted single nucleotide alterations in a precise and efficient manner, is base editing. Base editing is unique in that it avoids nucleic acid backbone cleavage and instead directly chemically modifies target nucleobases in the

process of genome and transcriptome editing (**Figure 1.1A and 1.1B**). Both DNA and RNA base editors (BEs) have been developed, and their rapid adoption by the genome and transcriptome editing communities is a clear demonstration of their value as tools to enable both basic science research and development of human therapeutics. BEs, of which there are many variants, can be sorted into two main categories: those targeting DNA and those targeting RNA. While the origins of base editing technology begin with RNA BEs decades ago, both categories have recently seen an explosion in development. DNA BEs can be further categorized as cytosine base editors (CBEs) or adenine base editors (ABEs). Both CBEs and ABEs are powerful tools for the permanent introduction of point mutations in DNA in living cells with high efficiency.

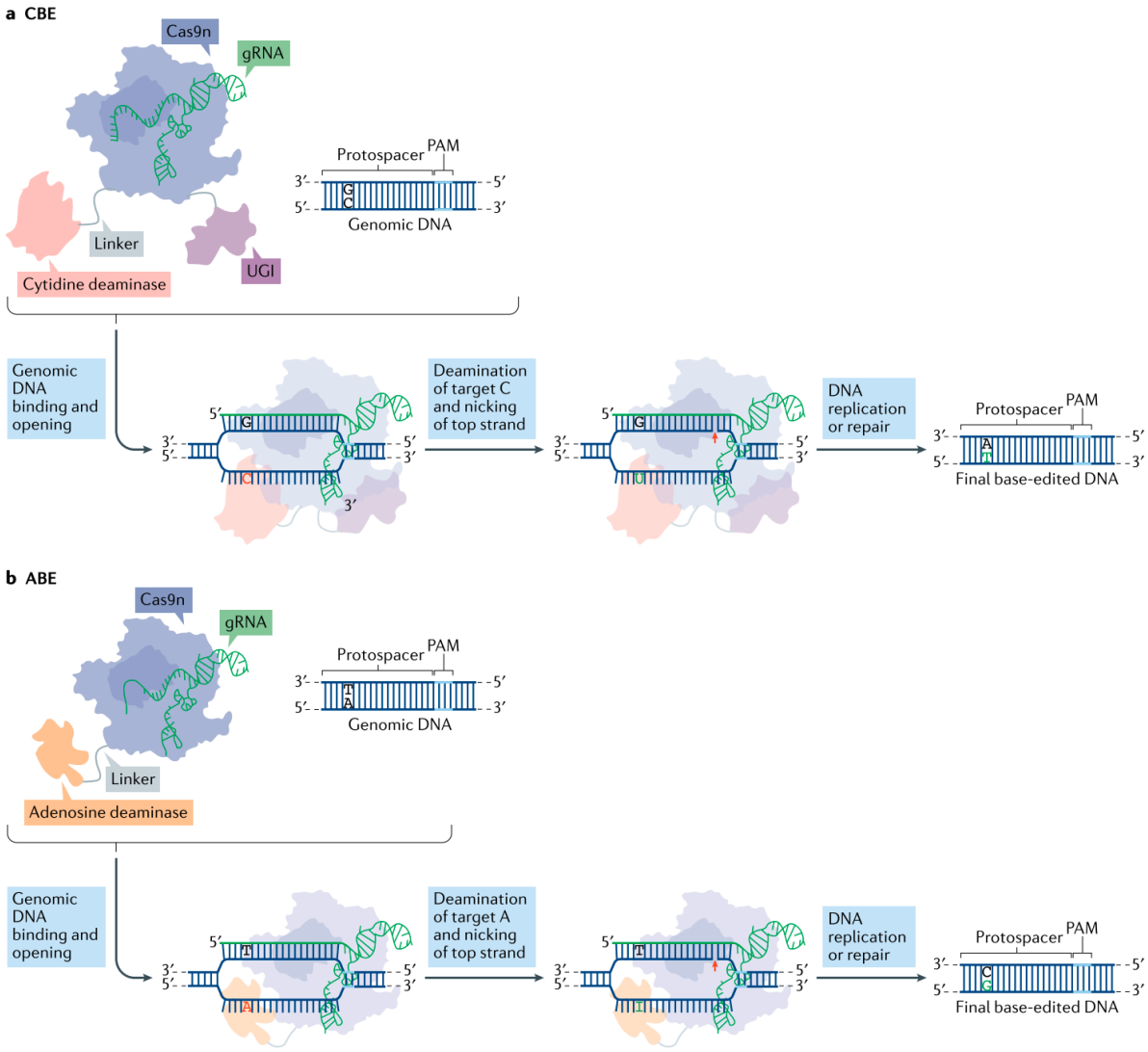


Figure 1.1 General overview of DNA base editing technologies. A) Cytosine base editor (CBE) mechanism. Principle components of the CBE are designated in colored text boxes. If uracil glycosylase inhibitor (UGI) is present (an optional component), it will 'protect' the U•G intermediate from excision by uracil DNA glycosylase (UDG) to boost efficiency of the final base-edited DNA outcome. The nickase version of Cas9 (Cas9n) nicks the top strand (red arrow) whereas the cytidine deaminase converts cytosine (red) to uracil (green). Ultimate conversion of a C•G to T•A base pair is achieved through the outlined steps. **B)** The adenine base editor (ABE) mechanism is similar to that of CBE, without possible inclusion of a UGI domain in the ABE architecture. Through ABE-mediated editing, an A•T to G•C base pair conversion is achieved via an inosine-containing inter- mediate. gRNA, guide RNA; PAM, protospacer adjacent motif; target A, ABE desired base substrate; target C, CBE desired base substrate.

1.2 C•G to T•A base editors (CBEs)

The first DNA BE was developed as a method to perform genome editing without using DSBs. A naturally occurring cytidine deaminase enzyme was used to convert target cytosines to uracil, which has the base pairing properties of thymine. This was expected to catalyze an overall C•G to T•A base pair conversion following the cell's use of uracil as a template for repair (**Figure 1.1A**).(25) The original prototype (named BE1, or first-generation base editor) used a catalytically dead version of the *Streptococcus pyogenes* Cas9 (dCas9) enzyme tethered to the single-strand DNA (ssDNA) specific cytidine deaminase enzyme APOBEC1 from *Rattus norvegicus* (rAPOBEC1) (**Figure 1.2A**). dCas9 binds to a target DNA locus of interest (the protospacer, **Figure 1.2B**) through canonical RNA-DNA base pairing between the gRNA and the genomic DNA. Sequence complementarity between the gRNA and the protospacer, and the presence of an NGG (where N = A/C/G/T) protospacer-adjacent motif (PAM) sequence are required for dCas9 binding to the target locus. Once dCas9 finds its target sequence, it will locally denature the dsDNA to generate an R-loop,(26) exposing a short stretch of ssDNA (positions 4 to 8 if the PAM is counted as positions 21-23, **Figure 1.2B**) on the non-complementary strand for deamination by the APOBEC1 enzyme.

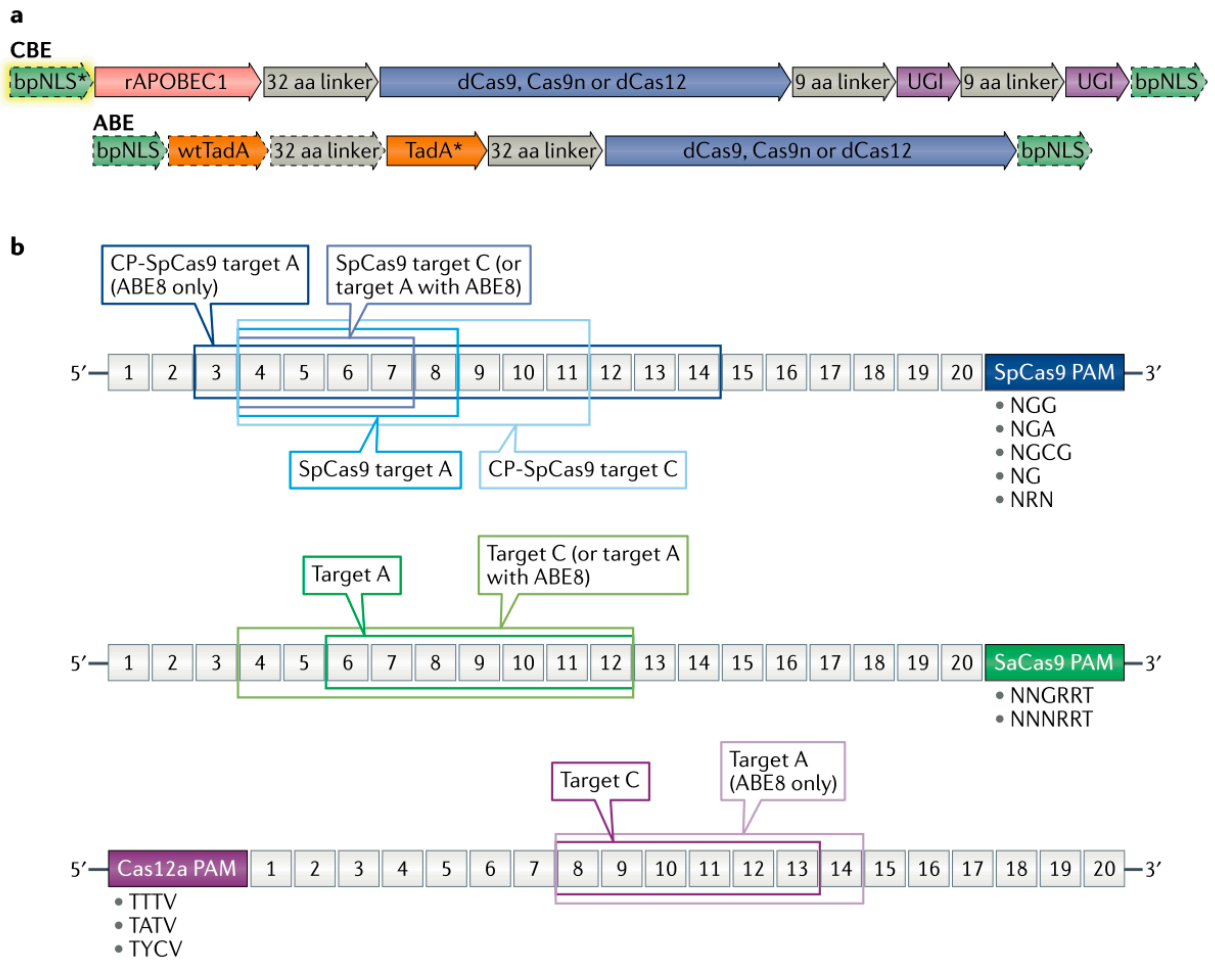


Figure 1.2 DNA base editor and protospacer design scheme. A) Construct maps of basic cytosine base editor (CBE) and adenine base editor (ABE) architectures. In the CBE architecture (top), solid line components make up the basis for the fourth-generation CBE, BE4, whereas dotted line components (bipartite nuclear localization signal (bpNLS); green) can be optionally added, to produce BE4max. The amino-terminal bpNLS* component is FLAG-tagged (yellow haze). In the ABE architecture (bottom), all solid line and dotted line components make up the basis for ABE7.10; the dotted lined components (wtTadA (orange) and one of the two 32-amino-acid (aa) linkers (grey)) are optional, and removal of these components results in a monomeric ABE construct with no reduction in on-target efficiency. For both CBE and ABE architectures, use of an appropriate nickase Cas variant is only possible with Cas9 (Cas9n; blue). **B)** Activity windows of base editors with the basic architecture from part a with the indicated Cas proteins (Streptococcus pyogenes Cas9 (SpCas9; blue), Staphylococcus aureus Cas9 (SaCas9; green) and Cas12a (purple)). Protospacer adjacent motifs (PAMs) associated with each Cas enzyme are listed. Base editor activity windows are shown over the 20-nucleotide protospacer sequence (corresponding colored box outlines).

BE1 could effectively convert cytosines to uracils *in vitro* in a programmable manner, but was significantly less effective at introducing C•G to T•A point mutations in live cells (5- to 36-fold decreases in efficiency were observed).(25) The large decrease in base editing efficiency was hypothesized to be partially due to high intracellular levels of uracil excision of the U•G intermediate by the base excision repair (BER) enzyme uracil DNA glycosylase (UDG). UDG catalyzes the removal of uracil in DNA to initiate the BER pathway, ultimately resulting in reversion to the original C•G base pair.(27, 28) To protect the uracil intermediate and boost base editing efficiencies, the phage polypeptide uracil glycosylase inhibitor (UGI) was added to the BE1 architecture, resulting in the second-generation base editor, BE2. Addition of UGI enhanced editing efficiencies approximately 3-fold compared to BE1.(25) In a final improvement to the BE architecture, the dCas9 portion of BE2 was replaced with the nickase version of Cas9 (Cas9n) to yield the third-generation editor BE3. In this new construct, Cas9n would nick the DNA backbone of the unedited, G-containing DNA strand, flagging it for removal by the eukaryotic mismatch repair (MMR) pathway and forcing the cell to use the uracil as a template during downstream repair (**Figure 1.1A**). This nicking strategy boosted efficiencies an additional 2- to 6-fold compared to BE2.(25) As the BE toolbox expanded, DNA BEs capable of facilitating C•G to T•A base pair conversions collectively became known as cytosine base editors (CBEs). It is important to note the dependence of this strategy on the single-stranded portion of the R-loop; ssDNA-specific cytidine deaminase fusions with other classes of genome editing agents such as ZFNs did not display such precision or efficiency.(29) However, very recently a dsDNA-specific cytidine deaminase was discovered and repurposed into a C•G to T•A base editor using TALEs.(30) In this system, the deaminase is split in half, with each half fused to a different TALE construct. The two TALEs bind to adjacent sites in DNA, bringing the two deaminase halves together where the

enzyme performs base editing chemistry. Notably, this new base editor, DdCBE, enabled efficient mitochondrial genome editing for the first time, as its reliance on TALEs instead of a Cas enzyme inherently overcame the previous challenges facing nucleic acid delivery to mitochondria.(30)

1.3 A•T to G•C base editors (ABEs)

Drawing inspiration from CBEs, it was quickly recognized that adenosine deamination chemistry would result in inosine, which is read by replication and transcription machinery as guanine. This theoretical adenine base editor (ABE) would therefore be capable of correcting C•G to T•A mutations, which represent the most common pathogenic SNVs reported in the ClinVar database.(31) Naturally occurring adenosine and adenine deaminase enzymes do exist, but their substrates are confined to various forms of RNA. In order to create an ABE, an adenosine deaminase acting on ssDNA needed to be generated (**Figure 1.1B, Figure 1.2A**). A variety of naturally occurring adenosine deaminases (such as *Escherichia coli* TadA, or ecTadA, human ADAR2, mouse ADA, and human ADAT2) were assayed for ABE activity, but none yield A•T to G•C base editing above background levels.(32) Therefore, directed evolution was employed to evolve the desired enzyme from ecTadA. Similarity between the desired substrate (ssDNA) and the wild-type substrate (all contacts between ecTadA and its tRNA substrate are localized to the single-stranded loop region of the tRNA), along with its shared homology with the APOBEC enzyme used in CBE, were among the main reasons why ecTadA was selected as the starting point for directed evolution.(33)

A total of seven rounds of directed evolution were performed, identifying 14 mutations in TadA to create the final ABE7.10 construct, consisting of a heterodimeric wtTadA-TadA* (* indicates the presence of mutations, the wild-type enzyme acts as a dimer to perform its chemistry

on tRNA) complex fused to Cas9n. ABE7.10 was demonstrated to introduce A•T to G•C point mutations in live cells with average editing efficiencies of 58% across 17 genomic loci, with an editing window of positions 4 to 7 within the protospacer (**Figure 1.2B**).(32) Unlike CBE, no DNA repair manipulation component (such as UGI) is required due to the infrequent nature of the inosine intermediate (intracellular inosine excision is much less efficient than that of uracil). Additionally, a variety of subsequent studies have suggested that the wtTadA component of ABEs is unnecessary and can be omitted without decreases in editing efficiency, indicating a fundamental difference in the enzyme's mechanism for performing chemistry on RNA versus DNA.(34, 35) Together, ABE and CBE are theoretically capable of correcting 63% of pathogenic SNVs reported in ClinVar.

1.4 Prime editors (PEs)

Prime editors (PEs) are a next step in the evolution of genomic medicine and have addressed some of the limitations of BEs (36). This new technology, like BEs, avoids the use of DSBs and therefore installs genomic modifications with high precision. PEs perform genome editing using a completely different mechanism than BEs, and the two technologies are therefore complementary to each other. PEs employ a reverse transcriptase (RT) fused to nCas9 and an extended gRNA, called a prime editing gRNA (pegRNA) that has a 3' extension. The pegRNA encodes both the location of editing (via the spacer sequence), and the edit to be introduced (via the 3' extension). Following DNA binding and nicking of the PAM-containing strand, the RT directly appends a portion of the 3' extension of the pegRNA sequence onto the broken DNA end. In this manner, PEs can install any type of small modification into the genome in a programmable and precise manner. The quick establishment of Prime Medicine to develop PEs into therapeutics is a sign of additional exciting clinical trials in the future.

1.5 Dissertation overview

The primary objectives of this dissertation seek to contribute to the field of genome editing through studies of two of the more recent entries to the field: base editing and prime editing. Primary scientific study will be addressed in Chapters 2 and 3, which detail a mechanistic investigation of PEs. Additional scientific study will be addressed in Chapter 4, which details attempted novel tool development for BEs, as well as future outlooks. Finally, Chapter 5 will briefly cover individual growth and development throughout graduate study away from the bench and traditional scientific research.

1.6 Acknowledgments

Chapter 1 is reproduced, in part, with permission, from: Porto, E. M., Komor, A. C., Slaymaker, I. M., and Yeo, G. W. (2020) Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.,* 19, 839-859. and Porto, E. M., and Komor, A. C. (2023). In the business of base editors: Evolution from bench to bedside. *PLOS Biology,* 21, e3002071. The dissertation author was the primary author on all reprinted materials.

Chapter 2

Studies of Point Mutation Introduction by PE2/PE3 and Cell Cycle Dependence

2.1 Introduction

Prime editing is a recently developed genome editing technology (36) which, like base editing, utilizes the programmable DNA binding capabilities of CRISPR/Cas9, but utilizes an additional enzyme to perform DNA modification chemistry. Prime editors (PEs) can achieve a wide range of user-defined edits including single base pair conversions (both transitions and transversions), small insertions/deletions (indels), and large insertions/deletions, which make them attractive candidates for basic research applications as well as potential therapeutics for treating

genetic diseases. Due to their relatively new addition to the field, there is much to be learned about the mechanics of PEs and their interactions within the cellular environment.

PEs are comprised of three core components: a Cas9 nickase (Cas9n), a reverse transcriptase (RT), and an engineered prime editing guide RNA (pegRNA) (**Figure 2.1A**). The original development of prime editing (36) included PE1 (Cas9n fused to the wild type Moloney murine leukemia virus [M-MLV] RT), PE2 (Cas9n fused to an engineered M-MLV RT), PE3 (PE2 used with the addition of a separate nicking gRNA), and PE3b (PE3 with a specific type of nicking gRNA). PEs function by Cas9n binding to a genomic locus of interest (the protospacer), facilitated by base-pairing with the 5' end of the pegRNA (the spacer) and the presence of a protospacer adjacent motif (PAM), a short DNA sequence recognized by the Cas9n protein (**Figure 2.1A**). Cas9n nicks the DNA strand opposite the protospacer, which can then anneal to the primer binding sequence (PBS) portion of the pegRNA 3' extension (**Figure 2.1A**), "priming" the DNA for reverse transcription. The RT then reverse transcribes the reverse transcription template (RTT) portion of the pegRNA 3' extension sequence directly onto the primed genomic DNA strand. This results in a "prime editing intermediate" (**Figure 2.1A**) comprised of a mismatch or bulge combined with a flap. PE3 and PE3b employ the same prime editing machinery but use an additional gRNA to nick the genomic DNA across from the strand that was edited (**Figure 2.1A**), increasing editing efficiencies but also resulting in increases in undesired indels.

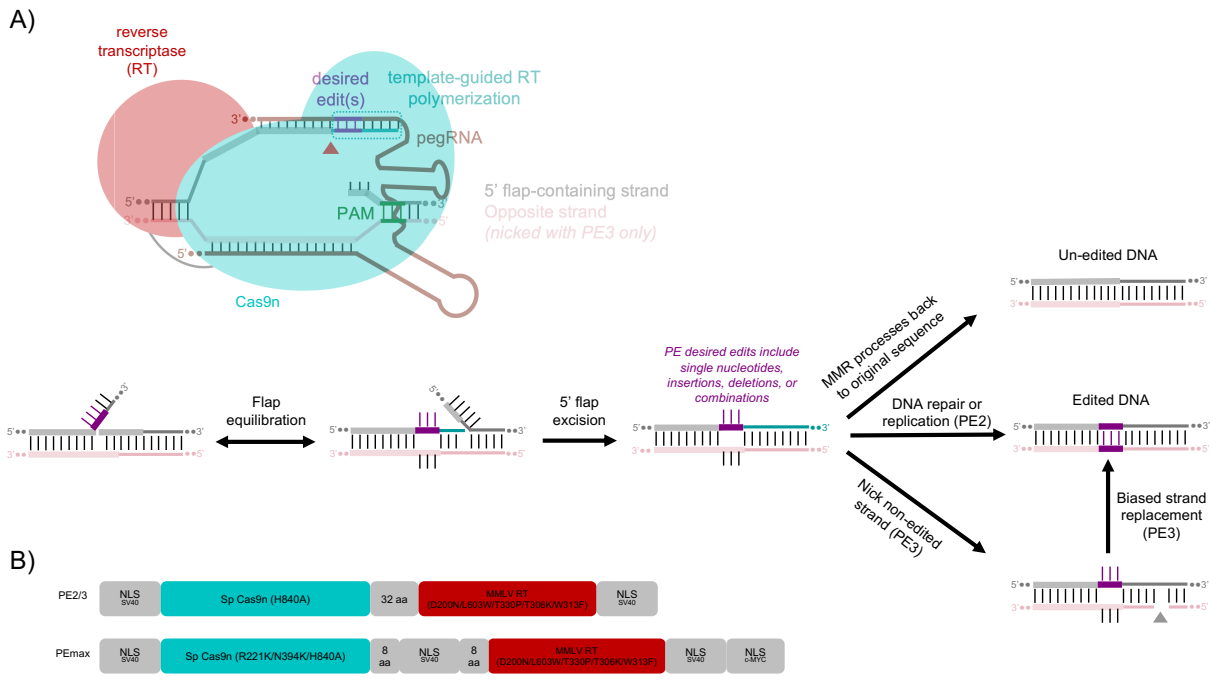


Figure 2.1. Overview of prime editing. A) The PE system components are shown, including the MMLV reverse transcriptase (RT, red), which is tethered to Cas9n (teal). Binding of Cas9n to the target site, or protospacer (pink strand), is dictated by the sequence of the 5' end of the pegRNA (brown), and the presence of a protospacer adjacent motif (PAM, green). The non-targeting strand (grey), which will be subsequently edited and contain a 5' flap, is nicked by Cas9n (red triangle). The 3' end of the pegRNA anneals to the nicked DNA strand, "priming" it for reverse transcription. The RT then synthesizes new DNA onto the 3' end of the grey nicked DNA (purple and teal), using the additional 3' end of the pegRNA as a template. After prime editing has occurred, the desired edit (purple) is present on a 3'-flap (bottom schematic), which can undergo flap equilibrium to produce an intermediate in which the newly synthesized DNA strand (teal) is "mismatched" with the original pink DNA strand, and the original grey strand is present as a 5' flap. When installing point mutations, this intermediate would contain a mismatch, and when installing insertions or deletions, this intermediate would contain a bulge. For the intended DNA edit to become permanently incorporated, the 5'- flap must be excised, producing a mismatch or bulge. The native cellular DNA repair machinery must then resolve this intermediate and replace the unedited pink DNA strand to make the desired edit permanent. In the case of PE3, the opposite strand (pink) is also nicked by an additional nicking gRNA to boost on-target editing efficiency through biased strand replacement. **B**) Construct map of the PE2 and PE2max editors used in this study.

During the development of improved PE variants, it was shown that knockdown of the mismatch repair (MMR) complex MutLα resulted in enhanced editing efficiencies by PE3 and an increase in product purity by PE3 (37, 38). MMR is an endogenous DNA repair pathway that

resolves DNA base mismatches and small bulges. Overall, it was suggested that the MMR machinery impedes prime editing activity by preferentially replacing the edited DNA strand. Notably, this work led to the development of PE4 and PE5, which PE2 or PE3 are co-expressed with a dominant negative mutant of MLH1 (MLH1dn) (37). MLH1 is one half of the MutLα heterodimer. Overall, PE4 and PE5 employ the same prime editing intermediates, but manipulate cellular processing of the prime editing intermediate. Notably, while this work identified a key mechanism of prime editing reversal, the mechanisms by which prime editing intermediates are processed into the desired outcomes are currently unknown.

Expression levels and therefore activities of different DNA repair proteins vary during the difference phases of the cell cycle (39), which include Gap 1 (G1-phase), Synthesis (S-phase), Gap 2 (G2-phase), and Mitosis (M). In G1, the cell grows and prepares for DNA synthesis, in S-phase, the cell duplicates the entirety of its genetic material, in G2-phase the cell prepares for division, and in M-phase, the cell divides. Because the activities of DNA repair proteins can be heavily influenced by the phases of the cell cycle, the efficiency and precision of genome editing agents can be impacted by cell cycle phases as well. For example, "traditional" genome editing agents that employ double-stranded DNA breaks (DSBs) rely on the homology-directed repair (HDR) pathway to install desired edits. As the HDR machinery is only expressed in the late S and G2 phases of the cell cycle, the use of DSB-reliant tools for precision editing is confined to actively dividing cells, which go through S and G2 phase. In contrast, we recently found that base editors, which install C·G to T·A and A·T to G·C point mutations via uracil and inosine intermediates, respectively, function independently of the cell cycle when the DNA strand opposite the uracil or inosine intermediate is nicked. The relationship between prime editing efficiency and the phases of the cell has not yet been explored. MMR, the repair pathway shown to revert prime editing

intermediates back to the unedited sequence, is active throughout the cell cycle, but upregulated during S- and G2-phases (39).

In this study, we use PE2 and PE3 to install point mutations and small indels at a variety of genomic loci in unsynchronized, G1-synchronized, and G2/M-synchronized cells to quantify the cell cycle dependence of prime editing. We chose to study PE2 and PE3 rather than PE4 and PE5 to avoid any potential confounding factors due to DNA repair manipulation. We were particularly interested in observing the differences in the cell cycle dependencies of PE2 versus PE3 that would be caused by the additional downstream nick, which boosts editing efficiency. Notably, in our study of the cell cycle dependence of base editors, we observed a dependence on S-phase in the absence of nicking of the unedited strand. Furthermore, our use of different pegRNAs targeting a variety of point mutations and indels at different genomic loci was intended to provide us with a broad exploration of how processing of the various types of prime editing intermediates can be impacted by the phases of the cell cycle. The results from our study provide insights into the DNA repair mechanisms governing processing of PE intermediates into desired outcomes and suggest that different PE intermediates may be processed by different pathways.

2.2 Results

2.2.1 Establishing a timeline of prime editor activity

To establish the experimental conditions for combining cell synchronization with prime editing, we first quantified the timeline of on-target editing activity by PE2. Similar to our previous work characterizing the timeline of BE activity (40), we conducted a time course experiment to monitor the efficiency of PE2 activity across multiple genomic loci. To accomplish this, we co-transfected HEK293T cells with plasmids encoding PE2 and one of three pegRNAs designed to

introduce point mutations at different genomic loci (see **Table 2.1** for full sequences) (36). We then lysed the cells and extracted genomic DNA (gDNA) at 6-, 12-, 18-, 24-, 36-, 48-, and 72-hours post-transfection. Genomic loci of interest were PCR amplified from the gDNA and subjected to high-throughput sequencing (HTS). HTS data was analyzed using CRISPResso2 to quantify on-target point mutation introduction efficiencies (41). We observed a gradual increase in editing at all three genomic loci, with editing peaking at 48 or 72 hours (**Figure 2.2A**). Editing at the *HEK3* and *EMX1* loci peaked at 72 hours post-transfection (2.9 ± 0.5% and 4.3 ± 0.8% respectively, mean ± SD for n = 3 biological replicates per site), while editing at the *RNF2* loci showed peak editing at 48 hours post-transfection (4.3 ± 0.6%), with editing at 72 hours within error of the 48-hour time point. Interestingly, this profile of editing activity mirrors that of adenine base editing, but contrasts with cytosine base editing, in which we observed editing peak around 36-48 hours post-transfection.

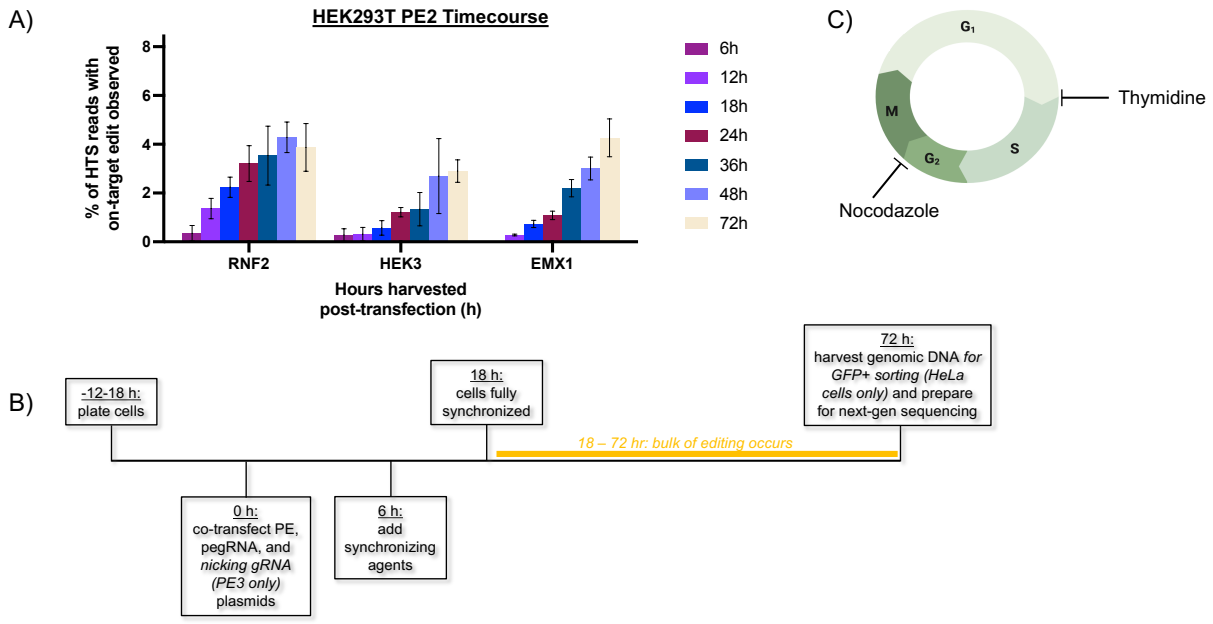


Figure 2.2. PE2 HEK293T time course experiment A). HEK293T cells were transfected with PE2 and pegRNAs (targeting single point mutations in *RNF2*, *HEK3*, and *EMX1* loci) and systematically lysed at 6-, 12-, 18-, 24-, 36-, 48-, and 72-hours post-transfection. Genomic DNA (gDNA) was extracted, target loci were PCR amplified and subjected to HTS. Prime editing efficiencies (percent of total HTS reads with the desired point mutations defined by the pegRNAs) were quantified with CRISPResso2. Resulting prime editing efficiencies are shown. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. **B**) Finalized experimental timeline for cell synchronization experiments. Cells were plated on 48-well (HEK293T cell experiments) or 6-well (HeLa cell experiments) plates 12-18 hours before transfection. Cells were transfected with PE (PE2 in HEK293T cells, PE2max in HeLa cells), pegRNA (targeting point mutations or small, 3-bp indels in *RNF2*, *HEK3*, *EMX1*, or *HEK4* loci), and nicking gRNA (PE3 only) plasmids. 6 hours after transfection cells were treated with the synchronization agents thymidine or nocodazole (unsynchronized cells were not treated). 12 hours after small molecule addition (18 hours post-transfection), the treated cells were fully synchronized. 72 hours after transfection cells were harvested, sorted for GFP positive fluorescence (HeLa only), and lysed. gDNA was extracted. and PCR amplified according to target loci and subjected to HTS. **C**) Stages of the cell cycle are shown. Thymidine synchronizes cells at the G1/S border, and nocodazole synchronizes cells at the G2/M border.

We previously demonstrated that HEK293T cells are fully synchronized at the G1/S boundary by thymidine and the G2/M boundary by nocodazole after 12 hours of small molecule treatment (**Figure 2.2C**) (40). Further, delaying the addition of small molecule by at least 6 hours after transfection was necessary to prevent decreases in transfection efficiency and genome editor

expression level (**Figure 2.2B**). Taking these HEK293T cell synchronization requirements into account, we elected to use a harvest time of 72 hours post-transfection for synchronization experiments to maximize the amount of prime editing that would occur between 18 hours post-transfection (when cells are fully synchronized) and harvesting. Specifically, 80 and 83% of overall editing occurred between 18 and 72 hours at the *HEK3* and *EMX1* loci, respectively, while 48% of overall editing occurred between 18 and 48 hours at the *RNF2* loci.

We elected to use HeLa cells as an orthogonal cell line to HEK293T, as the processing of PE3 intermediates was recently studied in this cell line (37). We thus characterized cell synchronization by thymidine and nocodazole in this cell line. We adapted our HEK293T synchronization protocol for HeLa cells (see Methods) and confirmed cell synchronization and cell viability at the 72-hour post-transfection harvest time. To analyze cell synchronization, we followed the experimental timeline indicated in **Figure 2.2B**, and at 72 hours post-transfection, cells were fixed, stained with propidium iodide, and analyzed by flow cytometry to determine DNA content, which is used to evaluate what phase of the cell cycle each cell is in. Unsynchronized cells had a cell cycle distribution similar to that observed for HEK293T cells (30.8% ± 2.89% of cells in G1 phase, 50.63% ± 3.75% of cells in S phase, and 15.03% ± 2.11% of cells in G2 phase, **Figure 2.3**). As expected, cells treated with thymidine were nearly evenly divided between G1 (50.17% ± 1.99%) and S (42.57% ± 1.19%), while cells treated with nocodazole were primarily in G2 (75.93% ± 5.09%) (**Figure 2.3**). Having confirmed that HeLa cells maintain cell synchronization at 72 hours post-transfection, we next analyzed the viability of cells at the 72-hour harvest timepoint. The experiment was repeated, except at harvest time, cells were examined for viability using a trypan blue stain. The viability of unsynchronized cells was high (89.29% ± 3.50%), with that of nocodazole-synchronized cells (57.20% ± 10.63%) and thymidine-

synchronized cells (72.71% ± 4.98%) slightly lower (**Figure 2.4**). The reduction in viability following extended (66 hours) cell synchronization with nocodazole was expected, as nocodazole achieves synchronization by disrupting microtubule dynamics (42). However, given the viability was >50% for all samples, we chose to proceed with this timeline.

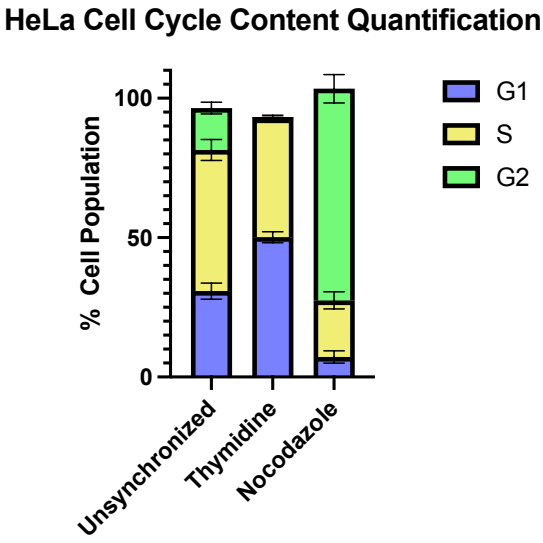


Figure 2.3. Example of HeLa cell synchronization quantification after 66 hours of small molecule treatment. Following the analysis outlined in Supplementary Figure 2, percentages of cell populations for unsynchronized, thymidine (G1/S-synchronized), and nocodazole (G2/M synchronized) treated cells are shown. Total cell populations exceed or fall short of 100% due to the manual gating strategy used which employed wide margins between each cell phase population. Graphed values for all conditions: mean ± SD for n = 3 biological replicates per condition.

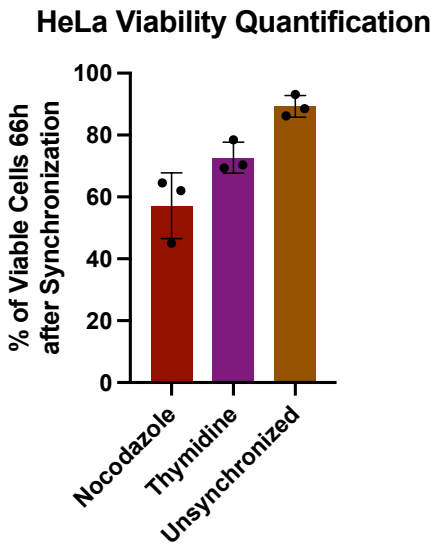


Figure 2.4. HeLa cell viability after 66 hours of synchronization. Cells were harvested at 72-hours post-transfection (equaling 66 hours total of exposure to the small molecular synchronizing agents thymidine and nocodazole) and stained with trypan blue. Cells were immediately placed on a hemocytometer and percentages of viable cells were determined and plotted above.

With the experimental timeline shown in **Figure 2.2B** verified for HEK293T and HeLa cells, we next set out to test how prime editing efficiency depends on the phase of the cell cycle across different genomic loci for the installation of point mutations as well as small insertion and deletion (indel) edits.

2.2.2 Testing synchronization effects on prime editing efficiencies by PE2 and PE3 to introduce point mutations and small indels

As previously mentioned, we sought to characterize the cell cycle dependence of prime editing efficiencies across different genomic loci, and for both point mutation and small indel edits by both PE2 and PE3 in both HEK293T and HeLa cell lines. This broad set of edit types and editors (point mutations and indels, by both PE2 and PE3) was included to cover a wide range of editing intermediates. Specifically, point mutation introduction involves a mismatch with a 5’ flap

with (PE3) or without (PE2) a nick on the strand opposite the 5' flap Indel introduction involves a bulge with a 5' flap with (PE3) or without (PE2) a nick on the strand opposite the 5' flap. We were interested in observing how cellular processing of each of these intermediates is impacted by the different phases of the cell cycle.

2.2.3 Point mutation introduction by PE2 (no nicking of the opposite strand)

We selected previously optimized pegRNAs (36) that install point mutations at four different genomic loci: *RNF2* (an exon-intron junction in the *RNF2* gene), *HEK3* (a non-coding region on chromosome 9), *EMX1* (a coding region in the *EMX1* gene), and *HEK4* (an enhancer region in chromosome 20). Following the experimental timeline and conditions developed previously (**Figure 2.2B**), we transfected HEK293T cells with plasmids encoding PE2 and one of the four selected pegRNAs, added synchronization agent 6 hours post-transfection, and lysed cells 72 hours post-transfection. The gDNA was harvested, and genomic loci of interest amplified and sequenced with HTS. CRISPResso2 was used to quantify point mutation introduction efficiencies. At two out of the four genomic loci tested (*RNF2* and *HEK4*), we observed no statistically significant changes in editing efficiencies with either synchronization condition, as determined by Welch's t-test (**Figure 2.5A**). At the other two sites (*HEK3* and *EMX1*), however, editing efficiencies decreased significantly upon G1/S (decreases of 3.5 ± 1.5-fold for *HEK3* and 3.8 ± 0.9-fold for *EMX1*, **Figure 2.5A**) and G2/M (decreases of 5.0 ± 0.9-fold for *HEK3* and 2.9 ± 0.4-fold for *EMX1*, **Figure 2.5A**) synchronization. Collectively, these data suggest a potential locus-, strand-, or mismatch type-specific dependency on S-phase for either installation or processing of PE intermediates. Notably, the *EMX1* and *RNF2* protospacers (which target actively transcribed regions) target different strands of the genomic DNA; in the *EMX1* sample, the sense strand will

contain the flap, while in the *RNF2* case the flap will be on the antisense strand (**Figure 2.6**). Furthermore, mismatch identity can impact repair efficiency by MMR, and three of the four sites we tested here proceed through different types of mismatches (**Figure 2.6A-D**). While HEK293T cells do not have a fully functional MMR pathway (they lack the MutLα heterodimer, which plays a role in sensing and installing nicks during MMR) (43), processing of these mismatch-containing intermediates to the desired outcome by an additional DNA repair pathway may also vary depending on the identity of the mismatch.

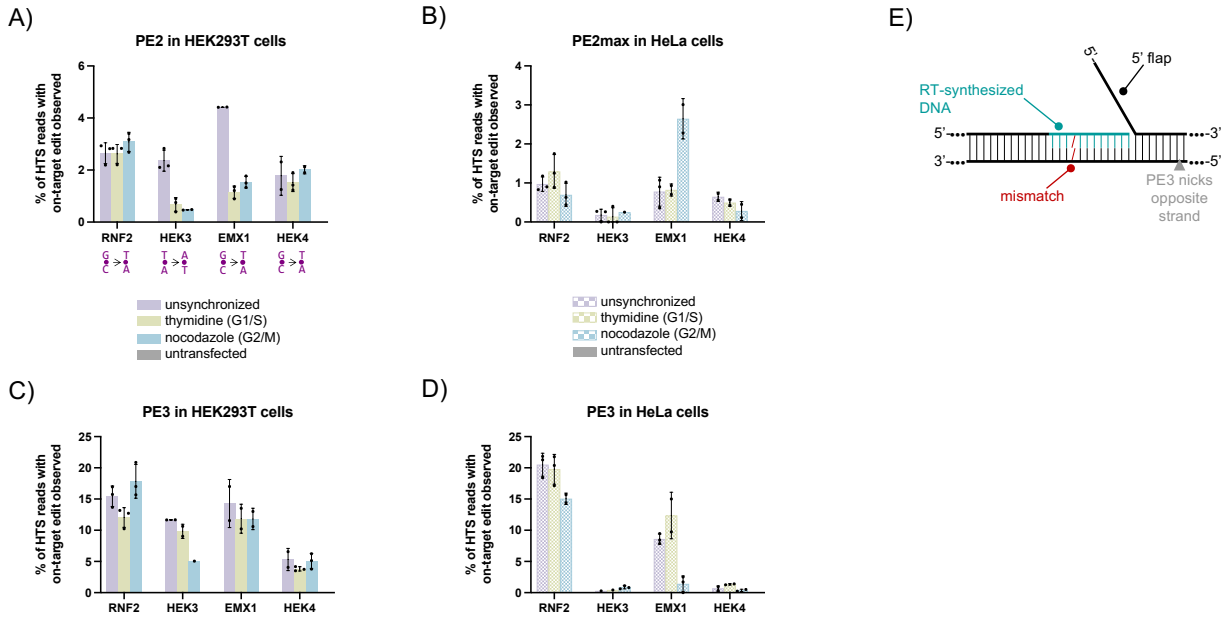


Figure 2.5. Cell cycle synchronization effects on point mutation introduction efficiencies by PE2 and PE3. Cells were transfected with PE2 (**A and C**) or PE2max (**B and D**), pegRNA, and nicking gRNA (**C and D only**), synchronization agents were added 6 hours post-transfection (thymidine for G1/S synchronization or nocodazole for G2/M synchronization), and cells were lysed at 72 hours. As a negative control, cells were untransfected. The genomic DNA was extracted, and target loci were PCR amplified and subjected to HTS. Prime editing efficiencies (percent of total HTS reads with the point mutations indicated in A introduced) were quantified with CRISPResso2. Prime editing efficiencies by PE2 in HEK293T cells (**A**), PE2max in HeLa cells (**B**), PE3 in HEK293T cells (**C**), and PE3 in HeLa cells (**D**) are shown. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. (**E**) General schematic of prime editing intermediates when installing point mutations is shown. For specific sequences, see **Figure 2.6A-D**.

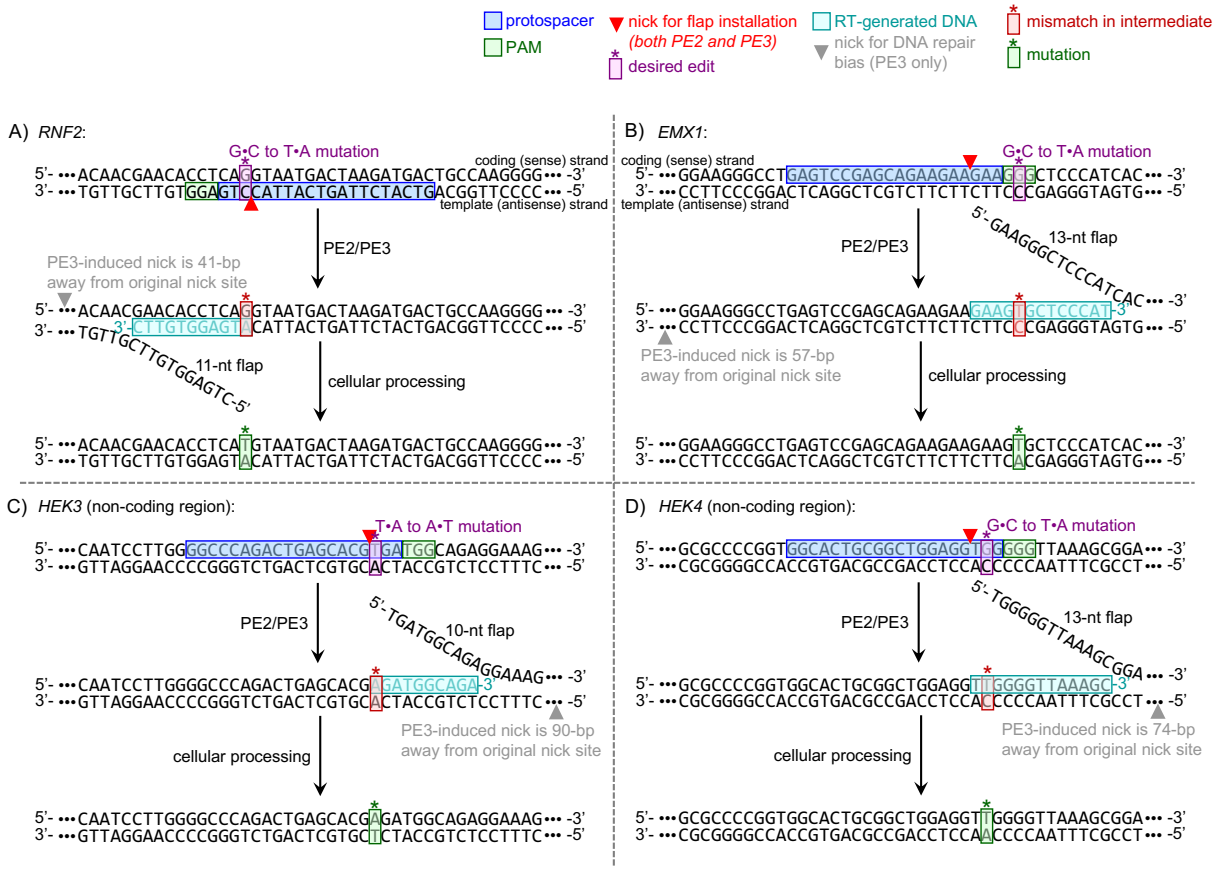


Figure 2.6. Sequences of prime editing intermediates for point mutation introductions. Following editing with the PE2 or PE3 system, the desired point mutation edit (purple) will be permanently incorporated in the DNA (green) through native cellular processing for the RNF2 (**A**), EMX1 (**B**), HEK3 (**C**) and HEK4 (**D**) loci. The excised 5' flap sequences resulting during the intermediate step are shown and vary in length according to designed pegRNAs.

We repeated this experiment in HeLa cells but used the PE2max variant (which improves nuclear localization, expression, and DNA nicking through architecture optimization, yet uses the same editing intermediates as PE2) (37), scaled up the transfection, and used fluorescence activated cell sorting (FACS) to sort for transfected cells to compensate for the much lower transfection efficiency in this cell line. Overall editing efficiencies in this line were still reduced approximately 2-fold compared to HEK293T cells, which may be due to the fully functional MMR pathway in HeLa cells. We again observed no statistically significant changes in editing efficiencies with either synchronization condition compared to unsynchronized cells for the *RNF2*

and *HEK4* samples (**Figure 2.5B**). Editing rates at the *HEK3* site were quite low (less than 0.25%) and we were therefore unable to robustly observe trends at this site. We observed no statistically significant changes in editing efficiency at the *EMX1* site upon synchronization in G1/S with thymidine (0.82% ± 0.16% versus 0.78% ± 0.37%), but we did observe a 3.25 ± 1.25-fold increase in editing at this site upon synchronization in G2/M with nocodazole (to 2.65% ± 0.52%). The difference in trends between the HEK293T and HeLa cell data at the *EMX1* site may be due to the differences in MMR functionality between the two cell lines. Nevertheless, these data do support our observations in HEK293T cells of a potential locus-, strand-, or mismatch type-specific dependency on the cell cycle, as for certain samples, prime editing efficiency was unaffected by cell cycle synchronization. We next sought to characterize how editing by PE3, which functions by nicking the DNA strand opposite the flap, changes with respect to cell synchronization.

2.2.4 Point mutation introduction by PE3 (with nicking of the opposite strand)

We next repeated our HEK293T experiments, but included an additional nicking gRNA for each experiment, which directs the PE to nick the DNA strand across from the edit. The locations of the nicks are indicated in **Figure 2.6A-D**. In unsynchronized HEK293T cells treated with PE3, we observed an overall increase in editing ranging from 3.0- to 5.8-fold as compared to the corresponding PE2 treated samples (**Figure 2.5C**). Interestingly, we observed no statistically significant changes in editing efficiencies upon either synchronization condition at three of the four sites (*RNF2*, *EMX1*, and *HEK4,* **Figure 2.5C**). Editing at the *HEK3* site decreased 2.3-fold upon synchronization in G2/M by nocodazole, with no statistically significant decrease upon G1/S synchronization. These data suggest a more robust and cell cycle-independent mechanism of PE3 compared to PE2. This is similar to our discovery that base editors that employ nicking of the

unedited DNA strand function independently of the cell cycle, while non-nicking base editors rely on S-phase (40).

We then repeated these experiments in HeLa cells, again using FACS to sort for transfected cells. Editing efficiencies at the two non-coding loci (*HEK3* and *HEK4*) were below 1.33% for both synchronized and unsynchronized samples, with editing reaching as low as 0.29%, so we were unable to observe trends at these two sites (**Figure 2.5D**). This may suggest high rates of repair of mismatches in these non-coding regions in HeLa cells, as editing efficiencies at these two loci were quite low with PE2 as well (**Figure 2.5B**), but we observed high editing rates at the HEK3 site in HeLa cells when introducing a 3-bp deletion (next section). In unsynchronized HeLa cells treated with PE3, we observed an overall increase in editing of 21.4- and 11.0-fold as compared to the corresponding PE2 treated samples at the *RNF2* and *EMX1* sites, respectively (**Figure 2.5D**). Editing at the *RNF2* site again remained unaffected by G1/S synchronization, but we did observe a small (1.4-fold) decrease in editing upon G2/M synchronization. In contrast to our PE2 data, editing at the *EMX1* site displayed a statistically significant decrease in editing upon synchronization in G2/M (editing decreased from 8.59% ± 0.81% to 1.41 ± 1.28%), and again showed no statistically significant changes upon synchronization in G1/S by thymidine (**Figure 2.5D**). Again, the differences in trends between the HeLa cells and the HEK293T cells (trends in editing efficiency upon synchronization at the *EMX1* site, and extremely low editing efficiencies at the *HEK3* and *HEK4* noncoding loci) may be due to the fully functioning MMR pathway in HeLa cells. Nevertheless, we will note that in general, point mutation introduction by PE3 is more robustly independent of the cell cycle than that by PE2. Intrigued by these data, we next sought to explore the cell cycle dependence of introducing small (3-bp) insertion and deletion edits by both PE2 and PE3 at the same genomic loci.

2.3 Discussion

In this work, we sought to observe how prime editing is impacted by the phases of the cell cycle, as this can help shed light on the mechanisms by which prime editing intermediates are processed by the cell. This work is continued in Chapter 3, however to this point we have made several key findings: the introduction of point mutations is affected by phases of the cell cycle, in general, point mutation editing efficiency is less affected by the phase of the cell cycle when using a PE3 editor, which nicks the DNA strand across from the edit, than when using a PE2 editor, and finally relative increases in editing efficiencies by nicking the unedited strand with PE3 are much larger in HeLa cells than HEK293T cells, which suggests a more multifaceted role by the mismatch repair (MMR) pathway in processing intermediates than previously assumed. These discussions continue in Chapter 3.3.

2.4 Methods

2.4.1 Molecular cloning

Prime editor-expressing plasmids were obtained from Addgene, pCMV-PE2-P2A-GFP (Addgene #132776) and pCMV-PEmax-P2A-GFP (Addgene #180020) and used for transfections. pegRNA plasmids were generated from corresponding gRNAs with the appropriate spacer sequences (**Table 2.2**), which have been previously used in the lab (44). Blunt end cloning was then used to install RT- and PBS- extension sequences, the sequences of which were obtained from previous studies that optimized these parameters (36). Primers were obtained from Integrated DNA Technologies and phosphorylated using T4 polynucleotide kinase (NEB #M0201) with T4 DNA Ligase Reaction Buffer (NEB #B0202S) by incubating at 37ºC for 20 minutes and then heat

inactivated at 95ºC for 5 minutes. PCR was then performed using the phosphorylated primers according to manufacturer instructions for Phusion High-Fidelity DNA Polymerase (NEB #M0530). PCR products were purified using the QIAquick PCR Purification kit (QIAGEN #28104) according to manufacturer instructions. Purified PCR products were then ligated using QuickLigase (NEB #M2200) with the associated QuickLigase Buffer by incubating for 15 minutes at room temperature. The reaction mixture was then immediately transformed into chemically-competent NEB 10β cells (NEB #C3019H) according to manufacturer instructions. Endotoxin-free plasmids were prepared using the Zymo Midiprep kit (Zymo #11-550B) per manufacturer instructions. Sanger sequencing was used to verify correct plasmid sequencing.

2.4.2 Cell culture and transfections

HEK293T (ATCC CRL-3216) and HeLa (ATCC CCL-2) cells were cultured at 37ºC in DMEM (Gibco #10566-016) supplemented with 10% FBS (Gibco #10437-028). HEK293T cells were plated at a density of 50,000 cells per well in a 48-well plate with a total volume of 250 µL per well 12-18 hours before transfection. HeLa cells were plated at a density of 300,000 cells per well in a 6-well plate with a total volume of 2.5 mL per well 12-18 hours before transfection. To transfect HEK293T cells, a 12.5 µL DNA-transfection mixture was prepared per well, which was composed of 1.5 µL of Lipofectamine 2000 (Invitrogen #11668-019), 750 ng of PE plasmid, 250 ng of pegRNA plasmid, and in the case of PE3 experiments, an additional 83 ng of nicking gRNA plasmid. The total volume of 12.5 µL was achieved with Opti-MEM (Gibco #31985-070). The DNA-transfection mixtures were incubated at room temperature for 15 minutes then added to HEK293T cells. To transfect HeLa cells, a 250 µL DNA-transfection mixture was prepared per well, which was composed of 7.5 µL of TransIT-LT1 (Mirus #MIR-2300), 1875 ng of PEmax

plasmid, 500 ng of pegRNA plasmid, and in the case of PE3 experiments, an additional 249 ng of nicking gRNA plasmid. The total volume of 250 µL was achieved with Opti-MEM. DNA-transfection mixtures were incubated at room temperature for 30 minutes then added to HeLa cells. HEK293T- and HeLa-transfected cells were incubated at 37ºC for an additional 72 hours for allow for genome editing to occur. At 6 hours post-transfection, selected cells were treated with Thymidine (Sigma CAS 50-89-5) at a final concentration of 5 mM per well or Nocodazole (Sigma CAS 31430-18-9) at a final concentration of 200 ng/mL per well. Treated cells were then returned to 37ºC for an additional 66 hours. At 72 hours post-transfection, cells were harvested: Well media was removed, HEK293T cells were washed with PBS, and subsequently lysed directly in the well with 100 µL of lysis buffer (1.25 µL Proteinase K (ThermoFisher #00-3011), 10 µL 10mM Tris Buffer, 5 µL 0.05% SDS, total volume of 1 mL achieved with sterile HyClone water (Cytiva #SH30538.02)). Genomic DNA from lysed cells was incubated at 37ºC for 1 hour, heat shocked at 80ºC for 30 minutes, and kept at 4ºC until preparation for HTS. HeLa cells required FACS; preparation described below.

2.4.3 Fluorescence activated cell sorting (FACS) and flow cytometry

Transfected HeLa cells were sorted using GFP fluorescence prior to high-throughput sequencing. HeLa cells were washed with 1 mL PBS (phosphate buffered saline, Gibco #10010-023) and detached from the plate using 150 µL of Accumax (Innovative-Cell Technology #AM-105). Cells were collected with an additional 350 µL of cold PBS for a total of 500 µL, filtered into FACS tubes (Falcon #352235), and kept on ice until sorting. Cells were sorted on a BioRad S3e cell sorter, calibrated and quality control checked prior to each batch of sorting. GFP positive samples were quantified using the 525/30 nm channel. Untransfected, no-color cells were used to

set up gates prior to sorting. Dead cells and doublets were excluded from the collected cell population. See **Figure 2.7** for gating examples. The upper limit of cells collected from sorting was 25,000. Cells were collected in 500 µL of PBS and subsequently spun down at 300 rcf for 10 minutes then resuspended in 10 µL of lysis buffer. Resuspended lysed cells were incubated at 37ºC for 1 hour, heat shocked at 80ºC for 30 minutes and kept at 4ºC until preparation for HTS.

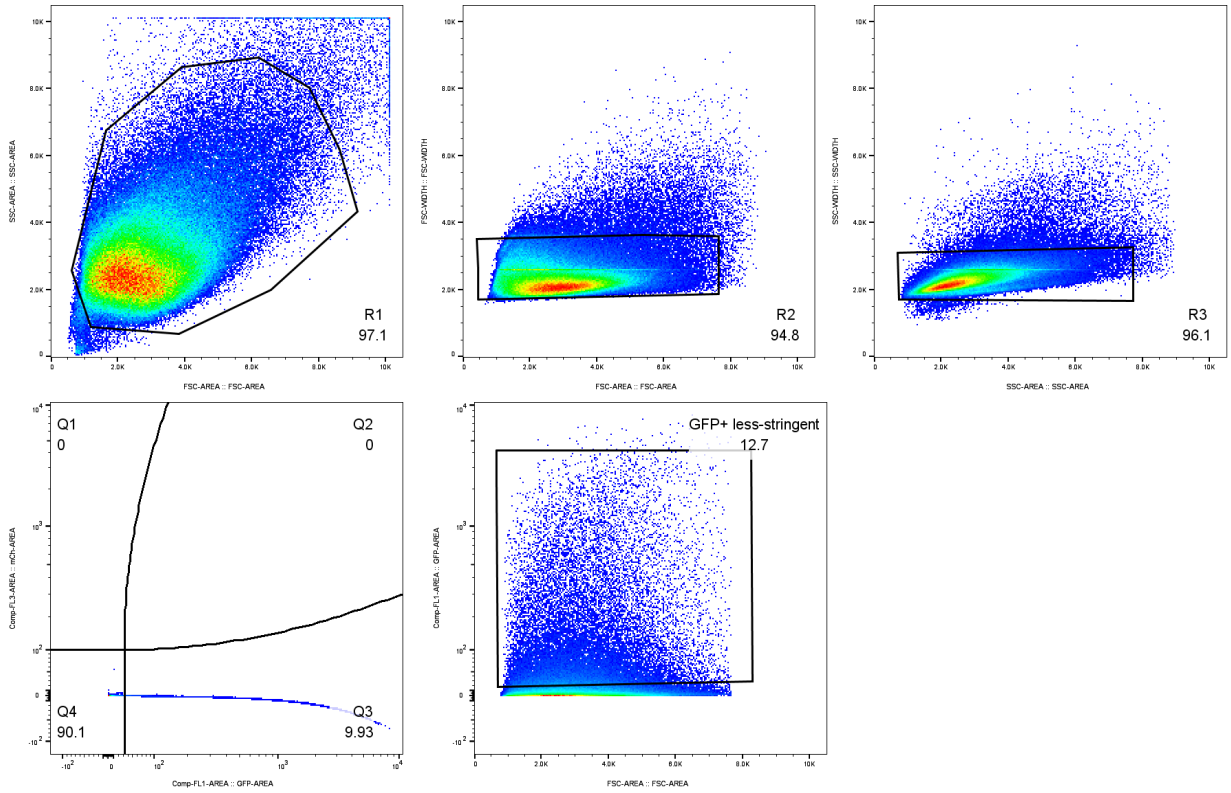


Figure 2.7. Example of fluorescence activated cell sorting (FACS) gates used for isolating GFP-positive HeLa cells. Gates were set using untransfected cells. Dead/non-viable cells were gated out using side scattering area against forward scattering area (top left). Doublets were gated out using forward scattering width against area (top middle) and side scattering width against area (top right). Gating of GFP-positive cells are shown in Q3 of the bottom left graph, which is 9.93% of the total population of cells in this example. To ensure all GFP-positive cells were collected, we expanded our gates slightly, resulting in an overall 'less-stringent' sort (bottom right).

HeLa cells measured for cell cycle content were detached from the plate using 150 µL of Accumax and transferred to 1.5 mL collection tubes. Cells were spun down at 200 rcf for 5 minutes

and washed with 1 mL of cold PBS and spun down again. Cells were resuspended in 500 µL of cold PBS at a concentration of $2x10^6$ cells/mL and vortexed gently. Cell suspension was added dropwise to an equal volume of 100% ethanol and stored at 4ºC for 24 hours. Cells were then spun down at 200 rcf for 10 minutes, washed with 1 mL cold PBS, spun down again, and resuspended in 300 µL of staining solution (2 mg DNAse-free RNAse A (ThermoFisher #EN0531), 1 mL Triton X-100 (Sigma CAS #9036-19-5), 10 mL PBS, 400 µL 0.5 mg/mL propidium iodide (ThermoFisher #P1304MP)). After staining, cells were filtered through FACS tubes and incubated at 37ºC while being protected from light by foil. Cells were then kept on ice until analyzed on the S3e cell sorter. Cells were analyzed similarly to those sorted; however the sorter was in acquisition mode rather than sort collection mode. Data was subsequently analyzed using FlowJo software.

2.4.4 High-throughput sequencing (HTS)

Following 72 hours of editing, harvesting, potential fluorescence sorting in the case of HeLa cells, and lysis, cells were prepared for HTS. PCR was used to amplify the genomic loci of interest using locus-specific primers (see **Table 2.1** for primer sequences) obtained from Integrated DNA Technologies. PCR was performed using Phusion High-Fidelity Polymerase according to manufacturer instructions, with the optional suggested modifications made to reduce primer dimers and denature high GC-content templates. A second round of PCR was then performed to barcode samples using unique primers. Following barcoding, PCR products were pooled according to size and gel extracted from a 1% agarose gel (QIAGEN #28704 gel extraction kit). These pooled libraries were purified a second time using the QIAquick PCR Purification kit (QIAGEN #28104). Purified libraries were quantified using the Qubit dsDNA high sensitivity kit (ThermoFisher #Q32851). Libraries were subsequently prepared according to the Illumina MiniSeq System

Denature and Dilute Libraries Guide (Illumina document #1000000002697). Libraries were sequenced on an Illumina MiniSeq using paired end sequencing.

2.4.5 Data analysis & statistics

FlowJo software (v10.9) was used to quantify cell cycle content. Gated populations of cells were manually fit according to the Dean-Jett-Fox Model to determine the percentage of cells within the gated population that were in G1, S, or G2 phase. See **Figure 2.8A-C** for an example.

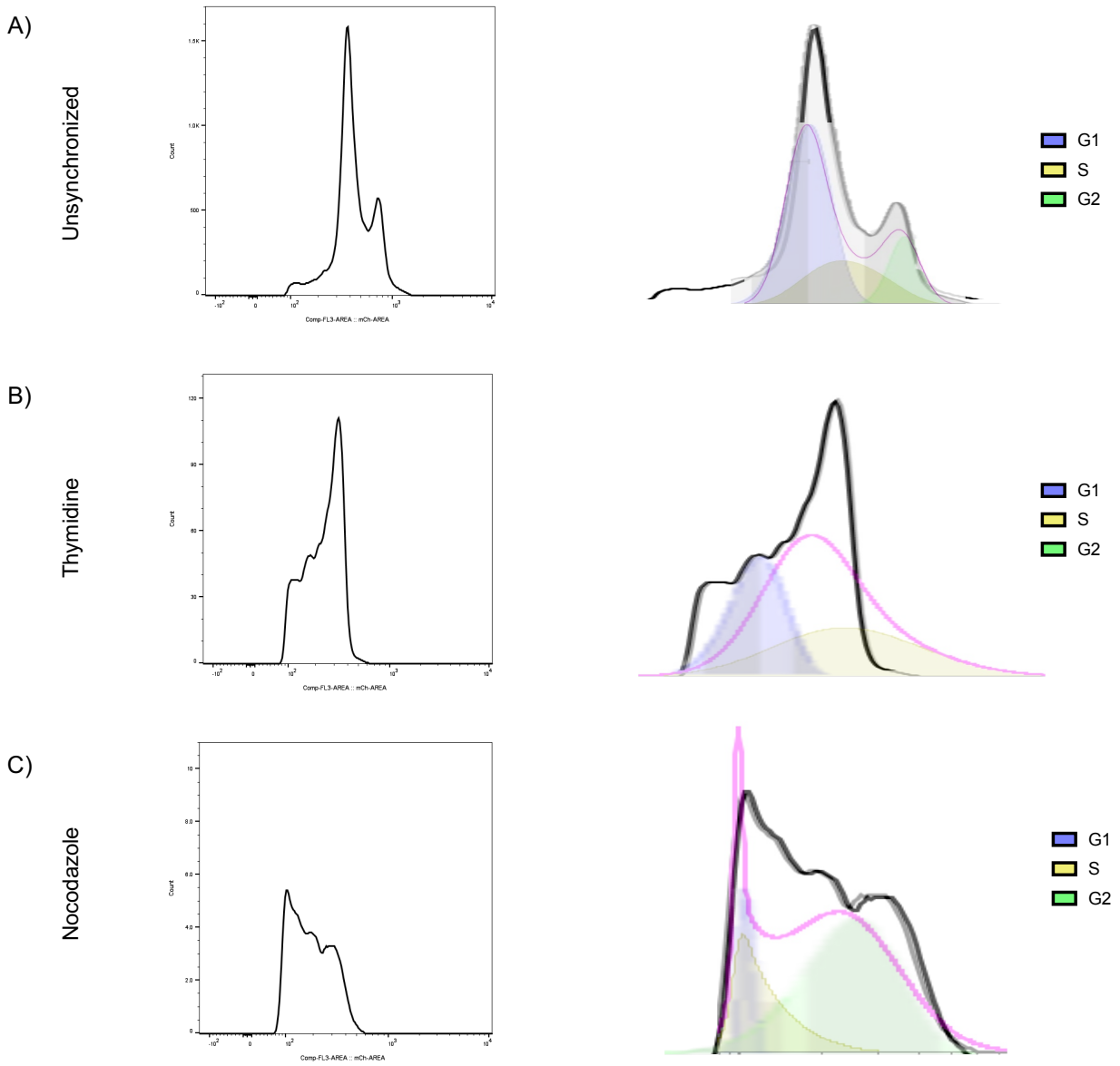


Figure 2.8. Example of gates used to quantify percent of cells in G1, S, and G2 phase of the cell cycle. Cells were fixed with ethanol and stained with propidium iodide to monitor cell cycle synchronization. Ungated cell distribution is shown at left for unsynchronized (**A**), thymidine treated (**B**), and nocodazole treated (**C**) cells. At right is an enlarged and transposed gating strategy for each cell condition. Cells in G1 phase are represented by blue peaks, cells in S-phase are represented by yellow peaks, and cells in G2 are represented by green peaks. Gates were manually fit for G1 and G2, and the sum of these gates is represented by the pink line. Remaining cells in S-phase are calculated by FlowJo software.

CRISPResso2 software was used off-line in batch mode to analyze Illumina MiniSeq-generated .fastq files. Files were sorted into batches according to the genomic loci of interest. Standard settings were used with a minor adjustment of lowering the default minimum align score from 60% homology to 50% homology. Amplicon sequence and guide RNA sequences were manually provided.

Statistical analysis on plots was performed using Prism v10, which auto-generated error bars according to standard deviation and mean for supplied values obtained from CRISPResso2 outputs.

Indel analysis plots were generated from individual indel histograms as part of the CRISPResso2 output. These values were binned according to sample and synchronization condition.

2.5 Acknowledgments

This work was supported by the National Science Foundation [Award # MCB-2048207]. E.M.P. was supported by the Molecular Biophysics Training Grant, National Institute of Health Grant T32 GM008326. E.M.P declares no competing financial interests. A.C.K. is a member of the SAB of Pairwise Plants, is an equity holder for Pairwise Plants and Beam Therapeutics, and receives royalties from Pairwise Plants, Beam Therapeutics, and Editas Medicine via patents licensed from Harvard University. A.C.K.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies.

Chapter 2 is reproduced, in full, with permission, from: Porto, E. M., and Komor, A. C. (2023). Studies of the relationship between early prime editors and cell cycle dependence. *Nucleic*

Acids Research, submitted. The dissertation author was the primary author on all reprinted materials.

Chapter 3

Studies of Small Insertion/Deletion Introduction by PE2/PE3 and Cell Cycle Dependence

3.1 Introduction

Continuing from the studies detailed in Chapter 2, we sought to expand on our work studying PE2 and PE3 variant-introduced intermediates. We next used previously designed and optimized pegRNAs for installing 3-bp insertions or deletions within the same genomic loci as our point mutation introductions (deletions at the *RNF2* and *HEK3* loci, and an insertion at the *EMX1* locus). Unlike point mutation installation, which proceeds through mismatch-containing

intermediates that are canonical substrates for the MMR pathway, the introduction of indels proceed through "bulged" or looped intermediates, the repair of which is less well-understood.

3.2 Results

3.2.1 Small insertion and deletion (indel) introduction by PE2 (no nicking of the opposite strand)

We repeated our synchronization experiments with PE2 in HEK293T cells and PE2max in HeLa cells. Notably, in HEK293T, we observed no statistically significant changes in editing efficiencies upon either synchronization condition at two of the three sites (*RNF2* and *HEK3*, **Figure 3.1A**), and only a modest 1.7-fold decrease in editing at the *EMX1* site upon G2/M synchronization (the editing efficiency in G1/S synchronized cells was not statistically significantly different to editing in unsynchronized cells at this site, **Figure 3.1A**). In HeLa cells, editing at the *RNF2* locus was less than 0.6%, preventing a rigorous analysis of editing trends at this site. Editing at the *HEK3* site decreased 2.6-fold upon G1/S synchronization and increased slightly but not statistically significantly upon G2/M synchronization. These data suggest that the installation of small insertions and deletions by PE2 is more consistently independent from the phases of the cell cycle than the installation of point mutations.

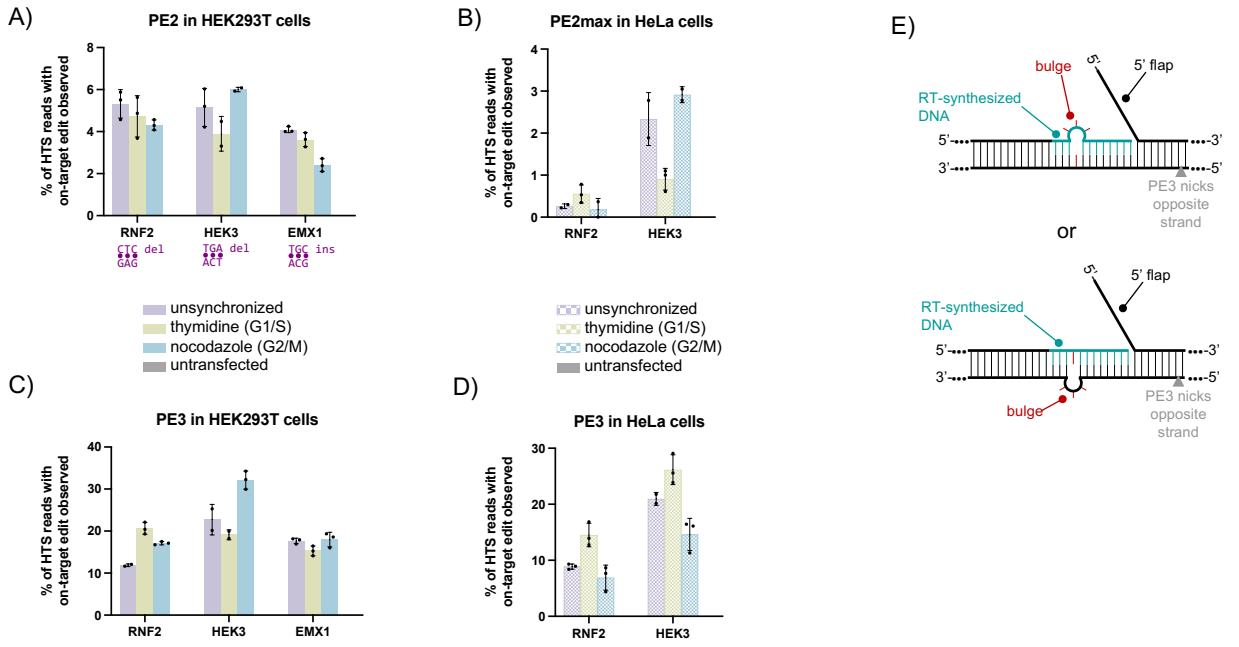


Figure 3.1. Cell cycle synchronization effects on small insertion and deletion (indel) introduction efficiencies by PE2 and PE3. Cells were transfected with PE2 (**A and C**) or PE2max (**B and D**), pegRNA, and nicking gRNA (**C and D only**), synchronization agents were added 6 hours post-transfection (thymidine for G1/S synchronization or nocodazole for G2/M synchronization), and cells were lysed at 72 hours. As a negative control, cells were untransfected. The genomic DNA was extracted, and target loci were PCR amplified and subjected to HTS. Prime editing efficiencies (percent of total HTS reads with the indels indicated in **A** introduced) were quantified with CRISPResso2. Prime editing efficiencies by PE2 in HEK293T cells (**A**), PE2max in HeLa cells (**B**), PE3 in HEK293T cells (**C**), and PE3 in HeLa cells (**D**) are shown. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. General schematics of prime editing intermediates when installing indels are shown (**E**).

3.2.2 Small insertion and deletion (indel) introduction by PE3 (with nicking of the opposite strand)

Finally, we used nicking gRNAs to nick the unedited strand for these 3-bp insertion and deletion edits and repeated our experiments in HEK293T and HeLa cells (**Figure 3.1E, Figure 3.2A-C**). Again, we observed much larger relative increases in editing efficiencies when comparing PE2 to PE3 for HeLa cells compared to HEK293T cells (in unsynchronized cells, these

increases ranged from 2.2- to 4.4-fold in HEK293T cells, while in HeLa cells they ranged from 9.0- to 34.2-fold), suggesting that MMR may play a more multifaceted roll in processing of PE3 intermediates than previous studies have revealed. In HEK293T cells, we observed minor changes in editing efficiencies upon synchronization at two of the three sites (changes ranged from 1.0-fold to 1.7-fold, **Figure 3.1C**). Similarly minor changes in editing efficiencies were also observed in HeLa cells at the *RNF2* and *HEK3* sites, where changes in editing efficiencies were within a 1.6-fold difference between synchronized and unsynchronized cells (**Figure 3.1D**). Interestingly, there was less of a difference in cell cycle dependence between PE2 and PE3 for indel introductions than point mutation introductions. This may be due to the more heterogeneous nature of the point mutation intermediates utilized here (as mentioned previously, the identity of the mismatch can impact DNA repair processing efficiencies) compared to our indel intermediates, which were all 3-nt bulges.

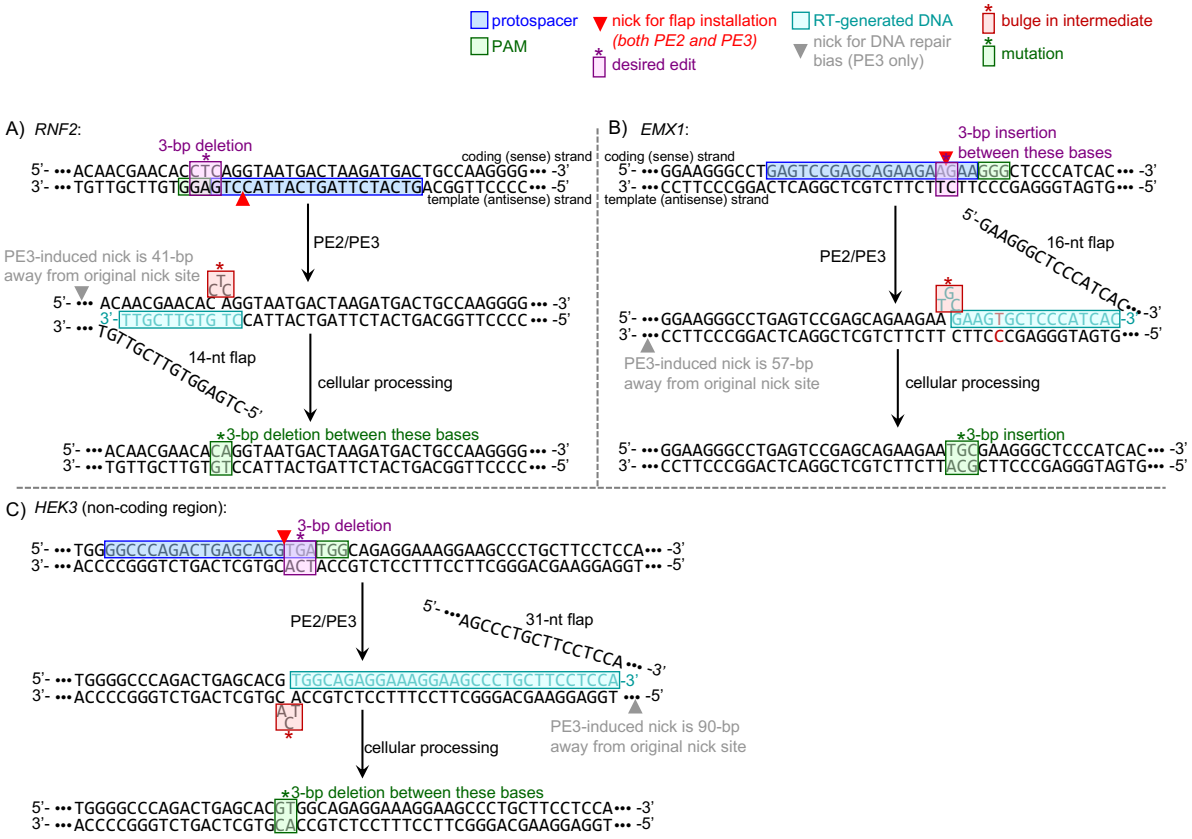


Figure 3.2. Sequences of prime editing intermediates for small insertion and deletion (indel) introductions. Following editing with the PE2 or PE3 system, the desired indel edit (purple) will be permanently incorporated in the DNA (green) through resolving the intermediate bulge and native cellular processing for the RNF2 (**A**), EMX1 (**B**), and HEK3 (**C**) loci. The excised 5' flap sequences resulting during the intermediate step are shown and vary in length according to designed pegRNAs

3.2.3 Examination of undesired indels introduced at edit sites

Finally, we examined undesired indel rates and sequences from all samples at the on-target loci (**Figure 3.3A-H**). Consistent with previous reports, we found that for matching PE2/2max and PE3 samples, undesired indel rates were higher in PE3 samples, with one exception (undesired indel rates in HeLa cells at the *HEK4* loci, for which we observed indel rates to be within error of each other, **Figure 3.3B and F**). Generally, we observed indel rates to be less than 1% for cells treated with PE2. For most sets of samples, we did not observe any statistically significant changes

in undesired indel introduction efficiencies for either synchronization conditions compared to unsynchronized cells (**Figure 3.3**). For sets of samples with statistically significant changes in overall undesired indels across different synchronization conditions, the fold changes were generally quite low, and there were no apparent trends (**Figure 3.3**). We additionally analyzed the sequences of these undesired indels and found that for all loci except for the *HEK4* site, there were no specific indel sequences that comprised greater than 1% of total sequencing reads for a given sample. Instead, many different indel sequences, each comprising less than 1% of total sequencing reads, collectively contributed to overall undesired indel rates. At the *HEK4* site, we consistently observed a 2-bp deletion, centered at the prime editor nicking location, for all samples including those treated with PE2 and PE3, and in both cell lines. Undesired indel rates at this locus were typically higher than the other loci and are likely caused by the nick that the prime editor installs within the protospacer.

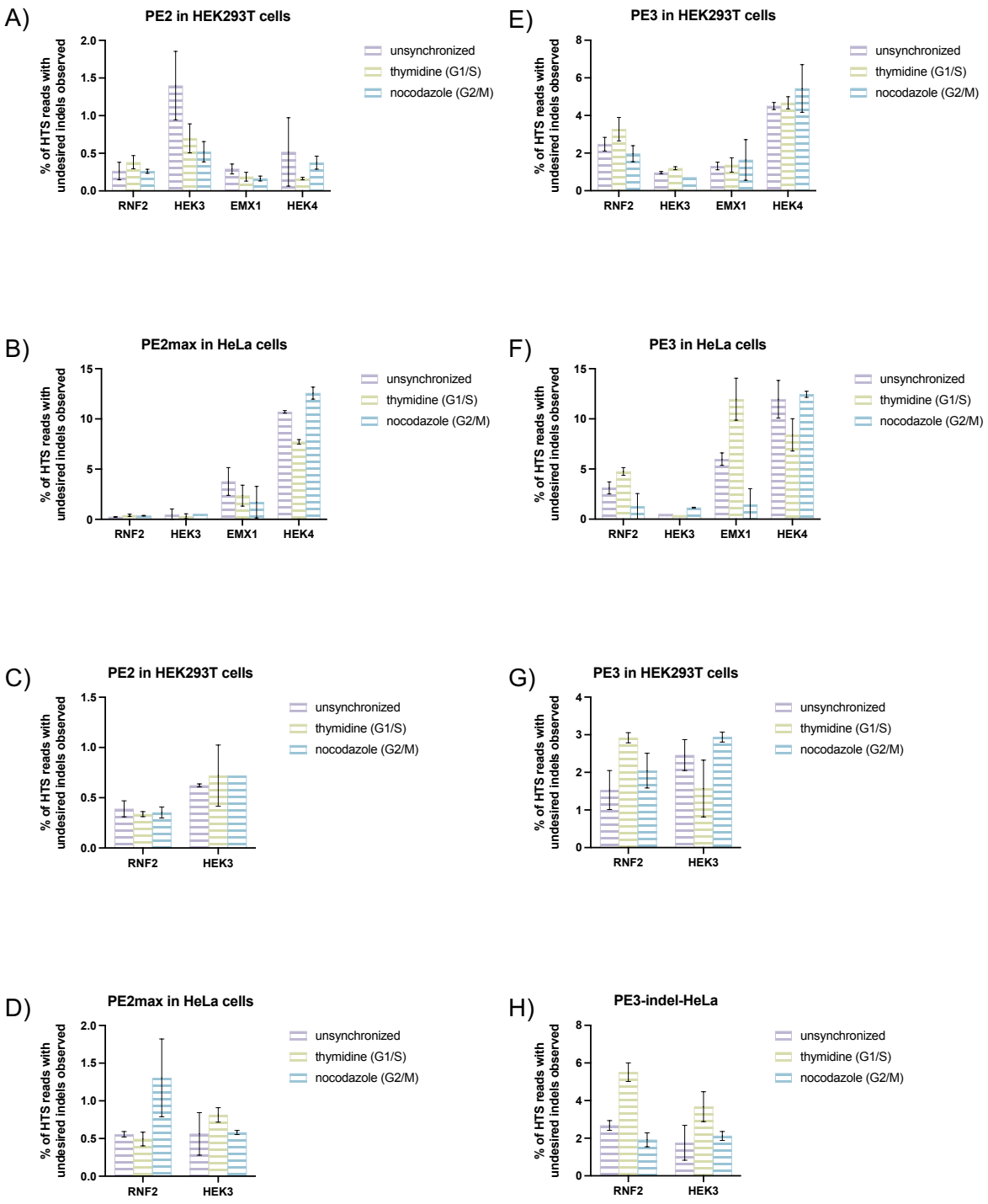


Figure 3.3. Undesired indel introduction efficiency quantification. Undesired indel rates were quantified by summing all individual indel rates from the CRISPResso2 indel histogram outputs. Shown are undesired indel rates for samples in which the desired edit was point mutations in HEK293T cells (PE2, **A** and PE3, **E**), point mutations in HeLa cells (PE2max, **B** and PE3, **F**), indels in HEK293T cells (PE2, **C** and PE3, **G**) and indels in HeLa cells (PE2max, **D** and PE3, **H**).

3.3 Discussion

We report here the first study to investigate the relationship between prime editing and the phases of the cell cycle. Precision genome editing with DSBs utilizes HDR, which is dependent on the late S and G2 phases of the cell cycle (45). This has resulted in major challenges when using DSB-mediated precision editing in non-dividing cells. The DNA repair factors that process prime editing intermediates are not currently fully elucidated and are likely to involve multiple pathways and vary depending on the type of edit that is being introduced (**Figure 2.5E and 3.1E**). In the absence of this mechanistic information, we utilized small molecule synchronization agents (**Figure 2.2C**) to synchronize and hold cells in late G1 or G2/M phase, performed prime editing, and quantified differences in efficiency and precision compared to unsynchronized controls. Overall, we observed that the introduction of small (3-bp) insertions and deletions by prime editing (with both PE2, which does not nick the unedited strand, and PE3, which does nick the unedited strand, **Figure 3.1E and Figure 3.2**) appears to be more consistently unaffected by cell synchronization conditions than the introduction of point mutations by prime editing (**Figure 2.5 versus Figure 3.1**).

The introduction of point mutations by prime editing is more consistently unaffected by cell synchronization conditions when using PE3 than PE2 (**Figure 2.5C-D versus Figure 3.1A-B**). We generally observed less consistent trends in cell cycle dependence with point mutation introduction than indel introduction, which may be due to the different types of mismatches present in the intermediates (**Figure 2.6A-D**), the different strands (sense versus antisense) being targeted, or the different loci (noncoding, un-transcribed regions versus coding regions) where editing is occurring. It is noteworthy that we used these same loci for our indel introduction samples, where we observed much less drastic differences in editing upon synchronization. Nevertheless, for

samples that displayed a cell cycle dependence in editing efficiency by PE2 in HEK293T cells (**Figure 2.5A**), editing efficiencies drastically decreased with both synchronization conditions, suggesting a reliance on S-phase. The locus-, strand-, or intermediate-specific dependence on the cell cycle is in contrast with our previous studies characterizing the cell cycle dependence of base editing, where both CBEs and ABEs that do not employ nicking of the unedited strand displayed a consistent dependence on S-phase across different loci, while CBEs and ABEs that employ nicking of the unedited strand were shown to function independently of the cell cycle. We also generally observed somewhat disparate trends in HeLa cells compared to HEK293T cells, which is likely due to the intact MMR pathway in HeLa cells. MMR has been shown to reduce prime editing efficiencies, and is generally expressed throughout the cell cycle, but is upregulated in S-phase. Differences in MMR processing efficiencies of these various intermediates in HeLa cells may be responsible for the disparate trends. In particular, we found much larger relative increases in editing efficiencies between PE2 and PE3 in HeLa cells compared to HEK293T cells. We attribute this to the additional nick in PE3 intermediates, which may be biasing MMR to replace the unedited strand at higher rates than the edited strand. Taken together with previous studies on the relationship between prime editing and MMR, these data suggest that MMR may play a more complex roll in processing prime editing intermediates that involves both reverting the intermediate back to the unedited sequence as well as converting it to the desired edit.

Just as PEs display a diversity in the types of edits they can facilitate, their intermediates are highly diverse as well (**Figures 2.6 and 3.2**). Our data here suggest that these different intermediates may be processed into their respective desired genome editing outcomes by different cellular factors. In particular, these results suggest that the intermediates used for point mutation introduction (mismatches with 5’ flaps) may be highly influenced by the identity of the mismatch.

Future studies to elucidate prime editing mechanisms may further shed light on these interesting observations.

3.4 Methods

3.4.1 Molecular cloning

Detailed in Chapter 2.4.1.

3.4.2 Cell culture and transfections

Detailed in Chapter 2.4.2.

3.4.3 Fluorescence activated cell sorting (FACS) and flow cytometry

Detailed in Chapter 2.4.3.

3.4.4 High-throughput sequencing (HTS)

Detailed in Chapter 2.4.4.

3.4.5 Data analysis and statistics

Detailed in Chapter 2.4.5.

3.5 Acknowledgments

This work was supported by the National Science Foundation [Award # MCB-2048207]. E.M.P. was supported by the Molecular Biophysics Training Grant, National Institute of Health

Grant T32 GM008326. E.M.P declares no competing financial interests. A.C.K. is a member of the SAB of Pairwise Plants, is an equity holder for Pairwise Plants and Beam Therapeutics, and receives royalties from Pairwise Plants, Beam Therapeutics, and Editas Medicine via patents licensed from Harvard University. A.C.K.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies.

Chapter 3 is reproduced, in full, with permission, from: Porto, E. M., and Komor, A. C. (2023). Studies of the relationship between early prime editors and cell cycle dependence. *Nucleic Acids Research, submitted*. The dissertation author was the primary author on all reprinted materials

Chapter 4

Studies of Novel Base Editor Expansion and Future Outlooks

4.1 Introduction

In recent years, the notion of "precision medicine" (treatment strategies based on a patient's individual characteristics, such as their genomic sequence), has exploded. At the same time, the field of genome editing (introduction of user-defined changes to the sequence of chromosomal DNA) has also seen rapid improvements which have been viewed as a major strategy to execute precise, therapeutic outcomes *in vivo*. The optimal genome editing tool would irreversibly edit any chromosomal position with high specificity and efficiency, and with zero undesired edits.(8) While

recent progress in the field has produced promising results, current genome editing methods are limited in their therapeutic relevance.

The use of the clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR-associated protein 9 (Cas9) system has become the standard method for specific genome editing. This system relies on the ability of the Cas9 endonuclease to introduce a double strand break (DSB) at a desired DNA sequence.(46, 47) When complexed with a short guide RNA (sgRNA) chimeric molecule through canonical RNA-DNA base pairing, Cas9 is guided to a target locus (the "protospacer").(46) Sequence complementarity between the sgRNA and protospacer, and the inclusion of a protospacer-adjacent motif (PAM) are requisites for DSB induction. Following induction, the DSB is resolved either by precise repair from an exogenous donor template through homology-directed repair (HDR) or non-homologous end joining (NHEJ). NHEJ is active during all phases of the cell cycle, thereby making it the more likely DSB repair pathway. As NHEJ does not use a template for repair, editing outcomes can result in high levels of undesired, random insertions and deletions (indels) at the site of the DSB.(48, 49) Ultimately, DSB-reliant genome editing is quite risky, thereby making it unsuitable for optimal, precise *in vivo* therapeutic development.

Base editing is an alternative genome editing technique that enables the direct, irreversible conversion of a single target DNA base in a precise, programmable manner without introducing a toxic DSB or requiring a donor template.(25) Given that approximately half of the known pathogenic genetic variants are caused by single nucleotide polymorphisms (SNPs), the need for tools capable of efficiently correcting SNPs is clear.(1) Base editors in their current scope meet that need, demonstrating the significance of this technology and emphasizing the importance of their continued development.

Current base editors employ a catalytically impaired Cas9 enzyme (dCas9 or Cas9n), where only DNA binding capability is maintained, fused to a ssDNA modifying enzyme.(25, 50) As in traditional CRISPR methodology, a sgRNA is used for complementary base pairing to the protospacer. After the base editor binds to its editing target, a small stretch of ssDNA is exposed, allowing for nucleotides within this "editing window" to be modified by the tethered enzyme. (25, 50) As it stands, only C•G to T•A and A•T to G•C base pair conversions have been achieved through the tethered cytidine- and adenosine- deaminase enzymes in base editor architectures. These base editors are referred to as cytosine base editors (CBEs) and adenine base editors (ABEs), respectively. Both CBE and ABE architectures employ the large *S. pyogenes* Cas9 (spCas9) enzyme, which at 4.1 kb, is incompatible with the optimal therapeutic delivery method, viral vectors. Furthermore, we are limited by the base pair conversions we can currently achieve through base editing (**Figure 4.1**).

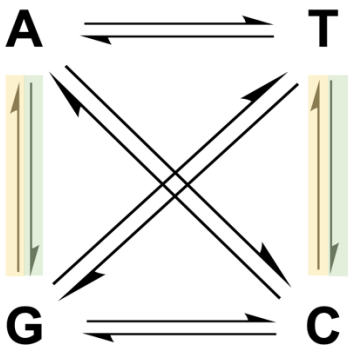


Figure 4.1. BE Conversions. Given the four DNA bases, a total of 12 base pair conversions are possible (represented by the 12 half-arrows). Green arrows represent base pair conversions achieved with ABEs, and yellow arrows are base pair conversions achieved with CBEs. Meanwhile, the 8 uncolored arrows represent substitutions we cannot yet achieve through base editing.

Viral vectors have the ability to package nucleic acids encoding for these programmable nucleases and carry them into target cells without degradation.(51) Common candidates for viral vectors are lentivirus, adenovirus, and adeno-associated virus (AAV). AAV is the preferred option as it does not integrate its own genome into the host nor elicit an immune response.(51, 52) However, AAV is limited by its small packaging capacity (~4.7 kb). Packaging large spCas9, along with a sgRNA, relevant promoters, and viral regulatory elements (all the requisite components for successful base editor delivery and activity) is therefore unsuitable for this optimal delivery method.

4.2 Development of a novel RlmN base editor

To expand the repertoire of base editors, we chose the radical S-adenosyl-L-methionine-dependent (SAM) methyltransferase RlmN for initial study. We believed the natural substrate promiscuity exhibited by RlmN makes it an interesting candidate base editor development. Such a base editor would utilize a modified methyl-2-adenosine (m^2A) base.(53) How exactly DNA polymerases handle the m^2A modification during replication has not been reported in the literature. This added methyl group is believed to impair A•U base pairing in RNA by forcing a Hoogsteen base pair conformation, suggesting the m^2A modified base would be mutagenic.(54) As a crucial first step, to determine the viability of an RlmN-derived base editor, we sought to characterize the mutagenicity of m^2A in mammalian cells.

Purchasing the m^2A base as a triphosphate was not a viable option, so we initiated a collaboration with the Tor lab at UCSD to generate m^2ATP. Following an extensive workup and purification process, we incorporated my purified, synthetic m^2ATP product into a plasmid backbone using terminal deoxynucleotidyl transferase (TdT) (**Figure 4.2A**). The final, ligated plasmid was subsequently transfected into HEK293T cells and genomic DNA (gDNA) was harvested 24 hours after transfection. gDNA was subsequently PCR amplified and prepared for next generation high throughput sequencing (HTS) on an

Illumina MiniSeq. Sequencing results were analyzed using a MATLAB script provided to the lab. As shown in **Figure 4.2B**, we saw a ~2.6% base pair conversion from T•m^{2}A to C•G. These mutagenicity results suggest that, when employed in the established base editor architecture, RlmN would result in the same editing outcome as the currently established ABE. Due to the high activity rate of the established ABE, as well as recent improvements to decrease its off-target editing activity, the need for an additional A to G base editor is nonexistent.(55, 56) As such, we are not further pursuing RlmN.

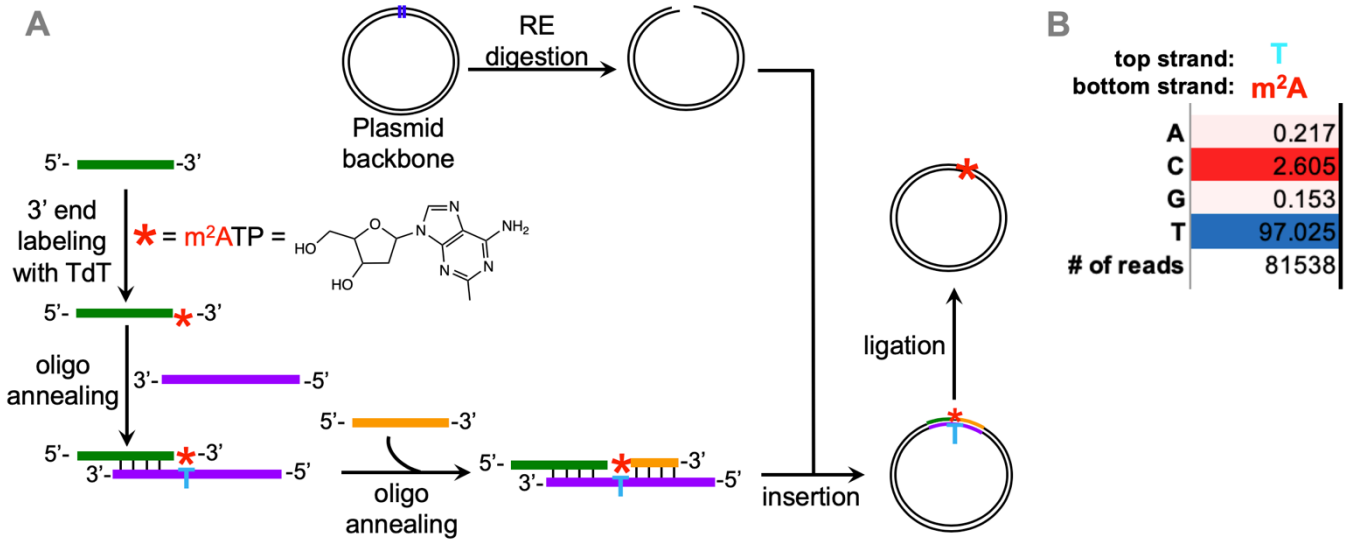


Figure 4.2. RlmN BE Development. A) Schematic of m^{2}A plasmid incorporation. Green, purple, and yellow oligos were purchased from IDT. TdT enzyme adds the m^{2}ATP base (red asterisk), which base pairs with a canonical thymine base (blue T) on a complementary oligo. 3-piece annealed oligo template is inserted into a restriction enzyme-digested plasmid backbone of choice. Final ligated plasmid is used for downstream study. **B)** HTS data shown for T•m^{2}A editing outcomes. Colorization corresponds to thresholds set during data analysis. Blue indicates an identity match while red indicates identity mismatch. Shading of the colors corresponds to numerical values assigned to dictate thresholds.

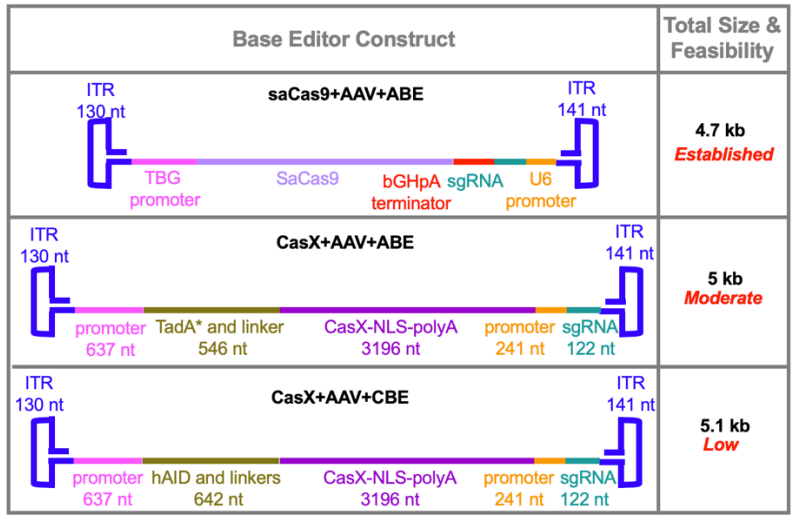
4.3 Development of a novel mini base editor

A second project was designed to address the delivery problem base editors face as we attempt to develop them as viable therapeutics. My efforts to shrink the base editor system to a degree compatible

with AAV has focused on incorporating Cas endonucleases derived from sources beyond *S. pyogenes*. In fact, *S. aureus* Cas9 (saCas9) has successfully been used for *in vivo* gene editing when delivered via AAV.(57, 58) However, the PAM requirement for saCas9 is far too restrictive for practical therapeutic use. As such, we have identified a Cas endonuclease, CasX (also known as Cas12e), which is 2,940 bp long and contains a PAM comparable to spCas9, as a potential candidate to include in base editor architecture that we believe would be compatible with an AAV delivery vehicle (**Figure 4.3A**).(59)

To determine whether CasX could function in the ABE architecture, we replaced spCas9 with catalytically dead CasX using standard USER-cloning to generate a CasX base editor plasmid. As we could not predict what the editing window would be for a CasX-ABE, multiple target sgRNA plasmids were generated to include target A bases in a variety of locations across multiple genomic loci (we chose genomic loci that have been reported in the literature, thereby establishing a direct comparison). To determine activity, we relied on HTS data to quantify base editing activity, similarly to our approach in aim 1. Despite our best efforts, no activity was seen for the CasX-ABE.

A



B

Cas Enzyme	Type	Subtype	Size	PAM (5' > 3')	Target
spCas9	II	IIA	1368 aa (4,104 bp)	NGG	dsDNA
CasX (Cas12e)	V	V-E	980 aa (2,940 bp)	TTCN	dsDNA
Cas12h1	V	V-H	870 aa (2,610 bp)	RTR	dsDNA
Cas14a (Cas12f1)	V	V-F1	529 aa (1,587 bp)	TTTR	ssDNA/dsDNA
Cas14b (Cas12f2)	V	V-F2	545 aa (1,635 bp)	TTAT	ssDNA/dsDNA

Figure 4.3. Mini BE Development. A) Base editor architecture shown for the established saCas9-ABE, suited for AAV delivery as well as proposed architectures for CasX-ABE and CasX-CBE. **B)** Categorization of recently identified Cas enzyme candidates. Size, PAM, and target columns are of special note, as these comprise the criteria we look at when selecting a candidate for study. spCas9 details provided for direct comparison.

4.4 Future outlooks

Overall, three projects were taken on to address identified weaknesses in the field of genome editing. Chapters 2 and 3 covered my work carrying out a mechanistic study of existing prime editors, Chapter 4.2 briefly details my work aimed at expanding the base editor toolkit by way of a novel RlmN base editor, and Chapter 4.3 briefly details my work aimed at shrinking traditional base editors to make them compatible with optimal viral therapeutic delivery vehicles.

The work detailed in Chapters 4.2 and 4.3 has exceled in other labs, while the work detailed in Chapters 2 and 3 has multiple paths forward within the Komor Lab, namely, expanding on intermediates tested. We anticipate future mechanistic studies to also address our

reported findings and overall shed light on the mechanisms of prime editor intermediate resolution.

Chapter 5

Non-Scientific Development

5.1 Invisible work

During my attendance at the 2022 American Chemical Society Annual Meeting held in San Diego, I was invited to participate in a panel titled *Affecting Change in Academic Departments through Strategic Diversity Leadership in the Chemical Sciences.* This panel consisted of a mix of students and faculty, presenting on their individual efforts within their departments at their home universities. One of the panelists, Dr. Safia Jilani, gave a truly memorable talk titled *Leading change through recognizing invisible work in a PhD dissertation and defense.* Dr. Jilani introduced

me to the concept of "invisible work," which is essential to the progress of science and academia, but is not formally recognized for professional or career advancement.

I spoke with Dr. Jilani at length following her talk and in the year following, and for the first time felt a path to recognize my non-scientific work throughout my tenure in graduate school. I was encouraged by Dr. Jilani to formally detail and recognize this work in my PhD dissertation, and am happy to be doing so at this point in time.

5.2 Chemistry Graduate Student Council (CGSC)

In today's ever-changing climate, systematic efforts enacting change for equity, diversity, and inclusion have never been more important. While changes within the higher institutions of academia have been highlighted at the university level, many grassroots-based efforts are evolving at the department levels. One such demographic that can greatly impact departments' efforts are graduate student organizations.

The Chemistry Graduate Student Council (CGSC) of the Department of Chemistry and Biochemistry at UC San Diego is one such organization committed to representing student interests at the department level. At its core, CGSC functions to bridge the communication gap between students and department administration. We work closely with sister student organizations such as the Society for Women in Graduate Studies (SWIGS), and the Graduate and Professional Student Association (GPSA) to ensure a thorough approach is taken to represent a diverse array of student interests.

Within our own representative body, CGSC requires participation from each cohort of graduate students in our department; from first years, to students in their "6^{th} +" year. We believe it is of great benefit to connect with students at every stage of their graduate student career as

unique challenges await them throughout their journey. Amplifying the voices and concerns of each group to our department administration will help ensure graduate student success.

Ensuring graduate student success is a key pillar for any academic department, however this can only be fully achieved with input and work from graduate students. Representation matters, student voices matter, and advocating on behalf of one's self matters. Similar to traditional political systems, we elect a small body of students to advocate on behalf of the larger student body. In doing so, we help expand the meaning of success in our department to include student belonging and well-being.

Throughout my 4 year tenure as a member of CGSC, 2 of which were spent as CGSC Chair, I interacted with a broader range of scientists than I did throughout my scientific studies. I was able to learn more about the inner workings of my department and build relationships with faculty and administration I otherwise never would have. It was a truly self-rewarding endeavor that produced tangible results in times when my scientific work did not. Overall, my time on CGSC was aimed at enriching the climate of our department and experience for other graduate students during their studies, and accomplishing these goals with equally committed individuals.

5.3 Future Outlook

As I conclude my PhD dissertation, and specifically this chapter on 'invisible work,' I would like to formally encourage all PhD candidates to include a similar chapter in their theses detailing any work they undertook during their tenure away from the bench that had an impact on their individual development. I am of the mindset that my work with CGSC was just as rewarding and beneficial as my scientific work, and deserves recognition. I hope my contributions to the Chemistry and Biochemistry department at UCSD as well as CGSC and other student

organizations will be long-lasting and built upon and improved by my colleagues, and I hope to see ‘invisible work’ chapters in their dissertations and at their defenses.

Table 2.1 Sequences of high throughput primers used in this study.

Loci	Direction	Sequence (5' > 3')
EMX1	Fwd.	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**NNNN**AGTTTCTCATCTGTGCCCCTCC
	Rev.	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTGCCCACCCTAGTCATTGGAG
HEK3	Fwd.	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**NNNN**AGCTTGGCATGAGAAACCTTGG
	Rev.	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTAGGAAAAGCTGTCCTGCG
HEK4	Fwd.	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**NNNN**CCTCCCTTCAAGATGGCTGACA
	Rev.	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTTTCAACCCGAACGGAGACAC
RNF2	Fwd.	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**NNNN**GCAGACAAACGGAACTCAACCA
	Rev.	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCACTGTTCACCCAGTACCT
NNNN *bases are necessary index tags required for Illumina MiniSeq HTS*		

Table 2.2 Sequences of pegRNAs used in this study.

Plasmid Type	Loci	Spacer sequence (5' > 3')	3' ext sequence (5' > 3')	PBS length (bp)	RT length (bp)
pegRNA	HEK3	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
pegRNA	EMX1	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGA	14	13
pegRNA	RNF2	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATG	15	11
pegRNA	HEK4	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGC	8	13
pegRNA	RNF2	GTCATCTTAGTCATTACCTG	AACGAACACAGGTAATGACTAAGATG	15	11
pegRNA	HEK3	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCACGTGCTCAGTCTG	13	31
pegRNA	EMX1	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTTCGCATTCTTCTGCTCGGA	14	19
nicking gRNA	RNF2	GTCAACCATTAAGCAAAACAT			
nicking gRNA	EMX1	GCCGTTTGTACTTTGTCCTC			
nicking gRNA	HEK3	GTCAACCAGTATCCCGGTGC			
nicking gRNA	HEK4	GAGACACACACACAGGCCTGG			

References

1. Sun,H. and Yu,G. (2019) New insights into the pathogenicity of non-synonymous variants through multi-level analysis. *Sci. Rep.*, **9**, 1–11.

2. Auton,A., Abecasis,G.R. and Consortium,T. 1000 G.P. (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.

3. Katsonis,P., Koire,A., Wilson,S.J., Hsu,T.-K., Lua,R.C., Wilkins,A.D. and Lichtarge,O. (2014) Single nucleotide variations: Biological impact and theoretical interpretation. *Protein Sci.*, **23**, 1650–1666.

4. Zhang,F. and Lupski,J.R. (2015) Non-coding genetic variants in human disease. *Hum. Mol. Genet.*, **24**, R102–R110.

5. Kircher,M., Witten,D.M., Jain,P., O'roak,B.J., Cooper,G.M. and Shendure,J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.

6. Bertucci,F., Ng,C.K.Y., Patsouris,A., Droin,N., Piscuoglio,S., Carbuccia,N., Soria,J.C., Dien,A.T., Adnani,Y., Kamal,M., Garnier,S., Meurice,G., Jimenez,M., Dogan,S., Verret,B., Chaffanet,M., Bachelot,T., Campone,M., Lefeuvre,C., Bonnefoi,H., Dalenc,F., Jacquet,A., De Filippo,M.R., Babbar,N., Birnbaum,D., Filleron,T., Le Tourneau,C. and André,F. (2019) Genomic characterization of metastatic breast cancers. *Nature*, **569**, 560–564.

7. Smith,C., Abalde-Atristain,L., He,C., Brodsky,B.R., Braunstein,E.M., Chaudhari,P., Jang,Y.Y., Cheng,L. and Ye,Z. (2015) Efficient and allele-specific genome editing of disease loci in human iPSCs. *Mol. Ther.*, **23**, 570–577.

8. Komor,A.C., Badran,A.H. and Liu,D.R. (2017) CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*, **168**, 20–36.

9. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).*, **337**, 816–821.

10. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 2579–2586.

11. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science (80-.).*, **339**, 823–826.

12. Martin,R.M., Ikeda,K., Cromer,M.K., Uchida,N., Nishimura,T., Romano,R., Tong,A.J.,

Lemgart,V.T., Camarena,J., Pavel-Dinu,M., Sindhu,C., Wiebking,V., Vaidyanathan,S., Dever,D.P., Bak,R.O., Laustsen,A., Lesch,B.J., Jakobsen,M.R., Sebastiano,V., Nakauchi,H. and Porteus,M.H. (2019) Highly Efficient and Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination. *Cell Stem Cell*, **24**, 821–828.

13. Paquet,D., Kwart,D., Chen,A., Sproul,A., Jacob,S., Teo,S., Olsen,K.M., Gregg,A., Noggle,S. and Tessier-Lavigne,M. (2016) Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*, **533**, 125–129.

14. Richardson,C.D., Ray,G.J., DeWitt,M.A., Curie,G.L. and Corn,J.E. (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.*, **34**, 339–344.

15. Yu,C., Liu,Y., Ma,T., Liu,K., Xu,S., Zhang,Y., Liu,H., La Russa,M., Xie,M., Ding,S. and Qi,L.S. (2015) Small molecules enhance crispr genome editing in pluripotent stem cells. *Cell Stem Cell*, **16**, 142–147.

16. Lin,S., Staahl,B.T., Alla,R.K. and Doudna,J.A. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*, **3**, e04766.

17. Song,J., Yang,D., Xu,J., Zhu,T., Chen,Y.E. and Zhang,J. (2016) RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.*, **7**, 10548.

18. Pinder,J., Salsman,J. and Dellaire,G. (2015) Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.*, **43**, 9379–9392.

19. Yeh,C.D., Richardson,C.D. and Corn,J.E. (2019) Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.*, **21**, 1468–1478.

20. Burmistrz,M., Krakowski,K. and Krawczyk-Balska,A. (2020) RNA-targeting CRISPR–cas systems and their applications. *Int. J. Mol. Sci.*, **21**, 1122.

21. Abudayyeh,O.O., Gootenberg,J.S., Essletzbichler,P., Han,S., Joung,J., Belanto,J.J., Verdine,V., Cox,D.B.T., Kellner,M.J., Regev,A., Lander,E.S., Voytas,D.F., Ting,A.Y. and Zhang,F. (2017) RNA targeting with CRISPR-Cas13. *Nature*, **550**, 280–284.

22. Gaj,T., Gersbach,C.A. and Barbas III,C.F. (2013) ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, **31**, 397–405.

23. Gupta,R.M. and Musunuru,K. (2014) Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Invest.*, **124**, 4154–4161.

24. Joung,J.K. and Sander,J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.*, **14**, 49–55.

25. Komor,A.C., Kim,Y.B., Packer,M.S., Zuris,J.A. and Liu,D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.

26. Nishimasu,H., Ran,F.A., Hsu,P.D., Konermann,S., Shehata,S.I., Dohmae,N., Ishitani,R., Zhang,F. and Nureki,O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935–949.

27. Kunz,C., Saito,Y. and Schär,P. (2009) Mismatched repair: Variations on a theme. *Cell. Mol. Life Sci.*, **66**, 1021–1038.

28. Fukui,K. (2010) DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids*, **2010**, 260512.

29. Yang,L., Briggs,A.W., Chew,W.L., Mali,P., Guell,M., Aach,J., Goodman,D.B., Cox,D., Kan,Y., Lesha,E., Soundararajan,V., Zhang,F. and Church,G. (2016) Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.*, **7**, 1–12.

30. Mok,B.Y., de Moraes,M.H., Zeng,J., Bosch,D.E., Kotrys,A. V, Raguram,A., Hsu,F., Radey,M.C., Peterson,S.B., Mootha,V.K., Mougous,J.D. and Liu,D.R. (2020) A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature*, **583**, 631–637.

31. Landrum,M.J., Lee,J.M., Benson,M., Brown,G., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Hoover,J., Jang,W., Katz,K., Ovetsky,M., Riley,G., Sethi,A., Tully,R., Villamarin-Salomon,R., Rubinstein,W. and Maglott,D.R. (2016) ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, **44**, D862–D868.

32. Gaudelli,N.M., Komor,A.C., Rees,H.A., Packer,M.S., Badran,A.H., Bryson,D.I. and Liu,D.R. (2017) Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.

33. Shi,K., Carpenter,M.A., Banerjee,S., Shaban,N.M., Kurahashi,K., Salamango,D.J., McCann,J.L., Starrett,G.J., Duffy,J. V., Demir,Ö., Amaro,R.E., Harki,D.A., Harris,R.S. and Aihara,H. (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.*, **24**, 131–139.

34. Grünewald,J., Zhou,R., Iyer,S., Lareau,C.A., Garcia,S.P., Aryee,M.J. and Joung,J.K. (2019) CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.*, **37**, 1041–1048.

35. Rallapalli,K.L., Komor,A.C. and Paesani,F. (2020) Computer simulations explain mutation-induced effects on the DNA editing by adenine base editors. *Sci. Adv.*, **6**, eaaz2309.

36. Anzalone,A. V., Randolph,P.B., Davis,J.R., Sousa,A.A., Koblan,L.W., Levy,J.M., Chen,P.J., Wilson,C., Newby,G.A., Raguram,A. and Liu,D.R. (2019) Search-and-replace genome

editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.

37. Chen,P.J., Hussmann,J.A., Yan,J., Knipping,F., Ravisankar,P., Chen,P.F., Chen,C., Nelson,J.W., Newby,G.A., Sahin,M., Osborn,M.J., Weissman,J.S., Adamson,B. and Liu,D.R. (2021) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*, **184**, 5635–5652.

38. Ferreira da Silva,J., Oliveira,G.P., Arasa-Verge,E.A., Kagiou,C., Moretton,A., Timelthaler,G., Jiricny,J. and Loizou,J.I. (2022) Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nat. Commun.*, **13**.

39. Branzei,D. and Foiani,M. (2008) Regulation of DNA repair throughout the cell cycle. *Nat. Rev. Mol. Cell Biol.*, **9**, 297–308.

40. Burnett,C.A., Wong,A.T., Vasquez,C.A., McHugh,C.A., Yeo,G.W. and Komor,A.C. (2022) Examination of the Cell Cycle Dependence of Cytosine and Adenine Base Editors. *Front. Genome Ed.*, **4**, 1–15.

41. Clement,K., Rees,H., Canver,M.C., Gehrke,J.M., Farouni,R., Hsu,J.Y., Cole,M.A., Liu,D.R., Joung,J.K., Bauer,D.E. and Pinello,L. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol. 2019 373*, **37**, 224–226.

42. Jordan,M.A., Thrower,D. and Wilson,L. (1992) Effects of vinblastine, podophyllotoxin and nocodazole on mitotic spindles: Implications for the role of microtubule dynamics in mitosis. *J. Cell Sci.*, **102**, 401–416.

43. Panigrahi,G.B., Slean,M.M., Simard,J.P. and Pearson,C.E. (2012) Human Mismatch Repair Protein hMutL Is Required to Repair Short Slipped-DNAs of Trinucleotide Repeats *. *J. Biol. Chem.*, **287**, 41844–41850.

44. Bodai,Z., Bishop,A.L., Gantz,V.M. and Komor,A.C. (2022) Targeting double-strand break indel byproducts with secondary guide RNAs improves Cas9 HDR-mediated genome editing efficiencies. *Nat. Commun.*, **13**.

45. Shrivastav,M., De Haro,L.P. and Nickoloff,J.A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res.*, **18**, 134–147.

46. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).*, **337**, 816–822.

47. Doudna,J.A. and Charpentier,E. (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).*, **346**.

48. Kosicki,M., Tomberg,K. and Bradley,A. (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **36**,

765–771.

49. Ihry,R.J., Worringer,K.A., Salick,M.R., Frias,E., Ho,D., Theriault,K., Kommineni,S., Chen,J., Sondey,M., Ye,C., Randhawa,R., Kulkarni,T., Yang,Z., McAllister,G., Russ,C., Reece-Hoyes,J., Forrester,W., Hoffman,G.R., Dolmetsch,R. and Kaykas,A. (2018) p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.*, 10.1038/s41591-018-0050-6.

50. Gaudelli, Nicole M.; Komor, Alexis C.; Rees, Holly A.; Packer, Michael S.; Badran, Ahmed H.; Bryson, David I.; Liu,D.R., Gaudelli,N.M., Komor,A.C., Rees,H.A., Packer,M.S., Badran,A.H., Bryson,D.I. and Liu,D.R. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.

51. Wu,Z., Yang,H. and Colosi,P. (2010) Effect of genome size on AAV vector packaging. *Mol. Ther.*, **18**, 80–86.

52. Naso,M.F., Tomkowicz,B., Perry,W.L. and Strohl,W.R. (2017) Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs*, **31**, 317–334.

53. Benítez-Páez,A., Villarroya,M. and Armengod,M.E. (2012) The Escherichia coli RlmN methyltransferase is a dual-specificity enzyme that modifies both rRNA and tRNA and controls translational accuracy. *RNA*, **18**, 1783–1795.

54. Ikehara, Morio; Hattori, Masao; Fukui,T. (1972) Synthesis and Properties of Poly 2-Methyladenylic Acid). *Eur. J. Biochem.*, **334**, 329–334.

55. Gaudelli,N.M., Lam,D.K., Rees,H.A., Solá-Esteves,N.M., Barrera,L.A., Born,D.A., Edwards,A., Gehrke,J.M., Lee,S.-J., Liquori,A.J., Murray,R., Packer,M.S., Rinaldi,C., Slaymaker,I.M., Yen,J., Young,L.E. and Ciaramella,G. (2020) Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.*, **38**, 892–900.

56. Richter,M.F., Zhao,K.T., Eton,E., Lapinaite,A., Newby,G.A., Thuronyi,B.W., Wilson,C., Koblan,L.W., Zeng,J., Bauer,D.E., Doudna,J.A. and Liu,D.R. (2020) Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol. 2020*, **38**, 883–891.

57. Nelson, C. E.; Hakim, C. H.; Ousterout, D. G.; Thakore, P. I.; Moreb, E. A.; Castellanos Rivera, R. M.; Madhavan, S.; Pan, X.; Ran, F. A.; Xan, W. X.; Asokan, A.; Zhang, F.; Duan, D.; Gersbach,C.A. (2016) In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science (80-.).*, **351**, 403–408.

58. Ran,F.A., Cong,L., Yan,W.X., Scott,D.A., Gootenberg,J.S., Kriz,A.J., Zetsche,B., Shalem,O., Wu,X., Makarova,K.S., Koonin,E. V., Sharp,P.A. and Zhang,F. (2015) In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, **520**, 186–191.

59. Burstein,D., Harrington,L.B., Strutt,S.C., Probst,A.J., Anantharaman,K., Thomas,B.C., Doudna,J.A. and Banfield,J.F. (2017) New CRISPR-Cas systems from uncultivated microbes. *Nature*, **542**, 237–241.