UNIVERSITY OF CALIFORNIA SAN DIEGO

# Investigations into the Cellular Mechanism of Base Editing

A thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Chemistry

by

Cameron Andrew Burnett

Committee in charge:

Professor Alexis Komor, Chair
Professor Colleen McHugh, Co-Chair
Professor Jeremy Klosterman

2019

The Thesis of Cameron Andrew Burnett is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____
Co-chair

_____
Chair

University of California San Diego

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor and chair of my committee Dr. Alexis Komor, for her vital advice and mentorship, through many rough drafts and constructive advice.

I would like to acknowledge my lab for their support and comradery ever since I have joined the lab.

I would like to acknowledge Dennis Young for helping me with the FACS sorting and cell cycle analysis.

ABSTRACT OF THE THESIS

Investigations into the Cellular Mechanism of Base Editing

by

Cameron Andrew Burnett

Master of Science in Chemistry

University of California San Diego, 2019

Professor Alexis Komor, Chair

Professor Colleen McHugh, Co-Chair

Base editing is a recent development in the genome editing field that allows for the introduction of single nucleotide changes in the genome with high efficiency. This is accomplished without the induction of double-strand breaks (DSBs), a hallmark of traditional Cas9 editing that suffers from modest editing efficiencies, and high frequencies of random insertions and deletions (indels) instead of the desired genome editing product. An additional major limitation of DSB-reliant genome editing methods is that their activity fluctuates in

different phases of the cell cycle. It is currently not known if base editors suffer from these same drawbacks. Additionally, base editing technology in its current form still suffers from shortcomings, such as undesired genome editing byproducts, that demand a more thorough understanding of its cellular mechanisms. To initiate our investigations of the cellular mechanism of base editing, we chemically synchronized cells into the various stages of the cell cycle and measured the effects this had on base editing efficiencies and levels of byproduct formation. We observed a decrease in base editing byproduct formation when cells were synchronized in G1, the cell cycle phase when cells are growing but not replicating their DNA. To further characterize base editing mechanisms, we have developed a genome-wide screen to identify the key proteins involved in repair of uracil-guanine mismatches, the intermediate of cytosine base editing.

# Chapter 1

## Introduction

The implementation of the CRISPR-Cas9 system has been widely used to introduce genetic alterations in a programmable manner. The CRISPR-associating (Cas) enzyme binds to short sequences of RNA (tracrRNA in the native bacterial setting and guide RNA for the repurposed biotechnological application), which guides the complex to a target sequence of interest. This protospacer sequence must have homology to the sgRNA and contain a protospacer-adjacent motif (PAM). The endonuclease cleaves both strands of DNA to yield a double-stranded break (DSB) and initiate the first step of genome editing. Cells respond to this event through either non-homologous end joining (NHEJ) which will result in random insertions or deletions (indels) of bases at the site of the DSB, or homology-directed repair (HDR), which will use an exogenous piece of DNA as a template for repair, resulting in precise DNA modification (Figure 1). A major limitation of DSB-reliant genome editing technologies is the competition between NHEJ and HDR, which results in mixtures of genome editing products when using the technology.

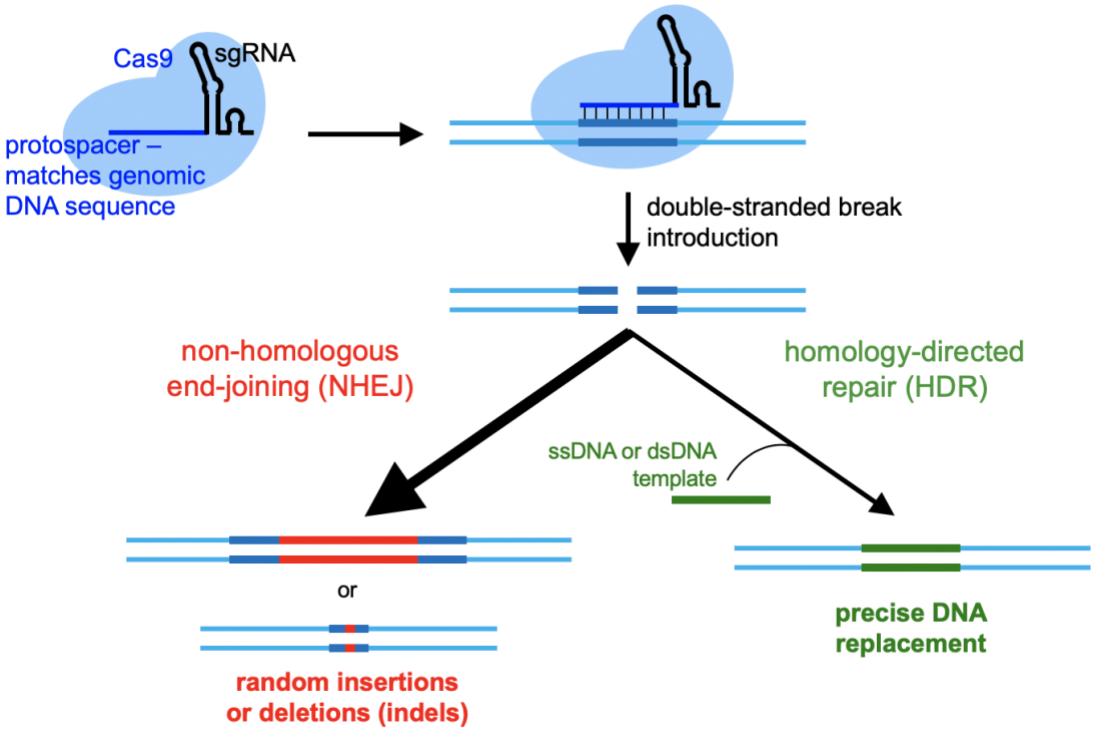



Figure 1: Traditional Cas9 Gene Editing Outcomes.
Cas9-gRNA complex targets the sequence of interest to induce a DSB, which is then repaired either by NHEJ or HDR. NHEJ is the dominant outcome, which frequently results in indel formation.

Additionally, these DSB repair pathways are cell cycle dependent, meaning not all cell types are amenable to precision (HDR-mediated) genome editing. Specifically, NHEJ is most active during the G1, S and G2 phases of the cell cycle. G1 is the growth or gap phase immediately after cell division and provides proteins and mRNA's necessary for DNA synthesis, which occurs in S phase (synthesis). G2 is a second gap phase where cells ensure genomic integrity for division in M phase (mitosis). HDR is restricted to late S and G2 phases as a mechanism to correct potential errors after DNA replication of sister chromatids as repair templates.

Previous groups have used chemical inhibitors to arrest HEK293T and fibroblast cells at G1, S, and M phases of the cell cycle prior to electroporation of Cas9:sgRNA complex (or ribonucleoprotein, RNP)[2]. There was a significant increase in HDR-mediated editing for cells

arrested by nocodazole in G2-M, which was unexpected as HDR machinery is most active in late S phase rather than at the G2/M border. Subsequent treatment with a G1/S arrested agent to prevent S-phase entry reduced the increase in HDR, supporting a mechanism where synchronous entry into S-phase amplifies editing, rather than the G2/M block itself[2]. Many therapeutic targets of genome editing are post-mitotic, and therefore do not undergo cell division, rendering HDR too inefficient for precision genome editing.

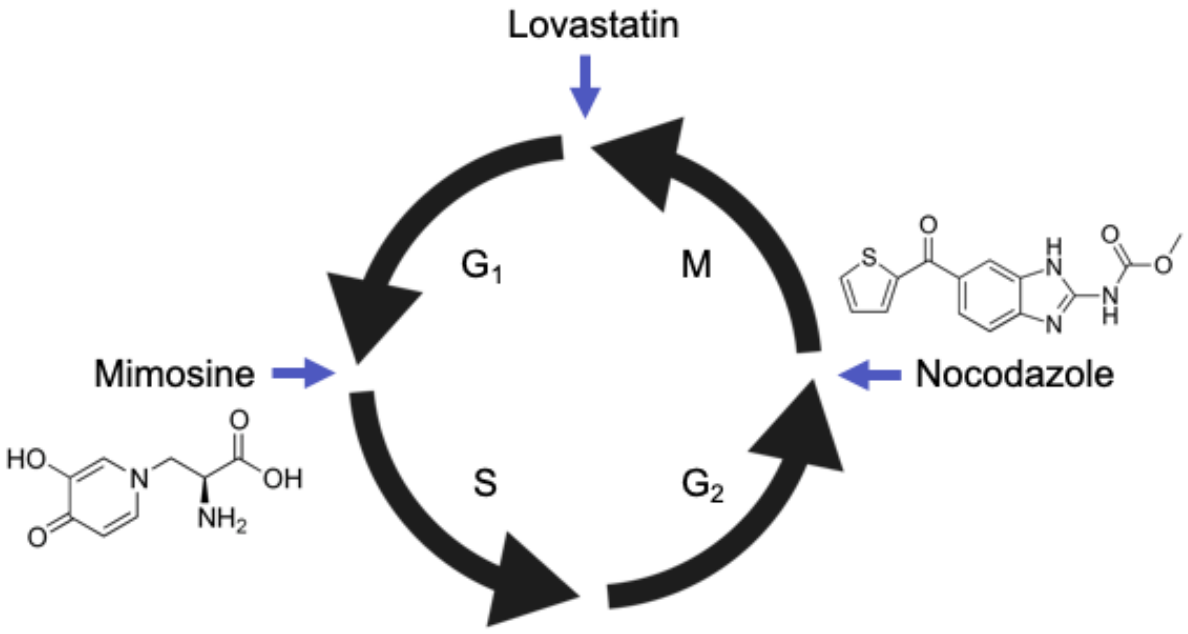



Figure 2: Chemical inhibitors and their target cell cycle arrest
Overview of the chemical inhibitory agents used in this thesis, their chemical structure and the cell cycle phase synchronization point.

A class of genome editing agents called base editors was developed that chemically modify target nucleobases on dCas9-exposed-single-stranded DNA (ssDNA) rather than cleaving both strands. Base editors consist of a catalytically impaired Cas9 (dCas9) tethered to either a ssDNA-specific cytidine deaminase or a variant of the tRNA adenine deaminase TadA evolved to act on ssDNA. Upon dCas9 binding to its target protospacer, a small window of

ssDNA is made accessible to the enzyme, which chemically modifies any cytidines or adenosines (depending on the enzyme) that are present within this window to initiate point mutation introduction. These two classes of editors catalyze C:G to T:A and T:A to G:C transitions via uracil and inosine intermediates, respectively (Figure 3). Initial cytosine base editors experienced low conversion efficiencies, likely from excision of the uracil intermediate by native uracil glycosylases. To counteract this, later base editor constructs incorporated a uracil glycosylase inhibitor (UGI) in their architecture, which increased the efficiency of C:G to T:A conversions[4]. Despite these engineering efforts, at certain loci high levels of unpredictable product distributions are observed, where C:G to non T:A point mutation introduction levels are significant. While it has been postulated that cellular replication and the mismatch repair pathway are responsible for converting the uracil and inosine intermediates into their respective desired base editing outcomes, no experiments have been performed to test this hypothesis. To initiate such a study, we sought to characterize the cell-cycle dependence of base editing. The results from this study will not only shed light on the cellular mechanism of base editing, but also will inform on the cell types most amenable to efficient base editing.

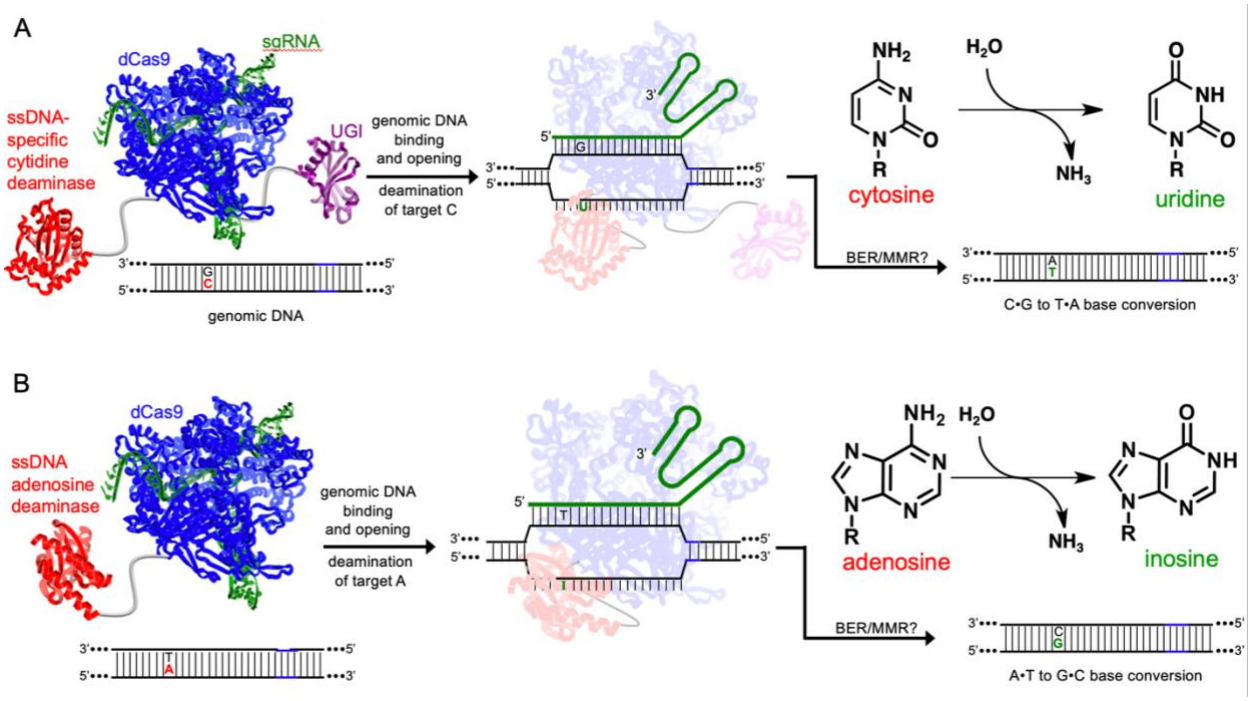



Figure 3: Base Editing Overview
Schematic of the mechanism behind cytosine and adenosine base editors. ssDNA modifying enzymes target exposed nucleotides to alter individual base pairs.

## Materials and Methods

### Preparation of Synchronizing agents

#### Mimosine

10 μM stocks solutions of mimosine (Sigma) were prepared in 1X PBS. Stock solution was added to DMEM media for a final concentration of 400 μM.

#### Lovastatin

52 mg of lovastatin (Sigma) was dissolved in 1.04 mL 95% ethanol. 813 μL 1 N NaOH was added and the PH was adjusted to 7.5. Working concentration was 40 μM.

#### Nocodazole

Nocodazole (Sigma) was prepared in DMSO to a concentration of 20 mg/mL. Working concentrations in DMEM media were 200 ng/mL.

### Synchronization Treatment

$3x10^5$ HEK293T cells were plated in a T75 flask in the presence of synchronizing agents as indicated above. After 12 and 17 hours, cells were fixed for PI staining.

### Cell cycle analysis

Cells in a T75 flask were washed with 10 mL PBS, detached with TrypLE, and collected by centrifugation for 10 minutes at 400 rcf. Cells were resuspended at 1x10^6 cells/mL in 1 mL cold PBS, then added to 9 mL cold 70% ethanol for ≥ 4 hours at -20º C.

After ethanol fixation, cells were centrifuged at 400 rcf, washed with cold PBS, then stained with PI solution (0.1% Triton X-100 (Sigma) 0.2 mg/mL RNAse (Sigma) 0.02 mg/mL PI in PBS). Cells were incubated at 37 º for 15 minutes, then put on ice for transport.

## Cell Culture

HEK293T cells (ATCC CRL-3216) were maintained in high glucose DMEM media supplemented with GlutaMAX (ThermoFisher Scientific), 10% (v/v) fetal bovine serum (ThermoFisher Scientific), and 100 U/mL Penicillin-Streptomycin (ThermoFisher Scientific), at 37º C with 5% $CO_2$. K562 cells (ATCC CRL-3344) were maintained in RPMI media (Life Sciences) supplemented as described above.

## Transfections

HEK293T cells were seeded in 48-well VWR Multiwell Cell Culture Plates at a density of 150,000 cells per well in 250 µL of media without Penicillin-Streptomycin. 1000 ng of BE plasmid and 250 ng of sgRNA plasmid were transfected into the cell suspension, using 1.5 µL of Lipofectamine 2000 (ThermoFisher Scientific) per well according to the manufacturer's protocol. Chemical inhibitors (Mimosine 400 µM, Nocodazole 20 ng/mL) were added to 48-well plate transfection 6 hours after addition of lipofectamine-plasmid complexes.

## Cloning

BE plasmids with a 2A-GFP promoter were obtained from the Broide lab. P2A-BE4ΔUGI was constructed using USER cloning[4] with P2A-BE4 as a template, using Phusion U Hot Start Polymerase (ThermoFisher Scientific). All sgRNA expression plasmids were generated using blunt-end cloning[1] with pFYF1230 (Addgene plasmid #47511) as a template, using Phusion High-Fidelity DNA Polymerase (New England BioLabs). Complete protospacer/PAM sequences are listed in Table 1. All DNA vector amplification was carried out using NEB 10 competent cells (New England BioLabs). All plasmids were purified using the ZymoPURE II Plasmid Midiprep Kit (Zymo Research).

High-throughput DNA sequencing (HTS) of genomic DNA

Transfected cell were rinsed with 150 µL per well of phosphate-buffered saline (ThermoFisher Scientific) 5 days after transfection. Cells were lysed on the plate by addition of 100 µL of lysis buffer (10 mM Tris, pH 7.5, 0.1% SDS, and 25 µg/mL Proteinase K). Lysed cells were then heated at 37º C for 1 hour, followed by 80º C for 20 minutes. Genomic loci of interest were PCR amplified with Phusion High-Fidelity DNA Polymerase (New England BioLabs) according to the manufacturer's protocol, with primers indicated in Table 2. 2 µL of genomic DNA mixture as a template, and 26 or fewer rounds of amplification. Unique forward and reverse combinations of Illumina adapter sequences were then appended with an additional round of PCR amplification with Phusion High-Fidelity DNA Polymerase (New England BioLabs) according to the manufacturer's protocol, using 1 µL of round 1 PCR mixture as a template and 15 rounds of amplification. The products were gel purified and quantified using NEBNext Ultra II DNA Library Prep Kit for Illumina. Samples were then sequenced on an Illumina MiniSeq according to the manufacturer's protocol.

HTS Data Analysis

Sequencing reads were demultiplexed in MiniSeq Reporter (Illumina), and individual FASTQ files were analyzed using a previously reported Matlab script[4].

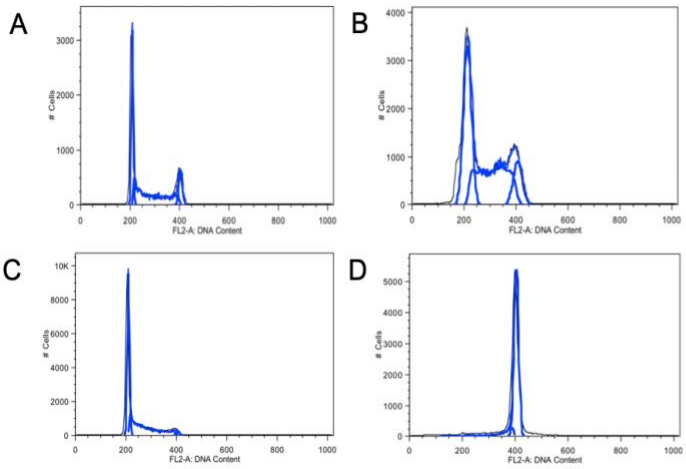
| Table 1: Protospacer sequences and PAM motifs | | |
|---|---|---|
| | Protospacer | PAM |
| HEK2 | GAAC$_4$AC$_6$AAAGCATAGACTGC | GGG |
| HEK3 | GGCC$_4$C$_5$AGACTGAGCACGTGA | TGG |
| RNF2 | GTC$_3$ATC$_6$TTAGTCATTACCTG | AGG |
| HIRA | GAAGA$_5$CCAAGGATAGACTGC | TGG |
| PSMB2 | GTAAA$_5$CA$_7$AAGCATAGACTGA | GGG |

| Table 2: First round genomic DNA PCR sequences | |
|---|---|
| Primer Name | Primer Sequence |
| HEK2-Fwd | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATTGTCCAGCCCCATCTGTCAA |
| HEK2 -Rev | TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCAAGTTACTGCAGCCCAAGC |
| HEK3-Fwd | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGACAGGGATCCCAGGGAAAC |
| HEK3-Rev | TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACCAG |
| RNF2-Fwd | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCAGACAAACGGAACTCAACCA |
| RNF2-Rev | TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCACCACTGTTCACCCCAGTACCT |
| HIRA-Fwd | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCATCATAGCGAGACCCTGTCT |
| HIRA-Rev | TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTGGCCCAATGACACCACATG |
| PSMB2 (pos 7)-Fwd | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACTGTGACTGGCCCCCAATATC |
| PSMB2 (pos 7)-Rev | TGGAGTTCAGACGTGTGCTCTTCCGATCTTACCCCTGTTCCTAAAGCCCAC |

## Results

### Effects of chemical inhibitors on HEK293T cells

Previously tested cell cycle inhibitors include aphidicolin, mimosine, thymidine, lovastatin, and nocodazole. Aphidicolin, mimosine, and thymidine all arrest at the G1/S phase border, lovastatin at the M/G1 border, and nocodazole at G2/M[6] (Figure 2). Mimosine was chosen as the G1/S arresting agent since it requires only one treatment period as opposed to aphidicolin and thymidine which require two 17-hour arrest blocks. Mimosine is a non-protein amino acid isolated from plants that arrests cells such as HeLa in late G1 prior to the onset of S phase[9]. It is believed to inhibit DNA biosynthesis, though its exact target and mechanism are still unclear. Lovastatin enriches early G1 phase cells by inhibition of HMG-CoA reductase, resulting in mevalonate depletion, which is necessary for cholesterol synthesis. Lovastatin has been used to synchronize breast cancer cells and Jurkat cells[6]. Nocodazole inhibits microtubule polymerization to synchronize cells in M phase. Nocodazole robustly arrests cells at G2/M border after 17 hours of treatment, as >70% demonstrate DNA content consistent with duplicated sister chromatids, an indicator of cells in G2, as compared to 14% in unsynchronized cells (Figure 1). Mimosine treatment enhances the fraction of cells in G1 (59% compared to 35% for unsynchronized cells) and severely inhibits G2 phase entry after 17 hours of treatment. Lovastatin had no significant effect on HEK293T cell cycle as compared to wild type asynchronous cells, consistent with previous literature[2].



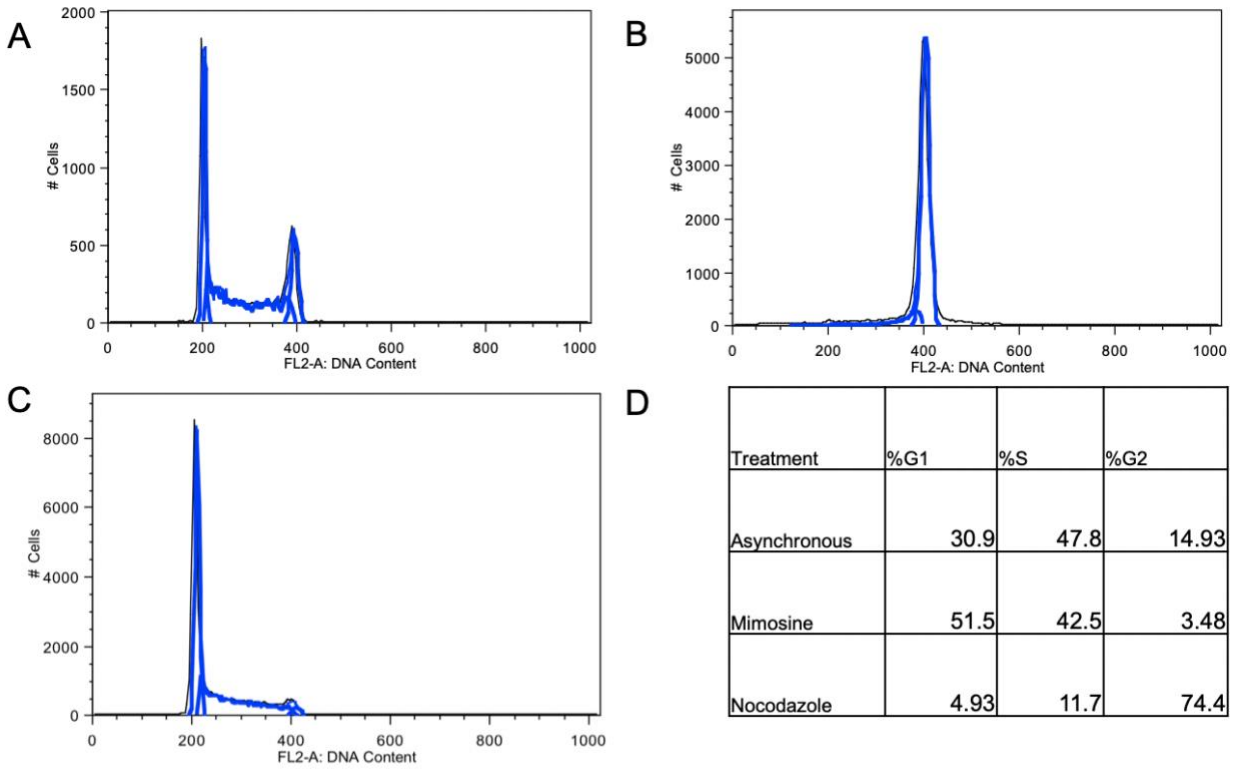

E

| Treatment | %G1 | %S | %G2 |
|---|---|---|---|
| Asynchronous | 37.1 | 45.3 | 14.9 |
| Asynchronous | 33.8 | 47.2 | 13.3 |
| | | | |
| Lovastatin | 40.8 | 42.5 | 13.4 |
| Lovastatin | 39.9 | 43.3 | 19.5 |
| | | | |
| Mimosine | 57.9 | 38.4 | 2.02 |
| Mimosine | 60.3 | 38.2 | 0 |
| Mimosine | 59.5 | 35 | 2.65 |
| | | | |
| Nocodazole | 0 | 13.1 | 74.5 |
| Nocodazole | 3.85 | 10.3 | 74.2 |
| Nocodazole | 7.38 | 12.7 | 69.2 |

Figure 4: DNA content following 17 hour treatment with chemical inhibitors
(A) Histogram representation of DNA content in asynchronous HEK293T cells based on propidium iodide (PI) staining. (B) Lovastatin (40 $\mu$M) treated HE293T cells. (C) Mimosine (400 $\mu$M) treated cells. (D) Nocodazole (200 ng/mL) treated cells. (E) Cell cycle distribution % by phase based on ModFit LT software analysis.

Optimizing timing of synchronization

Expression of base editors from transfected plasmids peaks at 24 hours, beginning at 12 hours and lasting up to 72 hours[7]. Pretreatment of cells with mimosine and nocodazole prior to transfection drastically reduces transfection efficiency with cationic lipid reagents. To control against differences in plasmid expression due to low levels of transfection, inhibitors were added 6 hours after transfection. GFP levels were comparable for each condition, indicating equivalent expression of base editor. Thus, timing peak base editor expression with the onset of cell cycle arrest best aligns the window of editing with synchronization. Six hours after transfection, mimosine and nocodazole were added to cells. PI staining at 12 hours post-synchronization shows robust nocodazole arrest (Figure 3), while mimosine treatment demonstrated a modest arrest in G1 (52% versus 31% for unsynchronized cells).




| Treatment | %G1 | %S | %G2 |
|---|---|---|---|
| Asynchronous | 30.9 | 47.8 | 14.93 |
| Mimosine | 51.5 | 42.5 | 3.48 |
| Nocodazole | 4.93 | 11.7 | 74.4 |

Figure 5: Treatment with Synchronizing Agents for 12 Hours
HEK293T cells were treated with synchronizing chemicals (A) Untreated control (B) 400 $\mu$M Mimosine and (C)200 ng/mL Nocodazole) for 12 hours and stained with PI.
(D) Cell cycle phase distribution as measured with ModFit LT software analysis.

## Effect of Synchronization on Editing

### BE4 and BE4ΔUGI

The fourth generation of cytosine base editor (BE4) and a variant lacking UGI (BE4ΔUGI) were used with previously established sgRNA's[1,3] targeting the genomic loci HEK293 site 2, HEK293 site 3, and RNF2. Six hours after transfection of BE and sgRNA plasmids, cells were synchronized with nocodazole or mimosine for 48 hours. 72 hours post-transfection, cells were lysed, the genomic DNA (gDNA) extracted, and the HEK2, HEK3, or RNF2 loci amplified by PCR. High-throughput sequencing (HTS) was then used to quantify base editing efficiencies at the various genomic loci.

### HEK2 site

The HEK2 site is on chromosome 5 and has two target Cs within the protospacer, one at position 4 and the other at position 6. The $C_6$ target is preferentially modified over the $C_4$ target, and high levels of C to non-T editing are generally observed at $C_6$. Analysis of HTS reads for overall c to T, as well as C to A, G, or T efficiencies, indicates that following cell cycle arrest, the total editing of targets $C_4$ and $C_6$ within the protospacer is reduced by up to 20% upon G1/S arrest with mimosine (Figure 6). Blocking in G2/M with nocodazole does not significantly impact the editing of either target C.

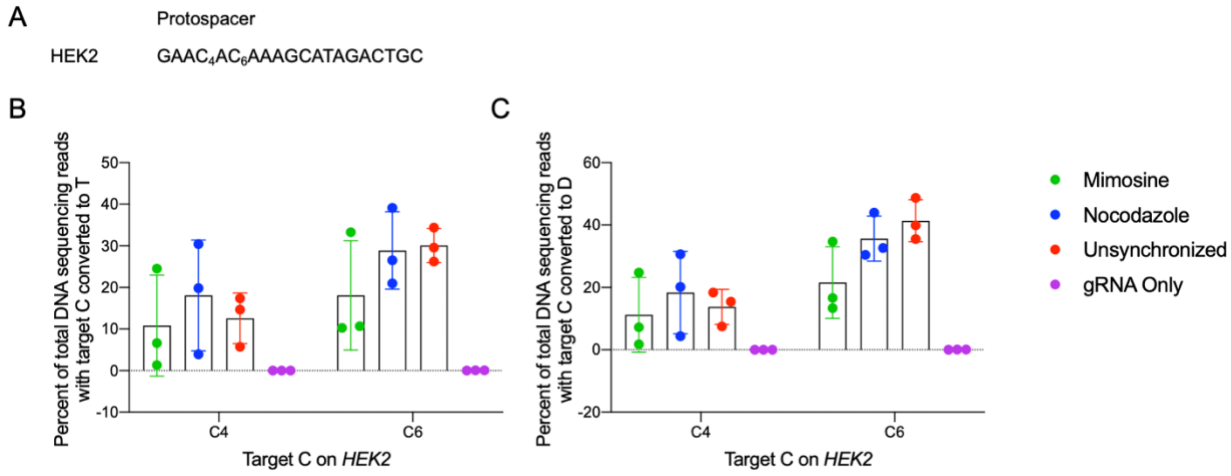



Figure 6: Effect of Synchronization on BE4 Editing at HEK293 site 2.
(A) Protospacer of HEK2 containing two editable C's within the target window. (B) C to T editing at $C_4$ and $C_6$ under different synchronization conditions with BE4. Mimosine arrest cells in G1/S phase, while nocodazole targets the G2/M border. (C)Total editing, where D is A,G, or T. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

HEK3 site

The HEK3site is on chromosome 9, and has two adjacent target Cs (at positions 4 and 5) in the protospacer. These are processed with similar efficiencies; treatment with both synchronizing agents resulted in a reduction in overall editing efficiencies, with a more pronounced decrease by 20% due to synchronization into G1/S phase by mimosine (Figure 6).

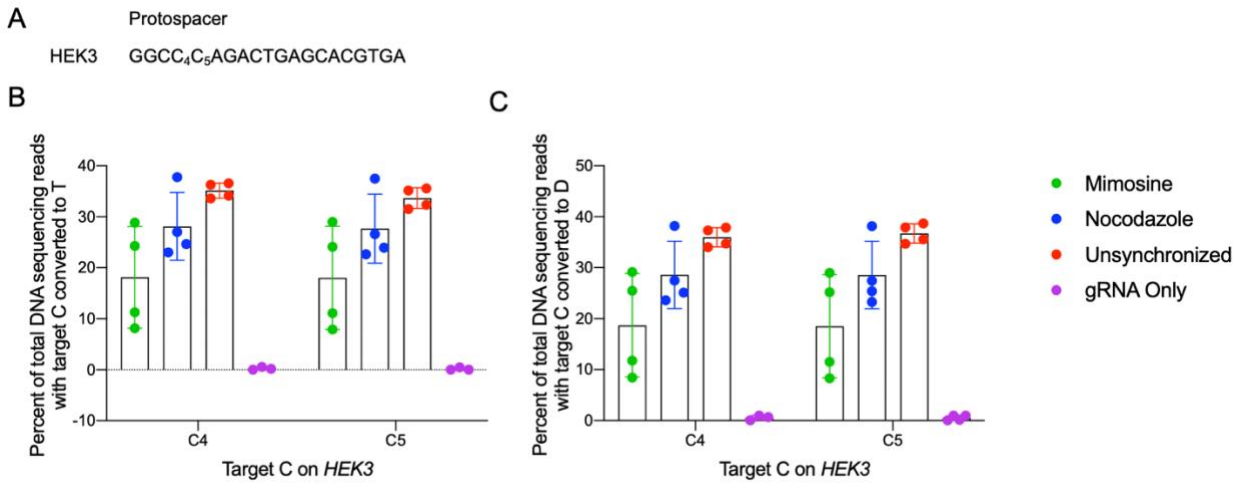



Figure 7: Effect of Synchronization on BE4 Editing at HEK293 site 3.
(A) Protospacer of HEK3 containing two editable C's within the target window. (B) C to T editing at $C_4$ and $C_5$ under different synchronization conditions with BE4. Mimosine arrest cells in G1/S phase, while nocodazole targets the G2/M border. (C)Total editing, where D is A,G, or T. Values and error bars reflect the means and SD of four independent biological replicates performed on different days.

RNF2 site

The RNF2 locus is on chromosome 1, and there are three target cytosines in the RNF2 protospacer. Overall editing for mimosine-treated cells arrested at G1/S by mimosine reduced total editing by 10% for $C_3$ and 17% for $C_6$. Editing at $C_{12}$ was significantly lower compared to $C_3$/$C_6$, and was comparable under all synchronizing conditions.

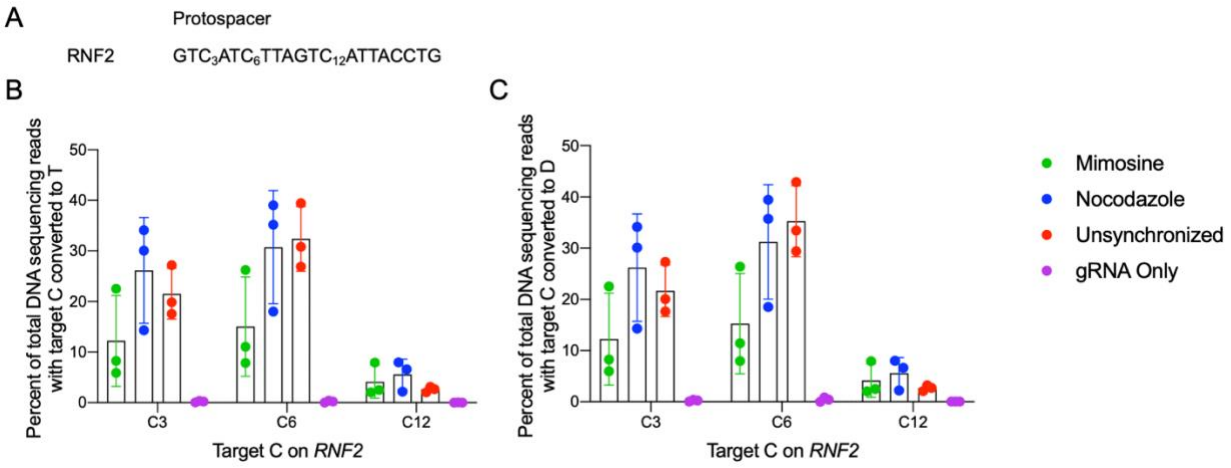



Figure 8: Effect of Synchronization on BE4 Editing at RNF2.
(A) Protospacer of RNF2 containing two editable C's within the target window. (B) C to T editing at $C_3$, $C_6$, and $C_{12}$ under different synchronization conditions. Mimosine arrest cells in G1/S phase, while nocodazole targets the G2/M border. (C)Total editing, where D is A,G, or T. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

BE4ΔUGI

A construct of BE4 lacking UGI was tested to evaluate base editing in synchronized cells without inhibiting uracil excision by UNG.

HEK2

The absence of UGI causes a significant increase in C to non-T editing products for BE4ΔUGI as compared to BE4. At the HEK2 site, desired C to T editing was reduced at C4 from 12% (Figure 5B) to below 4% (Figure 8B), and at $C_6$ from 30% to less than 3%. However, total editing from C to A, G, or T was comparable for BE4ΔUGI and BE4 in unsynchronized cells (Figure 5C, Figure 8C). Synchronizing of cells at the G2/M border with nocodazole decreased total editing

at both target Cs by 7% ($C_4$) and 18% ($C_6$) (Figure 8). Synchronizing cells into G1/S phase by mimosine treatment reduced overall editing from C to A, G, or T by 8% (C4) and 29% (C6), but retained comparable levels of C to T editing at $C_6$.

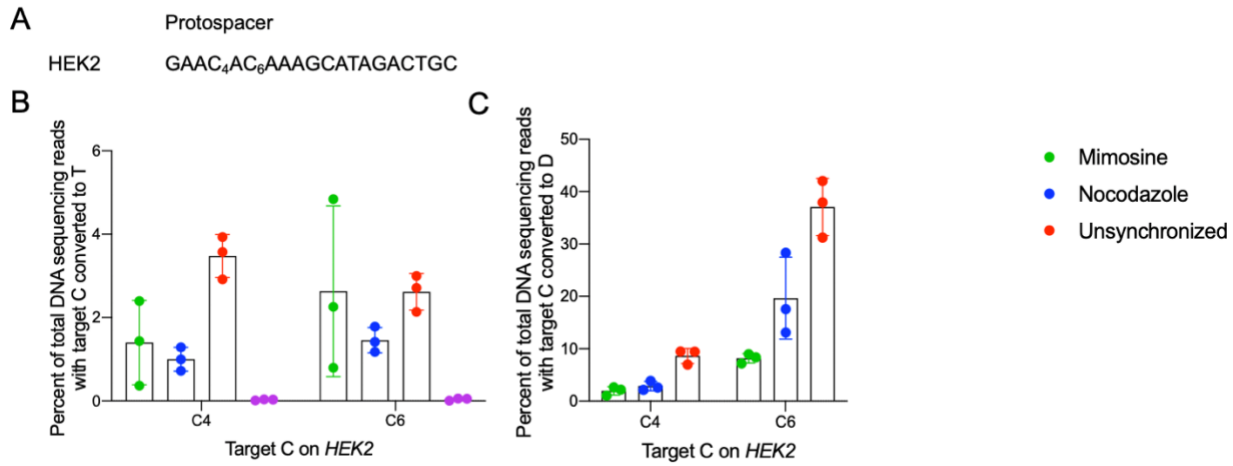



Figure 9: Effect of Synchronization on BE4ΔUGI Editing at HEK293 site 2.
(A) Protospacer of HEK2 containing two editable C's within the target window. (B) C to T editing at $C_4$ and $C_6$ under different synchronization conditions with BE4ΔUGI. Mimosine arrest cells in G1/S phase, while nocodazole targets the G2/M border. (C)Total editing, where D is A,G, or T. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

HEK3 site

Targeting BE4ΔUGI to the genomic loci HEK3 under synchronized conditions reduced both C to T and C to A, G, or T editing by 20% for cells arrested in G1/S and 10% under G2/M arrest. Unlike the HEK2 locus, where dramatic levels of Cto non-T editing are observed, the adjacent cytosine motif in the HEK3 protospacer experiences clean C to T editing with minimal undesired product distribution.

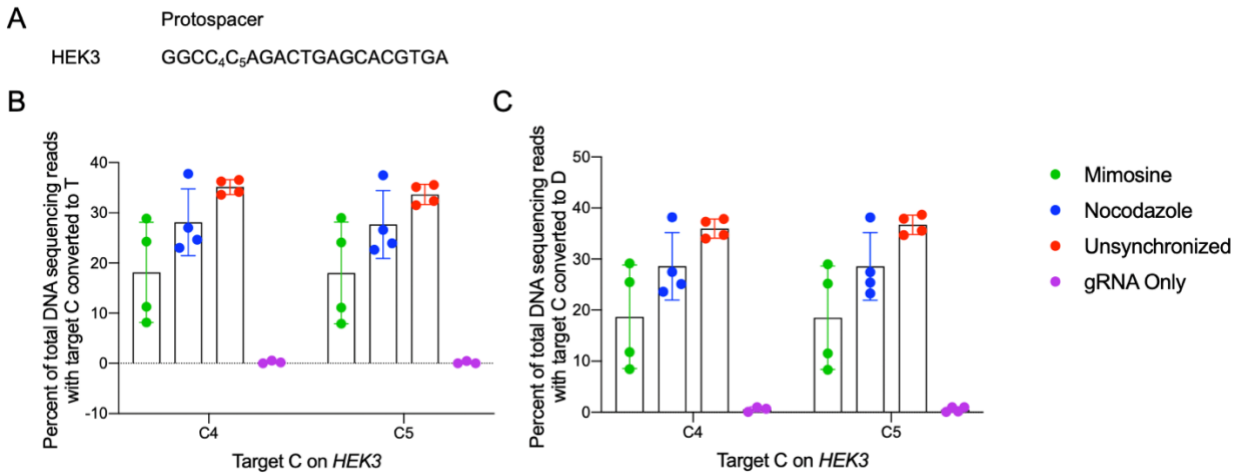



Figure 10: Effect of Synchronization on BE4ΔUGI Editing at HEK293 site 3.
(A) Protospacer of HEK3 containing two editable C's within the target window. (B) C to T editing at $C_4$ and $C_5$ under different synchronization conditions with BE4ΔUGI. Mimosine arrest cells in G1/S phase, while nocodazole targets the G2/M border. (C)Total editing, where D is A,G, or T. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

RNF2 site

Mimosine and nocodazole treatment reduced editing at all three C's within the RNF2 protospacer. At $C_3$, overall editing and C to T editing was reduced from approximately 15% to below 5% when synchronized with either mimosine or nocodazole (Figure 10B,C). $C_6$ total editing was reduced by over 20% for nocodazole arrest at G2/M, and 40% upon synchronizing in G1/S with mimosine. $C_{12}$ had much lower total editing efficiencies for all conditions, but modest decreases were still observed in synchronized cells.

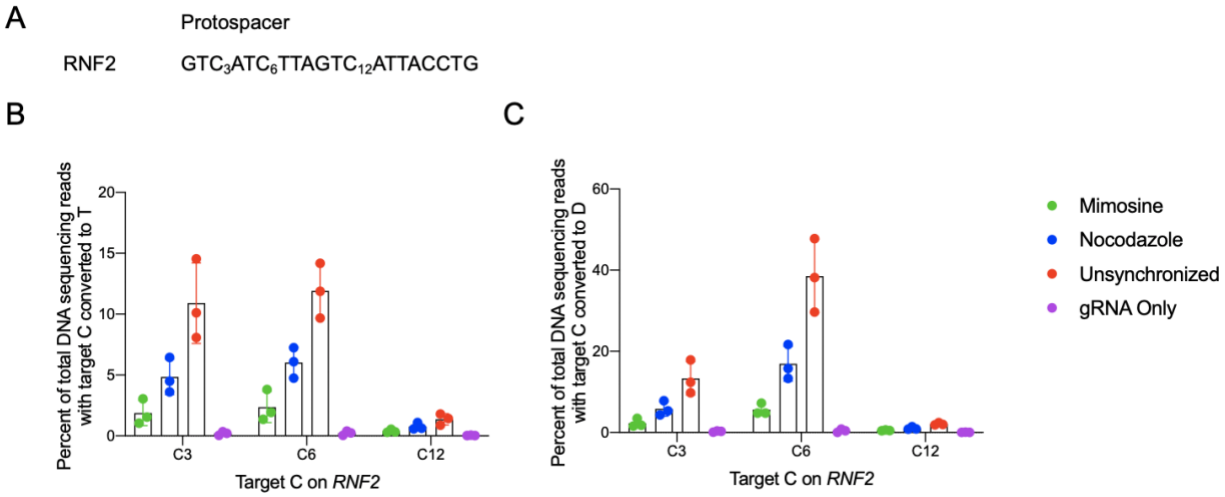



Figure 11: Effect of Synchronization on BE4ΔUGI Editing at RNF2.
(A) Protospacer of RNF2 containing three editable C's within the target window. (B) C to T editing at $C_3$, $C_6$, and $C_{12}$ under different synchronization conditions. Mimosine arrest cells in G1/S phase, while nocodazole targets the G2/M border. (C)Total editing, where D is A,G, or T. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

Product Distribution for BE4 and BE4ΔUGI

Edited reads across all sites for both BE4 and BE4ΔUGI were analyzed for their product distribution by calculating the percent of each individual nucleotide as a fraction of the total (i.e. non-C) reads (Figure 12). When BE4ΔUGI targets genomic locii HEK2 and RNF2, treatment of cells with mimosine reduced the percentage of undesired C to G edits by approximately 20%.

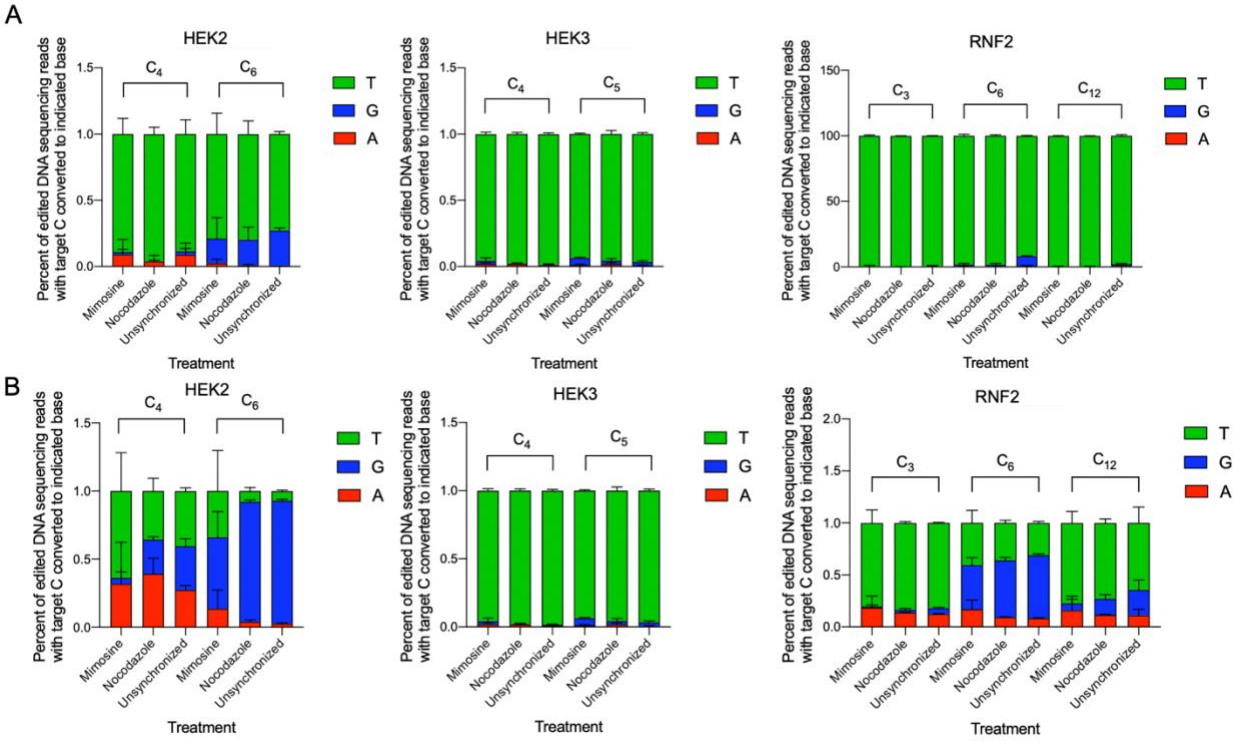



Figure 12: Product Distribution of BE4 and BE4ΔUGI
(A) Product distribution for BE4 targeting genomic loci HEK2, HEK3, and RNF2. Percent reads of A, G, and T nucleotides were calculated as a fraction of the total non-C editing outcomes. (B) Product distribution for BE4ΔUGI targeting genomic loci HEK2, HEK3, and RNF2. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

ABE treatment

Next, we examined the effect of cell cycle arrest on adensoine base editors (ABEs) at the HEK2, HIRA, and PSMB2 loci. There was no significant decrease in editing at these sites for synchronized cells compared with controls, though the distribution of editing efficiencies between duplicates require further experimentation.

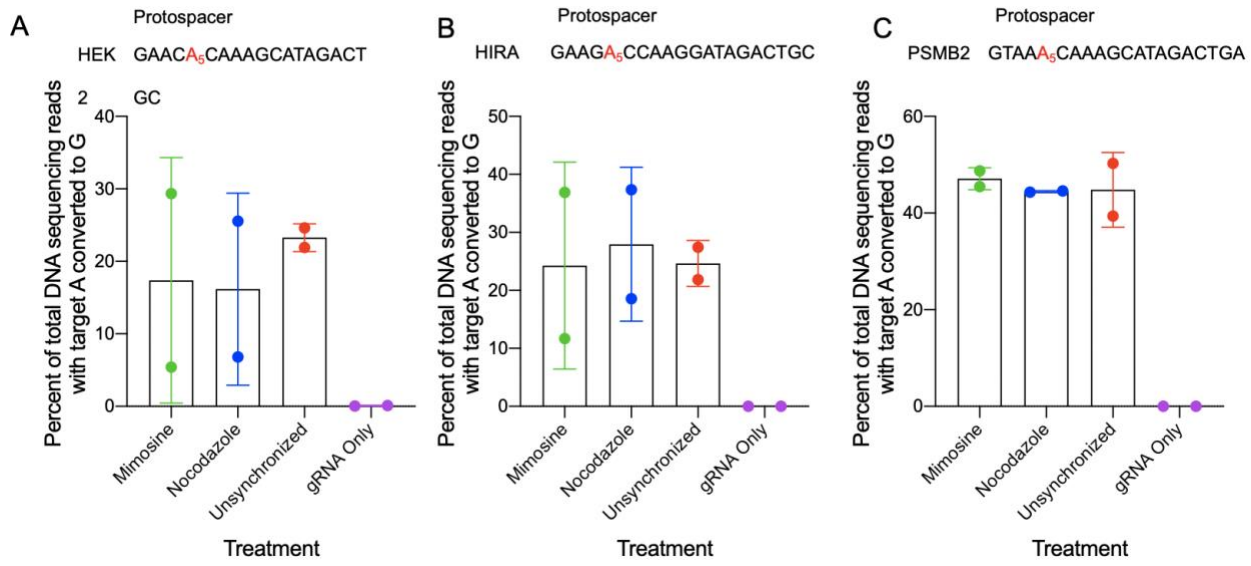



Figure 13: Effect of Synchronization on ABE Editing at HEK2, HIRA, and PSMB2
(A) Editing efficiencies for synchronized cells at the HEK2 genomic locus. Mimosine arrest cells in G1/S phase, while nocodazole targets the G2/M border.(B) HIRA locus. (C) PSMB2 locus.

## Discussion

### Synchronizing at G1/S reduces frequency of C>G conversions with BE4ΔUGI

High rates of C to G conversion are observed at certain sites due to base editing by BE4 and BE4ΔUGI, the mechanism of which is still poorly understood. As the addition of uracil glycosylase inhibitor decreases this phenomenon, likely the excision of uracil by UNG to yield an abasic site initiates this process. The downstream repair mechanism that causes this random mutagenesis is unknown. At three different genomic loci, the enhancement of G1 cells by mimosine reduces the frequency of G to C conversions, suggesting this downstream repair pathway is active during DNA replication and/or cellular division. These results suggest that cytosine base editing may be less error-prone when performed in post-mitotic cells, which are no longer replicating. Furthermore, editing with ABE does not seem to be cell-cycle dependent, as editing efficiencies were comparable across different synchronization conditions.

Chapter II

Introduction

Cytosine base editors (CBEs) consist of A CRISPR-Cas9 nickase tethered to a cytidine deaminase enzyme, APOBEC1. A single-guide RNA (sgRNA) interacts with the Cas9 enzyme to guide it to its DNA target with complementary sequence near a protospacer adjacent motif (PAM). nCas9 invades the double-stranded DNA and forms an R-loop; this displacement exposes a small window of single-stranded nucleotides for deamination by the APOBEC enzyme. This converts the C:G base pair to a U:G intermediate; replication or repair of this intermediate using the uracil-containing strand as a template will result in an overall C:G to T:A base pair conversion, as uracil has the same base pairing properties as thymine. Non-enzymatic hydrolytic deamination of cytosine occurs frequently in the genome; thus, the enzyme uracil-N-glycosylase (UNG) has evolved to combat this type of DNA damage and readily excises uracil present in DNA. As this enzyme was presumed to be responsible for reducing CBE efficiencies, later generation CBEs thus incorporated uracil-glycosylase inhibitors (UGI), which improved the editing efficiency. Unfortunately, UNG inhibition by the UGI component of CBEs is incomplete and causes unpredictable product distributions, defined as the frequency of conversion of C:G base pairs to a mixture of outcomes along with the desired T:A product, specifically high levels of C:G to G:C conversion (Chapter 1 Figure 6, and Figure 9). Previous studies demonstrated that upon knock-out of UNG, the product distribution of CBEs becomes >98% C:G to T:A. The mechanism by which the abasic site intermediate is converted to thee various outcomes (Figure 14) is poorly understood, and further knowledge of this phenomenon is required for base editors to be used in therapeutic settings.

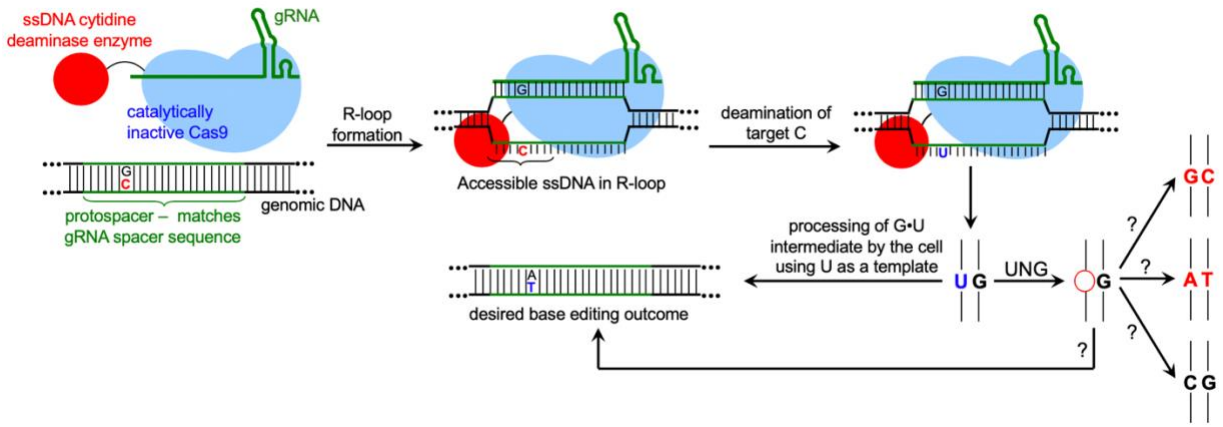



Figure 14: Cytosine Base Editor Mechanism and Repair Outcomes

We plan to use a CRISPRi screen to identify the genes involved in these various CBE outcomes. In CRISPRi whole-genome screens, cells are stably transfected with the transcriptional repressor dCas9-KRAB and a sgRNA library in which each cell targets dCas9-KRAB to the transcription start site of a different gene. The resulting population of cells therefore has a different gene knocked down in each cell, and researchers can use this library to determine what genes are involved in the cellular response to various "treatments". We have established a collaboration with Jacob Corn's lab and will obtain a dCas9-KRAB-expressing cell line with a custom sgRNA library that targets 2,000 genes involved in DNA processing, determined by gene ontology[13]. We plan to couple this CRISPRi system with "treatment" with selection plasmids containing synthetically incorporated U•G lesions. These selection plasmids will be designed such that processing of the U•G lesion by the cell to certain outcomes will result in resistance to the antibiotic Geneticin. Inhibition of genes involved in downstream uracil processing will fail to resolve the mutation as desired and render cells susceptible to Geneticin treatment. One population of wild-type containing cells will be harvested and sequenced to serve as the pre-screen sgRNA library. Genes involved in uracil processing will drop out of the screen

under dCas9-KRAB repression and will be underrepresented upon comparison of sgRNA with the wild type library via sequencing.

## Materials and Methods

### Cloning

Point mutations were installed in the Kanamycin resistance marker by blunt end cloning, using pRSF1030 as a template. Selection plasmids were cloned using blunt end cloning with primers targeting pcDNA3 GFP LIC (Addgene) using Phusion High Fidelity Polymerase. USER cloning removed additional BsaI sequence motifs.

### Restriction enzyme cloning

Plasmids were designed to contain dual BsaI sites at the targeted D190 and I207 residues. Using blunt end cloning, phosphorylated primers bearing the recognition sequence for BsaI amplified pcDNA GFP LIC (Addgene). DpnI treatment degraded the original template plasmid and the fragment was purified by column cleanup. QuickLigase (NEB) was used to ligate the phosphorylated ends, and the ligation mixture was transformed into Mach1 competent cells. Selection on carbenicillin ensured surviving colonies expressed the $Amp^R$ resistance marker and plasmid DNA was extracted by miniprep (Zymo). Sanger sequencing confirmed the presence of flanking GAGACC recognition sequence of BsaI.

### Restriction enzyme Digest

2 μg of selection plasmid was digested with 20 units RsrII for 30 min. at 37ºC in 1X CutSmart (NEB) buffer, then with 20 units rSAP (NEB) for 1 hour at 37ºC to dephosphorylate sticky ends and prevent religation. The fragment was then digested with 20 units BsaI (NEB) for 1 hour and heat inactivated for 30 min. at 37ºC.

### Insertion of uracil lesion

Oligonucleotides containing deoxyuracil were ordered from IDT. They were designed to anneal to one another with base pair overlaps complementary to the overhang sites resulting from BsaI digestion. Using RE-digested fragments described above, ligation was performed for 18 hours at 16ºC with T4 DNA Ligase (NEB).

| Table 3: Oligonucleotide Insert Sequences | |
|---|---|
| D190 WT Fwd. | **GGAT**CTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGG |
| D190 WT Rev. | **TTCC**ACCATGATATTCGGCAAGCAGGCATCGCCATGGGTCACGACGAG |
| D190 PC Rev. | **TTCC**ACCATGATATTCGGCAAGCAGGCATUGCCATGGGTCACGACGAG |
| D190 RE Fwd. | **GGAT**CTCGTCGTGACCCATGGCUATGCCTGCTTGCCGAATATCATGG |
| D190 RE Rev. | **TTC**CACCATGATATTCGGCAAGCAGGCATGGCCATGGGTCACGACGAG |
| I207 WT Fwd. | **TGGT**GGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGC |
| I207 WT Rev. | **GATA**GCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCC |
| I207 RE Fwd. | **TGGT**GGAAAATGGCCGCTTTTCTGGATTCAUCGACTGTGGCCGGCTGGGTGTGGCGGACCGC |
| I207 RE Rev. | **GATA**GCGGTCCGCCACACCCAGCCGGCCACAGTCGGTGAATCCAGAAAAGCGGCCATTTTCC |

## Minimum Inhibitory Concentration

### Bacterial MIC

Plasmids harboring mutations at Asp190, Phe206, Ile207, and Asp208 of the $Kan^R$ gene were transformed into Mach1 competent cells (Invitrogen) against spectinomycin selection. Cells were grown to saturation in 2XYT media + 50 ug/mL Spectinomycin for 17 hours and plated via serial dilution on various concentrations of kanamycin (0, 10, 20, 30, 40, 50 ug/mL).

### Mammalian Cell Culture and MIC Testing

K562 cells (ATCC CRL-3344) were maintained in RPMI media (Life Sciences) supplemented with GlutaMAX (ThermoFisher Scientific), 10% fetal bovine serum (ThermoFisher Scientific), and 100 U/mL Penicillin-Streptomycin (ThermoFisher Scientific), at 37º C with 5% $CO_2$. Plasmids harboring Asp190 and Ile207 mutations in the $G418^R$ gene were transfected into K562 cells in 48-well VWR Multiwell Cell Culture Plates at a density of 150,000 cells per well in 250 µL of RPMI media without Penicillin-Streptomycin. 500 ng of selection plasmid was transfected into the cell suspension, using 1.5 µL of Lipofectamine 2000 (ThermoFisher Scientific) per well according to the manufacturer's protocol. Geneticin (Life Sciences) was introduced at 500, 1000, and 2000 µg/mL, and cell viability was assessed 3 days post-transfection via Trypan Blue (ThermoFisher Scientific) dye exclusion.

| Table 4: Selection Scheme for Uracil Screen | | | |
|---|---|---|---|
| Selection | Mutated Residue | Uracil Incorporation | Possible Outcomes |
| U•G to C•G | Catalytic Residue Asp190 | coding strand GAT<br>template strand UTA<br>Asn | GAT<br>CTA Asp |
| | | | AAT<br>TTA Asn |
| | | | CAT<br>GTA His |
| U•G to G•C | Catalytic Residue Asp190 | coding strand UAT<br>template strand GTA<br>His | GAT<br>CTA Asp |
| | | | TAT<br>ATA Tyr |
| | | | CAT<br>GTA His |
| U•G to T•A | ATP Binding Residue Ile207 | coding strand AUC<br>template strand TGG<br>Thr | ATC<br>TAG Ile |
| | | | ACC<br>TGG Thr |
| | | | AGC<br>TCG Ser |

Results

Antibiotic Selection Scheme

To design a selection system that links U:G processing to cellular survival, we will incorporate the U:G mismatch is such a way that transcription across the template strand will result in a nonfunctional protein. APHIII encodes for an aminoglycoside phosphotransferase which inactivates aminoglycoside-based antibiotics, such as kanamycin (active against prokaryotic cells) and geneticin (active against both prokaryotic and eukaryotic cells). We sought to identify locations within the APHIII gene to incorporate U:G lesions that would inactivate the gene and require processing to C:G, G:C, T:A, or A:T to result in a functional copy of the gene.. Asp190, Phe206, Ile207, and Asp208 are highly conserved residues within the APH family (Figure 15) and could therefore be leveraged to create our selection system. Asp190 is located

within the catalytic loop and directly catalyzes the reaction to inactivate aminoglycosides[8]. Asp208 is also believed to help catalyze phosphate transfer. Phe206 and Ile207 coordinates the nucleobase within the binding pocket via hydrophobic interactions, based on structural data[9]. The codon for Asp190 on the coding strand is 5'-GAT-3', which has the complementary sequence 3'-CTA-5' on its template strand. Mutation of this G:C base pair to either A:T or C:G alters the encoded amino acid to asparagine or histidine respectively, which share different chemical properties from aspartic acid. The carboxylic side chain of aspartic acid facilitates phosphate transfer; the imidazole and carboxamide groups of His and Asn cannot catalyze this reaction and therefore either of these mutations should render the protein inactive (Table 4). Ile207 coordinates ATP within the nucleoside binding pocket via hydrophobic interactions to correctly orient the molecule[6]. Changing the nonpolar side chain of Ile207 to a polar residue (such as Thr, which can be done by installing a C:G to T:A point mutation) should impair phosphotransferase activity, though there have been no mutational studies to validate the structural data.

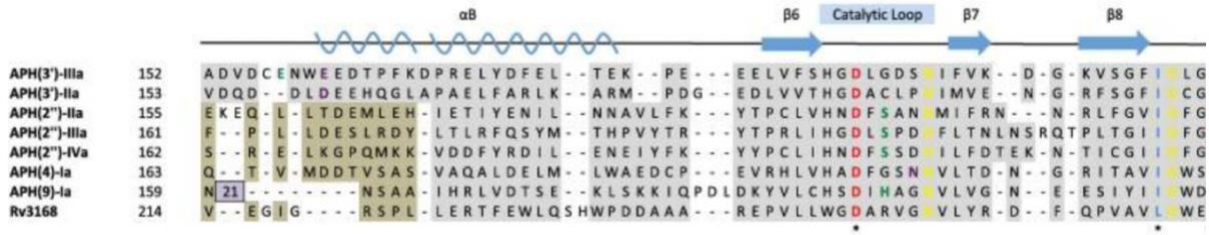


Figure 15: Sequence homology of APH. Asterisks indicate catalytic Asp190 and Ile207.

To confirm antibiotic susceptibility, the point mutants Asp190Asn, Phe206Asn, Phe206Thr, Ile207Thr, Ile207Ser, and Asp208Asn were cloned into a bacterial kanamycin resistance gene. At 40 µg/mL, growth of the Asp190 and Ile207 mutants was severely impaired compared to wild-type. We therefore chose these mutations for mammalian studies, and excluded Phe206, which had no effect on resistance, and Asp208, whose codon is identical to Asp190.

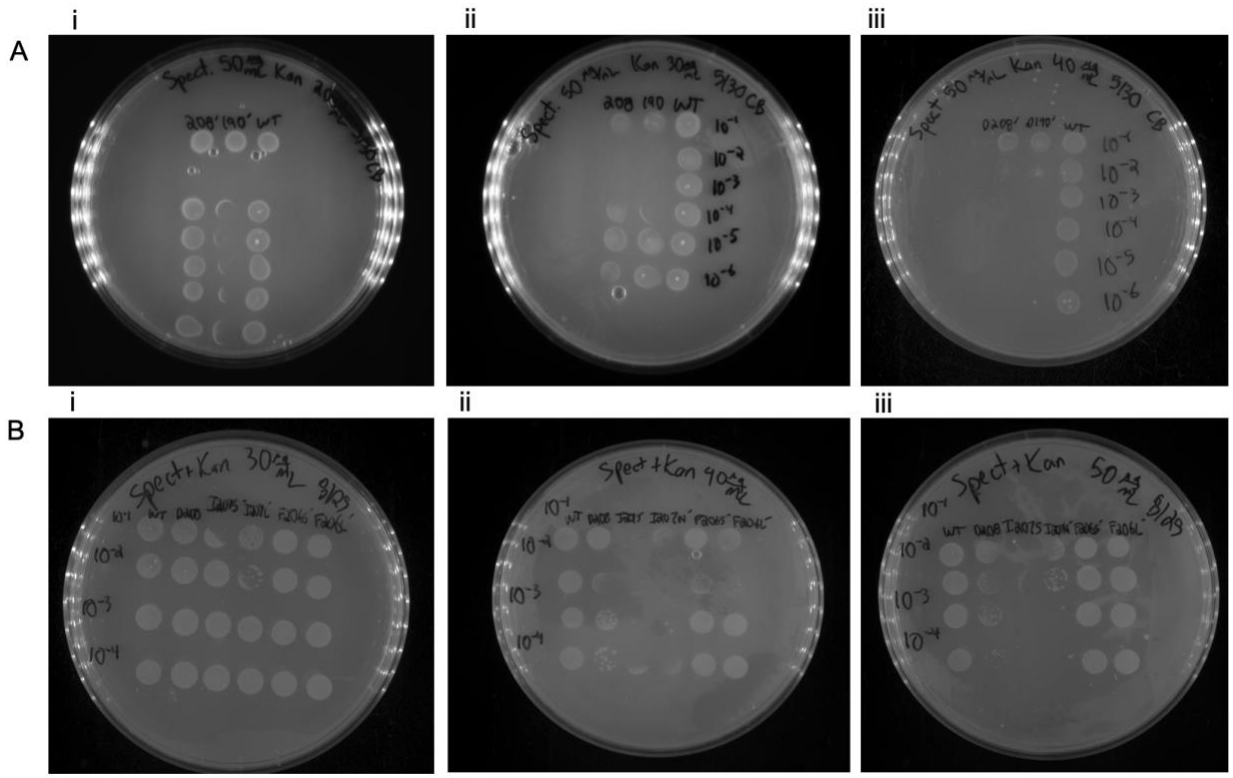



Figure 16: Kanamycin MIC Testing of D190 and I207 Point Mutants
WT, D190 and I207 mutants were plated on 2XYT+Spectinomycin (A) D208N, D190N, WT (left to right) were serially diluted on 20 $\mu$g/mL (i), 30 $\mu$g/mL (ii), and 40 $\mu$g/mL kanamycin (iii). (B) WT, D208N, I207S, I207T, F206N, F206T mutants (left to right) were plated on 30 $\mu$g/mL(i), 40 $\mu$g/mL(ii), and 50 $\mu$g/mL(iii) kanamycin

Next, these mutations were cloned into a mammalian vector, which contains the same APHIII gene with an SV40 promoter. Plasmids bearing Asp190 and Ile207 mutations were transfected into K562 cells. 24 hours post-transfection, cells were treated with geneticin at 500 µg/mL, 1000 µg/mL, and 2000 µg/mL for 72 hours.

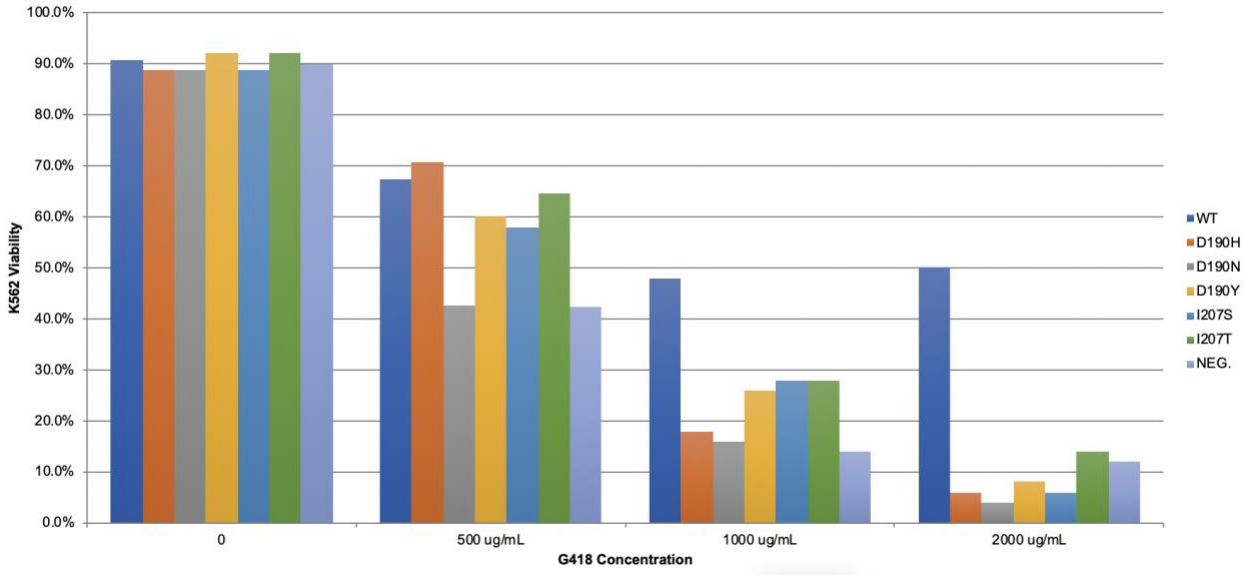



Figure 17: Viability of G418$^{R}$ point mutants in K562 cell line.
Plasmids harboring mutations in D190 and I207 were transfected and treated with increasing concentrations of Geneticin for three days. Cell viability was assessed with Trypan Blue.

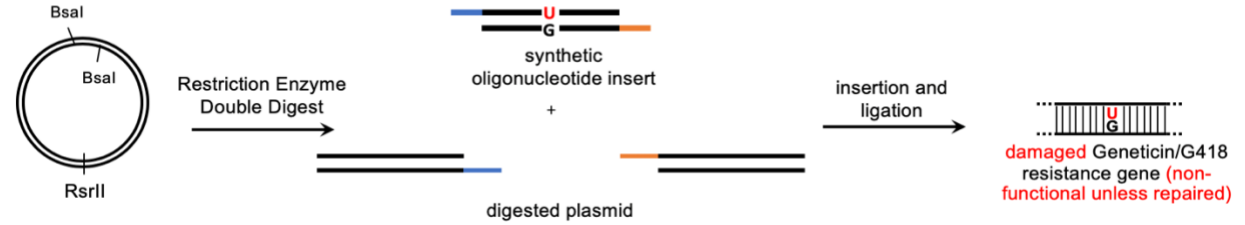
At 10000 µg/mL, cell viability is reduced by ≥20%, from 48% in wild type cells to below 30% for all mutants tested (Figure 14). Wild type viability is likely linked with transfection efficiency, as only cells that receive the G418$^{R}$ marker will be resistant to geneticin. At 2000 µg/mL, viability of D190 mutants were below 10% (D190H 6%, D190N 4%, D190Y 8%) and I207 mutants were below 15% (I207S 6% I207T 14%).

Ligation

Selection plasmids with synthetically incorporated U:G lesions will be generated via restriction enzyme cloning with synthetic DNA inserts (Figure 18). Golden Gate cloning exploits TypeII restriction enzyme's ability to cleave outside of its recognition sequence, and allows custom oligonucleotide incorporation. Proximal cut sites excise a fragment of DNA to generate a plasmid with 4 base pair overhangs; inserts with complementary sticky ends can be annealed and ligated into the vector. This cloning strategy will be used to selectively incorporate uracil-guanine base pairs at the sites of interest. Ligation of custom oligonucleotides bearing

deoxyuracil into linearized DNA will align the U:G base pairs at codons for essential amino acid residues in the APHIII gene.

To evaluate ligation efficiency, plasmids, vectors containing the dual BsaI sites at D190 and I207 were digested with BsaI and RsrII restriction enzymes to generate fragments that can be differentiated by size on an agarose gel with overhangs to anneal the insert (Figure 18). RsrII cut fragments were treated with rSAP (which dephosphorylates the cut ends) to prevent religation. Digested fragments were approximately 2.7 and 3.9 kb in size, that upon ligation produces a 6 kb fragment. Bands were present at 6 kb upon T4 ligation, but also appeared in the no insert control lane (Figure 14). Moving forward, we aim to redesign the BsaI sites to yield sticky ends with less complementarity to prevent religation, and/or increase the rSAP treatment time.



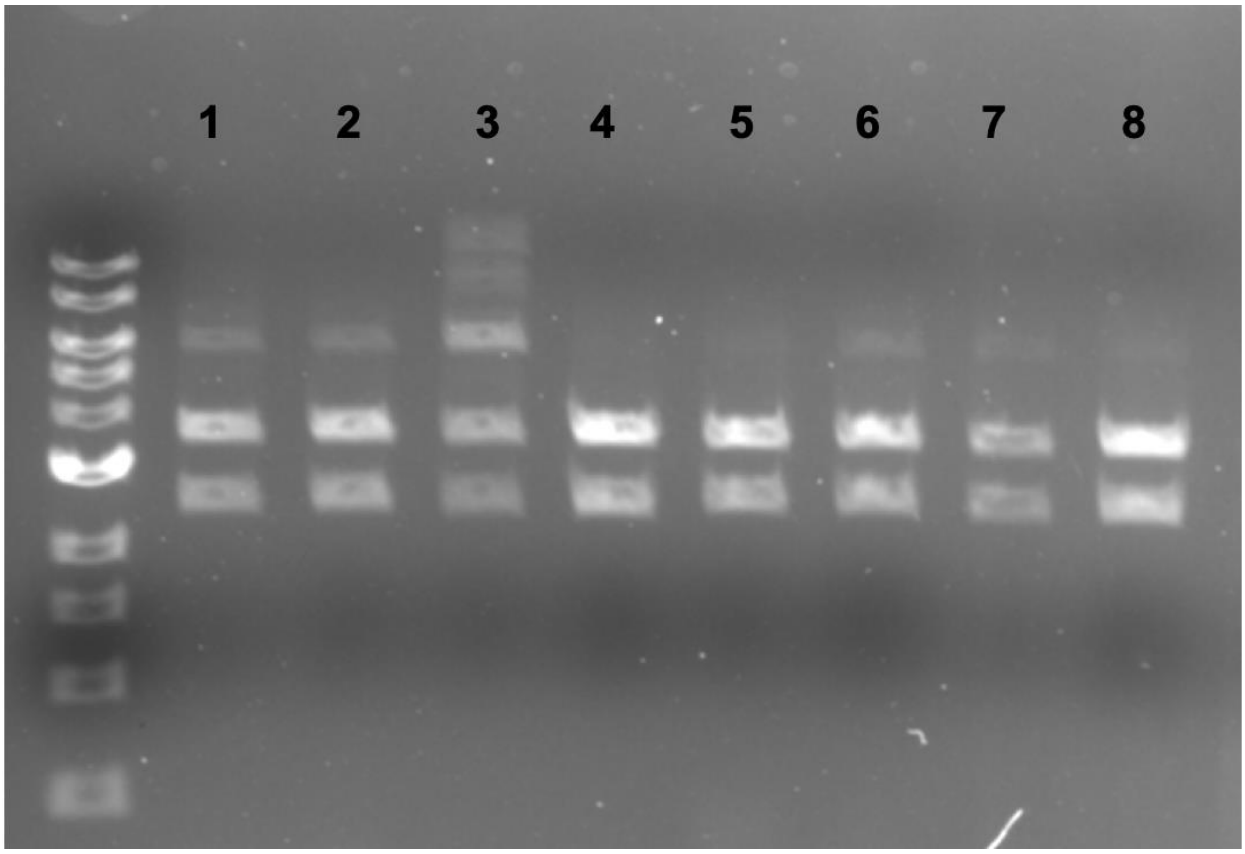




Figure 18: Ligation of Selection Plasmid with Insert
Top: Selection plasmid was double digested with RsrII and BsaI and treated with T4 DNA Ligase overnight. Bottom: (1) D190 Selection Plasmid + WT insert sequence. (2) D190 Selection Plasmid + uracil insert. (3) D190 Selection Plasmid - insert. (4) D190 Selection Plasmid – T4 DNA ligase. (5) I207 Selection Plasmid + WT insert sequence. (6) I207 Selection Plasmid + uracil insert. (7) I207 Selection Plasmid - insert. (8) D190 Selection Plasmid – T4 DNA ligase.

Discussion

Establishing a robust selection for the CRISPRi screen is essential for useful interpretation of dropout hits. Mutations mimicking possible repair outcomes severely reduce antibiotic resistance at both the Asp190 and Ile207 codon bacterial and mammalian cells. At

2000 ug/ML Geneticin, such mutations decrease viability by approximately 40%. Vectors with dual BsaI cut sites at both of these codons have been prepared, as well as oligonucleotides bearing targeted U:G base pairs that will generate lesions in frame with the coding sequence for $G418^R$ gene. Ligations of fragments produced by BsaI and RsrII double digest need to optimized with rSAP treatment to dephosphorylate sticky overhangs and prevent religation. If rSAP treatment is insufficient, then likely the site at D190 will have to be redesigned with less similar overhang sequences.

Table 5: Selection Scheme with MIC and Viability

| Selection | Mutated Residue | Uracil Incorporation | Possible Outcomes | Kan MIC (ug/mL) | G418 Viability (2000 ug/mL) |
|---|---|---|---|---|---|
| U•G to C•G | Catalytic Residue D190 | coding strand GAT<br>template strand UTA<br>Asn | GAT<br>CTA Asp | >100 | 50% |
| | | | AAT<br>TTA Asn | 40 | 4% |
| | | | CAT<br>GTA His | 40 | 6% |
| U•G to G•C | Catalytic Residue D190 | coding strand UAT<br>template strand GTA<br>His | GAT<br>CTA Asp | >100 | 50% |
| | | | TAT<br>ATA Tyr | 40 | 8% |
| | | | CAT<br>GTA His | 40 | 6% |
| U•G to T•A | ATP Binding Residue Ile207 | coding strand AUC<br>template strand TGG<br>Thr | ATC<br>TAG Ile | >100 | 50% |
| | | | ACC<br>TGG Thr | 40 | 14% |
| | | | AGC<br>TCG Ser | 40 | 12% |

References


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

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

3. Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., Liu, D. R. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Science advances*, *3*(8) (2017).

4. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. Programmable base editing of A T to G C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471(2017).

5. Badran AH, Guzov VM, Huai Q, Kemp MM, Vishwanath P, Kain W, Nance AM, Evdokimov A, Moshiri F, Turner KH, Wang P, Malvar T, Liu DR Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature 533*,7601: 58-63. (2016)

6. Jackman, J. & O'Connor, P. M. Methods for Synchronizing Cells at Specific Stages of the Cell Cycle. *Current Protocols in Cell Biology 00***,** 8.3.1-8.3.20 (1998).

7. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research* **24,** 1012–1019 (2014).

8. Shi, K., Caldwell, S. J., Fong, D. H. & Berghuis, A. M. Prospects for circumventing aminoglycoside kinase mediated antibiotic resistance. *Frontiers in Cellular and Infection Microbiology 3***,** (2013)

9. Nurizzo, D., Shewry, S.C., Perlin, M.H., Brown, S.A., Dholakia, J.N., Fuchs, R.L., Deva, T., Baker, E.N., and Smith, C.A. The crystal structure of aminoglycoside-3′-phospho-transferase-IIa, an enzyme responsible for antibiotic resistance. *Journal of Molecular Biology 327*, 491–506. (2003).

10. Krude, T. Mimosine arrests proliferating human cells before onset of DNA replication in a dose-dependent manner. *Experimental Cell Research 247***,** 148–159 (1999).

11. Peña-Diaz J, Hegre SA, Anderssen E, Aas PA, Mjelle R, Gilfillan GD, Lyle R, Drabløs F, Krokan HE, Sætrom P. Transcription profiling during the cell cycle shows that

subset of Polycomb-targeted genes is upregulated during DNA replication. *Nucleic acids research* vol. 41,5 (2013): 2846-56.

12. Stumpf, C. R., Moreno, M. V., Olshen, A. B., Taylor, B. S. & Ruggero, D. The translational landscape of the mammalian cell cycle. *Molecular Cell* 52**,** 574–582 (2013).

13. Richardson, C.D., Kazane, K.R., Feng, S.J., Zelin, E., Bray, N.L., Schäfer, A.J., Floor, S.N., and Corn, J.E. CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. Nature Genetics *50*, 1132–1139 (2018)