UNIVERSITY OF CALIFORNIA SAN DIEGO

Investigations into the Cellular Mechanisms of Base Editing

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biochemistry and Molecular Biophysics

by

Cameron Andrew Burnett

Committee in charge:

Professor Alexis Komor, Chair
Professor Partho Gosh
Professor Catriona Jameison
Professor Jeremy Klosterman
Professor Dong Wang

2023

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

University of California San Diego

2023

DEDICATION

To the family, friends, and associates who have supported me through the good and bad times, it means the world to me, and words cannot sufficiently express my gratitude.

TABLES OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

To my lab mates and colleagues, thank you for your vital assistance throughout the duration of my studies. I was part of the Komor group's foundation in 2017 as a Master's student with four other PhD students: Brodie, Elizabeth, Mallory and Sifeng, as well as Kartik from the Paesani lab. The first few years were demanding, balancing classes and teaching with establishing protocols, setting up instruments, and in general navigating the wild jungle that is a new lab. You never made me feel like I was any less for not being in the PhD program, and for that I am eternally grateful. It's reassuring even when I would feel lost or frustrated in graduate school, your kindness and empathy perpetually lessened those feelings. Carlos and Quinn, I knew during your rotations you would both make excellent additions to the lab. Whether spit balling ideas for future projects, coordinating who will fill up the liquid nitrogen, playing hoops or surfing, you further enhanced the good vibes of the lab. To the students who are the future of the lab, Ben, Kara, Natalie, Rachel, Sam, and Zulfiqar, as well as our resident postdoc Michael, you all are great people that ensure a bright future for the group and I'm glad to have gotten to know you. The day to day within the lab has been a joy as we all were able to be friends inside and out of the walls of Urey Hall. I was extremely fortunate to be a part of a fantastic group and wish you nothing but the best in your future endeavors.

To Alexis, I would not have made it this far without your continuous patience, guidance, and support. You always made time for us and our projects, brought us to fascinating conferences, and importantly helped to secure funding for my tuition remittance while I was still a master's student. Even in my shortcomings I never felt any judgement or disdain from you, you only encouraged the best for me and my future and were always

looking out for projects, collaborations, and networking with industry contacts. I know there were adversities around every corner of establishing the lab, but you always handled it well and showed tremendous leadership through the struggles.

To my committee thank you for your support during my studies. While a lot of this was via Zoom during the pandemic, the vital advice on the future directions of my projects assisted in my understanding of what it truly means to be a scientist. I really appreciate your perspective on how to be inquisitive and take everything into consideration for the iterative nature of our discipline. Your time and consideration have been extremely valuable and appreciated.

Chapters 2 and 3 have been adapted from: **Burnett CA**, Wong A, Vasquez C, McHugh CA, Yeo GW, and Komor AC. Examination of the Cell Cycle Dependence of Cytosine and Adenosine Base Editors. Frontiers in Genome Editing (2022). The dissertation author was the primary author of this paper.

VITA

2015 Bachelor of Science in Biochemistry, Boston College, Boston

2018 Master of Science in Biochemistry, University of California San Diego

2023 Doctor of Philosophy in Biochemistry and Molecular Biophysics, University of California San Diego

PUBLICATIONS

Burnett CA, Wong A, Vasquez C, McHugh CA, Yeo GW, and Komor AC. “Examination of the Cell Cycle Dependence of Cytosine and Adenosine Base Editors.” *Frontiers in Genome Editing* (2022).

Prugar,E*, **Burnett CA***, Chen X, and Hollingsworth NM. “Coordination of Double Strand Break Repair and Meiotic Progression in Yeast by a Mek1-Ndt80 Negative Feedback Loop.” *Genetics* (2017). *co-first authors

Rees HA, Minella AC, **Burnett CA**, Komor AC, and Gaudelli NM. "CRISPR-Derived Genome Editing Therapies: Progress from Bench to Bedside." *Molecular Therapy* (2021).

ABSTRACT OF THE DISSERTATION

Investigations into the Cellular Mechanisms of Base Editing

by

Cameron Andrew Burnett

Doctor of Philosophy in Biochemistry and Molecular Physics

University of California San Diego, 2023

Professor Alexis Komor, Chair

The development of CRISPR-derived genome editing technologies enables precise manipulation of DNA sequences within the human genome. One class of these editing tools is base editors, which install point mutations with high efficiency and specificity, without the disadvantages of traditional Cas9-based nucleases. This is due to the introduction of uracil and inosine intermediates, as opposed to breakage of the DNA backbone. As such, they rely on more ubiquitous DNA repair pathways than wild type Cas9. Cas9 is also highly reliant on cell cycle phase, depending on homology-directed repair (HDR), which is only active in late S and G2 phases. To date, no one has investigated the relationship of base editors to the cell cycle.

We examine how reliant base editors are on the cell cycle using chemical synchronization experiments. We optimize protocols to evaluate base editing under synchronization conditions. We find that nickase-derived BEs function independently of the cell cycle, while non-nicking variants are highly dependent on S-phase processes. We discover that G1 synchronization during cytosine base editing causes significant increases in C•G to A•T "byproduct" introduction rates, which can be leveraged to discover new strategies for precise C•G to A•T base editing.

We investigate the cause for C•G to A•T editing events and design an integrated fluorescent reporter system to identify genes implicated in these editing outcomes. We discover TLS polymerases POLI and POLK are likely responsible for these unique editing events. We further examine base editors' response to nucleotide concentration to identify if that alters editing outcome. Finally, we develop and test a cell-cycle regulated base editor to apply our discovery towards technology development.

CHAPTER 1

Introduction

Overview of Nucleic Acids and the Enzymes That Cut Them

Deoxyribonucleic acid, or DNA, is known as the code of life, stores heritable information to encode the essential building blocks of life[1]. The central dogma of biology dictates that DNA is transcribed into ribonucleic acid (RNA), ultimately translated into protein[2]. Differences in sequences of DNA determines amino acid identity within proteins, and in organism phenotypes on a larger scale. Examining genetic variance between humans, 96% of this is in the form of single nucleotide variance[3] bearing differences in a single letter of the genetic code. Development of highly efficient and thoroughly characterized gene editing tools can aid in the understanding of the 99% that lack clinical interpretation[3,4], as well as the development of therapeutics to treat life-threatening genetic diseases.

Since the advent of molecular cloning, there has been a great deal of scientific interest in altering DNA inside human cells as a therapeutic. Initial attempts focused on gene transfer therapy: insertion of foreign DNA into the genomes of cells, most commonly using a viral vector[5]. Nuclease-based genome editing in mammalian cells originated with meganucleases in the 1990s[6]. Rare-cutting endonucleases induced double-strand breaks (DSBs) in endogenous loci with a unique recognition sequence, detecting the resultant gene editing by DNA repair mechanisms. There are two major pathways that cells will use to process DSBs: non-homologous end-joining (NHEJ) and homology directed repair (HDR)[7,8]. Processing by NHEJ results in short insertion or deletion sequences and represents the dominant outcome following DSB induction[9,10]. The sequences of these indels are in general reproducible (and can, to certain degrees, be predicted), but they cannot be dictated and depend on the local sequence context

surrounding the DSB[11,12]. HDR can process DSB's using an exogenous donor template with homology to the region of interest. Editing outcomes following DSB induction can contain a mixture of editing products: templated repair and undesired outcomes (insertions and deletions). Later iterations of nuclease-based genome editing strategies include Zinc Finger Nucleases and TALEN systems, which advanced the targetability and proficiency of genome editing technology. Zinc Finger Nucleases function in an analogous manner as restriction enzymes: they possess recognition domains that specifically bind to a sequence of DNA and coupled Fok1 nucleases that introduce double-strand breaks (DSBs) at the site, where addition of a DNA template introduced gene replacement via recombination[13].

It was discovered the Cas9 enzyme could be reprogrammed to recognize, bind, and cleave a custom DNA sequence via base pairing between the desired locus and a piece of RNA, termed the guide RNA (gRNA)[14–16]. The target sequence, called the protospacer, requires the presence of a three base pair sequence called the protospacer adjacent motif (PAM). Following binding of the Cas9 protein, DNA unwinds for interaction with the D10 and H840 amino acid residues, where the DSB is initiated at the -4 position relative to the PAM[17]. The programmability and high efficiency across organisms, specifically mammalian cells, are major factors in the widespread and fast adoption of this technology for genome editing.

Rise of Base Editors

Base editing was developed in 2016 by Dr. Alexis Komor, utilizing the programmability of CRISPR-Cas9, but without reliance on DSBs[18]. The first cytosine base editor consisted of catalytically inactive Cas9 (dCas9), incapable of inducing DSBs, tethered to a DNA-modifying catalyst that acts upon single stranded DNA (ssDNA). The fused enzyme deaminates the exocyclic amine of the exposed DNA bubble. The cytidine deaminase rAPOBEC1, was used for the first generation of cytosine base editors (CBEs), and specifically deaminates cytosine to uracil. Following replication/repair, the uracil:guanine base pair is corrected to thymidine, resulting in an overall C>T edit. The first generation of base editors suffered from low efficiency, likely due to excision of uracil by uracil-N-glycosylase (UNG), which excised the uracil intermediate to restore the C:G base pair. This was confirmed in UNG knockout lines which drastically increased product purity of C>T editing outcomes. Second generation BE2's incorporated a UNG inhibitor (UGI) to preclude excision of the U:G intermediate. Third generation base editor restored the catalytically inactive dCas9 to a variant which nicks the opposite strand to target cellular repair responses towards nicked DNA strands, similar to processing of Okazaki fragments. Development of an adenine base editor (ABE) used seven rounds of extensive directed evolution to optimize an adenine base editor that recognizes DNA as a substrate, rather the native function of TadA as a tRNA deaminase[19] (Figure 1). Analogous to uracil intermediates installed by CBE, ABE7.10 converts A•T base pairs to G•C via inosine-containing intermediates. Lacking enzymes to act upon deoxyinosine bases, ABEs have high rates of efficiency, with negligible rates of indel introduction.

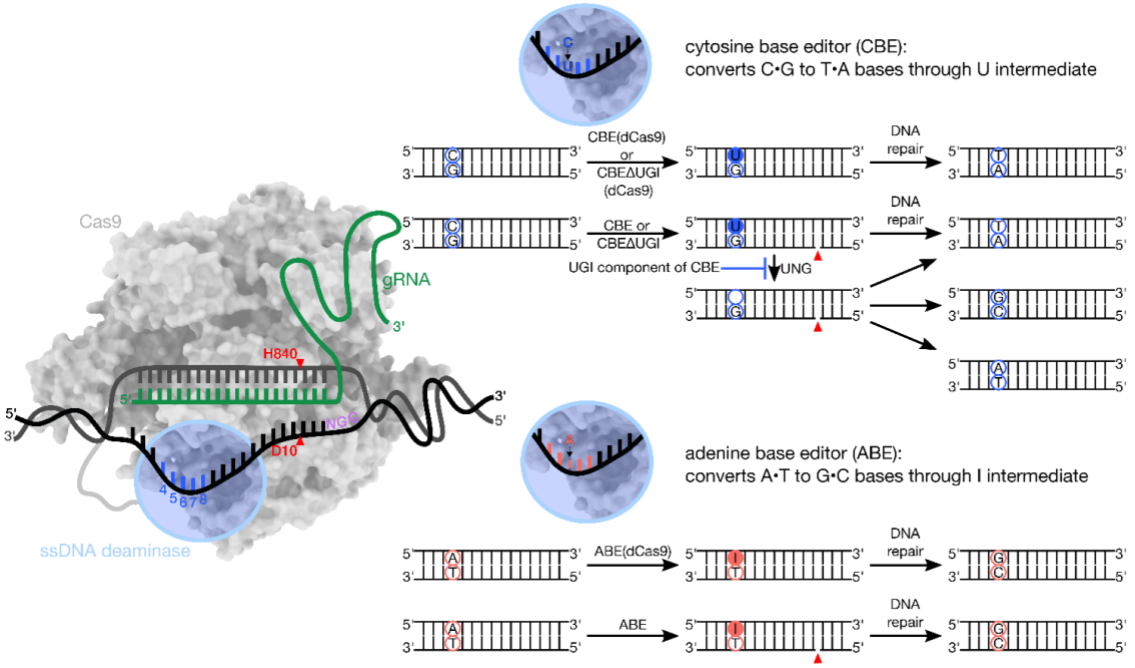

Figure 1.1: Overview of base editing. Base editors consist of a ssDNA modifying enzyme (blue) covalently tethered to a catalytically inactive or impaired Cas9 (dCas9 or Cas9n) protein (grey). The base editor binds to a genomic locus of interest via protospacer adjacent motif (PAM, shown in purple) recognition and Watson-Crick-Franklin base pairing between the gRNA and the protospacer. Local denaturation of the DNA during R-loop formation exposes ~5 nucleotides (protospacer positions 4-8) to the ssDNA modifying enzyme, which will deaminate the target base to the U or I intermediate. The modified base will get permanently converted to a canonical base pair following cellular replication or repair across the lesion.

Cytosine Base Editors and Uracil Repair

Spontaneous deamination of cytosine occurs frequently in the cellular environment and is repaired via base excision repair pathways. CBEs rely on uracil intermediates and unidentified DNA repair pathways that process uracil to achieve C>T editing. Traditionally, this uracil is repaired via base excision repair, either short patch or long patch, referring to either single nucleotide replacement or 2-15 base pairs, respectively[20]. Excision of uracil is primarily catalyzed by UNG, but there exist several other glycosylases, such as SMUG1 and TDG[21,22]. Resultant abasic sites are recognized by an apurinic/apyrimidinic endonuclease, likely in the APEX family, cleaving the sugar phosphate backbone of DNA[23]. Uracil repair is subject to base

excision repair processes, which revert the intermediate base U:G base pair to cytosine, but the presence of non C>T editing outcomes suggest alternative processing. This would include mismatch repair and/or translesion synthesis. Products of BE-dependent deamination can be considered mismatches, resembling the intermediates processed by noncanonical mismatch repair pathways (ncMMR). This lesion is recognized by the mismatch recognition complex MutSα, consisting of MSH2 and MSH6[24]. Translesion synthesis refers to DNA repair proteins capable of bypassing diverse DNA lesions[25]

CHAPTER 2

Introduction

We performed initial experiments by slightly modifying existing protocols[26], synchronizing cells prior to transfection with plasmid based delivery. We selected mimosine, thymidine, lovastatin, and nocodazole for synchronization of the cells at the G1/S, G0/M, and G2/M borders of the cell cycle, respectively. Mimosine is a non-protein amino acid isolated from plants that arrests cells in late G1 prior to the onset of S phase. 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. Thymidine halts cells in G1/S by blocking DNA replication via disruption of the deoxynucleotide metabolism pathway. Nocodazole inhibits mitotic spindle formation and blocks cells prior to M phase division. Evaluation of synchronization was assessed with propidium iodide (PI). PI intercalates nucleic acid, binding to the DNA of the ethanol-fixed, RNAse treated cells, producing a direct correlation of fluorescent signal to DNA content.

Observations of transfections of base editors to synchronized cells revealed BE expression, visualized by GFP transfection marker, is inhibited with prior synchronization. We tested the efficacy of synchronization, the viability of the treated cells, and the effect on transfection efficiency. Results of synchronizing HEK293T cells and K562 cells reveal robust synchronization for thymidine and nocodazole in the G1/S and G2/M phases of the cell cycle. Lovastatin did not have a significant difference in G1 content as compared to asynchronous controls. Mimosine exhibited less effective synchronization of cells in G1/S compared to thymidine, as well as interactions with iron chelating DNA repair proteins, and thus was excluded from analysis.

Protocol optimization revealed that delivery of transfection reagents occurs 6 hours following transfection is ideal for balancing expression of the base editor construct with synchronization. At 12 hours, synchronization was complete, and base editor expression, as revealed by GFP expression, was below saturation. To confirm that editing does not occur prior to synchronization, we conducted time course experiments for ABE7.10, BE4max, and BE4maxΔUGI, henceforth referred to as ABE, CBE, and CBEΔUGI. Cells were transfected in 48-well plate formats, and gDNA was collected at 12, 18, 24, 36, 48, 72, and 96 hour time points. Genomic loci were amplified, subjected to high throughput sequencing, and genome editing efficiency was analyzed via CRISPResso2.

Results

Timeline of Base Editing

To determine the optimal experimental conditions for combining cell synchronization with quantification of base editing outcomes, we first conducted a time course experiment to observe the kinetics of editing by CBE, CBEΔUGI, and ABE with established gRNAs (Figure 2.1). We transfected HEK293T cells with plasmids encoding base editor and gRNA and extracted genomic DNA (gDNA) at 12, 18, 24, 36, 48, 72, and 96 hours post-transfection. Genomic loci of interest were amplified, subjected to high throughput sequencing (HTS), and analyzed for genome editing efficiencies using CRISPResso2. We observed a gradual, consistent increase in editing efficiency over time by ABE (Figure 2.1) throughout the course of the entire experiment. In contrast, editing by both CBE variants appeared to peak at 36 hours, followed by a slight decrease at 48 hours which then recovered and increased through the end of the experiment. In fact, when each datapoint is normalized to the highest editing observed for each

locus, up to 50% of overall editing by CBE is observed within 18 hours of transfection. These differences in the timelines of editing between ABE and CBE could be due to inherent differences in how quickly the two intermediates are introduced by their respective enzymes, or due to differences in the timing of how the two intermediates are processed by the cell. Notably, on average 39 ± 24% of overall ABE editing and 39 ± 13 % (mean ± SD for n = 3 biological replicates per site, averaged over three different sites) of overall CBE editing (when normalized to the highest editing percentage observed for each locus) occurred between the 18- and 36-hour timepoints (Figure 2.1), which informed our experimental protocol to combine chemical synchronization with base editing.

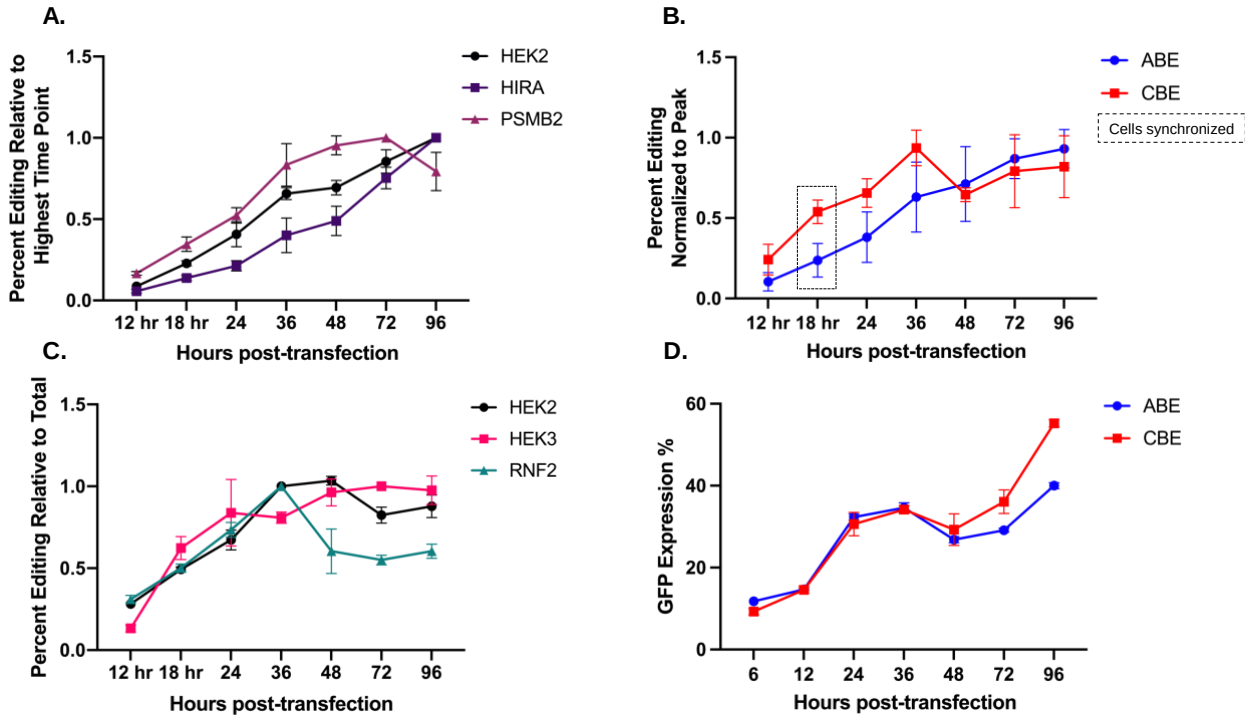

Figure 2.1: Time course of base editing. (a,b,c) Cells transfected with ABE7.10, BE4max, BE4maxΔUGI extracted genomic DNA at the indicated time points. Target sites were amplified via PCR and subjected to next-generation sequencing. The results were quantified as a fraction of the total max editing. For BE4ΔUGI, editing is defined as editing of C to any non-C base. (d) GFP expression over time. Cells were transfected at the indicated time points and analyzed for GFP expression via flow cytometry.

Chemical Inhibitors Arrest Cells After 12 Hours of Treatment

Cells can be arrested at cell cycle borders using chemical inhibitors such as lovastatin (M/G1 arrest), mimosine (G1/S), thymidine (G1/S) and nocodazole (G2/M)[27]. To determine the time frame of synchronization for our experiments, we treated HEK293T and K562 cells with each of these chemical inhibitors and monitored synchronization of the cells at 6, 12, and 18 hours by examining propidium iodide staining of fixed cells. Flow cytometry analysis allows for visualization of DNA content, which is then used to quantify the fraction of cells in the population that are in G1, S, and G2/M phases.

Unsynchronized HEK293T cells exhibit on average 43 ± 1 % of cells in G1, 42 ± 1 % of cells in S, and 15 ± 0.2 % of cells in G2/M (Figure 2.2). We observed a modest increase in synchronization of cells after 6 hours of nocodazole treatment (28 ± 3 % of cells in G1, 26 ± 1 % of cells in S, and 46 ± 1.5 % of cells in G2/M, Figure 2.2), which drastically increased by 12 hours post-treatment to 4 ± 4 % of cells in G1, 14 ± 1.5 % of cells in S, and 82 ± 4 % of cells in G2/M (Figure 2.2). Moderate synchronization was also observed after 6 hours of treatment with thymidine(60 ± 6 % of cells in G1, 40 ± 6 % of cells in S, and 0±0 % of cells in G2/M, Figure 2.2), but 12 hours of treatment was able to arrest cells at G1 and early S phase, with 80 ± 6 % of cells in G1, 19 ± 6 % of cells in S, and 0.7 ± 0.5 % of cells in G2/M (Figure 2.2). Mimosine treatment also arrested cells at the G1/S border after 12 hours, with 59 ± 1 % of cells in G1, 37 ± 1 % of cells in S, and 0.5±0.9(?) % of cells in G2/M. Analogous treatment of K562 cells with these three chemicals produced similar results (Figure 2.2). In contrast, lovastatin had no effect on either HEK293T or K562 cell cycle distribution with up to 18 hours of treatment (Figure 2.2). Cells held synchronized for more than 48 hours were determined to be non-viable by Trypan Blue staining (Figure 2.3A).

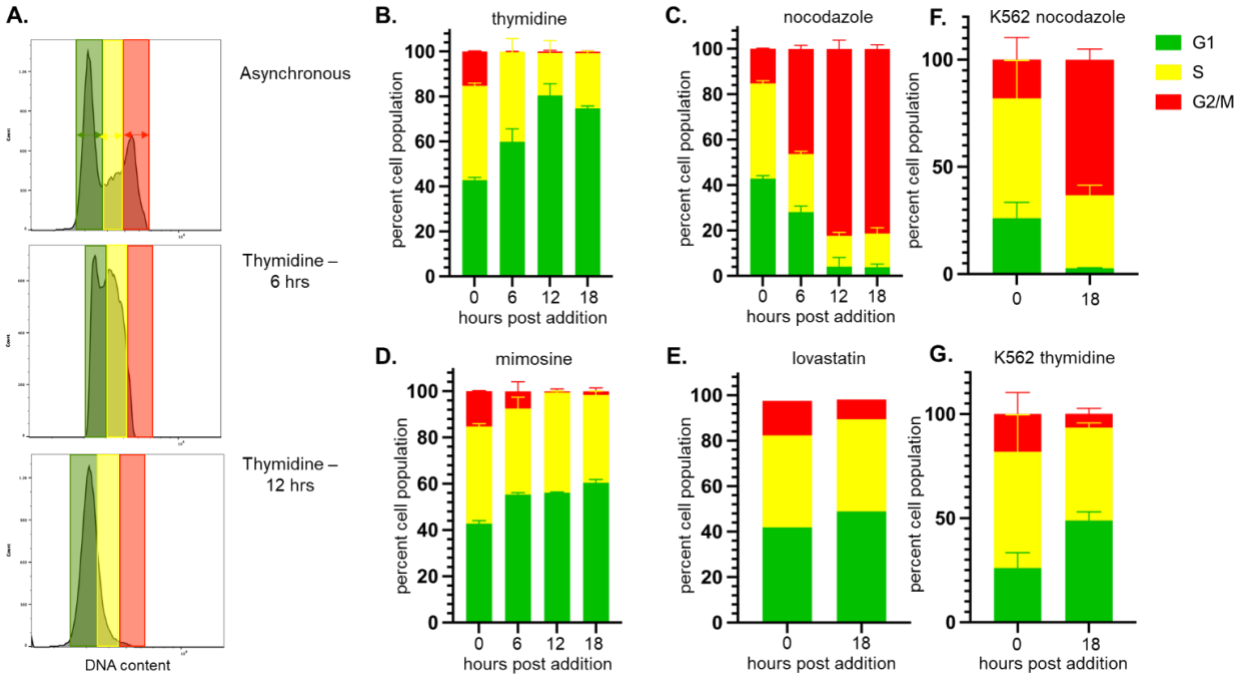

Figure 2.2: Cell synchronization quantification. HEK293T cells (A-E) or K562 cells (F-G) were treated with chemical inhibitors for 6, 12, and 18 hours and stained with propidium iodide (PI) after ethanol fixation. The percent of the total cell population in G1, S, or G2/M was determined based off PI fluorescence level, which quantifies DNA content. Representative PI stain plots are shown in (A), and quantification is shown for Thymidine (B and G), Nocodazole (C and F), Mimosine (D), and Lovastatin (E). Values and error bars reflect the means and SD of three independent biological replicates performed on different days, except for lovastatin which represents a single biological replicate.

Delaying Synchronization Is Required to Maintain Equal Base Editors Expression Levels

Synchronization agent was added first to HEK293T cells, followed by transfection of BE and gRNA plasmids after 17 h. At 24 h post-transfection, flow cytometry was used to measure and compare green fluorescent protein (GFP) expression levels of the synchronized samples to that of the unsynchronized samples (using BE-P2A-GFP constructs, in which BE and GFP are transcribed together but translated into separate proteins[28]). We observed drastic reductions in BE expression levels when cells were pre-synchronized with nocodazole or mimosine, most likely due to a reduction in either transfection efficiency or protein translation rates caused by synchronization (a decrease from an average of 60 ± 5% of cells exhibiting GFP fluorescence for

asynchronous cells to 32 ± 1% for G1 arrested cells by mimosine and 24 ± 2% for G2/M arrested cells by nocodazole, with thymidine within error at 57 ± 3% GFP, mean ± SD for n = 3 biological replicates, Figure 2.3). Adding synchronization agents at the time of transfection improved the percentage of GFP positive cells to 49 ± 5% for mimosine (G1 arrested), 43 ± 5% for nocodazole treated cells (G2/M arrested), and 57 ± 2% for thymidine treated cells (G1 arrested) at 24 h post-transfection. However, delaying the addition of synchronization agents until 6 h after transfection was determined to be the best balance between preserving GFP (and therefore BE) expression levels, while ensuring the majority of base editing activity (as determined by our time-course experiments) occurred when cells were synchronized. Specifically, 62 ± 3% of mimosine treated cells (G1 arrested, no statistically significant difference), 52 ± 5% of nocodazole treated cells (G2/M arrested, representing a 16 ± 8% decrease in GFP fluorescence compared to asynchronous cells), and 57 ± 2% of thymidine treated cells (G1 arrested, no statistically significant difference) exhibited GFP fluorescence at 24 h post-transfection (Figure 2.3B). We chose to move forward using thymidine as our G1 synchronization agent due to its more complete synchronization of cells in G1. By delaying addition of synchronization agents until 6 h after transfection, cells are fully synchronized by 18 h post-transfection, allowing us to observe changes in editing efficiency and precision that occur between 18 and 54 h post-transfection due to synchronization.

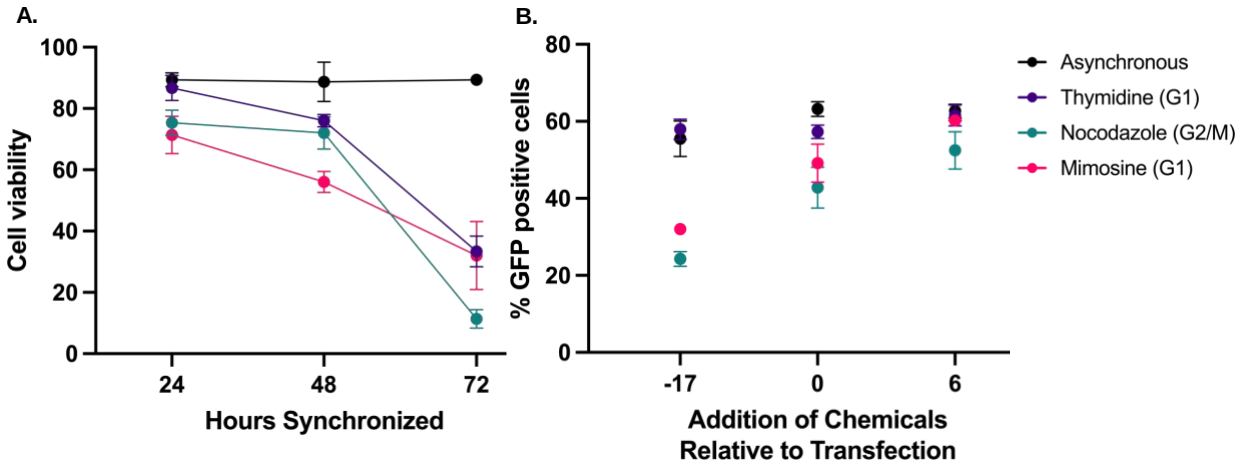

Figure 2.3: Synchronization effects on cell viability and BE expression levels. (A) Cell viability, as assessed via Trypan blue stain, relative to length of time that cells are held synchronized. (B) Percent of cells with GFP fluorescence relative to the time that synchronization agents were added (with respect to transfection of BE and gRNA). Cells were transfected with ABE-P2A-GFP plus gRNA and analyzed 24 hours post-transfection for GFP expression via flow cytometry.

Methods

Constructs and Molecular Cloning

All BE plasmids were constructed with USER cloning[29] with pCMV ABEmax_P2A_GFP (Addgene #112101) and pCMV_AncBE4max_P2A_GFP (Addgene #112100) plasmids as template, using Phusion U Hot Start Polymerase (ThermoFisher Scientific). All sgRNA expression plasmids were generated using blunt-end cloning with pFYF1230 (Addgene plasmid #47511) as a template, using Phusion High-Fidelity DNA Polymerase (New England BioLabs). All DNA vector amplification was carried out using NEB 10-β competent cells (New England BioLabs). All plasmids were purified using the ZymoPURE II Plasmid Midiprep Kit (Zymo Research D4200).

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

For all HEK293T cell transfections, 100,000 HEK293T cells in 250 µL of DMEM media without Penicillin-Streptomycin were added per well to 48-well VWR Multiwell Cell Culture Plates on top of lipofectamine/plasmid mixtures. For all K562 cell transfections, 50,000 K562 cells in 250 µL RPMI media without Penicillin-Streptomycin were added per well to 48-well VWR Multiwell Cell Culture Plates on top of lipofectamine/plasmid mixtures transfected at a density 50,000 cells per well in 250 µL RPMI media without Penicillin-Streptomycin. The lipofectamine/plasmid mixtures consisted of 1000 ng of BE plasmid, 250 ng of sgRNA plasmid, and 1.5 µL of Lipofectamine 2000 (ThermoFisher Scientific) in 25 µL of total volume, made up with Opti-MEM (Gibco #31985-070). Chemical inhibitors were added 6 hours after transfection from stock solutions (described below) to result in final concentrations of 5 mM (Thymidine), 800 µM (Mimosine), or 200 ng/mL (Nocodazole).

Preparation of Synchronizing Agents

Nocodazole (Sigma) was prepared in DMSO to a stock solution concentration of 20 mg/mL. This stock solution was diluted to 50 µg/mL immediately prior to addition to the cells,

and 1.1 µL of this diluted stock solution was added to the 275 µL of media in each well, for a final concentration of 200 ng/mL.

Thymidine (Sigma) was prepared in 1X PBS to a stock solution concentration of 50 mM. 30 µL of this stock solution was added to the 275 µL of media in each well, for a final concentration of 5 mM.

Mimosine (Sigma) was prepared in 1X PBS to a stock solution concentration of 10 mM. 24 µL of this stock solution was added to the 275 µL of media in each well, for a final concentration of 800 µM.

Lovastatin (Sigma) was prepared in 95% ethanol to a stock solution concentration of 70 mM with a pH of 7.5. This was diluted to 400 µM, then 30.5 µL of this stock solution was added to the 275 µL of media for a final concentration of 40 µM.

Flow Cytometry Analysis of Cell Synchronization

3 x 10^5 HEK293T or K562 cells were plated in a T25 flask in 5 mL of media and synchronizing agents were added to final concentrations as indicated above for 6, 12, or 18 hours. Cells were washed with 10 mL PBS, detached with 2 mL of TrypLE (HEK293T cells only), and collected by centrifugation for 10 minutes at 400g. Cells were resuspended at 1 x 10^6 cells/mL in 1 mL cold PBS, added to 9 mL cold 70% ethanol, and stored for at least 4 hours at -20º C. After ethanol fixation, cells were centrifuged at 400g, washed with 10 mL cold PBS, then stained with 400 µL PI solution (0.1% Triton X-100 [Sigma], 0.2 mg/mL RNAse [Sigma], 0.02 mg/mL PI [Sigma] in PBS). Cells were incubated at 37º C for 15 minutes prior to analysis. Cells were gated to exclude doublets and non-viable cells. Fluorescent signal from PI staining was analyzed via histogram on either a BLDSRFortessa or BioRad S3e cell sorter.

Flow Cytometry Analysis of GFP Fluorescence

For all GFP fluorescence measurements, 1 x 10^6 cells were resuspended in FACS buffer (1% FBS, 50 µM EDTA pH 8.0, 2 µg/mL PI [Sigma]) and filtered through a cell-strainer. Non-viable cells and doublets were eliminated via gating parameters. Flow cytometry was performed on a BioFortessa or S3e cell sorter (Bio-Rad).

Transfection Efficiency Quantification

Transfection efficiency was determined via flow cytometry analysis of cells 24 hours post-transfection. Chemical inhibitors were added to HEK293T cells at -17, 0 and 6 hours relative to transfection of BE and gRNA plasmids (as described above). 24 hours post-transfection, cells were washed with PBS, detached from the plate with 50 µL Accutase (Innovative Cell Technologies), resuspended in 250 µL FACS buffer, and analyzed by flow cytometry as described above. The percent of cells with GFP fluorescence was analyzed via Flowjo.

High-Throughput DNA Sequencing (HTS) of Genomic DNA

Transfected cells were rinsed with 150 µL PBS (ThermoFisher Scientific) per well at the indicated time points 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 bearing homology to the target site and relevant Illumina forward and reverse adapters (Table 2), 1 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 from 2% agarose gel with QIAquick Gel Extraction Kit (Qiagen) and quantified using NEBNext Ultra II DNA Library Prep Kit (NEB) on a CFX96 system (BioRad). Samples were then sequenced on an Illumina MiniSeq according to the manufacturer's protocol.

Chapters 2 has been adapted from: **Burnett CA**, Wong A, Vasquez C, McHugh CA, Yeo GW, and Komor AC. Examination of the Cell Cycle Dependence of Cytosine and Adenosine Base Editors. Frontiers in Genome Editing (2022). The dissertation author was the primary author of this paper.

CHAPTER 3

Introduction

Repair of DNA damage is mediated by several different repair pathways, which are active to varying degrees throughout different stages of the cell cycle[30]. For example, HDR machinery is primarily expressed during the late Synthesis (S) and Gap 2 (G2) phases of the cell cycle[26,30]. In fact, one strategy to improve HDR to NHEJ ratios in DSB-reliant genome editing experiments involved pre-synchronizing cells prior to delivery of Cas9:sgRNA ribonucleoprotein complex and donor template to coordinate the initial "burst" of genome editing activity with S- and/or G2/M- phase[26]. Base editing, in contrast, has been hypothesized to rely on the mismatch repair (MMR) and/or base excision repair (BER) pathways, which are thought to be less drastically regulated by the phases of the cell cycle than HDR. In support of this, several studies have reported successful base editing in post-mitotic, non-proliferating cell types[31,32]. However, no systematic investigation of the cell cycle-dependence of base editing has been conducted. A detailed understanding of how base editing outcomes can change with respect to the cell cycle would inform us on the cell types most amenable to efficient and precise base editing, reveal new strategies to enhance certain base editing outcomes, and provide more information about the DNA repair mechanisms by which these tools operate.

We examined how base editors respond to cell cycle arrest via chemical inhibitors that synchronized cells at various phases of the cell cycle. This has previously been done with traditional Cas9 genome editors, which rely on homology-directed repair for precise editing, but preferentially induce indel formation via NHEJ. As HDR is active predominantly in late S and G2 phases, enriched cell populations to transit through HDR-active phases synchronously, seeking to enhance precise repair outcomes. We applied this approach towards DSB-free base

editor technology, to see if there is cell cycle dependence for editing, for different types of Cas9-derived base editor, and to see how the editing outcomes could change for cytosine base editors. We examined the mechanistic underpinnings of C•G to A•T editing outcomes via RNAi knockdown experiments, nucleotide pool assays and development of a base editor with G1 restricted activity. Through investigations of certain facets of DNA repair involved in base editor intermediates, this work is substantial in understanding how to leverage these outcomes towards the development of better genome editing technologies.

Results

Synchronization effects on ABEs

After establishing the base editing and synchronization timeline, we conducted experiments in HEK293T cells using ABE7.10 constructs, both nickase and dCas9 variants. HEK293T cells were transfected with ABE7.10 or ABE(dCas9) and gRNA, then treated with synchronizing agents for 48 hours beginning 6 hours post-transfection. Cells were lysed, gDNA was extracted, and genomic loci of interest were amplified, subjected to HTS, and analyzed for genome editing efficiencies using CRISPResso2 (Figure 3.1A). These two editors differ in that ABE(dCas9) would produce an intermediate lacking a nick on the strand across from the inosine. While dCas9-derived BEs are used less frequently than their Cas9n counterparts due to their reduced overall efficiencies, we were interested in observing the impact that nicking of the unedited strand has on the cell cycle dependence of base editing. Notably, we observed no statistically significant changes in A•T to G•C editing efficiencies by ABE at any of the three sites (the *HEK2*, *HIRA,* and *PSMB2* loci) upon synchronization in G1 using thymidine relative to asynchronous populations ($P>0.05$ two-tailed Student's *t*-test, Figure 3.1B). When cells were

synchronized in G2/M using nocodazole, we observed a 37 ± 11% decrease in A•T to G•C editing efficiency at the *HEK2* site, a 45 ± 5% decrease at the *HIRA* site, and no reduction at the *PSMB2* site (Figure 3.1B). We repeated these experiments in K562 cells and observed the same overall trends, demonstrating the generality of these data (Figure 3.2A). In direct contrast, A•T to G•C editing efficiencies by ABE(dCas9), significantly decreased in HEK293T cells at all three sites with both synchronization conditions (Figure 3.1D). When synchronized in G1 by thymidine, A•T to G•C editing decreased by 63 ± 7% at the *HEK2* site, 67 ± 15% at the *HIRA* site, and 64 ± 16% at the *PSMB2* site (mean ± SD for n = 3 biological replicates). When synchronized in G2/M by nocodazole, A•T to G•C editing decreased by 70 ± 4% at the *HEK2* site, 68 ± 9% at the *HIRA* site, and 68 ± 12% at the *PSMB2* site (Figure 3.1D). These experiments were repeated in K562 cells and we observed the same drastic decreases in A•T to G•C editing efficiencies by ABE(dCas9) at two of the three genomic loci (Figure 3.3A). These data indicate clear differences in the mechanisms by which these two editors' intermediates are processed. Additionally, the nickase-derived ABE (which is more commonly used than the dCas9-derived ABE due to its higher efficiency) displays minimal cell cycle dependence, in direct contrast to DSB-reliant genome editing methods.

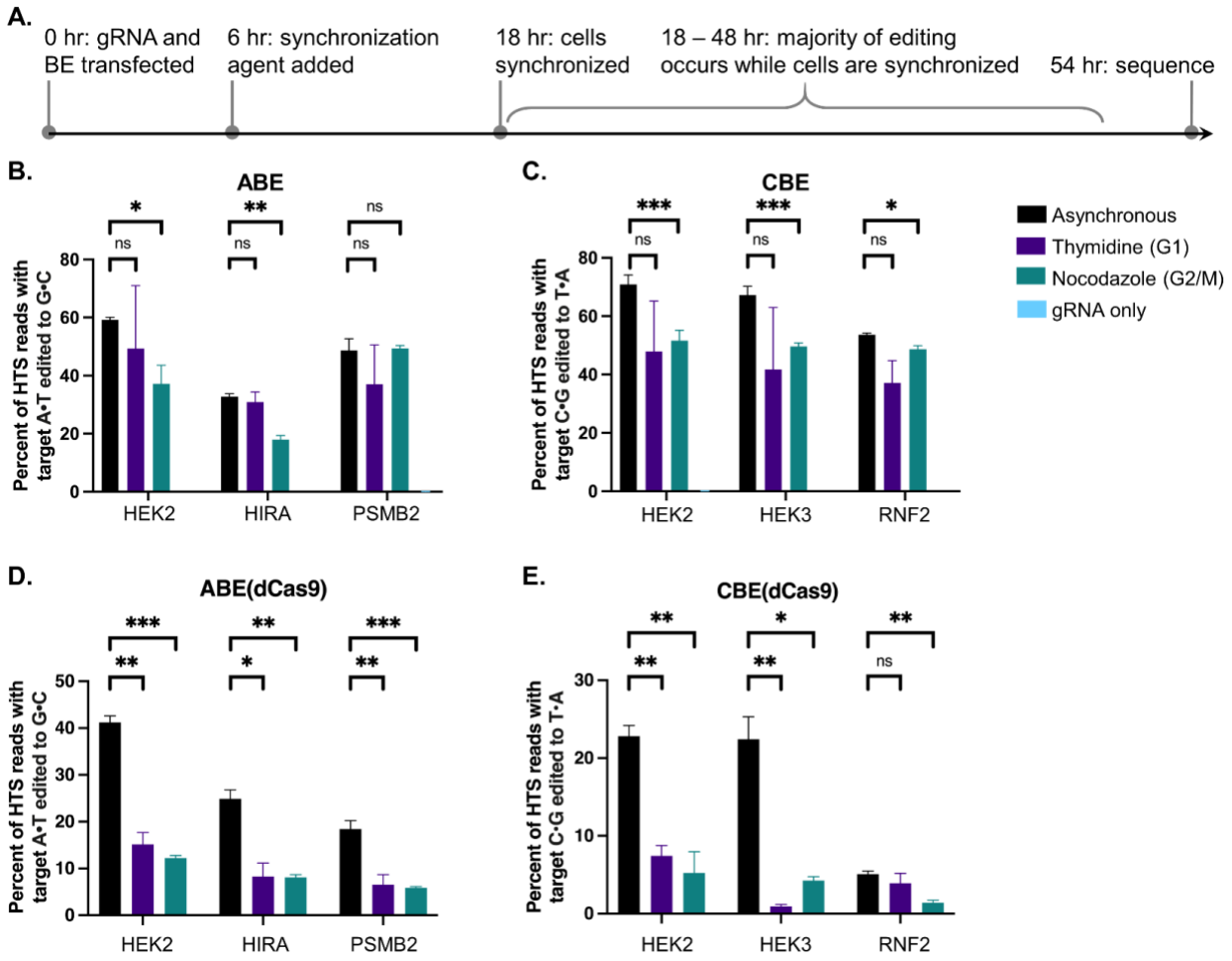

Figure 3.1: Cell cycle synchronization effects on base editing efficiencies and precision of ABE, ABE(dCas9), CBE, and CBE(dCas9) in HEK293T cells. (A) Cells were transfected with BE plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. As a negative control, cells were transfected with gRNA plasmid only (gRNA only sample). The genomic DNA was extracted, and target loci were amplified via PCR and subjected to high-throughput sequencing (HTS). Genome editing efficiencies [percent of total HTS reads with the target A•T base converted to G•C for ABE and ABE(dCas9), or percent of total HTS reads with the target C•G base converted to T•A for CBE and CBE(dCas9)] were quantified with CRISPResso2. Base editing efficiencies by ABE (B), CBE (C), ABE(dCas9) (D), and CBE(dCas9) (E) are plotted. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. Asterisks reflect p value calculations of unpaired *t* test, one tailed (ns indicates not significant, **p*<0.05, ***p*<0.01, ****p*<0.001).

Synchronization Effects on CBE Editing Efficiencies

We repeated these experiments with CBE and CBE(dCas9) as well. Editing by CBE produced mainly C•G to T•A editing products (product purity, the percent of edited reads in which the target C•G is edited to a T•A, was >80% at all three sites, Figure 3.5A), thus we analyzed only these outcomes. Upon G1 synchronization using thymidine, average C•G to T•A editing efficiencies by CBE decreased slightly at all three sites (Figure 3.1C), but these decreases were not statistically significant ($P>0.05$ two-tailed Student's *t*-test). Upon G2/M synchronization using nocodazole, editing decreased by 27 ± 7% at the *HEK2* site, 26 ± 5% at the *HEK3* site, and 5 ± 3% at the *RNF2* site (Figure 3.1C, mean ± SD for n = 3 biological replicates). However, these decreases could be attributed to the decrease in BE expression levels due to synchronization as described earlier (16 ± 8% decrease in GFP expression levels after synchronization with nocodazole). Analogous experiments in K562 cells again yielded comparable results (Figure 3.2B). C•G to T•A editing efficiencies by CBE(dCas9) drastically decreased at two of the three sites with both synchronization conditions (Figure 3.1E); when synchronized in G1 by thymidine editing decreased by 67 ± 9% at the *HEK2* site and 96 ± 18% at the *HEK3* site, and when synchronized in G2/M by nocodazole editing decreased by 77 ± 14% at the *HEK2* site and 81 ± 17% at the *HEK3* site. C•G to T•A editing efficiencies by CBE(dCas9) at the *RNF2* locus in asynchronous cells were quite low (5 ± 0.4%), but we observed a 72 ± 12% decrease upon synchronization in G2/M by nocodazole and no statistically significant decrease upon synchronization in G1 by thymidine. We repeated these experiments in K562 cells and again observed the same overall trends (Figure 3.2B). Taken together with the ABE synchronization results, these data suggest that both types of dCas9-derived BEs rely heavily on

S-phase-dependent pathways to incorporate their respective point mutations, while Cas9n-derived BEs are much less cell cycle dependent in comparison.

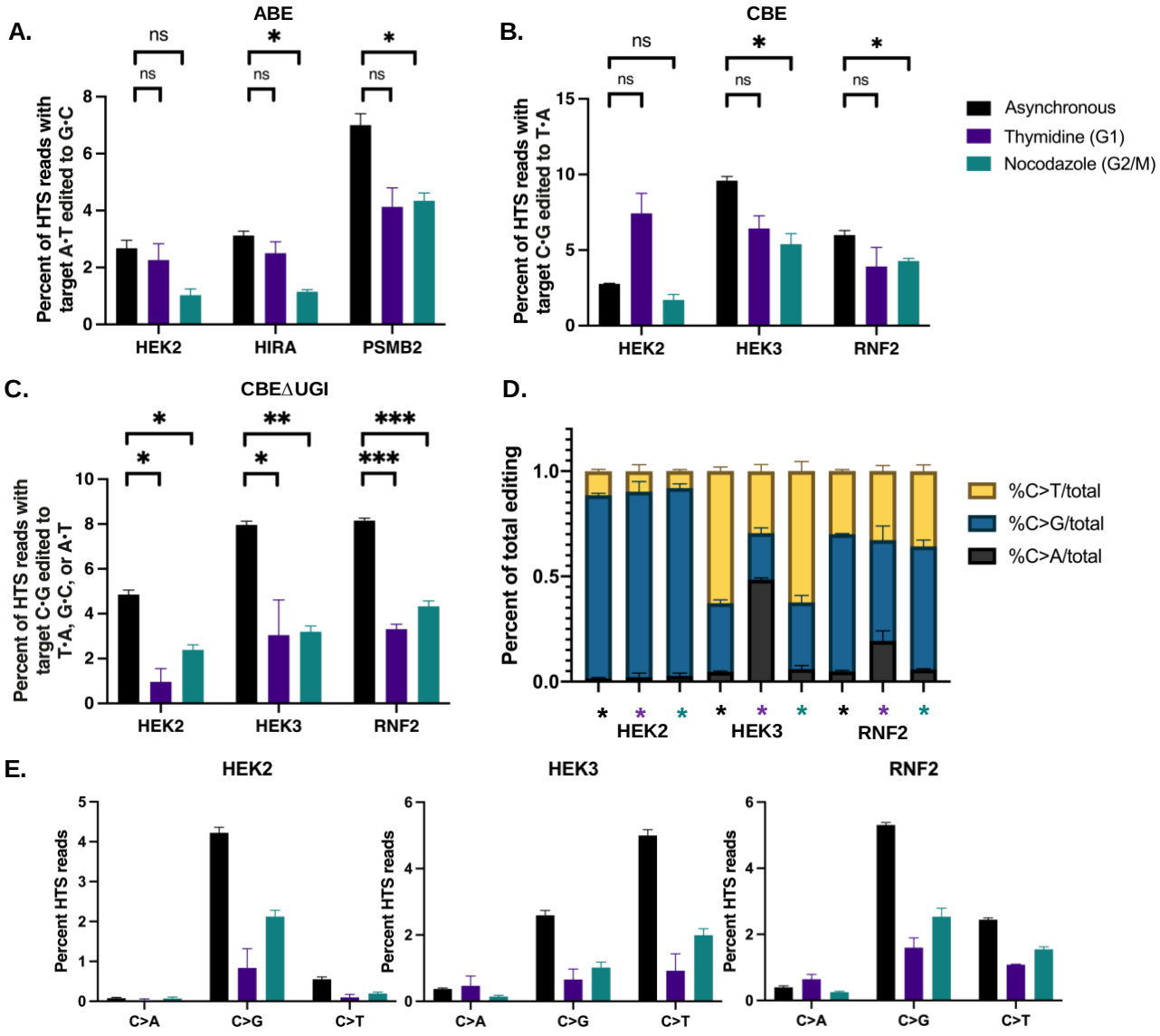

Figure 3.2: Cell cycle synchronization effects on base editing efficiencies and precision of Cas9n-derived BEs in K562 cells. Cells were transfected with ABE, CBE, or CBEΔUGI plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted, and target loci were amplified via PCR and subjected to next-generation sequencing. Genome editing efficiencies (percent of total HTS reads with the target A•T base converted to G•C for ABE, percent of total HTS reads with the target C•G base converted to T•A for CBE, or percent of total HTS reads with the target C•G base converted to T•A, G•C, or A•T for CBEΔUGI) were quantified with CRISPResso2. Base editing efficiencies by ABE (A), CBE (B), and CBEΔUGI (C) upon synchronization are plotted. (D) The product distribution, defined as the relative portion of edited sequencing reads (reads in which the target C•G is mutated to T•A, A•T, or G•C) that have been edited to each of the indicated outcomes, is plotted for CBEΔUGI. Absolute editing efficiencies (percent of total sequencing reads with target C•G mutated to T•A, A•T, or G•C) are plotted for CBEΔUGI at each target site.

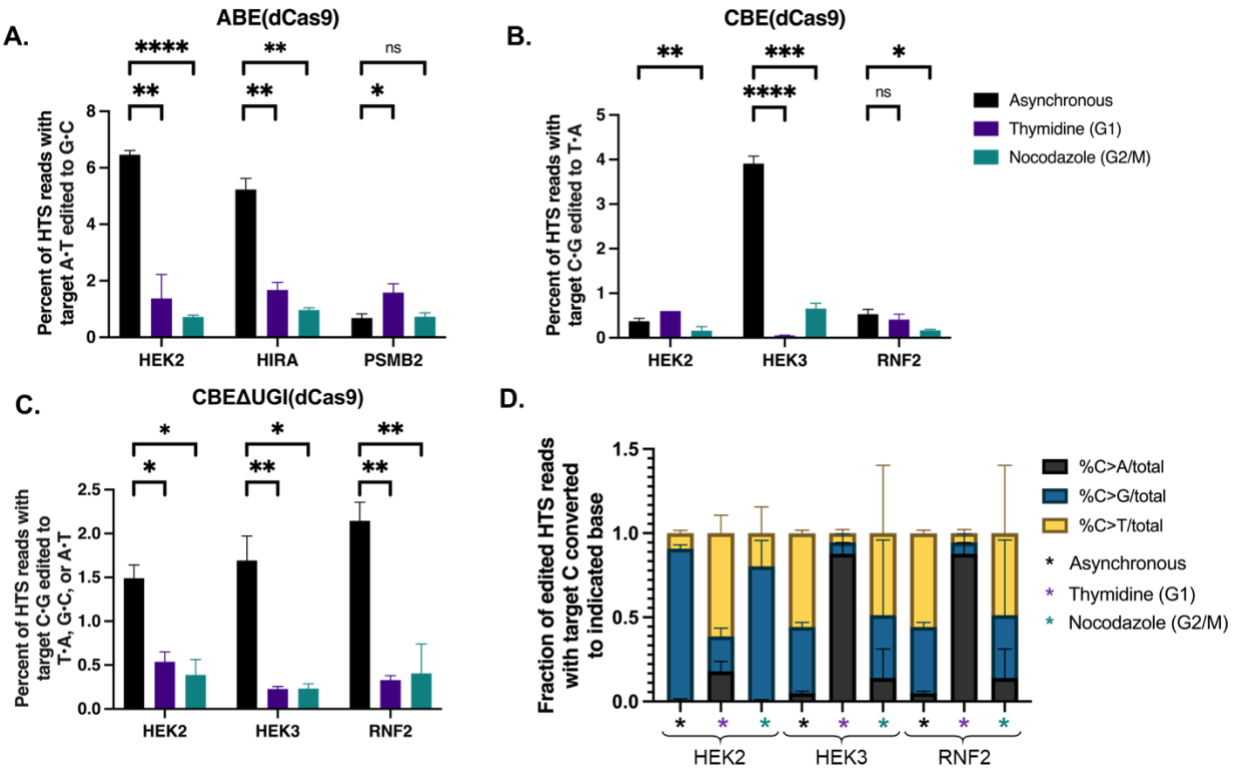

Figure 3.3: Cell cycle synchronization effects on base editing efficiencies by dCas9-derived BEs in K562 cells. Cells were transfected with ABE(dCas9), CBE(dCas9), or CBEΔUGI(dCas9) plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted and target loci were amplified via PCR and subjected to next-generation sequencing. Genome editing efficiencies (percent of total HTS reads with the target A•T base converted to G•C for ABE(dCas9), percent of total HTS reads with the target C•G base converted to T•A for CBE(dCas9), or percent of total HTS reads with the target C•G base converted to T•A, G•C, or A•T for CBEΔUGI) were quantified with CRISPResso2. Base editing efficiencies by ABE(dCas9) (A), CBE(dCas9) (B), and CBEΔUGI(dCas9) (C) upon synchronization are plotted. (D) The product distribution, defined as the relative portion of edited sequencing reads (reads in which the target C•G is mutated to T•A, A•T, or G•C) that have been edited to each of the indicated outcomes, is plotted for CBEΔUGI(dCas9). Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

Synchronization Effects on CBEΔUGI Product Purity and Editing Efficiency

Finally, we repeated these experiments with CBEΔUGI and CBE(dCas9)ΔUGI. Because these constructs lack the uracil glycosylase inhibitor component, excision of the uracil intermediate is quite efficient, resulting in high levels of C•G to non-T•A outcomes when using these constructs. This in turn allowed us to observe changes in base editing precision with respect to cell cycle synchronization. We analyzed overall editing efficiencies (percent of HTS reads with the target C•G edited to T•A, G•C, or A•T) and product distributions (the relative portion of edited sequencing reads in which the target C•G is edited to T•A, G•C, or A•T) of each of the samples treated with CBEΔUGI (Figure 3.4A-B). Consistent with previous studies[33], we observed high rates of C•G to G•C editing in asynchronous cells, particularly at the *HEK2* site, allowing for observation of changes in these relative efficiencies upon synchronization. To our surprise, cells synchronized in G1 using thymidine exhibited a significant increase in the relative fraction of edited reads with C•G to A•T mutations (which increased 19 ± 1.9 -fold at the *HEK2* site, 9 ± 0.8 -fold at the *HEK3* site, and 7 ± 0.4 -fold at the *RNF2* site, Figure 3.4B), with an accompanying decrease in relative C•G to G•C outcomes (which decreased by 1.6 ± 0.0 -fold at the *HEK2* site, 5 ± 0.10 -fold at the *HEK3* site, and 2.9 ± 0.0 -fold at the *RNF2* site, Figure 3.4B), and a minimal change in the relative fraction C•G to T•A edits (which was equivalent at the *HEK2* site, decreased by 1.9 ± 0.2 -fold at the *HEK3* site, and was equivalent at the *RNF2* site, Figure 3.4B). In addition to relative amounts, absolute C•G to A•T point mutation efficiencies also increased upon G1 arrest at all three sites (Figure 3.5C-E); absolute C•G to A•T efficiencies increased 17 ± 3-fold at the *HEK2* site, 6 ± 2-fold at the *HEK3* site, and 3 ± 1 -fold at the *RNF2* site). Upon G2/M synchronization, absolute C•G to A•T point mutation efficiencies were within error of asynchronous cells at the *HEK2* site, and slightly decreased at the *HEK3* and

RNF2 sites (1.7 ± 0.5 and 1.4 ± 0.3-fold reductions, respectively, Figure 3.5). Absolute C•G to G•C introduction efficiencies were highest in asynchronous cells, followed by G2/M synchronized (decreased by 36 ± 4% at the *HEK2* site, 56 ± 6% at the *HEK3* site, and 44 ± 7% at the *RNF2* site compared to asynchronous cells), and lowest in G1 synchronized cells (decreased by 46 ± 6% at the *HEK2* site, 84 ± 10% at the *HEK3* site, and 83 ± 10% at the *RNF2* site compared to asynchronous cells). These results were also observed in our K562 experiments (Figure 3.3), although the relative changes were less drastic, potentially due to the lower overall levels of editing in this cell line.

Editing percentages by CBEΔUGI(dCas9) decreased dramatically at all three sites under both synchronization conditions, consistent with the ABE(dCas9) and CBE(dCas9) results (editing decreased by 78 ± 7 % at the *HEK2* site, 78 ± 13 % at the *HEK3* site, and 80 ± 9 % at the *RNF2* site for cells synchronized in G1 by thymidine, and 79 ± 8% at the *HEK2* site, 85 ± 16% at the *HEK3* site, and 92 ± 11% at the *RNF2* site for cells synchronized in G2/M by nocodazole, Figure 3.6C). Even though overall editing efficiencies were below 5% at all three sites upon synchronization, we still observed the same trends in C•G to non-T•A editing outcomes upon synchronization (statistically significant increases in C•G to A•T introduction efficiencies upon G1 synchronization, and statistically significant decreases in C•G to G•C editing efficiencies upon G1 and G2/M synchronization, with higher absolute C•G to G•C editing efficiencies in G2/M synchronized cells compared to G1 synchronized cells, Figure 3.6D). Again, these experiments were repeated in K562 cells and comparable results were obtained (Figure 3.3D).

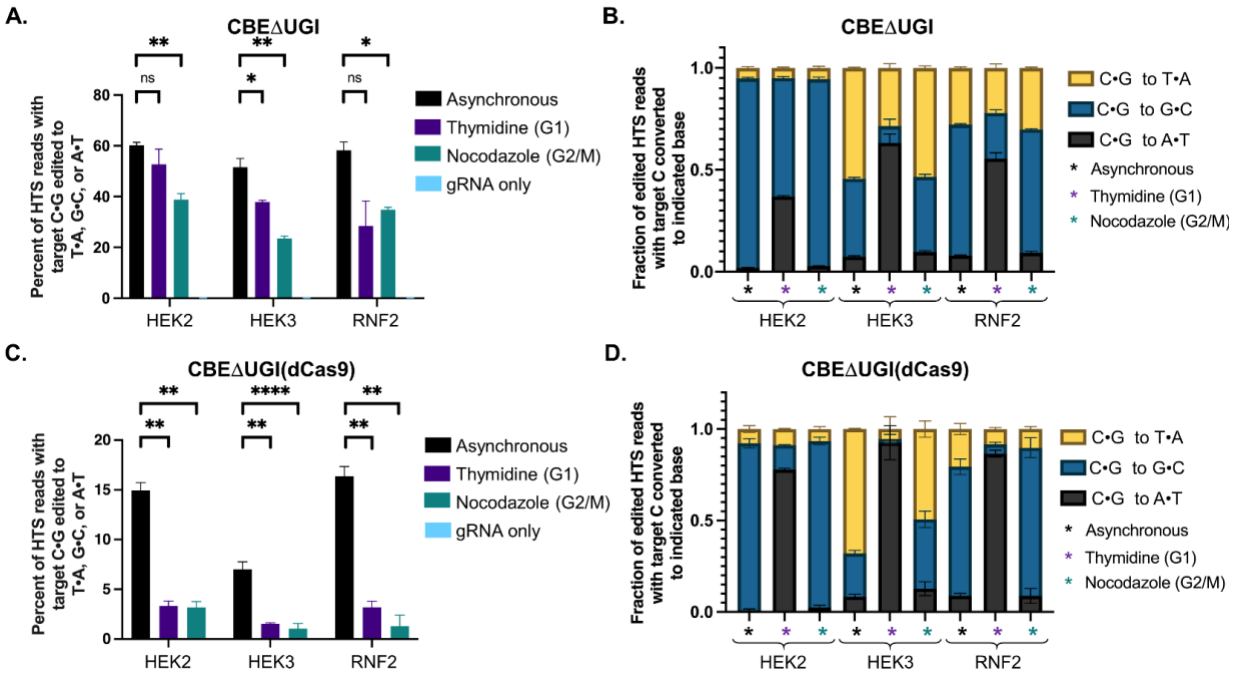

Figure 3.4: Cell cycle synchronization effects on base editing efficiencies of CBEΔUGI and CBEΔUGI(dCas9) in HEK293T cells. Cells were transfected with CBEΔUGI or CBEΔUGI(dCas9) plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted and target loci were amplified via PCR and subjected to HTS. Genome editing efficiencies (percent of total HTS reads with the target C•G base converted to T•A, G•C, or A•T) were quantified with CRISPResso2. Base editing efficiencies by CBEΔUGI (A) and CBEΔUGI(dCas9) (C) upon synchronization are plotted. (B and D) The product distribution, defined as the relative portion of edited sequencing reads (reads in which the target C•G is mutated to T•A, A•T, or G•C) that have been edited to each of the indicated outcomes, is plotted for CBEΔUGI (B) and CBEΔUGI(dCas9) (D). Values and error bars reflect the means and SD of three independent biological replicates performed on different days. Asterisks reflect p value calculations of unpaired *t* test, one tailed (ns indicates not significant, **p*<0.05, ***p*<0.01, ****p*<0.001, p****<0.0001).

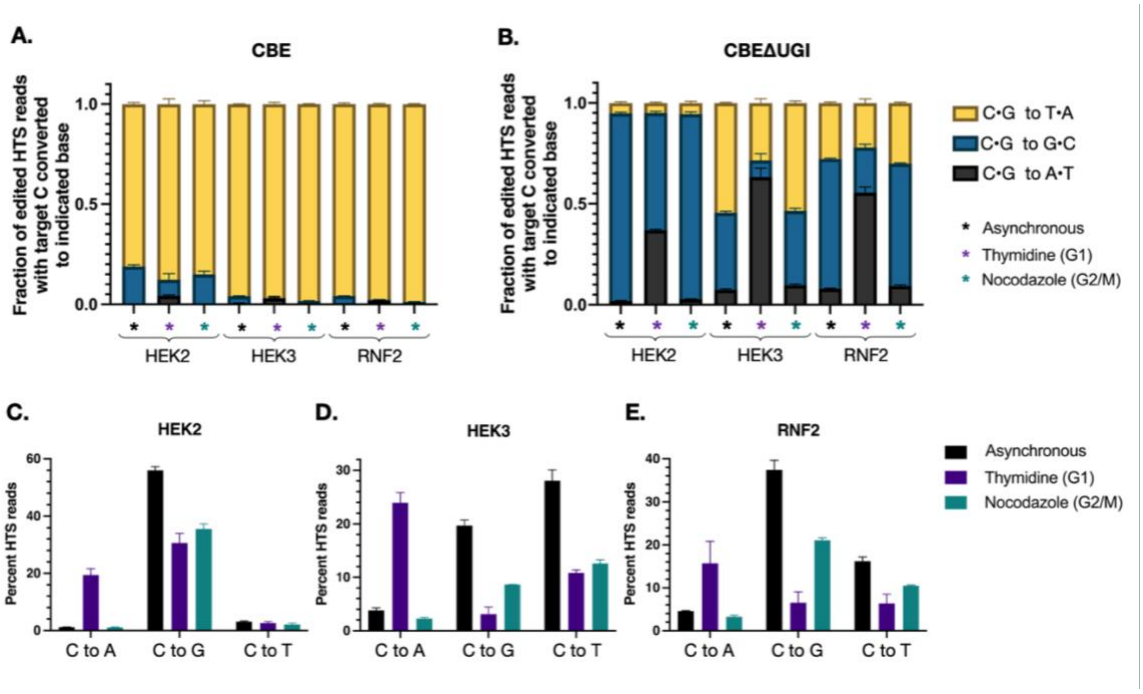

Figure 3.5: Product distribution of CBEs. Cells were transfected with CBE (A) or CBEΔUGI (B, C, D, E) plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted and target loci were amplified via PCR and subjected to next-generation sequencing. Genome editing efficiencies (percent of total sequencing reads with target C•G mutated to T•A, A•T, or G•C) were quantified with CRISPResso2. (A, B) The product distribution, defined as the relative portion of edited sequencing reads (reads in which the target C•G is mutated to T•A, A•T, or G•C) that have been edited to each of the indicated outcomes, is plotted for CBE (A) and CBEΔUGI (B). (C, D, E) Absolute editing efficiencies (percent of total sequencing reads with target C•G mutated to T•A, A•T, or G•C) are plotted for CBEΔUGI at each target site. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

To further confirm this phenomenon of drastic increases in C•G to A•T editing activity upon G1 synchronization, these experiments were repeated at three additional sites. At all three sites tested, there was a substantial increase in both the absolute and fractional C•G to A•T editing (Figure 3.7). These results indicate that the use of G1 synchronization agents can be used as a viable option for targeted C•G to A•T base editing. To control for changes that may be due to the chemical synchronization agent, we repeated these experiments using mimosine to synchronize the cells in G1 and observed the same increase in relative C•G to A•T rates (Figure 3.8). We will note that we observed drastic decreases in overall editing efficiencies by CBEΔUGI upon G1 synchronization with mimosine that were not observed upon G1 synchronization with thymidine. We attribute this to the differences in synchronization mechanisms by the two compounds; mimosine functions via chelation of iron, which many DNA repair proteins require for proper folding and function.

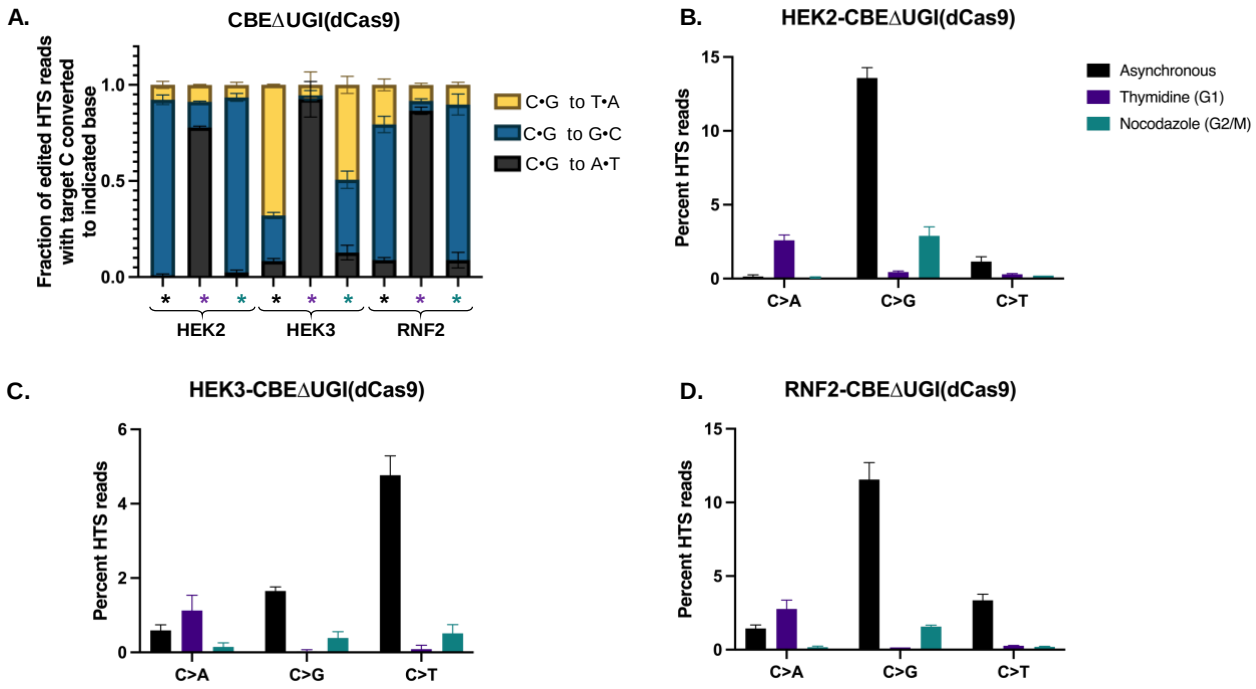

Figure 3.6: Cell cycle synchronization effects on base editing precision of dCas9-derived CBEs in HEK293T cells. Cells were transfected with CBEΔUGI(dCas9) plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted, and target loci were amplified via PCR and subjected to next-generation sequencing. Genome editing efficiencies (percent of total HTS reads with the target A•T base converted to G•C for ABE, percent of total HTS reads with the target C•G base converted to T•A for CBE, or percent of total HTS reads with the target C•G base converted to T•A, G•C, or A•T for CBEΔUGI) were quantified with CRISPResso2. (A) The product distribution, defined as the relative portion of edited sequencing reads (reads in which the target C•G is mutated to T•A, A•T, or G•C) that have been edited to each of the indicated outcomes, is plotted for CBEΔUGI. Absolute editing efficiencies (percent of total sequencing reads with target C•G mutated to T•A, A•T, or G•C) are plotted for CBEΔUGI at each target site (B,C,D). Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

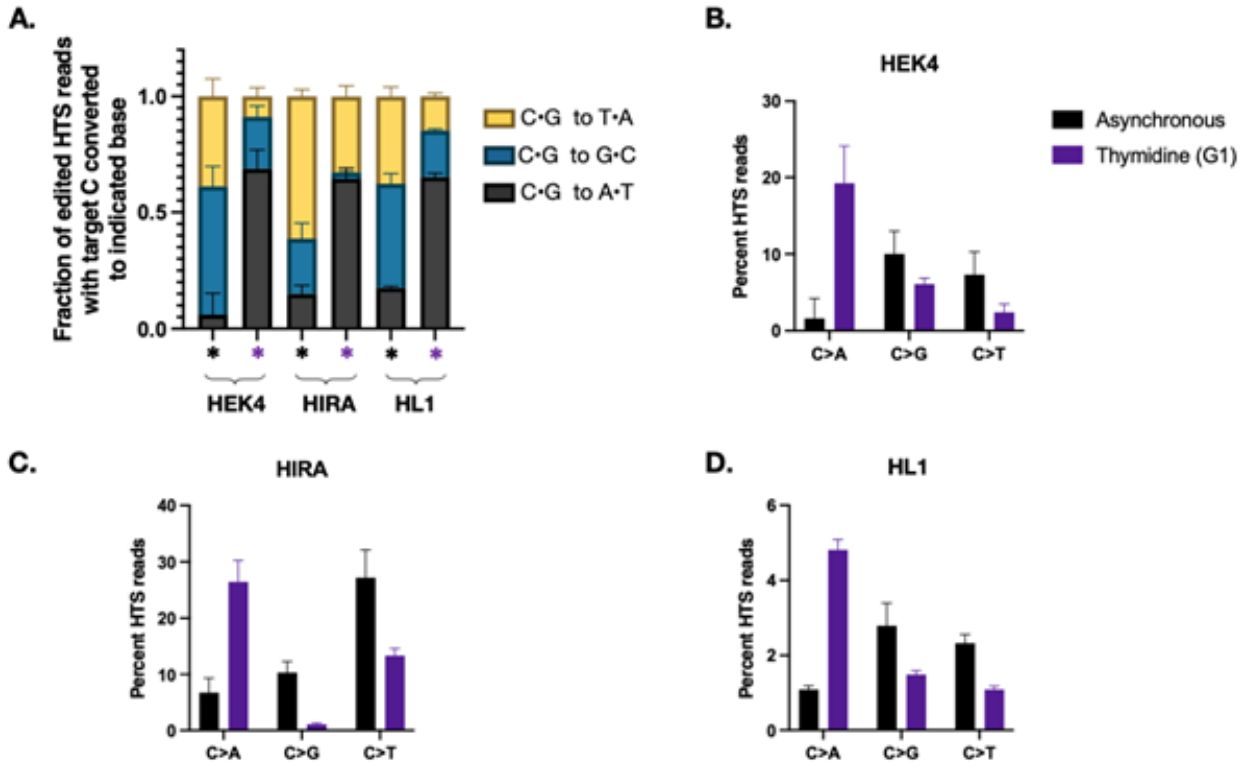

Figure 3.7: Product distribution of CBEΔUGI upon G1 synchronization at additional genomic loci. HEK293T cells were transfected CBEΔUGI plus gRNA (protospacer sequences indicated in Table 2), synchronization agent (thymidine) was added 6 hours post-transfection, and cells were lysed at 54 hours. The genomic DNA was extracted and target loci were amplified via PCR and subjected to HTS. Genome editing efficiencies (percent of total sequencing reads with target C•G mutated to T•A, A•T, or G•C) that have been edited to each of the indicated outcomes, is plotted for CBEΔUGI. (B,C,D) Absolute editing efficiencies (percent of total sequencing reads with target C•G mutated to T•A, A•T, or G•C) are plotted for CBEΔUGI at each target site.

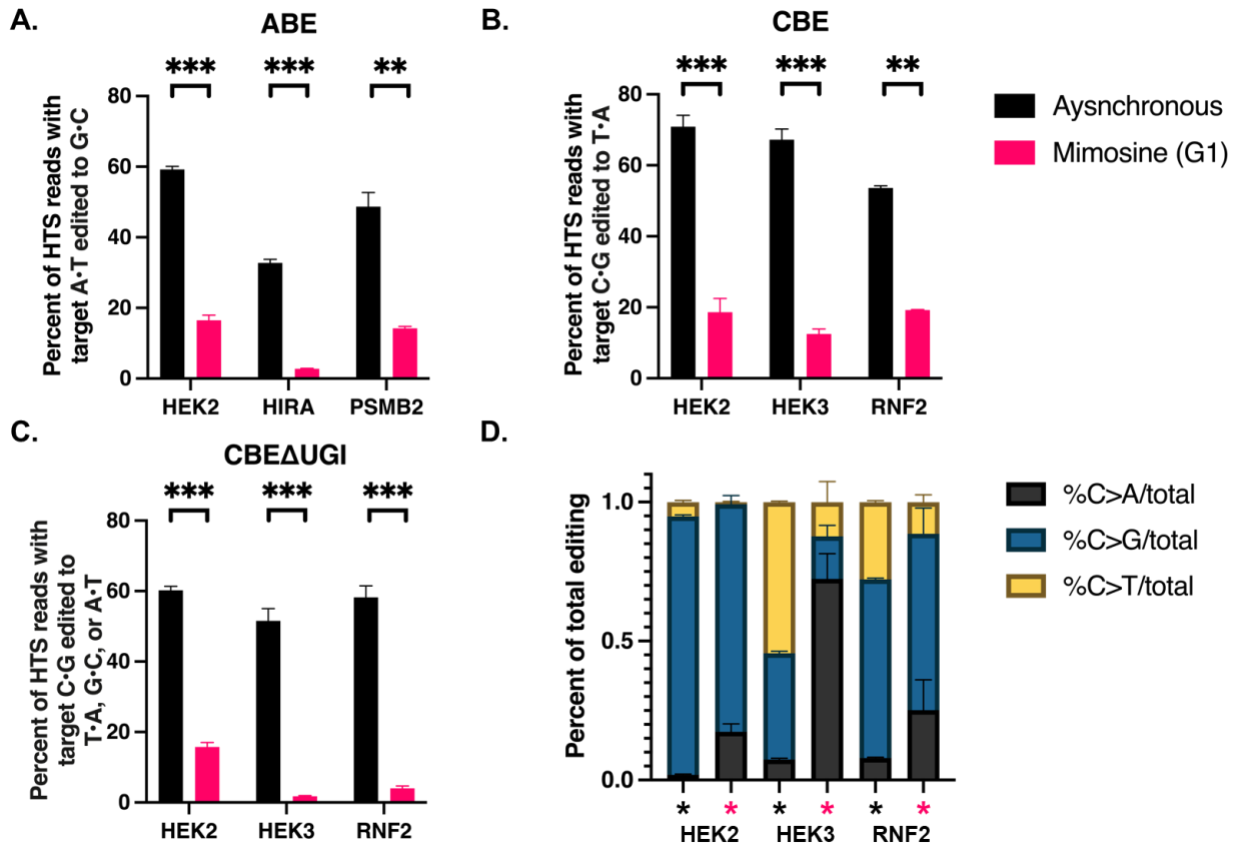

Figure 3.8: Effects on CBEΔUGI product distribution by G1 synchronization with mimosine. Cells were transfected with ABE, CBE, or CBEΔUGI plus gRNA (protospacer sequences indicated in Figure 1), mimosine was added 6 hours post-transfection (for G1/S synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted, and target loci were amplified via PCR and subjected to next-generation sequencing. Genome editing efficiencies (percent of total HTS reads with the target A•T base converted to G•C for ABE, percent of total HTS reads with the target C•G base converted to T•A for CBE, or percent of total HTS reads with the target C•G base converted to T•A, G•C, or A•T for CBEΔUGI) were quantified with CRISPResso2. Base editing efficiencies by ABE (A), CBE (B), and CBEΔUGI (C) upon synchronization are plotted. (D) The product distribution, defined as the relative portion of edited sequencing reads (reads in which the target C•G is mutated to T•A, A•T, or G•C) that have been edited to each of the indicated outcomes, is plotted for CBEΔUGI. Values and error bars reflect the means and SD of three independent biological replicates performed on different days.

Analysis of Changes in Indel Sequences and Introduction Efficiencies by Cas9n-Derived CBEs Following Synchronization

We additionally analyzed indel formation by all BEs in asynchronous and synchronized cells. Indel rates by ABE and dCas9-derived BEs were generally below 1% for all three sites at all conditions, consistent with previous reports (Figure 3.9A and C, and Figure 10A-D)[34]. Indel rates by CBE were higher than those by ABE, but still generally less than 1% except at the HEK2 site (Supplementary Figure 11A and C). However, indel rates by CBEΔUGI were on average 10 ± 3-fold higher than those by CBE at the exact same sites and under the same conditions, suggesting the involvement of UNG in CBE-induced indels (Figure 3.9A-B). Additionally, indel rates by CBEΔUGI were between 11- and 100-fold higher than those by CBEΔUGI(dCas9) at the exact same sites and under the same conditions (Figure 3.9B-C), suggesting the involvement of nicking of the unedited strand in CBE-induced indels as well. Due to the elevated rates of indel formation by CBEΔUGI, we focused our additional analyses on indels introduced by this construct.

We observed no consistent changes in absolute indel rates by CBEΔUGI upon synchronization with either agent (upon synchronization in G1 by thymidine, absolute indel rates by CBEΔUGI increased 2.4 ± 0.5-fold at the *HEK2* site, but did not change at the *HEK3* site or *RNF2* site, and upon synchronization in G2/M by nocodazole, absolute indel rates by CBEΔUGI were equivalent at the *HEK2* site, reduced by 1.4 ± 0.1 -fold at the *HEK3* site, and increased by 1.4 ± 0.1 -fold at the *RNF2* site, Figure 3.9B). This may be indicative of differences in DNA repair of non-coding versus coding regions of the genome, particularly with respect to their accessibility to glycosylases. We analyzed individual indel sequences and found that, among the most common sequences (those that represent >0.1% of total reads in two out of three of the

asynchronous, G1-synchronized, and G2/M synchronized samples, displayed in Figure 3.9D-F), deletion sequences were confined to the region between the deaminated target cytosine(s) and the location of the Cas9n-induced nick at all three sites. We observed no insertion sequences among the most common indel sequences. This is in direct contrast to DSB-mediated indels, which are centered around the Cas9 cut site. Interestingly, upon synchronization in either G1 or G2/M, the relative amounts of each indel sequence changed drastically (Figure 3.9D-F), indicating that certain indel sequences are preferentially produced during different phases of the cell cycle. Taken together, these observations suggest an inherently different mechanism of indel introduction by CBEs compared to DSB-reliant technologies.

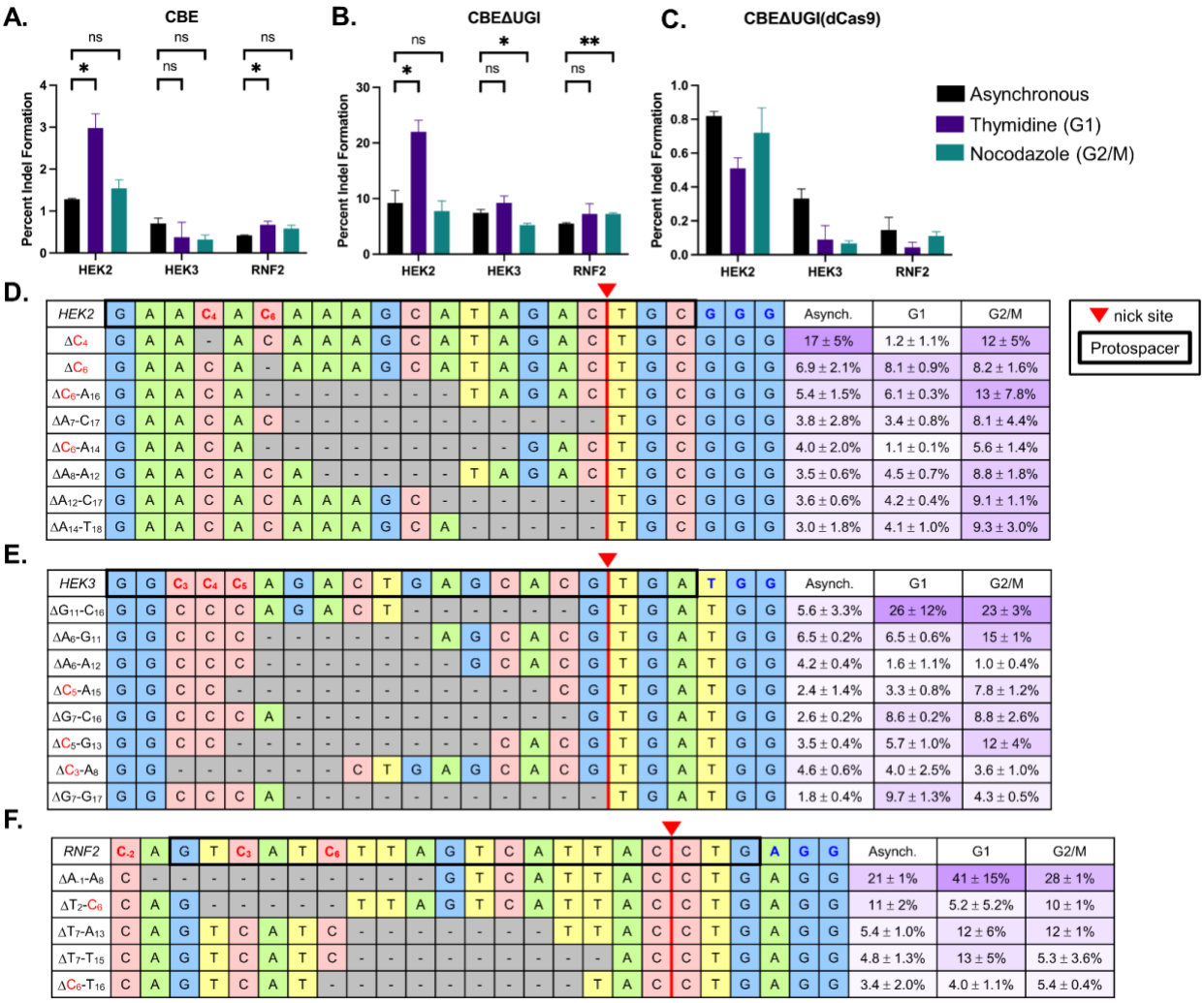

Figure 3.9: Indel analysis of CBEs in HEK293T cells. HEK293T cells were transfected with CBE (A), CBEΔUGI (B), or CBEΔUGI(dCas9) (C) plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted and target loci were amplified via PCR and subjected to HTS. Total indel introduction efficiencies for CBE (A), CBEΔUGI (B), and CBEΔUGI(dCas9) (C) were calculated as the percent of reads with insertions or deletions (determined via CRISPResso analysis) divided by the total number of HTS reads sequenced. (A-C) Effects of synchronization in G1 or G2/M on indel introduction efficiencies. (D, E, and F) The most common (defined as sequences that comprise greater than 0.1% of total reads in at least two out of three of the asynchronous, G1-synchronized, and G2/M-synchronized samples) indel sequences are shown with respect to the protospacer (bold outline), potential edited cytosines (indicated in red), nick site (red triangle), and PAM (indicated in blue) for the *HEK2* (C), *HEK3* (D), and *RNF2* (E) sites. The relative portion of total indel reads with each specific indel sequence is listed on the left in purple with respect to synchronization condition (note these are not absolute indel introduction efficiencies). Values and error bars reflect the means and SD of three independent biological replicates performed on different days. Asterisks reflect p value calculations of unpaired *t* test, one tailed (ns indicates not significant, **p*<0.05, ***p*<0.01, ****p*<0.001).

We additionally treated HEK293T cells with BEs optimized for reduced RNA off-target activity[28,35] (BE4-W90Y-R126E, BE4-W90Y-R126E-ΔUGI, and ABE7.10-F148A, referred to as CBE-YE1, CBEΔUGI-YE1, and ABE-F148, respectively, Figure 3.10E) as well as their catalytically inactivated deaminase counterparts (referred to as CBE-E63A-YE1[36], CBEΔUGI-E63A-YE1, and ABE-E59A-F148A[37], respectively,) and analyzed their rates of indel formation when targeting the *HEK2* site. We again observed a 12 ± 3-fold increase in indel rates in CBEΔUGI-YE1 treated cells compared to CBE-YE1 treated cells (Figure 3.10G-H). When comparing each catalytically active CBE variant to its respective catalytically inactive counterpart, we observed a 3 ± 1.8-fold decrease in indel introduction efficiency for CBE-YE1 (Figure 3.10G), and a 21 ± 5-fold decrease in indel introduction efficiency for CBEΔUGI-YE1 (Figure 3.10H), suggesting that nicking alone is insufficient for indel formation, but requires uracil introduction as well.

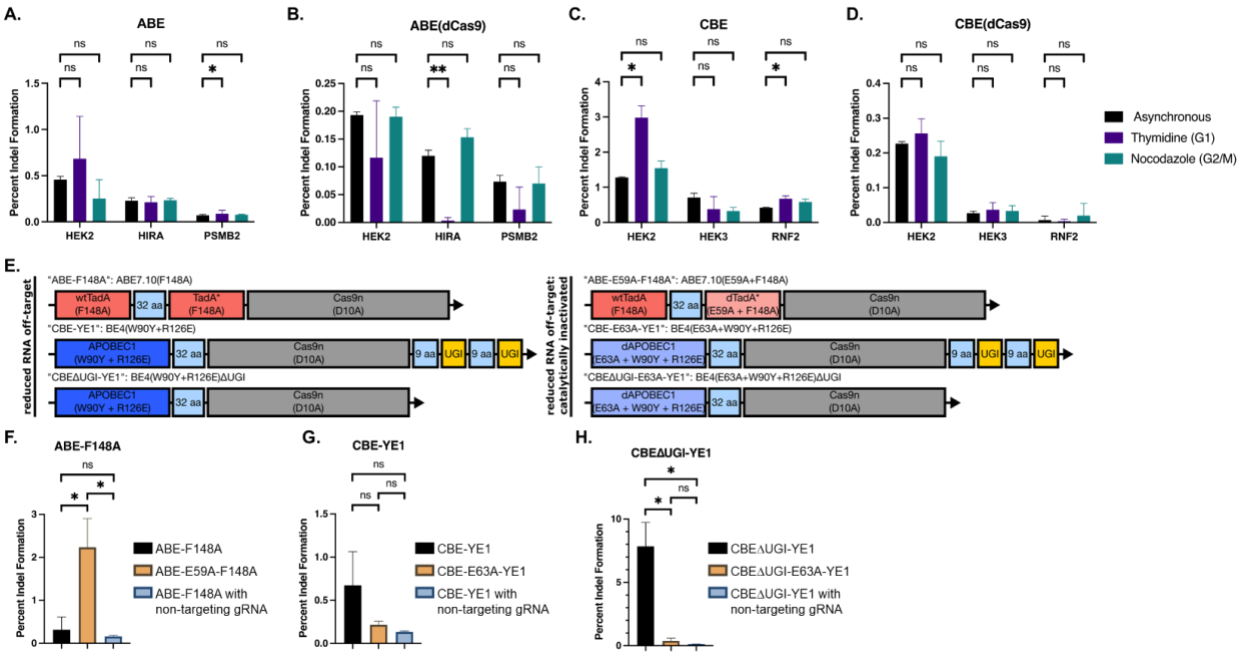

Figure 3.10: Cell cycle synchronization effects on indel introduction efficiencies in HEK293T cells. Effects of cell cycle synchronization and catalytic inactivation of the deaminase on indel introduction efficiencies in HEK293T cells. In (A) through (D), HEK293T cells were transfected with ABE (A), ABE(dCas9) (B), CBE (C), or CBE(dCas9) (D) plus gRNA (protospacer sequences indicated in Figure 1), synchronization agents were added 6 hours post-transfection (thymidine for G1 synchronization or nocodazole for G2/M synchronization), and cells were lysed at 54 hours. The genomic DNA was extracted and target loci were amplified via PCR and subjected to HTS. Total indel introduction efficiencies were calculated as the percent of reads with insertions or deletions (determined via CRISPResso analysis) divided by the total number of HTS reads sequenced. The data in (C) is also presented in Figure 3.9A for comparison purposes. (E) Construct maps and names of reduced-RNA editing base editor variants. In (F) through (H), HEK293T cells were transfected with the reduced-RNA editing variants ABE-F148A (F), CBE-YE1 (G), CBEΔUGI-YE1 (H), or their catalytically inactivated deaminase counterparts ABE-E59A-F148A (F), CBE-E63A-YE1 (G), CBEΔUGI-E63A-YE1 (H) and either a HEK2-targeting gRNA or a non-targeting gRNA. Cells were lysed 48 hours after transfection, the genomic DNA was extracted, and target loci were amplified via PCR and subjected to HTS. Total indel introduction efficiencies were calculated as the percent of reads with insertions or deletions (determined via CRISPResso analysis) divided by the total number of HTS reads sequenced. Values and error bars reflect the means and SD of three independent biological replicates performed on different days. Asterisks reflect p value calculations of unpaired *t* test, one tailed (ns indicates not significant, *p<0.05, **p<0.01).

Discussion

We report here the first systematic study of the cell cycle dependence of both adenine and cytosine BEs. Notably, we observed drastic differences in the mechanism by which Cas9n-derived BEs and dCas9-derived BEs function. Cas9n-derived BEs (which are the most commonly used BE variants) display minimal changes in overall point mutation introduction efficiencies upon synchronization in G1, with small (less than 25% for CBE, and less than 45% for ABE) decreases in efficiency upon synchronization in G2/M. The dCas9-derived BEs both exhibited drastic reductions in their respective point mutation introduction efficiencies upon both G1 (~65% reductions for ABE and greater than 70% reductions for CBE) and G2/M (~70% reductions for ABE and ~80% reductions for CBE) synchronization. These data demonstrate that Cas9n-derived BEs rely on more ubiquitous DNA repair pathways than both dCas9-derived BEs and DSB-reliant technologies. The observation that minimal decreases in both CBE and ABE editing efficiencies occur upon synchronization in G1 is particularly noteworthy, as this is strong mechanistic confirmation that BEs can function well in nondividing cells. The significant decrease in editing efficiencies by dCas9-derived BEs upon both G1 and G2/M synchronization suggests that these tools' intermediates are highly dependent on S-phase processes to be converted to desired editing outcomes. We suggest that these tools may rely heavily on DNA synthesis across their respective base intermediates to install point mutations.

Interestingly, we observed drastic increases in C•G to A•T editing efficiencies by CBEΔUGI upon G1 synchronization across multiple cell lines and using different G1 synchronization agents. C•G to non-T•A editing by CBEs has been hypothesized to occur from translesion synthesis (TLS) polymerases processing abasic sites generated from uracil excision by UNG. Our observations may be caused by a combination of cell cycle-dependent changes in

UNG and TLS polymerase expression levels, but additional experiments are necessary to further probe this discovery. While the general CBE architecture has recently been repurposed for precision C•G to G•C base editing through DNA repair factor manipulation[38], strategies for precision C•G to A•T base editing in mammalian cells do not currently exist. The use of G1 synchronization agents can be used as a starting point to generate more specialized and precise DNA repair manipulation strategies to generate mammalian cell C•G to A•T base editors.

We additionally performed a mechanistic study on base editor-induced indels. An examination of both indel introduction rates and indel sequences introduced by CBE variants show that the introduction of indels by CBEs is dependent on UNG, DNA nicking, and catalytically active deaminase. Specifically, removal of UGI, mutation of H840A in Cas9 (which converts Cas9n to dCas9), or mutation of E58A in rAPOBEC1 (which catalytically inactivates the deaminase) are each independently sufficient to reduce indel introduction rates 10-fold. An analysis of the indel sequences was consistent with these observations as well; CBE-induced indels were found to all be deletion sequences (no insertions were observed, in direct contrast with DSB-reliant genome editing tools). The deletion sequences were either deletions of a single base (a target cytosine) or deletions that spanned the region between the nick and a deaminated cytosine. Taken together, these data suggest CBE-induced indel sequences are likely caused by in situ-generated staggered DSBs, which are putatively formed following processing of UNG-generated abasic sites by endonucleases such as APEX1/2 (which cleave the DNA backbone at abasic sites).

In summary, we have conducted here one of the first mechanistic studies of base editors. We have quantified changes in editing efficiency and precision of both adenine and cytosine base editors with respect to cell cycle synchronization and thus provide key insights into the DNA

processing mechanisms of base editor intermediates. Changes in base editing efficiency with respect to cell cycle synchronization suggest nicking BEs rely on more ubiquitous DNA repair pathways than DSB-reliant technologies to introduce their respective point mutations, while non-nicking BEs are highly dependent on S-phase. These results in turn will guide future strategies to enhance base editing efficiency and/or precision and provide more mechanistic details regarding the robustness of nontraditional genome editing agents.

Methods

Constructs and Molecular Cloning

All BE plasmids were constructed with USER cloning[29] with pCMV ABEmax_P2A_GFP (Addgene #112101) and pCMV_AncBE4max_P2A_GFP (Addgene #112100) plasmids as template, using Phusion U Hot Start Polymerase (ThermoFisher Scientific). All sgRNA expression plasmids were generated using blunt-end cloning with pFYF1230 (Addgene plasmid #47511) as a template, using Phusion High-Fidelity DNA Polymerase (New England BioLabs). All DNA vector amplification was carried out using NEB 10-β competent cells (New England BioLabs). All plasmids were purified using the ZymoPURE II Plasmid Midiprep Kit (Zymo Research D4200).

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

For all HEK293T cell transfections, 100,000 HEK293T cells in 250 µL of DMEM media without Penicillin-Streptomycin were added per well to 48-well VWR Multiwell Cell Culture Plates on top of lipofectamine/plasmid mixtures. For all K562 cell transfections, 50,000 K562 cells in 250 µL RPMI media without Penicillin-Streptomycin were added per well to 48-well VWR Multiwell Cell Culture Plates on top of lipofectamine/plasmid mixtures transfected at a density 50,000 cells per well in 250 µL RPMI media without Penicillin-Streptomycin. The lipofectamine/plasmid mixtures consisted of 1000 ng of BE plasmid, 250 ng of sgRNA plasmid, and 1.5 µL of Lipofectamine 2000 (ThermoFisher Scientific) in 25 µL of total volume, made up with Opti-MEM (Gibco #31985-070). Chemical inhibitors were added 6 hours after transfection from stock solutions (described below) to result in final concentrations of 5 mM (Thymidine), 800 µM (Mimosine), or 200 ng/mL (Nocodazole).

Preparation of Synchronizing Agents

Nocodazole (Sigma) was prepared in DMSO to a stock solution concentration of 20 mg/mL. This stock solution was diluted to 50 µg/mL immediately prior to addition to the cells, and 1.1 µL of this diluted stock solution was added to the 275 µL of media in each well, for a final concentration of 200 ng/mL.

Thymidine (Sigma) was prepared in 1X PBS to a stock solution concentration of 50 mM. 30 µL of this stock solution was added to the 275 µL of media in each well, for a final concentration of 5 mM.

Mimosine (Sigma) was prepared in 1X PBS to a stock solution concentration of 10 mM. 24 µL of this stock solution was added to the 275 µL of media in each well, for a final concentration of 800 µM.

Lovastatin (Sigma) was prepared in 95% ethanol to a stock solution concentration of 70 mM with a pH of 7.5. This was diluted to 400 µM, then 30.5 µL of this stock solution was added to the 275 µL of media for a final concentration of 40 µM.

Flow Cytometry Analysis of Cell Synchronization

3 x 10^5 HEK293T or K562 cells were plated in a T25 flask in 5 mL of media and synchronizing agents were added to final concentrations as indicated above for 6, 12, or 18 hours. Cells were washed with 10 mL PBS, detached with 2 mL of TrypLE (HEK293T cells only), and collected by centrifugation for 10 minutes at 400g. Cells were resuspended at 1 x 10^6 cells/mL in 1 mL cold PBS, added to 9 mL cold 70% ethanol, and stored for at least 4 hours at -20º C. After ethanol fixation, cells were centrifuged at 400g, washed with 10 mL cold PBS, then stained with 400 µL PI solution (0.1% Triton X-100 [Sigma], 0.2 mg/mL RNAse [Sigma], 0.02 mg/mL PI [Sigma] in PBS). Cells were incubated at 37º C for 15 minutes prior to analysis. Cells were gated to exclude doublets and non-viable cells. Fluorescent signal from PI staining was analyzed via histogram on either a BLDSRFortessa or BioRad S3e cell sorter.

Flow Cytometry Analysis of GFP Fluorescence

For all GFP fluorescence measurements, 1 x 10^6 cells were resuspended in FACS buffer (1% FBS, 50 µM EDTA pH 8.0, 2 µg/mL PI [Sigma]) and filtered through a cell-strainer. Non-viable cells and doublets were eliminated via gating parameters. Flow cytometry was performed on a BioFortessa or S3e cell sorter (Bio-Rad).

Transfection Efficiency Quantification

Transfection efficiency was determined via flow cytometry analysis of cells 24 hours post-transfection. Chemical inhibitors were added to HEK293T cells at -17, 0 and 6 hours relative to transfection of BE and gRNA plasmids (as described above). 24 hours post-transfection, cells were washed with PBS, detached from the plate with 50 µL Accutase (Innovative Cell Technologies), resuspended in 250 µL FACS buffer, and analyzed by flow cytometry as described above. The percent of cells with GFP fluorescence was analyzed via Flowjo.

High-Throughput DNA Sequencing (HTS) of Genomic DNA

Transfected cells were rinsed with 150 µL PBS (ThermoFisher Scientific) per well at the indicated time points 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 bearing homology to the target site and relevant Illumina forward and reverse adapters, (1 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 from 2% agarose gel with QIAquick Gel Extraction Kit (Qiagen) and quantified using NEBNext Ultra II DNA Library Prep Kit (NEB) on a CFX96 system (BioRad). Samples were then sequenced on an Illumina MiniSeq according to the manufacturer's protocol.

HTS and Indel Analysis of Targeted Amplicon Sequencing Reads

Analysis of Illumina HTS sequencing readout was conducted with CRISPRessov2[18,39]. Specifically, for these analyses, fastq files were analyzed via scripts run on Docker, where the reads were analyzed against the entire amplicons, with outputs for the guide RNA and base editor (--guide_seq and –base_editor_output). Product distribution for CBE variants was determined by taking the fraction of individual A•T, G•C, and T•A reads and dividing by the sum. CRISPResso was also used to validate editing percentages and analyze indel frequency, where the total number of indel reads was obtained from the indel histogram output and expressed as the fraction of reads with indel over total reads. For analysis of indel sequences, specific reads constituting >4% of the total indels from the CRISPResso were compiled.

Chapter 3 has been adapted from: **Burnett CA**, Wong A, Vasquez C, McHugh CA, Yeo GW, and Komor AC. Examination of the Cell Cycle Dependence of Cytosine and Adenosine Base Editors. Frontiers in Genome Editing (2022). The dissertation author was the primary author of this paper.

CHAPTER 4

Introduction

The ability to achieve high rates of C•G to A•T editing under G1/S synchronization is unprecedented in mammalian base editor tool development. Furthermore, consistent with all previous literature and our data, the activity of UNG to induce abasic sites at the target cytosines is indispensable for C•G to G•C and C•G to A•T editing outcomes. Downregulating UNG via knockout, knockdown, or inhibitor protein fusion results in dominant C•G to T•A repair outcomes[33,40]. As previously mentioned in Chapter 3, CBE architecture has been repurposed for precision C•G to G•C base editing through DNA repair factor manipulation. In bacterial cells, the nCas9-deaminase-Ung construct induces C•G to A•T edits[41]. However, when transfected into mammalian cells this base editor generates C•G to G•C conversions, and constitutes the CGBE developed by the Liu lab[42]. Additionally, there is a reported C-to-G base editor with nCas9 fused to cytidine deaminase and rXRCC1[38]. For base editing experiments to explore "undesired" editing outcomes, BE4maxΔUGI is employed in the experiments investigating C•G to A•T conversion.

Thymidine synchronization leads to robust C•G to A•T editing activity with CBEΔUGI at 6 different target sites, across multiple cell lines, and with both nCas9 and dCas9 derived BEs (dCas9showing lower overall editing, but higher relative rates of C•G to A•T outcomes). Furthermore, the alternate G1 synchronizing agent, mimosine, also demonstrated increased relative C•G to A•T editing (Figure 3.). Thymidine synchronization is believed to function via inhibition of ribonuclease reductase (RNR)[43]. As thymidine is converted to dTTP, changes in nucleotide pool concentrations impede DNA synthesis via a negative feedback loop.

We hypothesized that synchronization in G1 phase leads upregulation of DNA repair pathways that convert uracil intermediates into adenine. We argue that it is independent of nucleotide pool composition, based on consistency across different chemical inhibitors. We created a base editor with a G1 synchronizing effector to temporally restrict enzyme activity to leverage into a C•G to A•T editor. We chose a protein that is degraded in S/G2 phase, the licensing factor Cdt1[44]. Cdt1 is degraded by ubiquitin-mediated proteolysis[45,46]. The fusion of Cdt1 to Cas9 has previously been used to restrict protein activity to G1 phase[47]. These constructs are shown to be degraded in S and G2 phases of the cell cycle. We cloned the N-terminal fragment (which possesses the ubiquitylation domain necessary for degradation) into our base editor constructs and transfected HEK293T cells with CBEΔUGI-Cdt1. There were no significant alterations to base editing outcomes with the Cdt1 construct, ruling out its possible use as a targeted C•G to A•T editor.

We utilized knockdown of certain DNA repair genes to identify which is responsible for C•G to A•T editing outcomes. We developed a fluorescent assay where repair of the target cytosine specifically to adenine results in GFP-turn on. Mutant mimics of C>G and C>T outcomes lacked fluorescence, confirming activity was exclusive to C•G to A•T conversion. Using this screen, we tested several families of DNA repair pathways, and identified TLS polymerases POLI and POLK as implicated in this unique repair. We finally sought to identify if nucleotide concentrations could be responsible alteration of editing outcomes by addition of dATP, dCTP, dGTP, and dTTP to the cell culture media. Given the lack of change to overall, editing outcome we conclude it is reliance on translesion synthesis, rather than nucleotide availability, that dictates C•G to A•T editing.

Results

Fluorescent Screens for Analyzing Uracil Repair Outcomes

We developed screens to identify proteins involved in determining uracil repair outcomes. First, we optimized selection systems that link the processing of U•G lesions to a fluorescent assay. Our selections are plasmids generated via restriction enzyme cloning with synthetic DNA inserts. Ligation of our custom oligonucleotides into the digested backbone yields a plasmid with a site-specifically incorporated U•G lesion. We designed three selections with U•G lesions incorporated at the mCherry gene such that transcription of the template DNA strand cannot encode a fluorescent protein.

Table 4.1: Fluorescent selection scheme of potential U:G repair outcomes.

Selection	Mutated Residue	Uracil Incorporation	Possible Outcomes
U•G to C•G	Chromophore Residue M71	coding strand ATG template strand TAU Met	ATG TGC Met
			ATC TGG Iso
			ATT TAA Iso
U•G to G•C	Chromophore Residue M71	coding strand ATU template strand TAG Iso	ATG TGC Met
			TGC ACG Iso
			CGC GCG Iso
U•G to T•A	Chromophore Residue M71	coding strand AUG template strand TGC Thr	ATG TAC Met
			ACG TGC Thr
			AGG TCC Arg

One selection implicates genes involved in repair of U:G to the original C:G base pair, as anticipated by native base excision repair excising the uracil to restoring the original DNA sequence. Repair of U•G to only the outcome of interest results in the translation of functional protein. This will distinguish hits for general BER-involved genes from the desired mechanism, which will be identified by the second selection (Table 1). The third selection will identify genes relevant to current base editors, by examining genes responsible for failure to resolve the anticipated base editing outcome, T:A. We cloned plasmids harboring a DNA lesions in a critical fluorescent residue of mCherry. While mutant outcomes proved non-fluorescent, the project was ultimately limited by key bottlenecks in efficiency. T4 DNA ligase is intended to be coupled with antibiotic selection and transformation into bacterial cells, where inefficiency is masked by selective pressure. Based on observations of ligation reactions, it is not sufficiently robust to be used for such purposes. Secondly, it is difficult to distinguish between partial linear ligation from complete circular ligation. Efforts to purify and transfect such ligation mixtures would still display GFP transfection marker, as mammalian cells can readily transcribe linear DNA construct. Third, mutagenesis rates of uracil bearing plasmids into mammalian cells are very low (0.3% Figure 4.1C), limiting interpretation signal to noise ratios. Extensive optimization could potentially bypass these bottlenecks, but we focused on pursuing the mechanisms of results detailed in Chapter 3.

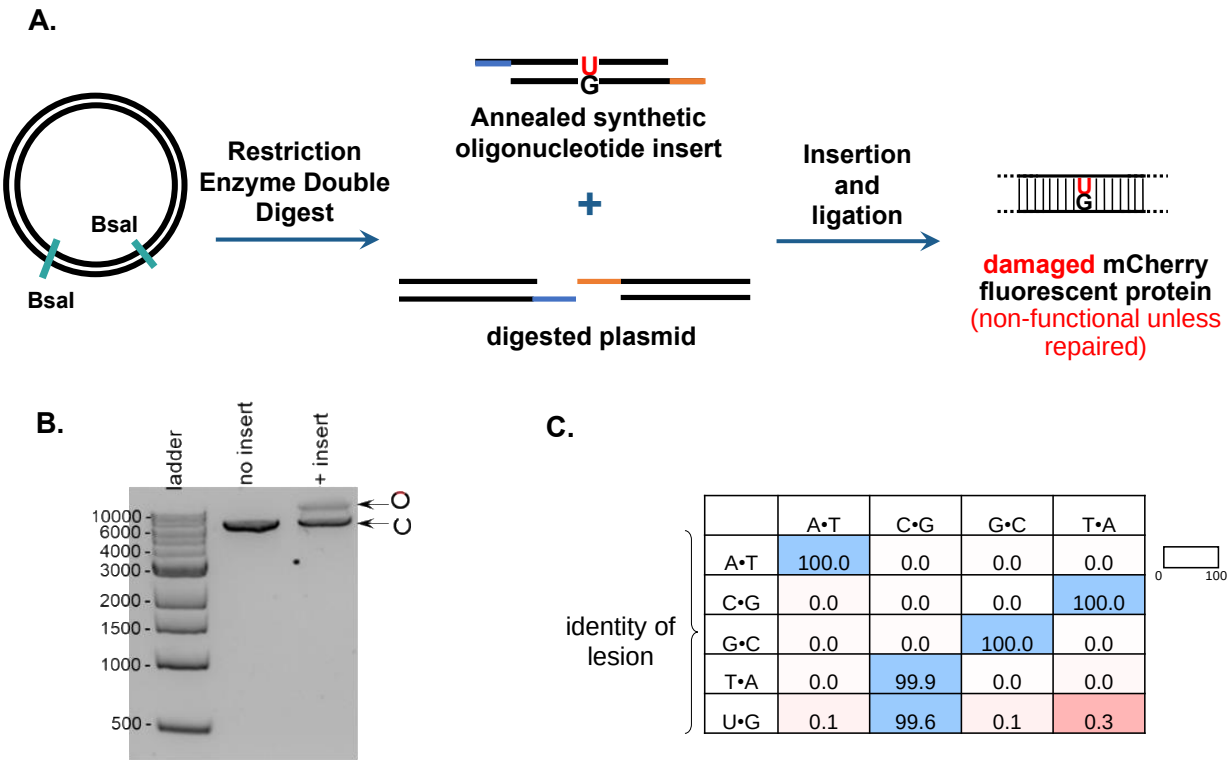

Figure 4.1: Ligation selection scheme and mutagenicity. (A) Design of ligation strategy for site-specific installation of uracil within mCherry gene. Plasmids bearing dual Bsa1 recognition sites are digested, purified, and ligated together with synthetic oligonucleotides with homology to the RE generated overhangs. (B) Agarose gel showing efficacy of oligonucleotide insertion. From left to right: DNA ladder, digested plasmid without inserts, fully ligated plasmid containing the annealed synthetic oligonucleotide of interest. (C) Mutagenesis rate of uracil in the context of ligation assay. The left column displays the identity of individual base pairs of the annealed oligonucleotide. Shown in the table is the distribution of the HTS readout.

C•G to A•T Fluorescent Screen Assay

Exploring the results of synchronization experiments, we sought to identify genes involved in C•G to A•T repair outcomes. We designed a GFP-based fluorescent screen to detect C•G to A•T editing events by CBEΔUGI, and identified the chromophore residue Phe66 for our selection, due to the presence of a targetable adenine in its codon. By installing a cytosine within this codon, we placed it within the target window of the protospacer for CBE editing. This mutant GFP construct will be referred to as dGFPY66S. We needed to

confirm this Y66S mutant, and the amino acids resultant from C>G (Y66F) or C>T (Y66C) editing, would be non-fluorescent. We cloned Y66S, Y66F, and Y66C GFP mutants and transfected them into HEK293T cells. At 24 hours post-transfection, only the plasmid with the original Y66 residue was GFP+, consistent with selectivity based on DNA repair outcome. Next, we tested if CBEΔUGI editing could restore fluorescence to the dGFP gene via C•G to A•T conversion. We transfected a BE4ΔUGI construct with no EGFP gene, along with the Y66S dGFP plasmid and corresponding gRNA. We observe GFP positive cells with C•G to A•T base editing, and signal was enhanced with the addition of thymidine (Figure 4.2B).

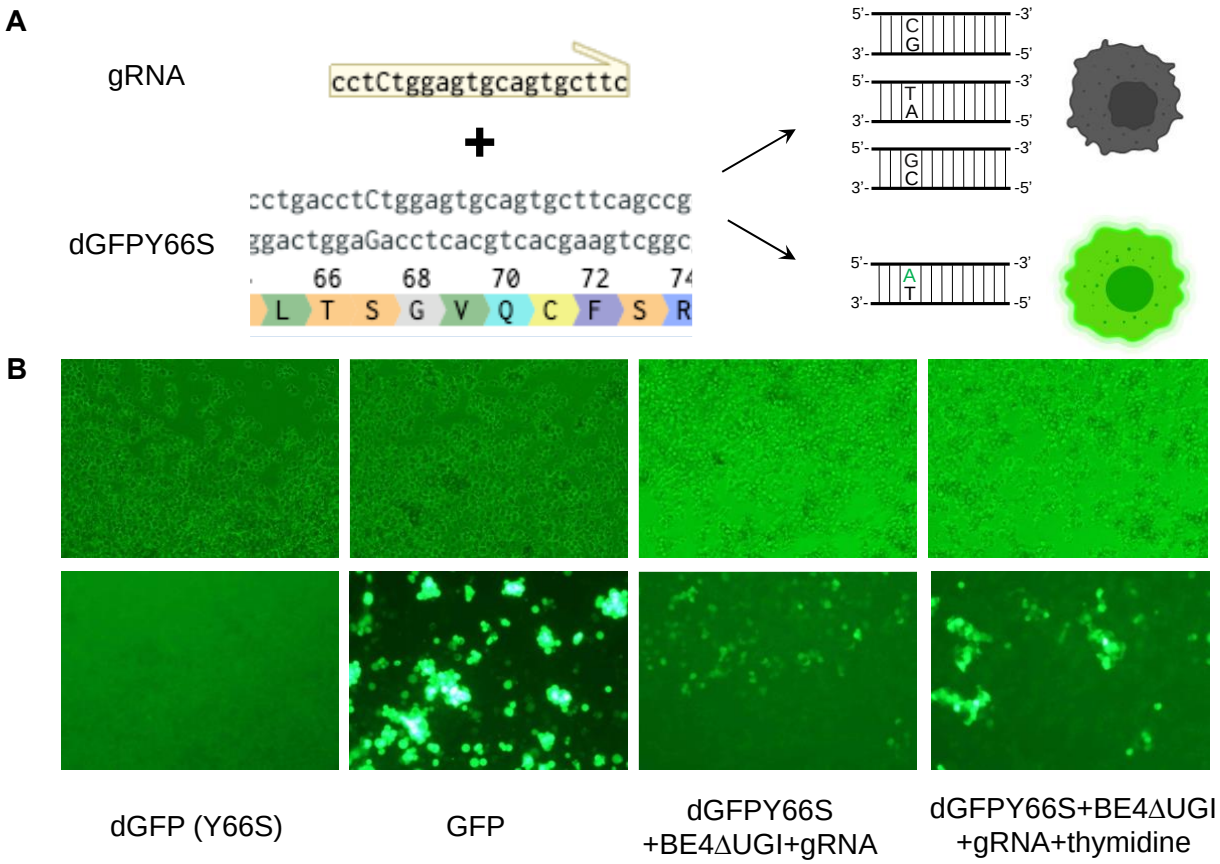

Figure 4.2: Fluorescent assay for C•G to A•T activity. (A) gRNA targeting the indicated sequence of the GFP gene. (B) Imagery of GFP expression for mutant dGFPY66S, positive GFP control, base editor conditions, and BE expression under thymidine synchronization.

After validating the screening conditions, we created a HEK293T cell line with an integrated CBEΔUGI under the control of the Tet promoter for inducible expression. After 5 days of hygromycin selection, we added doxycycline to induce base editor expression. We confirmed expression via mCherry visualization, and further established base editor activity with transfection of dGFP and gRNA (Figure 4.3). Next, we integrated the dGFP and protospacer into our inducible cell line with lentivirus. The construct bearing the A>C mutation at Y66 was cloned into a lentiviral screening construct, along with the appropriate

protospacer. This was transduced into HEK293T-CBEΔUGI cells and selected for positive integration via puromycin selection over five days.

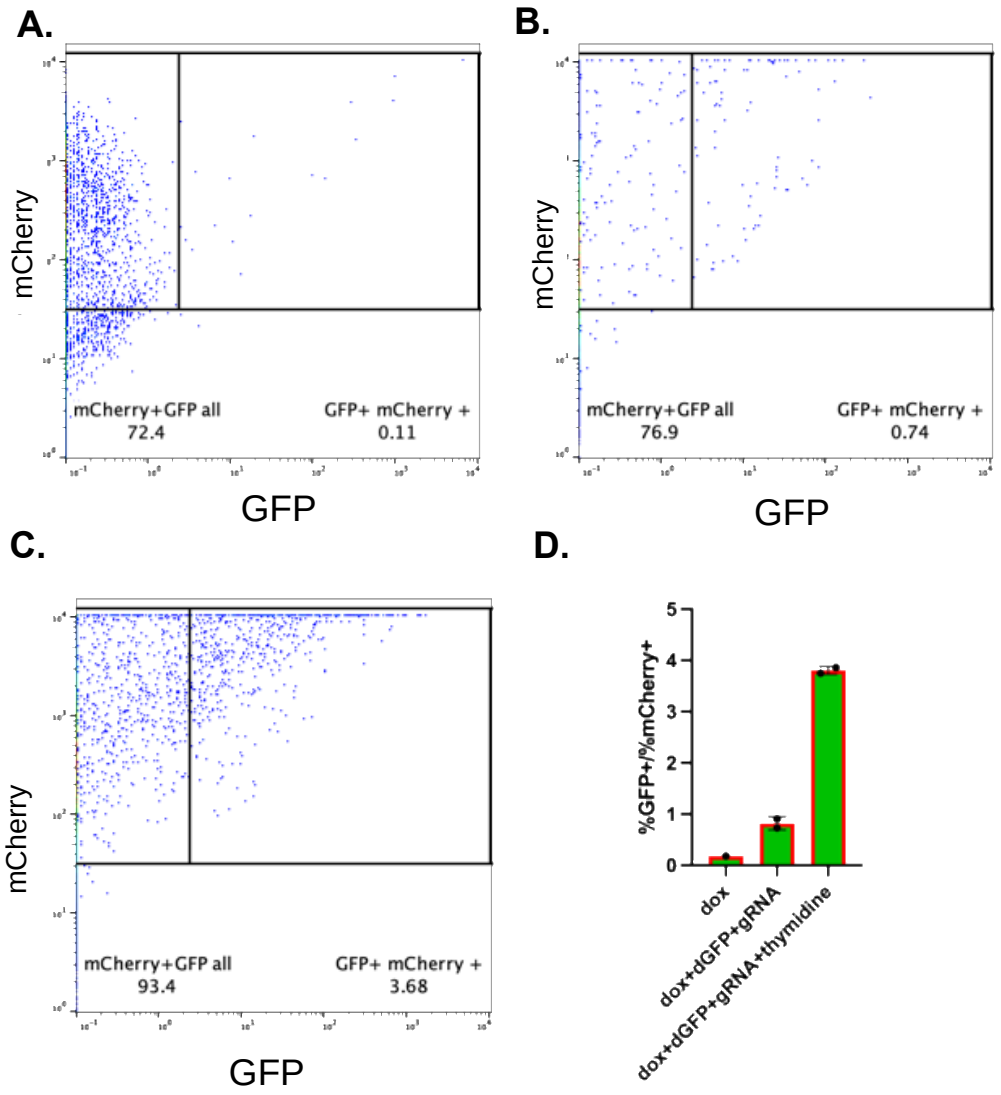

Figure 4.3: Flow cytometry validating C•G to A•T fluorescent turn-on screen. Integrated CBEΔUGI cell line was transfected with dGFPY66S and the gRNA targeting the corresponding protospacer sequence. 7 days post-transfection, cells were collected for flow cytometry analysis. (A) Samples without doxycycline induction (B) Under doxycycline induction (C) Dox induction plus 54-hour thymidine treatment, followed by synchronization release from G1 arrest into DMEM media.

siRNA Knockdown for C•G to A•T Screen

We decided to use our C•G to A•T screen assay in combination with siRNA knockdown to identify DNA repair pathways that convert U:G to A:T. siRNA lowers expression of genes via binding of the mRNA produced by the target gene, preventing the translation into protein, a process known as RNA interference (RNAi). We ordered commercially available and verified siRNA's that target several DNA repair pathways thought to act on base editing intermediates. These include base excision repair, single strand repair, mismatch repair, and translesion synthesis. Due to the redundancy of these targets, multiple genes in each pathway were targeted for knockdown to counteract the backup genes. We transfected in siRNA at day 0 to ensure complete repression prior to expression of base editor. After 24 hours, we added doxycycline to induce BE expression. Additionally, to enhance screening signal, we added thymidine to our cells for 54 treatment, as prior observations indicated this increases GFP signal. Cells were monitored and split before confluence in the 48-well plate format, and further passaged into 12-well plate format for flow cytometry analysis. We evaluated fluorescent signal by quantifying GFP expression only in mCherry+ cells, indicating active base editor. Transfection of siRNA itself decreased GFP+ slightly, when comparing non-targeting siRNA knockdown to untreated samples (**Figure 4.4**). This non-targeting siRNA was used as a negative control for our assay, while UNG knockdown was used as a positive control, given its established role in C•G to A•T editing. We additionally prepared samples for HTS analysis and qPCR by harvesting gDNA and mRNA in parallel (data to be collected prior to the defense).

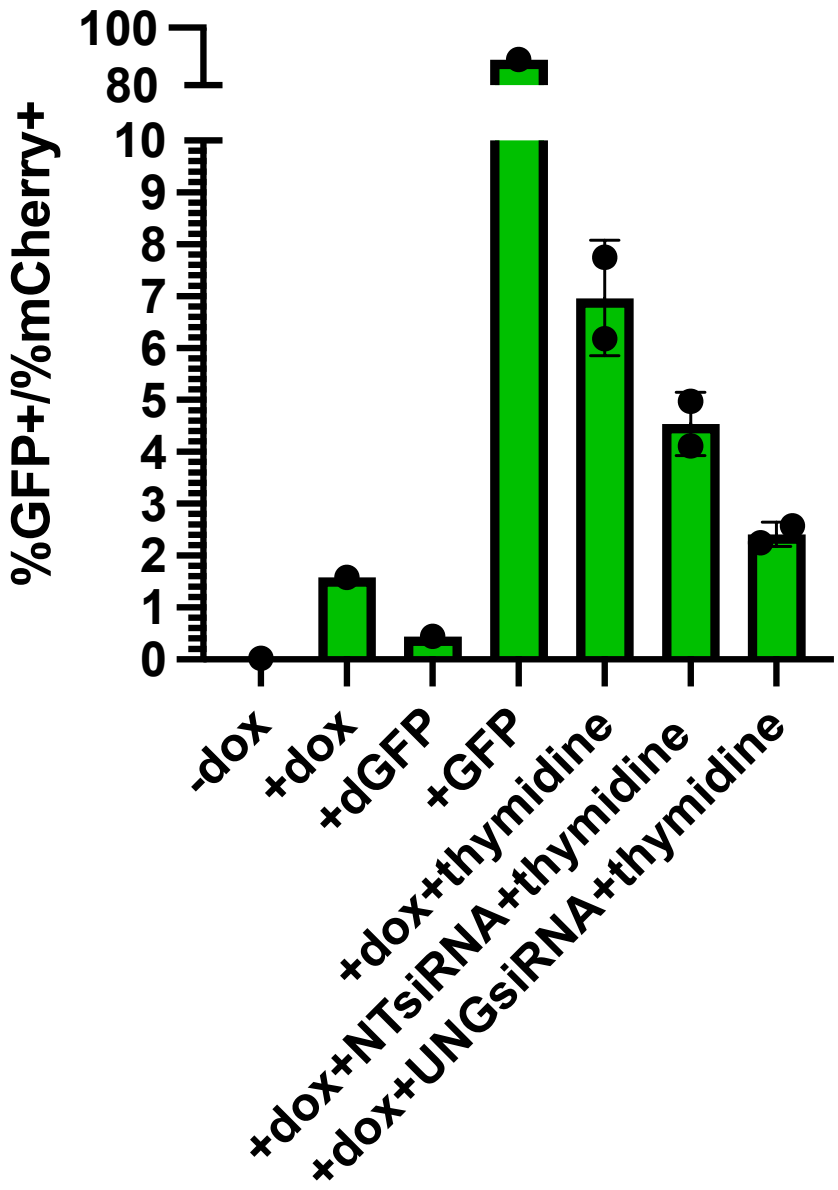

Figure 4.4: Control conditions for siRNA knockdown C•G to A•T editing assay. HEK293T cells with integrated CBEΔUGI were transfected with indicated siRNAs. 24 hours post-transfection, base editor expression was induced with the addition of doxycycline. Thymidine was added for 54 hours, then released into dox media for viability purposes. Cells were collected for FACS analysis on day 5.

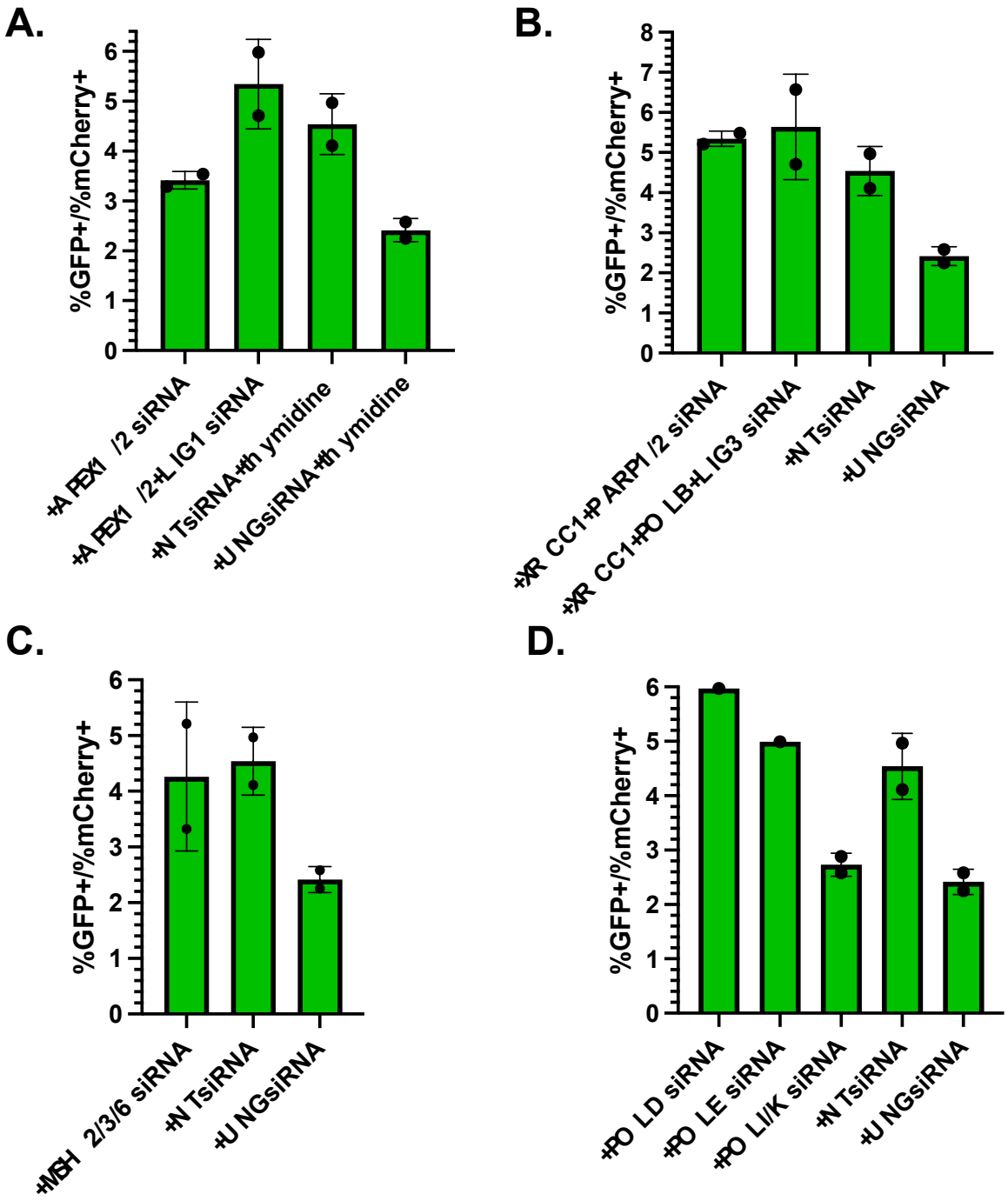

Figure 4.5: siRNA knockdown of DNA repair pathways' effect on C•G to A•T editing. HEK293T cells with integrated CBEΔUGI were transfected with indicated siRNAs. 24 hours post-transfection, base editor expression was induced with the addition of doxycycline. Thymidine was added for 54 hours, then released into dox media for viability purposes. Cells were collected for FACS analysis on day 5. (A) Base excision repair pathway knockdown including spBER (APEX1 and APEX2) and lpBER (APEX1,2 and LIG1) with NT siRNA negative control and UNG siRNA positive control. (B) Single-strand break repair pathways (XRCC1+PARP1+PARP2 and XRCC1+POLB+LIG3). (C) Mismatch repair (MSH2+MSH3+MSH6). (D) Translesion synthesis pathways (POLD, POLE, and POLI+POLK).

We find that siRNA knockdown of *UNG* leads to expected reductions in C•G to A•T editing levels (4.5% to 2.4%, Figure 4.5A). Additionally, decreasing expression of BER proteins APEX1 and APEX2 also reduces C•G to A•T editing efficiency to 3.2%, likely due to their processing of UNG-induced abasic sites. Interestingly, knockdown of LIG1 in combination with APEX1 and APEX2 restores C•G to A•T editing levels within error of negative controls (Figure 4.5A). For XRCC1-initiated single-strand break repair pathways, there was no observed difference in GFP signal (Figure 4.5B), likely excluding their role in this repair outcome. The mismatch repair pathway proteins MSH2, MSH3, and MSH6 demonstrated high variance between replicates (Figure 4.5C) and require further analysis to evaluate their role in C•G to A•T repair. Knockdown of the translesion synthesis protein POLD displays increased levels of C•G to A•T editing, while POLE had no significant difference from negative non-targeting control. Efforts to knockdown both POLD and POLE simultaneously proved lethal. Finally, reducing activity of TLS proteins POLI and POLK reduces C•G to A•T editing to 2.7%, comparable with our positive control kd of UNG. We conclude these proteins are likely responsible for dictating C•G to A•T repair outcomes, dependent upon UNG excision and APEX1/2 processing of uracil intermediates.

Creation of a G1-Regulated Base Editor

Given G1 synchronization is associated with distinct C•G to A•T editing patterns, we fused a degradation protein modifier to the CBEΔUGI construct to prevent base editing in S or G2 phase (Figure 4.6A). We cloned the degradation domain of Cdt1 to the C terminus of Cas9, analogous to addition of BE modifiers such as UGI. We transfected this construct into HEK293T cells targeting two different sites within the genome. We observed C>G and C>T editing at

similar levels to CBEΔUGI, establishing activity for a base editor degraded in S/G2 phase (Figure 4.6B). Unfortunately, there was minimal C•G to A•T editing activity, comparable to levels observed CBEΔUGI, with below 10% C•G to A•T editing as compared with 30% when synchronized with thymidine (Figure 4.6B). We hypothesized that the steric hindrance due to Cas9 might be impeding Cdt1-mediated degradation and added an XTEN linker sequence between Cas9 and Cdt1 to increase accessibility. We tested asynchronous and thymidine treated cells to examine the editing distribution at the *HEK2*, *HEK4* and *HIRA* sites. There was no change in C•G to A•T editing outcomes. Our results indicate fusion of Cdt1 to base editor only increased C•G to A•T outcomes in one of three sites tested, and thus is unlikely as a target to further develop CABE's.

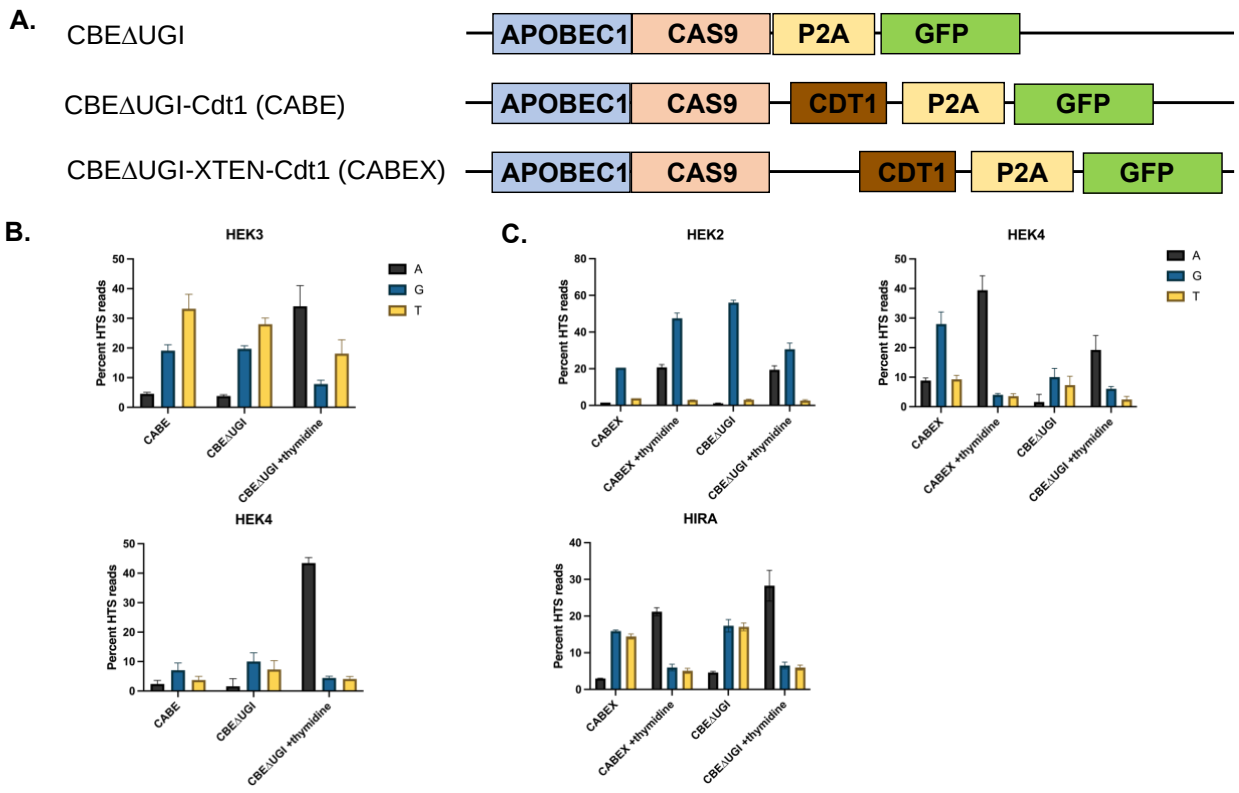

Figure 4.6: Design and implementation of a G1-restricted base editor. (A.) Construct maps of CBEΔUGI-Cdt1 fusions. (B.) CBEΔUGI-Cdt1 base editing efficiency in HEK293T cells at the HEK3 and HEK4 sites. Editing is displayed by nucleotide outcome. (C.) CBEΔUGI-XTEN-Cdt1 base editing efficiency in HEK293T cells at HEK2, HEK4 and HIRA sites.

Base Editing with Enriched Nucleotide Pools

We determined if local nucleotide concentration was responsible for changes in editing outcomes. First, we transfected cells with CBEΔUGI and gRNA, then spiked in dATP, dCTP, and dGTP at 5 mM at 6 hours post-transfection before collecting the gDNA at 54 hours for sequencing, consistent with our previous protocol. We analyzed overall editing efficiencies (percent of HTS reads with the target C•G edited to T•A, G•C, or A•T) and product distributions (the relative portion of edited sequencing reads in which the target C•G is edited to T•A, G•C, or A•T) of each of the samples treated with CBEΔUGI (Figure 3.6A-C). We observed slight reductions in editing efficiencies across all three sites tested, but no

significant alterations in product distribution for each of the editing outcomes. At the *HEK2* site, we again observed high rates of C•G to G•C editing, with minimal C•G to T•A editing. Addition of each individual nucleotide did not significantly change editing outcomes (Figure 3.6A), where approximately 90% of editing outcomes result in C•G to G•C conversions, independent of dNTP treatment. Interestingly, dTTP addition did not lead to higher rates of C•G to A•T editing, indicating reliance on G1 synchronization rather than local nucleotide pool concentrations that places a dTTP nucleotide across from the processed abasic site intermediate. Examining the other two tested sites, HIRA and RNF2, reveals similar trends (Figure 4.6B,C). At the *HIRA* site, there was approximately equivalent C•G to T•A and C•G to G•C editing, with < 10% C•G to A•T editing. At *RNF2*, C•G to G•C transitions constituted 61-74% of editing outcomes, with C•G to T•A representing 21-33%, and ~5% of edits from C•G to A•T. Upon addition of dTTP at *HIRA* and *RNF2*, there was a slight increase in proportional C•G to A•T editing (7.9 to 11.5% at *HIRA* and 5.4 to 8.6% at *RNF2*), but not as substantially as editing outcomes in the presence of thymidine alone. These observations suggest mild influence of nucleotide pool concentrations on editing outcomes, but likely more influenced by repair proteins, as will be explored by siRNA knockdown studies.

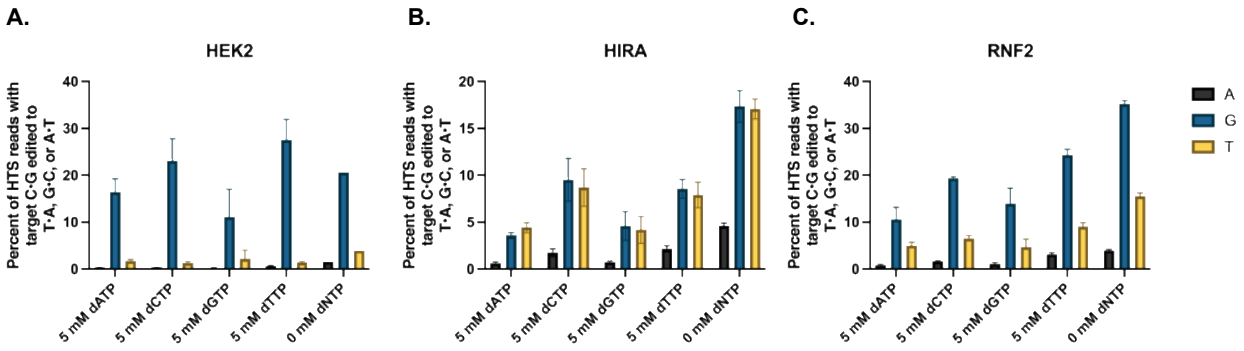

Figure 4.7: Enhancement of nucleotide pools with base editing. HEK293T cells were transfected with CBEΔUGI along with (A) HEK2, (B) HIRA, and (C) RNF2 gRNAs in the presence of 5 mM dATP, dCTP, dGTP, and dTTP. Genomic DNA was extracted and target loci were amplified via PCR and subjected to HTS. Genome editing efficiencies (percent of total HTS reads with the target C•G base converted to T•A, G•C, or A•T) were quantified with CRISPResso2.

DISCUSSION

Here we investigated the implications from our synchronization experiments, specifically the increase of C•G to A•T editing outcomes via thymidine-induced G1 synchronization. We excluded the role of dNTP concentration, given that addition of individual deoxynucleotides did not significantly impact editing outcomes. We acknowledge that dTTP did not replicate the results of thymidine synchronization as could be expected and suggest the conversion of thymidine to dTTP within HEK293T cells as more the mechanism behind synchronization in HEK293T cells.

We developed and tested a fluorescent screen for probing siRNA knockdown of DNA repair genes and discover additional TLS polymerases responsible for induction of C•G to A•T editing in addition to BER proteins responsible for abasic site induction. Abasic site introduction via UNG and APEX1/2 are essential for initiation of mutagenic outcomes, and this was further confirmed by our results. Individual knockdown of APEX1 and APEX2 could further confirm dependence on a single endonuclease, while lack of reduction in C•G to A•T editing would suggest redundancy or both interacting with DNA damage. Similarly, knockdown of POLI and POLK separately is suggested for further clarity of their role in mutagenic DNA repair. POLI is unique to mammalian cells, with structural homologs only present in mice and fruit flies, and not bacterial or yeast cells[48]. Observations that CGBEs in bacterial cells induce C•G to A•T edits suggest either a different mechanism of action, such as recruitment of POLZ[49]. POLK is a Y family TLS polymerase that is capable of inserting nucleotides across from abasic sites in vitro[50]. Furthermore, overexpression of POLK is linked with a 10-fold increase in mutagenesis[51]. Collectively, these observations suggest a future role in developing new classes of C•G to A•T base editors via fusion of POLI/K.

We test a novel cell cycle regulated base editor using a fusion of Cdt1 to CBEΔUGI constructs. Further confirmation of cell cycle dependent degradation via western blot analysis is lacking in our analysis and could explain the lack of changes in editing distribution. If the Cdt1 is indeed inducing S-phase degradation, then a possible explanation could be repair of uracil is conducted during the resultant S and G2 phases of the cell cycle, even in the absence of base editor. Other cell cycle regulated proteins could be explored as possible modifiers to create C to A base editors.

Methods

Constructs and Molecular Cloning

All BE plasmids were constructed with USER cloning[29] with pCMV ABEmax_P2A_GFP (Addgene #112101) and pCMV_AncBE4max_P2A_GFP (Addgene #112100) plasmids as template, using Phusion U Hot Start Polymerase (ThermoFisher Scientific). All sgRNA expression plasmids were generated using blunt-end cloning with pFYF1230 (Addgene plasmid #47511) as a template, using Phusion High-Fidelity DNA Polymerase (New England BioLabs). All DNA vector amplification was carried out using NEB 10-β competent cells (New England BioLabs). All plasmids were purified using the ZymoPURE II Plasmid Midiprep Kit (Zymo Research D4200).

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.

Transfections

For all HEK293T cell transfections, 100,000 HEK293T cells in 250 µL of DMEM media without Penicillin-Streptomycin were added per well to 48-well VWR Multiwell Cell Culture Plates on top of lipofectamine/plasmid mixtures. The lipofectamine/plasmid mixtures consisted of 1000 ng of BE plasmid, 250 ng of sgRNA plasmid, and 1.5 µL of Lipofectamine 2000 (ThermoFisher Scientific) in 25 µL of total volume, made up with Opti-MEM (Gibco #31985-070). Thymidine was added 6 hours after transfection from stock solutions (described below) to a final concentration of 5 mM.

High-Throughput DNA Sequencing (HTS) of Genomic DNA

Transfected cells were rinsed with 150 µL PBS (ThermoFisher Scientific) per well at the indicated time points 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 bearing homology to the target site and relevant Illumina forward and reverse adapters,1 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 from 2% agarose gel with QIAquick Gel Extraction Kit (Qiagen) and quantified using NEBNext Ultra II DNA Library Prep Kit (NEB) on a CFX96 system (BioRad). Samples were then sequenced on an Illumina MiniSeq according to the manufacturer's protocol.

HTS and Indel Analysis of Targeted Amplicon Sequencing Reads

Analysis of Illumina HTS sequencing readout was conducted with CRISPRessov2[18,39]. Specifically, for these analyses, fastq files were analyzed via scripts run on Docker, where the reads were analyzed against the entire amplicons, with outputs for the guide RNA and base editor (--guide_seq and –base_editor_output). Product distribution for CBE variants was determined by taking the fraction of individual A•T, G•C, and T•A reads and dividing by the sum. CRISPResso was also used to validate editing percentages and analyze indel frequency, where the total number of indel reads was obtained from the indel histogram output and expressed as the fraction of reads with indel over total reads. For analysis of indel sequences, specific reads constituting >4% of the total indels from the CRISPResso were compiled.

siRNA Knockdown

Dharmacon™ siRNAs were obtained from Horizon and resuspended in a 1x siRNA buffer, composed of 60 mM KCl, 6 mM HEPES-pH7.5, and 0.2 mM $MgCl_2$. siRNA was stored at 20 µM at -20ºC, until transfection. For transfections in 48-well plate format, 1 µL of 5 µM siRNA was combined with 1.5 µL Lipofectamine 2000 (ThermoFisher Scientific) in 25 µL of total volume, made up with Opti-MEM (Gibco #31985-070).

RNA Isolation and Purification

Total RNA was isolated via Zymo RNA extractions kit (Zymo, R1054), as per manufacturer's protocol. Cells were lysed with 300 µL RNA lysis buffer before addition of 300 µL 100% ethanol. The mixture was centrifuged in the Zymo-Spin IC Column at 16,000 rcf for 30 seconds. DNase treatment followed this by washing with 400 µL RNA wash buffer and treating with 5 uL DNase 1 (1 U/µL) and 35 µL DNA Digestion Buffer for 15 minutes at room temperature. 400 µL RNA Prep Buffer was added to the column, centrifuged, and washed 2x with RNA wash buffer prior to collection with 50 µL DNase/RNase-Free water. RNA was stored at -20º C before library preparation.

Flow Cytometry Analysis of GFP Fluorescence

For all GFP fluorescence measurements, 1 x 10^6 cells were resuspended in FACS buffer (1% FBS, 50 µM EDTA pH 8.0, 2 µg/mL PI [Sigma]) and filtered through a cell-strainer. Non-viable cells and doublets were eliminated via gating parameters. Flow cytometry was performed on a BioFortessa or S3e cell sorter (Bio-Rad).

REFERENCES

1. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
2. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).
3. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., & Abecasis, G.R. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
4. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102-110 (2015).
5. Blaese, R. M., Culver K., Miller, D., Carer, C., Fleisher, T., Clerici, M., Shearer, G., Chang, L., Chiang, Y., Tolstoshev, P., Greenblatt, J., Rosenberg, S., Klein, H., Berger, M., Muloen, C., Ramsey, W.J., Muul, L., Morgan, R., & Anderson, F. T Lymphocyte-Directed Gene Therapy for ADA− SCID: Initial Trial Results After 4 Years. *Science* **270**, 475–480 (1995).
6. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
7. Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res.* **18**, 134–147 (2008).
8. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair* **7**, 1765–1771 (2008).
9. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334 (2014).

10. Fishman-Lobell, J., Rudin, N. & Haber, J. E. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol. Cell. Biol.* **12**, 1292–1303 (1992).
11. Allen, F., Crepaldi, L., Alsinet, C., Strong, A., Kleschchevinkov, V., & Leopold, P. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–72 (2019).
12. Sergison, E., Dlventhal, K., Pruett-Miller, S., Modarai, S., Kmiec, E., Regan, M., & Miller, B. GERG Study 2019: Reproducibility of indel formation rates by comparing guideRNA format and delivery method. *J. Biomol. Tech. JBT* **31**, S9 (2020).
13. Carroll, D. Genome Engineering With Zinc-Finger Nucleases. *Genetics* **188**, 773–782 (2011).
14. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., & Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
15. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
16. Jinek, M., East, A., Cheng, A, Lin, S., Ma, E., & Doudna, J. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
17. Jiang, F. & Doudna, J. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, (2017).
18. 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).

19. Gaudelli, N. M., Komor, A.C., Rees, H.A., Packer, M., Badran, A., Bryson, D., & Liu, D. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
20. Krokan, H. E. & Bjørås, M. Base Excision Repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012583 (2013).
21. Hardeland, U. Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J., & Schar, P. Thymine DNA glycosylase. *Prog. Nucleic Acid Res. Mol. Biol.* **68**, 235–253 (2001).
22. Nilsen, H., Haushalter, K.A., Robins, P., Barnes, D.E., Verdine, G.L., & Lindahl, T. Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *EMBO J.* **20**, 4278–4286 (2001).
23. Demple, B. & Sung, J.-S. Molecular and biological roles of Ape1 protein in mammalian base excision repair. *DNA Repair* **4**, 1442–1449 (2005).
24. Zlatanou, A., Despras, E., Braz-Petta, T., Boubakour, I., Pouvelle, C. Stewart, G., & Nakajima, S. The hMsh2-hMsh6 Complex Acts in Concert with Monoubiquitinated PCNA and Pol η in Response to Oxidative DNA Damage in Human Cells. *Mol. Cell* **43**, 649–662 (2011).
25. Goodman, M. F. & Woodgate, R. Translesion DNA Polymerases. *Cold Spring Harb. Perspect. Biol.* **5**, a010363 (2013).
26. 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).
27. Jackman, J. & O'Connor, P. M. Methods for Synchronizing Cells at Specific Stages of the Cell Cycle. *Curr. Protoc. Cell Biol.* **00**, 8.3.1-8.3.20 (1998).

28. Kim, Y. B., Komor, A. C., Levy, J., Packer, M., Zhao, K.T., & Liu, D. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
29. Badran, A. H., Guzov, V., Huai, Q., Kemp, M., Vishwanath, P., Kain, W., Nance, A.M., Evdokimov, A., Moshiri, F., Turner, K.H., Wang, P., Malvar, T., & Liu, D. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **533**, 58–63 (2016).
30. Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nat. Rev. Mol. Cell Biol.* **9**, 297–308 (2008).
31. Yeh, W.-H., Chiang, H., Rees, H. A., Edge, A. S. B. & Liu, D. R. In vivo base editing of post-mitotic sensory cells. *Nat. Commun.* **9**, 2184 (2018).
32. Lim, C. K. W., Gapinske, M., Brooks, A., Woods, W., Powell, J.E., Zeballos, M.A., Winter, J., Perez-Pinera, P., & Gaj, T. Treatment of a Mouse Model of ALS by In Vivo Base Editing. *Mol. Ther.* **28**, 1177–1189 (2020).
33. Komor, A. C., Zhao, K.T., Packer, M., Gaudelli, N.M., Waterbury, A.L., Koblan, L.W., Kim, Y. B., Badran, A.H., & Liu, D. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
34. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
35. Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., Li, Y., Zhou, H., Guo, F., & Yang, H. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).

36. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).

37. Kim, J., Malashkevich, V., Roday, S., Lisbin, M., Schramm, V.L., & Almo, S.C. Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase. *Biochemistry* **45**, 6407–6416 (2006).

38. Koblan, L. W., Arbab, M., Shen, M.W., Hussmann, J.A., Anzalone A.V., Doman, J.L., Newby, G.A., Yang, D., Mok, B., Replogle, J.M., Xu, A., Sisley, T.A., Weissman, J.S., Adamson, B., Liu, & D.R. Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat. Biotechnol.* **39**, 1414–1425 (2021).

39. 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., & Pinello, L. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

40. Burnett, C. A., Wong, A.T., Vasquez, C.A., McHugh, C.A., Yeo, G.W., & Komor, A.C. Examination of the Cell Cycle Dependence of Cytosine and Adenine Base Editors. *Front. Genome Ed.* 39 (2022).

41. Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M.A., Rosser, S.J., Bi, C., & Zhang, X. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2021).

42. Kurt, I. C., Zhou, R., Iyer, S., Garcia, S.P., Miller, B.R., Langner, L.M., Grunewald, J., & Young, J.K. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).

43. Ligasová, A. & Koberna, K. Strengths and Weaknesses of Cell Synchronization Protocols Based on Inhibition of DNA Synthesis. *Int. J. Mol. Sci.* **22**, 10759 (2021).
44. Pozo, P. N. & Cook, J. G. Regulation and Function of Cdt1; A Key Factor in Cell Proliferation and Genome Stability. *Genes* **8**, 2 (2016).
45. Kim, Y. & Kipreos, E. T. Cdt1 degradation to prevent DNA re-replication: conserved and non-conserved pathways. *Cell Div.* **2**, 18 (2007).
46. Hu, J., McCall, C. M., Ohta, T. & Xiong, Y. Targeted ubiquitination of CDT1 by the DDB1-CUL4A-ROC1 ligase in response to DNA damage. *Nat. Cell Biol.* **6**, 1003–1009 (2004).
47. Matsumoto, D., Tamamura, H. & Nomura, W. A cell cycle-dependent CRISPR-Cas9 activation system based on an anti-CRISPR protein shows improved genome editing accuracy. *Commun. Biol.* **3**, 1–10 (2020).
48. Guo, C., Kosarek-Stancel, J. N., Tang, T.-S. & Friedberg, E. C. Y-family DNA polymerases in mammalian cells. *Cell. Mol. Life Sci.* **66**, 2363–2381 (2009).
49. Prakash, S. & Prakash, L. Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev.* **16**, 1872–1883 (2002).
50. Ohashi, E., Ogi, T., Kusumoto, R., Iwai, S., Masutani, C., Hanaoka, F., & Ohmori, H. Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa. *Genes Dev.* **14**, 1589–1594 (2000).
51. Ogi, T., Kato, T., Kato, T. & Ohmori, H. Mutation enhancement by DINB1, a mammalian homologue of the Escherichia coli mutagenesis protein DinB. *Genes Cells* **4**, 607–618 (1999).