

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Committee on Higher Degrees in Biological Sciences in Public Health
have examined a dissertation entitled
Engineering CRISPR-Based DNA Transversion Editing Technologies
presented by Ibrahim Cagri Kurt
candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature Curtis Huttenhower

Typed name: Prof. Curtis Huttenhower

Signature Brendan Manning

Typed name: Prof. Brendan Manning

Signature Lee Zou

Typed name: Prof. Lee Zou

Signature Jia Chen

Typed name: Prof. Jia Chen

Date: November 23, 2021

Engineering CRISPR-Based DNA Transversion Editing Technologies

A dissertation presented by

Ibrahim Cagri Kurt

to

The Committee on Higher Degrees in Biological Sciences in Public Health

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

November 2021



© 2021 Ibrahim Cagri Kurt

All rights reserved.

Engineering CRISPR-Based DNA Transversion Editing Technologies

Abstract

Adenine base editors (ABE) and cytosine base editors (CBE) can efficiently install A-to-G and C-to-T transition edits, respectively. However, until recently, the field lacked base editors to efficiently install C-to-G transversion edits in human cells in a programmable manner. We modified the existing CBE architectures to repurpose them into new C-to-G base editors (CGBEs). Our optimized CGBE induced C-to-G alterations with efficiencies as high as ~70%, particularly in AT- rich sequence contexts. CGBE exhibited a narrow editing window within the target site (positions 5-7 in the SpCas9 spacer sequence) but we also showed that incorporation of SpCas9 PAM recognition variants that can recognize NG or NGA PAMs into CGBE are functional, thereby expanding the range of targetable sites. CGBE displayed favorable Cas9-dependent DNA off-target profiles compared to the BE4max C-to-T base editor. We found in side-by-side comparison that CGBE achieves higher efficiencies of C-to-G edits compared with the recently described PE2 and PE3 prime editing technologies across different cell lines. CGBE described here expands the editing capabilities of the base editor platform for both research and therapeutic applications.

Chapter 1 lays the foundational introduction into the field of genome editing, prominent tools engineered in chronological order, off-target genome editing, and platforms and methods to assess specificity profiles of genome editing tools.

Chapter 2 describes the novel CRISPR-based C-to-G Base Editors (CGBE) that can efficiently install C-to-G transversion edits in human cells in context-dependent manner.

Chapter 3 describes the efforts and attempt towards engineering novel CRISPR-based Guanine Base Editors (GBE) that can introduce G-to-T transversion mutations.

Chapter 4 describes the efforts to integrate Tn5-based transposition system into the current frameworks of the gold standard *in vitro* and *in vivo* off-target methods to increase their scalability, sensitivity, and time-efficiency. Finally, this chapter describes a synthetic self-targeting in cellular off-target library that can be used to score precision or fidelity of CRISPR-based genome editing tools in a high-throughput manner.

Chapter 5 discusses the big picture and future of the genome editing field, bigger research relationships, and our contribution to the current field with the work outlined in this thesis.

Table of Contents

Abstract.....	iii
Table of Contents.....	v
Acknowledgements.....	viii
Chapter 1: Introduction.....	1
Advent of gene targeting.....	3
Advent of genome editing tools: Meganucleases.....	5
Repair outcomes of DSBs in mammalian cells.....	6
Advent of genome editing tools: Zinc-finger Nucleases (ZFN).....	10
Advent of genome editing tools: Transcription Activator-Like Effector Nucleases (TALEN).....	13
The Scissor: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR).....	14
The Pen: 2nd generation of CRISPR-based genome editing tools: Advent of Base Editors.....	18
The Word Processor: 3rd generation of CRISPR-based genome editing tools: Advent of Prime Editor.....	22
Genome-wide and Transcriptome-wide Off-target Editing of Genome Editing Tools.....	26
In Cellular Synthetic Off-target Libraries.....	29
Chapter 2: Engineering DNA Transversion C-to-G Base Editors (CGBE).....	34

CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells...	37
Discussion & Future Directions.....	50
Extended Data.....	59
Methods.....	67
Chapter 3: Engineering DNA Transversion G-to-T Guanine Base Editors (GBE).....	73
Introduction.....	73
Results.....	75
Discussion & Future Directions.....	82
Chapter 4: Developing Sensitive, Scalable and High-Throughput Off-target Analysis Platforms.....	87
Tn5-Assisted Circularization Sequencing (TACO-seq) & Sensitive Capture of Oligonucleotides Post-Editing Sequencing (SCOPE-seq).....	87
Introduction.....	87
Results.....	93
Tn5-Assisted Circularization Sequencing (TACO-seq).....	94
Sensitive Capture of Oligonucleotides Post-Editing Sequencing (SCOPE-seq).....	96
Discussion & Future Perspectives.....	105

In Cellulo Self-targeting Off-target Platform to Score Fidelity of CRISPR-based Genome Editing Tools.....	109
Introduction.....	109
Results.....	110
Discussion & Future Directions.....	120
Chapter 5: Discussion & Future Directions.....	123
Supplementary Figures.....	130
Supplementary Tables.....	146
Supplementary Table 1.....	147
Supplementary Table 2.....	150
Supplementary Table 3.....	162
References.....	185

Acknowledgements

إن الحمد لله نحمده ونستعينه ونستغفره ونعوذ بالله من شرور أنفسنا ومن سيئات أعمالنا من يهده الله فلا مضل له ومن يضلل فلا هادي له, وأشهد ألا إله إلا الله وحده لا شريك له وأشهد أن محمدا عبده ورسوله صلى الله وسلم عليه وعلى آله وصحبه أجمعين – أما بعد:

All praise is due to Allah ﷻ, the Master of all universes, the earth, skies and whatever is in and in between them. To Allah ﷻ we belong, to Him we shall return. We testify that there is no god but Allah ﷻ, and that Muhammed ﷺ is His final messenger and prophet. We ask Allah ﷻ the Exalted to send an abundance of prayers, salutations, and blessings upon our beloved Messenger Muhammed ﷺ, his family, and his companions. Allahumma ameen.

I would not have been able to complete ANY part of this hard thesis work without the unceasing support and nourishment from my Maker Allah ﷻ who created and originated the beautiful and powerful DNA, irreducibly complex enzymatic systems such as CRISPR-Cas9, and the meticulously detailed organisms that host them. So first and foremost, I thank Allah ﷻ. I can never thank you enough till my last breath.

I would not have been able to reach the finish line without the continuous support from my beautiful and caring mother Dr. Cemile Kurt, my generous and strong father Dr. Necmi Kurt, my smart sisters Dr. Kübra Neslihan Kurt Oktay and Melike Kurt, my brother Serdar Kurt, and my always-smiling brother-in-law Dr. Alper Oktay. Thank you, all of you, but especially my mother!

The momentum I gained from the happiness and joy I attained from my marriage with my incredibly smart and skillful wife Emine Karakaş Kurt, and the birth of my Zeyneb, is irreplaceable and priceless. I, firsthand, experienced and witnessed the saying that “behind every

great man is a great woman”. Thank you Emine and Zeyneb, for everything! Knowing your presence, being able to heatedly discuss the scientific projects with you in depth, and literally benefitting from your experimental suggestions always warmed my heart.

Last but not least, I would like to sincerely thank Dr. Keith Joung for his support and mentorship throughout my PhD studies. And a special thanks goes to my colleague Dr. Jason Gehrke with whom I was able to consult with any academic and experimental matter that bothered me during these tough years.



Chapter 1: Introduction

It has been 155 years since Gregory Mendel published his seminal work on peas in his monastery's backyard that revolutionarily described the inheritance patterns of phenotypic traits as transmittable distinct units (Mendel, 1866). Since then, DNA as the nucleic acid in nuclei of cells was isolated and named as nuclein by Frederick Miescher in 1869, chromosomes were observed to follow Mendelian inheritance pattern in 1902 by Walter Sutton, and genes were shown to be located in chromosomes while physically linked to each other by Thomas Hunt Morgan in 1911 through his seminal work with fruit flies (Miescher, 1869; Morgan, 1910; Sutton, 1902). Today, we know that the complete genetic material in an organism called genome holds every bit of information required for a cell/organism not only to function in a perfect homeostasis at cellular and tissue levels, but also to reproduce at systemic level to duplicate and pass its genome to generate its offspring. This complex set of functions are encoded in DNA molecules using molecular "letters" called nucleobases of adenine (A), cytosine (C), thymine (T), and guanine (G). The intricate cellular machinery that allows the smooth and responsive functioning of cells, and eventually copying of its genomic material in high-fidelity manner to divide, is governed by the pathway known as the central dogma. The central dogma briefly depicts that DNA stores the information for RNAs to be transcribed from itself, and that RNA encodes the information for proteins/enzymes to be translated from itself. This seemingly simple, yet delicate machinery ultimately gives rise to all of the natural diversity -in state of health and disease- as we see today in nature.

Being able to alter the wildtype sequences of genomic DNAs of organisms lying at the heart of this central dogma allows one to establish correlations between DNA and its cognate downstream RNA/protein counterparts at molecular level, and between DNA and phenotype at

cellular/organismal level. Such correlative knock-in/knock-out experiments constitute the core of deciphering molecular mechanisms playing behind the state of disease and health. Methods and tools to introduce such deliberate DNA insertions, deletions, substitutions, and structural rearrangements into the genomes of living organisms are collectively studied by the field of “genome editing” and hold tremendous potential for its applications across -virtually all- organisms, ranging from medicine, biotechnology to molecular biology and genetics. The ability to edit genome without limits would allow the generation of model cell lines and animals for different diseases paving the path towards therapeutic discoveries by drug screening, generation of engineered cell lines expressing novel therapeutics in the form of peptides/antibodies or chimeric receptors, and generation of transgenic animals with suppressed immunogenicity that holds extensive implications for organ transplantation. Finally, and arguably most importantly, genome editing tools and their applications empower scientists to potentially repair the pathogenic human DNA variations by either reversing/correcting them or knocking-out the sequences/genes responsible for the pathogenic phenotype.

The artificial selection of wild animals and plants carrying the desired traits by humans even prior to the advent of modern science and genomic information is a classic historical example of such a purposeful endeavor and application. Specifically, domestication of wolves in wild into dogs living with humans to be tamed and fed 20,000-40,000 years ago, and development of the modern maize from its ancient ancestor teosinte represent some of the aspirations of genome editing in agriculture and livestock, even in the absence of the technology and know-how (MacHugh et al., 2017; Meyer & Purugganan, 2013). Humans selected for teosintes plants that have larger ears and kernels for over numerous generations and ultimately domesticated them into modern maize as an important food source for humans and animals.

While we did not know the exact polymorphic genomic variants found within the genomes of wolves and teosintes plants that confer them the desirable traits, simple selective breeding of the organisms with the desired traits led to the gradual increase in prevalence of the desirable alleles in the population while the alleles of unwanted traits diminished.

Advent of gene targeting

The field of genetics for many years was contingent on the natural variants that were already present or on the natural mutations that arose in genomes of organisms during mitosis/meiosis. Therefore, we have been learning significant amount of details of cellular biology and physiology by identifying a phenotype first, and then retrospectively studying the DNA sequence variations that could have led to the phenotype in question. However, in the first half of the 20th century, HJ Muller in 1927 and C Auerbach in 1947 showed that you could artificially accelerate the rate of mutagenesis employing ionizing radiation (X-rays) or chemicals such as toxic mustard gas (Auerbach, 1945; Muller, 1927). Both groups were able to generate mutant *Drosophilas* with visible altered phenotypes upon mutagenic treatment. These mutations were arising faster than natural acquisition of single nucleotide polymorphisms through generations, but they were still random in nature and uncontrollable/unprogrammable. The efforts towards programmable and precise editing of genomic sequences were rampant in mid to late 1980s, especially after the advent of restriction endonuclease enzymes found in prokaryotes as part of a defense mechanism against the invading bacteriophages (Roberts, 2005). Restriction endonuclease enzymes were successfully purified and repurposed to be used in vitro for DNA cloning that led to the beginning of the era of recombinant DNA technologies. The main hallmark of restriction endonuclease enzymes was their short (usually 4-8 base pairs) DNA recognition motives. Fundamentally, the goal of the genome editing field was to be able to

translate these findings through different tools into *in cellulo/in vivo* settings, as the prize of such tools were highly rewardable.

One of the initial studies that set a major steppingstone for achieving this goal was the observation by Wigler and Axel in 1977 that one of the BamHI digested linear fragments of Herpes Simplex Virus (HSV) genome that carries the gene for thymidine kinase (tk) could be transfected into tk-deficient mammalian (mouse) cells in tissue culture (Wigler et al., 1977). Data strongly suggested that tk gene was randomly integrated into the host genome, while tk⁺ cells were stable and positively selected for over several hundred generations. Capecchi built on this observation and increased the efficiency of genomic integration by several folds of magnitude by carrying out microinjections directly into the nuclei of host mouse embryonic stem cells (mESCs) (Capecchi, 1980). However, his studies also revealed an interesting observation that was going to revolutionize the field. He demonstrated in certain cases, exogenous DNA was randomly integrated into host but in the form of a perfect concatemer of hundreds of same molecules. Further analysis led to the finding that homologous recombination (HR) was responsible for concatemerizing the exogenous DNA fragments, and that HR was functional in mammalian cells as well as yeast cells that was shown in parallel (Folger et al., 1982; Hinnen et al., 1978). These were exciting findings that led to the thinking that perhaps the endogenous homologous recombination machinery can be harnessed to carry out targeted (nonrandom) insertion of exogenous DNA molecules. To this end, the critical step in the right direction came from the studies of Capecchi introducing exogenous DNA fragments flanked by homologous arms (with endogenous target genomic sequences), allowing to virtually target any specific locus. However, this process was imperfect and leading to random off-target insertions. In order to select only for the cells carrying the on-target integration events, Capecchi devised a double

selection mechanism in which a neomycin resistance gene was located in between homology arms, and tk gene was located outside of the homology arms (Thomas et al., 1986). In the event of an imperfect off-target integration, tk gene would also be stably integrated into host genome. Double (positive via neomycin + negative via tk gene) selection by neomycin and the nucleoside analog that becomes toxic upon phosphorylation by tk gene, he was able to deplete the cells with the off-target integration events and enrich for the cells with the successful on-target integration events. This technique called “gene targeting” stood against the test of time and allowed the generation of tens of thousands of different knock-in and knock-out mESC lines and transgenic mice within two decades of its advent. The major drawback of gene targeting was the extremely low efficiency of successful integration events via HR leading to a single positive recombinant cell/clone in 10^3 - 10^9 cells depending on the cell type and state (Capecchi, 1989, 2005).

Advent of genome editing tools: Meganucleases

Scientists needed to discover alternative ways or engineer genome editing tools to increase the efficiency of HR events. Stimulating studies in eukaryotic *Saccharomyces cerevisiae* yeast cells in the late 20th century have shown that yeasts have two mating-types (MAT) that are dictated by the transcription factors called “alpha” and “a” encoded on separate DNA cassettes, and that yeasts can switch between those as efficient/fast as once per generation (Rudin et al., 1989). Subsequent studies revealed not only that this switch was HR-mediated, but also that the dramatically high efficiency in “gene conversion” was achieved via a site-specific homothallic switching (HO) endonuclease (Klar et al., 1984; Kostriken et al., 1983). HO endonuclease is a homing meganuclease that recognizes its 18-bp target site and introduces a staggered-end double-stranded break (DSB), thereby stimulating the HR directed repair between the homologous genes (Nickoloff et al., 1986). By 1989, scientists were able to artificially induce

high efficiency HR events in yeast cells by engineering extrachromosomal DNA fragments carrying HO recognition sites or knocking-in extra HO recognition sites in yeast genome (Rudin et al., 1989). Building up on this observation, the critical proof-of-concept experiments in mammalian cells came from Rouet et al. in 1994 by the Jasin lab. They seminally demonstrated that utilizing meganucleases with such long recognition sequences could be extremely useful for inducing site-directed HR mediated integrations in mammalian cells too (Rouet et al., 1994b). To this end, they employed a mitochondrially encoded 18-bp cutter yeast meganuclease that is structurally related to HO endonuclease, called I-SceI, as HO endonuclease was not commercially available in its purified form (Jasin & Haber, 2016). They showed that the same strategy yeast cells utilize to generate HR events via HO endonuclease at high efficiencies can be artificially reproduced in monkey kidney and mouse cells by either introducing extrachromosomal plasmids with I-SceI sites (Rouet et al., 1994a), or pre-integrating I-SceI sites into their genomes first (Rouet et al., 1994b). I-SceI-induced HR between the endogenous I-SceI sites was 100-fold more efficient compared to the control. Although the overall HR efficiencies were now higher in mammalian cells compared to the gene targeting (~ 1 in 10^2 cells vs 1 in 10^{3-9}), DNA sequencing analysis showed that majority of the repair outcomes were falling into the category of error-prone “nonhomologous recombination” -as was called at the time- as opposed to HR (Jasin & Haber, 2016).

Repair outcomes of DSBs in mammalian cells

Loosing majority of the clones/cells to the error-prone “nonhomologous recombination” as opposed to the HR-mediated precise events was undesirable and warranted immediate attention for finding ways to skew the ratio in favor of HR. Jasin and her team stipulated that using longer homologous arms could be tested to increase HR efficiencies. In order to tackle this

issue, we need to understand some of the major repair pathways active in mammalian cells that would eventually dictate the repair outcomes.

Mammalian cells in each moment encounter numerous mutagenic sources that can induce DNA damage in the form of oxygenation, alkylation, deamination, dimerization and replication errors that can ultimately lead to abasic sites, single-stranded breaks, or DSBs (Cooper, 2000). Among these damaging sources are replication stress, cell metabolism that eventually produces radical oxygen species (ROS), ionizing radiation and UV light, viral infection, or chemotherapy. Therefore, mammalian cells are equipped with multiple repair pathways of mismatch repair (MMR), base-excision repair (BER), nucleotide-excision repair (NER), and DSB repair to cope with the DNA damages that can potentially lead to detrimental outcomes such as apoptosis or tumorigenesis, as genomic instability is one of the hallmarks of cancer (Chatterjee & Walker, 2017; Cooper, 2000; Hanahan & Weinberg, 2011). Moreover, innate processes such as meiosis during gamete production and adaptive immune system at large depends on the controlled DSB formation via MMR to function correctly (Khan & Ali, 2017; G.-M. Li, 2007; Mehta & Haber, 2014). These processes generate the required DNA sequence diversity in lymphocytes by class switch recombination and somatic hypermutation in B-cells, and VDJ recombination in B and T-cells in order to compete with and defend against the invading foreign entities via antibody maturation. It is utmost critical to finely regulate these deliberate DNA damages and their repair outcomes as failed trials could lead to lymphomas (Khan & Ali, 2017; G.-M. Li, 2007).

Observation by Capecchi that any linear DNA molecule in nucleus has a higher chance of getting randomly integrated into genome as opposed to HR-driven on-target integration, and by Jasin that majority of the genomic DSB outcomes are imperfect re-ligation of the cut site strongly suggested that an alternative cellular pathway was much more dominant than HR to

resolve DSBs. Today, we know that a DSB in mammalian cells can be repaired by two competing cellular processes called non-homologous end joining (NHEJ) and homology-directed repair (HDR). The molecular components taking part in NHEJ are active and expressed throughout all cellular stages of G0/G1, S and G2, whereas proteins driving HDR are more effectively expressed at S and G2 phases, thereby limiting its activity to almost only mitotic cells (Moore & Haber, 1996). This is thought to be a natural protective mechanism for cells to minimize the unwanted HDR-driven large chromosomal recombinations outside of S and G2 phases. During S and G2 phases, HDR utilizes the availability of the sister chromatids in close proximity against any potential DNA damage and errors that could arise from the high-demand DNA synthesis in S phase, thereby achieving the high-fidelity duplication of genome. This, however in return leads to the unbalanced outcomes of genome editing attempts severely skewed towards NHEJ even in the presence of a donor\repair DNA template with longer homology arms (although higher frequencies achieved by using longer arms). (K. Li et al., 2014; Lin et al., 2014; Sansbury et al., 2019; H. Yang et al., 2020)

NHEJ repair cascade in mammalian cells starts by detection of the DSB by the Ku70/Ku80 heterodimer protein complex binding to free ends of the DNA on both sides. Ku70/80 then recruits DNA-dependent protein kinase DNA-PKcs to the DSB, whose main role is to phosphorylate and induce conformational changes in the downstream DNA modifying proteins Artemis, CtIP, MRN, Pol λ , Pol μ , and TdT (Jeggo, 1998; O'Driscoll & Jeggo, 2006). While DNA nuclease Artemis, CtIP and MRN induce resection of the free ends, low-fidelity polymerases Pol λ , Pol μ , and TdT competes to fill in the gaps to generate stable dsDNA ends (Sonoda et al., 2006; Van Dyck et al., 1999). Finally, DNA ligase 4, XRCC4 and XLF are recruited to the DSB to promote ligation of the free ends to each other (Chang et al., 2017).

NHEJ mechanism is innately error-prone and can lead to small insertion or deletion (indel) formation at the cut site, or to microhomology-mediated end joining (MMEJ) in the presence of two flanking homologous stretches of sequences on both ends of DSB leading to the semi-precise and predictable loss of sequence in between (Shen et al., 2018). Interesting to note that tumor-suppressor p53, that is responsible for eliminating potentially tumorigenic mammalian cells by inducing apoptosis in the presence of DSB, promotes NHEJ-mediated direct ligation of DSB ends by inhibiting excessive DNA resection (Bunting et al., 2010). NHEJ is active throughout the cell cycles and is the main and dominant repair pathway resolving the DSBs in mammalian cells (Mao et al., 2008).

On the other hand, HDR is promoted efficiently in S/G2 phases of the cell cycle and promoted by the recognition of the DSB by the MRN complex (Chapman et al., 2012). MRN complex consists of MRE11, RAD50, and NBS1 proteins and mainly catalyzes 5' to 3' DNA end resection generating 3' single-stranded overhangs on both sides of DSB (Khan & Ali, 2017). This is done specifically by the endonuclease activity of MRE11 nicking downstream of DSB followed by again its own exonuclease activity resecting until the nick. Exonucleases such as DNA2 and EXO1 are also implicated as key players to further resect starting from nick further downstream generating long-range 3' single-stranded overhangs as long as 1000 nt in length (S. Liu & Kong, 2020). Production of long 3' overhangs suppresses the NHEJ pathway by blocking the access of Ku proteins, and promotes downstream HDR pathways. Tumor-suppressors BRCA1/2 has been shown to promote HDR-mediated DSB resolution by inducing the end resecting nucleolytic processes (C.-C. Chen et al., 2018; Densham & Morris, 2019). The long 3' single-stranded overhangs (called ssDNA nucleofilament) are then bound by Replication protein A (RPA) proteins that protect the free 3' overhangs from exonucleases with 3' to 5' activity and

prevent them from coiling back. RPA proteins are then replaced by recombinase protein RAD51 with the aid of RAD52 promoting either double-strand break repair (DSBR) or synthesis-dependent strand annealing (SDSA) by searching for homologous DNA and initiating strand invasion (Stark et al., 2004). One of the free ssDNA nucleofilaments loaded with RAD51 invades the homologous DNA and forms a region of heteroduplex containing the invading strand and the complementary strand from the donor, forming the Displacement-loop (D-loop). The 3' of the invading strand then acts as a primer for the DNA polymerase delta to synthesize the missing resected complementary strand of the other end (Mehta & Haber, 2014). The replication bubble dissociates, the newly synthesized strand leaves the donor and is captured by the 3' single-stranded tail on the other side of the DSB. The gap and the nicks are then filled and sealed by DNA polymerase alpha and LIG1 and LIG3 respectively (Fan et al., 2004; Mortusewicz et al., 2006; Mortusewicz & Leonhardt, 2007; Puebla-Osorio et al., 2006). SDSA is the canonical repair pathway implicated as the main HDR pathway cell employs to resolve DSB in the presence of an exogeneous repair/donor template as double-stranded DNA (Nambiar et al., 2019). On the other hand, resolution of a DSB by DSBR is more complex, mainly due to the formation of an intermediate four-stranded Holliday Junction between damaged recipient dsDNA and donor homologous dsDNA by simultaneously capturing the other free 3' overhang. Holliday Junction can be resolved in two ways by Resolvase enzyme with nuclease activity leading to a final product with either crossover (like in meiosis) or non-crossover between donor and recipient dsDNA.

Advent of genome editing tools: Zinc-finger Nucleases (ZFN)

Meganucleases that can recognize 12-40 bps work efficiently in nature and increased the efficiency of HR events in artificial settings in mammalian cells. Meganucleases can further

function as a stable monomer, has a relatively compact size (as opposed to other genome editing tools that will be discussed later), and could have significantly high cleavage specificities. However, while having the aforementioned ideal features for a genome editing tool, engineering novel meganucleases with de novo recognition specificities has proven time-consuming and difficult, if not impossible (Adli, 2018). Although tens of novel meganucleases with different recognition sequences from nature and few engineered meganucleases have been catalogued, structure-guided protein engineering remains to be challenging mainly due to the fact that its DNA recognition and cleavage sites being interlaced (Arnould et al., 2011; Smith et al., 2006; Taylor et al., 2012; Zaman et al., 2019). Therefore, an ideal protein-based genome editing tool would have consisted of a modular and separate DNA cleavage and easily engineerable DNA recognition domains.

To this end, the advent of ZFN came from the studies on the structure of zinc finger protein motifs and the restriction enzyme FokI by Chandrasegaran and his colleagues showing that the recognition and cleavage domains of FokI were physically separable (L. Li et al., 1992). Zinc finger is one of the most abundant small eukaryotic protein motifs that can bind to its target 3 bps of DNA via major groove by chelating a zinc ion by coordinating two cysteine and two histidine residues. Efforts of Miller et al. in 1985 revealed that the abundant zinc finger transcription factor IIIA (TFIIIA) consisted of 12 consecutive repeats, each making up a unique zinc finger motif (J. Miller et al., 1985). Characterization of its primary amino acid sequence along with the seminal work of Pavletich and Pabo revealing the structure of a zinc finger bound to DNA in 1991 demonstrated that a single zinc finger is around 30 amino acids in length, 4 of which are the critical residues to mediate the unique DNA target specificity (Pavletich & Pabo, 1991). This made it possible to design programmable chimeric zinc finger proteins fused to non-

specific DNA cleavage domain of FokI restriction endonuclease. The advantages of employing DNA cleavage domain of FokI as the nuclease partner in a genome editing tool is twofold: 1) FokI cleavage domain needs to dimerize to introduce DSB, 2) Amino acid residues at the dimerization interface can be mutated yielding an obligate heterodimer FokI. These two features combined allowed scientists to engineer pairs of three-finger ZFNs each binding to the flanking opposite 9-bp region of a target site, dramatically increasing the specificity and viability of cells by reducing cytotoxic off-target activities (J. C. Miller et al., 2007; Pruett-Miller et al., 2008; Szczepek et al., 2007). In other words, a single ZFN alone bound to its target site is not sufficient to introduce DSB, rendering the system higher in fidelity. This pair design of ZFNs proved to be able to cut lambda phage DNA in vitro, a *Drosophila* gene in vivo, and a human gene in tissue culture (Bibikova et al., 2002; Y. G. Kim et al., 1996; Urnov et al., 2005). These were historical and landmark achievements towards programmable and efficient genome editing.

Engineering of a custom zinc finger for a custom 3 bp target could be achieved by affinity-based selection systems such as phage display libraries (Choo & Klug, 1994; Jamieson et al., 1994; Wu et al., 1995). However, further investigations and engineering efforts suggested that single zinc fingers were not truly modular in nature, but rather could be affected by the adjacent flanking zinc fingers and target nucleotides (Greisman & Pabo, 1997; Wolfe et al., 1999). This rendered the simple “mix and match” fusion of single zinc fingers that can be custom tailored for a target of interest difficult. Although extensive work by Dr. Keith Joung’s lab increased the throughput by which ZFNs could be engineered for a custom target DNA context (Maeder et al., 2008; Sander et al., 2010), certain target DNA sequences are not covered and targetable yet to this day. Moreover, such engineering platforms require deep expertise and

know-how to establish in a lab, rendering it virtually impossible for a common research lab to employ to design their own genome editing tools.

Advent of genome editing tools: Transcription Activator-Like Effector Nucleases (TALEN)

The TALENs arose from studies of gram-negative bacterial plant pathogens of the genus *Xanthomonas*. *Xanthomonas* produce Transcription Activator-Like Effector (TALE) proteins that are secreted into the plant host cell and regulate the transcription of host genome by binding to the upstream of host genes, promoting and “affecting” the bacterial infection (Boch & Bonas, 2010). The initial cloning and sequencing efforts of the first identified 17-repeat TALE protein AvrBs3 proved to be challenging due to its highly repetitive structure back in 1989 but revealed that each TALE unit consisted of ~34 amino acids (Bonas et al., 1989). Later in 2000, another TALE named AvrXa7 was shown to be capable of binding to dsDNA in vitro (B. Yang et al., 2000). Further studies in 2007 by Kay et al. and Römer et al. characterizing the TALE DNA binding sequences paved the way towards the ultimate deciphering of TALE DNA binding “code” in 2009 by the two seminal papers in Science (Boch et al., 2009; Kay et al., 2007; Moscou & Bogdanove, 2009; Römer et al., 2007).

As opposed to the zinc fingers, TALEs were truly modular in nature and each ~34 amino acid unit could independently recognize a single DNA base, albeit larger in size. It was not long before in 2010 Li et al. and Christian et al. published the functional architectures of TALEs fused to the same FokI DNA-cleavage domain used in ZFNs, leading to the generation of TALENs (Christian et al., 2010; T. Li et al., 2011). The initial computational work in 2011 by Cermak et al. batch analyzing the endogenous TALEs found in nature and their cognate binding sequences

laid out the theoretical framework and guidelines for successful/functional TALEN designs while drastically reducing the theoretical targeting scope of TALENs (Cermak et al., 2011). However, later empirical comprehensive work employing GFP reporter system to test tens of novel TALEN pairs showed that the computational guidelines by Cermak et al. mostly failed to predict the in vivo efficiency/outcome of TALENs partially alleviating the concerns regarding the targeting range (Reyon et al., 2012). Similar observations were also reported successfully targeting numerous human endogenous loci in the same work. Moreover, because of the robust and truly modular nature of TALE DNA recognition code, it was dramatically easier than ZFNs to design a novel TALEN for a given new DNA target sequence. However, two TALE repeats targeting two different bases were only 2 adjacent amino acids different from each other, while one needed to design 12-18 repeat-TALENs for successful genome editing applications. Therefore, the actual cloning of TALENs with immensely homologous repeats proved to be difficult for a common lab, although extensive and significant work has been done to devise cloning platforms and strategies by Dr. Keith Joung and his group with the expertise and niche equipment (Reyon et al., 2012).

By 2011, TALENs have proven its versatility and robustness -similar to ZFNs- by successfully engineering the genomes of various organisms including but not limited to zebrafish, rats, cultured human cells, and nematodes (Hockemeyer et al., 2011; Huang et al., 2011; Sander et al., 2011; Tesson et al., 2011; Wood et al., 2011).

The Scissor: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

Although the advent of meganucleases, ZFN, and TALEN significantly increased the feasibility and editing efficiency of genome editing applications, they were still not easily

employable in a cost-effective manner to a common scientific lab. The fundamental underlying reason for this was that all of the concurrent genome editing tools were facilitating the target DNA recognition, and cleavage solely via protein-DNA interactions. Therefore, the common limitation underpinning the inability of the concurrent genome editing tools to be widely adapted across globe was the expertise and know-how required to successfully engineer and clone proteins, namely meganucleases, ZFNs, or TALENs. Thus, a more simple, flexible, and effective reprogrammable genome editing tool would conduct its DNA binding through Watson-Crick base pairing with the target site, thereby bypassing the cumbersome need to re-engineer the protein scaffold.

Although CRISPR genomic locus in prokaryotes (*E. coli* specifically) was first time discovered and identified in 1987 by Yoshizumi Ishino, and in archaea in 1993 by Francisco Mojica, it took roughly another two decades for CRISPR to take to the stage of genome editing upon full elucidation (Y. Ishino et al., 1987; Yoshizumi Ishino et al., 2018; Lander, 2016; F. J. M. Mojica et al., 1993). Both groups recognized the peculiar dyad symmetry (palindrome) containing repeated elements separated by non-repeating unique spacer sequences. Mojica and others had later realized that the peculiar spacer sequences flanked by the palindromic DNA segments were homologous to the parts of invading bacteriophage genomes and externally transferred plasmids (Francisco J.M. Mojica et al., 2005). It had been further demonstrated that challenging a host bacterium carrying a CRISPR locus by a novel bacteriophage not only extended its CRISPR locus by adding a new spacer segment from the new invader's genome, but also conferred resistance to the host against the same pathogen for future challenges (Barrangou et al., 2007; Pourcel et al., 2005). All of these were pointing towards an adaptive immune system functional in prokaryotes, whereby the invading foreign nucleic acid would be subject to an

initial digest by the host CRISPR system (adaption), expression and maturation of the adapted spacer sequence, and finally applying the adaptive immunity by target degradation (interference). Interestingly, upon batch analysis of the spacers found in *S. thermophilus* CRISPR locus, Bolotin and his team realized that the viral protospacers (spacer's cognate name in the invading genome) always had an adjacent downstream sequence what is now called Protospacer Adjacent Motif (PAM) (Bolotin et al., 2005). In the meantime, The Human Genome Project brought about the wealth of genomic sequences of various organisms including many prokaryotes and archaea, enabling the collective search of CRISPR loci. Such analysis revealed that CRISPR loci included numerous open reading frames (ORF) encoding CRISPR-Associated (Cas) proteins. The first experimental proof that the large Cas9 protein was responsible for the adaptive immunity in *E. coli*, characterizing the endonuclease in the cascade, came in 2010 (Garneau et al., 2010). The experiments further demonstrated that Cas9 introduces DSB at the target site specifically 3 nucleotides upstream of the PAM sequence. We now classify Cas9 endonuclease as the effector protein of a Class II Type II CRISPR system (Chylinski et al., 2014; Makarova et al., 2020).

Seminal biochemical studies of labs of Jennifer Doudna and Emmanuelle Charpentier dissected out the delicate machinery behind the CRISPR-Cas9 system from the most studied organism *Streptococcus pyogenes* (SpCas9). Specifically, Emmanuelle Charpentier and her team in 2011 conducted RNA-seq experiments and characterized the two essential RNA components for CRISPR-Cas9 to be active, building upon the previous work of John van der Oost and his group: CRISPR RNAs (crRNAs) and trans-activating CRISPR RNA (tracrRNA) (Brouns et al., 2008; Deltcheva et al., 2011). It became evident that CRISPR-Cas9 was a ribonucleoprotein (RNP) complex (made up of RNA and protein) consisting of a Cas effector protein and CRISPR-related short RNAs that guide the Cas effector to its target via RNA-DNA Watson-Crick base

pairing, granting itself the RNA-Guided Nucleases (RGN) title. Most impressively, collaborative efforts of Doudna and Charpentier in 2012 showed that it was feasible to fuse the two CRISPR RNAs into a single synthetic guide that is now called a single guide RNA (sgRNA) or guide RNA (gRNA), eventually granting them the Nobel Prize in Chemistry in 2020 (Jinek et al., 2012). In short, Cas9 gRNA consists of a ~20 nt spacer sequence on its 5' end serving as the programmable address to the target locus, and a downstream ~100 nt constant gRNA scaffold region (handle) that Cas9 interacts with. They further displayed that Cas9 has two distinct protein domains, namely HNH and RuvC, that nicks the complementary and non-complementary sites, respectively. It was also possible to introduce point mutations in HNH and RuvC cleavage domains and generate impaired (nickase) or dead (inactive) Cas9 variants. Finally, it was not long before groups of Feng Zhang and George Church in parallel adapted the CRISPR-Cas9 platform into mammalian cells (human and mouse) and showed that it can be reprogrammed to target multiple loci (Cong et al., 2013; Mali et al., 2013).

The scientific journey elucidating the fine mechanism behind the CRISPR machinery of nearly three decades have now granted the scientific community at large the flexibility to design their own genome editing experiments at an unprecedented level of pace, cost, and ease of design in everyday experiments. This became possible mainly by the virtue of the ease of cloning new gRNA molecules without the need to re-engineer new CRISPR Cas effector proteins for each experiment, unlike the previous platforms of meganucleases, ZFNs, and TALENs. This in return had tremendous impact in both academic and industrial spheres leading to the burst of academic publications, spin-off startup companies, therapeutic applications, and clinical trials across the globe. We are still living in the era of CRISPR, and more precise editing tools based off of CRISPR, that has been tested to be functional across tens of species and organisms.

The Pen: 2nd generation of CRISPR-based genome editing tools: Advent of Base Editors

Although the CRISPR genome editing tools has clearly surpassed the other concurrent genome editing tools and lowered the climax of the mountain of feasibility and ease, in the end of the day it was still generating the same outcome of its predecessors at the target DNA, namely DSB, albeit more robustly and at higher efficiencies in a context dependent manner. In other words, CRISPR-Cas nuclease effectors have been robust tools to mainly knockout genes or disrupt coding or noncoding elements mainly by introducing indel mutations, thus frameshifts or nonsense mutations in the case of ORFs, but was still falling short to effectively introduce precise alterations across cell lines and organisms. Therefore, the end users were still solely dependent on the ineffective HDR machinery along with ssDNA or dsDNA repair/homology templates exogenously provided in order to install precise and user-defined alterations into target genome of interest. As was discussed earlier, this has been the main bottleneck and drawback of the genome editing tools as the nuclease-induced DSBs are resolved dominantly by the error-prone NHEJ pathway even in the presence of the HDR machinery, while HDR is virtually nonexistent in nondividing post-mitotic cells. Moreover, and most importantly, CRISPR-Cas9 nuclease has been shown to be capable of inducing unintended on-target effects of chromothripsis, megabase-scale deletions, translocations, inversions and complex rearrangements that can have serious adverse consequences for patients (Leibowitz et al., 2021). Therefore, there was still the unmet need to devise a genome editing tool that can efficiently reverse pathogenic point mutations in human without inducing DSBs, as the majority of the pathogenic alterations catalogued in human are single nucleotide variations (SNV) (Landrum et al., 2016; Rees & Liu, 2018).

To this end, the second revolution within the CRISPR era happened at the advent of the second generation of CRISPR-based genome editing tools called Base Editors (BE) upon the creative work of David Liu and his group at Harvard Medical School. BE is a CRISPR-based fusion protein that can precisely install alterations by chemically altering nucleotide(s) at the target site without inducing DSB, even at a single nucleotide resolution in a context dependent manner. The first canonical base editor engineered in 2016 by Komor et al. employed a nickase SpCas9 (D10A mutant in the RuvC cleavage domain) fused to the rat cytosine deaminase APOBEC1 (rAPOBEC1) and further fused to Uracil DNA glycosylase inhibitor (UGI) from bacteriophage PBS1 (Komor et al., 2016). This architecture called BE3 became one of the early canonical Cytosine BEs (CBE) in the field and is thought to work in the following manner: 1) nCas9 continuously scans the genome for target sites and cognate PAM elements, 2) upon base pairing of spacer-protospacer between gRNA and target DNA, the R-loop is generated, 3) as gRNA exclusively base pairs with only the target strand, the nontarget strand is partially exposed as ssDNA in the solution, 4) rAPOBEC1 gains access to the ssDNA exposed to potentially deaminate the target cytosine(s) in the editing window into uracil(s), 5) nCas9 in parallel nicks the target strand triggering BER or MMR-induced 3' to 5' trimming & removing the original guanine(s) found in the opposite strand of the newly formed uracil(s), 6) DNA polymerases re-fill the newly generated gap while placing adenine(s) instead of the original guanine(s) opposite to the uracil(s) leading to permanent mutation of the deaminated residues, 7) UGI continuously inhibits cellular uracil DNA glycosylase (UNG) in the vicinity thereby actively preventing BER machinery from accessing the uracil(s). This mechanism of action not only allows for high efficiency (>70%) and pure C-to-T edits (thereby G-to-A edits in the opposite strand) in a context dependent manner, but also virtually eliminates the indel byproducts by bypassing the

NHEJ altogether via not employing an active/nuclease Cas9 that can induce DSBs. Later, Komor et al. and Koblan et al. improved the architecture of BE3 rendering it more efficient and more precise (less byproducts/purer results) by appending an extra UGI domain, optimizing the codon usage and nuclear localization signal, respectively, generating BE4 and BE4max canonical CBEs (Koblan et al., 2018; Komor et al., 2017). Interestingly, Wang et al. in 2017 demonstrated that an unlinked/free UGI overexpression along with the remaining nCas9-deaminase construct (via P2A self-cleaving system) could significantly increase the base editing efficiencies as an alternative to appending multiple UGI moieties (L. Wang et al., 2017). Extensive further work by Jia Chen and his team also led to the engineering of LbCpf1-based CBE genome editing tools expanding the functional toolbox and targeting range from GC-rich (PAM preference of SpCas9) to AT-rich (PAM preference of Cpf1/Cas12a) target loci (X. Li et al., 2018).

Since then, CBEs has been employed widely across globe to successfully introduce precise alterations into human, rat, mouse, rabbit, and zebrafish embryos/zygotes, and mouse inner ear cells in vivo (K. Kim et al., 2017; Liang et al., 2017; Z. Liu et al., 2018; Tanaka et al., 2018; Yeh et al., 2018; Y. Zhang et al., 2017). Although, CBEs in theory can be harnessed to reverse 300 to 900 clinically relevant human point mutations by their C-to-T activities, it still covers only 14% of the total pathogenic SNVs in human that needs to be reversed (Komor et al., 2016; Rees & Liu, 2018). As one of the most common DNA damages in our cells is the spontaneous deamination of cytosines into uracils, this leads to the unbalanced accumulation of pathogenic point mutations that needs an A-to-G transition base editor to fix, constituting 47% of the total mutations. To this end, David Liu and his lab concentrated their subsequent efforts in 2017 on engineering an Adenine Base Editor (ABE) that can efficiently install A-to-G transition edits at target sites (Gaudelli et al., 2017).

It is important to note that the generation of a CBE transition editor without extensive rounds of protein engineering efforts was possible mainly because of the presence and knowledge of cytidine base editors that can act on ssDNA as substrate. However, that was not the case for the ABE efforts, as there was not any known enzyme that could deaminate adenine as part of a ssDNA substrate. To this end, Gaudelli et al. evolved a functional deoxyadenosine deaminase that can work on ssDNA starting from the *E. coli* tRNA-specific adenosine deaminase (TadA) through a series of positive selection in bacteria (Gaudelli et al., 2017). 16 rounds of protein engineering and directed evolution efforts resulted in the mutant TadA (TadA*) that can deaminate adenines into inosines in ssDNA in mammalian cells. Similar to CBEs, fusion of impaired Cas9 (nickase Cas9) to TadA* via a peptide linker led to high efficiency edits with high product purity in context dependent manner. Noteworthy is that, unlike CBEs, ABEs has proven to be not in need of an additional inhibitor of the cellular repair machinery -like UGI in CBEs- to work efficiently in human cells, reflecting the fact that A-to-I spontaneous deamination is rarer in human cells and therefore its repair machinery is not as robust. Additionally, this less acute response of cells to the A-to-I deamination leads to even lower indel byproduct frequencies observed compared to CBE editing. Although an inosine can be recognized as multiple bases in the wobble position of a tRNA, it is specifically misrecognized as a guanine during DNA polymerization by polymerases (Alseth et al., 2014; Srinivasan et al., 2021). Therefore, ABE-induced A-to-I deamination in the target window has proven effective at installing A-to-G transition alterations across numerous sites in various cell lines and organisms (Hua et al., 2018, 2019; Kang et al., 2018; C. Li et al., 2018; Ma et al., 2018; Ryu et al., 2018; Yan et al., 2018).

The Word Processor: 3rd generation of CRISPR-based genome editing tools: Advent of Prime Editor

While BEs indisputably raised the bar of precision genome editing significantly to an unprecedented level, BEs were still far from being the most ideal genome editing tools. Specifically, BEs were inherently limited by the targeting range of the CRISPR effector protein it was built on, exacerbated by the narrow editing window the deaminase could access to edit. Moreover, in the presence of more than one target base along with the desired target base, BEs could imprecisely deaminate and edit multiple undesired target bases (bystander editing), although significant work has been done to increase the precision and lower bystander editing of BEs (Gehrke et al., 2018; Y. B. Kim et al., 2017). Most importantly, BEs are limited to installing C-to-T and A-to-G transition alterations robustly. Combining all of the aforementioned problems, an ideal genome editing tool would have more flexible targeting scope and editing window, would be able to precisely install a desired edit with minimal byproducts and bystander editing, and install virtually any type of genomic alteration (insertion, deletion, substitution, combinations of thereof, structural variations) as an edit.

The recent advent of the CRISPR-based Prime Editor (PE) genome editing tool by David Liu and his team was so far the longest step towards the right direction addressing the needs of an ideal genome editing tool (Anzalone et al., 2019). The canonical PE (called PE2) RNP is composed of a nickase Cas9 (H840A in HNH cleavage domain), a pentamutant hyperactive reverse transcriptase (RT) from Moloney murine leukaemia virus (MMLV), and a gRNA with a 3' extension with a particular design called prime editor gRNA (pegRNA). 3' extension of the pegRNA encodes the reverse transcription template (RTT) and primer binding site (PBS) downstream of the conventional gRNA scaffold from 5' to 3' respectively. The mechanism of

action of PE2 is thought to happen in the following manner: 1) nCas9 binds to its cognate target site similar to the previous technologies, 2) R-loop is formed upon spacer-protospacer base pairing, 3) nCas9 nicks the nontarget strand (as opposed to the target strand in the case of BEs), 4) 3' end of the nicked and freed nontarget ssDNA base pairs with the complementary PBS sequence found on the 3' most end of pegRNA, 5) Formed RNA-DNA hybrid serves as the primer for MMLV-RT to polymerize a flap 5'-to-3' using RTT as the template, 6) Extended 3' flap competes with the nonextended/original 5' flap, 7) 3' newly synthesized flap wins over at some frequency and the opposite strand goes through DNA repair (possibly via FEN1 or EXO1 endo/exonuclease activities) leading to permanent installation of the edit (Keijzers et al., 2015; Y. Liu et al., 2004).

A recent study has fused chromatin-modulating peptides to PE construct and employed proximal dead sgRNAs in parallel (both ideas stem from original work with nuclease Cas9 to increase its cleavage activity) and showed that this enhanced PE system has higher activity, especially when targeted to heterochromatin (closed chromatin) regions, presumably by opening up the target region for PE to access (F. Chen et al., 2017; Ding et al., 2019; Park et al., 2021). Hypothesizing that 3' end of pegRNA is vulnerable to RNase degradation in human cells, it has been shown that extending the 3' end of pegRNA with a refractory RNA structural motif, a pseudoknot, can increase the stability of pegRNA (now called epegRNA) and thereby increase the PE editing efficiencies (J. W. Nelson et al., 2021). David Liu and his team further conducted a CRISPRi knockdown screen for the DNA repair enzymes and identified that mismatch repair pathway is mainly responsible for reversing the prime edit introduced and thereby reduce on-target efficiencies. To overcome this, they fused a dominant negative mutant form of an MMR protein (MLH1) to PE architecture to suppress local MMR transiently, generating PE4 and PE5

with substantially increased PE editing levels (P. J. Chen et al., 2021). Moreover, they introduced the previously identified mutations that add more positive charges to SpCas9's DNA-interaction interface that enhance SpCas9's nuclease activity, optimized codon usage of RT, peptide linker length, nuclear localization signal (NLS), and employed epegRNA, generating PE4max and PE5max with even increased PE editing levels, up to 72-fold higher activity in context dependent manner (P. J. Chen et al., 2021; Spencer & Zhang, 2017). Addressing one of the limitations of PE system -that is the upper limit to an indel that can be introduced-, recent efforts called PEDAR, twinPE, PASTE and PRIME-Del have employed dual PEs in parallel targeting adjacent sites to install precise large genomic deletions (up to 10 kb) and replacements (Anzalone et al., 2021; Choi et al., 2021; Ioannidi et al., 2021; Jiang et al., 2021). TwinPE and PASTE platforms also showed that PE-mediated integration of Bxb1 recognition sites in a programmable manner enables site-specific serine recombinases to install precise large genomic insertions and inversions that are up to ~40 kb (Anzalone et al., 2021; Ioannidi et al., 2021).

Fundamentally, by solely altering the RTT sequence, PE in theory can install any small indels, any transition or transversion substitutions, and combinations thereof, thereby making PE the most versatile and flexible genome editing tool that has ever been engineered in the history of genome editing. It is also important to note that PE could also be the most accurate CRISPR-based genome editing tool, if not the most robust/active, due to the fact that a potential off-target locus would have to pass through three checks (base pairing between spacer and protospacer, between PBS and 3' nontarget nicked and freed ssDNA, between RTT-templated newly polymerized 3' flap and endogenous sequence downstream of the nick on the target strand) as opposed to a single check (base pairing between spacer and protospacer) for a conventional CRISPR-based RGN or BE. This is confirmed by multiple reports in multiple settings

demonstrating extremely low off-target activities by PE tools (Jin et al., 2021; D. D. Y. Kim et al., 2020; Schene et al., 2020).

Nonetheless, PE has its own limitations such as its large size and can suffer significantly from low activity in a cell line- and organism-dependent manner, especially while installing single nucleotide substitutions (Kurt et al., 2020; Petri, Zhang, et al., 2021). Although RNA-seq experiments demonstrated moderately minimal perturbations at transcriptome level upon PE overexpression, it is of utmost importance to design critical experiments to test potential negative effects of the exogenous overexpression of a viral hyperactive RT found in the PE architecture, especially given the fact that more than half of human genome consists of transposable “jumping” elements some of which rely on an intermediate RT-mediated reverse transcription to transpose itself into new loci (Anzalone et al., 2019; Kramerov & Vassetzky, 2011). Particularly, endogenous RT activity of retrotransposon long interspersed element 1 (LINE1) in human genome can lead to retrogene production, that is reverse transcription and genomic integration of endogenous mRNAs, along with integration of itself, potentially leading to a list of human diseases (Elbarbary et al., 2016). Moreover, although remains controversial, a recent study by Rudolf Jaenisch at Whitehead Institute/MIT demonstrated that even SARS-COV-2 genomic elements can be integrated stably into human genome by exogenous LINE1 RT overexpression (L. Zhang et al., 2021). If anything, all premises combined strongly warrants to develop sensitive ways to test how the genomic pattern of LINE/SINE/Alu retrotransposons change upon prime editor treatment, especially of Alu elements that are enriched at genic regions (S. Kim et al., 2016).

Genome-wide and Transcriptome-wide Off-target Editing of Genome Editing Tools

Specificity is a crucial aspect of genome editing tools that must be vetted thoroughly in order to characterize the context-dependent genome-wide specificity profiles thereof. Specificity for the engineered nucleases of TALENs, ZFNs, and CRISPR-based RGNs can be described as the innate tendency towards inducing an unwanted DSB at an off-target locus that is partially homologous to the on-target site. Comprehensive characterization of the specificity profile of a genome editing tool is utmost critical, especially in the clinical context, considering the fact that many *ex vivo* and *in vivo* therapeutic applications entail treatment of hundreds of millions to billions of cells drastically increasing the odds of generating the rare clones with deleterious or oncogenic mutations (Chu et al., 2021; Frangoul et al., 2021; Koblan, Erdos, et al., 2021; Leibowitz et al., 2021; Musunuru et al., 2021).

Engineered nucleases of TALENs and ZFNs engage with their target sites via noncovalent protein-DNA interactions at the major groove. Therefore, exceeding the energetic threshold of binding at any genomic loci potentially constitutes an off-target site. Correspondingly, TALENs and ZFNs have the potential to recognize and interact with off-target loci that are up to 12 and 3 base pairs different from the intended on-target sites, respectively (Gabriel et al., 2011; Hockemeyer et al., 2011; Osborn et al., 2013; Pattanayak et al., 2011, 2014). Splitting the homodimer FokI unspecific nuclease into a pair of TALENs or ZFNs, and introducing mutations on the interaction interface of each half of FokI domains conferring obligate heterodimerization properties has been employed to mitigate off-target editing of the engineered nucleases (Doyon et al., 2010; Gupta & Musunuru, 2014; J. C. Miller et al., 2007; Szczepek et al., 2007). As an alternative strategy, targeted replacement of cationic amino acids of arginine and lysine with uncharged amino acid glutamine in the C-terminus of TALENs

improves the specificity profiles by reducing the excessive nonspecific DNA-binding energy while preserving the on-target activity levels (Guilinger et al., 2014). This proves that it is possible to increase the fidelity of a genome editing tool by hampering the nonspecific interactions with the negatively charged backbone of target DNA molecule, while still largely retaining the on-target activity.

CRISPR-based RGNs, on the other hand, bind to their target sites in a threefold mechanism. First, Cas9 scans the genome for the cognate PAM sequences via its PAM-interacting Domain (PID), second, Watson-Crick base pairing between the guide RNA and protospacer or an off-target site takes place, third, Cas9 undergoes a conformational change that activates its endonucleolytic cleavage (J. S. Chen et al., 2017). Therefore, specificity profile of RGNs is contingent on both noncovalent protein-DNA (PID-to-PAM and other nonspecific DNA interactions) and RNA-DNA (gRNA-target site) interactions. Initial ChIP-seq experiments showed that the degree of homology between a target site and Cas9 is less stringent presumably during the PAM-scanning period, with as high as 10 mismatches out of 20 nucleotides are accommodated for target binding (Kuscu et al., 2014). However, later work proved that only a small fraction of this superset is actually bonafide off-target sites that are cleaved at detectable levels (Tsai et al., 2014). Nuclease Cas9 has been demonstrated to be capable of introducing DSBs at the off-target sites with up to 6 mismatches in vivo and 7 mismatches in vitro, especially in the PAM-distal region (Pattanayak et al., 2013; Tsai et al., 2014). Similar to TALENs, structure-informed protein engineering led to the advent of high-fidelity nuclease Cas9 variants by diminishing the nonspecific protein-DNA interactions or altering the conformational thresholds, albeit usually at the context-dependent cost of on-target activity (J. S. Chen et al., 2017; N. Kim et al., 2020; Kleinstiver et al., 2016; Rees & Liu, 2018; Slaymaker et al., 2016).

By the same token, truncation of the 20 nucleotide spacer region of a guide RNA down to 17 or 18 nucleotides have also demonstrated substantially improved CRISPR-based RGN specificity profiles, suggesting that there could be excessive binding energy that can be reduced at the RNA-DNA interface as well (Fu et al., 2014).

In the case of CRISPR-based BEs, BEs not only possess the inherited specificity features of the parental CRISPR effector it is built upon, but also have its distinct modalities that affect the specificity profiles thereof. Specifically, on top of the CRISPR-guided off targets due to the protein-DNA and RNA-DNA interaction kinetics, there are also potential unwanted CRISPR-independent spurious off-target activities at the genomic and transcriptomic levels that can be induced by the deaminase or other DNA repair partners in the construct (Doman et al., 2020; Grünewald, Zhou, Garcia, et al., 2019; Grünewald, Zhou, Iyer, et al., 2019; Jin et al., 2019; Yu et al., 2020; E. Zuo et al., 2019). There has been significant protein engineering efforts, especially by Dr. Keith Joung's lab, towards generating higher-fidelity BEs. Incorporating R33A and other amino acid alterations into rat APOBEC1 component of CBE, authors were able to generate SECURE-CBE variants with significantly lower transcriptome-wide off-target activities (Grünewald, Zhou, Garcia, et al., 2019). Likewise, the same group engineered a mini-ABE variant with substantially diminished transcriptome-wide off-target RNA editing activities by removing the wildtype TadA deaminase domain and introducing V82G alteration into the remaining engineered TadA domain (Grünewald, Zhou, Iyer, et al., 2019). Later, SECURE-CBE variants have been shown to also have favorable CRISPR-dependent DNA off-target activities at 20 sites tested (Doman et al., 2020). Presenting an alternative strategy, Gaudelli and her team at Beam Therapeutics has scanned the protein databases and mined out alternative APOBEC paralogues with innately lower spurious DNA and RNA off-target profiles as distinct CBE

deamination partners (Yu et al., 2020). Most recently, Jia Chen and his group engineered a robust cytosine base editing system called transformer BE (tBE) that displays background level DNA and RNA off-target editing. The system employs a distinct cytosine deaminase with a self-regulatory/inhibitory moiety that would only be cleaved off at the on-target site, transforming the tBE into active state only at the desired locus, eliminating the genome-wide and transcriptome-wide off-target mutations (L. Wang et al., 2021).

In Cellular Synthetic Off-target Libraries

Translation of the genome editing tools into clinic mandates comprehensive characterization of their specificity profiles and the intended and unintended editing outcomes at the on-target and off-target sites. This is corroborated by the frequent necessity of treating hundreds of millions to billions of cells *in vivo* or *ex vivo* as part of gene therapy programs employing the genome editing tools (Frangoul et al., 2021; Mancuso et al., 2020). In other words, it is possible to generate unintended clones of cells carrying the rare deleterious or tumorigenic off-target mutations that could lead to adverse outcomes in clinical trials similar to the retrovirally-induced T-cell leukemia cases in past (Hacein-Bey-Abina et al., 2003).

The off-target analysis methods can be broadly categorized into two complementary categories by two approaches: 1) *in vitro* or *in vivo*, 2) natural or synthetic. *In vitro* natural methods leverage the controllable biochemical environment, such as the concentrations of the purified genomic DNA and the editor, and molecular enrichment steps to substantially increase the sensitivity (Bao et al., 2020; Tsai & Joung, 2016). However, the purified genomic DNA employed in *in vitro* techniques is stripped from its chromatin structures such as nucleosomes (histones) that form the epigenetic landscape -although an effort was put forth to directly work

with chromatin DNA with histones intact-, and therefore the output off-target sites make up a ranked nominated superset that needs to be validated in cells/in vivo (Bao et al., 2020; Horlbeck et al., 2016; D. Kim & Kim, 2018). In vivo validation of the nominated sites above the determined cutoff by targeted deep sequencing are then promoted to bona fide off-target sites (Atkins et al., 2021). In vivo natural off-target platforms, on the other hand, captures the endogenous biochemical and epigenetic environment within the innate genomic space in the nucleus during Cas9's activity and lifetime at the cost of sensitivity (Bao et al., 2020; X. Chen et al., 2017). This is particularly important given the observation that Cas9 is more active at the open chromatin (euchromatin) regions such as active enhancers and promoters, and actively transcribed genes (Lazzarotto et al., 2020).

In vitro and in vivo synthetic off-target platforms employ synthetic custom libraries to expand the targeting space outside of natural genomic sequence borders. In vitro synthetic off-target platforms are similar in nature to the aforementioned in vitro natural methods with the exception of the source and complexity of the target DNA molecules. In vitro efforts to comprehensively characterize the specificity profiles of ZFNs, TALENs, and CRISPR-Cas9 started with highly complex and randomized off-target libraries with as many members as $>10^{11}$ that were generated by synthetic randomized short primers-driven PCR (Pattanayak et al., 2011, 2013, 2014). Later efforts generated complex in vitro nonrandom off-target libraries for CRISPR-based genome and base editing technologies (up to 80k members with up to 6-8 mismatches) that are designed against single guide RNAs (Petri, Kim, et al., 2021).

In vivo synthetic off-target platforms, on the other hand, integrate the artificial library into the genomes of living cells, thereby generating the custom targeting space of interest in nuclei. In vivo synthetic off-target platforms also permit alteration of cell types while keeping

the library constant. It is further possible to subsequently introduce the genome editing tool of interest in the forms of plasmid DNA, purified mRNA, or RNP complexes. This approach also opens the doors of off-the-shelf production of off-target cells carrying a genome-integrated synthetic library that can be frozen and transferred across labs with ease. Therefore, genome-integrated in cellular off-target libraries are extremely powerful and flexible tools to engineer artificial on-target or off-target sites that normally do not exist in the genomic space and yet leverage the natural biochemical environment that genome editing tools would interact with in clinical applications. Nonetheless, it must be noted that the artificial libraries are exogenously supplied via retroviral integration or transposase transposition and mostly integrated into DNase-sensitive open and accessible parts of the chromatin with active or bivalent histone modifications (Schröder et al., 2002; Yoshida et al., 2017).

Specificities and activity levels of range of genome editing tools can be assessed and scored by massively parallel quantification of editing outcomes of thousands of on-target, off-target, or combination thereof integrated-sites (Arbab et al., 2020; N. Kim et al., 2020; Tycko et al., 2018). Moreover, the uniform data generated from a single experimental platform paved the path towards improved understanding of the dynamics governing the on-target and off-target interactions, and machine learning-based off-target scoring and on-target outcome predicting algorithms (Marquart et al., 2021).

To this end, Patrick Hsu and his team set out to characterize the specificity profile of *Staphylococcus aureus* Cas9 (SaCas9) nuclease by lentivirally integrating a complex library of 88,692 self-targeting members designed for 73 different gRNAs/spacers, containing all single and double mismatches and DNA and RNA bulges as off-target sites (Tycko et al., 2018). Hyongbum Kim and his group designed strictly on-target pairwise libraries with ~15,000

members for both AsCpf1 and SpCas9 nucleases and trained deep learning algorithms to predict the outcomes of CRISPR nuclease effectors at a given site (H. K. Kim et al., 2018, 2019). In parallel, Richard Sherwood and his team reported their neural-network-learned editing outcome predictions for SpCas9 nuclease using an on-target library with ~2000 members, seminally demonstrating that microhomology-mediated end-joining (MMEJ) is a major determinant in predicting SpCas9 nuclease outcomes (Shen et al., 2018). In this work, Shen et al. employed a Tol2 transposome system to integrate their library members into the cells, as opposed to the conventional lentiviral integration. Later in 2020, Hyongbum Kim and his group assigned specificity scores to a diverse set of engineered high-fidelity variants of SpCas9 nuclease that show lower off-target activities. To this end, they employed three complex self-targeting lentiviral libraries with 11k, 23k and 7k pairs of sgRNA and their cognate off-target sites to develop deep learning models to predict activity levels of SpCas9 variants at a given target locus (N. Kim et al., 2020).

Recently, in vivo synthetic on-target library platforms have been started to be employed to also assess the specificities and activities of base editing technologies in cells. The platform called TRAP-seq by Yonglun Luo and his team generated an on-target lentiviral library with 12k members designed to test CBE (BE4-Gam) and ABE (ABE7.10) editors along with the conventional SpCas9 nuclease with the goal of revealing features affecting editing dynamics, and designing machine learning-based prediction tools for designing efficient and accurate guide RNAs (Xiang et al., 2020). David Liu and his group later published a Tol2-based on-target library with up to ~40k members and ran it with 6 CBE and 2 ABE variants (BE4max, BE4-CP, evoA-BE4, AID-BE4, CDA1-BE4max, eA3A-BE4, ABEmax, ABE-CP) and Liu CGBE variants later in 2021 (Arbab et al., 2020; Koblan, Arbab, et al., 2021). This comprehensive BE-based

library work led to the deep conditional autoregressive machine-learning model called BE-Hive that can predict base editing yield and bystander editing patterns in a cell type specific manner. Finally, two more machine learning models (attention-based model & two-hidden layer convolutional neural network framework) that are capable of predicting base editing outcomes, called BE-DICT and DeepBaseEditor, were trained employing ABE7.10, ABEmax, ABE8e, BE4, CBE4max, and Target-AID base editor tools with on-target lentiviral libraries containing around 20k and 15k members, respectively (Marquart et al., 2021; Song et al., 2020).

Chapter 2: Engineering DNA Transversion C-to-G Base Editors (CGBE)

Ibrahim C. Kurt^{1,2,3,5}, Ronghao Zhou^{1,2,5}, Sowmya Iyer¹, Sara P. Garcia¹, Bret R. Miller^{1,2}, Lukas M. Langner^{1,2}, Julian Grünewald^{1,2,4,6}, J. Keith Joung^{1,2,4,6}

¹Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA

²Center for Cancer Research and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA

³Biological Sciences in Public Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁴Department of Pathology, Harvard Medical School, Boston, MA, USA

⁵These authors contributed equally to this work

⁶These authors jointly supervised this work

This chapter contains work from the manuscript “CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells”. The text and figures were modified to fit the format of this dissertation. The manuscript was co-written by Julian Grunewald, Ibrahim C. Kurt, Ronghao Zhou, and J. Keith Joung. Ibrahim C. Kurt later amended the discussion paragraphs with the recent publications and elaborated them into the “Discussion & Future Directions” subchapter for the thesis. Specific author contributions for each figure are broken down under “Author Contributions” subsection below.

Acknowledgements

This work was supported by the National Institutes of Health (RM1 HG009490 and R35 GM118158) and by the Defense Advanced Research Projects Agency Safe Genes Program (HR0011-17-2-0042). J.G. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 416375182. L.M.L. was supported by a Boehringer Ingelheim Fonds MD fellowship. J.K.J. is additionally supported by the Robert B. Colvin, M.D. Endowed Chair in Pathology and the Desmond and Ann Heathwood MGH Research Scholar Award. We thank B. Kleinstiver and R. Walton for providing a codon-optimized version of VRQR-SpCas9 (D10A). We thank M.K. Clement for technical advice, K. Petri, P. Cabeceiras, J. Angstman, J. Hsu, C. Lareau, and M. Aryee for discussions and technical advice, and L. Pottenplackel for help with editing the manuscript. J.K.J. dedicates this work to the memory of Ki Sub Joung with gratitude for his lifelong advice, support, and inspiration.

Author Contributions

S.I. and S.P.G. contributed equally to this work and are co-second authors. Wet lab experiments were performed by I.C.K., R.Z., B.R.M., and L.M.L., while S.P.G. and S.I. performed computational analysis of the data. I.C.K., R.Z., J.G., and J.K.J. conceived of and designed the study. J.G. and J.K.J. supervised the work. I.C.K., R.Z., J.G., and J.K.J. co-wrote the initial manuscript draft and all authors contributed to the editing of the manuscript. I. C. K. amended the discussion paragraphs with the recent publications and elaborated them into the “Discussion & Future Directions” subchapter for the thesis. Table below specifically breaks down the contribution of each author for the raw data generated for the given figure(s):

Fig. 1	R.Z.
Fig. 2	I.C.K., R.Z.
Fig. 3	B.R.M.
Fig. 4	N/A (Koblan, Arbab, et al., 2021)
Fig. 5*	I.C.K.
Ext. Fig. 1	R.Z.
Ext. Fig. 2	R.Z.
Ext. Fig. 3	I.C.K., R.Z.
Ext. Fig. 4	I.C.K., R.Z.
Ext. Fig. 5	I.C.K., R.Z., B.R.M.
Ext. Fig. 6	L.M.L.
Ext. Fig. 7 a-e	I.C.K., R.Z., B.R.M., L.M.L.
Ext. Fig. 7 f	I.C.K.
Supp. Fig. 1	R.Z.
Supp. Fig. 2	B.R.M.
Supp. Fig. 3	I.C.K.
Supp. Fig. 4	R.Z.
Supp. Fig. 5	I.C.K., R.Z., B.R.M., L.M.L.

*Plot generated from bioinformatics analysis; no biological data was generated.

Competing Financial Interests Statement

J.K.J. has financial interests in Beam Therapeutics, Editas Medicine, Excelsior Genomics, Pairwise Plants, Poseida Therapeutics, Transposagen Biopharmaceuticals, and Verve Therapeutics (f/k/a Endcadia). J.K.J.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. J.K.J. is a member of the Board of Directors of the American Society of Gene and Cell Therapy. J.G., I.C.K, R.Z., and J.K.J. are co-inventors on patent applications that have

been filed by Partners Healthcare/Massachusetts General Hospital on engineered C-to-G transversion base editors and other related base editor technologies.

CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells

Abstract

CRISPR-guided DNA cytosine and adenine base editors (CBEs and ABEs) are widely used for many applications (Gaudelli et al., 2017; Komor et al., 2016; Nishida et al., 2016; Rees & Liu, 2018) but primarily create DNA base transitions (i.e., pyrimidine-to-pyrimidine, or purine-to-purine). Here we describe the engineering of two base editor architectures that can efficiently induce targeted C-to-G base transversions, with reduced levels of unwanted C-to-W (W = A or T) and indel mutations. One of these C-to-G base editors (CGBE1), consists of an RNA-guided Cas9 nickase, an *E. coli*-derived uracil DNA N-glycosylase (eUNG), and a rat APOBEC1 cytidine deaminase variant (R33A) previously shown to have reduced off-target RNA and DNA editing activities (Doman et al., 2020; Grünwald, Zhou, Garcia, et al., 2019). We show that CGBE1 can efficiently induce C-to-G edits, particularly in AT-rich sequence contexts in human cells. We also removed the eUNG domain to yield miniCGBE1, which reduced indel frequencies but only modestly decreased editing efficiency. CGBE1 and miniCGBE1 enable C-to-G edits and will serve as a basis for optimizing C-to-G base editors for research and therapeutic applications.

Our efforts to develop a C-to-G base editor were motivated by our previous observation that the A-to-G editor ABEmax could induce unexpected C-to-G edits at sites in which a C was present at position 6 of the protospacer (Grünwald, Zhou, Iyer, et al., 2019) (numbering starting from the position most distal to the protospacer adjacent motif [PAM]). (These results were recently confirmed by another group who reported C-to-G edits for Cs at positions 5, 6, and 7 of the protospacer by the same ABE (H. S. Kim et al., 2019).) We hypothesized that the heterodimeric TadA-based adenosine deaminase domain present in ABEmax might deaminate the C at position 6 to a U with subsequent creation of an abasic site by cellular UNG (base excision repair) followed by the preferential introduction of a G at that position by an as-yet undefined mechanism (**Fig. 1a**). Consistent with this, we found that the addition of two uracil glycosylase inhibitor (**UGI**) domains to ABEmax resulted in a reduction of C-to-G edits and indels, and an increase in C-to-T edits (**Extended Data Figs. 1 and 2a, Supplementary Table 1**), presumably due to reduced efficiency of uracil excision.

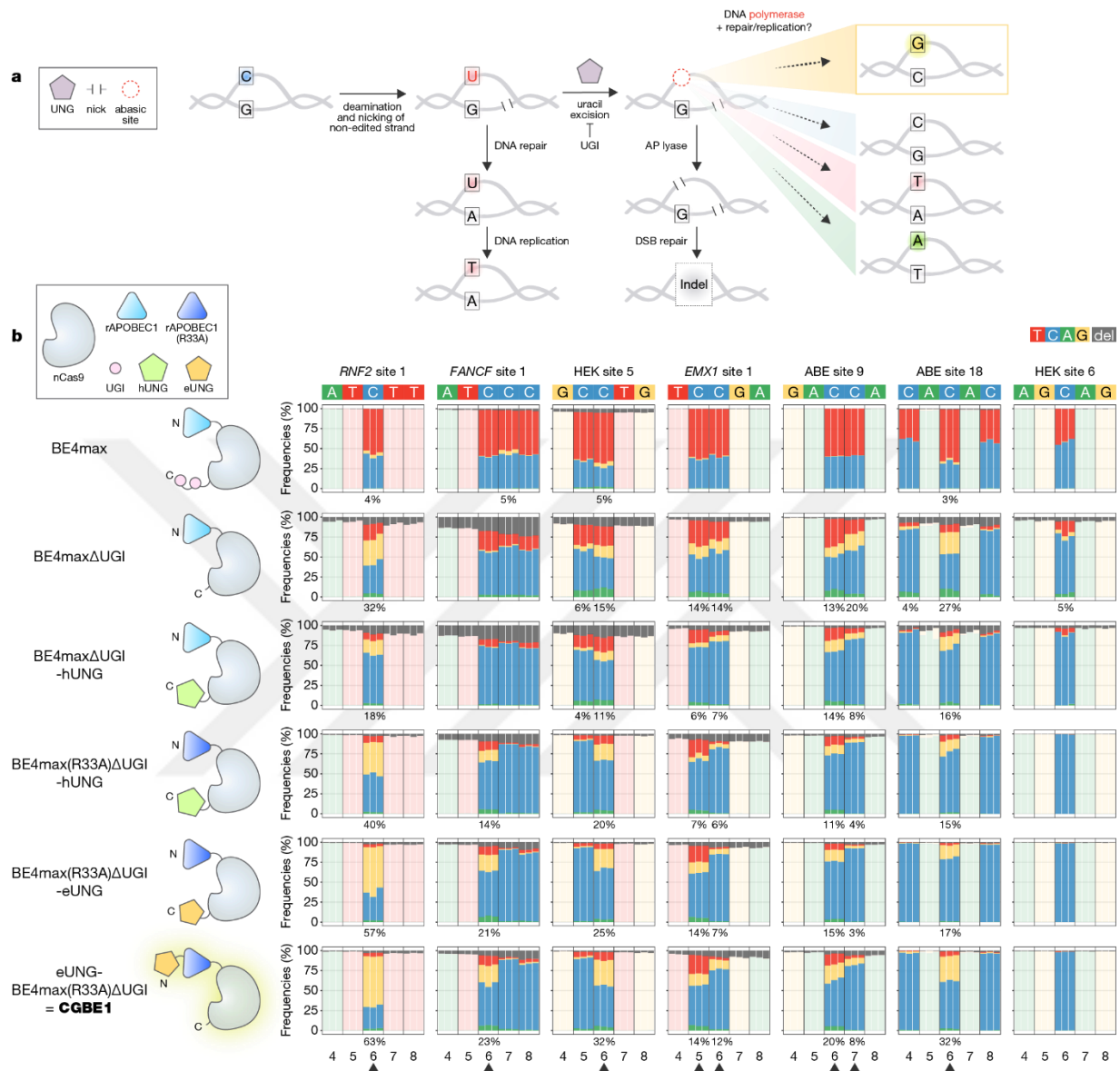


Figure 1: Engineering of a C-to-G base editor. (a) Schematic of potential cellular mechanisms and outcomes downstream of cytosine deamination by base editors. Uracil excision by endogenous uracil N-glycosylase (UNG, purple pentagon), nicking on the non-edited strand by Cas9 nickase, followed by DNA repair and replication can lead to diverse editing outcomes. C, Cytosine; G, Guanine; U, Uracil; T, Thymine; A, Adenine; UGI, uracil glycosylase inhibitor; AP lyase, apurinic/apyrimidinic site lyase; DSB, double strand break. (b) Bar plots showing on-target DNA base editing frequencies with various base editor architectures using seven gRNAs targeting genomic sites in HEK293T cells. N and C in the base editor illustrations indicate amino-terminal and carboxy-terminal ends, respectively. Gray overlay bars at top represent deletions at each nucleotide. Target cytosines are highlighted. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side-by-side. Percentage values below specific cytosine bases indicate the average C-to-G editing observed (values below 3% not reported). Numbering on the bottom indicates position of the base in the protospacer with 1 being the most PAM-distal base. Arrowheads indicate cytosines with C-to-G edits.

Given this observation about ABE-mediated C-to-G alterations, we wondered whether we could induce these edits more efficiently by modifying the BE4max CBE (Koblan et al., 2018; Komor et al., 2017), which harbors an enzyme actually intended to deaminate cytosines (the rat APOBEC1 cytidine deaminase). Removal of the two UGIs from BE4max to create BE4max Δ UGI resulted in an increase in C-to-G (and to a lesser degree C-to-A) edits relative to wild-type BE4max when tested with seven different gRNAs targeted to sites with Cs at protospacer positions 5, 6, and 7 in HEK293T cells (**Fig. 1b**). In general, C-to-G editing was observed with BE4max Δ UGI at Cs that were preceded by A, C, or T, with the most efficient editing generally observed with Cs at protospacer position 6 (**Fig. 1b**). We also observed a substantially higher frequency of indels with BE4max Δ UGI relative to BE4max (**Fig. 1b**, **Extended Data Fig. 2b**), consistent with the idea that this fusion is likely more efficient at creating abasic sites (Komor et al., 2016, 2017). Reasoning that generation of an abasic site is important for increased C-to-G editing, we further hypothesized that adding human UNG (**hUNG**) enzyme to BE4max Δ UGI might enhance the frequency of desired edits. However, a BE4max Δ UGI-hUNG fusion possessed somewhat decreased C-to-G editing activity and did not induce appreciably changed frequencies of indels with the seven gRNAs tested (although it did show decreased C-to-T editing activity) (**Fig. 1b**, **Extended Data Fig. 2b**). Similar results were obtained when hUNG was fused at the N-terminus of BE4max Δ UGI (**Extended Data Fig. 1**). Fusion of UNG to ABEmax did not yield enhanced C-to-G editing compared to ABEmax (**Extended Data Fig. 1**). We also tested a variety of CBEs that are based on non-APOBEC1 deaminase architectures, such as human APOBEC3A (**A3A**), engineered A3A-BE3 (Gehrke et al., 2018), human AID-BE3 (Komor et al., 2017), and the *Petromyzon marinus* CDA1(**pmCDA1**)-based Target-AID (Nishida et al., 2016), as well as variants thereof lacking

UGIs and having added UNGs. Among this larger ensemble of variants, none consistently showed higher C-to-G editing activity than the BE4maxΔUGI-hUNG editor (**Supplementary Fig. 1**).

We also investigated whether introducing mutations into the APOBEC1 part of BE4maxΔUGI-hUNG might further increase the frequency of C-to-G editing. Although we do not have a mechanistic understanding of how C-to-G edits are induced, we reasoned that altering the deamination dynamics of APOBEC1 might also influence the editing outcome. We focused on the APOBEC1 R33A mutation, a substitution we previously showed can decrease off-target RNA editing while substantially preserving the efficiency and increasing the precision of on-target DNA editing by CBEs (Grünewald, Zhou, Garcia, et al., 2019). We found that introduction of R33A into BE4maxΔUGI-hUNG increased C-to-G editing frequencies with three of the seven gRNAs tested in HEK293T cells while leaving editing frequencies essentially unaltered with the other four (**Fig. 1b**). The effect of the R33A variant was most striking with the *FANCF* site 1 gRNA, which had shown virtually no C-to-G editing with any of the other editors we tested but now showed a mean editing frequency of 14.0% (**Fig. 1b**). Interestingly, BE4max(R33A)ΔUGI-hUNG on average showed lower indel byproducts with 6 out of 7 gRNAs compared to BE4maxΔUGI-hUNG (**Extended Data Fig. 2b**).

We additionally explored whether replacing the hUNG present in the BE4max(R33A)ΔUGI-hUNG editor with an orthologous UNG from *Escherichia coli* (**eUNG**) might increase the efficiency of C-to-G edits. We created two additional editors: BE4max(R33A)ΔUGI-eUNG and eUNG-BE4max(R33A)ΔUGI with an eUNG added to the carboxy- or amino-terminal ends, respectively. Testing of these fusions in HEK293T cells revealed that both induced C-to-G edits with higher frequencies than BE4max(R33A)ΔUGI-

hUNG for six out of seven gRNAs tested (mean editing frequencies ranging from 3.3-57.0% and 8.5-62.6% for BE4max(R33A) Δ UGI-eUNG and eUNG-BE4max(R33A) Δ UGI, respectively) (**Fig. 1b**). Indel frequencies with both fusions were generally comparable to those observed with BE4max(R33A) Δ UGI-hUNG (**Extended Data Fig. 2b**). Given its higher C-to-G editing activity, we chose the eUNG-BE4max(R33A) Δ UGI fusion (hereafter referred to as C-to-G Base Editor 1 (**CGBE1**)) for additional characterization.

To more comprehensively characterize CGBE1, we tested its activity with 18 additional gRNAs in human HEK293T cells. 12 of the sites targeted by these 18 gRNAs have a C at position 6 (“C6-sites”) (**Fig. 2a** and **Extended Data Fig. 3**) and 6 have a C at positions 4, 5, 7, or 8 (“non-C6-sites”) (**Fig. 2b** and **Extended Data Fig. 4a**). For 16 of the 18 sites, CGBE1 induced C-to-G edits with substantially higher frequencies than what was observed with its parental CBE control (BE4max(R33A)) (**Fig. 2a** and **b**). Highly efficient C-to-G edits were observed for 4 of the 18 sites (ABE site 7, ABE site 8, HEK site 2, and *PPP1R12C* site 6), with mean editing frequencies ranging from 41.7 to 71.5% (**Fig. 2a** and **b**). C-to-G edits were by far the most efficiently induced edits at these 4 sites with only very low levels of C-to-T or C-to-A byproducts observed (**Fig. 2a** and **b**). C-to-G was also the most efficiently induced edit for 6 additional sites albeit at lower frequencies (three C6-sites and three non-C6-sites) (**Fig. 2a** and **b**). In total, when combined with the results obtained with the initial seven gRNAs described above (**Fig. 1b**), CGBE1 induced C-to-G editing with mean frequencies of 20% or higher at 14 of the 25 sites tested (**Figs. 1b, 2a** and **b**). Notably, C-to-G editing was most efficient for Cs embedded in an AT-rich sequence context (**Figs. 1b, 2a** and **b**). Analysis of the spatial distribution of editing across all 25 sites tested shows that the mean frequency of C-to-G editing

was highest at position 6 and that indels (Extended Data Figs. 2b and 4b) were distributed throughout the protospacer (Fig. 2c).

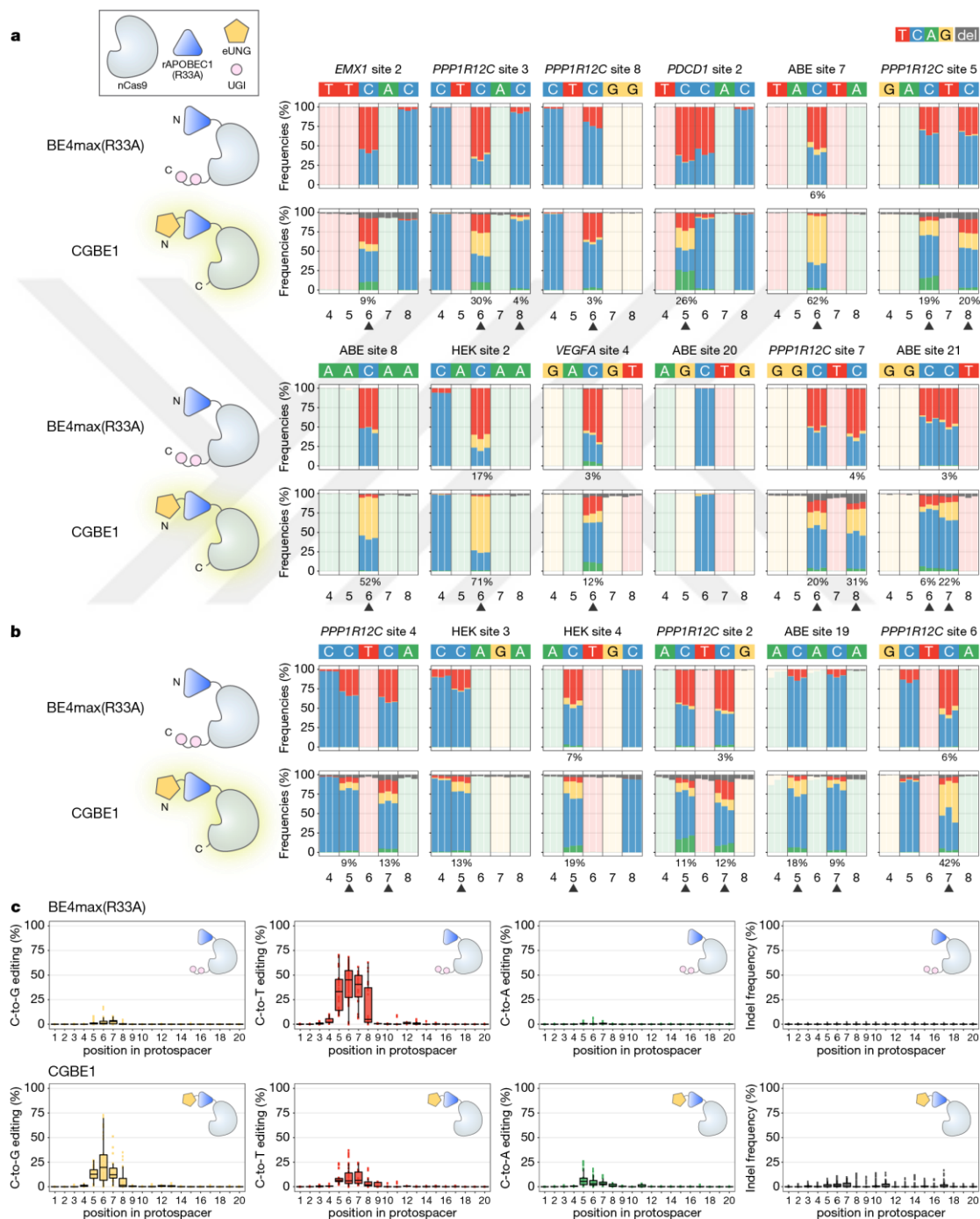


Figure 2: Additional characterization of CGBE1 on-target editing activities in HEK293T cells. a,b, Bar plots showing the on-target DNA base editing frequencies induced by

Figure 2 (Continued)

BE4max(R33A) and CGBE1 using 12 gRNAs for sites with a C at position 6 (C6-sites; **(a)**) and 6 gRNAs for sites with a C at position 4, 5, 7, or 8 (non-C6-sites; **(b)**) in HEK293T cells. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side-by-side. **c**, Dot and box plots representing the combined distribution of C-to-G (yellow), C-to-T (red), C-to-A (green), and indel (gray) frequencies per nucleotide across the entire protospacer from experiments performed with BE4max(R33A) and CGBE1 using 25 guides. Boxes span the interquartile range (IQR; 25th to 75th percentile), horizontal lines indicate the median (50th percentile), and whiskers extend to $\pm 1.5 \times \text{IQR}$. Data points in plots represent full range of values plotted. Single dots represent individual replicates ($n = 3$ independent replicates per site). The graphs were derived from the data shown in Figs. 1b, 2a and b, and Extended Data Fig. 1.

We next explored the impact of deleting the eUNG domain from the CGBE1 editor on its activity. This particular editor architecture, which we named miniCGBE1 (**Fig. 3**), had not been made or tested over the course of the stepwise progression from BE4max to CGBE1 and also had the added advantage of being smaller in size. Side-by-side comparisons of miniCGBE1 with CGBE1 at the same 25 sites we had previously tested showed that the frequencies of editing observed with miniCGBE1 were comparable but moderately lower at 6 out of 25 sites tested (mean editing frequencies across all 25 sites of 14.4% and 13% with CGBE1 and miniCGBE1, respectively), whereas the indel frequencies induced by miniCGBE1 were lower at 15 out of 25 sites (mean indel frequencies of 10.4% and 8.5% for CGBE1 and miniCGBE1, respectively; **Fig. 3, Supplementary Fig. 2**).

To more fully characterize the positional preferences within the editing windows of CGBE1 and miniCBGE1, we tested these two editors side-by-side with BE4max and BE4max(R33A) using 23 additional gRNAs that target sites with cytosines at protospacer positions 4, 5, 7, and 8 (**Supplementary Fig. 3**). The targets of these 23 gRNAs included six sites with a C5, five with a C7, four with a C8, and eight with two Cs at various positions (C4 and C7, C4 and C8, C5 and C7, C5 and C8, and C7 and C8). Mean editing frequencies induced

by CGBE1 were comparable to those of miniCGBE1: 1.7% and 1.5% at C4, 7.3% and 6.7% at C5, 16.0% and 13.5% at C7 and 3.4% and 2.9% at C8 for CGBE1 and miniCGBE1, respectively (**Supplementary Fig. 3a**). In addition, indel frequencies induced by CGBE1 and miniCGBE1 were comparable at 10 sites, lower with CGBE1 at five sites, and lower with miniCGBE1 at eight sites (**Supplementary Fig. 3b**). Collectively, our testing of CGBE1 and miniCGBE1 with 48 different gRNAs demonstrates that both have a favorable editing window for cytosines at positions 5-7 in the protospacer with those at position 6 being edited most efficiently (**Extended Data Fig. 5**). This finding is consistent with our previously published studies showing that a CBE with the APOBEC1-R33A variant edits optimally on positions 5-7 of the protospacer and more weakly on positions 4 and 8 (Grünewald, Zhou, Garcia, et al., 2019).

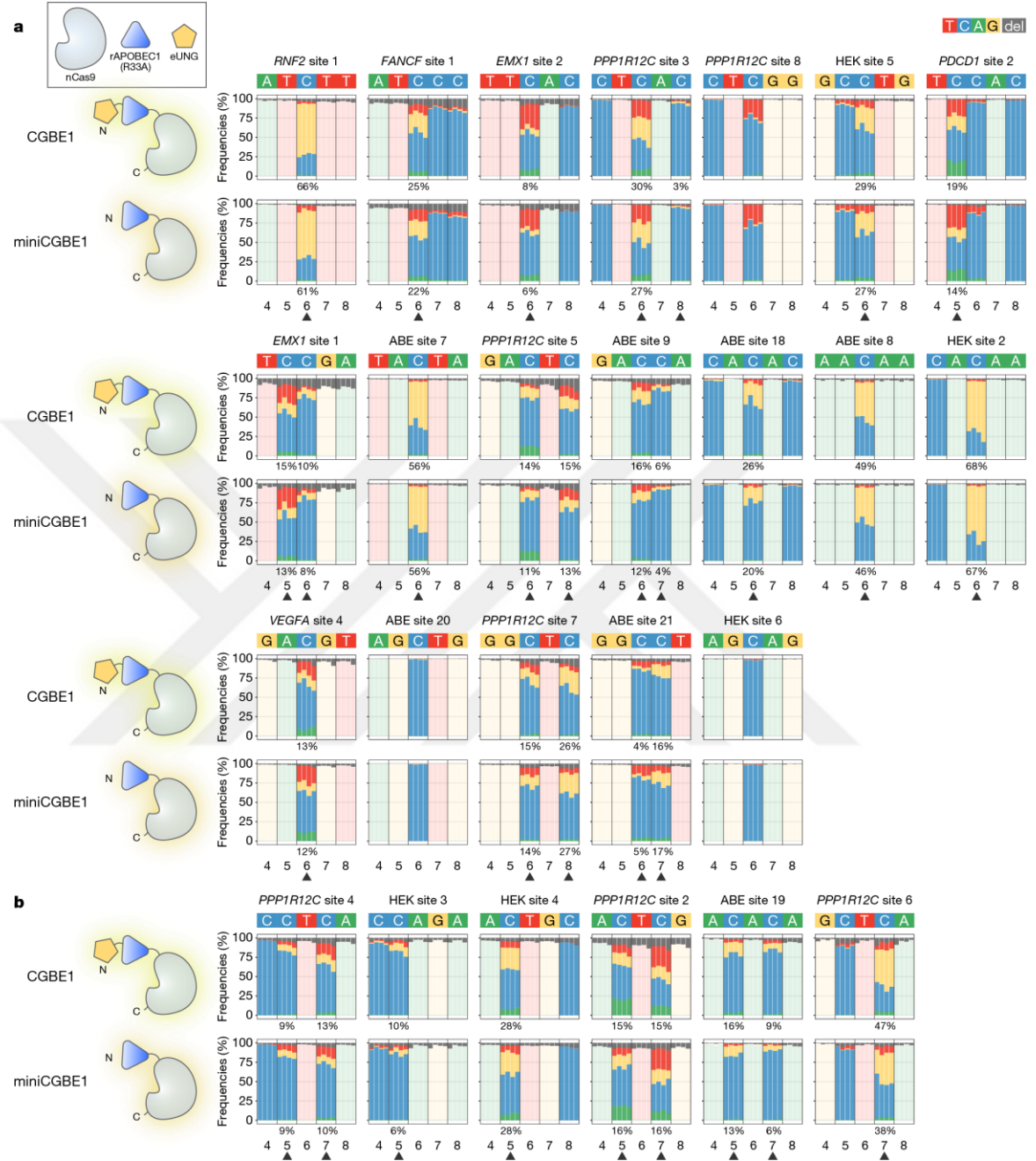


Figure 3: Comparison of CGBE1 and miniCGBE1 on-target editing activities in HEK293T cells. a,b, Bar plots showing the on-target DNA base editing frequencies of CGBE1 and miniCGBE1 using 19 gRNAs for sites with a C at position 6 (C6-sites; **a**) and 6 gRNAs for sites with a C at position 4, 5, 7, or 8 (non-C6-sites; **b**) in HEK293T cells. Editing frequencies of four independent replicates ($n = 4$) at each base are displayed side-by-side.

To characterize gRNA-dependent DNA off-target profiles of CGBE1 and miniCGBE1, we assessed their editing activities side-by-side with BE4max and BE4max(R33A) at 23 known SpCas9 off-target sites of five different gRNAs (previously identified by GUIDE-seq (Tsai et al., 2014)) (**Supplementary Fig. 4**). BE4max induced C-to-D (D = A, G, or T) edits at 15 of the 23 off-target sites, while BE4max-R33A induced edits at lower frequencies across all of these 15 sites (**Supplementary Fig. 4a**). Similarly, both CGBE1 and miniCGBE1 showed lower C-to-D off-target editing at 14 out of the 15 off-target sites that were edited by BE4max (**Supplementary Fig. 4a**). As expected, off-target indel frequencies were higher with CGBE1 and miniCGBE1 relative to BE4max at 18 out of 23 sites, although miniCGBE1 again showed reduced activity compared with CGBE1 at 14 out of these 18 sites (**Supplementary Fig. 4b**). Overall, this assessment of gRNA-dependent DNA off-target editing shows that CGBE1 and miniCGBE1 induce fewer off-target DNA base edits than BE4max, that CGBE-induced indels can occur at off-target sites, and that indels are reduced with miniCGBE1 relative to CGBE1.

We additionally tested whether we could improve the somewhat more restricted targeting range of CGBEs by using previously described SpCas9-NG and SpCas9-VRQR variants that recognize shorter NG (Nishimasu et al., 2018) and alternative NGA (Kleinstiver et al., 2016) PAMs, respectively. We targeted six sites with NGT PAMs using modified CGBE1-NG and miniCGBE1-NG variants and six sites with NGAG PAMs using CGBE1-VRQR and miniCGBE1-VRQR variants. Each of these 12 sites have a cytosine at position 6 embedded within an AT-rich sequence context to provide an optimal target for C-to-G editing (**Extended Data Fig. 6**). At these target sites, CGBE1-NG and miniCGBE1-NG induced C-to-G edits with frequencies as high as 27% and 26%, respectively, and CGBE1-VRQR and miniCGBE1-VRQR induced C-to-G edits with frequencies of up to 31% (**Extended Data Fig. 6**). These results show

that the targeting range of CGBE constructs can be expanded by using Cas9 variants with altered or relaxed PAM recognition specificities.

Lastly, we compared our CGBEs with new Prime Editing (PE) methods that can introduce a diverse range of different edits and that were published (Anzalone et al., 2019) while we were completing this project. The PE2 system uses two components: (1) a Prime Editor fusion protein and (2) a prime editing gRNA (**pegRNA**) (**Extended Data Fig. 7a**) (Anzalone et al., 2019). The more efficient PE3 system adds a secondary “nicking gRNA” (**ngRNA**) that directs a nick to the non-edited DNA strand, thereby increasing editing efficiency (**Extended Data Fig. 7a**) (Anzalone et al., 2019). We performed side-by-side comparisons of our CGBEs with PE2 and PE3 systems for installing four different C-to-G edits, assessing frequencies of these alterations across four different human cell lines (HEK293T, K562, U2OS, and HeLa cells). Positive control experiments we performed in all four cell lines re-confirmed that two other previously described pegRNAs could induce a G-to-T transversion in *FANCF* site 1 and a CTT insertion in HEK site 3, that PE3 outperforms PE2, and that the highest prime edit frequencies are observed in HEK293T cells (**Extended Data Fig. 7b-d**). For all four C-to-G edits (which we had already established could be efficiently induced by CGBEs in HEK293T cells), we found that both PE2 and PE3 were substantially less efficient than CGBE and miniCGBE1 across all four cell lines (**Extended Data Fig. 7e, Supplementary Fig. 5a**). Importantly, these data also show that our CGBEs can function robustly and efficiently across multiple human cancer cell lines. In addition, we found that the frequencies of unwanted indels were lower with prime editors compared to the CGBEs in all four cell lines, an unsurprising finding given that on-target editing of PEs was also much less efficient (**Supplementary Fig. 5b**). To rule out that the pegRNAs and ngRNAs we designed were inactive or unable to interact

with Cas9, we tested their abilities to induce Cas9-mediated indels at their target sites in HEK293T cells (note that we could not assess the activity of the HEK site 3 ngRNA due to its overlap with a required PCR primer). The indel frequencies induced by these pegRNAs and ngRNAs were comparable to those observed with the two positive controls (**Extended Data Fig. 7f**).

Our development and characterization of CGBE1 and miniCGBE1 provide an important proof-of-concept for a modified base editor architecture that can be used to induce programmable C-to-G transversion mutations. Additional optimization and engineering will be required to enable a more robust and efficient CGBE platform. Although highly efficient C-to-G editing with high product purity could be observed at some sites with a C at position 6 in protospacer, not all C6 sites showed equally high activities. At least some of this variability may be influenced by the identities of bases flanking C6. For example, C6 sites with the most efficient editing and optimal product purity had As and/or Ts at the flanking 5 and 7 positions. In addition, C6 sites with a preceding G showed consistently absent or very low C-to-G editing, a known limitation of the rat APOBEC1 cytidine deaminase (Komor et al., 2016). In addition, our combined data from 48 different sites demonstrated that C-to-G editing by CGBE constructs was favored when the target C was in position 6 and less favored when in positions 5, 7 and 8. Collectively, these observations show that the targeting range of the CGBE constructs is constrained by both the location of the target C and its sequence context, although additional work could further refine these parameters. The use of CGBEs harboring Cas9 PAM recognition variants with altered (Kleinstiver et al., 2015, 2016) or broadened (Hu et al., 2018; Nishimasu et al., 2018; Walton et al., 2020) PAM specificities may help to further alleviate some of these targeting range limitations.

Among four target sites we tested, we also found that our CGBE variants generally induced higher frequencies of desired C-to-G edits than PE2 or PE3 prime editors directed to induce the same alterations. This was consistently observed for these four C-to-G edits across four different human cancer cell lines. Although we designed the pegRNAs and ngRNAs following parameters previously described by the Liu group (Anzalone et al., 2019), it is possible that additional optimization (e.g., of the pegRNA target site choice, primer binding site (PBS) or reverse transcription template (RTT) lengths, distance between the pegRNA and ngRNA nicking sites) might lead to more efficient prime editing outcomes. For the four sites we tested, prime editors appear to induce lower frequencies of unwanted indels relative to our CGBEs. All of these findings are consistent with the conclusions of Liu and colleagues that base editors induce a more restricted range of edits than prime editors but also generally show more efficient editing if the target base is positioned optimally (Anzalone et al., 2019).

Discussion & Future directions

Although the exact mechanism behind the preferential C-to-G editing activity needs to be elucidated, findings of our work and the others collectively suggest that uracil-initiated BER is fundamentally involved in the preferential installation of a guanine at the CGBE-induced abasic sites (L. Chen et al., 2021; Koblan, Arbab, et al., 2021; D. R. Liu & Koblan, 2018; D. Zhao et al., 2020). Specifically, Zhao et al. back-to-back with our work demonstrated that fusing *E. coli* Ung to the wildtype rAPOBEC1-nCas9 Δ UGI construct can install C-to-G alterations with up to 92% product purity and up to 53% efficiency in a context-dependent manner in human HEK293T cells (D. Zhao et al., 2020). Chen et al. screened the downstream BER proteins DNA Polymerase

β (only its DNA binding and lyase domain), DNA Ligase III, and XRCC1 and found that only XRCC1 as a fusion partner increased the mean C-to-G editing levels when fused to the wildtype rAPOBEC1-nCas9 Δ UGI construct, compared to the wildtype rAPOBEC1-nCas9 Δ UGI alone (L. Chen et al., 2021). XRCC1 is a scaffold protein that can recruit a suit of downstream BER proteins including DNA Ligase III, DNA polymerase β , AP endonuclease APE1, Poly (ADP-ribose) polymerase (PARP) (Nazarkina et al., 2007). Wildtype rAPOBEC1-nCas9-xRCC1 construct as a CGBE candidate displayed 15.4% mean C-to-G editing activity with 37% reported as the highest activity achieved at the target sites tested (L. Chen et al., 2021). Finally, Koblan et al. showed that the individual knockdown of the BER pathway components UNG, PCNA, POLD, REV1, along with other non-BER DNA repair proteins, substantially reduced C-to-G editing activity levels (Koblan, Arbab, et al., 2021; Yoshihiro Matsumoto, 2001; Nicolas et al., 2016; Prasad et al., 2016). Koblan et al. also employed an UNG ortholog (UdgX) and removed UGI as basis for most of their candidate CGBE constructs. A direct comparison of all 18 CGBE constructs described by four different groups have shown that our CGBE tools outperformed the other tools in two sites, performed comparable to the other tools in six sites, and performed worse in three sites in terms of product yield and purity, as assessed by Koblan et al. (**Fig. 4**) (Koblan, Arbab, et al., 2021). Applying CGBE-Hive experimental workflow directly on our CGBE constructs should reveal the complex sequence features and dynamics that regulate C-to-G editing outcomes (Koblan, Arbab, et al., 2021). Together, all of these studies strongly suggest that uracil-initiated BER could be responsible for the preferential C-to-G installation by the CGBE constructs, and that CGBE editing yield and product purity is context- and construct-dependent.

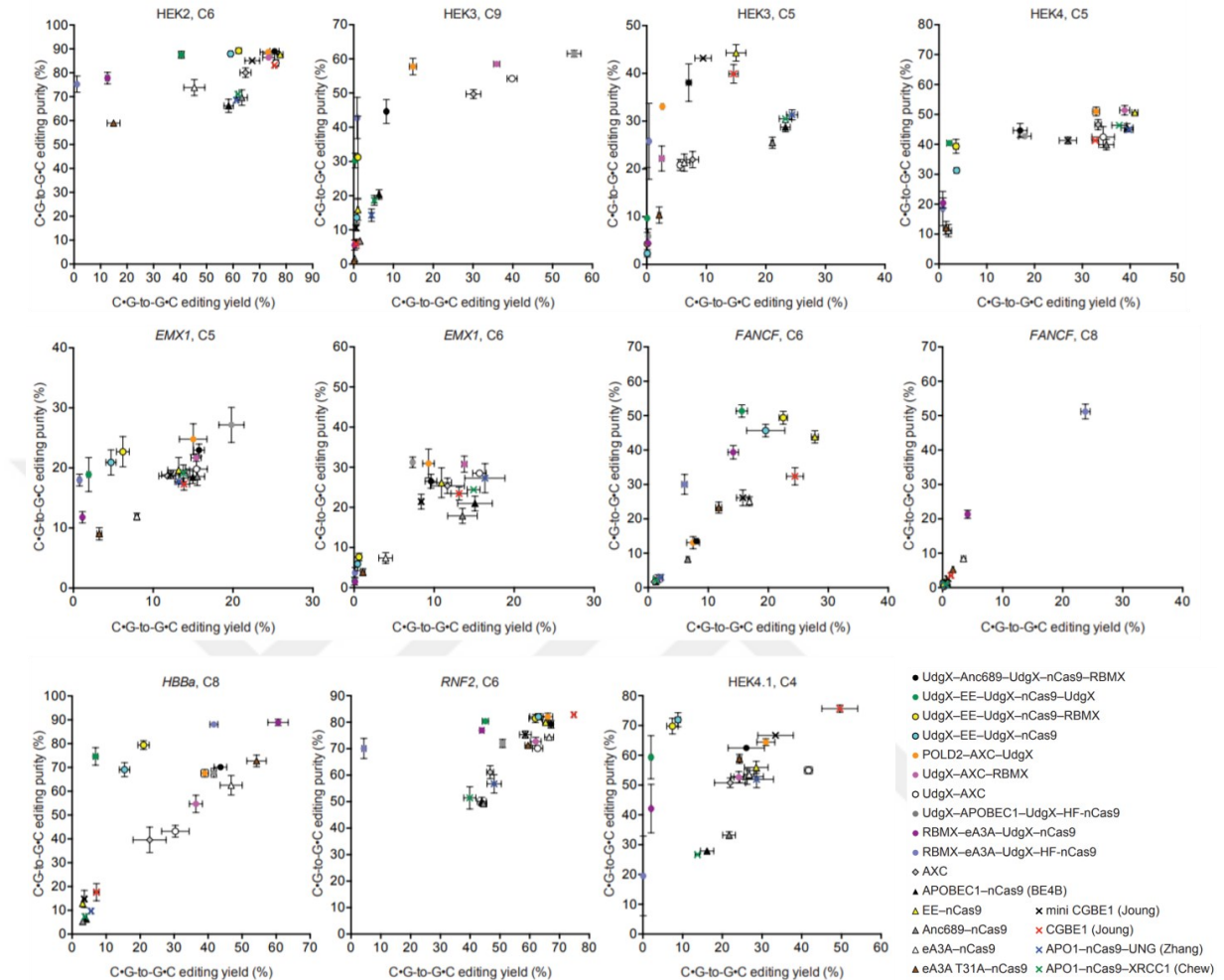


Figure 4: Direct comparison of CGBEs (by Koblan et al.) described in four different studies in terms of product purity and yield. This figure is a direct adaption from Supplementary Figure 13 of the study of Koblan et al. (Koblan, Arbab, et al., 2021). X-axis displays mean C-to-G editing yield, whereas Y-axis displays mean C-to-G editing product purity for three biological replicates.

Although we do not yet understand why the R33A mutation in rat APOBEC1 increases the efficiency of C-to-G editing, this observation strongly suggests that altering the activities of rat APOBEC1 with other mutations might yield further desirable increases in editing activities. This suggestion is corroborated by the findings that using engineered human APOBEC3A

(eA3A), engineered rat APOBEC1 with EE mutations, ancestrally reconstructed rat APOBEC1 (Anc689) as the cytidine deaminase fusion partner increase the C-to-G editing efficiency and product purity levels (Gehrke et al., 2018; Y. B. Kim et al., 2017; Koblan et al., 2018; Koblan, Arbab, et al., 2021).

We also found that C-to-G editing frequencies showed cell-type-dependence across the four cell lines tested, a finding consistent with the previous reports, which described that even standard CBEs can induce higher efficiencies of unwanted C-to-G edits in certain cell types (Gehrke et al., 2018; Komor et al., 2017). Likewise, other CGBE candidates engineered by the other groups also showed cell-type dependent C-to-G activity levels (L. Chen et al., 2021; Koblan, Arbab, et al., 2021). A better understanding of the mechanistic parameters that govern both the frequencies and the product purity of C-to-G edits may suggest additional strategies to further increase the targeting range and efficiency of CGBEs. Running the recently described pairwise CRISPRi library targeting 476 DNA repair genes that can link the C-to-G editing outcomes to individual DNA repair gene that is knocked down, the next steps and fusion partners to engineer CGBE2 from CGBE1 could be potentially identified (Husmann et al., 2021).

We do not yet know the reasons behind the preference for the cytosines at position six. Complete crystal structure of the ternary Cas9-gRNA-DNA complex demonstrated that the nucleotides in position 14 to 20 in nontarget strand are buried within Cas9 protein, while nucleotides in position 1 to 13 are not enclosed but still partially bound to the outside surface of the Cas9 molecule (Huai et al., 2017). Moreover, rat APOBEC1 with R33A mutation used in our constructs has been shown to harbor a tighter editing window (position 5-7) compared to its wildtype counterpart (position 4 to 8) (Grünwald, Zhou, Garcia, et al., 2019). It could be hypothesized that due to the resultant specific ternary conformation, cytosines at position six are

the most accessible to the cellular or Cas9-bound UNG. Cas9 can remain bound to its target site for extended periods of time both in vitro and in vivo, even post-cleavage in the case of nuclease (Brinkman et al., 2018). Incorporation of the high-fidelity Cas9 mutations that lower the excess DNA binding energy of Cas9 has been shown to substantially improve the C-to-G editing profiles at a cytosine at position 9, indicating that altering protein-DNA interactions of the CRISPR effector can affect the editing rates and purity. Although high fidelity mutations has been shown not to alter Cas9's DNA binding affinity for its targets in short term, a high fidelity variant Cas9 can still exhibit shorter dwelling time in long term possibly exposing deeper parts of the nontarget strand for longer durations (J. S. Chen et al., 2017; M. Yang et al., 2018). Alternative CGBE designs with altered peptide linker length and composition, or targeted UNG-recruitment via stem-loop of gRNA instead of peptide-fusion could yield shifted or expanded targeting window, as was the case with the previous base editors (Komor et al., 2016; Yanhong Wang et al., 2020).

Although we do not yet know the reasons behind the preference for the AT-rich sequence flanking the target cytosines for high product yield and purity, we can hypothesize several theories following the lines of published body of evidence. It could be hypothesized that either the cellular UNG or the DNA polymerase that fill the opposite strand of the lesion could have preference for AT-rich context around the lesion. Although there are conflicting reports, it has been hypothesized and reported that UNG needs to bend its target site for catalysis and the energetic cost of bending for a target uracil in double-stranded DNA with adjacent adenine or thymine is lower in vitro (Hölz et al., 2019; Seibert et al., 2002). UNG from herpes simplex virus in vitro also exhibits more efficient excision of uracil substrate in single-stranded DNA with AT-rich flanking sequences, compared to GC-rich (Bellamy & Baldwin, 2001). Although whether

CGBE-induced uracil in nontarget strand is still bound to Cas9-gRNA complex and thus in single-stranded state, or in double-stranded state, must be determined, it could be hypothesized that UNG sequence preference could play a role in the outcomes observed. Uracil-induced BER pathway is expected to generate an abasic site in nontarget strand, while nCas9 concomitantly nicks the target strand. This leads to the hypothesis that REV1 translesion DNA polymerase with dCMP transferase activity opposite abasic site could be responsible for the preferential C-to-G installation by CGBE constructs (Gajula, 2019; D. R. Liu & Koblan, 2018; Masuda et al., 2002; J. R. Nelson et al., 1996). Interestingly, meta-analysis of human cancer genomes indicated that a mutational signature of TCW (TCA or TCT) motif was widely mutated in C-to-T or C-to-G manners by the aberrant activities of endogenous human APOBEC3A and 3B enzymes, and C-to-G activity was dependent on the catalytic activity of REV1 translesion DNA polymerase (Y. Yang et al., 2018). An activity similar to the DNA polymerase eta (POLH) during somatic hypermutation in B cells could also be proposed as an alternative mechanism that could explain the AT-rich preference behind the optimal CGBE activity. POLH has been shown to be responsible for placing a wrong guanine opposite endogenous thymine in WT (AT or TT) context during AID-induced mismatch repair (Y. Zhao et al., 2013). It is rationalized that the preceding adenine or thymine right before the template thymine constitutes a weak base pairing at the primer end (opposite to the logic of favorable GC-clamp for primer design) increasing the likelihood of POLH misplacing a guanine (Y. Zhao et al., 2013). Although it needs to be tested to what extent this contributes to the AT-rich context preference of CGBEs, nonetheless, rat APOBEC1, including the R33A variant, strongly favors cytosines preceded by a thymine (TC) and disfavors cytosines preceded by a guanine (GC) (Doman et al., 2020). Finally, WCW motif as the sequence preference for efficient and pure C-to-G editing has been observed and

recapitulated by two other groups describing alternative CGBE architectures (L. Chen et al., 2021; Koblan, Arbab, et al., 2021). Further in vitro experiments are warranted to elucidate the exact reasons behind the WCW motif preferred by CGBEs, as in vivo knockout of DNA repair enzymes usually are confounded by the redundancy and compensatory mechanisms in cells (Koblan, Arbab, et al., 2021).

CGBEs described in this work outperformed the PE2 and PE3 tools for single nucleotide substitutions tested at four sites in four human cell lines. Further optimization of PBS, RTT, and nicking gRNA parameters for PE3, using the recently described PE4 and PE5 technologies along with the enhanced pegRNAs and proximal dead gRNAs, and installing a second synonymous mutation to bypass MMR machinery could lead to higher or comparable C-to-G editing activity levels with PE technology (P. J. Chen et al., 2021; J. W. Nelson et al., 2021; Park et al., 2021). Moreover, genome- and transcriptome-wide spurious off-target editing of CGBE1 and miniCGBE1 must be assessed as the follow-up experiments, although rat APOBEC1 harboring the R33A mutation employed in our constructs has been shown to possess substantially lower propensity to induce guide-independent transcriptome-wide off-target editing, guide-independent and -dependent DNA off-target editing at select sites (Doman et al., 2020; Grünewald, Zhou, Iyer, et al., 2019).

The availability of efficient and programmable C-to-G base editing should further expand both research and therapeutic applications of base editing. CGBE-NG variants with relaxed PAM specificities described in our work can theoretically be employed to target and correct 81 disease-causing SNVs (pathogenic or likely-pathogenic cytosines at position 6 of spacers) as catalogued in ClinVar database (**Supplementary Table 2**). C-to-G editing permits the introduction of new codon and sequence changes not possible with CBEs and ABEs (**Fig. 5**). C-

to-T base edits performed by CBE at the third nucleotide (wobble) position of codons can induce only two amino acid alterations Met→Ile (ATG→ATA) and Trp→Stop (TGG→TGA). Similarly, ABE-induced A-to-G edits can induce only Ile→Met (ATA→ATG) and Stop→Trp (TGA→TGG) alterations. By contrast, C-to-G editing by CGBE at the third nucleotide position of codons enables 16 different amino acid alterations not possible with CBEs or ABEs. C-to-G alterations at the first and second nucleotide positions of codons also enable additional amino acid substitutions. Furthermore, transversion mutations introduced into transcription factor binding sites would be expected to have more pronounced effects on binding and gene expression as compared to transition mutations (Guo et al., 2017). Finally, and most importantly, the ability to install C-to-G edits will permit the correction of additional disease-causing mutations in both coding and non-coding regions that cannot be accessed with existing technologies. As was the case with both CBEs and ABEs (Y. B. Kim et al., 2017; Koblan et al., 2018; Komor et al., 2017; Thuronyi et al., 2019), we envision that our initial description of CGBE1 and miniCGBE1 will spur further optimization and development of next-generation C-to-G editors with more efficient and robust activities.

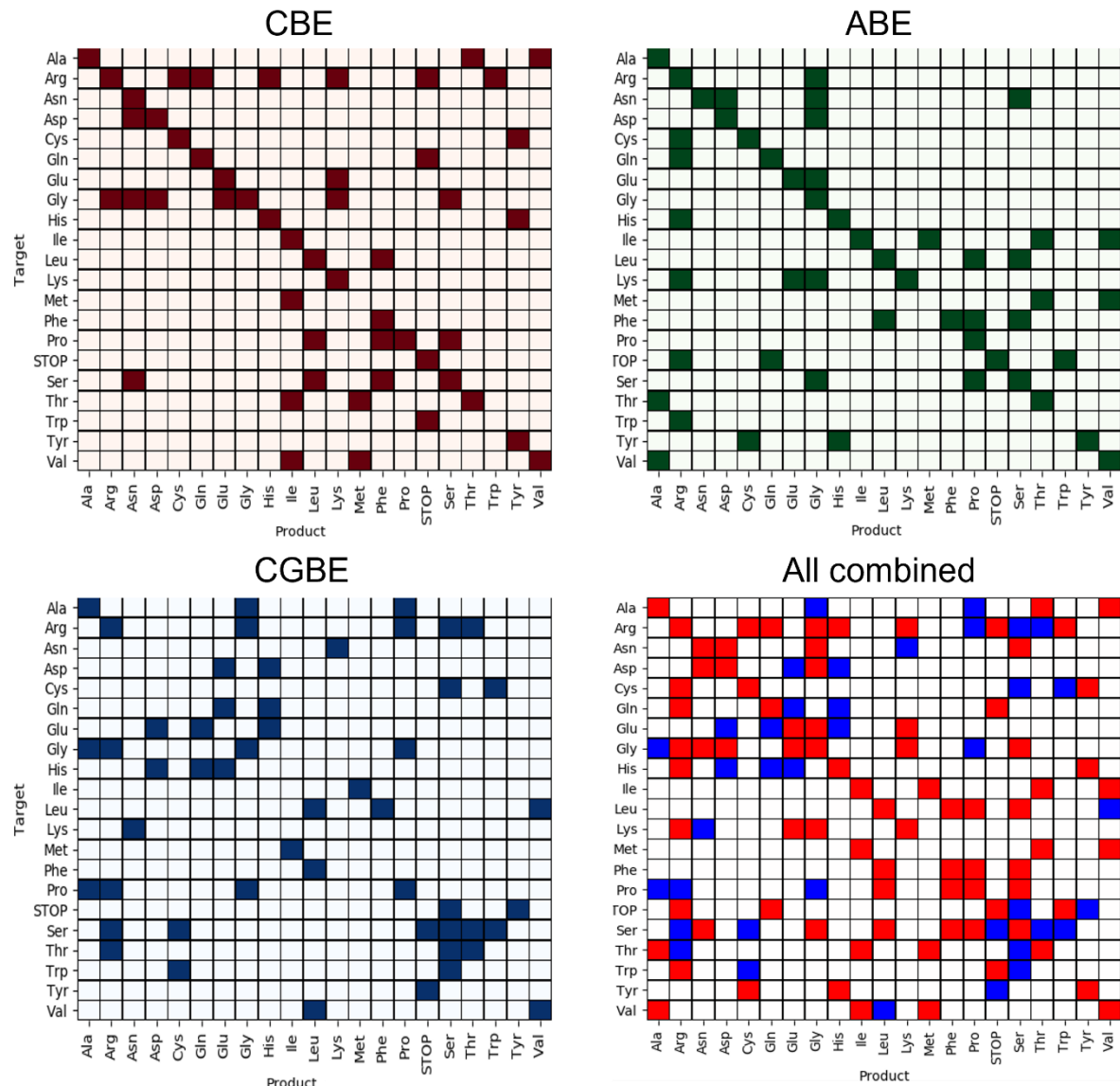
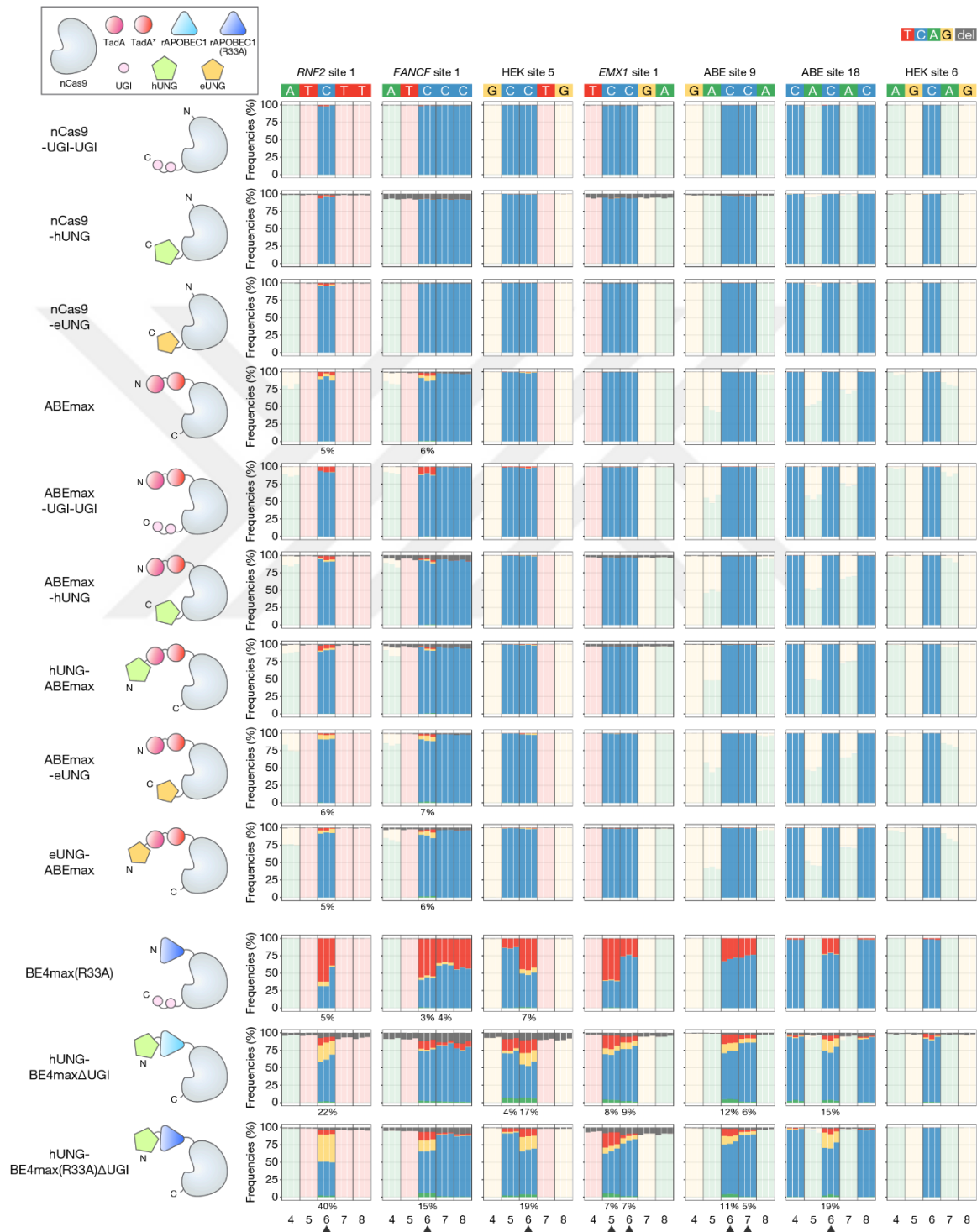


Figure 5: Comparison of theoretical amino acid alterations that can be installed using CBEs, ABEs, or CGBEs. Plots demonstrating the amino acids that can be targeted (y-axis) using CBEs (top left), ABEs (top right), CGBEs (bottom left) and converted into product amino acids (x-axis). Filled boxes represent possible alterations that can be induced with the respective base editor. Bottom right plot displays all plots overlayed wherein blue boxes represent the newly “unlocked” alterations that were not possible before the engineering of CGBEs, whereas red boxes represent the alterations that could be installed using the previous base editors.

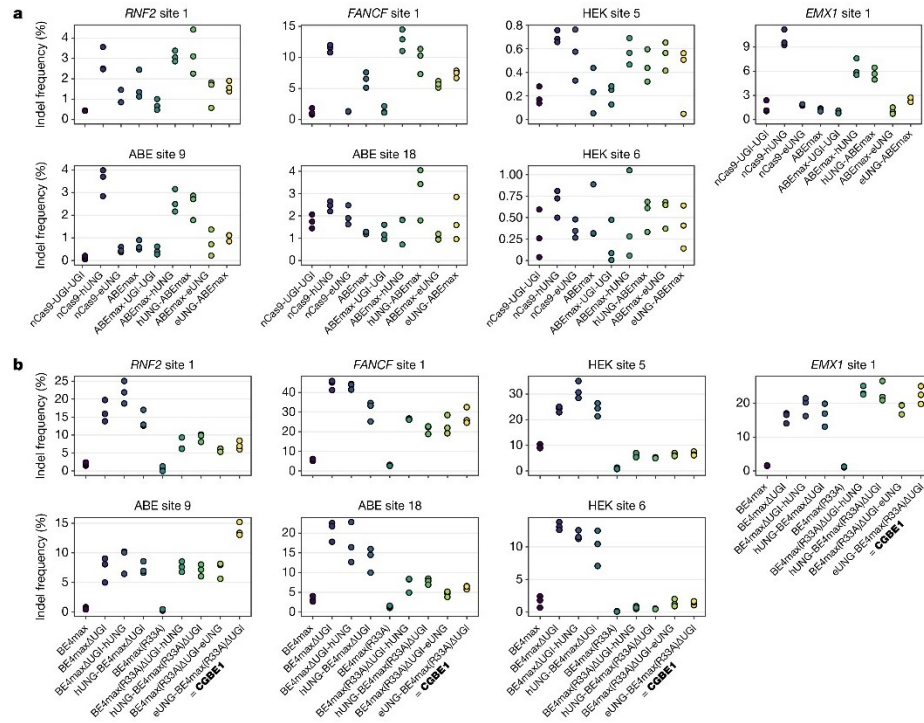
Extended Data



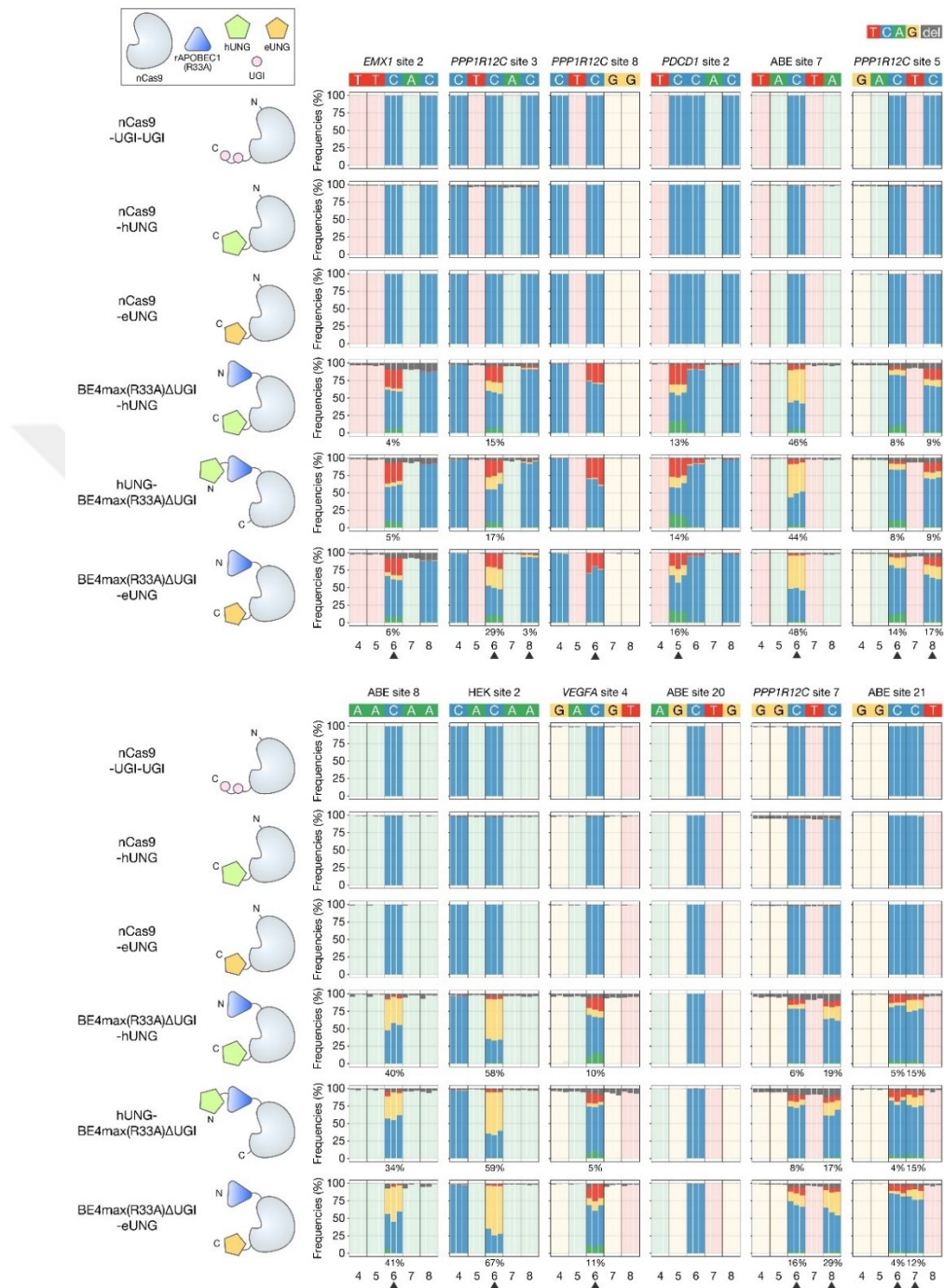
Extended Data Figure 1: On-target activities of nCas9 controls, ABE variants and more CBE variants tested for C-to-G editing in HEK293T cells. Bar plots showing the on-target DNA base

Extended Data Figure 1 (Continued)

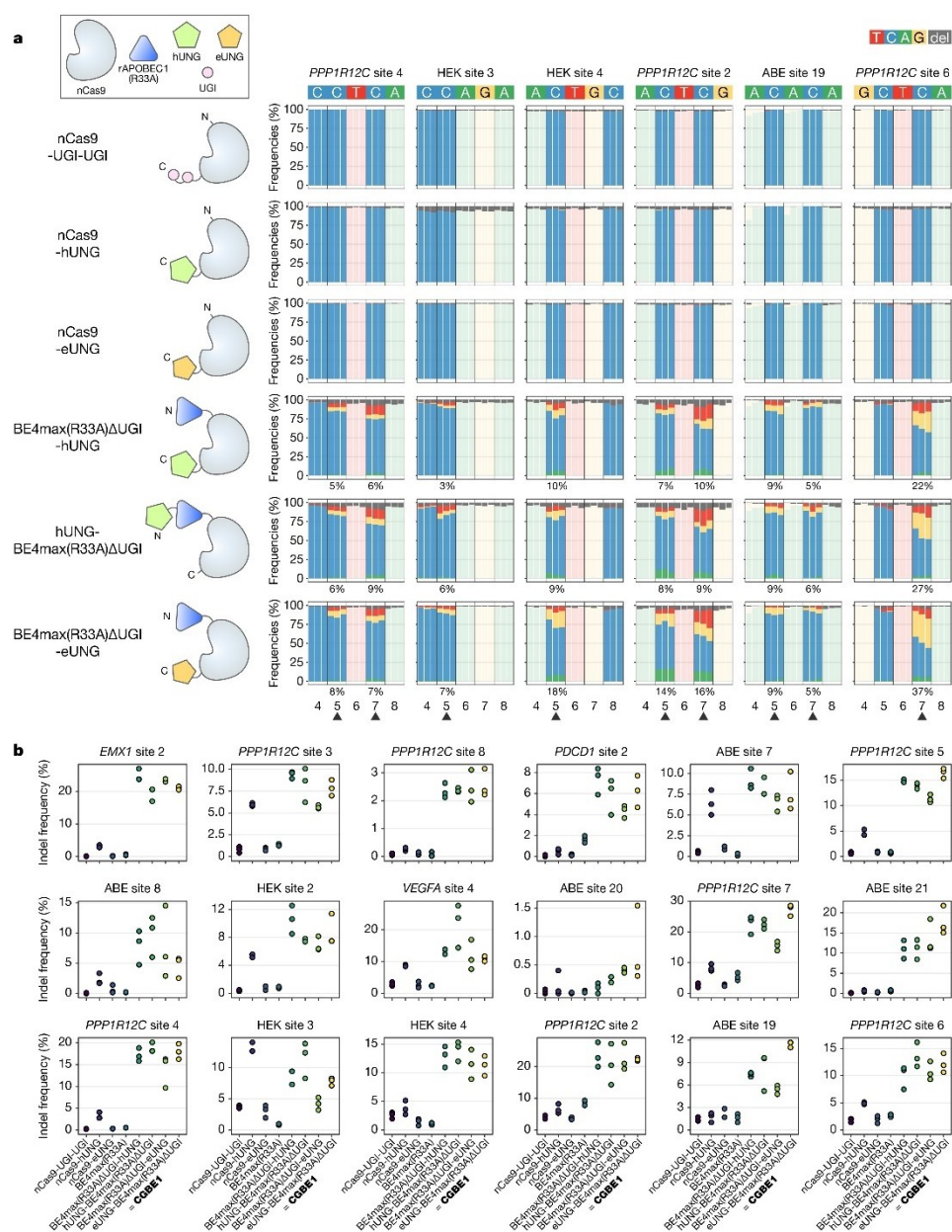
editing frequencies induced by nCas9 negative controls, ABE and ABE variants and other CBE variants with seven gRNAs in HEK293T cells. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side-by-side. Arrowheads indicate cytosines showing C-to-G edits by CGBE1.



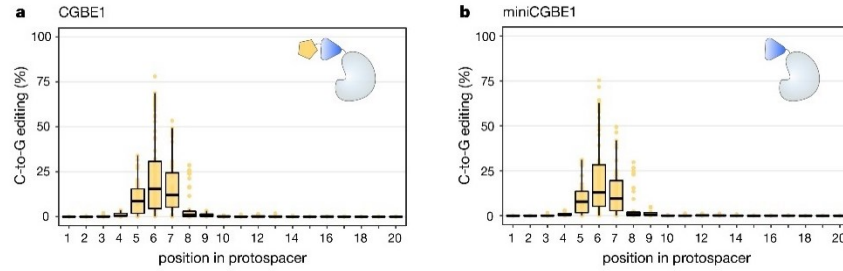
Extended Data Figure 2: Indel frequencies of nCas9 controls, ABE variants and CBE variants tested for C-to-G editing in HEK293T cells. a,b, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with various base editor architectures reported in (a) Extended Data Fig. 1 or (b) Fig. 1b and Extended Data Fig. 1. Single dots represent individual replicates ($n = 3$ independent replicates).



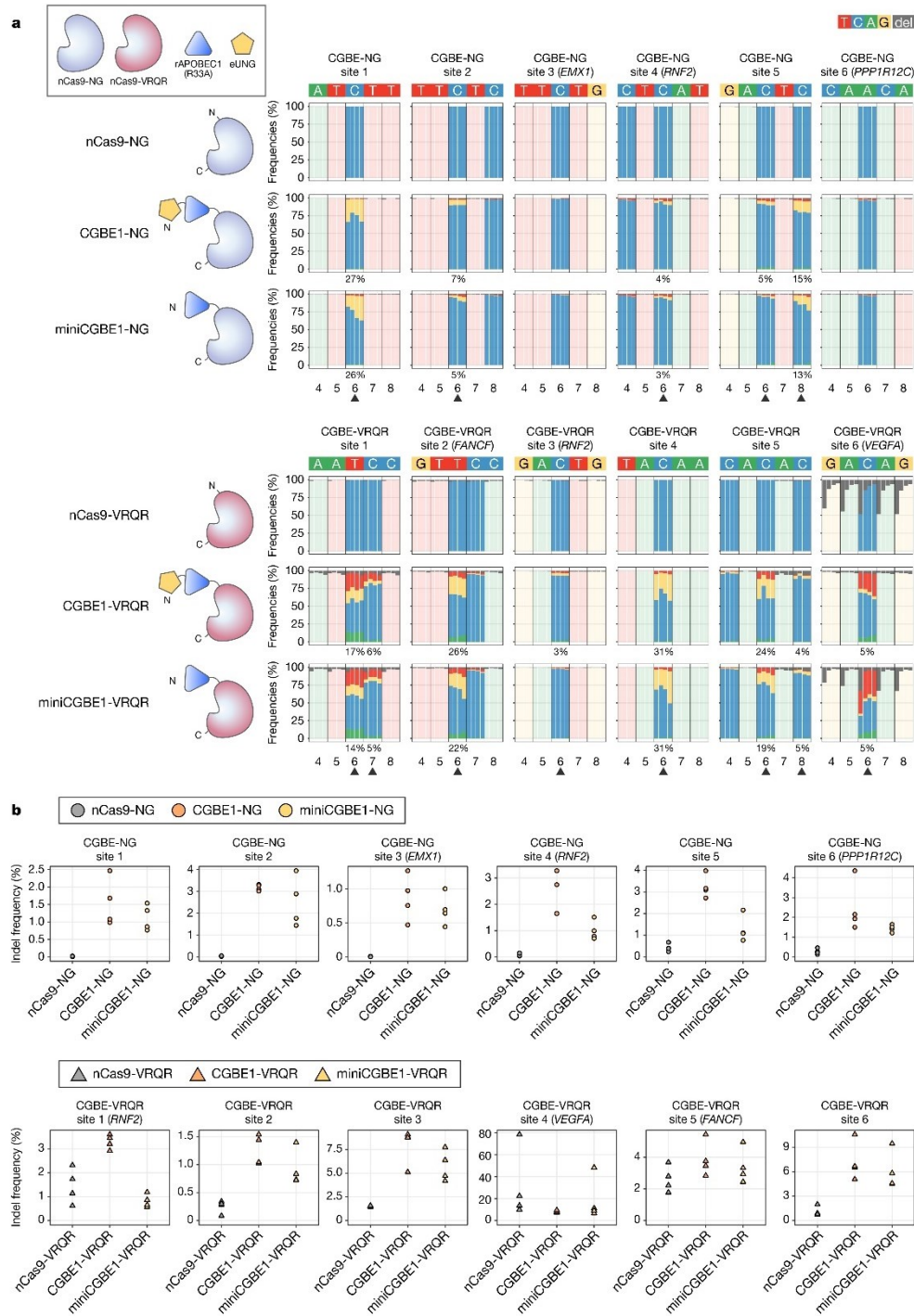
Extended Data Figure 3: On-target activities of nCas9 controls and CGBE1-related variants with 12 C6 gRNAs in HEK293T cells. Bar plots showing the on-target DNA base editing frequencies of nCas9 controls and CGBE1-related variants using 12 gRNAs for sites with a C at position 6 (C6-sites) in HEK293T cells. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side-by-side.



Extended Data Figure 4: On-target activities of nCas9 controls and CGBE1-related variants with 6 non-C6 gRNAs in HEK293T cells and indel frequencies across 18 targeted sites. a, Bar plots showing the on-target DNA base editing frequencies of nCas9 controls and CGBE1-related variants using 6 gRNAs for sites with a C at position 4, 5, 7, or 8 (non-C6-sites) in HEK293T cells. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side-by-side. **b**, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with CGBE1-related variants reported in Fig. 2a and b and Extended Data Figs. 3 and 4a. Single dots represent individual replicates ($n = 3$ independent replicates).



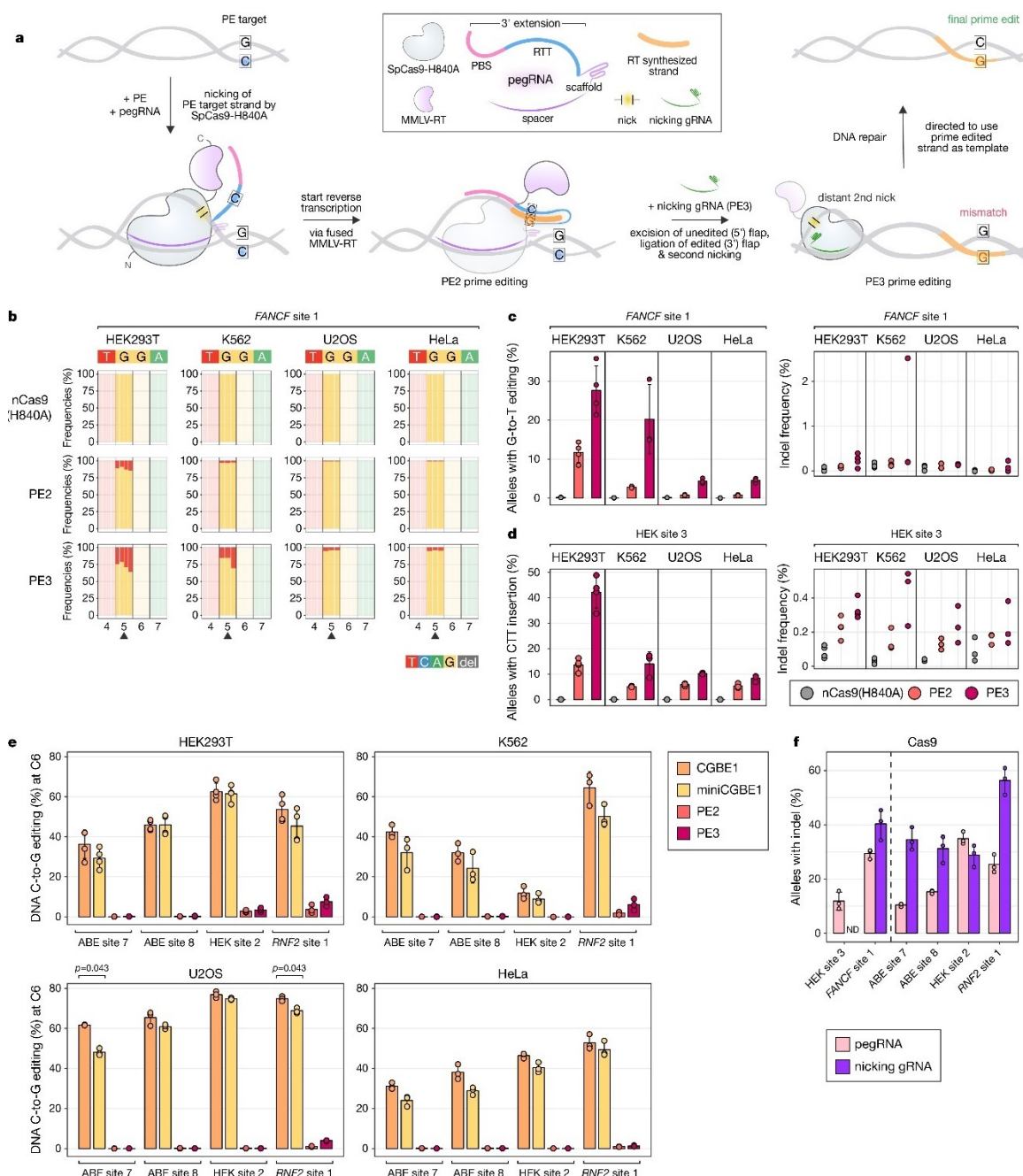
Extended Data Figure 5: Aggregated distribution of C-to-G editing frequencies across protospacer with CGBE1 and miniCGBE1 in HEK293T cells. a,b, Dot and box plots representing the aggregate distribution of C-to-G (yellow) editing frequencies per nucleotide across the entire protospacer from experiments performed with CGBE1 (a) and miniCGBE1 (b) with all 48 tested gRNAs. Boxes span the interquartile range (IQR; 25th to 75th percentile), horizontal lines indicate the median (50th percentile), and whiskers extend to $\pm 1.5 \times \text{IQR}$. Data points in plots represent full range of values plotted. Single dots represent individual replicates. The graphs were derived from the data shown in Fig. 3a,b ($n = 4$ independent replicates per site), and Supplementary Fig. 3a ($n = 3$ independent replicates per site).



Extended Data Figure 6: On-target DNA editing activities of NG and VRQR variants of CGBE1 and miniCGBE1 in HEK293T cells. a, Bar plots showing the on-target DNA base editing frequencies induced by NG and VRQR variants of nCas9, CGBE1, and miniCGBE1 using 6 gRNAs that target AT-rich genomic loci with PAMs that are compatible with SpCas9-NG (NGT) and SpCas9-VRQR (NGAG)

Extended Data Figure 6 (Continued)

variants in HEK293T cells. Editing frequencies of four independent replicates ($n = 4$) at each base are displayed side-by-side. **b**, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with NG and VRQR variants of CGBE1 and miniCGBE1 reported in **a**. Single dots or triangles represent individual replicates ($n = 4$ independent replicates).



Extended Data Figure 7: Comparing the editing activities of CGBEs and PEs in multiple human cell lines. a, Schematic of prime editing (PE) used to install a C-to-G substitution. PE fusion protein

Extended Data Figure 7 (Continued)

consists of an SpCas9-H840A nickase fused to an engineered Moloney murine leukemia virus reverse transcriptase (MMLV-RT). The prime editing guide RNA (pegRNA) consists of a standard targetable SpCas9 gRNA that also harbors a 3' extension containing a primer binding site (PBS) and a reverse transcription template (RTT) that encodes the desired edit. The PE2 system encompasses the prime editor fusion protein and a pegRNA. The PE3 system additionally includes a nicking gRNA (ngRNA). **b**, Bar plots showing the on-target DNA prime editing frequencies induced by nCas9(H840A), PE2 and PE3 using a pegRNA that targets *FANCF* site 1 across four human cancer cell lines. Gray overlay bars at top represent deletions at each nucleotide. Editing frequencies of four independent replicates ($n = 4$) for HEK293T cells or three independent replicates ($n = 3$) for K562, U2OS, and HeLa cells at each base are displayed side-by-side. Numbering on the bottom indicates the position of the base with 1 being the first nucleotide 3' of the pegRNA/Cas9-induced nick. Arrowheads indicate guanines that exhibit desired G-to-T prime edits. **c,d**, Bar and dot plots representing the average on-target DNA prime editing and indel frequencies of PE2 and PE3 targeting *FANCF* site 1 for G-to-T prime editing (**c**; data from the same experiment as **b**) and HEK site 3 for PE-induced CTT insertion (**d**) in 4 cell lines. Single dots represent individual replicates ($n = 4$ for HEK293T and $n = 3$ for K562, U2OS, and HeLa cells). Error bars represent standard deviation (s.d.). Measure of center for the error bars = mean. **e**, Bar and dot plots showing the average on-target DNA C-to-G base or prime editing frequencies induced by CGBE1, miniCGBE1, PE2 or PE3 on four genomic target loci. Single dots represent individual replicates ($n = 4$ for HEK293T and $n = 3$ for K562, U2OS, and HeLa cells). A two-tailed Student's t-test with p -values adjusted for multiple testing was used to calculate the shown p -values ($p = 0.043$ for both). Error bars represent (s.d.). Measure of center for the error bars = mean. **f**, Bar and dot plots representing the average frequency of alleles with indels (%) induced by pegRNAs and nicking gRNAs used in the experiments shown above (and *FANCF* site 1 +21 ngRNA control, Supplementary Table 3) with wild-type SpCas9 in HEK293T. pegRNAs/ngRNAs designed by Anzalone *et al.* (left) and by us (right) are separated by the dashed line. Single dots represent individual replicates ($n = 3$ independent replicates). Error bars represent (s.d.). Measure of center for the error bars = mean. ND, not done.

Methods

Molecular Cloning

All base editor (BE) and prime editor (PE) constructs (PE2, Addgene #132775) used in this study (**Supplementary Table 1**) were cloned into a mammalian expression plasmid backbone under the control of a pCMV promoter (AgeI and NotI restriction digest of parental plasmid Addgene #112101). The wild-type SpCas9 construct (SQT 817; Addgene #53373) used in **Extended Data Fig. 7f** is expressed under the control of a CAG promoter. All BE and PE constructs were encoded as P2A-eGFP fusions for co-translational expression of the base/prime editors and eGFP. Gibson fragments with matching overlaps were PCR-amplified using Phusion High-fidelity polymerase (NEB). Fragments were gel-purified and assembled for 1 hour at 50°C and transformed into chemically competent *E. coli* (XL1-Blue, Agilent). The UNGs used in our experiments originated either from *E. coli* (eUNG; UniProtKB-P12295) or *Homo sapiens* (hUNG; UniProtKB-P13051), were codon-optimized for expression in human cells and synthesized as gblocks (IDT). All guide RNA (gRNA) constructs (**Supplementary Table 3**) used in this study were cloned into a BsmBI-digested pUC19-based entry vector (BPK1520, Addgene #65777) with a U6 promoter driving gRNA expression. We designed the pegRNAs (**Supplementary Table 3**) to implement the same C-to-G changes that the CGBE constructs would install and followed previously described default design rules for designing pegRNAs and ngRNAs (Anzalone et al., 2019). PegRNAs were cloned into the BsaI-digested pU6-pegRNA-GG-acceptor entry vector (Addgene #132777) and ngRNAs were cloned into the above-mentioned BsmBI-digested entry vector BPK1520. For pegRNA cloning, oligos containing the spacer, the 5'phosphorylated pegRNA scaffold, and the 3' extension sequences were annealed to

form dsDNA fragments with compatible overhangs and ligated using T4 ligase (NEB). All plasmids used for transfection experiments were prepared using Qiagen Midi or Maxi Plus kits.

Cell Culture

STR-authenticated HEK293T (CRL-3216), K562 (CCL-243), HeLa (CCL-2), and U2OS cells (similar match to HTB-96; gain of #8 allele at the D5S818 locus) were used in this study. HEK293T and HeLa cells were grown in Dulbecco's Modified Eagle Medium (DMEM, Gibco) with 10% heat-inactivated fetal bovine serum (FBS, Gibco) supplemented with 1% penicillin-streptomycin (Gibco) antibiotic mix. K562 cells were grown in Roswell Park Memorial Institute (RPMI) 1640 Medium (Gibco) with 10% FBS supplemented with 1% Pen-Strep and 1% GlutaMAX (Gibco). U2OS cells were grown in DMEM with 10% FBS supplemented with 1% Pen-Strep and 1% GlutaMAX. Cells were grown at 37°C in 5% CO₂ incubators and periodically passaged upon reaching around 80% confluency. Cell culture media supernatant was tested for mycoplasma contamination every 4 weeks using the MycoAlert PLUS mycoplasma detection kit (Lonza) and all tests were negative throughout the experiments.

Transfection & Electroporation Experiments

HEK293T cells were seeded at 1.25×10^4 cells per well into 96-well flat bottom cell culture plates (Corning) for DNA on-target experiments or at 6.25×10^4 cells per well into 24-well cell culture plates (Corning) for DNA off-target experiments. 24 hours post-seeding, cells were transfected with 30 ng of control or base/prime editor plasmid and 10 ng of peg or gRNA plasmid (and 3.3 ng nicking gRNA plasmid for PE3) using 0.3 μ L of TransIT-X2 (Mirus) lipofection reagent for experiments in 96-well plates, or 150 ng control or base editor plasmid and 50 ng gRNA, and 1.5 μ L TransIT-X2 for experiments in 24-well plates. K562 cells were

electroporated using the SF Cell Line Nucleofector X Kit (Lonza), according to the manufacturer's protocol with 2×10^5 cells per nucleofection and 800 ng control or base/prime editor plasmid, 200 ng gRNA or pegRNA plasmid, and 83 ng nicking gRNA plasmid (for PE3). U2OS cells were electroporated using the SE Cell Line Nucleofector X Kit (Lonza) with 2×10^5 cells and 800 ng control or base/prime editor plasmid, 200 ng gRNA or pegRNA, and 83 ng nicking gRNA (for PE3). HeLa cells were electroporated using the SE Cell Line 4D-Nucleofector X Kit (Lonza) with 5×10^5 cells and 800 ng control or base/prime editor, 200 ng gRNA or pegRNA, and 83 ng nicking gRNA (for PE3). 72 hours post-transfection, cells were lysed for extraction of genomic DNA (gDNA).

DNA Extraction

HEK293T cells were washed with 1X PBS (Corning) and lysed overnight by shaking at 55°C with 43.5 µL of gDNA lysis buffer (100 mM Tris-HCl at pH 8, 200 mM NaCl, 5 mM EDTA, 0.05% SDS) supplemented with 5.25 µL of 20 mg/ml Proteinase K (NEB) and 1.25 µL of 1M DTT (Sigma) per well for experiments in 96-well plates, or with 174 µL DNA lysis buffer, 21 µL Proteinase K, and 5 µL 1M DTT per well for experiments in 24-well plates. K562 cells were centrifuged for 5 min, media removed, and lysed overnight by shaking at 55°C with 174 µL DNA lysis buffer, 21 µL Proteinase K, and 5 µL 1M DTT per well in 24-well plates. U2OS cells and HeLa cells were washed with 1X PBS and lysed overnight shaking at 55°C with 174 µL DNA lysis buffer, 21 µL Proteinase K, and 5 µL 1M DTT per well in 24-well plates. Subsequently, gDNA was extracted from lysates using 1-2X paramagnetic beads as previously described (Grünwald, Zhou, Garcia, et al., 2019) and eluted in 45 µL of 0.1X EB buffer. DNA extraction was performed using a Biomek FX^P Laboratory Automation Workstation (Beckman Coulter).

Targeted Amplicon Sequencing

DNA targeted amplicon sequencing was performed as previously described (Grünewald, Zhou, Garcia, et al., 2019). Briefly, extracted gDNA was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher). Amplicons were constructed in 2 PCR steps. In the first PCR, regions of interest (170-250 bp) were amplified from 5-20 ng of gDNA with primers containing Illumina forward and reverse adapters on both ends (**Supplementary Table 3**). PCR products were quantified on a Synergy HT microplate reader (BioTek) at 485/528 nm using a Quantifluor dsDNA quantification system (Promega), pooled and cleaned with 0.7X paramagnetic beads, as previously described (Grünewald, Zhou, Garcia, et al., 2019). In a second PCR step (barcoding), unique pairs of Illumina-compatible indexes (equivalent to TruSeq CD indexes, formerly known as TruSeq HT) were added to the amplicons. The amplified products were cleaned up with 0.7X paramagnetic beads, quantified with the Quantifluor or Qubit systems, and pooled before sequencing. The final library was sequenced on an Illumina MiSeq machine using the Miseq Reagent Kit v2 (300 cycles, 2x150bp, paired-end). Demultiplexed FASTQ files were downloaded from BaseSpace (Illumina).

Analysis and Plotting of NGS data

Amplicon sequencing data were analyzed with CRISPResso2 version 2.0.31 in batch mode (Clement et al., 2019). Downstream analysis was performed using R version 3.5.1 with data sourced from 'Nucleotide_percentage_summary.txt' for nucleotide distribution barplots and 'MODIFICATION_PERCENTAGE_SUMMARY.txt' tables for indels. Each point in position-wise indel plots (Fig. 2c indel frequency dot plots) is the sum of “Insertions_Left” and “Deletions” columns in the text file for each position in the protospacer. CRISPResso2 output

table “CRISPRessoBatch_quantification_of_editing_frequency.txt”, was used to quantify percentage of alleles that contain an insertion or deletion across the protospacer sequence for base/prime editor experiments, with window parameters set to -wc 10 -w 10 (Extended Data Figs. 2, 4b, 6b, and 7c, d and Supplementary Figs. 1b, 2c, 3b, 4b, 5b). For WT-Cas9 experiments (Extended Data Fig. 7f) default window parameters were used (-wc 3 -w 1) to quantify indels around the expected cut site. Each point on the plot is calculated from columns in the above table as follows: $\text{Insertions} + \text{Deletions} - \text{'Insertions and Deletions'}/ \text{Reads_aligned}$. Extended Data Fig. 7d: From tables “Alleles_frequency_table_around_sgRNA_*.txt” that are part of CRISPResso2 output, percentages of alleles that contained a CTT insertion in positions 18, 19 and 20 of the protospacer were summed and reported in Extended Data Fig. 7d (% alleles bar plot). Sums of percentages of other indels besides the above insertions were plotted separately in Extended Data Fig. 7d (indel frequency dot plot).

Statistics & Data Reporting

Boxes span the interquartile range (IQR; 25th to 75th percentile), horizontal lines indicate the median (50th percentile), and whiskers extend to $\pm 1.5 \times \text{IQR}$. Data points in box and dot plots represent full range of values plotted (Fig. 2c and Extended Data Fig. 5). A two-tailed Student’s t-test with *p*-values adjusted for multiple testing was used to calculate the *p*-values in Extended Data Fig. 7e. The error bars in all dot and bar plots show the standard deviation (s.d.) and were plotted with ggplot (R-3.5.1). The measure of center for the error bars is the mean. We did not predetermine sample sizes based on statistical methods. Investigators were not blinded to experimental conditions or assessment of experimental outcomes.

Data availability

Plasmids encoding CGBE1 (Addgene #140252) and miniCGBE1 (Addgene #140253), as well as other constructs used in this work are available on Addgene via https://www.addgene.org/Keith_Joung/. Targeted amplicon sequencing data (obtained from Illumina Basespace) have been deposited at the Sequence Read Archive (SRA): <https://www.ncbi.nlm.nih.gov/sra/PRJNA622835>. All other relevant data are available from the corresponding authors upon request.

Chapter 3: Engineering DNA Transversion G-to-T Guanine Base Editors (GBE)

Introduction

Genome editing tools are designed to enable users to install a specific mutation with high efficiency in a precise manner without inducing unwanted byproducts. The first-generation genome editing tools (meganucleases, TALENs, ZFNs, and CRISPR nucleases) function by introducing double stranded breaks (DSB) at the target loci in a programmable manner. However, this strategy is dramatically low in precision as the majority of the DSBs are resolved by the error-prone nonhomologous end joining (NHEJ) repair pathway leading to semi-stochastic, unpredictable (except microhomology-mediated end joining (MMEJ)), and thus uncontrollable outcomes mainly consisting of insertion/deletion (indel) mutations (Shen et al., 2018). While efficient introduction of indel mutations at target loci is a robust strategy to knock out genes by introducing frameshift mutations and disrupt noncoding regulatory elements by altering chromatin modifying enzyme or transcription factor binding sites, it proved impracticable to install precise alterations. To this end, homology-directed repair (HDR) cellular pathway that is mostly active in G2 and S phases of mitotic cells has been harnessed by the users to introduce predictable and precise alterations via introducing exogenous repair templates, albeit at low yield. Moreover, and most importantly, CRISPR-Cas9 nuclease has been shown to induce unintended on-target effects of chromothripsis, megabase-scale deletions, translocations, inversions and complex rearrangements that can have serious adverse consequences for patients (Leibowitz et al., 2021).

The second-generation CRISPR-based genome editing tools called base editors (BEs) drastically mitigated the precision problem by utilizing a nickase Cas9 (nCas9) (as opposed to a

nuclease Cas9) along with a diverse set of deaminases to modify target bases. Employing nCas9 virtually abolished the imprecise indel formation by eliminating generation of DSBs, and deaminases allowed chemical alteration of the target base(s) in a relatively tighter editing window. Until recently, BEs were available in two formats: cytosine base editors (CBEs) that can install cytosine-to-thymine (C-to-T) alterations, and adenine base editors (ABEs) that can install adenine-to-guanine (A-to-G) alterations. The underlying reason to why CBEs and ABEs were the initial BEs that were successfully engineered is at the heart of the problem for engineering a functional guanine base editor (GBE).

More specifically, spontaneous deamination of cytosines into uracils and adenines to inosines (albeit at lower frequency) occur across our genomes every day. Cells are equipped with the appropriate repair machinery called base excision repair (BER) to tackle such potential mutagenic lesions by excising and re-synthesizing the deaminated bases. It was feasible to engineer a CBE by employing the cytidine deaminase enzymes in nature that can already act on single-stranded DNA molecules. Therefore, fusion of a cytidine deaminase to an nCas9 was almost sufficient for generating a functional CBE (Komor et al., 2016). On the other hand, engineering of an ABE required directed evolution efforts (starting with the *E. coli* tRNA-specific adenine deaminase called tadA) to engineer a novel adenine deaminase that is functional on single-stranded DNA, since there were not any known enzymes in the nature that could catalyze this reaction (Gaudelli et al., 2017).

In this work, we sought to engineer a functional guanine base editor (GBE) by making use of the catalogued guanine lesions in nature. Fusing guanosine deaminase from *A. thaliana* and xanthine dioxygenase from *A. nidulans* to nCas9, we preliminarily observed intriguing C-to-

A (thus G-to-T) editing and indel activities at VEGFA site 1 in HEK293T cells and at three de novo sites in OGG1^{-/-} MUTYH^{-/-} CHO cells, respectively.

Results

Among different classes of modifications, we focused on the highly mutagenic 8-oxoguanine (8-OG) that is the product of spontaneous oxidation of guanines in DNA and RNA potentially leading to a G-to-T transversion mutation (Ba et al., 2014; Hsu et al., 2004; Yanli Wang & Schlick, 2007). Specifically, oxidation of C8 in guanine nucleotide via the unsaturated bond between C8 and N7 leads to protonation of N7 into NH, revealing an alternative scheme of hydrogen bond donors and acceptors that can Hoogsteen mispair with an adenine nucleotide (**Fig. 6**) (Faucher et al., 2012). We hypothesized that engineering an architecture that can oxidate genomic deoxyguanine nucleotides in a programmable manner could lead to a functional transversion GBE. However, there are no enzymes characterized up to date that can directly catalyze oxidation of guanine nucleotides into 8-OG.

We decided to consider alternative ways to generate 8-OG or a similar oxidized molecule that has the potential to behave similarly such as uric acid. Uric acid is the C2 deaminated form of 8-OG molecule that is the product of either oxidation of xanthine or direct deamination of 8-OG (R. S. Hall et al., 2010). It is therefore a necessary intermediate step to deaminate guanine into xanthine first to be able to generate uric acid via its further oxidation. Guanine as a free monomer in cytosolic pools can be deaminated into xanthine via guanine deaminase (guanase) enzymes. Although humans have a guanine deaminase called GDA, it is excessively large in size (454 amino acids) and can only deaminate guanines but not guanosines (nucleobase form) unlike the deaminase partners in functional CBEs and ABEs that can deaminate cytidines and

adenosines, respectively (Bucardo et al., 2021). Furthermore, we did not identify homology between human GDA and the deaminase enzymes used in CBE and ABE architectures (APOBECs and tadA, respectively) considering the homology as a premise for functionality (data not shown). Searching for a deaminase that can deaminate guanosine, we identified guanosine deaminase from *Arabidopsis thaliana* called GSDA that is 185 amino acids in size, can already interact with deoxyguanosine instead of guanine base, and has partial homology with tadA deaminase (**Fig. 7**) (Dahncke & Witte, 2013).

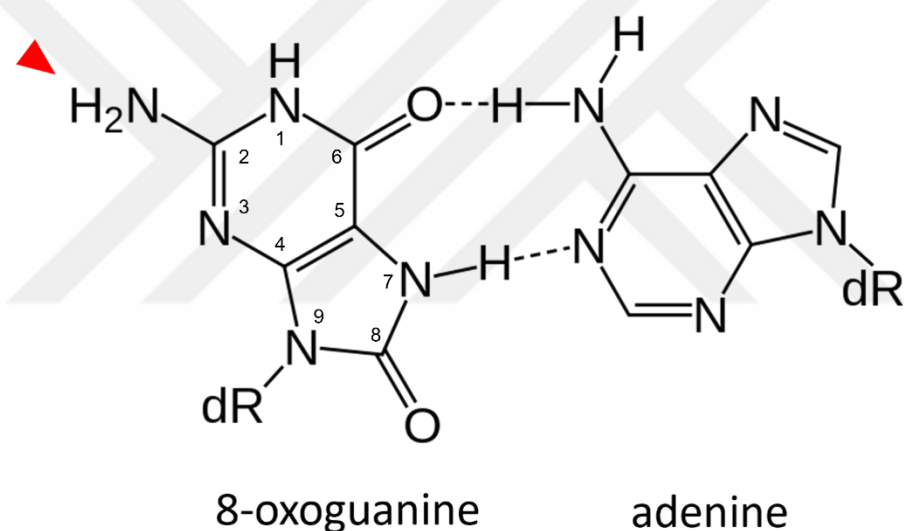


Figure 6. Hoogsteen mis-base pairing between 8-oxoguanine and adenine bases with the C2 amino group that is deaminated in uric acid marked with red triangle. Modified from illustration of Fvasconcellos with the author's common permission.

```

>Protein alignment Copy Alignment of 2 sequences: e coli tadA trna specific ade
nosine deaminase, GSDA_ARATH Guanosine deaminase OS=Arabidopsis thaliana

Identities = 48/193 (24%),
Positives = 79/193 (40%), Gaps = 34/193 (17%)

e coli tadA trna specific adenosine deaminase      1 MSEVEF-----SHEYWMRHALTLAKRAWDERE-VPVGAVLVHN  37
M E +                      S  ++ A+ A + D + P GAV+VHN
GSDA_ARATH Guanosine deaminase OS=Arabidopsis thaliana 1 MEEAKVEAKDGTISVASAFSGHQAVHDSHKKFLTQAVEEAYKGVDCGDDGGPFGAVIVHN  60

e coli tadA trna specific adenosine deaminase      38 NRVIGEGWNRPIGRHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHS  97
N V+   N +   DPTAHAE+ A+R+   +   L + +Y + EPC MC GA+ S
GSDA_ARATH Guanosine deaminase OS=Arabidopsis thaliana 61 NEVVASCHNMVLYKTDPTAHAEVTAIREACKKLNKIELSECEIYASCEPCMCFGAHLS 120

e coli tadA trna specific adenosine deaminase      98 RIGRVVFGARDAKTGAAG---SLMDVLHHFGMNRHVEITEGILADECAALLSDFFRMRRQ 154
R+ R+V+GA+   A G   + D L  G+ + +   AA+   F+ ++
GSDA_ARATH Guanosine deaminase OS=Arabidopsis thaliana 121 RLKRLVYGAKAEAAIAIGFDDFIADALRGTVGYQKSSEIKKADGNGAAIAEQVFQNTKE 180

e coli tadA trna specific adenosine deaminase      155 EIKAQKKAQSSTD 167
+ +
GSDA_ARATH Guanosine deaminase OS=Arabidopsis thaliana 181 KFRLLY----- 185

```

Figure 7. Protein alignment result of *E. coli* tadA and *A. thaliana* GSDA demonstrating 24% identical amino acids and 40% positive (identical or with similar chemical properties) amino acids between them.

Afterwards, we started searching for appropriate enzymes that can catalyze the second step of oxidation of xanthine into uric acid. Although human cells have xanthine dehydrogenase/oxidase (XDH) taking part in purine catabolism that can oxidate hypoxanthine to xanthine and xanthine to uric acid, it is 1333 amino acids in size making it as large as the SpCas9 protein itself (Saksela & Raivio, 1996). Whereas, Xanthine dioxygenase enzyme (XanA) from *Aspergillus nidulans* can also produce uric acid by oxidating xanthine via hydroxylation and is only 370 amino acids in size (Cultrone et al., 2005; Montero-Morán et al., 2007). Therefore, we cloned nCas9 (D10A mutant in RuvC domain) fused to GSDA from *Arabidopsis thaliana* and XanA from *Aspergillus nidulans* as the first candidate GBE transversion editor (called as GBE from now on). We hypothesized that higher local concentration of GSDA at the target site could force deaminate target guanine(s) into xanthine(s), and subsequently XanA could force oxidate target xanthine(s) into uric acids, that in return could be recognized as thymine(s) by DNA

polymerases during base excision repair induced by the Cas9-induced-nick in the target strand (Fig. 8).

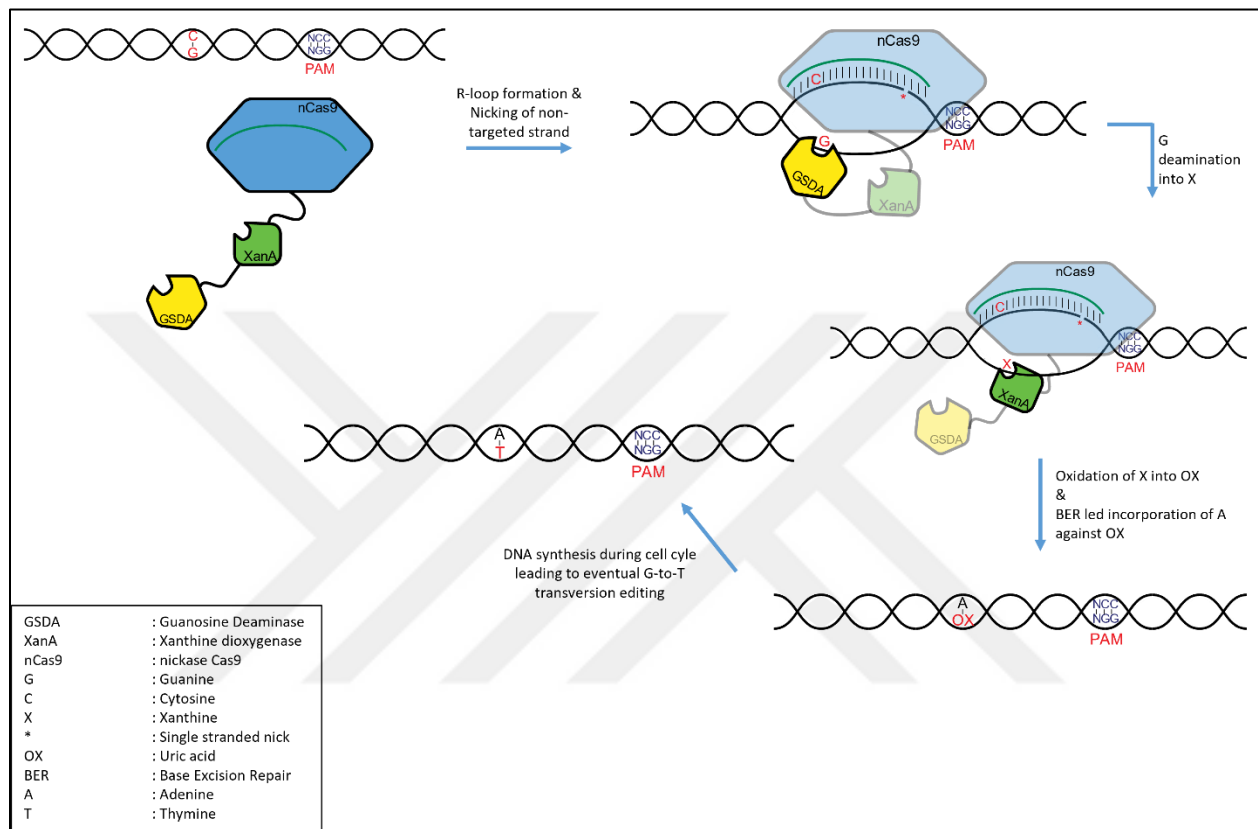


Figure 8. Schematic describing the hypothesized mechanism of action of GBE architecture as GSDA-XanA-nCas9 fusion deaminating guanine (G) into xanthine (X), and oxidating xanthine (X) into uric acid (OX). XO is recognized as thymine by the base excision repair machinery induced translesion polymerase, leading to stable genomic G-to-T mutation.

8-OG is one of the most common and damaging lesions found in human genome (Hsu et al., 2004; Van Loon et al., 2010). Therefore, cells have respective tight countermeasures to repair 8-OG back to guanine robustly and efficiently before the irreversible G-to-T transversion mutation takes place (Markkanen et al., 2013). 8-oxoguanine glycosylase (OGG1) in humans targets 8-OG robustly and excise it out to generate an abasic site (Faucher et al., 2012; Jayanth &

Puranik, 2017). This paves the path for the short-patch base excision repair (SP-BER) pathway that would faithfully reconstitute the original guanine back into its place. If DNA replication happens before the OGG1-induced SP-BER takes place, then an adenine in front of 8-OG can be misplaced by the cellular polymerases with varying rates (Personal contact with Dr. Bogdan I. Fedeles). Nonetheless, bypassing OGG1 and having an adenine misplaced at the opposite strand of the 8-OG lesion may not be sufficient to lead to an irreversible G-to-T transversion mutation, as there is an extra layer of protection built into the cellular machinery via mutY DNA glycosylase (MUTYH) in humans (Kundu et al., 2009). MUTYH, as opposed to OGG1, acts on the mis-incorporated adenine in the opposite strand. MUTYH excises the mis-base paired adenine from the A:8-OG base pair, and initiate the long-patch BER (LP-BER) pathway that would ultimately repair the lesion back to the C:8-OG stage (Markkanen et al., 2013). Therefore, we hypothesized that OGG1 and MUTYH enzymes need to be suppressed in order to efficiently generate G-to-T transversion mutations via the oxidation pathway.

To test our hypothesis and see whether our first-generation GBE construct works, we decided to target the GC-rich promoter region of VEGFA (VEGFA site 1 gRNA) gene in human HEK293T cells in the presence and absence of an OGG1 small chemical inhibitor (SML1697) simultaneously (H. Wang et al., 2020). Preliminary targeted deep sequencing data demonstrated a range of 0.20-0.31% alleles with C-to-A (therefore G-to-T in the opposite strand) alterations at VEGFA site 1 only in the presence of gRNA targeting VEGFA site 1 and OGG1-inhibitor (**Fig. 9**). The guanines that showed G-to-T alterations span a region of position -2 to +8 of the protospacer (numbering starting from the position most distal to the PAM) (**Fig. 9**). The average G-to-T editing frequency for all of the alleles with a G-to-T edit within the -2 to +8 window was 0.27%, whereas the total editing frequency of alleles thereof combined was 1.87%.

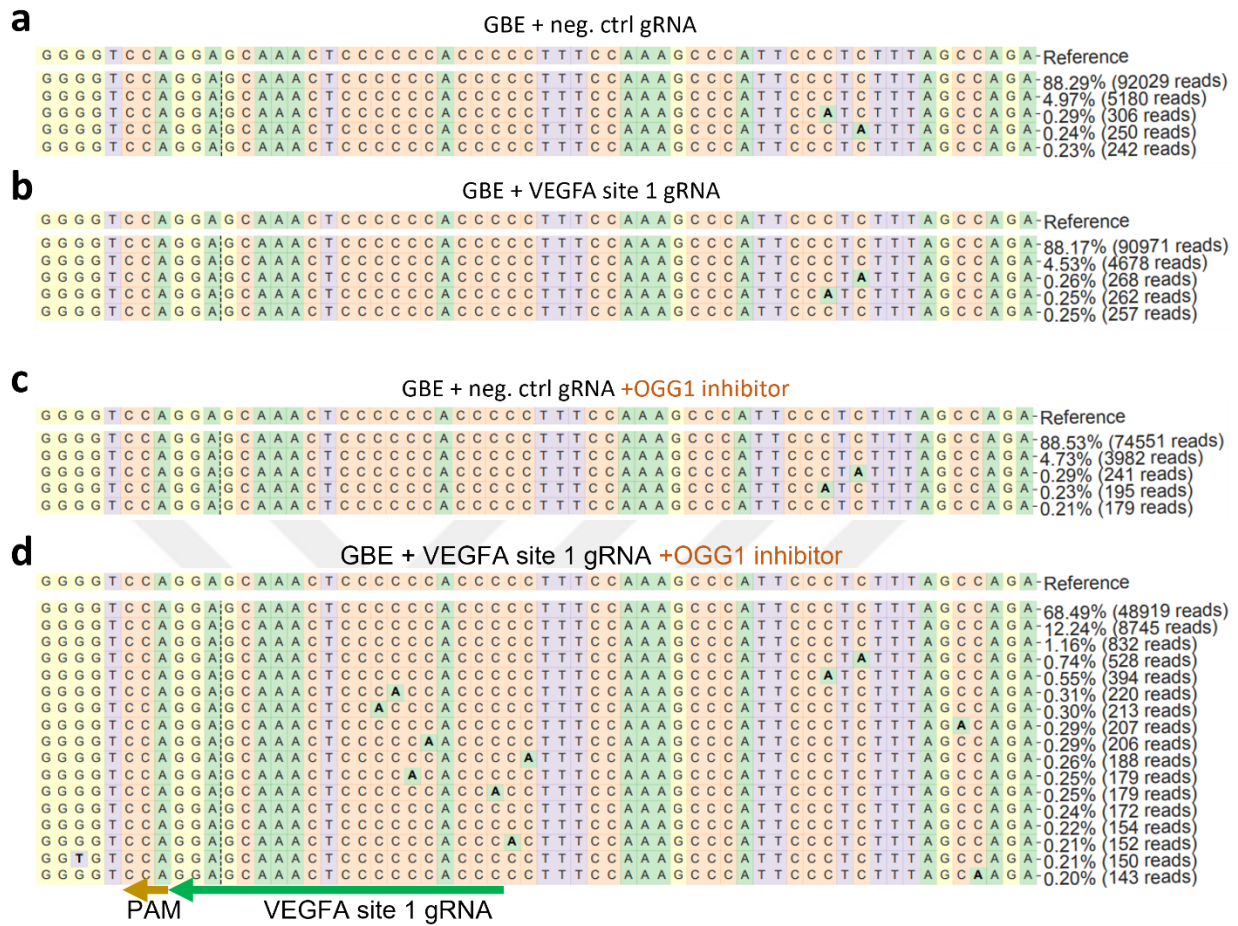


Figure 9. CRISPResso2 analysis of HEK293T cells treated with a) GBE (pCAG-wtGSDA-wtXanA-nSpCas9) + gRNA backbone (negative control without a spacer cloned in), b) GBE + VEGFA site 1 gRNA, c) GBE + negative control gRNA + OGG1 inhibitor (500 nM), d) GBE + VEGFA site 1 gRNA + OGG1 inhibitor (500 nM).

Given this observation in HEK293T cells treated with OGG1 inhibitor, we decided to attempt to completely bypass the cellular 8-OG repair machinery. To this end, we employed a OGG1 ^{-/-} MUTYH ^{-/-} double knockout CHO cell line, that is devoid of the major glycosylases to carry out base excision of 8-OG lesion (Tajai et al., 2018) (cells were kindly gifted by Jutamaad Satayavivad, John M. Essigmann, and Bogdan I. Fedeles). To carry out further testing of our GBE candidate, we designed 3 *de novo* guide RNAs (CHO site 1-3) targeting G-rich loci

found in Chinese hamster genome and treated the cells with the GBE candidate without the OGG1 inhibitor. All three sites demonstrated the hallmarks of efficient double stranded break formation at around the canonical cut site of SpCas9 (3 nucleotides upstream of PAM), although our GBE construct was carrying an impaired nicking SpCas9 with D10A mutation (**Fig. 10**). CHO site 1 displayed deletions spanning a region of 38 base pairs centered around 1 base upstream of the nCas9-induced expected nick with deletion efficiencies as high as 43.6%. Whereas CHO site 2 and 3 showed deletions spanning a region of 15 base pairs centered around 1 base downstream of the nCas9-induced expected nick with deletion efficiencies as high as 22.3% and 13.7% respectively, while CHO site 3 also displayed ~12% insertion frequencies right at the nick site (**Fig. 10**).

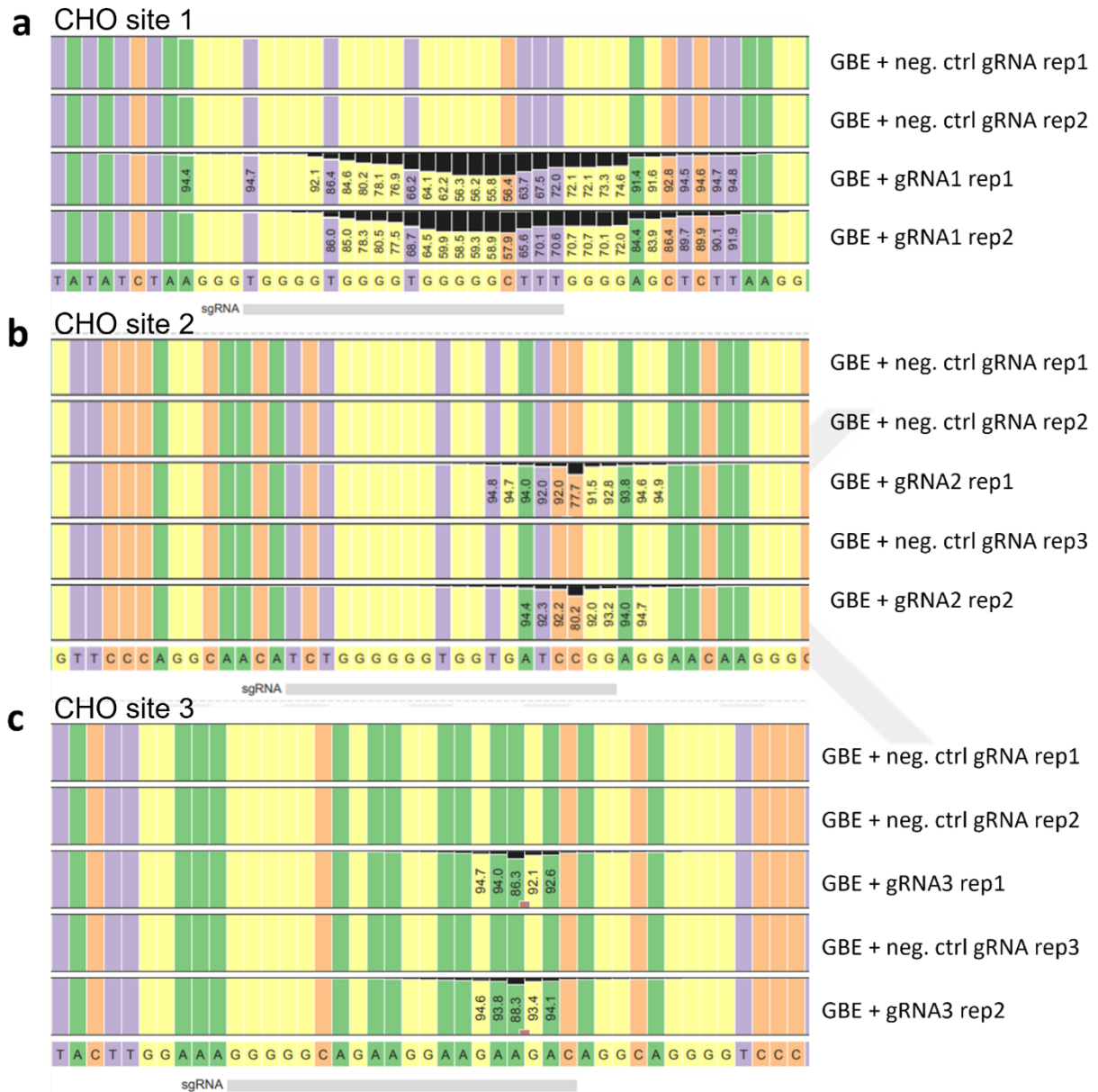


Figure 10. CRISPResso2 analysis of 3 genomic sites in OGG1 ^{-/-} MUTYH ^{-/-} double knockout CHO cells treated with GBE and cognate gRNA or negative control gRNA. a) CHO site 1, b) CHO site 2, c) CHO site 3. Black bars represent the deletions and brown bars protruding from bottom in plot c represent the insertions.

Discussion & Future Directions

Although our preliminary data warrants further experimentation at different target sites in different cell lines, the modest G-to-T activity in human HEK293T cells in the presence of

OGG1 inhibitor and the indel formation in double knockout CHO cells could possibly indicate that GBE is deaminating and oxidating the target guanines and xanthines thereof respectively, and thereby leading to either G-to-T alterations or indel formation. The tentative editing window observed in HEK293T cells also resembled the canonical editing window that has been reported with the CBEs and ABEs (Molla et al., 2021; Bei Yang et al., 2019).

We do not know whether uric acid as a nucleotide in genome would be recognized and pursued by the BER pathways similar to an 8-OG lesion. Carbonyl side chain of glutamine at position 315 in OGG1 enzyme serves as a hydrogen bond acceptor and hydrogen bonds with the NH₂ amine moiety at C2 of 8-OG (Faucher et al., 2012; Koval et al., 2014). Although it needs to be tested empirically, because glutamine can serve both as a hydrogen bond donor and acceptor via its amine and carbonyl side chains, it could be possible to still maintain the hydrogen bonding with the oxygen moiety at C2 of uric acid via its amine side chain (Shimoni & Glusker, 1995). Therefore, the main reason for the inefficient G-to-T editing in human HEK293T cells could be the robust repair machineries against 8-OG lesions active in mammalian cells. 8-OG is the most frequent (~4 lesions/10⁶ bases) and lethal lesion with higher fitness cost, as transversion mutations are likely to have more pronounced effects both in open reading frames and noncoding regions compared to transition mutations (Gedik & Collins, 2005; Guo et al., 2017). Nonetheless, human lymphoblastoid cells have been reported to be able to faithfully reconstitute 94.1% of the synthetic 8-OGs in exogenously supplied reporter genes, i.e., only 5.9% of the 8-OG lesions led to stable G-to-T transversion mutations (Yasui et al., 2014), demonstrating the extreme fidelity of 8-OG repair machineries in human cells. Therefore, it is plausible to hypothesize that inhibiting OGG1 alone is insufficient or incomplete to reach higher levels of G-to-T editing. Co-treating the cells with siRNA or shRNA against MUTYH along with the OGG1 inhibitor could

reveal whether there will be synergy between MUTYH knockdown and OGG1 suppression. Since nCas9 (D10A) employed in this study concomitantly nicks the target strand (that is the non-modified strand) with the aim of rapidly inducing BER to install a mis-paired adenine opposite to the uric acid, it can be further hypothesized that inhibiting MUTYH is more pertinent compared to OGG1 to increase G-to-T editing efficiency, as MUTYH is mainly responsible for excising the adenine in a A:8-OG mis-pair (Markkanen et al., 2013). This hypothesis needs to be assessed by testing our GBE construct in cell lines with homozygous MUTYH-knockout or MUTYH-knockdown at transcriptional level via si/shRNA or MUTYH-suppressed by small-molecule inhibitors (Tahara et al., 2018, 2019; Visnes et al., 2018).

Indel formation as unwanted byproducts can be observed with CBE, ABE, and CGBE base editors (Gaudelli et al., 2017; Komor et al., 2016; Kurt et al., 2020). Because nCas9 component of BEs nick target-strand, and deaminase component deaminates the nontarget-strand, successful AP-site generation at the modified base followed by backbone digestion will lead to DSB formation at varying degrees depending on the rate of deaminase, presence of glycosylase inhibitor moieties or other BER components such as UNG. The reason behind the differential indel patterns at different CHO sites can be due to existence of differential microhomology sequences flanking the theoretical cut site. Significant +1 base insertion at cut site at CHO site 3 could also be induced by the previously reported context-dependent Cas9-catalyzed staggered ends upon DNA cleavage (Z. Zuo & Liu, 2016). InDelphi computational method can be employed to dissect out the potential microhomologies that could be driving the patterns of indel outcomes and predict 1-bp insertion frequencies at the target loci (Shen et al., 2018).

If OGG1 and MUTYH are the two main enzymes responsible for repairing 8-OG lesions via BER pathways, then what is the rationale behind the robust indel formation induced by GBE in OGG1 $-/-$ MUTYH $-/-$ double knockout CHO cell line? This could be explained by the vast redundancy observed in between BER pathways and overlapping substrate ranges of pyrimidine and purine glycosylases (Hazra et al., 2007; Hegde et al., 2008; Krokan & Bjørås, 2013). As an example, uracil-DNA glycosylase (UNG), thymine-DNA glycosylase (TDG), and single-strand selective monofunctional uracil DNA glycosylase (SMUG1) have common substrate of uracil in genomic DNA with varying efficiencies, while still retaining their distinct substrate preferences (Darwanto et al., 2009). Correspondingly, there is also substantial redundancy of mammalian DNA glycosylases for oxidized bases; human genes of mismatch-repair pathway component MutS/MSH2, Nei endonuclease VIII/NEIL1 and NEIL2, OGG2, and NTH1 has been reported to possess varying levels of 8-OG excision activities, especially towards 8-OG:A mis-pair (Hazra et al., 2002; Hegde et al., 2008; Markkanen et al., 2013; Yukiko Matsumoto et al., 2001). Explicit redundancy that works as a back-up mechanism for crucial oxidized lesions such as 8-OG can also be inferred from single and double knock out mice models. OGG1 and MUTYH single and double knockout mice are completely viable and fertile (Isogawa, 2004; Xie et al., 2004). Moreover, genetic compensation upon gene knockout has been observed in multiple organisms and cell lines that could also account for the efficient indel formation in double knockout CHO cell line upon GBE treatment (El-Brolosy & Stainier, 2017). This hypothesis can be tested by assessing the transcriptional or translational expression level changes of the glycosylases that can excise 8-OG in parental wildtype and double knock out CHO cell lines side by side via qPCR or western blotting, respectively.

There are several limitations of this study. Having included an nCas9-only negative control in CHO cell experiments could have informed us whether GSDA or XanA enzymatic components of our GBE construct was directly responsible for the observed high frequency indel formation across the three sites. Moreover, targeting the GC-rich sites selectively could bias the editing outcomes as GG or GGG stretches of multiple guanines has been reported to be more amenable to oxidation, reflecting the sequence context dependent manner of 8-OG production (Faucher et al., 2012).

We envision that our work will spur the field in future to engineer functional and well-characterized GBE architectures building upon our initial hypothesis. We further believe that, similar to the directed evolution of tadA from *E. coli* to make it accept adenosines in single stranded genomic DNA as part of the ABE work, rationale design and directed evolution strategies could yield an evolved XanA enzyme that can directly oxidate guanosines into 8-OG (Gaudelli et al., 2017). Ability to install G-to-T transversion alterations would unlock a new suit of amino acid alterations that could not be installed using the current transition BEs, and empower users to introduce nucleotide variations into the noncoding genome with more pronounced effects compared to transition mutations (Guo et al., 2017). Most importantly, a functional GBE tool would enable the correction of additional pathogenic single nucleotide variations that otherwise would not be accessible with the current base editor technologies (Rees & Liu, 2018).

Chapter 4: Developing Sensitive, Scalable and High-Throughput Off-target Analysis Platforms

Tn5-Assisted Circularization Sequencing (TACO-seq) & Sensitive Capture of Oligonucleotides Post-Editing Sequencing (SCOPE-seq)

TACO-seq method was conceived by Ibrahim C. Kurt. All experiments were designed, and data were generated by I.C.K. SCOPE-seq project was conceived by Jonathan Y. Hsu and conducted in collaboration with. I.C.K. I.C.K. purified and prepared the Tn5 complexes and optimized Tn5 tagmentation parameters, designed the ddPCR primers/probes for pre-NGS QC. J.Y.H. carried out all steps of GUIDE-seq and SCOPE-seq experiments starting with the pre-loaded Tn5 complexes, including PCRs, NGS and data analysis. Specifically, Fig. 19 was generated by J.Y.H.

Introduction

Off-target analysis methods to assess the specificity of widely adapted CRISPR-Cas9 nuclease has been instrumental in the field to characterize fidelity thereof, especially in the clinical setting. Understanding and minimizing off-target dynamics of Cas9 are crucial as in vivo and ex vivo gene therapy applications often involve batch treatment of hundreds of millions to billions of cells, maximizing the odds for any unwanted generation of clones with deleterious or tumorigenic off-target mutations (Frangoul et al., 2021; Mancuso et al., 2020).

CRISPR off-target analysis methods can be broadly categorized into two complementary categories as in vitro or in vivo. In vitro methods leverage the controllable biochemical

environment, such as the concentrations of the purified genomic DNA and Cas9/gRNA RNP complex, and rationally designed downstream adapter ligation and enrichment steps to substantially increase the sensitivity (Bao et al., 2020; Tsai & Joung, 2016). However, the purified genomic DNA employed in *in vitro* techniques is stripped from its chromatin structures such as nucleosomes that form the epigenetic landscape -although an effort was put forth to use intact chromatin DNA with histones-, and therefore the output off-target sites make up a ranked nominated superset that needs to be validated in cells/*in vivo* (Bao et al., 2020; Horlbeck et al., 2016; D. Kim & Kim, 2018). *In vivo* validation of the nominated sites above the determined cutoff by targeted deep sequencing are then promoted to bona fide off-target sites (Atkins et al., 2021). *In vivo* off-target platforms, on the other hand, captures the natural biochemical and epigenetic environment within the nucleus during Cas9's activity and lifetime at the cost of sensitivity (Bao et al., 2020; X. Chen et al., 2017). This is particularly important given the observation that Cas9 is more active at the open chromatin (euchromatin) regions such as active enhancers and promoters, and actively transcribed genes (Lazzarotto et al., 2020).

Digenome-seq was the first unbiased *in vitro* CRISPR off-target platform to assess the genomic off-target profiles of nuclease Cas9 (D. Kim et al., 2015). Digenome-seq lacked an enrichment step and relied on whole genome sequencing (WGS) the Cas9-digested fragmented genomic DNA to bioinformatically identify the signatures of Cas9-cleaved sites. Therefore, it required around 400 million raw reads per sample, rendering it a cost-ineffective technique outside of reach of a common lab.

CIRCLE-seq method later addressed this drawback by introducing a subsequent enrichment step based on circularization of the target genomic DNA molecules pre-Cas9 treatment (Tsai et al., 2017). The circularized DNA molecules were then digested with purified

Cas9/gRNA RNP complex to cleave and linearize the on-target and potential off-target sites, allowing free-ends to be selectively captured for deep sequencing. CIRCLE-seq thereby reduces the required sequencing depth by two orders of magnitude (around 4 million raw reads per sample) compared to Digenome-seq. However, CIRCLE-seq is labor and material intensive with 2-days long protocol with long hands-on timing and therefore hard to scale.

On the other side, genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq) has remained the gold standard in vivo off-target assessment technique since its advent in 2014 (Chaudhari et al., 2020; Frangoul et al., 2021; Tsai et al., 2014). GUIDE-seq entails co-transfection of short blunt-ended double-stranded oligodeoxynucleotides (dsODN) as baits and plasmids encoding Cas9 and gRNA. Cas9-induced DSBs in the population of cells incorporate the dsODN baits via NHEJ pathway at certain frequency. Genomic DNA from the treated cells is then isolated, sheared by sonication, end-repaired, A-tailed, and adapter ligated for the subsequent two-sided dsODN-specific PCR amplifications, identifying the locations of the Cas9-cleaved off-target sites. Although GUIDE-seq has proved itself to be the most accurate (low false positive-rate) and sensitive (can detect off-target sites with >0.1% indel frequency) unbiased genome-wide in vivo off-target profiling method to date, GUIDE-seq systematically suffers from low signal-to-noise ratio (~0.25% on-target/total raw NGS reads) (Tsai et al., 2014) (**Fig. 11**).

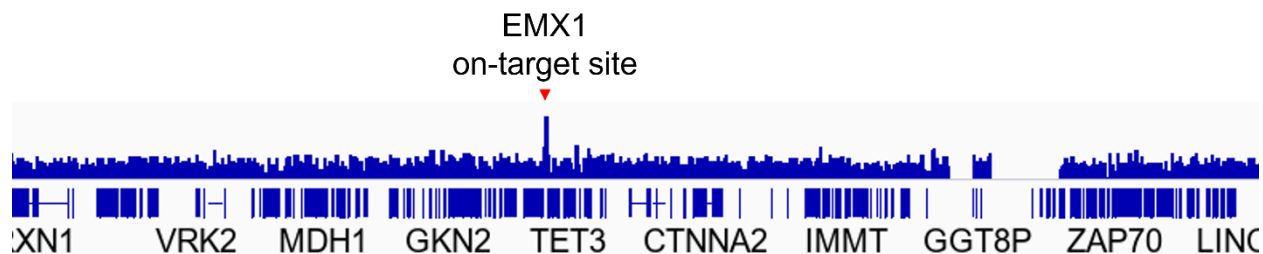


Figure 11. Integrative Genomics Viewer (IGV) screenshot displaying the mapped NGS reads from a GUIDE-seq sample treated with EMX1 gRNA + Cas9 nuclease, mapping across chromosome 2. Higher signal peak at the EMX1 on-target site is marked by the red triangle. Y-axis, # of reads mapped to region; x axis, genes.

Transposases are enzymes that can recognize its cognate transposon DNA (cargo) and translocate it to another genomic position in a cut-and-paste mechanism. Tn5 transposase binds to its cargo via one of the flanking 19-bp end sequences (ES), forms a dimer complex, excises the cargo out, and integrates the cargo into a new genomic locus in a virtually random manner (although conflicting reports exist regarding target specificity and preference) (Adey et al., 2010; Gunasekera et al., 2021). Although Tn5 is virtually inactive in normal conditions, modifying the ES into mosaic end sequences (ME) and introducing E54K and L372P mutations into Tn5 leads to drastically higher transposition frequencies (Picelli et al., 2014). Hyperactive Tn5 has been repurposed by loading it with short dsODN adapters with ME sequences at the end and harnessed as a tool to fragment high molecular weight (HMW) DNA molecules, and been commercialized in parallel (Gunasekera et al., 2021). Tn5 has been employed to fragment genomic DNA of prokaryotes and eukaryotes with as small input DNA as picograms, while combining fragmentation and adapter-ligation into a single quick step (Parkinson et al., 2012; Picelli et al., 2014; Vonesch et al., 2021). This process called enzymatic shearing or tagmentation is rapid, controllable, scalable, and automatable.

In this project, we sought to integrate the Tn5-based tagmentation into the frameworks of CIRCLE-seq and GUIDE-seq methods with the goals of increasing scalability and sensitivity, respectively. For CIRCLE-seq, we observed that the time-consuming physical shearing (sonication), end-repair, A-tailing, and adapter ligation consecutive steps with wash steps in between each lead to substantial material loss at each step and in return demands as high as 25 micrograms of starting input genomic DNA material from hundreds of millions of cells (Tsai et al., 2017). We hypothesized therefore that combining all of the aforementioned steps into a single 7-minute enzymatic reaction would not only make the protocol significantly more time-efficient, but also substantially reduce the amount of input genomic DNA required (**Fig. 12**).

For GUIDE-seq, we hypothesized that the low signal-to-noise ratio is mainly due to the presence of excess free (uncaptured) dsODNs floating in sample solution throughout the protocol, randomly ligating to the background genomic DNA fragments along with the appropriate PCR-adapters. Adapter-ligated PCR-compatible background genomic DNA molecules would then dominate the subsequent enrichment PCR reactions leading to the observed extremely low signal-to-noise ratio (Tsai et al., 2014) (**Fig. 11**). We hypothesized that employing Tn5 complexes loaded with PCR-compatible adapters to directly tagment the isolated GUIDE-seq-treated HMW (unsheared) genomic DNA would substantially bypass the background-generating ligation step and thus largely prevent the excess free dsODNs from ligating to the PCR adapters, leading to a gain in signal and loss in noise.

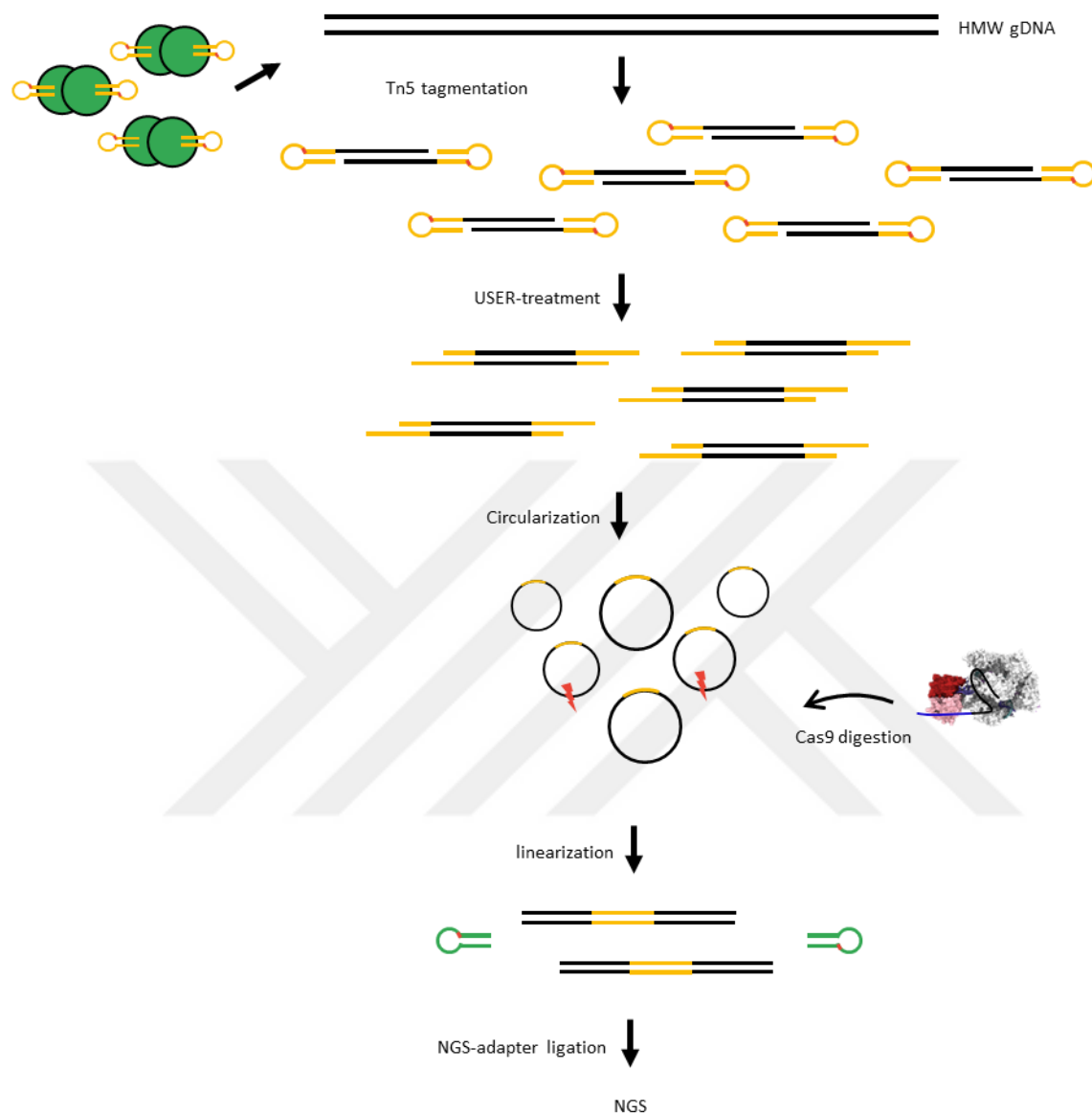


Figure 12. Schematic overview of the TACO-seq approach. High-molecular weight genomic DNA undergoes tagmentation, during which DNA is simultaneously fragmented and tagged with hairpin adapters containing uracil (shown as red marks). Subsequent USER-treatment opens up the stem-loop and exposes complementary sticky ends. Ligated and circularized DNA molecules are treated with Cas9-gRNA RNP complex. Cas9-cleaved target sites are captured by subsequent NGS-compatible adapter ligation step.

Results

Purification and optimization of homemade Tn5 transposase

We purified hyperactive Tn5 using chitin column elution system and ensured that the final elute was free from other visible contaminant proteins from the producer bacteria (**Fig. 13**). Then, we loaded the purified Tn5 transposases with custom adapters containing MEs on their ends and verified the functionality of custom Tn5 transposome complexes in a time-course experiment (**Fig. 14**). This experiment also ensured that the final buffer of Tn5 was free from DNase contamination.

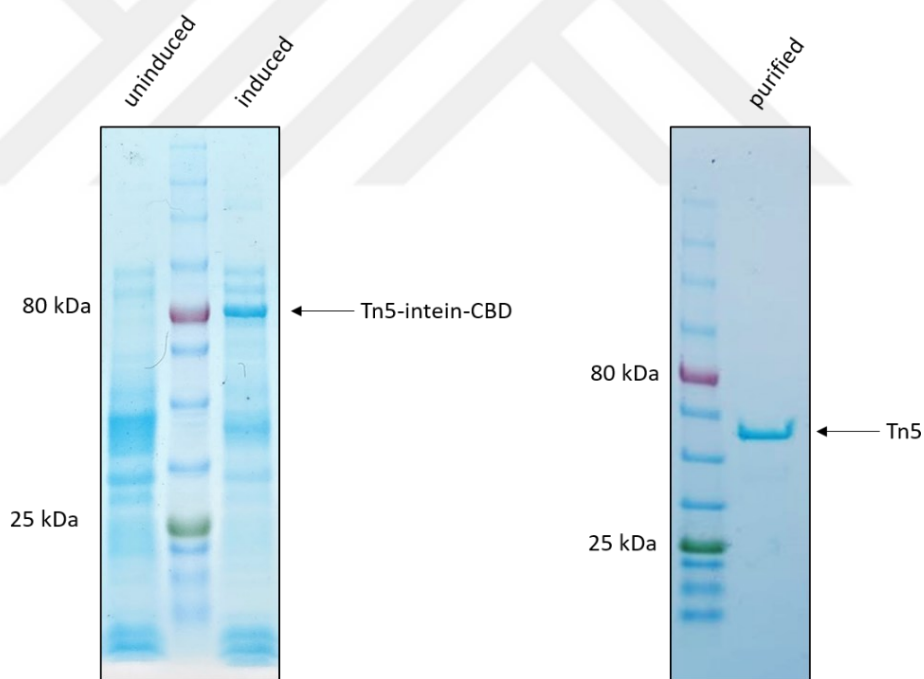


Figure 13. Coomassie-stained SDS-PAGE protein analysis of the crude bacterial lysate (on left) comparing the IPTG-induced and uninduced bacteria transformed with the Tn5 construct, with protein ladder in between. Chitin column-purified Tn5 in the final elute (on right) after removal of the intein-CBD domain, with protein ladder on left.

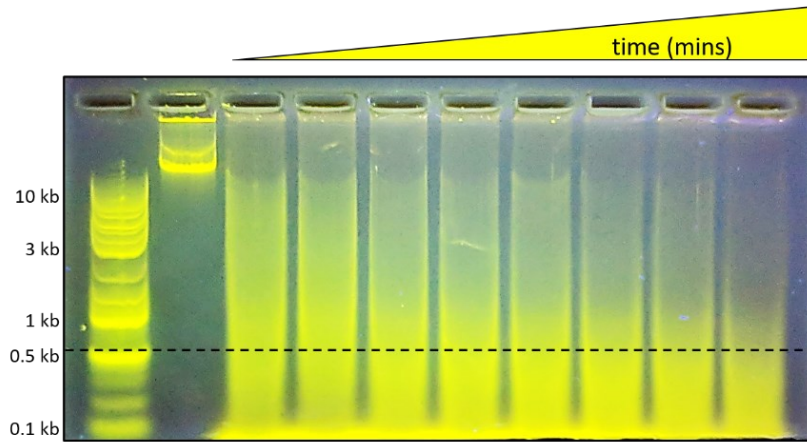


Figure 14. Time-course experiment showing SYBR gold nucleic acid-stained agarose gel electrophoresis of 50 ng untreated HMW control (well #2) and Tn5-treated genomic DNA at 55°C for 1, 2, 3, 4, 5, 6, 7 and 8 minutes (well #3 to #10 from left to right). Well #1 is loaded with DNA ladder. Well #2 is gDNA with Tn5 only without the adapters loaded as negative control.

Tn5-Assisted Circularization Sequencing (TACO-seq)

Upon characterization of our homemade Tn5, we wanted to test its utility by integrating it into the framework of CIRCLE-seq. CIRCLE-seq begins with randomly shearing 25 micrograms of genomic DNA from cells to an average of ~300 bp fragments. This is followed by end-repair, A-tailing and hairpin adapter ligation with a uracil-containing stem-loop. Fragments that have failed to be flanked by the covalently closed DNA adapters are then removed by exonuclease treatment. Subsequent USER treatment excises the uracils in the flanking stem-loops of the adapters, generating new sticky ends that are compatible for subsequent intra-molecular ligation. DNA ligation step generates the final circular DNA molecules.

We designed custom uracil-containing Tn5 hairpin adapters replacing the complimentary region with Tn5 ME sequence that would generate 4 nucleotide 3' sticky end upon USER treatment similar to the CIRCLE-seq scheme (**Fig. 15**). Since transposition of the Tn5

transposome complex at target site generates 9 nucleotide 5' overhangs, this leads to 9 base gaps between the 5' ends of the cargo and 3' ends of the tagmented genomic DNA (Chaudhari et al., 2020). Therefore, we included an immediate subsequent gap-filling and nick-sealing step before proceeding with the remaining circularization steps.

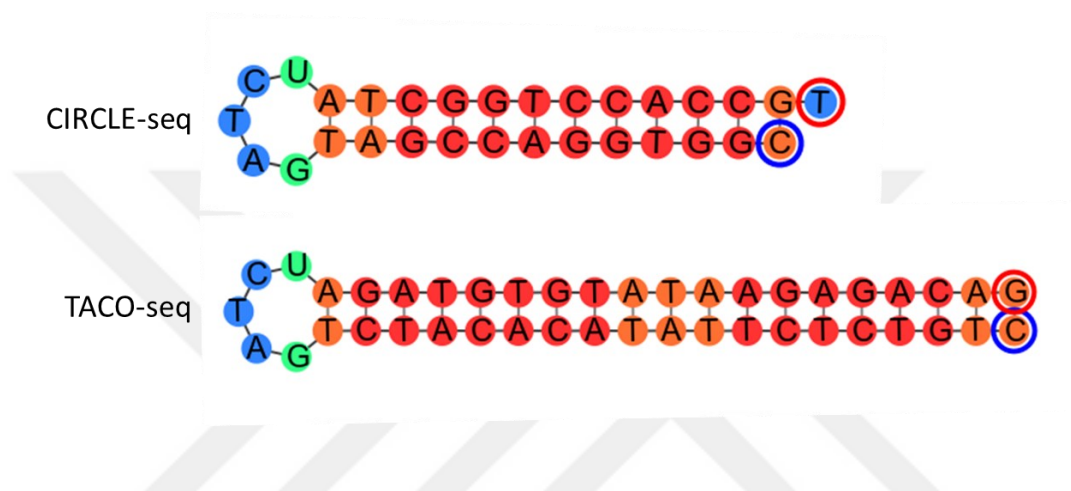


Figure 15. DNA fold schematic of the uracil containing stem loop adapters used in CIRCLE-seq and TACO-seq. Red and blue circles mark the 5' and 3' ends of the adapters, respectively. Note the sticky end of CIRCLE-seq adapter that is compatible with A-tailed genomic DNA fragments.

We wanted to test whether TACO-seq method rendered the system more efficient by quantifying the amount of genomic DNA recovered post-tagmentation and clean-up, and found that TACO-seq was 6.5-fold more efficient compared to CIRCLE-seq. This increase in efficiency allowed us to start with 4 micrograms of input genomic DNA material instead of 25 micrograms CIRCLE-seq employs. To test the sensitivity of TACO-seq compared to CIRCLE-seq, we ran TACO-seq and CIRCLE-seq side-by-side using purified genomic DNA from HEK293T cells targeting EMX1 site with Cas9/gRNA RNP complex. Digital droplet PCR (ddPCR) analysis using different sets of probes and primers designed for the genomic EMX1 protospacer and for the flanking NGS adapters, we indirectly quantified the number of Cas9-cleaved on-target events and total NGS library size within the end products. Compared to

TACO-seq, CIRCLE-seq yielded 4.5-fold higher on-target signal within around half amount of the total NGSable fragment size of TACO-seq. Adapter-dimers comprised ~23.8% and ~7.9% of the NGS library prepared by CIRCLE-seq and TACO-seq, respectively (**Fig. 16**).

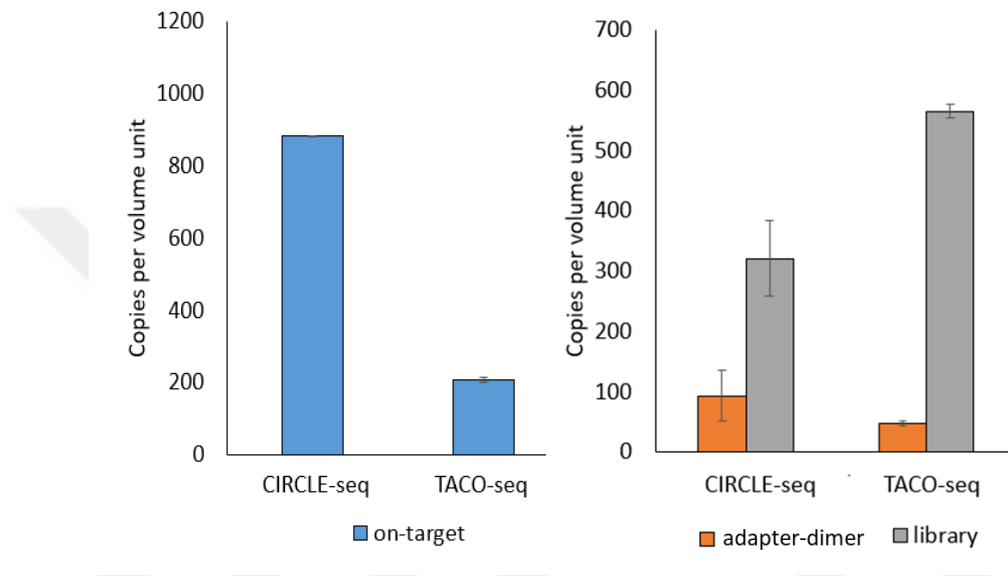


Figure 16. ddPCR analysis of CIRCLE-seq and TACO-seq end products using HEK293T genomic DNA that was digested with Cas9/gRNA RNP complex targeting EMX1 site. Blue bars represent the on-target signal, orange bars represent the adapter dimers, gray bars represent the total NGSable fragments (NGS library size).

Sensitive Capture of Oligonucleotides Post-Editing Sequencing (SCOPE-seq)

As part of a separate project, we also wanted to test our hypothesis that prepping the GUIDE-seq-treated purified genomic DNA by tagmentation post-dsODN capture within cells would lead to a more sensitive in vivo off-target analysis method compared to GUIDE-seq (**Fig. 17**). GUIDE-seq begins by nucleofecting target cells with Cas9 and gRNA encoding plasmids along with dsODN baits. 72-hours post-transfection, genomic DNA from population of cells is harvested, mechanically sheared down to ~500 bp fragments, end-repaired, and A-tailed for the ultimate ligation of Illumina TruSeq P5 NGS adapter.

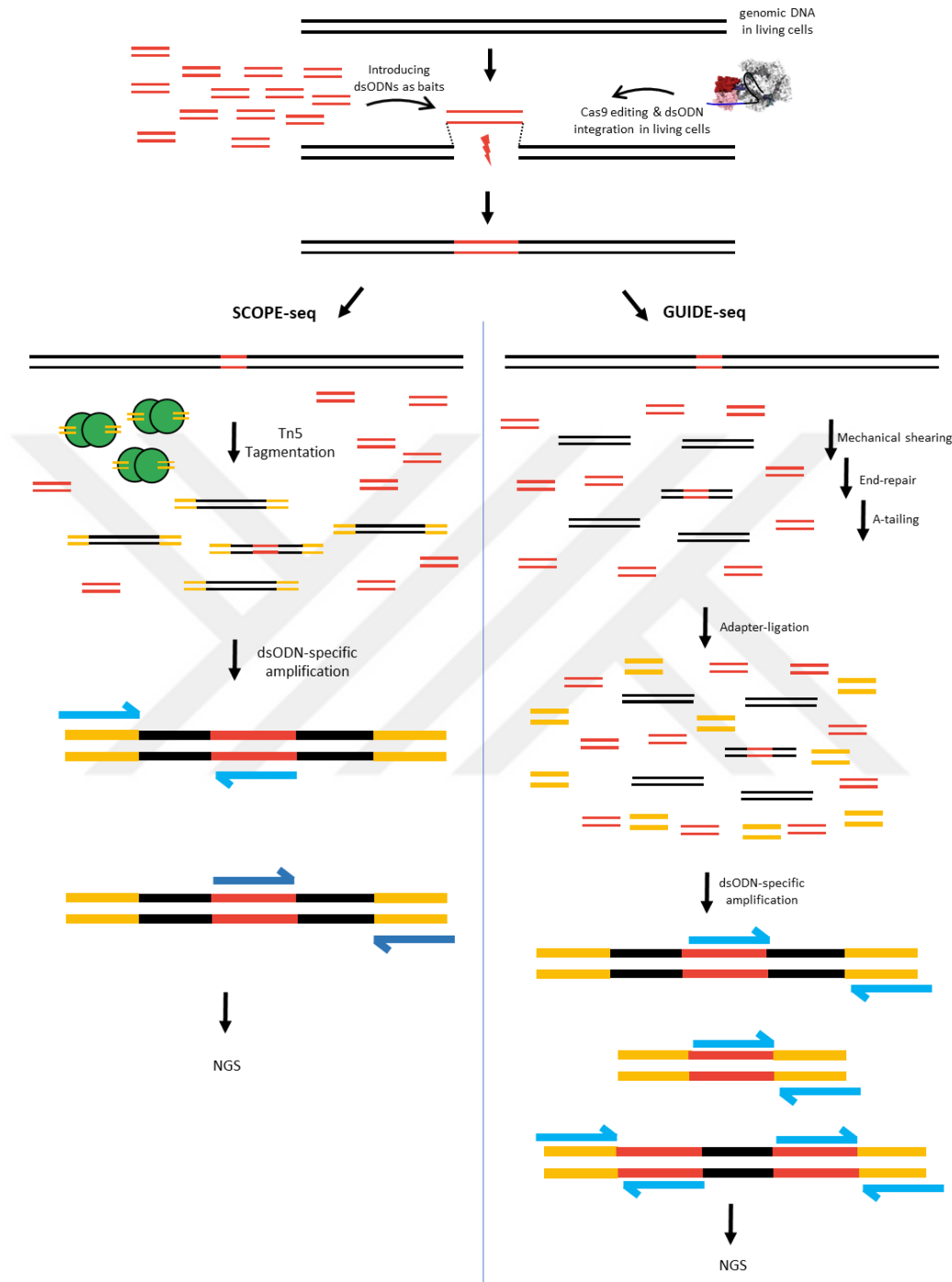


Figure 17. Schematic overview of SCOPE-seq and GUIDE-seq methods. Living cells are co-transfected with plasmids encoding Cas9 and guide RNA, and dsODN baits. dsODNs are captured at DSB sites cleaved by Cas9. Purified genomic DNA is then enzymatically sheared and tagged with adapters in a single step using Tn5 tagmentation in SCOPE-seq framework, followed by dsODN-specific two-sided PCR and NGS. Purified genomic DNA, during GUIDE-seq on the other hand, is mechanically sheared, end-repaired, A-tailed, and adapter ligated in the

Figure 17 (Continued)

presence of excess free dsODNs carried over from cells. Finally, ligated fragments are amplified using dsODN-specific primers for the subsequent NGS application. Note the hypothetical adapter-dsODN-adapter byproducts and dsODN-flanked nonspecific genomic DNA fragments that could lead to low on-target signal and high genomic background during NGS.

As an adaptation in SCOPE-seq framework, we designed custom P5 NGS adapters replacing the Illumina TruSeq Read 1 sequence with the Illumina Nextera Read 1 sequence that is Tn5 ME compatible (**Fig. 18**). This way we were able to use the same NGS platform and run the GUIDE-seq and TACO-seq samples side-by-side in the same flow cell without the need to spike in custom sequencing oligos.

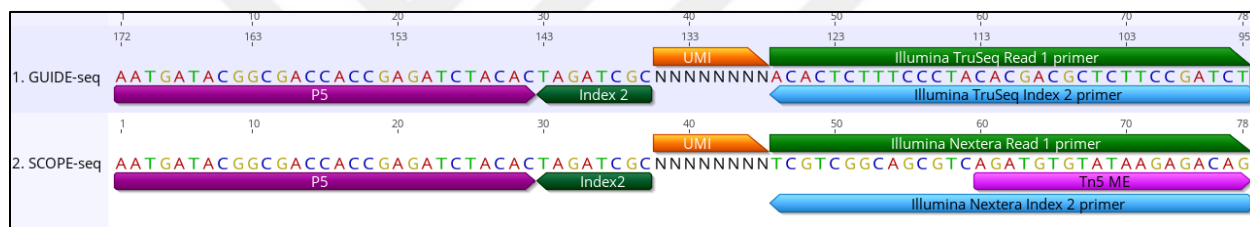


Figure 18. Side-by-side comparison of the P5 NGS adapters containing 8-mer Unique Molecular Indices used in GUIDE-seq and SCOPE-seq frameworks, respectively. Light purple element highlights the 19-nucleotide long Tn5 ME element. Dark purple highlights the Illumina P5 flow cell attachment sequence. Adapter sequences are shown as 5' to 3' from left to right.

We tested our hypothesis by nucleofecting HEK293T and K562 human cell lines with dsODNs along with plasmids containing Cas9 and gRNA targeting RNF2 or EMX1 or HEK site 4 genomic sites (stringent, medium, and promiscuous guide RNAs, respectively, in terms of specificity as previously defined by GUIDE-seq (Tsai et al., 2014)). We ran GUIDE-seq and SCOPE-seq methods in parallel using the same harvested genomic DNA per cell line and per target site. Post-NGS QC demonstrated that sequencing depths of samples were comparable (**Fig. 19**). Analyzing the generated NGS data through GUIDE-seq bioinformatics pipeline showed increased on-target signal and off-target detection rates in samples prepped with SCOPE-seq

compared to GUIDE-seq in 2 out of 3 sites in K562 cells, and 3 out of 3 sites in HEK293T cells (**Fig. 20-21**). There were roughly two-fold of more off-target sites identified in samples prepped with SCOPE-seq compared to GUIDE-seq in HEK293T cells treated with EMX1 and HEK site 4 gRNAs, and in K562 cells treated with HEK site 4 gRNA, while GUIDE-seq detected one more off-target site at RNF2 site in K562 cells compared to SCOPE-seq (**Fig. 20**). Samples prepped with SCOPE-seq yielded 1.5-fold more normalized on-target reads in average across 6 samples (range of 0.8-2.2-fold). All samples treated with SCOPE-seq, except RNF2 site in K562 cells, had higher normalized on-target signal (8562 vs. 15891, and 11887 vs. 12676 reads for GUIDE-seq vs SCOPE-seq at EMX1 site in HEK293T and K562 cells, respectively. 18734 vs. 25742, and 23850 vs. 18263 reads for GUIDE-seq vs SCOPE-seq at RNF2 site in HEK293T and K562 cells, respectively. 7012 vs. 15460, and 7392 vs. 11469 reads for GUIDE-seq vs SCOPE-seq at HEK site 4 in HEK293T and K562 cells, respectively.) (**Fig. 21**). Whereas, GUIDE-seq treated sample at RNF2 site in K562 cells had 1.3-fold more normalized on-target signal compared to SCOPE-seq.

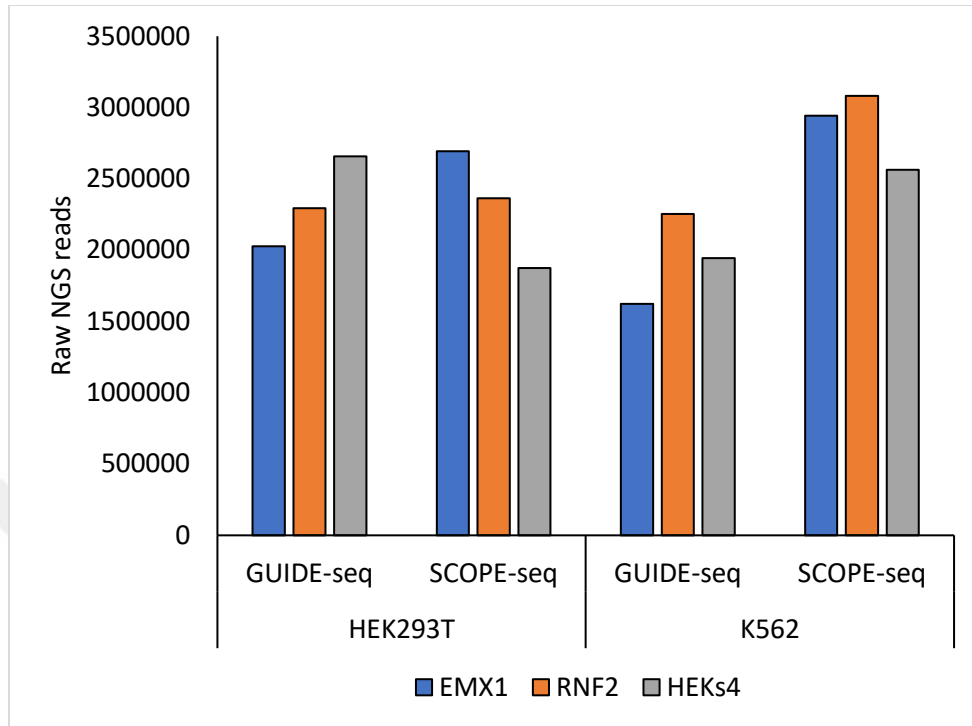


Figure 19. Side-by-side comparison of # of total raw NGS reads (sequencing depth) for GUIDE-seq and SCOPE-seq prepped samples.

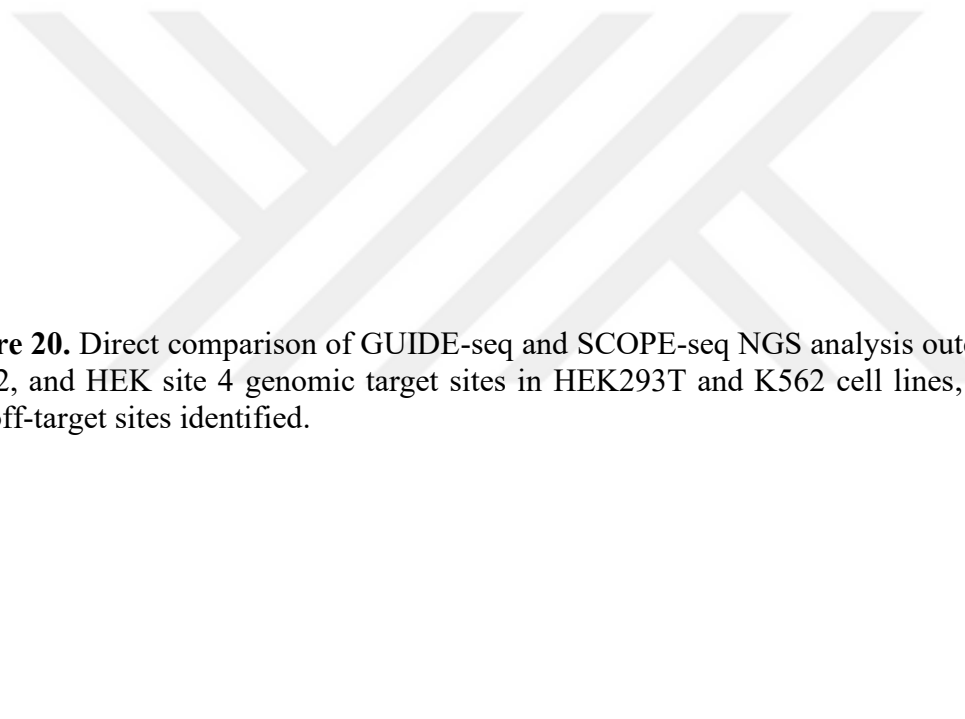


Figure 20. Direct comparison of GUIDE-seq and SCOPE-seq NGS analysis outcomes at EMX1, RNF2, and HEK site 4 genomic target sites in HEK293T and K562 cell lines, listing on-target and off-target sites identified.

Figure 20 (Continued)

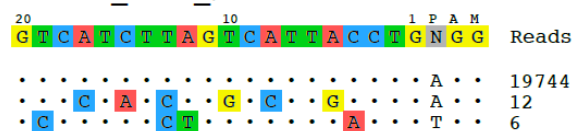
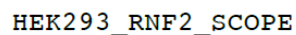
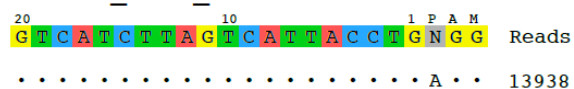
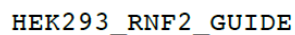
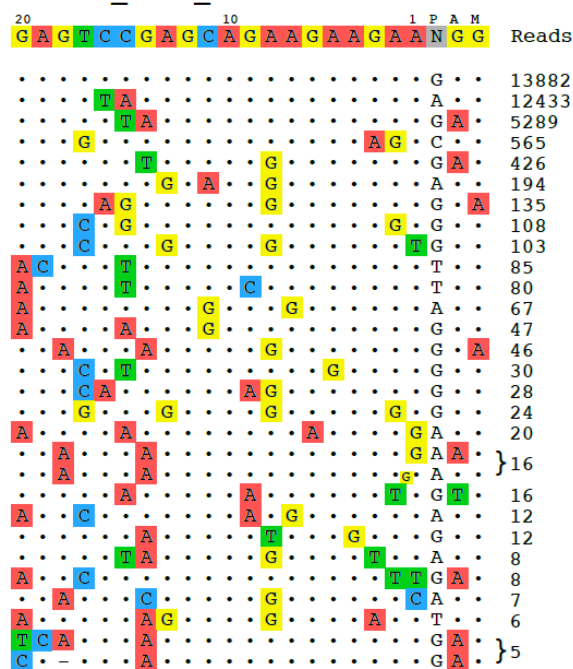
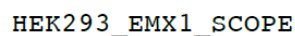
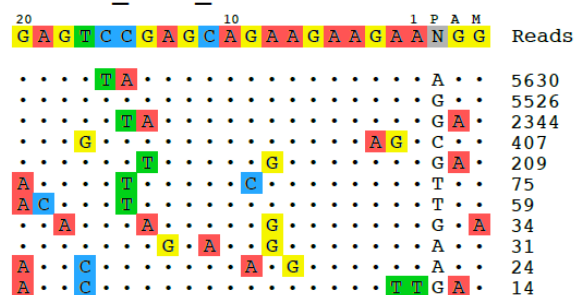
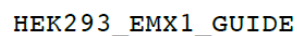


Figure 20 (Continued)

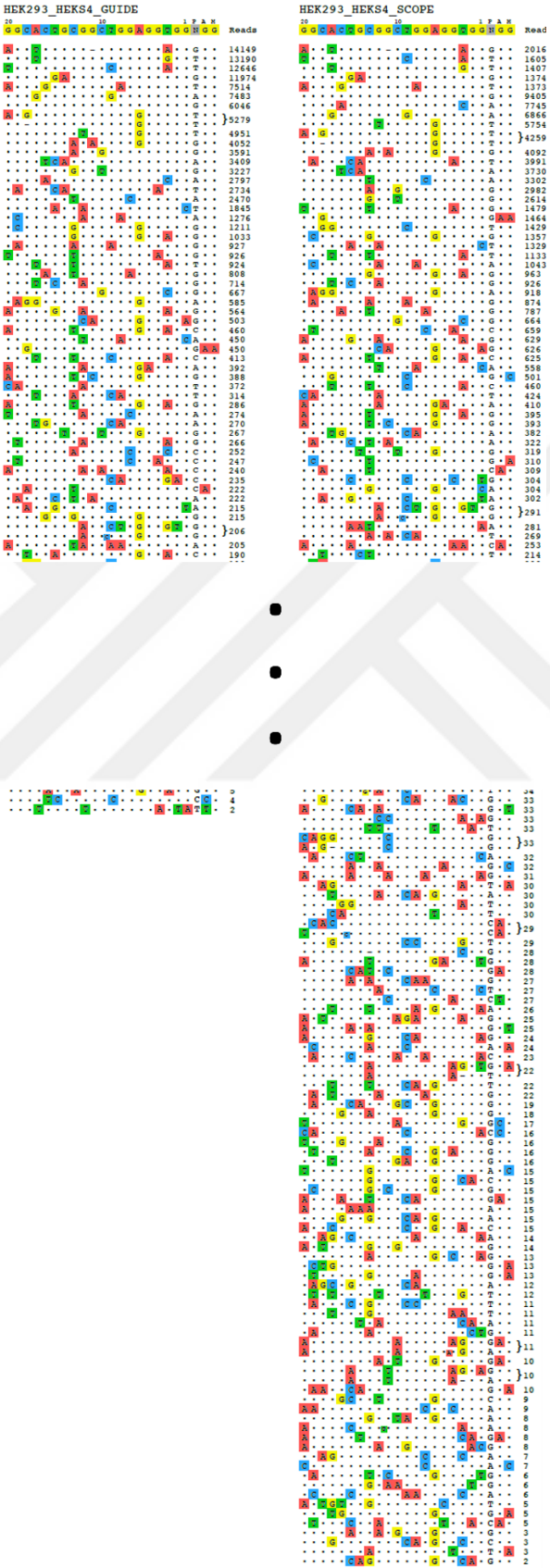


Figure 20 (Continued)

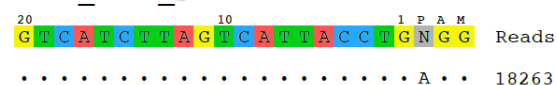
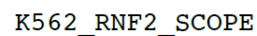
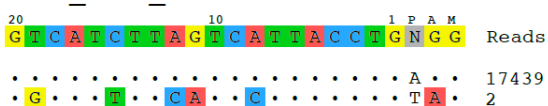
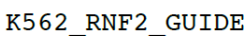
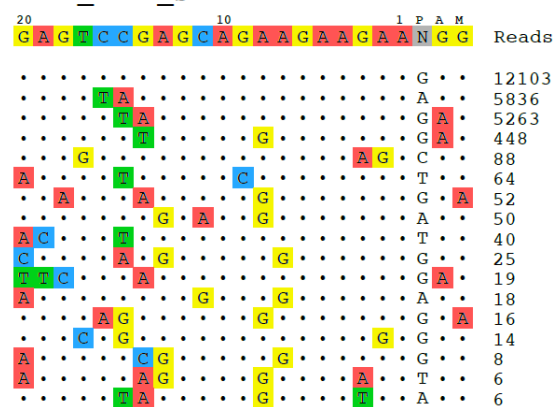
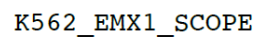
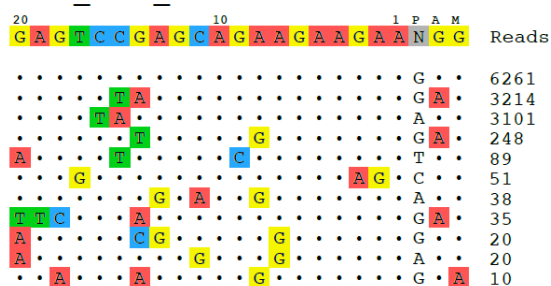
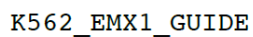
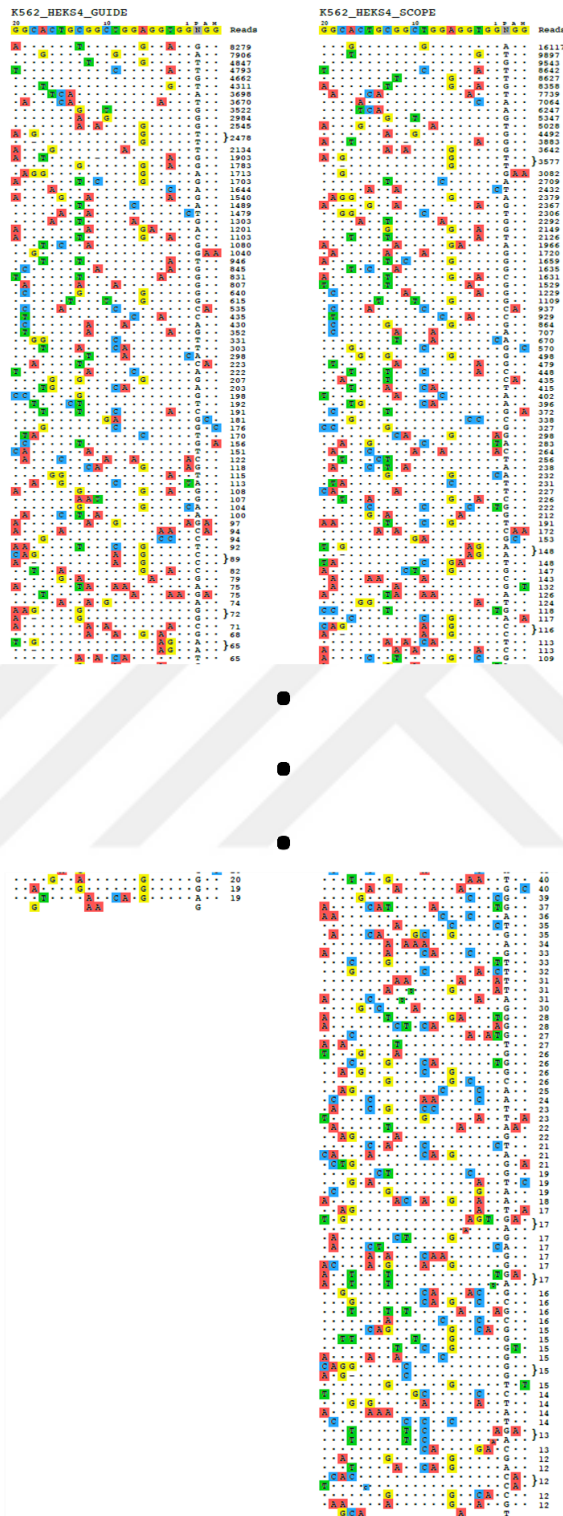


Figure 20 (Continued)



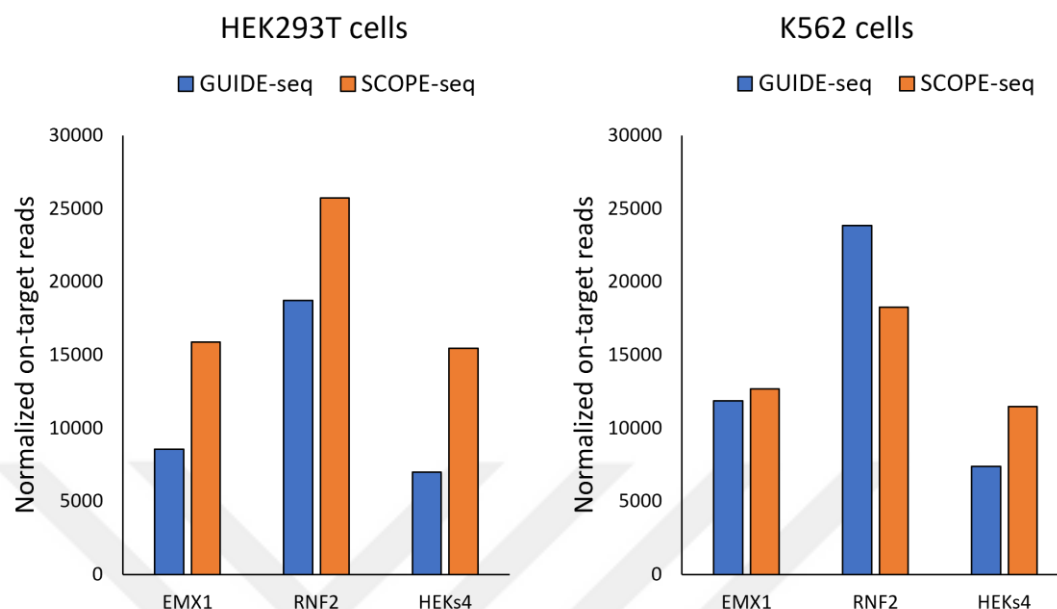


Figure 21. Normalized on-target reads from samples prepped with GUIDE-seq and SCOPE-seq at EMX1, RNF2, and HEK site 4 genomic target sites in HEK293T and K562 cell lines.

Discussion & Future Perspectives

Our preliminary data from TACO-seq suggested that Tn5-assisted circularization-based off-target characterization of Cas9 has the potential to substantially reduce the amount of the input genomic DNA used (>5-fold) and protocol time (2-fold), while increasing the scalability, albeit at the cost of sensitivity. TACO-seq in its first-pass form demonstrated an order of magnitude less sensitivity compared to CIRCLE-seq, although the unwanted adapter-dimer formation was significantly improved presumably due to the complete elimination of the unspecific and robust adapter ligation step (Tsai et al., 2017). It must be noted that ddPCR is a proxy method for quantifying the number of Cas9-cleaved on-target molecules and NGSable total fragments. Samples need to be tested via NGS analysis for the most accurate head-to-head comparison. Further optimization for TACO-seq was warranted, but we decided not to pursue

the project further. Two years later in 2020, a Tn5-tagmentation adaption of CIRCLE-seq called CHANGE-seq was published at Nature Biotechnology (Lazzarotto et al., 2020). Interestingly, CHANGE-seq also decreased the amount of input genomic DNA 5-fold down to 5 micrograms, while preserving the sensitivity levels compared to CIRCLE-seq. The reason that could account for the higher sensitivity in CHANGE-seq could be the blunt-ended linear scheme of the uracil containing Tn5 adapters the authors utilized, as opposed to our hairpin adapters, that could be more efficiently digested by USER enzyme mix.

On the other hand, SCOPE-seq results from two cell lines across three sites demonstrated that employing Tn5-based tagmentation strategy can increase the sensitivity of the platform and identify novel candidate off-target sites compared to the current gold-standard GUIDE-seq method (Tsai et al., 2014). This is presumably largely due to the complete elimination of the critical ligation step for adding the P5 side of the NGS adapters. In the conventional GUIDE-seq protocol, fragmented and repaired genomic DNA molecules along with the free/uncaptured bait dsODNs ($\sim 10^8$ dsODNs per cell transfected) carried over from cells are incubated with $\sim 10^{24}$ number of adapter molecules with T4 DNA ligase for 60 minutes. T4 DNA ligase is one of the most efficient ligases that is used widely in molecular cloning techniques with fast turnover rate and high catalytic efficiency reported (Lohman et al., 2011). We hypothesized that this is the main step generating the excess amount of background genomic reads by allowing free dsODNs to ligate to NGS-compatible adapters and genomic DNA fragments in an unspecific manner, therefore diminishing the Cas9-induced signal-to-noise ratio. Tn5-based enzymatic shearing of the HMW genomic DNA is a promising technique to minimize the ligation of carried-over free-dsODNs to the NGS adapters and to the genomic DNA fragments, as the minimum insert length of Tn5 transposome complexes reported is ~ 35 base pairs while dsODNs are 34 base pairs (Adey

et al., 2010; Tsai et al., 2014). The novel off-target sites that were identified by SCOPE-seq need to be validated by targeted amplicon sequencing. Previously, it has been shown that 80% of the GUIDE-seq-identified off-target sites can be validated by targeted amplicon sequencing, while the main bottleneck has been thought to be related to the sensitivity of sequencing (Tsai et al., 2014). The prediction that incorporation of the Tn5 system into the GUIDE-seq framework would enhance the sensitivity of the platform was also reported before (Giannoukos et al., 2018). Around a year after we tested SCOPE-seq, a scalable Tn5-based in vivo off-target method called TTISS built upon GUIDE-seq was developed and published at Molecular Cell (Schmid-Burgk et al., 2020). Although they reported that TTISS demonstrated comparable sensitivity to GUIDE-seq, TTISS was able to detect more off-target sites at 60 target sites in multiplexed manner (i.e. co-transfecting 60 gRNAs and dsODNs in a single experiment).

Although Tn5 has been commercialized as part of a series of NGS library preparation kits (amplicon, whole genome, and whole exome sequencing) by Illumina for over a decade now, Tn5 insertion preferences and potential context-dependent biases with conflicting reports in the literature must be still vetted carefully (Gunasekera et al., 2021). Secondary comparative NGS analysis focusing on the background reads from samples prepped by sonication or Tn5-based tagmentation could reveal potential coverage biases.

In Cellulo Self-targeting Off-target Platform to Score Fidelity of CRISPR-based Genome Editing Tools

This project was conceived by Ibrahim C. Kurt and was carried out in collaboration with bioinformatician M. Nafiz Hamid and lab technician Patrick Exconde. I.C.K. designed the experiments, the in silico off-target sequences and the off-target libraries. I.C.K. cloned and viral packaged the viral-library and generated the stable library lines. P.E. transfected the stable library lines with the editors, cloned the Tol2-library, and carried out NGS experiments under supervision and guidance from I.C.K. M.N.H. carried out the bioinformatic analysis post-NGS and prepared the plots in Fig. 22-31 except Fig. 26. Vikram Pattanayak co-supervised the work and provided a list of potential off-target sites found in human genome for our guides.

Introduction

Approximately half of the pathogenic alleles catalogued in human are single nucleotide variations (SNV) (Landrum et al., 2016; Stenson et al., 2013). CRISPR-based base editing tools hold tremendous potential towards precisely correcting the pathogenic SNVs (Rees & Liu, 2018). In order for base editing technologies to fully realize its therapeutic potential, high fidelity base editors with minimal genome-wide and transcriptome-wide off-target editing activities must be engineered. Protein engineering and data mining efforts have led to the advent of engineered cytosine and adenine base editors with minimized CRISPR-independent DNA and RNA off-target events (Grünwald, Zhou, Garcia, et al., 2019; Grünwald, Zhou, Iyer, et al., 2019; Yu et al., 2020). However, higher fidelity base editors have mostly been tested for their unguided or CRISPR-independent off-target editing activities of the deaminases they harbor. To be able to

translate the promising higher fidelity base editors into clinic, their guided or CRISPR-dependent off-target editing tendencies and dynamics must be characterized systematically.

Although significant work has been done using massively multiplexable synthetic in vivo on-target libraries to characterize the hidden features governing the base editor and on-target dynamics by training machine learning models, no one has designed in cellular off-target libraries for base editing tools, yet. To this end, in this work, we sought to design and develop complex in cellular self-targeting off-target library containing 5000 members to dissect out the rules governing optimal guide RNA design for highest specificity, and to assign overall DNA fidelity scores to the base editor tools reported to have minimal spurious off-target editing activities.

Results

We designed a lentiviral self-targeting (paired) off-target library with 5,000 members containing spacer sequences of 6 guides (EMX1, RNF2, FANCF, HEK2, HEK3, HBBgX19) along with their cognate on-targets, in silico-designed off-targets with single mismatches, and up to ~3 mismatches and ~2 bulges found in human genome (**Fig. 22**). This lentiviral library was integrated into host HEK293T cells at MOI 10 and selected against antibiotic resistance to yield a stable cell line carrying the library (**Fig. 23**).

To test the stable cell line carrying the off-target library, we treated the cells with SpCas9 nuclease, the canonical CBE BE4max, and high-fidelity CBE variant SECUREmax (R33A) (Grünwald, Zhou, Iyer, et al., 2019). NGS data was generated and reads failing to pass the filter were discarded. This analysis demonstrated that ~25-40% of reads by average failed to pass the QC filter (**Fig. 24**). The distribution of the library members undivided, or as divided into sub-libraries per guide, showed patterns of long tails and high skew-ratios of 5-11 (described as the ratio of 90th/10th percentile) (**Fig. 25**).

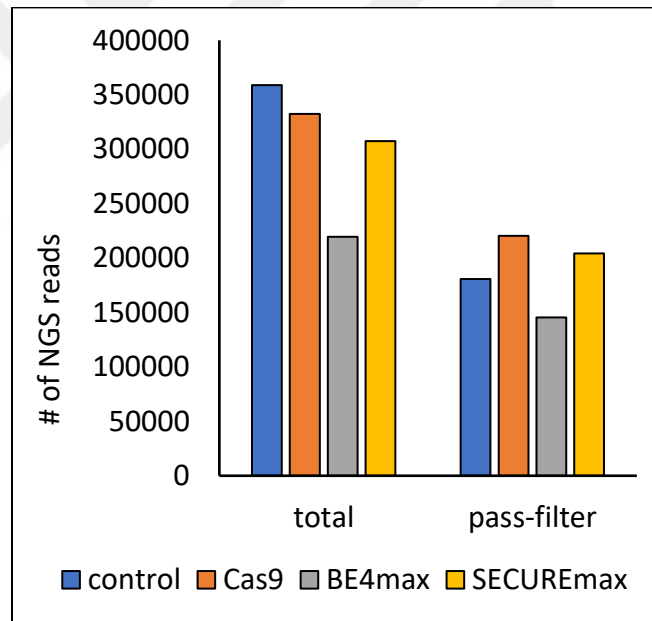


Figure 24. Bar plot showing the average total number of reads and average number of reads passed filter for the treated or untreated samples (n=3). Reads passed filter had spacer-to-first barcode present with up to 2 or less mismatches and both barcodes present with up to 2 or less mismatches.

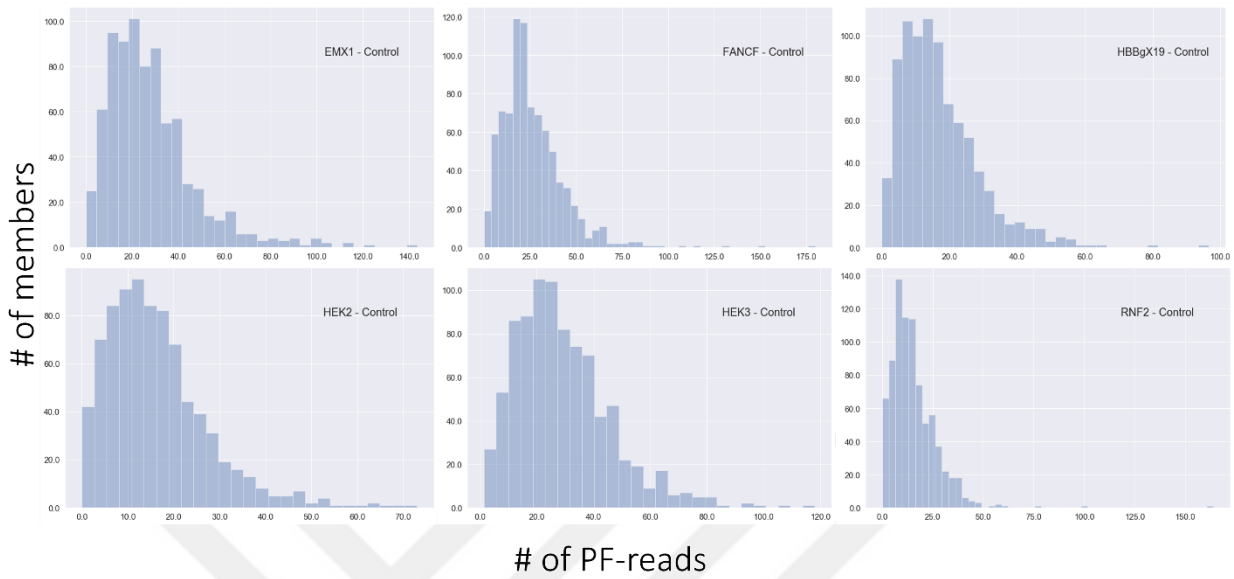


Figure 25. Histograms showing the distribution of the PF-reads across the lentiviral off-target library members in the control sample. X-axis: # of PF-reads, Y-axis: # of library members.

Plotting the library members in EMX1 and FANCF sub-libraries as swarm plots showing the average percentage of library members edited demonstrated the absence of on-target editing in the control sample, comparable on-target editing across CRISPR effectors, but also unexpectedly high noise in the control sample (Fig. 26-27). Moreover, we observed a clear separation of the library members with single mismatches from the other off-target members with more complex mismatches and bulges and combination thereof (Fig. 26-27). Mean EMX1 site on-target editing frequencies were 2.3%, 52.6%, 66.3%, and 55.3% for the control, Cas9, BE4max, and SECUREmax treated samples, respectively (Fig. 26). A library member within the EMX1 sub-library with a combination of mismatches and bulges of 5 nucleotides unexpectedly displayed high editing rates (97.3%, 78.6%, 92%, 90% in control, Cas9, BE4max, SECUREmax, respectively) across all samples, in the absence of undersampling (supported by 17, 13, and 14 NGS reads in the control biological replicates) (Fig. 26).

Even in the presence of high background masking the real signal, we plotted the specificity/tolerance heat-map for the off-by-1 library members (**Fig. 28**). This analysis recapitulated the NGG PAM preference of SpCas9 used across the editors. 2D-scatter plot analysis of the library members as whole treated with either SECUREmax or BE4max demonstrated a general tendency towards off-target library members getting edited at higher frequencies with BE4max and less with SECUREmax (**Fig. 29**). In-depth analysis of the editing outcomes of a single HEK3 off-target library member with a single mismatch revealed that BE4max installed C-to-T edits at high frequencies (86.7% for cytosine at position 5) in the editing window, whereas SECUREmax did not induce any off-target edits (**Fig. 30**).

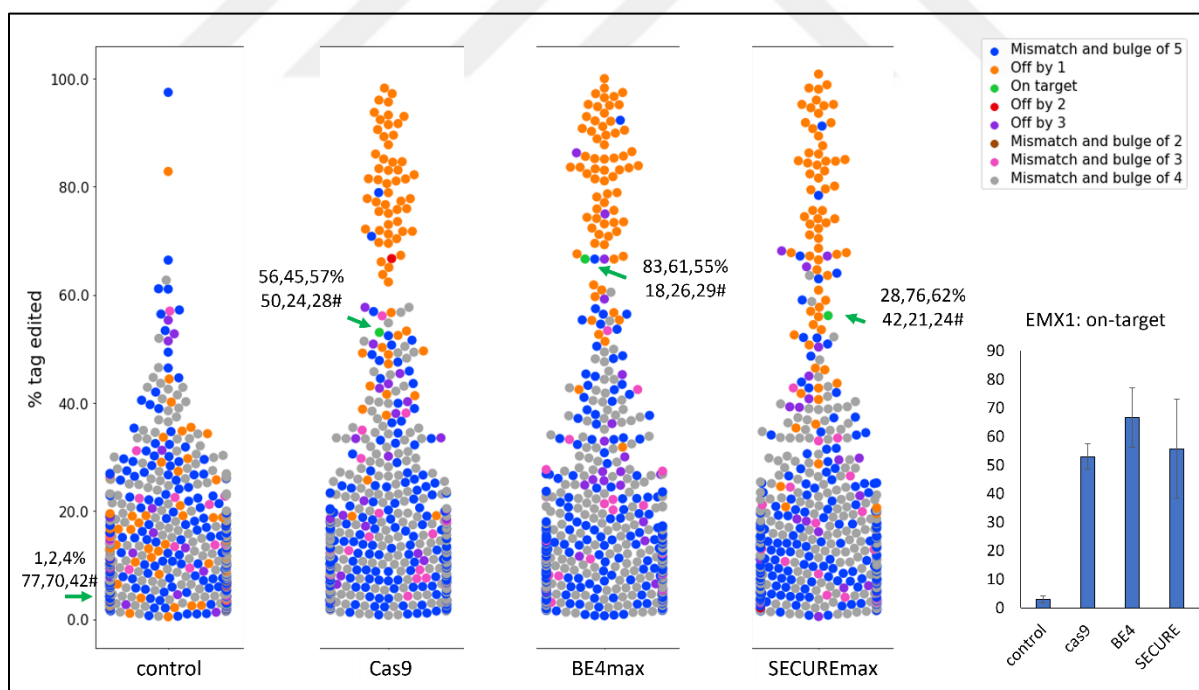


Figure 26. Swarm plots showing mean editing rates across EMX1 sub-library members in control, Cas9, BE4max, and SECUREmax treated samples. On-target member is pointed with a green arrow across all samples, along with the number of NGS reads supporting the editing rate across three replicates. Bar plot at the bottom right shows the mean EMX1 on-target editing rates across samples. (n=3 biological replicates)

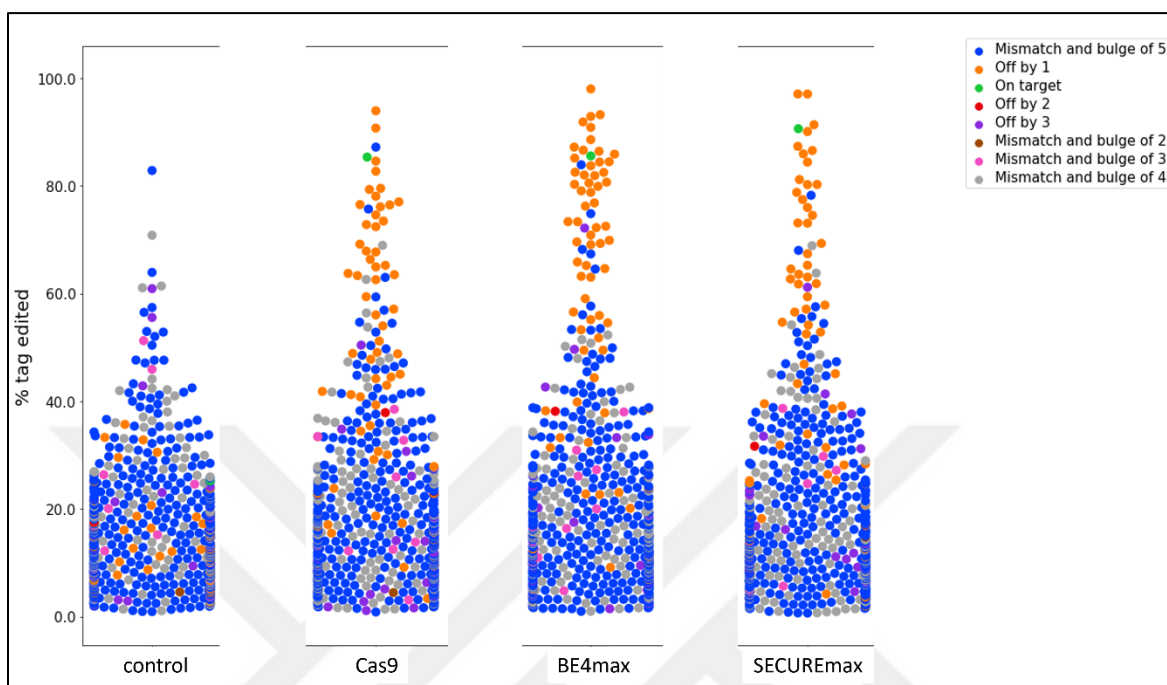


Figure 27. Swarm plots showing mean editing rates across FANCF sub-library members in control, Cas9, BE4max, and SECUREmax treated samples. (n=3 biological replicates)

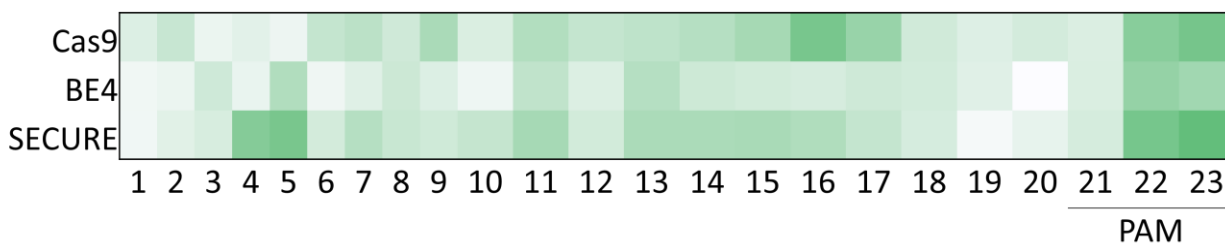


Figure 28. Average mismatch tolerance/specificity heat-map quantified for the EMX1 sub-library in the control, Cas9, BE4max, and SECUREmax samples. Position 21-23 marks the location of PAM sequence. Darker green: less tolerant to mismatches, lighter green: more tolerant to mismatches. (n=3 biological replicates)

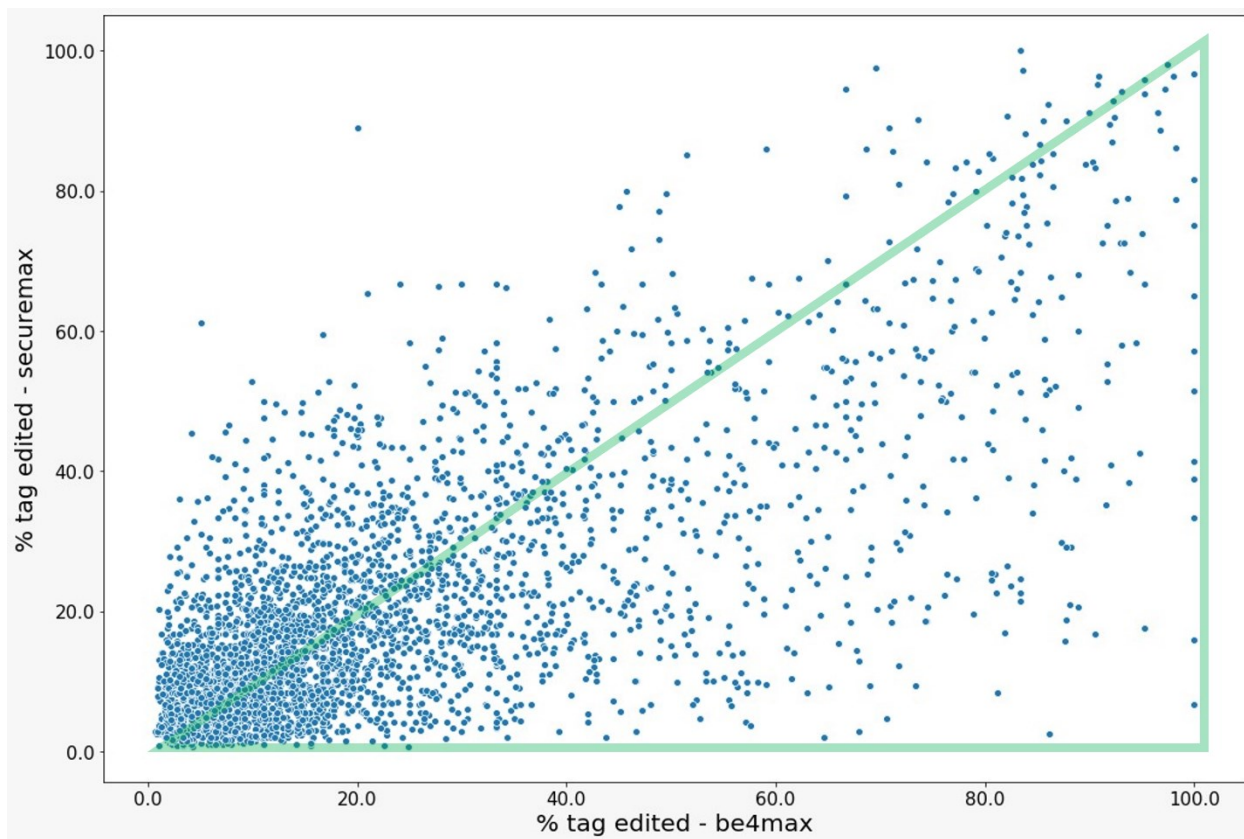


Figure 29. 2D scatter plot displaying mean editing rates for all off-target library members in samples treated with SECUREmax and BE4max (n=3 biological replicates).

into the middle across all library members could maximize the PCR chimera formation between different library members through large homology/complementary between fragments (Qiu et al., 2001). Therefore, we decided to change our library design and cloning strategy with the goal of reducing the high background across samples.

We decided to test an adaptation of Vidigal et al.'s cloning method optimized for complex libraries with cognate two variable sequences separated by a short linker (Shen et al., 2018; Vidigal & Ventura, 2015). By employing this method, we were able to design the library oligos without the long gRNA scaffold, thereby effectively bypassing the excessive homology between the members during the initial PCR amplification. We also decided to alter the modes of delivery and integration from lentiviral to Tol2 transposase (Arbab et al., 2020; Shen et al., 2018).

We cloned the new Tol2-based library successfully and conducted post-cloning NGS QC analysis to check the distribution of the new library members and background editing levels before integrating the library into the genome of the host cells. Data showed that ~20% of the NGS reads were getting filtered out with the most stringent filter (filter #3) (**Fig. 31**). The library as whole and as divided into six sub-libraries displayed skew-ratios as small as ~1.5 with short distribution tails (**Fig. 32**). Side-by-side comparison of the lentiviral off-target library and Tol2 off-target library demonstrated that the background editing rates were substantially reduced in the new Tol2 library (**Fig. 33**). Manual closer inspection of the majority of the outliers were supported by 6 or less NGS reads (**Fig. 33**).

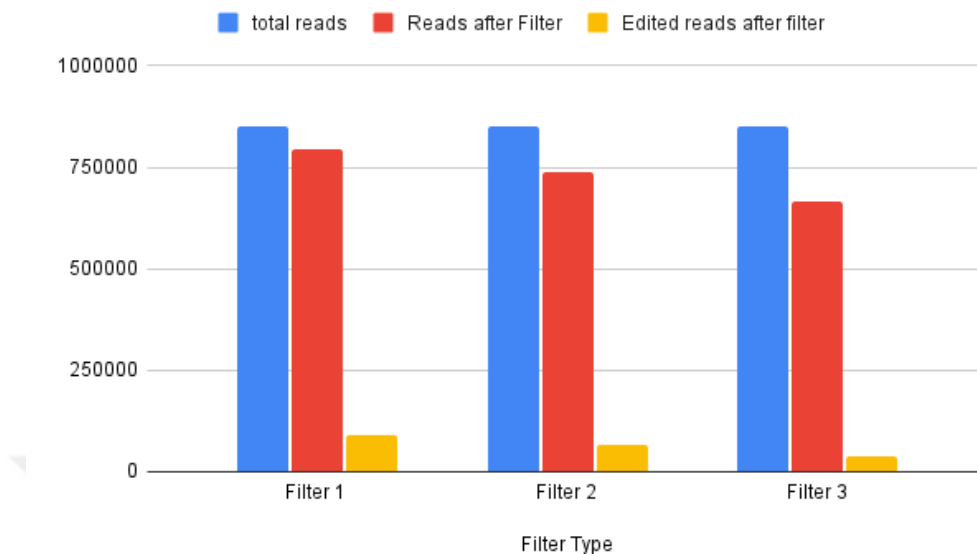


Figure 31. Bar graph showing the number of NGS reads pre-filter, post-filter, and edited for the Tol2-library across three filters. Filter 1 is both barcodes present with up to 2 or less mismatches tolerated. Filter 2 is the stringency of filter 1 and spacer present with up to 2 or less mismatches tolerated. Filter 3 is both barcodes with zero mismatches and spacer-to-first barcode present with up to 1 mismatch tolerated.

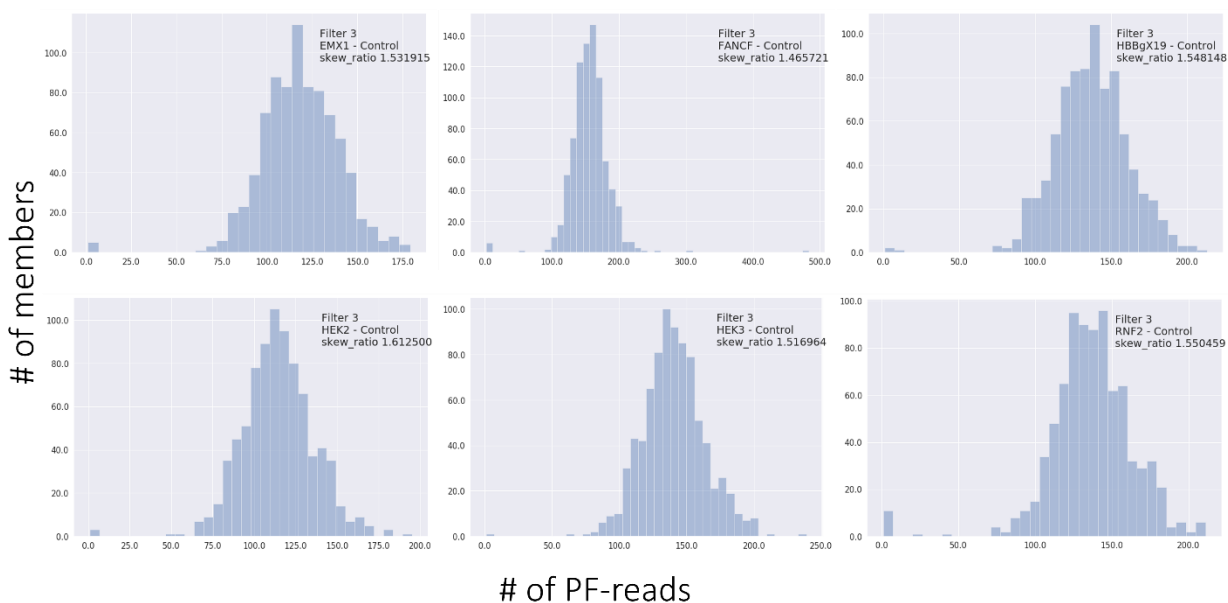


Figure 32. Histograms showing the distribution of the library members across the 6 spacer sub-libraries in the post-cloning control sample of the Tol2-based off-target library. X-axis: # of PF-reads, Y-axis: # of library members.

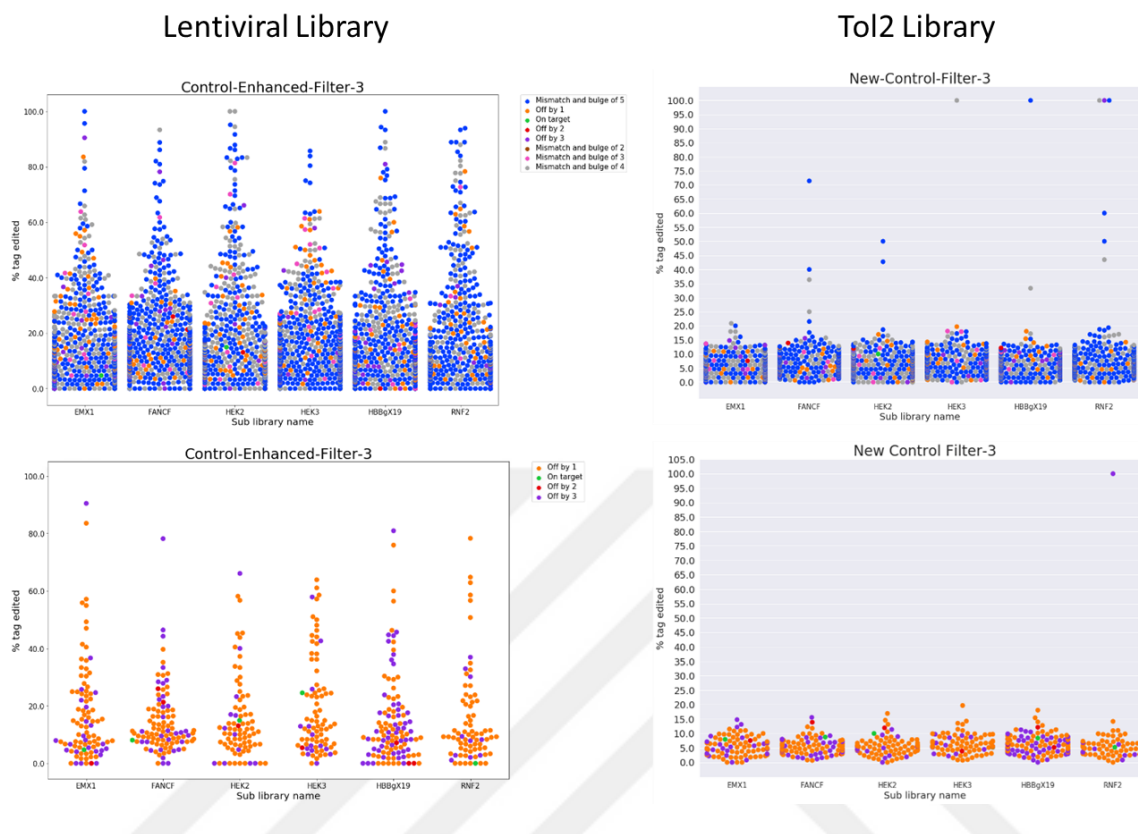


Figure 33. Side-by-side comparison of the swarm plots demonstrating the editing rates between the lentiviral library and Tol2 library across sub-libraries containing all members (top) and containing only on-target, off-by-1, off-by-2, and off-by-3 members (bottom).

Discussion & Future Directions

CRISPR-based base editing technologies are promising tools for therapeutic applications in clinic for reversing the pathogenic SNVs. Guided and unguided DNA off-target editing of the base editing technologies remains to be a major concern along the path towards the full realization of therapeutic potential thereof. There have been numerous cytosine, adenine, and cytosine-to-guanine base editors engineered with altered precision, specificity, activity and targeting scope harboring alternative or engineered CRISPR effectors, deaminases or other DNA repair domains (Gaudelli et al., 2020; Gehrke et al., 2018; Grünewald, Zhou, Iyer, et al., 2019;

Kurt et al., 2020; Rees et al., 2017; Rees & Liu, 2018; Yu et al., 2020). This diverse set of tools renders selecting the most optimal base editing tool and designing specific cognate guides based on editor's specific off-target interaction dynamics hard. Although there have been significant efforts to characterize the genome-wide off-target editing of BEs in vitro and in vivo, systematic testing of the specificity of BEs at thousands of cognate off-target sites in massively multiplexed manner has not been conducted yet (Jin et al., 2019; D. Kim et al., 2017, 2019; Yu et al., 2020).

Although preferentially integrated at the open chromatin regions, lentivirus- and Tol2-based in cellular on-target and off-target systems constitute one of the most realistic biochemical and epigenetic environments for genome editing tools to be assessed, while retaining the synthetic flexibility of custom designs. Our initial lentiviral library design harbored a long ~100 nucleotide SpCas9 guide RNA scaffold constant sequence, and up to ~40% of its NGS reads failed to pass our QC-filter. PCR amplification of homologous DNA sequences with extended constant sequences in between has been reported to produce up to 40% chimeric PCR reads demanding optimized designs and PCR conditions (Omelina et al., 2019; Potapov & Ong, 2017). By altering our cloning strategy and re-designing the library without the guide RNA scaffold sequence, we were able to bypass the high homology between the DNA templates, substantially reduce the number of reads failing to pass the QC-filter, and reduce background editing rates observed in the control samples. A further application of a hard threshold for the number of NGS reads required to support the library members can further improve the background profile of the control samples by removing the undersampled members.

For multiple off-target sub-libraries, we observed that SECUREmax had lower tendencies to induce off-target base editing compared to its parental BE4max (**Fig. 26-27, 29-30**). While the Cas9-dependent off-target base editing activities of SECUREmax was completely

abolished at certain loci at low sequencing depth in our library, it will be of interest to validate these observations as separate experiments by targeted amplicon sequencing.

It will be of interest to employ our Tol2-based library with reduced background in the future to assign fidelity scores to the engineered high-fidelity CBE and ABE variants of SECURE-CBE, SECURE-ABE, HF-BE3, engineered and wildtype versions of the non-rAPOBEC1-containing CBEs (BE4-PpAPOBEC1, BE4-RrA3F, BE4-AmAPOBEC1, BE4-SsAPOBEC3B), along with the extremely hyperactive ABE variants of ABE8 and ABE8e (Gaudelli et al., 2020; Grünewald, Zhou, Garcia, et al., 2019; Grünewald, Zhou, Iyer, et al., 2019; Rees et al., 2017; Richter et al., 2020; Yu et al., 2020). Moreover, such uniform and multiplexed data from a single platform hold potential to train machine learning algorithms to dissect out the features governing the editor-specific off-target interaction dynamics, paving path towards interactive apps for assisted guide-design for striking a balance between specificity and activity for a given site and application of interest. We believe that our work, if finished in the future, will help the community to navigate through the diverse set of base editing technologies and decide on the most specific and optimal one to employ for a given application and scenario, and help train machine-learning algorithms for designing custom tailored specific guides per editor while retaining the on-target activity levels.

Chapter 5: Discussion & Future Directions

Moving forward, undoubtedly, we will continue living in the CRISPR-era with the unprecedented feasibility that it brought into the scientific literature. Although the designer nucleases before the advent of CRISPR tools have already empowered scientists to dissect out the genomic elements and their cognate phenotypic effects to a certain extent, the accessibility and programmability of the genome editing tools have witnessed an unparalleled level of ease with the advent of CRISPR nucleases, mainly due to the difference that a CRISPR nuclease is a ribonucleoprotein complex consisting of an RNA molecule. While programmability of CRISPR nucleases is contingent on cloning and re-cloning of new RNA molecules, of the predecessors were contingent on engineering and re-engineering of new protein architectures, substantially rendering the platforms inaccessible for scientific labs without the niche expertise (Adli, 2018).

While there have been, and there are number of clinical trials and tremendous excitement from the preliminary outcomes thereof for the use of designer nucleases in the clinical setting, later recent work have unequivocally showed that CRISPR nuclease SpCas9 harbors detrimental potential to induce unwanted on-target activities such as chromothripsis, megabase-scale deletions, translocations, inversions and complex rearrangements that can have serious adverse consequences for patients (Frangoul et al., 2021; Leibowitz et al., 2021). Although CRISPR nucleases will undoubtedly remain to be the main genome editing tool in the R&D settings for robustly knocking out genes by introducing frameshifts or disrupting noncoding genomic elements, slowly but steadily it became clearer that the future of therapeutic application of genome editing tools will lie in the hands of next-generation CRISPR-based tools with substantially increased precision and lowered byproducts (mainly the indels by double-stranded breaks). Even the demand for and use of CRISPR nucleases for deleting or knocking in large

segments of genomic DNA by introducing exogenous template DNA via harnessing the HDR pathway seems to be destined to ultimately diminish and make way for the recently described twin prime-editing and site-specific recombinases that can efficiently and precisely delete up to 10 kb and install large genomic insertions and inversions up to 40 kb (Anzalone et al., 2021; Choi et al., 2021; Ioannidi et al., 2021; Jiang et al., 2021).

Moving forward, the advent of CRISPR-based base editors successfully addressed the heart of the problem by virtually bypassing the introduction of double-stranded breaks. While cytosine and adenine base editors have addressed the need to install transition mutations at high frequencies in programmable manner, the work outlined in this thesis have recently addressed the need to install transversion mutations at high frequencies in a context-dependent manner by engineering cytosine-to-guanine base editor architectures. As has been always the case, the field will unquestionably continue to put efforts towards engineering novel CRISPR-based base editors with improved versatility with respect to the types of edits that can be induced, on-target precision, lowered degree of off-target editing, altered cell type-dependent activity, altered or broadened targeting scope; all of which are paramount to the successful implementation thereof in clinical and R&D settings.

On the other hand, as has been the case with the designer nucleases and even more complex than them, CRISPR-based base editors are also prone to induce off-target mutagenesis. CRISPR-based base editors not only possess the inherited specificity features of the parental CRISPR effector it is built upon, but also have its distinct modalities that can affect the specificity profiles thereof. Specifically, on top of the CRISPR-guided off-target editing due to the protein-DNA and RNA-DNA interaction kinetics, there are also potential unwanted CRISPR-independent spurious off-target activities at the genomic and transcriptomic levels that can be

induced by the deaminase or other DNA repair partners in the construct (Doman et al., 2020; Grünewald, Zhou, Garcia, et al., 2019; Grünewald, Zhou, Iyer, et al., 2019; Jin et al., 2019; Yu et al., 2020; E. Zuo et al., 2019). While there have been tremendous protein engineering and mining efforts at the frontline of increasing fidelity of base editors, it seems impractical to entirely bypass the spurious off-target activities induced by the overexpressed DNA repair components of base editing technologies especially in settings where it may be necessary to fully avoid off-target editing (Grünewald, Zhou, Garcia, et al., 2019; Grünewald, Zhou, Iyer, et al., 2019; Yu et al., 2020). As the recent work has corroborated, it may be necessary to have a two-layered system via co-expression of a secondary regulatory/inhibitory protein to bring genome-wide and transcriptome-wide off-target editing activities of base editing technologies to background levels (L. Wang et al., 2021).

Currently, the prime editing technology holds tremendous potential to address both the precision and specificity concerns of the CRISPR-based genome editing tools. It is in fact the most versatile and flexible genome editing tool that has ever been engineered in the history of genome editing. It is also important to note that prime editors could also be the most accurate CRISPR-based genome editing tool, if not the most robust/active in a context-dependent manner, due to the fact that a potential off-target locus would have to pass through three checks (base pairing between spacer and protospacer, between PBS and 3' nontarget nicked and freed ssDNA, between RTT-templated newly polymerized 3' flap and endogenous sequence downstream of the nick on the target strand) as opposed to a single check (base pairing between spacer and protospacer) for a conventional CRISPR-based genome editing tool or a base editor. This is indeed confirmed by multiple reports in multiple settings demonstrating the extremely low off-

target activity of the prime editing tools (Jin et al., 2021; D. D. Y. Kim et al., 2020; Schene et al., 2020).

Nonetheless, in future, the potential genome-wide and transcriptome-wide off-target activities of the prime editing technology needs to be vetted carefully by alternative approaches other than the widely employed short-read DNA sequencing. Although the RNA-seq experiments demonstrated moderately minimal perturbations at the transcriptome level upon prime editor overexpression, it is of utmost importance to design critical experiments to test potential negative effects of the exogenous overexpression of a viral hyperactive RT found in the prime editor architecture, especially given the fact that more than half of human genome consists of transposable “jumping” elements some of which rely on an intermediate RT-mediated reverse transcription to transpose itself into new loci (Anzalone et al., 2019; Kramerov & Vassetzky, 2011). Particularly, endogenous RT activity of retrotransposon long interspersed element 1 (LINE1) in human genome can lead to retrogene production, that is reverse transcription and genomic integration of endogenous mRNAs, along with the integration of itself, potentially leading to a list of human diseases (Elbarbary et al., 2016). Moreover, although remains controversial, a recent study by Rudolf Jaenisch at Whitehead Institute/MIT demonstrated that even SARS-COV-2 genomic elements can be integrated stably into human genome by exogenous LINE1 RT overexpression (L. Zhang et al., 2021). Moving forward, all premises combined strongly warrants to develop sensitive ways to test how the genomic pattern of LINE/SINE/Alu retrotransposons change upon prime editor treatment, especially of Alu elements that are enriched at genic regions (S. Kim et al., 2016).

The transversion base editors described in this thesis constitute a foundation for systems to install cytosine-to-guanine and guanine-to-thymine in a programmable manner in human cells.

CGBEs leverage the endogenous repair machinery and a modified cytidine deaminase to install cytosine-to-guanine edits at high frequency in a context-dependent manner. The additional CGBE literature published at the same time or after our work outlined in this thesis has collectively shown that uracil-induced UNG-driven base excision repair is necessary for the observed preferential cytosine-to-guanine installation (L. Chen et al., 2021; Koblan, Arbab, et al., 2021; D. Zhao et al., 2020). Although we do not yet know the exact downstream mechanism(s) -though discussed extensively in the Discussion of Chapter 2- leading to the preferential cytosine-to-guanine mutagenesis, or the exact reasons behind the increased editing activities with mutant cytidine deaminases, cell-type dependence of editing outcomes, preference for the cytosine at or around position six, and AT-rich sequence context for higher yield and purity, the recent CGBE works by other groups have recapitulated and corroborated our findings. CGBEs outlined in this thesis out of 18 different CGBE constructs described by others have been recently shown to have the capacity to outperform the others at two sites, to perform comparable to others at six sites, and perform worse at three sites in terms of product yield and purity, demonstrating that it may be possible to employ different CGBEs in different contexts for the optimal outcomes (**Fig. 4**). Employing the recently described pairwise CRISPRi library targeting 476 DNA repair genes individually while linking the cytosine-to-guanine editing outcomes to the individual DNA repair gene that is knocked down, the answers to some of the research questions posed above may be answered and exact fusion partners to engineer CGBE2 from CGBE1 could be identified (Husmann et al., 2021). Applying CGBE-Hive experimental workflow directly on our CGBE constructs should also reveal the complex sequence features and dynamics that regulate cytosine-to-guanine editing outcomes (Koblan, Arbab, et al., 2021).

The ability to target guanines and mutate them into thymines in a programmable manner via base editing technology will also carry importance moving forward. The GBE constructs outlined and preliminarily tested in this thesis can lead to engineering of efficient GBE architectures. Critically, engineering novel enzymes with ability to recognize guanines directly to oxidate them into mutagenic 8-oxoguanine will be pivotal moving forward towards efficient guanine-to-thymine transversion mutagenesis, similar to the work of Gaudelli et al. engineering a novel adenosine deaminase with altered substrate preference (Gaudelli et al., 2017). Specifically, the work outlined in this thesis may aid in rational protein design studies to alter the key amino acid residues in the active site of alpha-ketoglutarate-dependent xanthine dioxygenase to render guanine, instead of its cognate deamination product xanthine, as a substrate leading to the direct oxidation thereof into 8-oxoguanine. To this end, Gln99-to-Glu99 mutation in XanA may be a feasible approach to re-establish hydrogen bonding with the hypothetical guanosine substrate (Fig. 34).

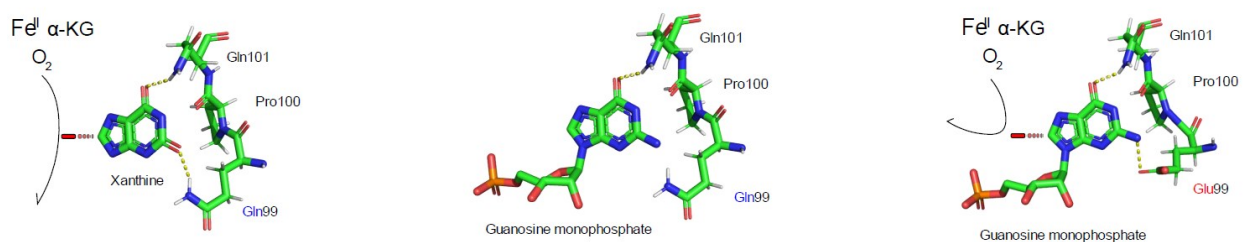
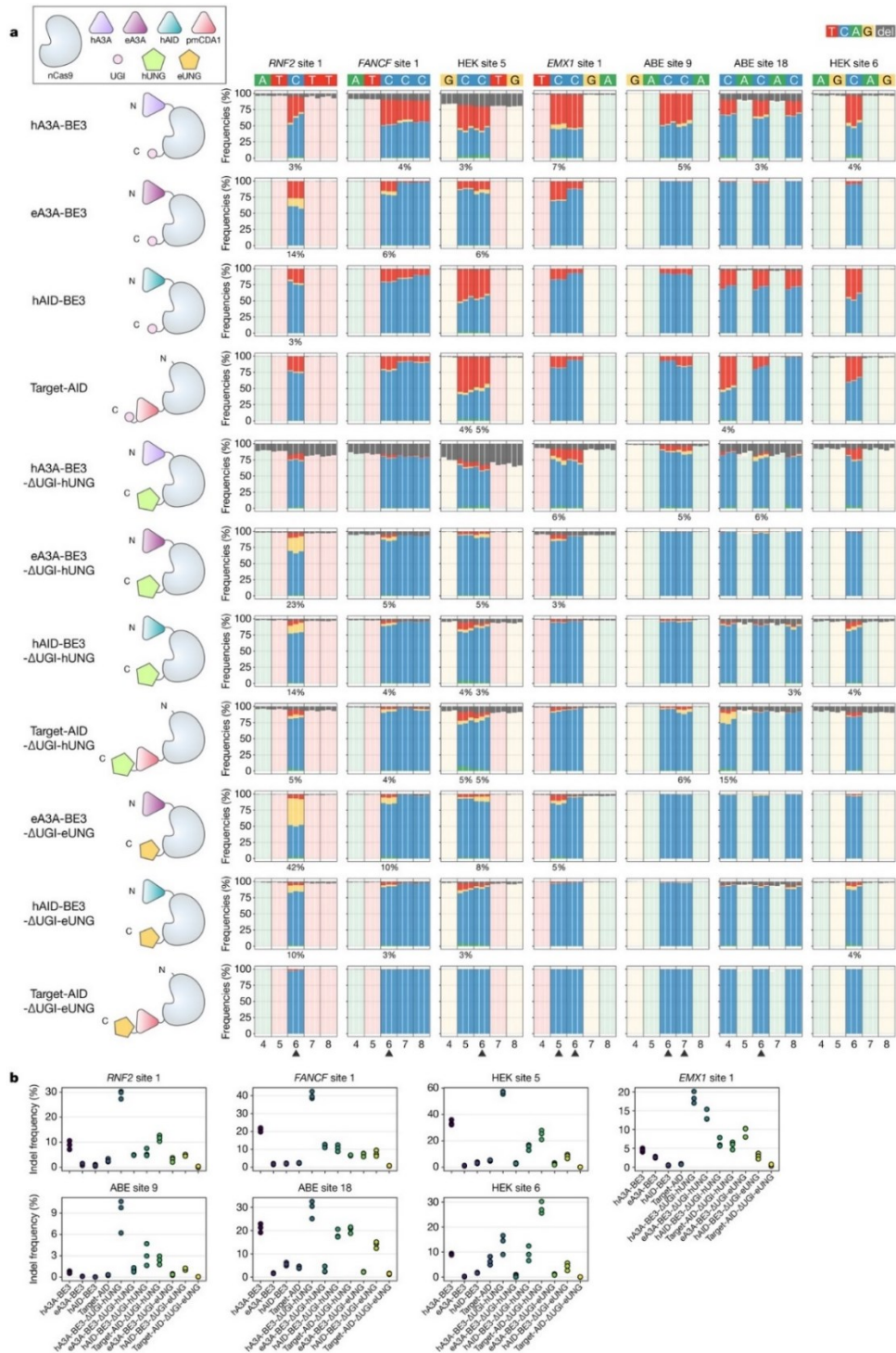


Figure 34. Schematic demonstrating the key amino acids in the active site of XanA that takes place in stabilization of xanthine substrate during its alpha-ketoglutarate and iron-dependent oxidation into uric acid (left most). Replacing Gln99 that serves as a hydrogen donor via its amine moiety with Glu99 could be able to re-establish hydrogen bonding with the C2 amine moiety in guanosine.

The work outlined in this thesis will enable the researchers to install transversion mutagenesis and codon/amino acid alterations that were not possible before with the previous base editing tools, thereby expanding the therapeutic and research applications of base editing toolbox. As has been the case with the previous genome editing tools, we believe that our work sets the stage to spur other researchers to engineer more efficient transversion base editor tools. It will be also utmost critical to thoroughly assess the genome-wide and transcriptome-wide off-target base editing profiles of the transversion base editors described in our work, with the ultimate goal of characterizing their potential in their transition towards use in therapeutic genome editing.

Supplementary Figures

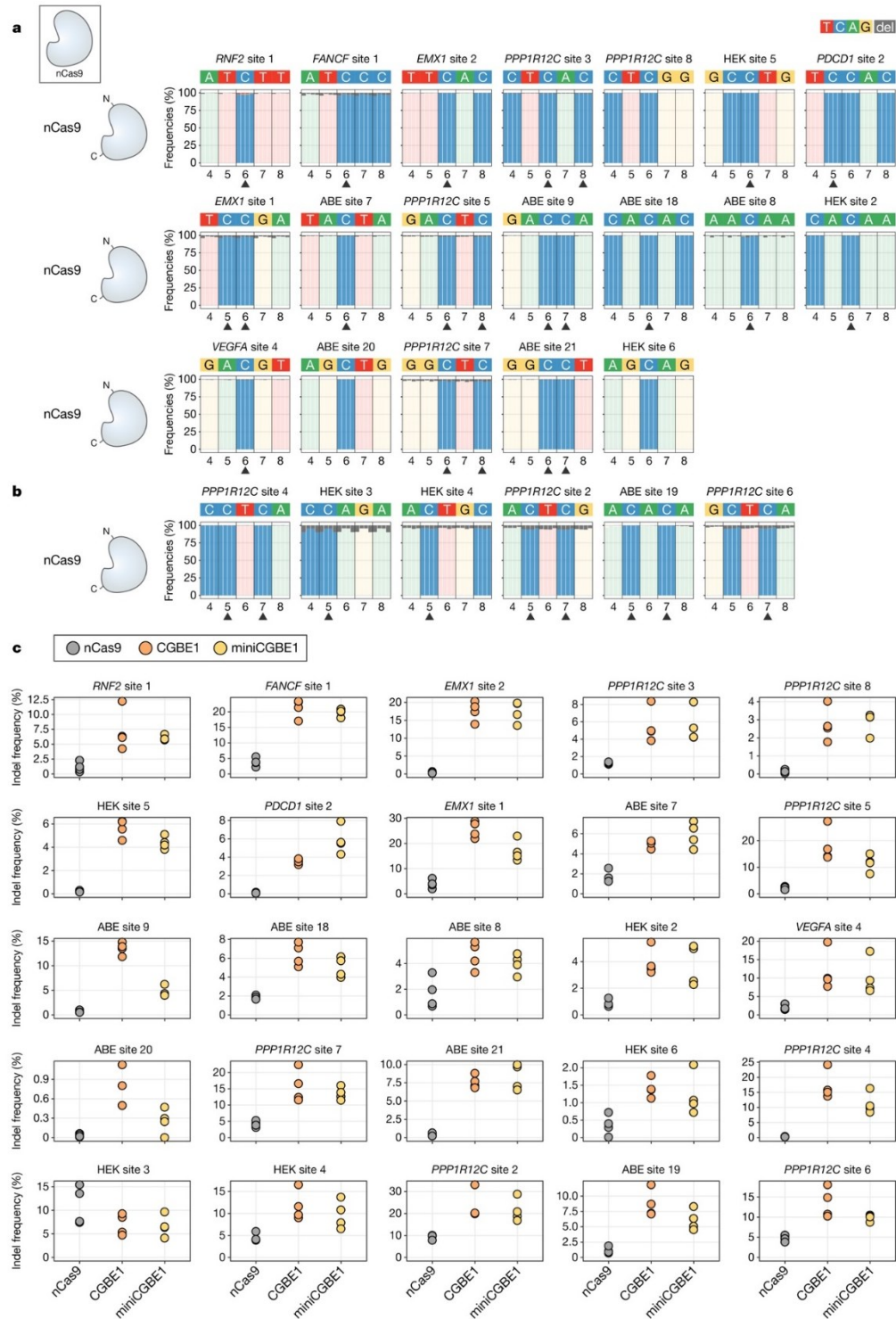


Supplementary Figure 1. On-target activities of non-APOBEC1 CBE variants tested for C-to-G editing in HEK293T cells. a, Bar plots showing the on-target DNA base editing frequencies induced by

Supplementary Figure 1 (Continued)

non-APOBEC1 CBEs and their variants with h/eUNG and seven gRNAs in HEK293T cells. N and C indicate amino-terminal and carboxy-terminal ends, respectively, of the various base editors. Gray overlay bars at top represent deletions at each nucleotide. Editing frequencies of three independent replicates ($n=3$) at each base are displayed side-by-side. Percentage values below specific cytosine bases indicate the average C-to-G editing observed (values below 3% not reported). Numbering on the bottom indicates position of the base in the protospacer with 1 being the most PAM-distal base. Arrowheads indicate cytosines showing C-to-G edits by CGBE1. **b**, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with non-APOBEC1 CBE variants reported in **a**. Single dots represent individual replicates ($n = 3$ independent replicates).





Supplementary Figure 2. Additional data comparing on-target editing activities of CGBE1 and miniCGBE1 in HEK293T cells. a,b, Bar plots showing the on-target DNA base editing frequencies observed with expression of a nCas9 negative control using 19 gRNAs for sites with a C at position 6 (C6-sites; **a**) and 6 gRNAs for sites with a C at position 4, 5, 7, or 8 (non-C6-

Supplementary Figure 2 (Continued)

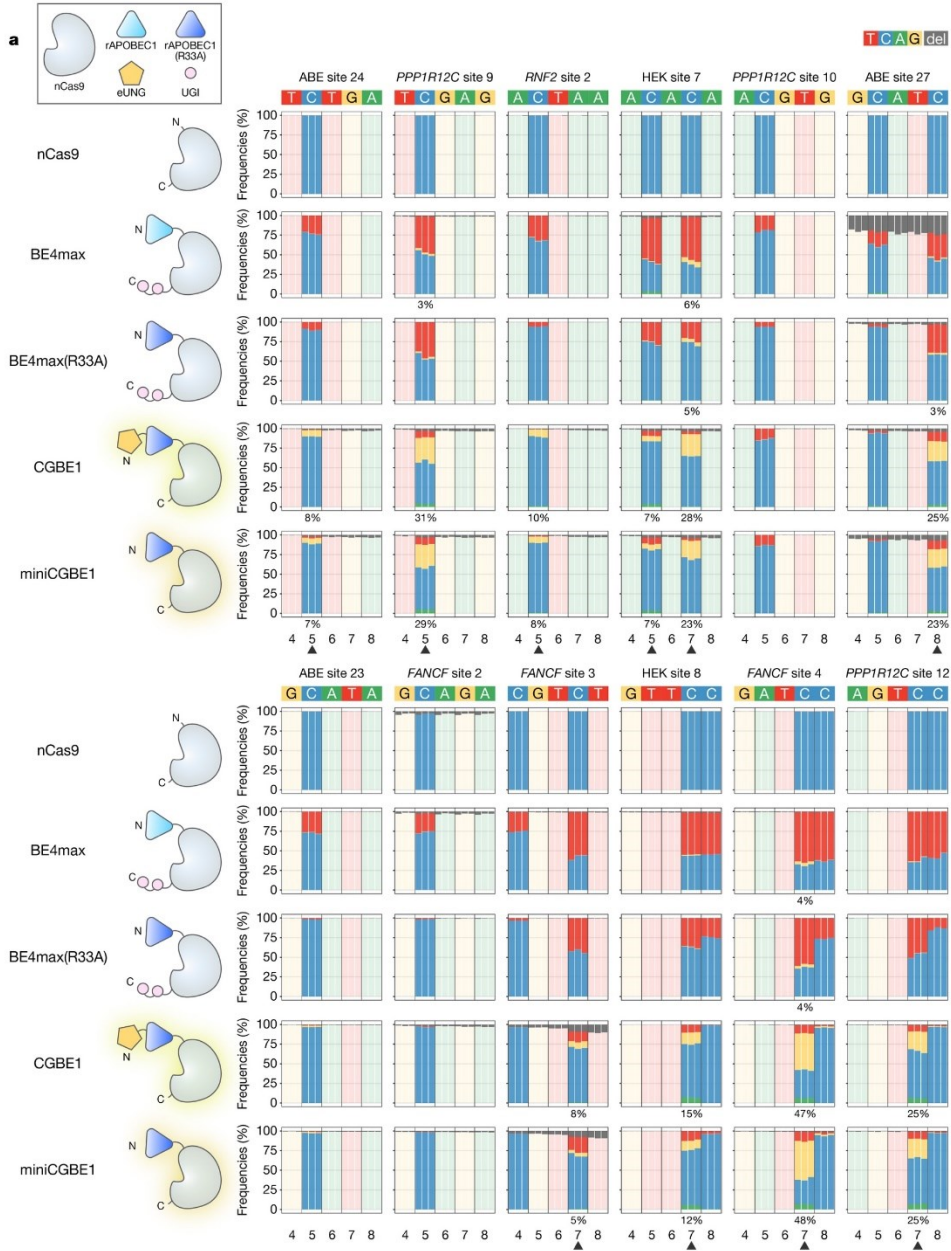
sites; **b**) in HEK293T cells. These are negative control experiments related to the data shown in Fig. 3. Editing frequencies of four independent replicates ($n = 4$) at each base are displayed side-by-side. **c**, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with CGBE1 and miniCGBE1 reported in Fig. 3 and control experiments reported in **a** and **b**. Single dots represent individual replicates ($n = 4$ independent replicates).



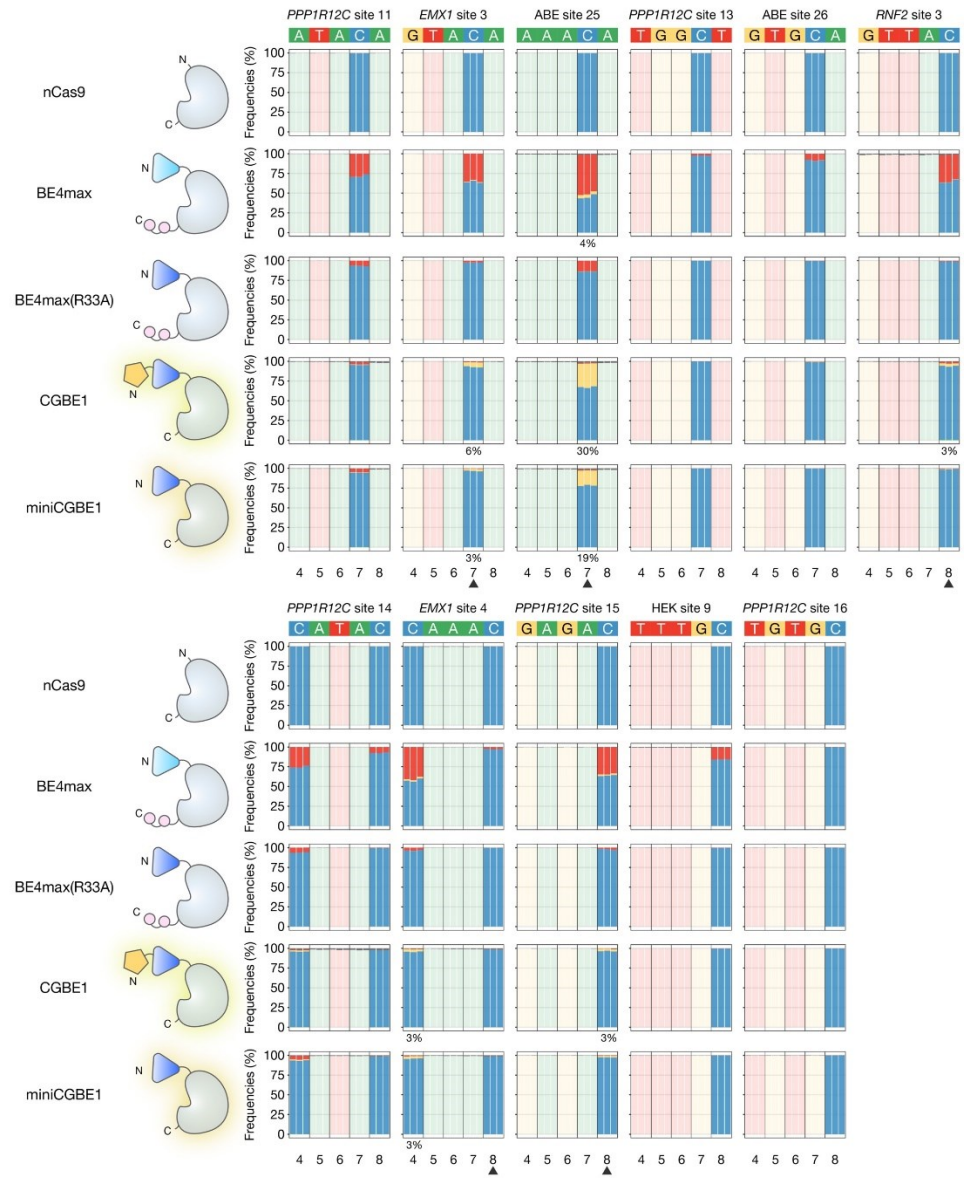


Supplementary Figure 3. Additional comparison of CGBE1 and miniCGBE1 on-target editing activities with 23 non-C6 gRNAs in HEK293T cells. **a**, Bar plots showing the on-target DNA base editing frequencies induced by nCas9 control, BE4max, BE4max(R33A), CGBE1, and miniCGBE1 with 23 gRNAs for sites with a C at position 4, 5, 7, or 8 (non-C6 sites) in HEK293T cells. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side-by-side. **b**, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with BE4max, BE4max(R33A), CGBE1 and miniCGBE1 reported in **a**. Single dots represent individual replicates ($n = 3$ independent replicates).

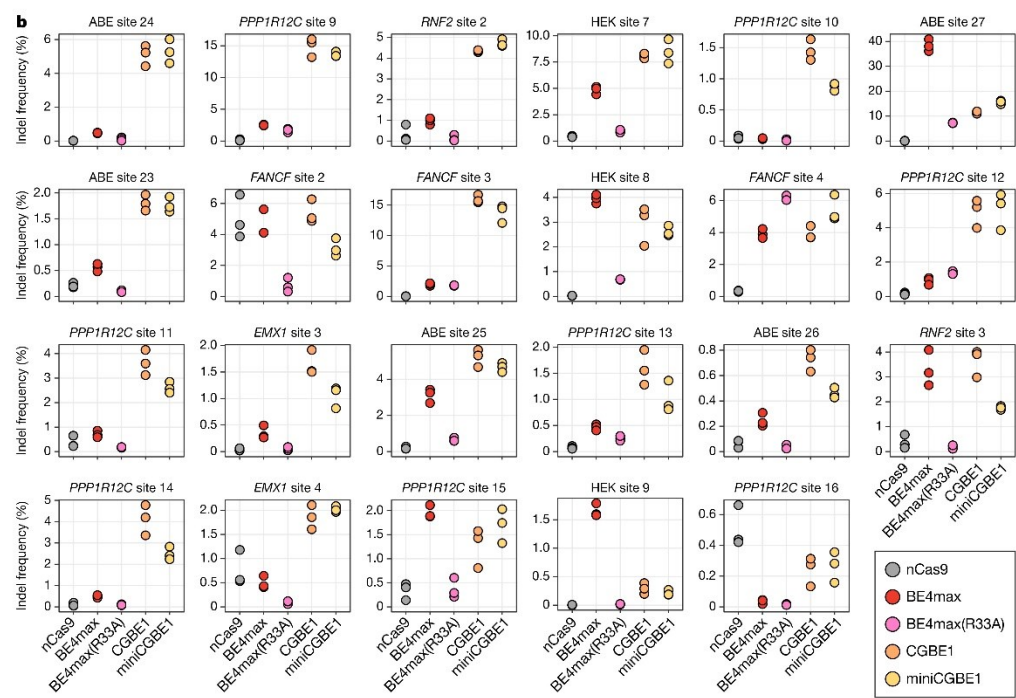
Supplementary Figure 3 (Continued)



Supplementary Figure 3 (Continued)



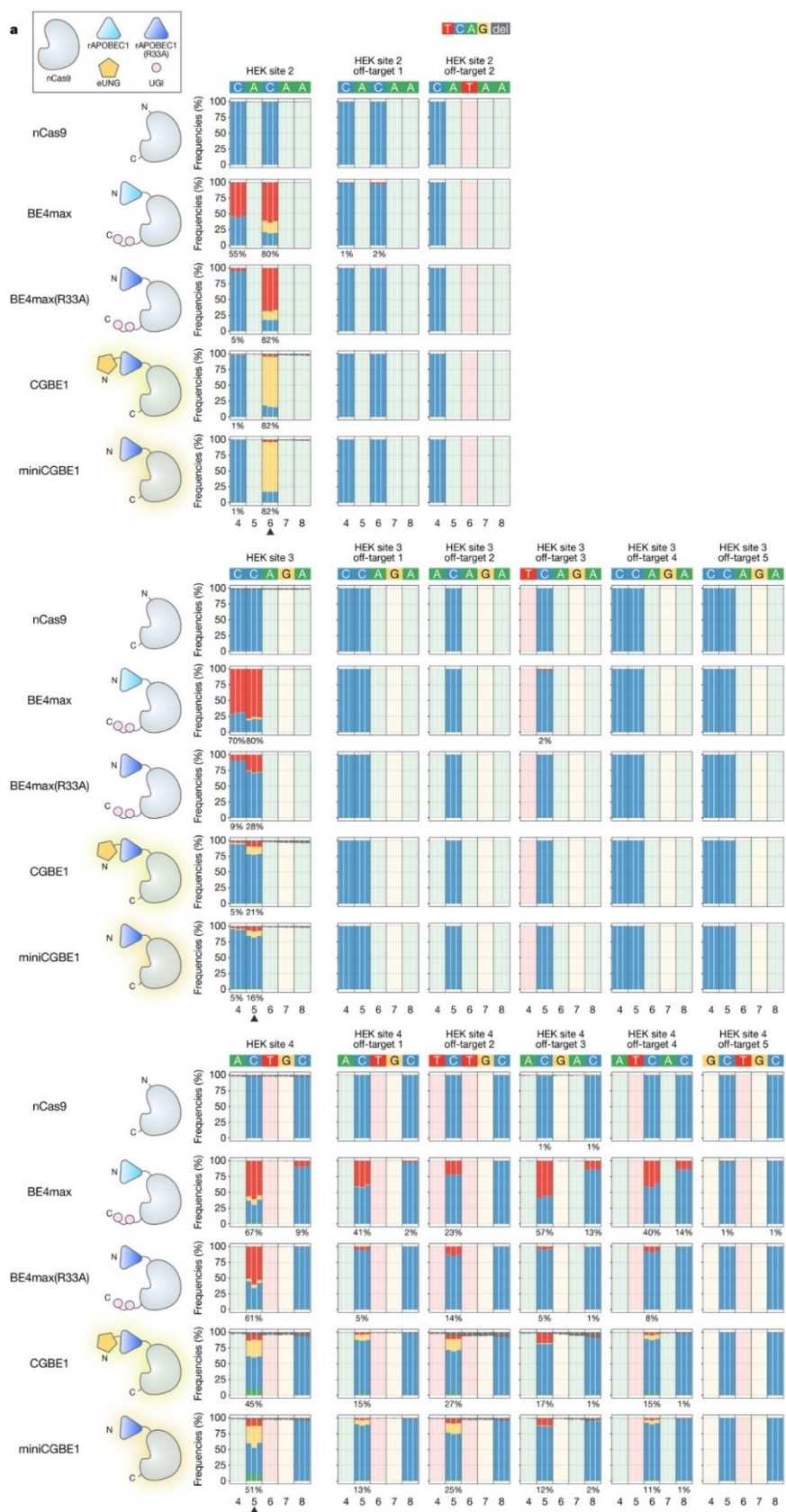
Supplementary Figure 3 (Continued)



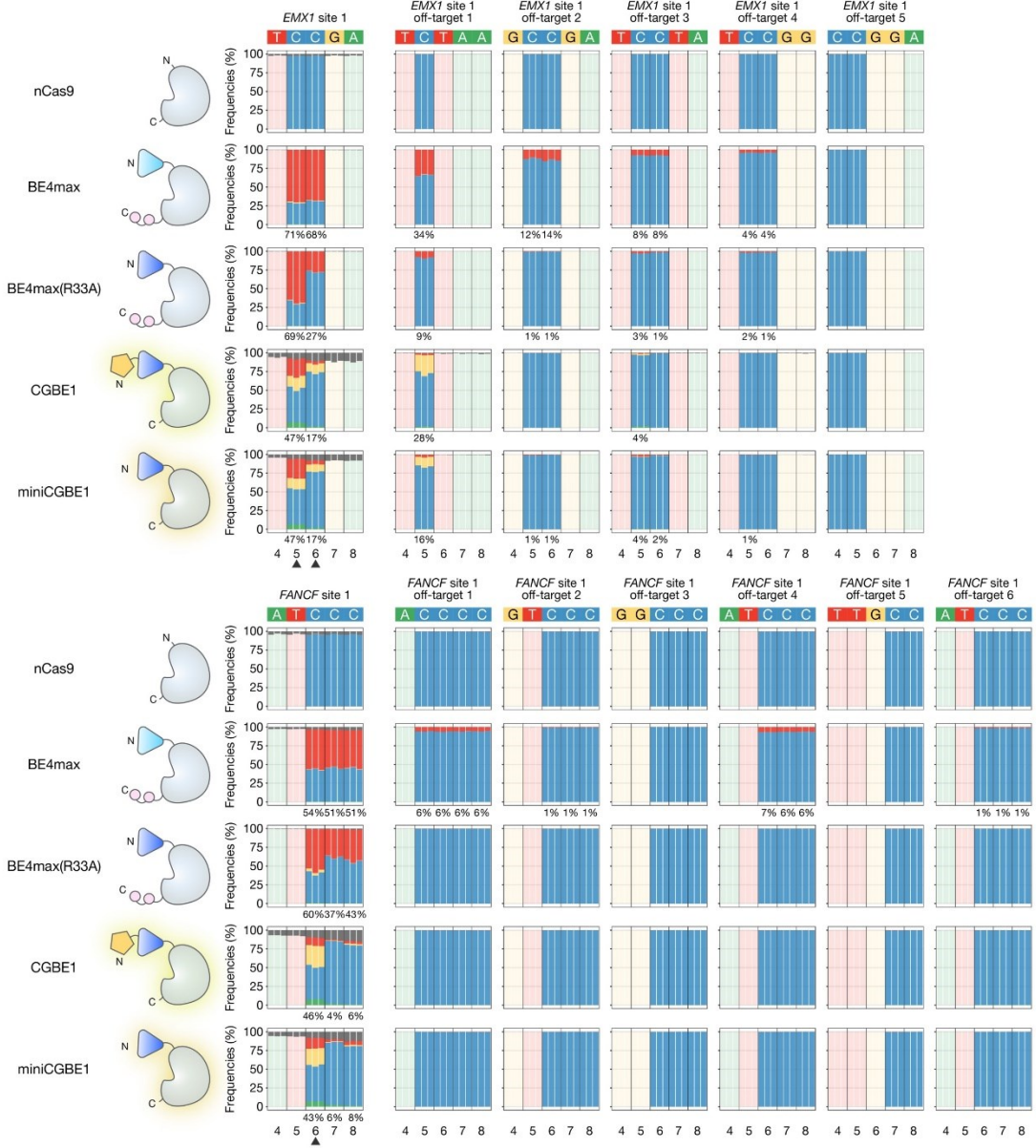


Supplementary Figure 4. Off-target DNA editing activities of CGBE1 and miniCGBE1 in HEK293T cells. **a**, Bar plots showing the off-target DNA base editing frequencies induced by nCas9 control, BE4max, BE4max(R33A), CGBE1, and miniCGBE1 using HEK site 2, HEK site 3, HEK site 4, *EMX1* site 1, and *FANCF* site 1 gRNAs in HEK293T cells. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side-by-side. Percentage values below specific cytosine bases indicate the average C-to-D (D = A/T/G) editing observed (values below 1% not reported). **b**, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with BE4max, BE4max(R33A), CGBE1 and miniCGBE1 reported in **a**. Single dots represent individual replicates ($n = 3$ independent replicates).

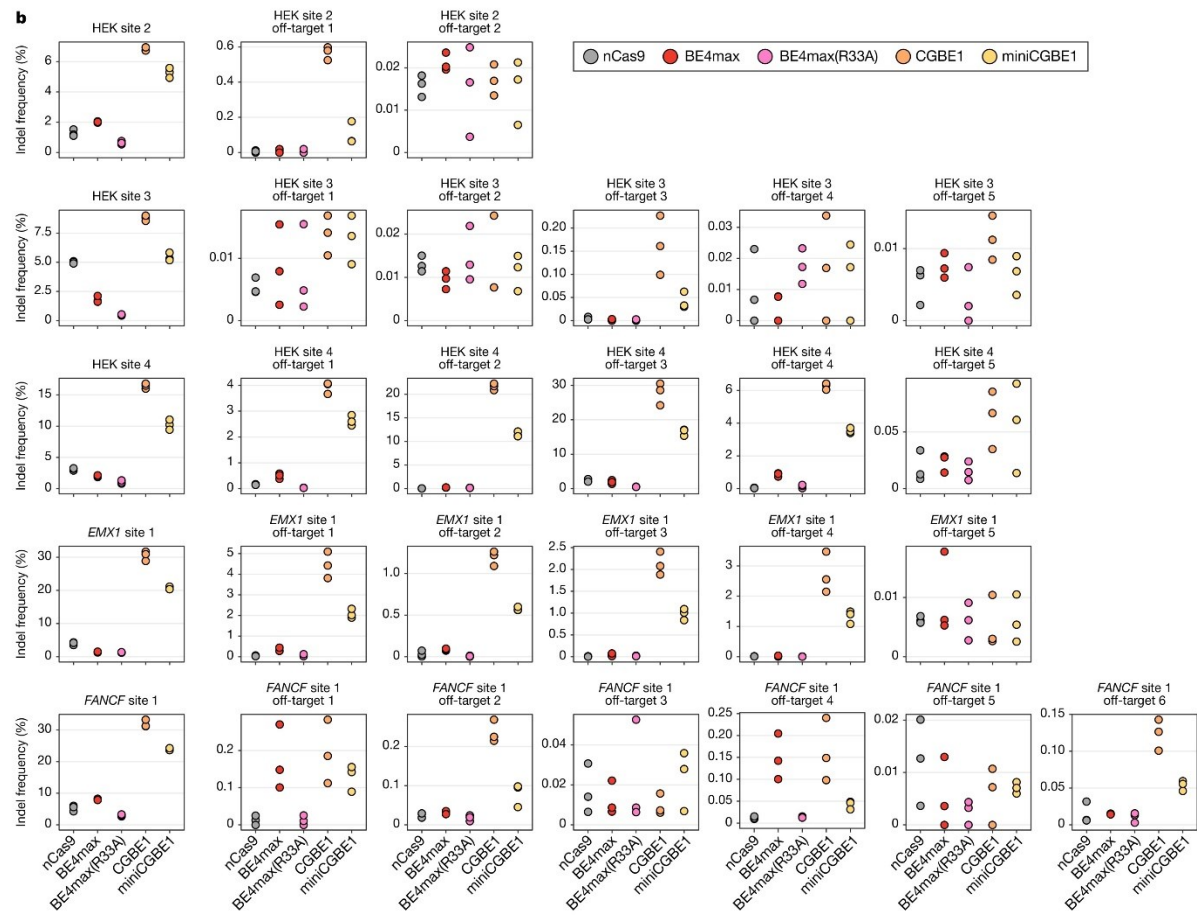
Supplementary Figure 4 (Continued)



Supplementary Figure 4 (Continued)

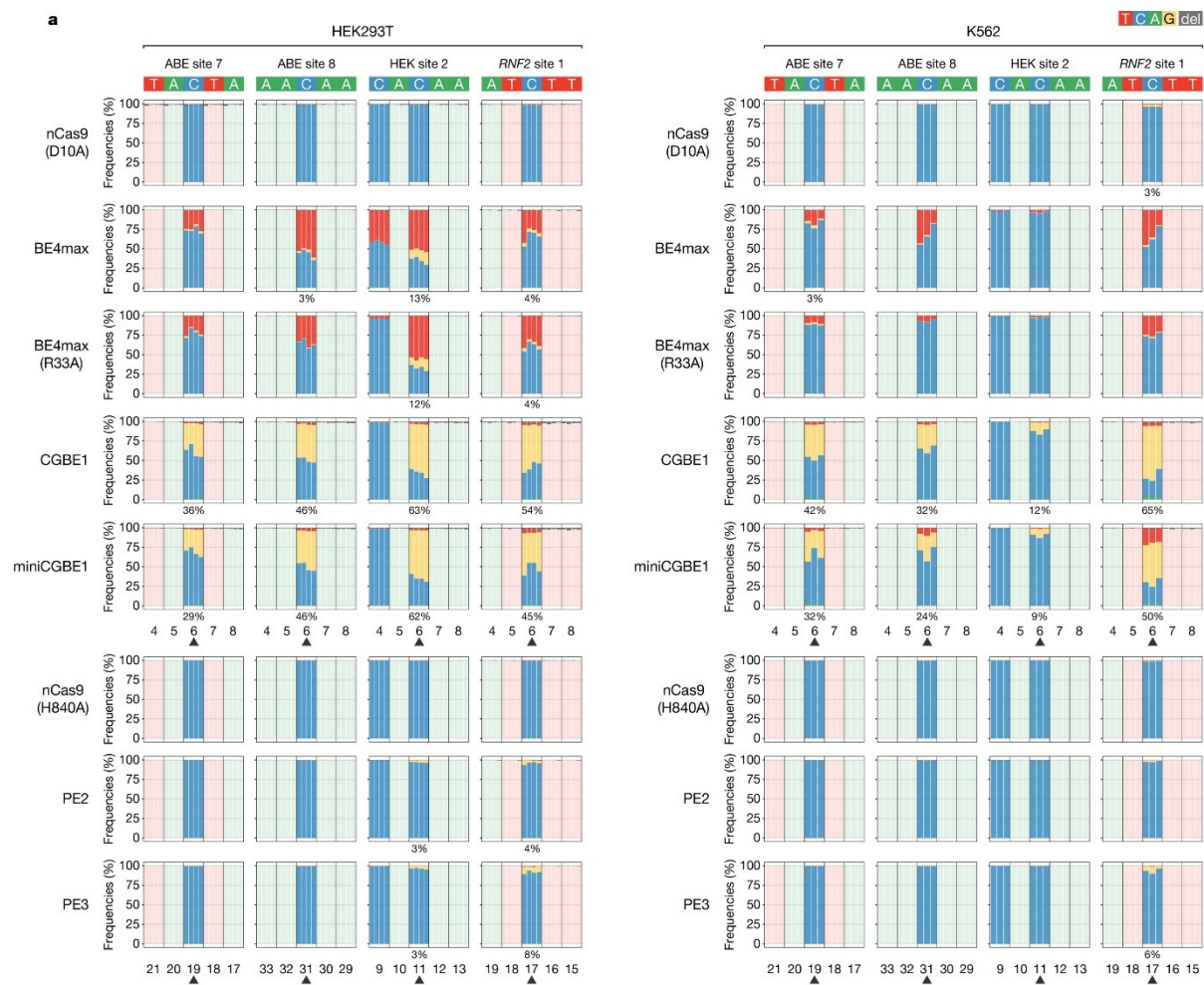


Supplementary Figure 4 (Continued)

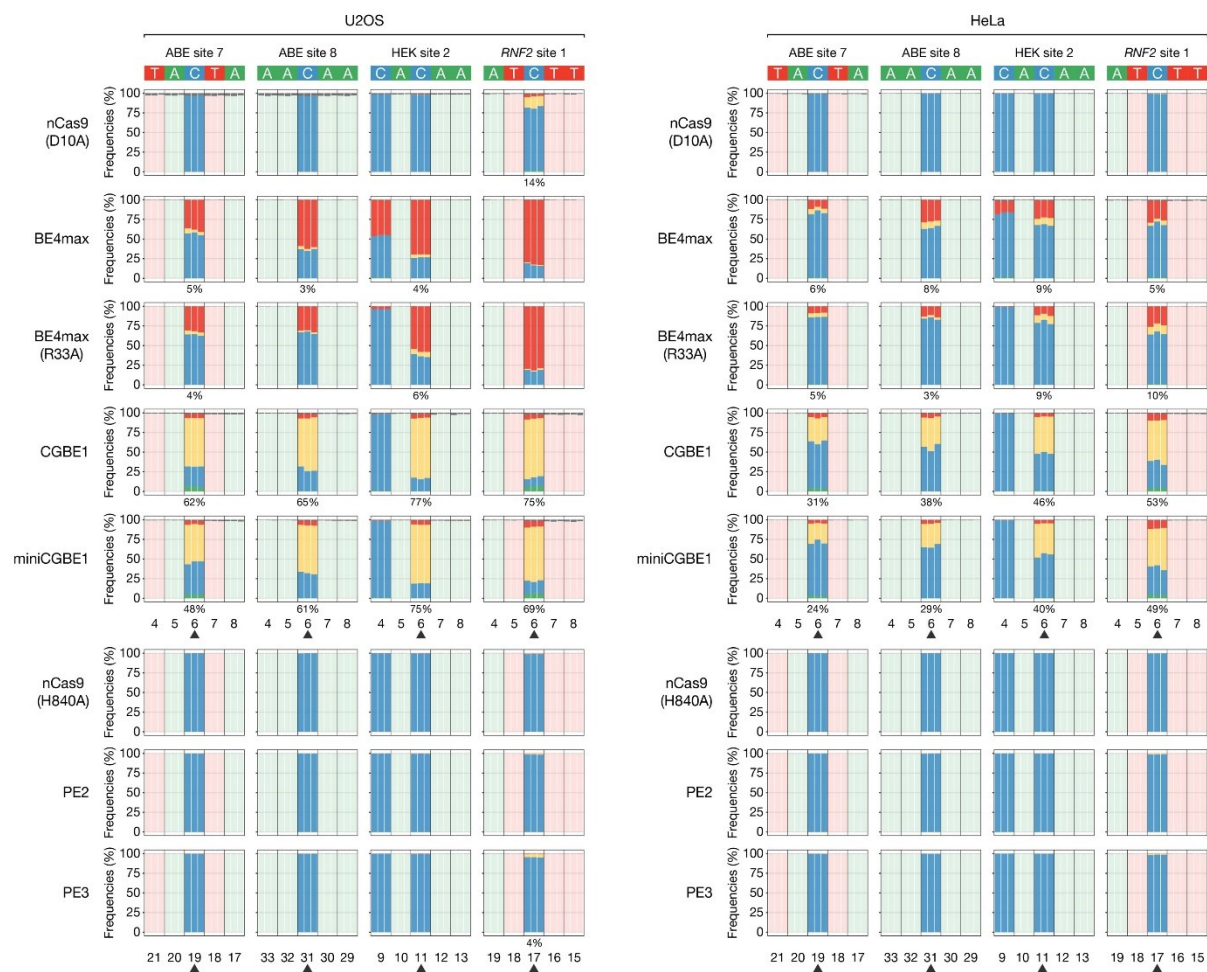


Supplementary Figure 5. Additional comparisons of CGBE1, miniCGBE1, PE2, and PE3 on-target editing activities in HEK293T, K562, U2OS, and HeLa cells. **a**, Bar plots showing the on-target DNA editing frequencies induced by nCas9 controls, CGBE1, miniCGBE1, PE2, and PE3 with four gRNAs (CGBEs), four pegRNAs (PE2), or 4 pegRNA/nicking gRNA combinations (PE3), designed to install a C-to-G substitution at the same cytosine at four genomic loci in four cell lines. Gray overlay bars at top represent deletions at each nucleotide. Editing frequencies of four independent replicates ($n = 4$) for HEK293T cells or three independent replicates ($n = 3$) for K562, U2OS, and HeLa cells at each base are displayed side-by-side. Percentage values below cytosine bases reflect the average C-to-G editing observed (values below 3% not reported). Numbering on the bottom indicates position of the base with 1 being the most PAM-distal base for base editors, or the first nucleotide 3' of the pegRNA/Cas9-induced nick for prime editors. Arrowheads indicate cytosines with C-to-G edits by CGBEs. Data shown are from the same experiment as shown in Extended Data Fig. 7e. **b**, Dot plots representing percentage of alleles that contain an insertion or deletion across the entire protospacer from experiments with CGBE1, miniCGBE1, PE2, and PE3 reported in Extended Data Fig. 7e and **a**. Single dots represent individual replicates ($n = 4$ for HEK293T and $n = 3$ for K562, U2OS, and HeLa cells).

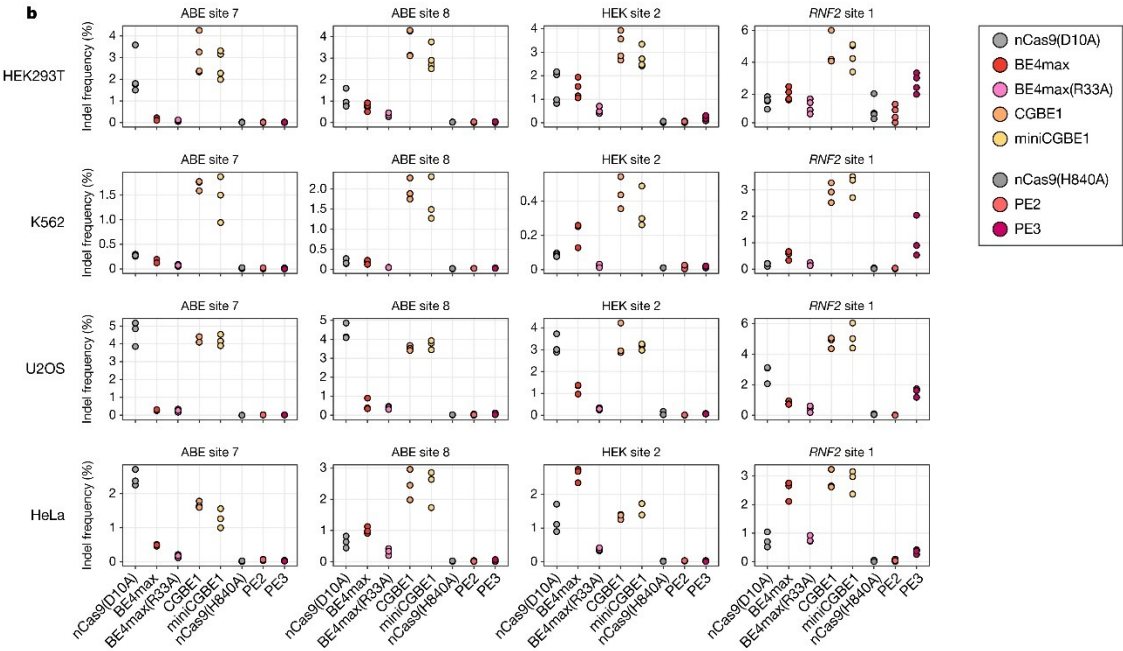
Supplementary Figure 5 (Continued)



Supplementary Figure 5 (Continued)



Supplementary Figure 5 (Continued)



Supplementary Tables

Supplementary Table 1: List of all base and prime editor constructs and controls used in this study.

Supplementary Table 2: List of pathogenic/likely-pathogenic disease-causing cytosine SNPs at 6th position with NG PAMs that can be targeted and reversed using CGBEs

Supplementary Table 3: Sequences of gRNAs and pegRNAs, primers, and amplicons used in this study.

Supplementary Table 1: List of all base and prime editor constructs and controls used in this study.

All plasmids (except pSQT817) are in a pCMV backbone (digest from ABEmax, Addgene #112101). All base editor constructs are fused to P2A-EGFP. All constructs are suitable for mammalian expression. Growth in bacteria: 37°C, resistance: Ampicillin. The denomination "Cas9" in this file refers to *Streptococcus pyogenes* Cas9 (SpCas9).

Plasmid	Abbreviation	Construct	Promoter	
pRZ70	nCas9	bpNLS-nCas9-bpNLS-P2A-eGFP-NLS	CMV	based on Addgene #123616
pBM529	nCas9-UGI-UGI	bpNLS-nCas9-UGI-UGI-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pRZ2910	nCas9-eUNG	bpNLS-nCas9-UNG(E.coli)-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pRZ3980	nCas9-hUNG	bpNLS-nCas9-UNG(H.sapiens)-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pBM539	BE4max	bpNLS-rAPOBEC1-nCas9-UGI-UGI-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pBM608	BE4max(R33A)	bpNLS-rAPOBEC1(R33A)-nCas9-UGI-UGI-bpNLS-P2A-eGFP	CMV	Addgene #140251
pRZ3144	BE4max-ΔUGI	bpNLS-rAPOBEC1-nCas9-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pBM1219	BE4max(R33A)-ΔUGI = miniCGBE1	bpNLS-rAPOBEC1(R33A)-nCas9-bpNLS-P2A-eGFP	CMV	Addgene #140253
pRZ4037	BE4max(R33A)-ΔUGI-eUNG	bpNLS-rAPOBEC1(R33A)-nCas9-UNG(E.coli)-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pRZ3885	eUNG- BE4max(R33A)-ΔUGI = CGBE1	bpNLS-UNG(E.coli)-rAPOBEC1(R33A)-nCas9-bpNLS-P2A-eGFP	CMV	Addgene #140252
pRZ3956	BE4max-ΔUGI-hUNG	bpNLS-rAPOBEC1-nCas9-UNG(H.sapiens)-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pRZ4115	hUNG-BE4max-ΔUGI	bpNLS-UNG(H.sapiens)-rAPOBEC1-nCas9-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pRZ3964	BE4max(R33A)-	bpNLS-rAPOBEC1(R33A)-nCas9-	CMV	based on Addgene #112093

Supplementary Table 1 (Continued)

	Δ UGI-hUNG	UNG(H.sapiens)-bpNLS-P2A-eGFP		
pRZ4133	hUNG- BE4max(R33A)- Δ UGI	bpNLS-UNG(H.sapiens)-rAPOBEC1(R33A)- nCas9-bpNLS-P2A-eGFP	CMV	based on Addgene #112093
pBM550	hA3A-BE3	hA3A-nCas9-UGI-NLS-P2A-eGFP	CMV	based on Addgene #131314
pBM563	eA3A-BE3	eA3A-nCas9-UGI-NLS-P2A-eGFP	CMV	based on Addgene #131315
pBM574	hAID-BE3	hAID-nCas9-UGI-NLS-P2A-eGFP	CMV	based on Addgene #131316
pBM584	Target-AID	NLS-nCas9-NLS-SH3-3xFLAG-pmCDA1-NLS- UGI-P2A-eGFP	CMV	based on Addgene #131300
pRZ2920	eA3A-BE3- Δ UGI-eUNG	eA3A-nCas9-UNG(E.coli)-NLS-P2A-eGFP	CMV	based on Addgene #131315
pRZ3246	hAID-BE3- Δ UGI-eUNG	hAID-nCas9-UNG(E.coli)-NLS-P2A-eGFP	CMV	based on Addgene #131316
pRZ4510	Target-AID- Δ UGI-eUNG	NLS-nCas9-NLS-SH3-3xFLAG-pmCDA1-NLS- UNG(E.coli)-P2A-eGFP	CMV	based on Addgene #131300
pRZ4758	hA3A-BE3- Δ UGI-hUNG	hA3A-nCas9-UNG(H.sapiens)-NLS-P2A-eGFP	CMV	based on Addgene #131314
pRZ2983	eA3A-BE3- Δ UGI-hUNG	eA3A-nCas9-UNG(H.sapiens)-NLS-P2A-eGFP	CMV	based on Addgene #131315
pRZ2988	hAID-BE3- Δ UGI-hUNG	hAID-nCas9-UNG(H.sapiens)-NLS-P2A-eGFP	CMV	based on Addgene #131316
pRZ3002	Target-AID- Δ UGI-hUNG	NLS-nCas9-NLS-SH3-3xFLAG-pmCDA1-NLS- UNG(H.sapiens)-P2A-eGFP	CMV	based on Addgene #131300
pJUL1459	ABEmax	bpNLS-TadA-TadA*-nCas9-bpNLS-P2A-eGFP- NLS	CMV	Addgene #112101
pRZ3748	ABEmax-UGI- UGI	bpNLS-TadA-TadA*-nCas9-UGI-UGI-bpNLS- P2A-eGFP-NLS	CMV	based on Addgene #112101
pRZ3026	ABEmax-eUNG	bpNLS-TadA-TadA*-nCas9-UNG(E.coli)-bpNLS- P2A-eGFP-NLS	CMV	based on Addgene #112101
pRZ3043	eUNG-ABEmax	bpNLS-UNG(E.coli)-TadA-TadA*-nCas9-bpNLS- P2A-eGFP-NLS	CMV	based on Addgene #112101

Supplementary Table 1 (Continued)

pRZ3989	ABEmax-hUNG	bpNLS-TadA-TadA*-nCas9-UNG(H.sapiens)- bpNLS-P2A-eGFP-NLS	CMV	based on Addgene #112101
pRZ4178	hUNG-ABEmax	bpNLS-UNG(H.sapiens)-TadA-TadA*-nCas9- bpNLS-P2A-eGFP-NLS	CMV	based on Addgene #112101
pJUL2448	nCas9(H840A)	bpNLS-nCas9(H840A)-bpNLS-P2A-eGFP	CMV	based on Addgene #132775
pJUL2440	PE2	bpNLS-nCas9(H840A)-MMLVrt-bpNLS-P2A- eGFP	CMV	based on Addgene #132775
pRZ4302	nCas9_NG	bpNLS-nCas9_NG-bpNLS-P2A-eGFP-NLS	CMV	nCas9_NG variant based on Addgene #125615
pBM1246	CGBE1_NG	bpNLS-UNG(E.coli)-rAPOBEC1(R33A)- nCas9_NG-bpNLS-P2A-eGFP	CMV	Addgene #140256
pBM1182	miniCGBE1_NG	bpNLS-rAPOBEC1(R33A)-nCas9_NG-bpNLS- P2A-eGFP	CMV	Addgene #140257
pRZ4327	nCas9_VRQR	bpNLS-nCas9_VRQR-bpNLS-P2A-eGFP-NLS	CMV	
pBM1192	CGBE1_VRQR	bpNLS-UNG(E.coli)-rAPOBEC1(R33A)- nCas9_VRQR-bpNLS-P2A-eGFP	CMV	Addgene #140254
pBM1173	miniCGBE1_VR QR	bpNLS-rAPOBEC1(R33A)-nCas9_VRQR- bpNLS-P2A-eGFP	CMV	Addgene #140255
pSQT817	Cas9	Cas9-NLS-3xFLAG	CAG	Addgene #53373

Supplementary Table 2: List of pathogenic and likely-pathogenic disease-causing cytosine SNVs at protospacer position 6 with NG PAMs that in principle can be targeted and reversed using CGBE-NG constructs.

chr	start	end	snpId	fix	gRNA+PAM	geneId	hgvsCod	hgvsProt	phenotypeList
chr 1	45333 512	45333 513	37012 4822	C>G	GGTATCAGGAGAC AGAGGCCTGC	4595 M UTYH	NM_001048171.1: c.206C>G	NP_001041636.1: p.Ser69Ter	Hereditary cancer- predisposing syndrome,MYH- associated polyposis,not provided
chr 2	14940 10	14940 11	12190 8088	C>G	CATCTCCTTCCCA ATGAGACAGG	7173 TP O	NM_000547.5:c.1 978C>G	NP_000538.3:p.Gl n660Glu	Deficiency of iodide peroxidase,not provided
chr 2	73950 573	73950 574	53674 6349	C>G	TGGAAGTGGGATG GGAATGGGGA	1716 DG UOK	NM_080916.1:c.4 44-11C>G		Progressive external ophthalmoplegia with mitochondrial DNA deletions,autosomal recessive 4,not provided
chr 2	15152 5264	15152 5265		C>G	TTTTTCTAATTTGT CTAAATAGA	4703 NE B	NM_001271208.1: c.22275C>G	NP_001258137.1: p.Tyr7425Ter	Nemaline myopathy,Nemaline myopathy 2
chr 2	17856 4810	17856 4811	55731 2035	C>G	AGATGCTAGCCAA TAATTTTGGT	7273 TT N	NM_001267550.2: c.81321C>G	NP_001254479.2: p.Tyr27107Ter	Dilated cardiomyopathy 1G,Limb-girdle muscular dystrophy,type 2J,Primary dilated cardiomyopathy
chr 2	21475 2471	21475 2472	58778 1707	C>G	AGGATCATTCATT CTTCTCTGGT	580 BAR D1	NM_000465.3:c.1 652C>G	NP_000456.2:p.Se r551Ter	Breast cancer,susceptibility to,Familial cancer of breast,Hereditary cancer- predisposing syndrome,not provided
chr 3	33072 617	33072 618	39812 3350	C>G	CGGGACTAGTGAA TGCTTCCTGA	2720 GL B1	NM_000404.3:c.1 71C>G	NP_000395.2:p.Ty r57Ter	GM1 gangliosidosis type 2,GM1 gangliosidosis type 2,Gangliosidosis

Supplementary Table 2 (Continued)

									GM1 type 3,Gangliosidosis GM1 type 3,Infantile GM1 gangliosidosis,Infantile GM1 gangliosidosis,Mucopolysaccharidosis,MPS-IV-B,Mucopolysaccharidosi s,MPS-IV-B,not provided
chr 3	17921 8305	17921 8306	12191 3286	C>G	CTCCTCCTCAGTG ATTTCAGAGA	5290 PI K3CA	NM_006218.3:c.1 636C>G	NP_006209.2:p.Gl n546Glu	Adenocarcinoma of prostate,Adenocarcinom a of stomach,Breast adenocarcinoma,Gliobla stoma,Malignant melanoma of skin,Malignant neoplasm of body of uterus,Medulloblastoma, Neoplasm of brain,Neoplasm of the breast,Neoplasm of the large intestine,Squamous cell carcinoma of the head and neck,Transitional cell carcinoma of the bladder,Uterine Carcinosarcoma,Uterine cervical neoplasms
chr 5	60918 363	60918 364	14336 7518	C>G	ACACTCTATCTGT GAACATCAGG	1161 ER CC8	NM_000082.3:c.3 00C>G	NP_000073.1:p.Ty r100Ter	Cockayne syndrome type A,not provided
chr 5	11276 7278	11276 7279	74953 290	C>G	GGCTTCATACAGA TCCTTCCCGG	324 APC	NM_000038.5:c.3 11C>G	NP_000029.2:p.Se r104Ter	Hereditary cancer- predisposing syndrome,not provided,not specified

Supplementary Table 2 (Continued)

chr 8	89943 319	89943 320	73088 1857	C>G	GATCTCATCCTCC AATGATGTGT	4683 NB N	NM_002485.4:c.2 117C>G	NP_002476.2:p.Se r706Ter	Hereditary cancer- predisposing syndrome, Microcephaly, normal intelligence and immunodeficiency, not provided
chr X	71223 914	71223 915	87885 3697	C>G	GATGGCGAAGAAT TGGTCATAGC	2705 GJ B1	NM_001097642.2: c.208C>G	NP_001091111.1: p.Pro70Ala	Charcot-Marie-Tooth Neuropathy X, not provided
chr 11	52271 57	52271 58	33941 377	C>G	CCTAGCGTGTGGC TCCACAGGGT	3043 HB B	NM_000518.5:c.- 137C>G		Beta thalassemia intermedia, Beta-plus- thalassemia, beta Thalassemia, not provided
chr 11	36575 651	36575 652		C>G	TAGCTCAGACCCC TTTCACCCGA	5896 RA G1	NM_000448.2:c.2 348C>G	NP_000439.1:p.Se r783Ter	Alpha/beta T-cell lymphopenia with gamma/delta T-cell expansion, severe cytomegalovirus infection, and autoimmunity, Combined cellular and humoral immune defects with granulomas, Combined cellular and humoral immune defects with granulomas, Histiocytic medullary reticulosis, Severe immunodeficiency, autos omal recessive, T-cell negative, B-cell negative, NK cell- positive, Severe immunodeficiency, autos omal recessive, T-cell

Supplementary Table 2 (Continued)

									negative,B-cell negative,NK cell- positive
chr 11	47342 718	47342 719	39751 5905	C>G	CTCCCCGGTCAGC TCCACCCCGT	4607 M YBPC3	NM_000256.3:c.1 483C>G	NP_000247.2:p.Ar g495Gly	Cardiovascular phenotype,Familial hypertrophic cardiomyopathy 1,Hypertrophic cardiomyopathy,Primary familial hypertrophic cardiomyopathy,not provided
chr 11	77162 844	77162 845	10575 17774	C>G	GGGGACTGGAGG GGCAGCTGTGA	4647 M YO7A	NM_000260.3:c.1 555-8C>G		Deafness,autosomal recessive 2,Usher syndrome,type 1,not provided
chr 11	10827 9577	10827 9578	58777 9833	C>G	TTCAACTATGCAT TTTCAAAAGC	472 AT M	NM_000051.3:c.3 372C>G	NP_000042.3:p.Ty r1124Ter	Ataxia-telangiectasia syndrome,Ataxia- telangiectasia syndrome,Familial cancer of breast,Hereditary cancer- predisposing syndrome,not provided
chr 12	49185 575	49185 576	13785 3043	C>G	GATGCCGGGATAG GGCACCAGGT	7846 TU BA1A	NM_006009.4:c.7 90C>G	NP_006000.2:p.Ar g264Gly	Tubulinopathies,not provided
chr 12	49186 819	49186 820	10575 20574	C>G	GGATGCAGATGCA CTCACGCTGT	7846 TU BA1A	NM_006009.4:c.1 7C>G	NP_006000.2:p.Se r6Cys	Tubulinopathies,not provided
chr 12	12073 7363	12073 7364	74949 1616	C>G	CCCAGCTAGAGAG ACTAGGAGGA	35 ACA DS	NM_000017.3:c.3 69C>G	NP_000008.1:p.Ty r123Ter	Deficiency of butyryl- CoA dehydrogenase,not provided
chr 15	89318 736	89318 737	20173 2356	C>G	CACACCGCTGGTC ATAAACTGGG	5428 PO LG	NM_002693.2:c.3 286C>G	NP_002684.1:p.Ar g1096Gly	not provided
chr 15	89318 737	89318 738	76164 9878	C>G	ACACGCCTGGTCA TAAACTGGGA	5428 PO LG	NM_002693.2:c.3 285C>G	NP_002684.1:p.Se r1095Arg	POLG-Related disorder,Progressive

Supplementary Table 2 (Continued)

									sclerosing poliodystrophy,not provided
chr 15	89329 050	89329 051	76941 0130	C>G	AAGCTCCTTAGCC CTGAGATGGC	5428 PO LG	NM_002693.2:c.9 15C>G	NP_002684.1:p.Se r305Arg	Progressive sclerosing poliodystrophy,not provided
chr 19	10823 885	10823 886	58778 3598	C>G	TCTCGCGGTAGAC GCCAGCTCGG	1785 DN M2	NM_001005360.2: c.1880C>G	NP_001005360.1: p.Pro627Arg	Myopathy,centronuclear, not provided
chr 19	11089 413	11089 414	87925 4375	C>G	GCAGGCGGAGGA GTTTGCAGTGG	3949 LD LR	NM_000527.4:c.- 135C>G		Familial hypercholesterolemia
chr 19	11105 406	11105 407	75259 6535	C>G	TTGTCCCAGGCCC ACAGCTGGGG	3949 LD LR	NM_001195800.1: c.314-1985C>G		Familial hypercholesterolemia
chr 19	11107 485	11107 486	87598 9909	C>G	GACCACTCCCGGC AGTCTCTAGC	3949 LD LR	NM_000527.4:c.9 12C>G	NP_000518.1:p.As p304Glu	Familial hypercholesterolemia,Fa miliar hypercholesterolemi as
chr 19	35033 653	35033 654	10489 4718	C>G	ACGTGCCACTCGT AGTCGCCCGA	6324 SC N1B	NM_199037.4:c.3 63C>G	NP_950238.1:p.C ys121Trp	Atrial fibrillation,familial,13,At rial fibrillation,familial,13,Br ugada syndrome 5,Brugada syndrome 5,Epileptic encephalopathy,early infantile,52,Generalized epilepsy with febrile seizures plus,Generalized epilepsy with febrile seizures plus,type 1,Generalized epilepsy with febrile seizures plus,type 1,Seizures,not provided
chr	44636	44636	12190	C>G	GACATCCAGTTCC	100 AD	NM_001322050.1:		Severe combined

Supplementary Table 2 (Continued)

20	278	279	8725		ACCTGCAAGG	A	c.-247C>G		immunodeficiency due to ADA deficiency,not provided
chr 20	50892556	50892557	587777526	C>G	TGCTCCTAAGTGC GCTTCACAGG	23394 A DNP	NM_015339.4:c.2157C>G	NP_056154.1:p.Tyr719Ter	Helsmoortel-van der aa syndrome,Inborn genetic diseases,not provided
chr 1	2229114	2229115	869312901	G>C	TGGGACGCGAGA AGCGCCTGTGT	6497 SK I	NM_003036.3:c.349G>C	NP_003027.1:p.Gly117Arg	Shprintzen-Goldberg syndrome,not provided
chr 1	17028736	17028737	397516833	G>C	TTTCACGCATCTG TGGCTCTTGT	6390 SD HB	NM_003000.2:c.287-1G>C		Gastrointestinal stroma tumor,Hereditary Paranglioma-Pheochromocytoma Syndromes,Hereditary cancer-predisposing syndrome,Parangliomas as 4,Parangliomas 4,Pheochromocytoma,not provided
chr 1	155236297	155236298	398123527	G>C	AGAGTCTGCGGCT AGGCTCCTGG	2629 GB A	NM_001005741.2:c.1171G>C	NP_001005741.1:p.Val391Leu	Gaucher disease,Gaucher's disease,type 1
chr 1	156115274	156115275	794728589	G>C	GCGCGCTGAGTTC GCCCAGGTGG	4000 LM NA	NM_170707.3:c.356+1G>C		Primary dilated cardiomyopathy,not provided
chr 1	156130756	156130757	267607570	G>C	TCTGCCGGGCCAG GTGGCCAAGG	4000 LM NA	NM_001282624.1:c.254G>C	NP_001269553.1:p.Arg85Pro	Cardiovascular phenotype,Charcot-Marie-Tooth disease,type 2,Dilated cardiomyopathy 1A,not provided
chr 1	215680376	215680377	397517977	G>C	TTACACGGAACAA GCCATCAAGC	7399 US H2A	NM_206933.2:c.12067-1G>C		Retinitis pigmentosa 39,Usher syndrome,type 2A,Usher syndrome,type 2A
chr	21641	21641	87665	G>C	CCTAACGTGTGTT	7399 US	NM_206933.2:c.4		Retinitis pigmentosa

Supplementary Table 2 (Continued)

1	8679	8680	7730		ATAGAAAAGA	H2A	86-1G>C		39,Usher syndrome,type 2A,Usher syndrome,type 2A,not provided
chr 2	26192308	26192309	1167218743	G>C	GAAGTCTAAGTCC GTTTGAGTGG	3030 HA DHA	NM_000182.4:c.2000+1G>C		Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency,Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency,Mitochondrial trifunctional protein deficiency
chr 2	178546048	178546049	869320742	G>C	GCCTGCAGCCTTC CGCAGGAAGA	7273 TT N	NM_001256850.1:c.90264G>C	NP_001243779.1:p.Trp30088Cys	Hereditary myopathy with early respiratory failure,not provided
chr 2	178609682	178609683	727504799	G>C	CATTGCTAAGTTT TGTTTCATGG	7273 TT N	NM_001267550.2:c.51739+1G>C		Cardiomyopathy,Primary dilated cardiomyopathy
chr 2	240786338	240786339	672601366	G>C	ACAAGCCCAGGTG GGTGGAGGGA	547 KIF 1A	NM_001244008.1:c.604G>C	NP_001230937.1:p.Ala202Pro	Mental retardation,autosomal dominant 9
chr 5	177267718	177267719	587784141	G>C	TACAGCTAAGCCT GAAGAATAGC	64324 N SD1	NM_022455.4:c.5303+1G>C		Beckwith-Wiedemann syndrome,Sotos syndrome 1
chr 5	177269600	177269601	863224905	G>C	TCCCACGTGGTGG CCAGCTGAGA	64324 N SD1	NM_022455.4:c.5304-1G>C		Beckwith-Wiedemann syndrome,Sotos syndrome 1
chr 7	5999275	5999276	988423880	G>C	AAACACGAGTATG CCAAAATGGT	5395 PM S2	NM_000535.6:c.538-1G>C		Hereditary cancer-predisposing syndrome,Lynch syndrome
chr 7	16406337	16406338	767978961	G>C	TTTTACAGTATGT TGGATAAAGG	729920 CRPPA	NM_001101426.3:c.258-1G>C		Congenital muscular dystrophy-dystroglycanopathy with brain and eye

Supplementary Table 2 (Continued)

									anomalies,type A7,Muscular dystrophy-dystroglycanopathy (limb-girdle),type c,7,not provided
chr 7	14080 1516	14080 1517	13259 51163	G>C	TTGTCCAAAGCTG CTTTTCCAGG	673 BRA F	NM_004333.5:c.7 55G>C	NP_004324.2:p.Ar g252Pro	Global developmental delay,not provided
chr 8	99823 978	99823 979	75601 3171	G>C	AGGAGCTAAGCA AATCATAACGA	157680 VPS13B	NM_017890.4:c.9 405+1G>C		Cohen syndrome,not provided
chr 9	34648 113	34648 114	39812 3181	G>C	TGATACATCTTTG AAAACAAAGG	2592 GA LT	NM_000155.3:c.5 08-1G>C		Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase,not provided
chr 9	10890 0302	10890 0303	13785 3022	G>C	TTCACCGATTGTC ACTGTTGTGC	8518 EL P1	NM_003640.4:c.2 087G>C	NP_003631.2:p.Ar g696Pro	Familial dysautonomia
chr X	10140 1751	10140 1752	10489 4845	G>C	CCTGCCCAGGCTT CCCTGGGAGT	2717 GL A	NM_000169.2:c.4 27G>C	NP_000160.1:p.Al a143Pro	Fabry disease,not provided
chr 11	53428 8	53428 9	10489 4229	G>C	GCGCCCGCGGTGT GGGCAAGAGT	HRAS,L RRC56	NM_005343.3:c.3 4G>C	NP_005334.1:p.Gl y12Arg	Neoplasm of the large intestine,Neoplasm of the thyroid gland
chr 11	25729 78	25729 79	12007 4186	G>C	GTGGTCGGGGGTG GTAAGTCGGA	3784 KC NQ1	NM_181798.1:c.5 33G>C	NP_861463.1:p.Tr p178Ser	Cardiovascular phenotype,Congenital long QT syndrome,Jervell and Lange-Nielsen syndrome 1,not provided
chr 11	66187 75	66187 76	12190 8195	G>C	AATACCGTGCCTT TTGGGACTGA	1200 TP P1	NM_000391.3:c.2 29G>C	NP_000382.3:p.Gl y77Arg	Ceroid lipofuscinosis neuronal 2,not provided
chr 11	17408 517	17408 518	10575 17420	G>C	TTCCACATCATTG CCATGAAGGA	6833 AB CC8	NM_000352.4:c.2 695-1G>C		Familial hyperinsulinism,Persistent hyperinsulinemic hypoglycemia of infancy
chr 11	66521 268	66521 269	74852 3268	G>C	TTGCACATGAGCC TTCCCAGCGT	582 BBS 1	NM_024649.4:c.7 24-1G>C		Bardet-Biedl syndrome 1

Supplementary Table 2 (Continued)

chr 11	68047 479	68047 480	13785 3150	G>C	TGTTCCGGGATGT GGGCCACGGG	10312 T CIRG1	NM_006019.3:c.1 213G>C	NP_006010.2:p.Gl y405Arg	Osteopetrosis autosomal recessive 1
chr 11	10825 7479	10825 7480	87665 9710	G>C	CAATACTCTCTAA TGCAATGTGC	472 AT M	NM_000051.3:c.2 251-1G>C		Ataxia-telangiectasia syndrome, Hereditary cancer-predisposing syndrome
chr 11	10829 2617	10829 2618	75952 0465	G>C	TATTACGCCTTCTT GTATCATGG	472 AT M	NM_000051.3:c.4 437-1G>C		Ataxia-telangiectasia syndrome, Hereditary cancer-predisposing syndrome, not provided
chr 12	10953 3532	10953 3533	53940 7162	G>C	CATCCCCTGCGTG GAGGTGTCGG	89910 U BE3B	NM_130466.3:c.2 990G>C	NP_569733.2:p.Ar g997Pro	Inborn genetic diseases, Kaufman oculocerebrofacial syndrome
chr 12	11091 1175	11091 1176	19947 4813	G>C	TTGCACGTTGACC AGATGTTTCGC	4633 M YL2	NM_000432.3:c.4 03-1G>C		Familial hypertrophic cardiomyopathy 10, not provided
chr 13	32319 329	32319 330	81002 840	G>C	GGTAACTAATGCA ATATGGTAGA	675 BRC A2	NM_000059.3:c.3 16+5G>C		Familial cancer of breast, Hereditary breast and ovarian cancer syndrome, not provided
chr 13	32326 612	32326 613	80358 871	G>C	TCATACGTAATAA TAGCAAATGT	675 BRC A2	NM_000059.3:c.6 31G>C	NP_000050.2:p.V al211Leu	Breast-ovarian cancer, familial 2, Hereditary cancer- predisposing syndrome, not provided
chr 14	49621 978	49621 979	73088 2218	G>C	TGGAACCTGCATT TTGTGTGGGA	4247 M GAT2	NM_002408.3:c.7 11G>C	NP_002399.1:p.Ly s237Asn	Abnormal facial shape, Abnormal glycosylation (CDG IIa), Carbohydrate- deficient glycoprotein syndrome type II, Global developmental delay
chr 15	40407 639	40407 640	22293 11	G>C	GCTTCCTCAGACC ATGGCTAAGT	3712 IV D	NM_002225.4:c.1 49G>C	NP_002216.3:p.Ar g50Pro	Isovaleryl-CoA dehydrogenase

Supplementary Table 2 (Continued)

									deficiency
chr 15	42410957	42410958	778768583	G>C	ACTTTCACAGTTT CATCTGCTGC	825 CAP N3	NM_000070.2:c.2 338G>C	NP_000061.1:p.As p780His	Limb-girdle muscular dystrophy,type 2A,not provided
chr 16	16163100	16163101	63750473	G>C	CCAGGCCAGCACA GTGGTCCGGG	368 ABC C6	NM_001171.5:c.3 398G>C	NP_001162.4:p.Gl y1133Ala	Pseudoxanthoma elasticum,not provided
chr 16	89513036	89513037	912983346	G>C	ACGAACGTATATT CATCTGATGT	6687 SP G7	NM_003119.3:c.3 76G>C	NP_003110.1:p.Gl u126Gln	Hereditary spastic paraplegia,not provided
chr 17	7674892	7674893	587778720	G>C	TTTTCCACATAGT GTGGTGGTGC	7157 TP 53	NM_000546.5:c.6 38G>C	NP_000537.3:p.Ar g213Pro	Adenocarcinoma of prostate,Adenocarcinom a of stomach,Adenoid cystic carcinoma,Adrenocortica l carcinoma,Carcinoma of esophagus,Glioblastoma, Hepatocellular carcinoma,Hereditary cancer-predisposing syndrome,Lung adenocarcinoma,Maligna nt melanoma of skin,Malignant neoplasm of body of uterus,Nasopharyngeal Neoplasms,Neoplasm of brain,Neoplasm of the breast,Neoplasm of the large intestine,Ovarian Serous Cystadenocarcinoma,Pan creatic adenocarcinoma,Renal cell carcinoma,papillary,1,Sq

Supplementary Table 2 (Continued)

									uamous cell carcinoma of the head and neck,Squamous cell carcinoma of the skin,Squamous cell lung carcinoma,Transitional cell carcinoma of the bladder,Uterine Carcinosarcoma,not specified
chr 17	31336331	31336332	1555534596	G>C	AACTACATTACAG ATCTGCTTGA	4763 NF1	NM_001042492.2: c.6007-1G>C		Hereditary cancer-predisposing syndrome,not provided
chr 17	43124015	43124016	80358010	G>C	ATCTGCTAAGTCA GCACAAGAGT	672 BRCA1	U14680.1:n.199+1 G>C		Breast-ovarian cancer,familial 1,Hereditary cancer-predisposing syndrome
chr 18	23538608	23538609	80358254	G>C	CTCAGCGGGGAGA CTTCATGAGA	4864 NPC1	NM_000271.4:c.2 974G>C	NP_000262.2:p.G1 y992Arg	Niemann-Pick disease type C1,Niemann-Pick disease,type C,not provided
chr 19	11105407	11105408	200727689	G>C	CCTGCCACAACGA CCCCGACTGC	3949 LDLR	NM_001195800.1: c.314-1984G>C		Familial hypercholesterolemia
chr 19	11105437	11105438	879254565	G>C	GCTCGCATGAGTG GCCGCAGCGC	3949 LDLR	NM_001195800.1: c.314-1954G>C		Familial hypercholesterolemia
chr 19	11107514	11107515	879254729	G>C	GTGCGCTGAGTCT CGGTGCAGGC	3949 LDLR	NM_000527.4:c.9 40+1G>C		Familial hypercholesterolemia
chr 19	11113387	11113388	121908036	G>C	CTCTGCACACGGA GGTGGCCAGC	3949 LDLR	NM_000527.4:c.1 297G>C	NP_000518.1:p.As p433His	Familial hypercholesterolemia
chr 19	11116200	11116201	28942082	G>C	CAATGCCATCACC CTAGGTATGT	3949 LDLR	NM_000527.4:c.1 694G>C	NP_000518.1:p.G1 y565Ala	Familial hypercholesterolemia
chr 19	11116212	11116213	875989926	G>C	CCTAGCTATGTTC GCAGGACAGC	3949 LDLR	NM_000527.4:c.1 705+1G>C		Familial hypercholesterolemia
chr 19	38526991	38526992		G>C	CCCCACAAAGACA CAGATGAGGA	6261 RYR1	NM_000540.2:c.1 0627-1G>C		not provided

Supplementary Table 2 (Continued)

chr 19	41422 369	41422 370	39812 3508	G>C	GCATTCGTATGGG CTCTGCTGGC	593 BCK DHA	NM_000709.3:c.8 53G>C	NP_000700.1:p.A1 a285Pro	Maple syrup urine disease,not provided
chr 19	55154 093	55154 094	39751 6354	G>C	GGCCCCGGCTAAG GAGTCCCTGG	7137 TN NI3	NM_000363.4:c.4 85G>C	NP_000354.4:p.Ar g162Pro	Hypertrophic cardiomyopathy,Primary familial hypertrophic cardiomyopathy
chr 21	43774 759	43774 760	14748 4110	G>C	GTCTACGTGAGGT CCCAGCTTGA	1476 CS TB	NM_000100.3:c.6 7-1G>C		Epilepsy,progressive myoclonic 1A (Unverricht and Lundborg),Inborn genetic diseases,Progressive myoclonic epilepsy,Unverricht- Lundborg syndrome,not provided

Supplementary Table 3: Sequences of gRNAs and pegRNAs, primers, and amplicons used in this study.

All gRNAs were of the form 5’-

NNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGCTTTTTTT-3’ (from BsmBI digest of BPK1520, Addgene #65777).

Shown below are the protospacer regions (NNNNNNNNNNNNNNNNNNNNNNNN for these gRNAs), primers (including illumina-compatible adapters) to amplify genomic loci for these gRNAs, and the genomic amplicons. ABE site 1-19 gRNAs and primers were described by Gaudelli et al, Nature 2017. Most gRNAs are widely used in the field (e.g. described in Kleinstiver et al, Nature 2016 - PMID 26735016 or Komor et al, Sci Adv 2017 - PMID 28875174.) ABE site 20 and 21, HEK site 5 and 6, PDCD1 site 2, PPP1R12C site 7 and 8, and VEGFA site 4 are new gRNAs used in this study to expand sequence contexts around C6. All HEK site 2, 3, and 4 off target sites were described by Gaudelli et al, Nature 2017 and Rees et al, Sci Adv 2019. All EMX1 site 1 and FANCF site 1 off target sites were described by Tsai et al, Nat Biotechnol 2014.

target	spacer	PAM	Fwd PCR1	Rev PCR1	Amplicon
ABE site 7	GAATACT AAGCATA GACTCC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGTT CCACCGCCTTGT T TACTGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCAGCCTCTTTCTTCGC CCTC	GTTCCACCGCCTTGTTTACTGCTCAGTCC CCCAGAATGGCAGGCACGTAGTAGGGCT CAGTAAATATTTGCTGAATGAAGTTATGA TCGAAA GAATACTAAGCATAGACTCCA GGATAAGGTAAGTAATATAGAGTTCCAC AGTGGGCAGGTCTATGCAAACCTACCCC AAAGTCCGAGGAAGCTGAGAGGGCGAAG AAAGAGGCTGAG
ABE site 8	GTAAACA AAGCATA GACTGA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAT ACTTAAAGTAAA GCAGAAGGAAT AACAGTGC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG AAGTTTCATTCCACTAT AGAAGTGGTTCTG	CATACTTAAAGTAAAGCAGAAGGAATAA CAGTGCCCCAAATTATTTACTACCCCTGT TCCTAAAGCCCACCATCTTTTGTACTCTC AGAGTAAACAAAGCATAGACTGAGGGG TACAATCCTACTCTAGTCCCTTTTCCTCATG CAACAGCTATGAAATAGCTGGTTAGTACT GACAGCAATTCTTGGCCACAAACAGAAC CAGTTCTATAGTGAATGAACTTC
ABE site 9	GAAGACC AAGGATA	TGG	ACACTCTTTCCC TACACGACGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA	ATCATAGCGAGACCCTGTCTCAGCAACA ACAACAAAACGGACAGTGAAGAGAGTAT

Supplementary Table 3 (Continued)

	GA CTGC		TTCCGATCT GAG TACTCTCCAATG ACACCTTTGGC	TCATAGCGAGACCCTG TCTCAGC	TTGAAGGAGAGGGGAACAGGCAGCGTAT TGCTTATTGCTGAGGGGCAAAGT GAAGA CCAAGGATAGACTGCTGG GCTTGACAG CATGGAGGGTGCTGGGGGCCTAGGCAAG TGCAATATTCATGTGGTGTCTATTGGGCCA AAGGTGTCATTGGAGAGTACTC
ABE site 18	ACACACA CACTTAG AATCTG	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCT ATC TCAGCGCTTTTCG TCCAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCATTTCCTCCACTCCCT CC	ATCTCAGCGCTTTCGTCCACCACCCCTCT ACACACACACAGAC ACACACACTTAG AATCTGTGG GAGTGGAGCCCAGCAATCT GTGTTTGCACAAACCTCCAGGGAATTCT GATGCCCCTGAAGTTTGAGAACTAGAG GTAAACTATACCTTTCTGCAGGCGAGAAC CTGGACCAAGGAGGGAGTGGGGAAATGA G
ABE site 19	CACACAC ACTTAGA ATCTGT	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCT ATC TCAGCGCTTTTCG TCCAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCATTTCCTCCACTCCCT CC	ATCTCAGCGCTTTCGTCCACCACCCCTCT ACACACACACAGAC CACACACTTAG AATCTGTGG AGTGGAGCCCAGCAATCT GTGTTTGCACAAACCTCCAGGGAATTCT GATGCCCCTGAAGTTTGAGAACTAGAG GTAAACTATACCTTTCTGCAGGCGAGAAC CTGGACCAAGGAGGGAGTGGGGAAATGA G
ABE site 20	TTAAGCT GTAGTAT TATGAA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCT TGA TCCACCTGCCTC AGCCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG TGCTAAACAGAGAGTT ACTGCTCAGAC	GTGCTAAACAGAGAGTTACTGCTCAGAC ATGTAATAATAATAAATAACACATCAAA TAACCATAACCATT TTAAGCTGTAGTATT ATGAAGGG AAATCTGGAGCAAAGAGAA TAGACTGTAGGAAACCAGTTAAGAAAT AGGACATGGAGGCTAGGTGCAGTGGCTC ACGCCTGTAATCGCAGCACTTTGGGAGGC TGAGGCAGGTGGATCA
ABE site 21	CCTGGCC TGGGTCA ATCCTT	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCT CAG GGAAACGCCCAT GCAATTAGTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TGCACATACTAGCCCC TGTCTAGG	CAGGGAAACGCCCATGCAATTAGTCTATT TCTGCTGCAAGTAAGCATGCATTTGTAGG CTTGATGCTTTTTTTCTGCTTCTCCAGCC CC TGGCCTGGGTCAATCCTTGGG GCCCAG ACTGAGCACGTGATGGCAGAGGAAAGGA AGCCCTGCTTCCTCCAGAGGGCGTCGCAG

Supplementary Table 3 (Continued)

					GACAGCTTTTCCTAGACAGGGGCTAGTAT GTGCAG
<i>EMX1</i> site 1	GAGTCCG AGCAGAA GAAGAA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCA GTGGCTGCTCTG GGGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC GATGTCCTCCCCATTG GCCTG	CCAGTGGCTGCTCTGGGGGCTCCTGAGT TTCTCATCTGTGCCCTCCCTCCCTGGCCC AGGTGAAGGTGTGGTTCCAGAACCGGAG GACAAAGTACAAACGGCAGAAGCTGGAG GAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACATCAACCGGT GGCGCATTGCCACGAAGCAGGCCAATGG GGAGGACATCG
<i>EMX1</i> site 2	GTATTCA CCTGAAA GTGTGC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTAT CTCCAGGCTCCT GTCCATTCTGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC AAACAAGCAAATATG GCTCCAGCC	TATCTCCAGGCTCCTGTCCATTCTGGGAT ATTTGAAATCTTTGAAGACAGGACACGT ATTCACCTGAAAGTGTGCAGGGCCCTG GAGCCACAGACTTTTCCATTTGCAGGAGT GGTGAAAAAGAAGGGATCCAGACCTCGT TCATTCAGAATAGCAGATTATTCCAAAGT AAAGTCTGATTAGTTCTGAAATATCGGCT GGAGCCATAGTTTGCTTGTTTG
<i>FANCF</i> site 1	GGAATCC CTTCTGC AGCACC	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGAC GTAGGTAGTGCT TGAGACCGCC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA GTTGCCAGAGTCAAG GAACACGG	AGTTGCCCAGAGTCAAGGAACACGGATA AAGACGCTGGGAGATTGACATGCATTTC GACCAATAGCATTGCAGAGAGGCGTATC ATTTGCGGGATGTTCCAATCAGTACGCAG AGAGTCGCCGTCTCCAAGGTGAAAGCGG AAGTAGGGCCTTCGCGCACCTCATGGAA TCCCTTCTGCAGCACCTGGATCGCTTTT CCGAGCTTCTGGCGGTCTCAAGCACTACC TACGTC
HEK site 2 (ABE site 1)	GAACACA AAGCATA GACTGC	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTACA AGACCTGGCTGA GCTAACTGTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAGCCCCATCTGTCAA ACTGTG	ACAAGACCTGGCTGAGCTAACTGTGACA GCATGTGGTAATTTTCCAGCCCGCTGGCC CTGTAAAGGAACTGGAACACAAAGCAT AGACTGCGGGCGGGCCAGCCTGAATA GCTGCAAACAAGTGCAGAATATCTGATG ATGTCATACGCACAGTTTGACAGATGGG GCTGG
HEK site 3	GGCCCAG ACTGAGC	TGG	ACACTCTTTCCC TACACGACGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTT	GCCCATGCAATTAGTCTATTTCTGCTGCA AGTAAGCATGCATTTGTAGGCTTGATGCT

Supplementary Table 3 (Continued)

	ACGTGA		TTCCGATCTGCC CATGCAATTAGT CTATTTCTGCTG	GCAGGAGCTGCACATA CTAGCC	TTTTTTCTGCTTCTCCAGCCCTGGCCTGGG TCAATCCTTGGGGCCAGACTGAGCAC GTGATGGCAGAGGAAAGGAAGCCCTGCT TCCTCCAGAGGGCGTCGCAGGACAGCTTT TCCTAGACAGGGGCTAGTATGTGCAGCTC CTGC
HEK site 4	GGCACTG CGGCTGG AGGTGG	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTTC AAGATGGCTGAC AAAGGCCG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTT CCTTTCAACCCGAACG GAGACAC	TTCAAGATGGCTGACAAAGGCCGGGCTG GGTGGAAGGAAGGGAGGAAGGGCGAGG CAGAGGGTCCAAAGCAGGATGACAGGCA GGGGCACC GCGCGCCCCGGTGGCACTG CGGCTGGAGGTGGGGTTAAAGCGGAG ACTCTGGTGCTGTGTGACTACAGTGGGGG CCCTGCCCTCTCTGAGCCCCCGCCTCCAG GCCTGTGTGTGTGTCTCCGTTTCGGGTTGA AAGGA
HEK site 5	CTGGCCT GGGTCAA TCCTTG	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGCC CATGCAATTAGT CTATTTCTGCTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTT GCAGGAGCTGCACATA CTAGCC	GCCCATGCAATTAGTCTATTTCTGCTGCA AGTAAGCATGCATTTGTAGGCTTGATGCT TTTTTTCTGCTTCTCCAGCCCTGGCCTGG GTCAATCCTTGGGGCCCAGACTGAGCAC GTGATGGCAGAGGAAAGGAAGCCCTGCT TCCTCCAGAGGGCGTCGCAGGACAGCTTT TCCTAGACAGGGGCTAGTATGTGCAGCTC CTGC
HEK site 6	CAAAGCA GGATGAC AGGCAG	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTTC AAGATGGCTGAC AAAGGCCG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTT CCTTTCAACCCGAACG GAGACAC	TTCAAGATGGCTGACAAAGGCCGGGCTG GGTGGAAGGAAGGGAGGAAGGGCGAGG CAGAGGGTCCAAAGCAGGATGACAGGC AGGGGCACCGCGGCGCCCCGGTGGCACT GCGGCTGGAGGTGGGGGTAAAGCGGAG ACTCTGGTGCTGTGTGACTACAGTGGGGG CCCTGCCCTCTCTGAGCCCCCGCCTCCAG GCCTGTGTGTGTGTCTCCGTTTCGGGTTGA AAGGA
PDCD1 site 2	ACTTCCA CATGAGC GTGGTC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGG TACCGCATGAGC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCTCTTTGATCTGCGCC TTGGG	TGGTACCGCATGAGCCCCAGCAACCAGA CGGACAAGCTGGCCGCCTTCCCCGAGGA CCGCAGCCAGCCCGGCCAGGACTGCCGC TTCCGTGTCACACA ACTGCCCAACGGGGC

Supplementary Table 3 (Continued)

			CCCAG		TG ACTTCCACATGAGCGTGGTCAGG GC CCGGCGCAATGACAGCGGCACCTACCTCT GTGGGGCCATCTCCCTGGCCCCCAAGGCG CAGATCAAAGAGAG
<i>PPP1R1</i> 2C site 2	G GCACTC G GGGGCG A AGAGGA	GGG	A CACTCTTTCC T ACACGACGCTC T TCCGATCTACC TGAGCCAGCTCC CATAGC	G ACTGGAGTTCAGACG T GTGCTCTTCCGATCTG GGAAGGGGAAACAGT GGGC	GGGAAGGGGAAACAGTGGGCCAGAGGC GCGCAGAAGCCAGTAGAGCTCAAAGTGG TCCGGACTCAGGAGAGAGACGGCACGGC AGCGTTAGAGGGCAGAGTTCCGGCGGCA CAGCAAG G GCACT C GGGGGG C AGAG G A GGGCAGCGCAAAGTGACAATGGCCAGG GCCAGGCAGATAGACCAGA
<i>PPP1R1</i> 2C site 3	G AGCTCA C TGAACG C TGGCA	TGG	A CACTCTTTCC T ACACGACGCTC T TCCGATCTGCC TGGCCTTGGGAG CTTCC	G ACTGGAGTTCAGACG T GTGCTCTTCCGATCTC GGCTCTGCCATGGATC TGCCG	GCCTGGCCTTGGGAGCTTCTGTACCTC ACA G AGCT C ACTGAACGCTGGC A TGG C TCTAGTGCTTTCCAGGGGCACGTACTCCC CCGAGCTCCTGACAGCACAATCCACGTGT CACGCATGACCCCTCATAAGTTGCGCAC CATAACATTCCCATTGACAGATAAGGAA ACTGAGGCCCGGAAAGGAAAGTGGCTGG CTGAGACTCACGGCAGATCCATGGCAGA GCCG
<i>PPP1R1</i> 2C site 4	G ACCCTC A GCCGTG C TGCTC	GGG	A CACTCTTTCC T ACACGACGCTC T TCCGATCTGGA ATTCTGCATCAT GTGGGCGTGTCC	G ACTGGAGTTCAGACG T GTGCTCTTCCGATCTG GATCAGCCAGGAACGA AGACCACTGG	GGATCAGCCAGGAACGAAGACCACTGGG GCCACTGGAAA G ACCCT C AGCCGT G CT G CT C GGGTGGGACTCTGCGGCCCCAAA TCACAAGGCGACTGCCTGTGACACCCTGT GACACCCGGCTCCACACAGGAGCCTCAG AGGGGCTGCAGTGCAGCAGGACCTGGATC ACAGGCTGCACCCAGCTCTAAGGAGGG CGGGTCTTTTGACCCCTGCAGGACACG CCCACATGATGCAGAATTCC
<i>PPP1R1</i> 2C site 5	G CTGACT C AGAGAC C CTGAG	TGG	A CACTCTTTCC T ACACGACGCTC T TCCGATCTCCA CACCTAGGAGGA TAAGAGGCACG	G ACTGGAGTTCAGACG T GTGCTCTTCCGATCTG GAAAGGCCTGTGATCT CCGTTTCC	GGAAAGGCCTGTGATCTCCGTTTTCCAG TGCTTAGC G CT G ACT C AGAGACC T G A G TGGGACACACGACTGCATGGAGTACAG TTGGGAAACAACCTGGGCTAGTCTGGGCCT CCTGAGTCTCAGAGGGGAGCAGGGGGAA GAGCTCTGGGCAAGTCACTGGGCTGAAC CCTGGCCAGCCCAGTGTATGACCCTGGGT

Supplementary Table 3 (Continued)

					ATTTCAAGGTCATGGGGTGTGAGTCCTACC TCGTGCCTCTTATCCTCCTAGGTGTGG
<i>PPP1R1</i> 2C site 6	GGGGCTC AACATCG GAAGAG	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGC CAGGCAGATAG ACCAGACTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CTCTCTTCCGCATTGGA GTCGC	GGCCAGGCAGATAGACCAGACTGAGCTA TGGGAGCTGGCTCAGGTTTCAAGAGAGGG CAGGGCAGGGAAGGAGACAAAGTCCAGG ACCGGCTGGAAGGGCTCAACATCGGAA GAGGGGAAGTCGAGGGAGGGATGGTAA GGAGGACTGCATGGGTGAGCACAGGCTG CCAAAGCCAGGGCCAGTTAAAGCGACTC CAATGCGGAAGAGAGT
<i>PPP1R1</i> 2C site 7	GCTGGCT CAGGTTT AGGAGA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGCC AGGCAGATAGA CCAGACTGAGC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG CCTCCTTACCATTCCCC TTCGACC	GCCAGGCAGATAGACCAGACTGAGCTAT GGGAGCTGGCTCAGGTTTCAAGAGAGG GCAGGGCAGGGAAGGAGACAAAGTCCA GGACCGGCTGGAGGGGCTCAACATCGGA AGAGGGGAAGTCGAGGGAGGGATGGTAA GGAGGACTGCATGGGTGAGCACAGGCTG CCAAAGCCAGGGCCAGTTAAAGCGACTC CAATGCGGAAGAGAGTAGGTCTGAAGGGG AATGGTAAGGAGGC
<i>PPP1R1</i> 2C site 8	CTGCTCG GGGTGGG ACTCTG	CGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGA ATTCTGCATCAT GTGGGCGTGTCC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG GATCAGCCAGGAACGA AGACCACTGG	GGATCAGCCAGGAACGAAGACCACTGGG GCCACTGGAAAGACCCTCAGCCGTGCTG CTCGGGGTGGGACTCTGCGGCCCCAAA TCACAAGGCGACTGCCTGTGACACCCTGT GACACCCGGCTCCACACAGGAGCCTCAG AGGGGCTGCAGTGCAGCAGGACCTGGATC ACAGGCTGCACCCAGCTCTAAGGAGGG CGGGTCTTTTGCACCCCTGCAGGACACG CCCACATGATGCAGAATTCC
<i>RNF2</i> site 1	GTCATCT TAGTCAT TACCTG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGT GCAGACAAACG GAACTCAACC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAACATACAGAAGTCA GGAATGCTTGA	CCAACATACAGAAGTCAGGAATGCTTGA ATATAAATTTATTATTACTCTATGTTCTAT TTAAGTTTTCATGTTCTAAAAATGTATCC CAGTTTACACGTCTCATATGCCCTTGGC AGTCATCTTAGTCATTACCTGAGGTGTT CGTTGTAACATCATATAAACTGAGTTCCCA TGTTTTGCTTAATGGTTGAGTTCCGTTTGT CTGCAC

Supplementary Table 3 (Continued)

VEGFA site 4	GAGGACG TGTGTGT CTGTGT	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGT GCAGACGGCAGT CACTA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG GGGAGAGGGACACAC AGAT	CAGACGGCAGTCACTAGGGGGCGCTCGG CCACCACAGGGAAGCTGGGTGAATGGAG CGAGCAGCGTCTTCGAGAGT GAGGACGT GTGTGTCTGTGTGGG TGAGTGAGTGTGT GCGTGTGGGGTTGAGGGCGTTGGAGCGG GGAGAAGGCCAGGGGTCACTCCAGGATT CCAATAGATCTGTGTGTCCCTCT
ABE site 23	TAAGCAT AGACTCC AGGATA	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGTT CCACCGCCTTGT TACTGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCAGCCTCTTTCTTCGC CCTC	GTTCCACCGCCTTGTTTACTGCTCAGTCC CCCAGAATGGCAGGCACGTAGTAGGGCT CAGTAAATATTTGCTGAATGAAGTTATGA TCGAAAGAATACT TAAGCATAGACTCCA GGATAAGG TAAGTAATATAGAGTTCCAC AGTGGGCAGGTCTATGCAAACCTACCCC AAAGTCCGAGGAAGCTGAGAGGGCGAAG AAAGAGGCTGAG
ABE site 24	TACTCTG AGTGTAC AAAAGA	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAT ACTTAAAGTAAA GCAGAAGGAAT AACAGTGC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG AAGTTTCATTCCACTAT AGAACTGGTTCTG	GAAGTTTCATTCCACTATAGAACTGGTTC TGTTTGTGGCCAAGAATTGCTGTCACTAC TAACCAGCTATTTTCATAGCTGTTGCATGA GGAAAGGGACTAGAGTAGGATTGTACCC CTCAGTCTATGCTTTGTT TACTCTGAGTG TACAAAAGATGG TGGGCTTTAGGAACAG GGGTAGTGAATAATTGGGGCACTGTTAT TCCTTCTGCTTTACTTTAAGTATG
ABE site 25	AGTAAAC AAAGCAT AGACTG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAT ACTTAAAGTAAA GCAGAAGGAAT AACAGTGC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG AAGTTTCATTCCACTAT AGAACTGGTTCTG	CATACTTAAAGTAAAGCAGAAGGAATAA CAGTGCCCCAAATTATTCCTACCCCTGT TCCTAAAGCCCACCATCTTTTGTACACTC AG AGTAAACAAAGCATAGACTGAGGGG TACAATCCTACTCTAGTCCCTTTCCTCATG CAACAGCTATGAAATAGCTGGTTAGTACT GACAGCAATTCTTGGCCACAAACAGAAC CAGTTCTATAGTGGAATGAACTTC
ABE site 26	TTTGTGC AAACACA GATTGC	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTATC TCAGCGCTTTCG TCCAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCATTTCCCCACTCCCT CC	CTCATTTCCCCACTCCCTCCTTGGTCCAG GTTCTCGCCTGCAGAAAGGTATAGTTTAC CTCTAGTTCTCAAACCTTCAGCGGGCATCA GAATTCCCTGGAGGG TTTGTGCAAACAC AGATTGCTGGG GCTCCACTCCCACAGATT

Supplementary Table 3 (Continued)

					CTAAGTGTGTGTGTGTCTGTGTGTGTGTA GAGGGGTGGTGGACGAAAGCGCTGAGAT
ABE site 27	CGGGCAT CAGAATT CCCTGG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTATC TCAGCGCTTTCG TCCAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCATTTCCTCCACTCCCT CC	CTCATTTCCTCCACTCCCTCCTTGGTCCAG GTTCTCGCCTGCAGAAAGGTATAGTTTAC CTCTAGTTCTCAAACCTTCAGCGGGCATC AGAATTCCTGGAGGGTTTGTGCAAACA CAGATTGCTGGGCTCCACTCCCACAGATT CTAAGTGTGTGTGTGTCTGTGTGTGTGTA GAGGGGTGGTGGACGAAAGCGCTGAGAT
EMX1 site 3	AAAGTAC AAACGGC AGAAGC	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCA GTGGCTGCTCTG GGGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC GATGTCTCTCCCATTTG GCCTG	CCAGTGGCTGCTCTGGGGGCCTCCTGAGT TTCTCATCTGTGCCCCCTCCCTCCCTGGCCC AGGTGAAGGTGTGGTTCCAGAACCGGAG GACAAAGTACAAACGGCAGAAGCTGGA GGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACATCAACCGGT GGCGCATTGCCACGAAGCAGGCCAATGG GGAGGACATCG
EMX1 site 4	GTACAAA CGGCAGA AGCTGG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCA GTGGCTGCTCTG GGGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC GATGTCTCTCCCATTTG GCCTG	CCAGTGGCTGCTCTGGGGGCCTCCTGAGT TTCTCATCTGTGCCCCCTCCCTCCCTGGCCC AGGTGAAGGTGTGGTTCCAGAACCGGAG GACAAA GTACAAACGGCAGAAGCTGGA GGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACATCAACCGGT GGCGCATTGCCACGAAGCAGGCCAATGG GGAGGACATCG
FANCF site 2	GCTGCAG AAGGGAT TCCATG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGAC GTAGGTAGTGCT TGAGACCGCC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA GTTGCCAGAGTCAAG GAACACGG	GACGTAGGTAGTGCTTGAGACCGCCAGA AGCTCGGAAAAGCGATCCAGGTGCTGCA GAAGGGATTCCATGAGGTGCGCGAAGG CCCTACTTCCGCTTTCACCTTGGAGACGG CGACTCTCTGCGTACTGATTGGAACATCC GCGAAATGATACGCCTCTCTGCAATGCTA TTGGTCGAAATGCATGTCAATCTCCCAGC GTCTTTATCCGTGTTCCCTTGACTCTGGGC AACT
FANCF site 3	CGCCGTC TCCAAGG	CGG	ACACTCTTTCCC TACACGACGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA	AGTTGCCAGAGTCAAGGAACACGGATA AAGACGCTGGGAGATTGACATGCATTTC

Supplementary Table 3 (Continued)

	TGAAAG		TTCCGATCTGAC GTAGGTAGTGCT TGAGACCGCC	GTTGCCCAGAGTCAAG GAACACGG	GACCAATAGCATTGCAGAGAGGCGTATC ATTTTCGCGGATGTTCCAATCAGTACGCAG AGAGT CGCCGTCTCCAAGGTGAAAGCG G AAGTAGGGCCTTCGCGCACCTCATGGA ATCCCTTCTGCAGCACCTGGATCGCTTTT CCGAGCTTCTGGCGGTCTCAAGCACTACC TACGTC
<i>FANCF</i> site 4	AGCGATC CAGGTGC TGCAGA	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGAC GTAGGTAGTGCT TGAGACCGCC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA GTTGCCCAGAGTCAAG GAACACGG	GACGTAGGTAGTGCTTGAGACCGCCAGA AGCTCGGAAA AGCGATCCAGGTGCTGC AGAAGG GATTCCATGAGGTGCGCGAAGG CCCTACTTCCGCTTTCACCTTGGAGACGG CGACTCTCTGCGTACTGATTGGAACATCC GCGAAATGATACGCCTCTCTGCAATGCTA TTGGTCGAAATGCATGTCAATCTCCCAGC GTCTTTATCCGTGTTCCCTTGACTCTGGGC AACT
HEK site 7	GGAACAC AAAGCAT AGACTG	CGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTACA AGACCTGGCTGA GCTAACTGTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAGCCCCATCTGTCAA ACTGTG	ACAAGACCTGGCTGAGCTAACTGTGACA GCATGTGGTAATTTTCCAGCCCGCTGGCC CTGTAAAGGAAACT GGAACACAAAGCA TAGACTGCGG GGCGGGCCAGCCTGAATA GCTGCAAACAAGTGCAGAATATCTGATG ATGTCATACGCACAGTTTGACAGATGGG GCTGG
HEK site 8	TGTGTTT CAGTTTC CTTTAC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTACA AGACCTGGCTGA GCTAACTGTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAGCCCCATCTGTCAA ACTGTG	CCAGCCCCATCTGTCAAACCTGTGCGTATG ACATCATCAGATATTCTGCACTTGTTTGC AGCTATTCAGGCTGGCCCCGCCCGCAGTC TATGCTTT TGTGTTCCAGTTTCCTTTACA GG GCCAGCGGGCTGGAAAATTACCACAT GCTGTCACAGTTAGCTCAGCCAGGTCTTG T
HEK site 9	TTGTTTG CAGCTAT TCAGGC	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTACA AGACCTGGCTGA GCTAACTGTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAGCCCCATCTGTCAA ACTGTG	CCAGCCCCATCTGTCAAACCTGTGCGTATG ACATCATCAGATATTCTGCACT TTGTTTGC AGCTATTCAGGCTGG CCCCGCCCGCAGT CTATGCTTTGTGTTCCAGTTTCCTTTACAG GGCCAGCGGGCTGGAAAATTACCACATG CTGTCACAGTTAGCTCAGCCAGGTCTTGT

Supplementary Table 3 (Continued)

<i>PPPIR1</i> 2C site 9	AAGTCGA GGGAGGG ATGGTA	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGC CAGGCAGATAG ACCAGACTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CTCTCTTCCGCATTGGA GTCGC	GGCCAGGCAGATAGACCAGACTGAGCTA TGGGAGCTGGCTCAGGTTTCAGGAGAGGG CAGGGCAGGGAAGGAGACAAAGTCCAGG ACCGGCTGGAGGGGCTCAACATCGGAAG AGGGGAAGTCGAGGGAGGGATGGTAA GGAGGACTGCATGGGTCAGCACAGGCTG CCAAAGCCAGGGCCAGTTAAAGCGACTC CAATGCGGAAGAGAGT
<i>PPPIR1</i> 2C site 10	GACACGT GGATTGT GCTGTC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGCC TGGCCTTGGGAG CTTCC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC GGCTCTGCCATGGATC TGCCG	CGGCTCTGCCATGGATCTGCCGTGAGTCT CAGCCAGCCACTTTCTTTCCGGGCCTCA GTTTCCTTATCTGTCAAATGGGAATGGTA TGGTGC GCAACTGTATGAGGGGTCATGC GTGACACGTGGATTGTGCTGTCAGGAG CTCGGGGGAGTACGTGCCCTGGAAAGC ACTAGAGCCATGCCAGCGTTCAGTGAGCT CTGTGAGGTGACAGGAAGCTCCCAAGGC CAGGC
<i>PPPIR1</i> 2C site 11	GTCATAC ACTGGGC TGGCCA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCA CACCTAGGAGGA TAAGAGGCACG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG GAAAGGCCTGTGATCT CCGTTTCC	CCACACCTAGGAGGATAAGAGGCACGAG GTAGGACTCACACCCCATGACCTGAAAT ACCCAGGATCATACTGGGCTGGCCA GGTTTCAGCCCAGTGACTTGCCCAGAGC TCTTCCCCCTGCTCCCCTCTGAGACTCAG GAGGCCAGACTAGCCCAGTTGTTTCCCA ACTGTACTCCATGCAGTCGTGTGTCCCAC TCAGGGTCTCTGAGTCAGCGCTAAGCACT GGGAAAACGGAGATCACAGGCCTTCC
<i>PPPIR1</i> 2C site 12	CAAAGTC CAGGACC GGCTGG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGC CAGGCAGATAG ACCAGACTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CTCTCTTCCGCATTGGA GTCGC	GGCCAGGCAGATAGACCAGACTGAGCTA TGGGAGCTGGCTCAGGTTTCAGGAGAGGG CAGGGCAGGGAAGGAGACAAAGTCCAG GACCGGCTGGAGGGCTCAACATCGGA AGAGGGGAAGTCGAGGGAGGGATGGTAA GGAGGACTGCATGGGTCAGCACAGGCTG CCAAAGCCAGGGCCAGTTAAAGCGACTC CAATGCGGAAGAGAGT
<i>PPPIR1</i> 2C site	GCATGGC TCTAGTG	AGG	ACACTCTTTCCC TACACGACGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC	GCCTGGCCTTGGGAGCTTCTGTACCTC ACAGAGCTCACTGAACGCTGGCATGGCT

Supplementary Table 3 (Continued)

13	CTTTCC		TTCCGATCTGCC TGGCCTTGGGAG CTTCC	GGCTCTGCCATGGATC TGCCG	CTAGTGCTTTCCAGGGGCACGTACTCCC CCGAGCTCCTGACAGCACAATCCACGTGT CACGCATGACCCCTCATAAGTTGCGCAC CATAACATTCCCATTGACAGATAAAGGAA ACTGAGGCCCGAAAGGAAAGTGGCTGG CTGAGACTCACGGCAGATCCATGGCAGA GCCG
<i>PPP1R1</i> 2C site 14	GGTCATA CACTGGG CTGGCC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCA CACCTAGGAGGA TAAGAGGCACG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG GAAAGGCCTGTGATCT CCGTTTTCC	CCACACCTAGGAGGATAAGAGGCACGAG GTAGGACTCACACCCCATGACCTGAAAT ACCCAGGGTCATACACTGGGCTGGCCA GGTTTCAGCCAGTGACTTGCCCAGAGC TCTTCCCCCTGCTCCCCTCTGAGACTCAG GAGGCCAGACTAGCCAGTTGTTTCCCA ACTGTACTCCATGCAGTCGTGTGTCCAC TCAGGGTCTCTGAGTCAGCGCTAAGCACT GGGAAAACGGAGATCACAGGCCTTTCC
<i>PPP1R1</i> 2C site 15	AAGGAGA CAAAGTC CAGGAC	CGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGC CAGGCAGATAG ACCAGACTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CTCTCTTCCGCATTGGA GTCGC	GGCCAGGCAGATAGACCAGACTGAGCTA TGGGAGCTGGCTCAGGTTTCAGGAGAGGG CAGGGCAGGGAAAGGAGACAAAGTCCAG GACCGGCTGGAGGGGCTCAACATCGGAA GAGGGGAAGTCGAGGGAGGGATGGTAAG GAGGACTGCATGGGTTCAGCACAGGCTGC CAAAGCCAGGGCCAGTTAAAGCGACTCC AATGCGGAAGAGAGT
<i>PPP1R1</i> 2C site 16	GATTGTG CTGTCAG GAGCTC	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGCC TGGCCTTGGGAG CTTCC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC GGCTCTGCCATGGATC TGCCG	CGGCTCTGCCATGGATCTGCCGTGAGTCT CAGCCAGCCACTTTCTTTCCGGGCCTCA GTTTCCTTATCTGTCAAATGGGAATGGTA TGGTGCGCAACTGTATGAGGGGTCATGC GTGACACGTGGATTGTGCTGTCAGGAG CTCGGGGGAGTACGTGCCCTTGAAAGC ACTAGAGCCATGCCAGCGTTCAGTGAGCT CTGTGAGGTGACAGGAAGCTCCCAAGGC CAGGC
<i>RNF2</i> site 2	ATGACTA AGATGAC TGCCAA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGT	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAACATACAGAAGTCA	TGTGCAGACAAACGGAACTCAACCATTA AGCAAAACATGGGAAGTCAGTTTATATG AGTTACAACGAACACCTCAGGTAATGAC

Supplementary Table 3 (Continued)

			GCAGACAAACG GAACTCAACC	GGAATGCTTGA	TAAGATGACTGCCAAGGGGCATATGAG ACGTGTAAACTGGGATACATTTTTAGAAC ATGAAAACCTTAAATAGAACATAGAGTAA TAATAAATTTATATTCAAGCATTCCTGAC TTCTGTATGTTGG
RNF2 site 3	TGAGTTA CAACGAA CACCTC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGT GCAGACAAACG GAACTCAACC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAACATACAGAAGTCA GGAATGCTTGA	TGTGCAGACAAACGGAACCTCAACCATTA AGCAAAACATGGGAACTCAGTTTATATG AGTTACAACGAACACCTCAGGTAATGA CTAAGATGACTGCCAAGGGGCATATGAG ACGTGTAAACTGGGATACATTTTTAGAAC ATGAAAACCTTAAATAGAACATAGAGTAA TAATAAATTTATATTCAAGCATTCCTGAC TTCTGTATGTTGG
CGBE_N G site 1	ACCATCT TTTGTA ACTCAG	AGT	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAT ACTTAAAGTAAA GCAGAAGGAAT AACAGTGC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG AAGTTTCATTCCACTAT AGAACTGGTTCTG	CATACTTAAAGTAAAGCAGAAGGAATAA CAGTGCCCCAAATTATTCACTACCCCTGT TCCTAAAGCCCACCATCTTTTGTACACT CAGAGTAAACAAAGCATAGACTGAGGG GTACAATCCTACTCTAGTCCCTTTCCTCAT GCAACAGCTATGAAATAGCTGGTTAGTA CTGACAGCAATTCTTGGCCACAAACAGA ACCAGTTCTATAGTGGAATGAACTTC
CGBE_N G site 2	CACTTCT CTTCCTG CCCTCT	GGT	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAA CTAAAGTCACTC TGCTGTGCTATT AGATAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC AGAAAGACAAGTTTCA CATGTTCTCACTC	CAACTAAAGTCACTCTGCTGTGCTATTAG ATACTAGATCTTACTCATTCTATCTAACT GTATTTTGCATACATTAACCATCCCCAC ATTATCCTCCACCTCCCCCACTTCTCTTCC TGCCCTCTGGTAACCATCATTCTATTCTT TATGTTTCATGAGTTCAACTGTTTTAATTT CAGCTCCACATATGAGTGAGAACATGT GAAACTTGTCTTTCTG
CGBE_N G site 3 (EMX1)	AGCTTCT GCCGTTT GTACTT	TGT	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCA GTGGCTGCTCTG GGGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC GATGTCTCCCCATTG GCCTG	CGATGTCTCCCCATTGGCCTGCTTCGTG GCAATGCGCCACCGTTGATGTGATGGG AGCCCTTCTTCTTCTGCTCGGACTCAGGC CCTTCCTCCTCCAGCTTCTGCCGTTTGT ACTTTGTCTCCTCCGTTCTGGAACCACACC TTCACCTGGGCCAGGGAGGGAGGGGCAC AGATGAGAACTCAGGAGGCCCCAGAG

Supplementary Table 3 (Continued)

					CAGCCACTGG
CGBE_N G site 4 (<i>RNF2</i>)	CGTCTCA TATGCCC CTTGGC	AGT	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGT GCAGACAAACG GAACTCAACC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAACATACAGAAGTCA GGAATGCTTGA	CCAACATACAGAAGTCAGGAATGCTTGA ATATAAATTTATTATTACTCTATGTTCTAT TTAAGTTTTTCATGTTCTAAAAATGTATCC CAGTTTACAC CGTCTCATATGCCCCCTTGG CAGT CATCTTAGTCATTACCTGAGGTGTT CGTTGTAACATCATATAAACTGAGTTCCCA TGTTTTGCTTAATGGTTGAGTTCCGTTTGT CTGCAC
CGBE_N G site 5	ATAGACT CCAGGAT AAGGTA	AGT	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGTT CCACCGCCTTGT TTACTGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCAGCCTCTTTCTTCGC CCTC	GTTCCACCGCCTTGTTTACTGCTCAGTCC CCCAGAATGGCAGGCACGTAGTAGGGCT CAGTAAATATTTGCTGAATGAAGTTATGA TCGAAAGAATACTAAGC ATAGACTCCAG GATAAGGTAAGT AATATAGAGTTCCACA GTGGGCAGGTCTATGCAAACCTACCCCA AAGTCCGAGGAAGCTGAGAGGGCGAAGA AAGAGGCTGAG
CGBE_N G site 6 (<i>PPP1R1</i> 2C)	CTCAACA TCGGAAG AGGGGA	AGT	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGC CAGGCAGATAG ACCAGACTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CTCTCTTCCGCATTGGA GTCGC	GGCCAGGCAGATAGACCAGACTGAGCTA TGGGAGCTGGCTCAGGTTTCAAGAGAGGG CAGGGCAGGGAAGGAGACAAAGTCCAGG ACCGGCTGGAGGGG CTCAACATCGGAA GAGGGGAAGT CGAGGGAGGGATGGTAA GGAGGACTGCATGGGTGAGCACAGGCTG CCAAAGCCAGGGCCAGTTAAAGCGACTC CAATGCGGAAGAGAGT
CGBE_V RQR site 1	TCAATCC TTGGGGC CCAGAC	TGAG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGCC CATGCAATTAGT CTATTTCTGCTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTT GCAGGAGCTGCACATA CTAGCC	GCCCATGCAATTAGTCTATTTCTGCTGCA AGTAAGCATGCATTTGTAGGCTTGATGCT TTTTTTCTGCTTCTCCAGCCCTGGCCTGGG TCAATCCTTGGGGCCCAGACTTGAG CAC GTGATGGCAGAGGAAAGGAAGCCCTGCT TCCTCCAGAGGGCGTCGCAGGACAGCTTT TCCTAGACAGGGGCTAGTATGTGCAGCTC CTGC
CGBE_V RQR site	ATGTTCC AATCAGT	AGAG	ACACTCTTTCCC TACACGACGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA	AGTTGCCCAGAGTCAAGGAACACGGATA AAGACGCTGGGAGATTGACATGCATTTC

Supplementary Table 3 (Continued)

2 (<i>FANCF</i>)	ACGCAG		TTCCGATCTGAC GTAGGTAGTGCT TGAGACCGCC	GTTGCCCAGAGTCAAG GAACACGG	GACCAATAGCATTGCAGAGAGGCGTATC ATTTTCGCGGATGTTCCAATCAGTACGCA GAGAGTCGCCGTCTCCAAGGTGAAAGCG GAAGTAGGGCCTTCGCGCACCTCATGGA ATCCCTTCTGCAGCACCTGGATCGCTTTT CCGAGCTTCTGGCGGTCTCAAGCACTACC TACGTC
CGBE_V RQR site 3 (<i>RNF2</i>)	GATGACT GCCAAGG GGCATA	TGAG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGT GCAGACAAACG GAACTCAACC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CAACATACAGAAGTCA GGAATGCTTGA	GTGCAGACAAACGGAACCTCAACCATTAA GCAAAACATGGGAACTCAGTTTATATGA GTTACAACGAACACCTCAGGTAATGACT AAGATGACTGCCAAGGGGCATATGAGA CGTGTAAACTGGGATACATTTTTAGAAC TGAAAACCTAAATAGAACATAGAGTAAT AATAAATTTATATTCAAGCATTCTGACT TCTGTATGTTGG
CGBE_V RQR site 4	AAGTACA AGCACTC AATGTG	TGAG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTAGT GTTGGGATTACA GCCTGA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA ACCTGAAGCCTTTCCC CAA	AGTGTTGGGATTACAGCCTGAGCCACCGT GCCCCGCCAAAGATGGAGTATTCTCTAAC CTGTGTGACACTTGGCATAAGGAAAGTA CAAGCACTCAATGTGTGAGTTTCATTCC CACAGTCTATGGTCTTTAGCCCAGTCTCT CCAGATGCCTATTGAGTTCTTCAAGCATC CAAATGGTTTGGGGAAAGGCTTCAGGTT
CGBE_V RQR site 5	ACACACA CTTAGAA TCTGTG	GGAG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTATC TCAGCGCTTTCG TCCAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TCATTTCCCCACTCCCT CC	TCAGCGCTTTCGTCCACCACCCCTCTACA CACACACAGACACACACACTTAGAAT CTGTGGGAGTGGAGCCCAGCAATCTGTG TTTGCACAAACCCTCCAGGGAATTCTGAT GCCCCGTGAAGTTTGAGAACTAGAGGTA AACTATACCTTTCTGCAGGCGAGAACCTG GACCAAGGAGGGAGTGGGGAA
CGBE_V RQR site 6 (<i>VEGFA</i>)	GCGGACA GTGGACG CGGCGG	CGAG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCGC GCTGACGGACAG ACAG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG CTCCTCCGAAGCGAGA ACAG	CGCGCTGACGGACAGACAGACAGACACC GCCCCAGCCCCAGCTACCACCTCCTCCC CGGCCGCGCGCGGACAGTGGACGCGGG GGCGAGCCGCGGGCAGGGGCGGAGCC CGCGCCCGGAGGCGGGGTGGAGGGGGTC GGGGCTCGCGGCGTCGCACTGAAACTTTT CGTCCAACCTTCTGGGCTGTTCTCGCTTCG

Supplementary Table 3 (Continued)

					GAGGAGC
HEK site 2 off 1	GAACACA ATGCATA GATTGC	CGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGTG TGGAGAGTGAGT AAGCCA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CGGTAGGATGATTTC GGCA	GTGTGGAGAGTGAGTAAGCCA GAACACA ATGCATAGATTGCCGG TAAATAGGTTTA GATTCATCCATTTTTAAAAAATGGTGTGG GAGCATTAAATATGTATATAGTAGATATG GAAAAATGATTCTCATAATAACTGACATT TCTGTTTCACAAGAAAATTATTTTACATT ATATGTATATTTTACATAAATTATACATA GTCATTTAAAAAGCTCAAATAGTGCAAA AACAATATGGAGAATTGCCTGAAATCAT CCTACCGT
HEK site 2 off 2	AAACATA AAGCATA GACTGC	AAA	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAC AAAGCAGTGTAG CTCAGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTT TTTTGGTACTCGAGTGT TATTCAG	CACAAAGCAGTGTAGCTCAGGGAAGGAG CAGTGAGTTTGGGCACTTGTGACAGAATA GTGGGACTATGCCAGAGATACACAGGAG GAGGTGGTACCTTCTAGCTCCCCCTCAA AA ACATAAAGCATAGACTGCAAA GTACTCC CAAGCAGGCTGAATAACACTCGAGTACC AAAAA
HEK site 3 off 1	CACCCAG ACTGAGC ACGTGC	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTCC CCTGTTGACCTG GAGAA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC ACTGTACTTGCCCTGA CCA	TCCCCTGTTGACCTGGAGAAGCATGAACC AGTCAAAAAGTTTAAAGACAAGAGCATT AACTGCACCAGTGGGCAGCTCAGCTCAG ACACCAGTAGCGTGGG CACCCAGACTGA GCACGTGCTGG AGCCCAAGAAATGCAG AGACCTGTGCACCTCTGGTCAGGGCAAGT ACAGTG
HEK site 3 off 2	GACACAG ACTGGGC ACGTGA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTTG GTGTTGACAGGG AGCAA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TGAGATGTGGGCAGAA GGG	TTGGTGTGACAGGGAGCAACTTCACAGT CCCAGGCATCAG GACACAGACTGGGCA CGTGAGGG AAGCCCAAGGGAGAGGACT GGTGTAAATCGAGGCTGACTCCACTTTTAA TGTTTGACTGATGATAGGTTTCAAGTCTC ACTAAGTCTCCTTCCCCTTCTGCCACAT CTCAG
HEK site 3 off 3	AGCTCAG ACTGAGC AAGTGA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTGA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG TCCAAAGGCCCAAGAA	TGAGAGGGGAACAGAAGGGCTAAGACTAA AAGGAACAGAGGAGTTCATAGTGAGCGG TAAAG AGCTCAGACTGAGCAAGTGAGG

Supplementary Table 3 (Continued)

			GAGGGAACAGA AGGGCT	CCT	GGCTCAGCCTCCCATGGAGGACAGGGGG CTGGGGCCCCTGGCTGATGTCTGGACTGA AGCCCCACGCCAGAGGTTCTTGGGCCT TTGGAC
HEK site 3 off 4	AGACCAG ACTGAGC AAGAGA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTCC TAGCACTTTGGA AGGTGC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG CTCATCTTAATCTGCTC AGCC	TCCTAGCACTTTGGAAGGTCGAAGCGGC AGGATGGCTTCAACCCAGGAGTTCGAGA CCAGACTGAGCAAGAGAGGGAGAGTGT CTGTATTAACAACAAACAAACAAACAAA AAACTAAACTAAAAGAACTGTGGTGTA TAATATAAAATTCTGGCTGAGCAGATTAA GATGAGC
HEK site 3 off 5	GAGCCAG AATGAGC ACGTGA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTAAA GGAGCAGCTCTT CCTGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG TCTGCACCATCTCCCA CAA	AAAGGAGCAGCTCTTCCTGGTGGAAATT GCGAGCAGAGGCTGCGTGAGTTCCGTAA CTCGCACACAGCCTCCATTTGGAGCCAG AATGAGCACGTGAGGGACCCCGGGCAG AGGGGCCAGTGCTGACATTATGCTCCATG CAACCTCCCATCCTGTTGTGGGAGATGGT GCAGAC
HEK site 4 off 1	TGCACTG CGGCCGG AGGAGG	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGC ATGGCTTCTGAG ACTCA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG TCTCCCTTGCACTCCCT GTCTTT	GGCATGGCTTCTGAGACTCATAGCTGGGG CTGAAGATCCCTAGGGGGGCTCTGCTGG GCTCACTGCTCTCCAGAGTGGTCCAGCCC GGCTGCAGGGTGCTGCTTCCAGCTTGGTG CACTGCGGCCGGAGGAGGTGGAGGAT GGAAAGTAAGATTCAAAGACAGGGAGTG CAAGGG
HEK site 4 off 2	GGCTCTG CGGCTGG AGGGGG	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTTT GGCAATGGAGG CATTGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG AAGAGGCTGCCCATGA GAG	TTTGGCAATGGAGGCATTGGGCAGGGGA AGCCTGTCTTCAGGGCACATGCACGTGCG CAGGGCTCTGCGGCTGGAGGGGGTGG GGTTGCTGTTAGTGACAGGGGCCCCAGCC AGGCAGGTTTCAGGATTGGGGAGCACTT GCTTCGGCTCCCTTGCTCTCATGGGCAGC CTCTTC
HEK site 4 off 3	GGCACGA CGGCTGG AGGTGG	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGT CTGAGGCTCGAA	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC TGTGGCCTCCATATCC CTG	GGTCTGAGGCTCGAATCCTGGCAGCAGG TCCTTCATGGCAAGGCGGGAAAAGAGAA AAGCCAACGGGTTCTCATGCTGGGAAAA GATGCCGGGCACGACGGCTGGAGGTGG

Supplementary Table 3 (Continued)

			TCCTG		GGGGTTGGGAGTGGGTGGGATGCTTGCG TGCCCTGCATGAGGTGCAGGGATATGGA GGCCACAG
HEK site 4 off 4	GGCATCA CGGCTGG AGGTGG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTTTC CACCAGAACTCA GCCC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CTCGGTTCTCCACAA CAC	TTTCCACCAGAACTCAGCCCAGGCTGCTG TGGGATGGAATCACCTGCACCCGGATGTT CTTTCTGGGCTGGTACATACAGGCAAGG CATCACGGCTGGAGGTGGAGGGGCCT AACCCGGGGTTGCCAGGAAGGGGTTTG CACATGGATTCTGGTGTGTGGAGGAAC CGAGG
HEK site 4 off 5	GGCGCTG CGGCGGG AGGTGG	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAC GGGAAGGACAG GAGAAG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG CAGGGGAGGGATAAA GCAG	CACGGGAAGGACAGGAGAAGGTGCTGGA CCGCCTGGACTTTGTGCTGACCAGCCTTG TGGCGCTGCGGCGGGAGGTGGAGGAG CTGAGAAGCAGCCTGCGAGGGCTTGCGG GGGAGATTGTTGGGGAGGTCCGGTGAGT AATGCGGCTTCTTCTCCTGCTTTATCCCTC CCCTGC
EMXI site 1 off 1	GAGTCTA AGCAGAA GAAGAA	GAG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAC CTGGGCGAGAA AGGTAACCTATG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC AGTGTTGGCTTTTACA AGGATGC	CACCTGGGCGAGAAAGGTAACCTTATGTTT CAGTAGCCTCTTTCTCAATGTGCTTCAAC CCATCACGGCCTTTGCAAATAGAGCCCTT TATTCATAGTAGACAAAGAGTCTAAGCAG AAGAAGAAGAGAGCCACTACCCAACCAT CTACTCTTCTAATGGTGTGTTTCCTACAAA GGCCAAGTCATGAGACTGCATCCTTGTGA AAGCCAACACTG
EMXI site 1 off 2	GAGGCCG AGCAGAA GAAAGA	CGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTAGA AGGAGGTGCAG GAGCTAGAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CCTCAGCCACTTTATT CATCCCC	AGAAGGAGGTGCAGGAGCTAGACTCCGA GGGGAGGCTGCGAGCCGCAAGCGCAGGA GCCGGGTGGGAGAGAGACCCCTTCTTCTG CAAATGAGGAGGCCGAGCAGAAGAAAG ACGGCGACAGATGTTGGGGGGAGGGGAC GGTTTGTGAGGGATAGGGAGAGAAAGTC TAAGTGAGAGCAGGACGAGGAGGGGAG GGTGGTTGGGGAAGGGGATGAAATAAAG TGGCTGAGGG
EMXI site 1 off	GAGTCCT AGCAGGA	GAG	ACACTCTTTCCC TACACGACGCTC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC	TATAATAATGTCCCTGCCCCGGGAGCTGTT CAGAGGCTGAAGAGGAAGACCAGACTCA

Supplementary Table 3 (Continued)

3	GAAGAA		TTCCGATCTTAT AATAATGTCCCT GCCCCGGGAGC	TTCTTTAGGGCCCAGC TGTGC	GTAAAGCCTGGAGGCTGCCAGGTAGGGC TGGGGCCAGCATGACCTGAGTCCTAGCA GGAGAAGAAGAGGCAGCCTAGAGTCTTC TGTGAAGTGCACATAGAAGAGAGACTGG GGCCAAGCCACAAAAGATAGAATGCACA GCTGGGCCCTAAAGAAG
EMXI site 1 off 4	GAGTCCG GGAAGGA GAAGAA	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGGT AGGAATAAAGG CGAGGAAGCG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA GCCCAAGAGGGCCAAG TCC	AGCCCAAGAGGGCCAAGTCCTGGCTGTC TGCCTCTGACGACGAGCAAGGTGGAGGC CCTTGTTAGCAGGATGGGTGGTGAGGA GTCCGGGAAGGAGAAGAAAGGCTCAGC GCGGCTTGCCTGAGCCTCCCTCCTCCCAG CTCCCGGCCCTGCTGCCGGCGGCTGTCA CTCCTCGCTGTGGTGCCAGCTGGCTCCCG CTTCCTCGCCTTTATTCTCTACC
EMXI site 1 off 5	GAGCCGG AGCAGAA GAAGGA	GGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTGAG CGAGCGAGGAG GGAGG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CCATTCACTCCACCTG ATCTCGG	ACCATTCACTCCACCTGATCTCGGGGCGC TGTGCGCTGAGGAAGGCGCGGGCGAGCC GGAGCAGAAGAAGGAGGAGCC AGCCGCTGCAGCCACCACCGCCACCATGT CCTACCAAGGCAAGAAGAACATCCCGCG GATCACGGTGAGTCCGGGCGCCGCTGCTC CCTCCCTCCTCGCTCGCTC
FANCF site 1 off 1	GGAACCC CGTCTGC AGCACC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCGC CTGCTGCACTCT CTGAG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA CAGTTGGTGTCAATAC AACTCCCC	ACAGTTGGTGTCAATACAACCTCCCCTTAT TTTCTCTCAAGTCACCTGGATCGTCCTGA CCCCGGGAACCCCGTCTGCAGCACCAG GCCCCCTCCGTGGAGAAAAGATGGAGCC GGATTAAGCACCCAGTGCTAAGGCGACT AAGACGCCACTGCCCCGAGGCCCTGCCG GAAAATACTCAGAGAGTGCAGCAGGCG
FANCF site 1 off 2	GGAGTCC CTCCTAC AGCACC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCT TACCAGATGGAG GACAGTGACC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA GACCTCTGACCTCCAC AACTGG	CCTTACCAGATGGAGGACAGTGACCCAG GTCCAGTGTTCCTCATCCCCAACACAGTG ACAGAAGGCAGCCAAGGAATCCTCATTC CTGTCCTGGAACCTACAAGAGTCCCTCCT ACAGCACCAGGTGTATTTCATCTTCTGTTG TTGCTATAACAAAATTACCACAACTTAG TGGCTTAAGTAACTACACATTTATTATTT TCCAGTTGTGGAGGTCAGAGGTCT

Supplementary Table 3 (Continued)

<i>FANCF</i> site 1 off 3	AGAGGCC CCTCTGC AGCACC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCC ACAGGTCCTTCT GGAAGGAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTA AATCACAAACCGGCAG GAGCC	CCCACAGGTCCTTCTGGAAGGACTCAGGC AGGAGTTAGGAGGCTCCCGGGGTCAGGC TTCTGGGTCTAGATTTCCAGAGGCCCT CTGCAGCACCAGCATTTCGCCTCTAGGA GTCATCGCTCTTCAGCGGATCCTGCAGCC CTTGCGGATGCTCAGAGTGAACGCGTTAC CCCGCCAGCCCCCTCTGCCGGCTCCTGC CGGTTTGTGATT
<i>FANCF</i> site 1 off 4	ACCATCC CTCCTGC AGCACC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCAG GCTAGGTTTAGA GCCAGTGAAC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC CATGTGCCTTATCCTAT TGTGAATGAG	CAGGCTAGGTTTAGAGCCAGTGAAGTAG AGTGGCATGCAACCTAGAGAGATACTCA CTGGCATGACTGAGAGATGGTTTGTACC ATCCCTCCTGCAGCACCAGCAGCAAA ACTCATAGCAACTAAAGTGACTCATTTCT CCTGCTTAAGGAAAAGAAATCAAAGTAT AAAGAGGATTTTGTCTTGCATCTTGATA CCATCTCATTACAATAGGATAAGGCACA TGG
<i>FANCF</i> site 1 off 5	GGATTGC CATCCGC AGCACC	TGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTTCA CGCCCGCCCTCC TTG	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTG GGACGGCCTTACCAGA TCAATG	GGGACGGCCTTACCAGATCAATGAACGT CAGGAATTTCCAGCTCCCTCCGTAGGTCT GGTGTGAGGGCATTTCGATGGCCTTGTA TTACACAGGATAGAGCAGTAAGACAGCA GGATTGCCATCCGCAGCACCCTGGCAGG GGACAAGCGCCATGTTCCGCCAGAGGCCT GGACCGGCTGGGCGGCAAGGAGGGCGGG CGTGA
<i>FANCF</i> site 1 off 6	TGAATCC CATCTCC AGCACC	AGG	ACACTCTTTCCC TACACGACGCTC TTCCGATCTCCA GGGTTTCAAGTC CTTTCAACACC	GACTGGAGTTCAGACG TGTGCTCTTCCGATCTC AGCCAGGCAGGTACTA TTATTATCTCC	CCAGGGTTTCAAGTCCTTTCAACACCTCT GCCCTCCGGGGGCCATTTCAGAAACTG GAGACCTCCTGGTTAAGAGCATGGGCC GTGGAGTCCAACCAACGGGGATGTGAAT CCCATCTCCAGCACCAGGTGCTGTGTA CTTGAGCAAACCTGCTTAACCTCTCTGAG TCTCAGCTTCTCATCTGTGAAATGGAGAT AATAATAGTACCTGCCTGGCTG

Supplementary Table 3 (Continued)

All pegRNAs were of the form 5’-

NNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCNNNNNNNNNNNNNNNNNNNNNTTTTTTT-3’ (from BsaI digest of pU6-pegRNA-GG-acceptor, Addgene #132777).

HEK_site 3_CTTins, FANCF_site 1_GtoT, and FANCF_site 1_+21 were described in Anzalone et al, Nature 2019 - PMID 31634902. All other pegRNAs were designed based on Anzalone et al for C-to-G substitution for the same C as edited by CGBEs.

target	spacer	PAM	3' extension sequence	Primer Binding Sequence	RT Template (prime edit in red)	PE3 nicking guide RNA	Fwd PCR1	Rev PCR1	Amplicon
ABE_site 7_CtoG	CTATA TTACT TACCT TATCC	TGG	GAATAGT AAGCAT AGACTCC AGGATA AGGTAA GTAATAT	ATATTA CTTACC TTA	TCCTGG AGTCTA TGCTTA CTATTC	AGAAT GGCAG GCACG TAGTA	ACACTCT TTCCCTA CACGAC GCTCTTC CGATCTG TTCCACC GCCTTGT TTACTGC TC	GACTGGA GTTTCTA CGTGTGC TCTTCCG ATCTCTC AGCCTCT TTCTTCGC CCTC	GTTCCACCGCCTTGT TACTGCTCAGTCCCC AGAATGGCAGGCACG TAGTAGGGCTCAGTAA ATATTTGCTGAATGAA GTTATGATCGAAAGAA TA ^C TAAGCATAGACTC CAGGATAAGGTAAGT AATATAGAGTTCCACA GTGGGCAGGTCTATGC AAACCTACCCCAAAGT CCGAGGAAGCTGAGA GGGCGAAGAAAGAGG CTGAG
ABE_site 8_CtoG	ATGA GGAA AGGG ACTA GAGT	AGG	GTAAAG AAAGCA TAGACTG AGGGGT ACAATCC TACTCTA GTCCCTT	GAGGAA AGGGAC TAG	AGTAGG ATTGTA CCCCTC AGTCTA TGCTTT CTTTAC	TGCAA CAGCT ATGAA ATAGC	ACACTCT TTCCCTA CACGAC GCTCTTC CGATCTC ATACTTA AAGTAA	GACTGGA GTTTCTA CGTGTGC TCTTCCG ATCTGAA GTTTCATT CCACTAT	CATACTTAAAGTAAAG CAGAAGGAATAACAG TGCCCCAAATTATTCA CTACCCCTGTTCTTAA AGCCCACCATCTTTTG TACACTCAGAGTAAAC ^C AAAGCATAGACTGAG

Supplementary Table 3 (Continued)

			TCCTC				AGCAGA AGGAAT AACAGT GC	AGAACTG GTTCTG	GGGTACAATCCTACTC TAGTCCCTTTCCTCAT GCAACAGCTATGAAAT AGCTGGTTAGTACTGA CAGCAATTCTTGGCCA CAAACAGAACCAGTTC TATAGTGGAATGAAAC TTC
HEK_site 2_CtoG	GCTG GCCC TGTA AGGA AAC	TGG	GCTTTCT GTTCCAG TTTCCTT TACAGG GCCA	TGGCCC TGTA GGA	AACTGG AACAGA AAGC	TTGTT TGCAG CTATT CAGGC	ACACTCT TTCCCTA CACGAC GCTCTTC CGATCTA CAAGAC CTGGCTG AGCTAAC TGTG	GA CTGGA GTT CAGA CGTGTGC TCTTCCG ATCTCCA GCCCCAT CTGTCAA ACTGTG	ACAAGACCTGGCTGA GCTAACTGTGACAGCA TGTGGTAATTTTCCAG CCCGCTGGCCCTGTAA AGGAACTGGAACAC AAAGCATAGACTGCG GGGCGGGCCAGCCTG AATAGCTGCAAACAA GTGCAGAATATCTGAT GATGTCATACGCACAG TTTGACAGATGGGGCT GG
RNF2_site 1_CtoG	TGAG TTACA ACGA ACAC CTC	AGG	GTCATGT TAGTCAT TACCTGA GGTGTTC GTTGTAA CT	AGTTAC AACGAA CAC	CTCAGG TAATGA CTAACA TGAC	TACAC GTCTC ATATG CCCCT	ACACTCT TTCCCTA CACGAC GCTCTTC CGATCTT GTGCAG ACAAAC GGAAGTC AACC	GA CTGGA GTT CAGA CGTGTGC TCTTCCG ATCTCCA ACATACA GAAGTCA GGAATGC TTGA	CCAACATACAGAAGTC AGGAATGCTTGAATAT AAATTTATTATTACTC TATGTTCTATTTAAGT TTTCATGTTCTAAAAA TGTATCCCAGTTTACA CGTCTCATATGCCCCT TGGCAGTCATCTTAGT CATTACCTGAGGTGTT CGTTGTAACATATA AACTGAGTTCCCATGT TTTGCTTAATGGTTGA GTTCCGTTTGTCTGCA C
HEK_site 3_CTTins	GGCC CAGA	TGG	TCTGCCA TCAAAGC	CAGACT GAGCAC	CTTTGA TGGCAG	GTCAA CCAGT	ACACTCT TTCCCTA	GA CTGGA GTT CAGA	GCCCATGCAATTAGTC TATTTCTGCTGCAAGT

Supplementary Table 3 (Continued)

	CTGA GCAC GTGA		GTGCTCA GTCTG	G	A	ATCCC GGTGC	CACGAC GCTCTTC CGATCTG CCCATGC AATTAGT CTATTTT TGCTG	CGTGTGC TCTTCCG ATCTTGC AGGAGCT GCACATA CTAGCC	AAGCATGCATTTGTAG GCTTGATGCTTTTTTTC TGCTTCTCCAGCCCTG GCCTGGGTCAATCCTT GGGGCCCAGACTGAG CACGTGATGGCAGAG GAAAGGAAGCCCTGC TTCCTCCAGAGGGCGT CGCAGGACAGCTTTTC CTAGACAGGGGCTAGT ATGTGCAGTCTCTGC
<i>FANCF</i> _site 1_GtoT	GGAA TCCCT TCTGC AGCA CC	TGG	GGAAAA GCGATCA AGGTGCT GCAGAA GGGA	TCCCTT CTGCAG C	ACCTTG ATCGCT TTTCC	GAAGC TCGGA AAAGC GATCA	ACACTCT TTCCCTA CACGAC GCTCTTC CGATCTG ACGTAG GTAGTGC TTGAGAC CGCC	GACTGGA GTTTCTG CGTGTGC TCTTCCG ATCTAGT TGCCAG AGTCAAG GAACACG G	AGTTGCCCAGAGTCAA GGAACACGGATAAAG ACGCTGGGAGATTGAC ATGCATTTTCGACCAAT AGCATTGCAGAGAGG CGTATCATTTCGCGGA TGTTCCAATCAGTACG CAGAGAGTCGCCGTCT CCAAGGTGAAAGCGG AAGTAGGGCCCTTCGCG CACCTCATGGAATCCC TTCTGCAGCACCTGGA TCGCTTTTCCGAGCTT CTGGCGGTCTCAAGCA CTACCTACGTC
<i>FANCF</i> _site 1_+21	-	-	-	-	-	GCTTG AGACC GCCAG AAGCT	ACACTCT TTCCCTA CACGAC GCTCTTC CGATCTT GTGCAG ACAAAC GGAATC AACC	GACTGGA GTTTCTG CGTGTGC TCTTCCG ATCTCCA ACATACA GAAGTCA GGAATGC TTGA	CCAACATACAGAAGTC AGGAATGCTTGAATAT AAATTTATTATTACTC TATGTTCTATTAAAGT TTTCATGTTCTAAAAA TGTATCCCAGTTTACA CGTCTCATATGCCCT TGGCAGTCATCTTAGT CATTACCTGAGGTGTT CGTTGTAACATCATATA

Supplementary Table 3 (Continued)

									AACTGAGTTCCCATGT TTTGCTTAATGGTTGA GTTCCGTTTGTCTGCA C
--	--	--	--	--	--	--	--	--	---

References

- Adey, A., Morrison, H. G., (no last name), A., Xun, X., Kitzman, J. O., Turner, E. H., Stackhouse, B., MacKenzie, A. P., Caruccio, N. C., Zhang, X., & Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology*, 11(12), R119. <https://doi.org/10.1186/gb-2010-11-12-r119>
- Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-04252-2>
- Alseth, I., Dalhus, B., & Bjørås, M. (2014). Inosine in DNA and RNA. *Current Opinion in Genetics & Development*, 26, 116–123. <https://doi.org/10.1016/J.GDE.2014.07.008>
- Anzalone, A., Gao, X., Podracky, C., Nelson, A., Koblan, L., Raguram, A., Levy, J., Mercer, J., & Liu, D. (2021). Programmable large DNA deletion, replacement, integration, and inversion with twin prime editing and site-specific recombinases. *BioRxiv*, 2021.11.01.466790. <https://doi.org/10.1101/2021.11.01.466790>
- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576(7785), 149–157. <https://doi.org/10.1038/s41586-019-1711-4>
- Arbab, M., Shen, M. W., Mok, B., Wilson, C., Matuszek, Ż., Cassa, C. A., & Liu, D. R. (2020). Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell*, 182(2), 463–480. <http://www.cell.com/article/S0092867420306322/fulltext>

- Arnould, S., Delenda, C., Grizot, S., Desseaux, C., Pâques, F., Silva, G. H., & Smith, J. (2011). The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. *Protein Engineering, Design and Selection*, 24(1–2), 27–31. <https://doi.org/10.1093/PROTEIN/GZQ083>
- Atkins, A., Chung, C.-H., Allen, A. G., Dampier, W., Gurrola, T. E., Sariyer, I. K., Nonnemacher, M. R., & Wigdahl, B. (2021). Off-Target Analysis in Gene Editing and Applications for Clinical Translation of CRISPR/Cas9 in HIV-1 Therapy. *Frontiers in Genome Editing*, 0, 16. <https://doi.org/10.3389/FGED.2021.673022>
- Auerbach, C. (1945). The Induction by Mustard Gas of Chromosomal Instabilities in *Drosophila melanogaster*. *Proceedings of the Royal Society of Edinburgh, Section B: Biological Sciences*, 62(3), 307–320. <https://doi.org/10.1017/S0080455X00130097>
- Ba, X., Aguilera-Aguirre, L., Rashid, Q. T. A. N., Bacsı, A., Radak, Z., Sur, S., Hosoki, K., Hegde, M. L., & Boldogh, I. (2014). The Role of 8-Oxoguanine DNA Glycosylase-1 in Inflammation. *International Journal of Molecular Sciences*, 15(9), 16975–16997. <https://doi.org/10.3390/IJMS150916975>
- Bao, X. R., Pan, Y., Lee, C. M., Davis, T. H., & Bao, G. (2020). Tools for experimental and computational analyses of off-target editing by programmable nucleases. *Nature Protocols*, 16(1), 10–26. <https://doi.org/10.1038/S41596-020-00431-Y>
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709–1712. <https://doi.org/10.1126/SCIENCE.1138140>

- Bellamy, S. R. W., & Baldwin, G. S. (2001). A kinetic analysis of substrate recognition by uracil-DNA glycosylase from herpes simplex virus type 1. *Nucleic Acids Research*, 29(18), 3857–3863. <https://doi.org/10.1093/NAR/29.18.3857>
- Bibikova, M., Golic, M., Golic, K. G., & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, 161(3), 1169. [/pmc/articles/PMC1462166/?report=abstract](https://pmc/articles/PMC1462166/?report=abstract)
- Boch, J., & Bonas, U. (2010). Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. *Annual Review of Phytopathology*, 48, 419–436. <https://doi.org/10.1146/ANNUREV-PHYTO-080508-081936>
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959), 1509–1512. <https://doi.org/10.1126/SCIENCE.1178811>
- Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151(8), 2551–2561. <https://doi.org/10.1099/MIC.0.28048-0>
- Bonas, U., Stall, R. E., & Staskawicz, B. (1989). Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Molecular and General Genetics MGG* 1989 218:1, 218(1), 127–136. <https://doi.org/10.1007/BF00330575>
- Brinkman, E. K., Chen, T., Haas, M. de, Holland, H. A., Akhtar, W., & Steensel, B. van. (2018). Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Molecular Cell*, 70(5), 801-813.e6. <https://doi.org/10.1016/J.MOLCEL.2018.04.016>

- Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slikhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960–964. <https://doi.org/10.1126/SCIENCE.1159689>
- Bucardo, M. S., Wu, Y., Paul T. Ludford, I., Li, Y., Fin, A., & Tor, Y. (2021). Real-Time Monitoring of Human Guanine Deaminase Activity by an Emissive Guanine Analog. *ACS Chemical Biology*, *16*(7), 1208–1214. <https://doi.org/10.1021/ACSCHEMBIO.1C00232>
- Bunting, S. F., Callén, E., Wong, N., Chen, H. T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., Xu, X., Deng, C. X., Finkel, T., Nussenzweig, M., Stark, J. M., & Nussenzweig, A. (2010). 53BP1 Inhibits Homologous Recombination in Brca1-Deficient Cells by Blocking Resection of DNA Breaks. *Cell*, *141*(2), 243–254. <https://doi.org/10.1016/J.CELL.2010.03.012>
- Capecchi, M. R. (1980). High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell*, *22*(2), 479–488. [https://doi.org/10.1016/0092-8674\(80\)90358-X](https://doi.org/10.1016/0092-8674(80)90358-X)
- Capecchi, M. R. (1989). Altering the genome by homologous recombination. *Science*, *244*(4910), 1288–1292. <https://doi.org/10.1126/SCIENCE.2660260>
- Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature Reviews Genetics* *2005* 6:6, *6*(6), 507–512. <https://doi.org/10.1038/NRG1619>
- Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia,

- N. V., Bogdanove, A. J., & Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, 39(12), e82–e82. <https://doi.org/10.1093/NAR/GKR218>
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology* 2017 18:8, 18(8), 495–506. <https://doi.org/10.1038/nrm.2017.48>
- Chapman, J. R., Taylor, M. R. G., & Boulton, S. J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell*, 47(4), 497–510. <https://doi.org/10.1016/J.MOLCEL.2012.07.029>
- Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*, 58(5), 235–263. <https://doi.org/10.1002/EM.22087>
- Chaudhari, H. G., Penterman, J., Whitton, H. J., Spencer, S. J., Flanagan, N., Zhang, M. C. L., Huang, E., Khedkar, A. S., Toomey, J. M., Shearer, C. A., Needham, A. W., Ho, T. W., Kulman, J. D., Cradick, T. J., & Kernysky, A. (2020). Evaluation of Homology-Independent CRISPR-Cas9 Off-Target Assessment Methods. *The CRISPR Journal*, 3(6), 440–453. <https://doi.org/10.1089/CRISPR.2020.0053>
- Chen, C.-C., Feng, W., Lim, P. X., Kass, E. M., & Jasin, M. (2018). Homology-Directed Repair and the Role of BRCA1, BRCA2, and Related Proteins in Genome Integrity and Cancer. <https://doi.org/10.1146/Annurev-Cancerbio-030617-050502>, 2, 313–336. <https://doi.org/10.1146/ANNUREV-CANCERBIO-030617-050502>

- Chen, F., Ding, X., Feng, Y., Seebeck, T., Jiang, Y., & Davis, G. D. (2017). Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nature Communications*, 8(1), 1–12. <https://doi.org/10.1038/ncomms14958>
- Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A., & Doudna, J. A. (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*, 550(7676), 407–410. <https://doi.org/10.1038/nature24268>
- Chen, L., Park, J. E., Paa, P., Rajakumar, P. D., Prekop, H.-T., Chew, Y. T., Manivannan, S. N., & Chew, W. L. (2021). Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nature Communications*, 12(1), 1–7. <https://doi.org/10.1038/S41467-021-21559-9>
- Chen, P. J., Hussmann, J. A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.-F., Chen, C., Nelson, J. W., Newby, G. A., Sahin, M., Osborn, M. J., Weissman, J. S., Adamson, B., & Liu, D. R. (2021). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*. <https://doi.org/10.1016/J.CELL.2021.09.018>
- Chen, X., Liu, J., Janssen, J. M., & Gonçalves, M. A. F. V. (2017). The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. *Molecular Therapy - Nucleic Acids*, 8, 558–563. <https://doi.org/10.1016/J.OMTN.2017.08.005>
- Choi, J., Chen, W., Suiter, C. C., Lee, C., Chardon, F. M., Yang, W., Leith, A., Daza, R. M., Martin, B., & Shendure, J. (2021). Precise genomic deletions using paired prime editing. *Nature Biotechnology*, 1–9. <https://doi.org/10.1038/S41587-021-01025-Z>

- Choo, Y., & Klug, A. (1994). Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proceedings of the National Academy of Sciences*, 91(23), 11163–11167. <https://doi.org/10.1073/PNAS.91.23.11163>
- Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., & Voytas, D. F. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*, 186(2), 757–761. <https://doi.org/10.1534/GENETICS.110.120717>
- Chu, S. H., Packer, M., Rees, H., Lam, D., Yu, Y., Marshall, J., Cheng, L.-I., Lam, D., Olins, J., Ran, F. A., Liquori, A., Gantzer, B., Decker, J., Born, D., Barrera, L., Hartigan, A., Gaudelli, N., Ciaramella, G., & Slaymaker, I. M. (2021). Rationally Designed Base Editors for Precise Editing of the Sickle Cell Disease Mutation. *The CRISPR Journal*, 4(2), 169–177. <https://doi.org/10.1089/CRISPR.2020.0144>
- Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research*, 5, 1–15. <https://doi.org/10.1093/nar/gku241>
- 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. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, 37(3), 224–226. <https://doi.org/10.1038/s41587-019-0032-3>
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, 339(6121), 819–823. <https://doi.org/10.1126/science.1231143>

- Cooper, G. M. (2000). DNA Repair. In *The Cell: A Molecular Approach* (2nd ed.). Sinauer Associates. <https://www.ncbi.nlm.nih.gov/books/NBK9900/>
- Cultrone, A., Scazzocchio, C., Rochet, M., Montero-Morán, G., Drevet, C., & Fernández-Martín, R. (2005). Convergent evolution of hydroxylation mechanisms in the fungal kingdom: molybdenum cofactor-independent hydroxylation of xanthine via α -ketoglutarate-dependent dioxygenases. *Molecular Microbiology*, 57(1), 276–290. <https://doi.org/10.1111/J.1365-2958.2005.04686.X>
- Dahncke, K., & Witte, C.-P. (2013). Plant Purine Nucleoside Catabolism Employs a Guanosine Deaminase Required for the Generation of Xanthosine in Arabidopsis. *The Plant Cell*, 25(10), 4101–4109. <http://www.ncbi.nlm.nih.gov/pubmed/24130159>
- Darwanto, A., Theruvathu, J. A., Sowers, J. L., Rogstad, D. K., Pascal, T., Goddard, W., & Sowers, L. C. (2009). Mechanisms of Base Selection by Human Single-stranded Selective Monofunctional Uracil-DNA Glycosylase *. *Journal of Biological Chemistry*, 284(23), 15835–15846. <https://doi.org/10.1074/JBC.M807846200>
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602–607. <https://doi.org/10.1038/nature09886>
- Densham, R. M., & Morris, J. R. (2019). Moving Mountains—The BRCA1 Promotion of DNA Resection. *Frontiers in Molecular Biosciences*, 0, 79. <https://doi.org/10.3389/FMOLB.2019.00079>

- Ding, X., Seebeck, T., Feng, Y., Jiang, Y., Davis, G. D., & Fuqiang, C. (2019). Improving CRISPR-Cas9 Genome Editing Efficiency by Fusion with Chromatin-Modulating Peptides. *The CRISPR Journal*, 2(1), 51–63. <https://doi.org/10.1089/CRISPR.2018.0036>
- Doman, J. L., Raguram, A., Newby, G. A., & Liu, D. R. (2020). Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nature Biotechnology*, 38(5), 620–628. <https://doi.org/10.1038/s41587-020-0414-6>
- Doyon, Y., Vo, T. D., Mendel, M. C., Greenberg, S. G., Wang, J., Xia, D. F., Miller, J. C., Urnov, F. D., Gregory, P. D., & Holmes, M. C. (2010). Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature Methods*, 8(1), 74–79. <https://doi.org/10.1038/nmeth.1539>
- El-Brolosy, M. A., & Stainier, D. Y. R. (2017). Genetic compensation: A phenomenon in search of mechanisms. *PLOS Genetics*, 13(7), e1006780. <https://doi.org/10.1371/JOURNAL.PGEN.1006780>
- Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science*, 351(6274). <https://doi.org/10.1126/SCIENCE.AAC7247>
- Fan, J., Otterlei, M., Wong, H., Tomkinson, A. E., & Wilson, D. M. (2004). XRCC1 co-localizes and physically interacts with PCNA. *Nucleic Acids Research*, 32(7), 2193–2201. <https://doi.org/10.1093/NAR/GKH556>
- Faucher, F., Doublié, S., & Jia, Z. (2012). 8-Oxoguanine DNA Glycosylases: One Lesion, Three Subfamilies. *International Journal of Molecular Sciences*, 13(6). <http://www.ncbi.nlm.nih.gov/pubmed/22837659>

- Folger, K. R., Wong, E. A., Wahl, G., & Capecchi, M. R. (1982). Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Molecular and Cellular Biology*, 2(11), 1372–1387. <https://doi.org/10.1128/MCB.2.11.1372-1387.1982>
- Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., Foell, J., Fuente, J. de la, Grupp, S., Handgretinger, R., Ho, T. W., Kattamis, A., Kernytsky, A., Lekstrom-Himes, J., Li, A. M., Locatelli, F., Mapara, M. Y., Montalembert, M. de, Rondelli, D., ... Corbacioglu, S. (2021). CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia. *The New England Journal of Medicine*, 384(3), 252–260. <https://doi.org/10.1056/NEJMOA2031054>
- Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, 32(3), 279–284. <https://doi.org/10.1038/nbt.2808>
- Gabriel, R., Lombardo, A., Arens, A., Miller, J. C., Genovese, P., Kaepfel, C., Nowrouzi, A., Bartholomae, C. C., Wang, J., Friedman, G., Holmes, M. C., Gregory, P. D., Glimm, H., Schmidt, M., Naldini, L., & Von Kalle, C. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature Biotechnology*, 29(9), 816–823. <https://doi.org/10.1038/nbt.1948>
- Gajula, K. S. (2019). Designing an Elusive C•G→G•C CRISPR Base Editor. *Trends in Biochemical Sciences*, 44(2), 91–94. <https://doi.org/10.1016/J.TIBS.2018.10.004>
- Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune

- system cleaves bacteriophage and plasmid DNA. *Nature* 2010 468:7320, 468(7320), 67–71.
<https://doi.org/10.1038/nature09523>
- Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, 551(7681), 464–471. <https://doi.org/10.1038/nature24644>
- Gaudelli, N. M., Lam, D. K., Rees, H. A., Solá-Esteves, N. M., Barrera, L. A., Born, D. A., Edwards, A., Gehrke, J. M., Lee, S.-J., Liquori, A. J., Murray, R., Packer, M. S., Rinaldi, C., Slaymaker, I. M., Yen, J., Young, L. E., & Ciaramella, G. (2020). Directed evolution of adenine base editors with increased activity and therapeutic application. *Nature Biotechnology*, 38(7), 892–900. <https://doi.org/10.1038/s41587-020-0491-6>
- Gedik, C. M., & Collins, A. (2005). Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study. *FASEB Journal*, 19(1), 82–84. <https://doi.org/10.1096/FJ.04-1767FJE>
- Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., & Joung, J. K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nature Biotechnology*, July. <https://doi.org/10.1038/nbt.4199>
- Giannoukos, G., Ciulla, D. M., Marco, E., Abdulkerim, H. S., Barrera, L. A., Bothmer, A., Dhanapal, V., Gloskowski, S. W., Jayaram, H., Maeder, M. L., Skor, M. N., Wang, T., Myer, V. E., & Wilson, C. J. (2018). UDiTaSTM, a genome editing detection method for indels and genome rearrangements. *BMC Genomics*, 19(1), 212.
<https://doi.org/10.1186/s12864-018-4561-9>

- Greisman, H. A., & Pabo, C. O. (1997). A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites. *Science*, 275(5300), 657–661.
<https://doi.org/10.1126/SCIENCE.275.5300.657>
- Grünewald, J., Zhou, R., Garcia, S. P., Iyer, S., Lareau, C. A., Aryee, M. J., & Joung, J. K. (2019). Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*, 569(7756), 433–437. <https://doi.org/10.1038/s41586-019-1161-z>
- Grünewald, J., Zhou, R., Iyer, S., Lareau, C. A., Garcia, S. P., Aryee, M. J., & Joung, J. K. (2019). CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nature Biotechnology*, 37(9), 1041–1048. <https://doi.org/10.1038/s41587-019-0236-6>
- Guilinger, J. P., Pattanayak, V., Reyon, D., Tsai, S. Q., Sander, J. D., Joung, J. K., & Liu, D. R. (2014). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature Methods*, 11(4), 429–435.
<https://doi.org/10.1038/NMETH.2845>
- Gunasekera, S., Abraham, S., Stegger, M., Pang, S., Wang, P., Sahibzada, S., & O’Dea, M. (2021). Evaluating coverage bias in next-generation sequencing of Escherichia coli. *PLOS ONE*, 16(6), e0253440. <https://doi.org/10.1371/JOURNAL.PONE.0253440>
- Guo, C., McDowell, I. C., Nodzenski, M., Scholtens, D. M., Allen, A. S., Lowe, W. L., & Reddy, T. E. (2017). Transversions have larger regulatory effects than transitions. *BMC Genomics*, 18(1), 1–9. <https://doi.org/10.1186/S12864-017-3785-4>
- Gupta, R. M., & Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of Clinical Investigation*, 124(10), 4154–4161.

<https://doi.org/10.1172/JCI72992>

Hacein-Bey-Abina, S., Kalle, C. Von, Schmidt, M., McCormack, M. P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C. S., Pawliuk, R., Morillon, E., Sorensen, R., Forster, A., Fraser, P., Cohen, J. I., Basile, G. de Saint, Alexander, I., Wintergerst, U., Frebourg, T., Aurias, A., ... Cavazzana-Calvo, M. (2003). LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science*, 302(5644), 415–419.

<https://doi.org/10.1126/SCIENCE.1088547>

Hall, R. S., Fedorov, A. A., Marti-Arbona, R., Fedorov, E. V., Kolb, P., Sauder, J. M., Burley, S. K., Shoichet, B. K., Almo, S. C., & Raushel, F. M. (2010). The Hunt for 8-Oxoguanine Deaminase. *Journal of the American Chemical Society*, 132(6), 1762–1763.

<https://doi.org/10.1021/JA909817D>

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>

Hazra, T. K., Das, A., Das, S., Choudhury, S., Kow, Y. W., & Roy, R. (2007). Oxidative DNA damage repair in mammalian cells: A new perspective. *DNA Repair*, 6(4), 470–480.

<https://doi.org/10.1016/J.DNAREP.2006.10.011>

Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P., Dizdaroglu, M., & Mitra, S. (2002). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proceedings of the National Academy of Sciences*, 99(6), 3523–3528. <https://doi.org/10.1073/PNAS.062053799>

Hegde, M. L., Hazra, T. K., & Mitra, S. (2008). Early steps in the DNA base excision/single-

- strand interruption repair pathway in mammalian cells. *Cell Research*, 18(1), 27–47.
<https://doi.org/10.1038/cr.2008.8>
- Hinnen, A., Hicks, J. B., & Fink, G. R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences*, 75(4), 1929–1933. <https://doi.org/10.1073/PNAS.75.4.1929>
- Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., Zeitler, B., Cherone, J. M., Meng, X., Hinkley, S. J., Rebar, E. J., Gregory, P. D., Urnov, F. D., & Jaenisch, R. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotechnology*, 29(8), 731–734.
<https://doi.org/10.1038/nbt.1927>
- Hölz, K., Pavlic, A., Lietard, J., & Somoza, M. M. (2019). Specificity and Efficiency of the Uracil DNA Glycosylase-Mediated Strand Cleavage Surveyed on Large Sequence Libraries. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-019-54044-x>
- Horlbeck, M. A., Witkowsky, L. B., Guglielmi, B., Replogle, J. M., Gilbert, L. A., Villalta, J. E., Torigoe, S. E., Tjian, R., & Weissman, J. S. (2016). Nucleosomes impede cas9 access to DNA in vivo and in vitro. *ELife*, 5(MARCH2016). <https://doi.org/10.7554/ELIFE.12677>
- Hsu, G. W., Ober, M., Carell, T., & Beese, L. S. (2004). Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* 2004 431:7005, 431(7005), 217–221. <https://doi.org/10.1038/nature02908>
- Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., & Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, 556(7699), 57–63. <https://doi.org/10.1038/nature26155>

- Hua, K., Tao, X., Yuan, F., Wang, D., & Zhu, J. K. (2018). Precise A·T to G·C Base Editing in the Rice Genome. *Molecular Plant*, 11(4), 627–630.
<https://doi.org/10.1016/J.MOLP.2018.02.007>
- Hua, K., Tao, X., & Zhu, J.-K. (2019). Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnology Journal*, 17(2), 499–504. <https://doi.org/10.1111/PBI.12993>
- Huai, C., Li, G., Yao, R., Zhang, Y., Cao, M., Kong, L., Jia, C., Yuan, H., Chen, H., Lu, D., & Huang, Q. (2017). Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nature Communications*, 8(1), 1–9. <https://doi.org/10.1038/s41467-017-01496-2>
- Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., & Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. *Nature Biotechnology* 2011 29:8, 29(8), 699–700.
<https://doi.org/10.1038/nbt.1939>
- Hussmann, J. A., Ling, J., Ravisankar, P., Yan, J., Cirincione, A., Xu, A., Simpson, D., Yang, D., Bothmer, A., Cotta-Ramusino, C., Weissman, J. S., & Adamson, B. (2021). Mapping the genetic landscape of DNA double-strand break repair. *Cell*, 184(22), 5653–5669.e25.
<https://doi.org/10.1016/J.CELL.2021.10.002>
- Ioannidi, E. I., Yarnall, M. T. N., Schmitt-Ulms, C., Krajeski, R. N., Lim, J., Villiger, L., Zhou, W., Jiang, K., Roberts, N., Zhang, L., Vakulskas, C. A., Walker, J. A., Kadina, A. P., Zepeda, A. E., Holden, K., Gootenberg, J. S., & Abudayyeh, O. O. (2021). Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases. *BioRxiv*, 2021.11.01.466786. <https://doi.org/10.1101/2021.11.01.466786>
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakamura, A. (1987). Nucleotide

- sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 169(12), 5429–5433. <https://doi.org/10.1128/JB.169.12.5429-5433.1987>
- Ishino, Yoshizumi, Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *Journal of Bacteriology*, 200(7). <https://doi.org/10.1128/JB.00580-17>
- Isogawa, A. (2004). Functional cooperation of Ogg1 and Mutyh in preventing G: C-->T: A transversions in mice. *Fukuoka Igaku Zasshi*, 95(1), 17–30. <https://pubmed.ncbi.nlm.nih.gov/15031996/>
- Jamieson, A. C., Kim, S.-H., & Wells, J. A. (1994). In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry*, 33(19), 5689–5695. <https://doi.org/10.1021/BI00185A004>
- Jasin, M., & Haber, J. E. (2016). The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA Repair*, 44, 6–16. <https://doi.org/10.1016/J.DNAREP.2016.05.001>
- Jayanth, N., & Puranik, M. (2017). Mechanism of Discrimination of 8-Oxoguanosine versus Guanosine by *Escherichia coli* Fpg. *The Journal of Physical Chemistry B*, 121(23), 5679–5687. <https://doi.org/10.1021/acs.jpcb.7b00205>
- Jeggo, P. A. (1998). 5. *DNA Breakage and Repair* (J. C. Hall, J. C. Dunlap, T. Friedmann, & F. Giannelli (eds.); Vol. 38, pp. 185–218). Academic Press. [https://doi.org/https://doi.org/10.1016/S0065-2660\(08\)60144-3](https://doi.org/https://doi.org/10.1016/S0065-2660(08)60144-3)

- Jiang, T., Zhang, X.-O., Weng, Z., & Xue, W. (2021). Deletion and replacement of long genomic sequences using prime editing. *Nature Biotechnology*, 1–8. <https://doi.org/10.1038/S41587-021-01026-Y>
- Jin, S., Lin, Q., Luo, Y., Zhu, Z., Liu, G., Li, Y., Chen, K., Qiu, J.-L., & Gao, C. (2021). Genome-wide specificity of prime editors in plants. *Nature Biotechnology* 2021, 1–8. <https://doi.org/10.1038/s41587-021-00891-x>
- Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.-L., Zhang, F., & Gao, C. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*, 364(6437), 292–295. <https://doi.org/10.1126/SCIENCE.AAW7166>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816–821. <https://doi.org/10.1126/science.1225829>
- Kang, B.-C., Yun, J.-Y., Kim, S.-T., Shin, Y., Ryu, J., Choi, M., Woo, J. W., & Kim, J.-S. (2018). Precision genome engineering through adenine base editing in plants. *Nature Plants* 2018 4:7, 4(7), 427–431. <https://doi.org/10.1038/S41477-018-0178-X>
- Kay, S., Hahn, S., Marois, E., Hause, G., & Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science*, 318(5850), 648–651. <https://doi.org/10.1126/SCIENCE.1144956>
- Keijzers, G., Bohr, V. A., & Rasmussen, L. J. (2015). Human exonuclease 1 (EXO1) activity characterization and its function on flap structures. *Bioscience Reports*, 35(3), 1–13.

<https://doi.org/10.1042/BSR20150058>

Khan, F. A., & Ali, S. O. (2017). Physiological Roles of DNA Double-Strand Breaks. *Journal of Nucleic Acids*, 2017. <https://doi.org/10.1155/2017/6439169>

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., Hwang, J., Kim, J.-I., & Kim, J.-S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods*, 12(3), 237–243. <https://doi.org/10.1038/nmeth.3284>

Kim, D. D. Y., Moon, S. Bin, Ko, J.-H., Kim, Y.-S., & Kim, D. D. Y. (2020). Unbiased investigation of specificities of prime editing systems in human cells. *Nucleic Acids Research*, 48(18), 10576–10589. <https://doi.org/10.1093/NAR/GKAA764>

Kim, D., Kim, D., Lee, G., Cho, S.-I., & Kim, J.-S. (2019). Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nature Biotechnology*, 37(4), 430–435. <https://doi.org/10.1038/s41587-019-0050-1>

Kim, D., & Kim, J.-S. (2018). DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. *Genome Research*, 28(12), 1894–1900. <https://doi.org/10.1101/GR.236620.118>

Kim, D., Lim, K., Kim, S.-T., Yoon, S., Kim, K., Ryu, S.-M., & Kim, J.-S. (2017). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nature Biotechnology*, 35(5), 475–480. <https://doi.org/10.1038/nbt.3852>

Kim, H. K., Kim, Y., Lee, S., Min, S., Bae, J. Y., Choi, J. W., Park, J., Jung, D., Yoon, S., & Kim, H. H. (2019). SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. *Science Advances*, 5(11), eaax9249.

<https://doi.org/10.1126/sciadv.aax9249>

- Kim, H. K., Min, S., Song, M., Jung, S., Choi, J. W., Kim, Y., Lee, S., Yoon, S., & Kim, H. (2018). Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nature Biotechnology*, 36(3), 239–241. <https://doi.org/10.1038/nbt.4061>
- Kim, H. S., Jeong, Y. K., Hur, J. K., Kim, J.-S., & Bae, S. (2019). Adenine base editors catalyze cytosine conversions in human cells. *Nature Biotechnology*, 37(10), 1145–1148. <https://doi.org/10.1038/s41587-019-0254-4>
- Kim, K., Ryu, S.-M., Kim, S.-T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S., & Kim, J.-S. (2017). Highly efficient RNA-guided base editing in mouse embryos. *Nature Biotechnology* 2017 35:5, 35(5), 435–437. <https://doi.org/10.1038/NBT.3816>
- Kim, N., Kim, H. K. H. H., Lee, S., Seo, J. H., Choi, J. W., Park, J., Min, S., Yoon, S., Cho, S.-R. R., & Kim, H. K. H. H. (2020). Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nature Biotechnology*, 38(11), 1328–1336. <https://doi.org/10.1038/s41587-020-0537-9>
- Kim, S., Cho, C.-S., Han, K., & Lee, J. (2016). Structural Variation of Alu Element and Human Disease. *Genomics & Informatics*, 14(3), 70–77. <https://doi.org/10.5808/GI.2016.14.3.70>
- Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology*, 35(4), 371–376. <https://doi.org/10.1038/nbt.3803>
- Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger

- fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, 93(3), 1156–1160. <https://doi.org/10.1073/PNAS.93.3.1156>
- Klar, A. J. S., Strathern, J. N., & Abraham, J. A. (1984). Involvement of double-stranded chromosomal breaks for mating-type switching in *Saccharomyces cerevisiae*. *Cold Spring Harbor Symposia on Quantitative Biology*, VOL. 49, 77–88. <https://doi.org/10.1101/SQB.1984.049.01.011>
- Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), 490–495. <https://doi.org/10.1038/nature16526>
- Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. <https://doi.org/10.1038/nature14592>
- 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. (2021). Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nature Biotechnology*, 1–12. <https://doi.org/10.1038/S41587-021-00938-Z>
- Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Maianti, J. P., Tay, T., Newby, G. A., Raguram, A., & Liu, D. R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology*, May. <https://doi.org/10.1038/nbt.4172>

- Koblan, L. W., Erdos, M. R., Wilson, C., Cabral, W. A., Levy, J. M., Xiong, Z.-M., Tavarez, U. L., Davison, L. M., Gete, Y. G., Mao, X., Newby, G. A., Doherty, S. P., Narisu, N., Sheng, Q., Krilow, C., Lin, C. Y., Gordon, L. B., Cao, K., Collins, F. S., ... Liu, D. R. (2021). In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature*, 589(7843), 608–614. <https://doi.org/10.1038/s41586-020-03086-7>
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), 420–424. <https://doi.org/10.1038/nature17946>
- Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., Kim, Y. B., Badran, A. H., & Liu, D. R. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C : G-to-T : A base editors with higher efficiency and product purity. *Science Advances*, August, 1–10.
- Kostriken, R., Strathern, J. N., Klar, A. J. S., Hicks, J. B., & Heffron, F. (1983). A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell*, 35(1), 167–174. [https://doi.org/10.1016/0092-8674\(83\)90219-2](https://doi.org/10.1016/0092-8674(83)90219-2)
- Koval, V. V., Knorre, D. G., & Fedorova, O. S. (2014). Structural features of the interaction between human 8-oxoguanine DNA glycosylase hOGG1 and DNA. *Acta Naturae*, 6(22), 52–65. <https://doi.org/10.32607/20758251-2014-6-3-52-65>
- Kramarov, D. A., & Vassetzky, N. S. (2011). Origin and evolution of SINEs in eukaryotic genomes. *Heredity* 2011 107:6, 107(6), 487–495. <https://doi.org/10.1038/hdy.2011.43>
- Krokan, H. E., & Bjørås, M. (2013). Base Excision Repair. *Cold Spring Harbor Perspectives in*

Biology, 5(4), 1–22. <https://doi.org/10.1101/CSHPERSPECT.A012583>

- Kundu, S., Brinkmeyer, M. K., Livingston, A. L., & David, S. S. (2009). Adenine removal activity and bacterial complementation with the human MutY homologue (MUTYH) and Y165C, G382D, P391L and Q324R variants associated with colorectal cancer. *DNA Repair*, 8(12), 1400–1410. <https://doi.org/10.1016/J.DNAREP.2009.09.009>
- Kurt, I. C., Zhou, R., Iyer, S., Garcia, S. P., Miller, B. R., Langner, L. M., Grünwald, J., & Joung, J. K. (2020). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nature Biotechnology*, 39(1), 41–46. <https://doi.org/10.1038/s41587-020-0609-x>
- Kuscu, C., Arslan, S., Singh, R., Thorpe, J., & Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology*, 32(7), 677–683. <https://doi.org/10.1038/nbt.2916>
- Lander, E. S. (2016). The Heroes of CRISPR. *Cell*, 164(1), 18–28. <https://doi.org/10.1016/J.CELL.2015.12.041>
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., & Maglott, D. R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, 44(D1), D862–D868. <https://doi.org/10.1093/NAR/GKV1222>
- Lazzarotto, C. R., Malinin, N. L., Li, Y., Zhang, R., Yang, Y., Lee, G. H., Cowley, E., He, Y., Lan, X., Jividen, K., Katta, V., Kolmakova, N. G., Petersen, C. T., Qi, Q., Strelcov, E.,

- Maragh, S., Krenciute, G., Ma, J., Cheng, Y., & Tsai, S. Q. (2020). CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nature Biotechnology*, 38(11), 1317–1327. <https://doi.org/10.1038/s41587-020-0555-7>
- Leibowitz, M. L., Papathanasiou, S., Doerfler, P. A., Blaine, L. J., Sun, L., Yao, Y., Zhang, C. Z., Weiss, M. J., & Pellman, D. (2021). Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nature Genetics*, 53(6), 895–905. <https://doi.org/10.1038/s41588-021-00838-7>
- Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., Zhang, R., & Gao, C. (2018). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biology* 2018 19:1, 19(1), 1–9. <https://doi.org/10.1186/S13059-018-1443-Z>
- Li, G.-M. (2007). Mechanisms and functions of DNA mismatch repair. *Cell Research* 2008 18:1, 18(1), 85–98. <https://doi.org/10.1038/cr.2007.115>
- Li, K., Wang, G., Andersen, T., Zhou, P., & Pu, W. T. (2014). Optimization of Genome Engineering Approaches with the CRISPR/Cas9 System. *PLOS ONE*, 9(8), e105779. <https://doi.org/10.1371/JOURNAL.PONE.0105779>
- Li, L., Wu, L. P., & Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. *Proceedings of the National Academy of Sciences*, 89(10), 4275–4279. <https://doi.org/10.1073/PNAS.89.10.4275>
- Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P., & Yang, B. (2011). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Research*, 39(1), 359–372.

<https://doi.org/10.1093/NAR/GKQ704>

- Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X., Yang, L., & Chen, J. (2018). Base editing with a Cpf1–cytidine deaminase fusion. *Nature Biotechnology* 2018 36:4, 36(4), 324–327. <https://doi.org/10.1038/NBT.4102>
- Liang, P., Ding, C., Sun, H., Xie, X., Xu, Y., Zhang, X., Sun, Y., Xiong, Y., Ma, W., Liu, Y., Wang, Y., Fang, J., Liu, D., Songyang, Z., Zhou, C., & Huang, J. (2017). Correction of β -thalassemia mutant by base editor in human embryos. *Protein & Cell* 2017 8:11, 8(11), 811–822. <https://doi.org/10.1007/S13238-017-0475-6>
- Lin, S., Staahl, B. T., Alla, R. K., & Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *ELife*, 3, e04766. <https://doi.org/10.7554/ELIFE.04766>
- Liu, D. R., & Koblan, L. W. (2018). *Cytosine to guanine base editor*. Patentscope.
- Liu, S., & Kong, D. (2020). End resection: a key step in homologous recombination and DNA double-strand break repair. *Genome Instability & Disease* 2020 2:1, 2(1), 39–50. <https://doi.org/10.1007/S42764-020-00028-5>
- Liu, Y., And, H.-I. K., & Bambara, R. A. (2004). Flap Endonuclease 1: A Central Component of DNA Metabolism. *Annual Review of Biochemistry*, 73, 589–615. <https://doi.org/10.1146/ANNUREV.BIOCHEM.73.012803.092453>
- Liu, Z., Chen, M., Chen, S., Deng, J., Song, Y., Lai, L., & Li, Z. (2018). Highly efficient RNA-guided base editing in rabbit. *Nature Communications* 2018 9:1, 9(1), 1–10. <https://doi.org/10.1038/S41467-018-05232-2>

- Lohman, G. J. S., Chen, L., & Evans, T. C. (2011). Kinetic Characterization of Single Strand Break Ligation in Duplex DNA by T4 DNA Ligase. *Journal of Biological Chemistry*, 286(51), 44187–44196. <https://doi.org/10.1074/JBC.M111.284992>
- Ma, Y., Yu, L., Zhang, X., Xin, C., Huang, S., Bai, L., Chen, W., Gao, R., Li, J., Pan, S., Qi, X., huang, X., & Zhang, L. (2018). Highly efficient and precise base editing by engineered dCas9-guide tRNA adenosine deaminase in rats. *Cell Discovery* 2018 4:1, 4(1), 1–4. <https://doi.org/10.1038/S41421-018-0047-9>
- MacHugh, D. E., Larson, G., & Orlando, L. (2017). Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annual Review of Animal Biosciences*, 5, 329–351. <https://doi.org/10.1146/ANNUREV-ANIMAL-022516-022747>
- Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., Townsend, J. A., Unger-Wallace, E., Sander, J. D., Müller-Lerch, F., Fu, F., Pearlberg, J., Göbel, C., Dassie, J. P., Pruett-Miller, S. M., Porteus, M. H., ... Joung, J. K. (2008). Rapid “Open-Source” Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. *Molecular Cell*, 31(2), 294–301. <https://doi.org/10.1016/J.MOLCEL.2008.06.016>
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., ... Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. In *Nature Reviews Microbiology* (Vol. 18, Issue 2, pp. 67–83). Nature Research. <https://doi.org/10.1038/s41579-019-0299-x>

- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, 339(6121), 823–826. <https://doi.org/10.1126/SCIENCE.1232033>
- Mancuso, P., Chen, C., Kaminski, R., Gordon, J., Liao, S., Robinson, J. A., Smith, M. D., Liu, H., Saryer, I. K., Saryer, R., Peterson, T. A., Donadoni, M., Williams, J. B., Siddiqui, S., Bunnell, B. A., Ling, B., MacLean, A. G., Burdo, T. H., & Khalili, K. (2020). CRISPR based editing of SIV proviral DNA in ART treated non-human primates. *Nature Communications*, 11(1), 1–11. <https://doi.org/10.1038/s41467-020-19821-7>
- Mao, Z., Bozzella, M., Seluanov, A., & Gorbunova, V. (2008). DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Http://Dx.Doi.Org/10.4161/Cc.7.18.6679*, 7(18), 2902–2906. <https://doi.org/10.4161/CC.7.18.6679>
- Markkanen, E., Dorn, J., & Hübscher, U. (2013). MUTYH DNA glycosylase: the rationale for removing undamaged bases from the DNA. *Frontiers in Genetics*, 4(FEB), 18. <http://journal.frontiersin.org/article/10.3389/fgene.2013.00018/abstract>
- Marquart, K. F., Allam, A., Janjuha, S., Sintsova, A., Villiger, L., Frey, N., Krauthammer, M., & Schwank, G. (2021). Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. *Nature Communications*, 12(1), 1–9. <https://doi.org/10.1038/s41467-021-25375-z>
- Masuda, Y., Takahashi, M., Fukuda, S., Sumii, M., & Kamiya, K. (2002). Mechanisms of dCMP Transferase Reactions Catalyzed by Mouse Rev1 Protein *. *Journal of Biological Chemistry*, 277(4), 3040–3046. <https://doi.org/10.1074/JBC.M110149200>

- Matsumoto, Yoshihiro. (2001). Molecular mechanism of PCNA-dependent base excision repair. *Progress in Nucleic Acid Research and Molecular Biology*, 68, 129–138.
[https://doi.org/10.1016/S0079-6603\(01\)68095-4](https://doi.org/10.1016/S0079-6603(01)68095-4)
- Matsumoto, Yukiko, Zhang, Q.-M., Takao, M., Yasui, A., & Yonei, S. (2001). Escherichia coli Nth and human hNTH1 DNA glycosylases are involved in removal of 8-oxoguanine from 8-oxoguanine/guanine mispairs in DNA. *Nucleic Acids Research*, 29(9), 1975–1981.
<https://doi.org/10.1093/NAR/29.9.1975>
- Mehta, A., & Haber, J. E. (2014). Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*, 6(9), a016428.
<https://doi.org/10.1101/CSHPERSPECT.A016428>
- Mendel, G. (1866). Versuche über Pflanzen-Hybriden. *Verhandlungen Des Naturforschenden Vereines, IV*, 3–47. https://www.deutschestextarchiv.de/mendel_pflanzenhybriden_1866/20
- Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* 2013 14:12, 14(12), 840–852.
<https://doi.org/10.1038/nrg3605>
- Miescher, F. (1869). Letter I; to Wilhelm His; Tübingen, February 26th, 1869. *Die Histochemischen Und Physiologischen Arbeiten von Friedrich Miescher-Aus Dem Wissenschaftlichen Briefwechsel von F. Miescher, I*, 33–38.
- Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O., & Rebar, E. J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing.

- Nature Biotechnology* 2007 25:7, 25(7), 778–785. <https://doi.org/10.1038/nbt1319>
- Miller, J., McLachlan, A. D., & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *The EMBO Journal*, 4(6), 1609.
[/pmc/articles/PMC554390/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/15791728/)
- Mojica, F. J. M., Juez, G., & Rodriguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Molecular Microbiology*, 9(3), 613–621. <https://doi.org/10.1111/J.1365-2958.1993.TB01721.X>
- Mojica, Francisco J.M., Diez-Villasenor, C., Garcia-Martinez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, 60(2), 174–182.
<http://www.ncbi.nlm.nih.gov/pubmed/15791728>
- Montero-Morán, G. M., Li, M., Rendón-Huerta, E., Jourdan, F., Lowe, D. J., Stumpff-Kane, A. W., Feig, M., Scazzocchio, C., & Hausinger, R. P. (2007). Purification and Characterization of the FeII- and α -Ketoglutarate-Dependent Xanthine Hydroxylase from *Aspergillus nidulans*. *Biochemistry*, 46(18), 5293. <https://doi.org/10.1021/BI700065H>
- Moore, J. K., & Haber, J. E. (1996). Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* 1996 383:6601, 383(6601), 644–646.
<https://doi.org/10.1038/383644a0>
- Morgan, T. H. (1910). Sex Limited Inheritance in *Drosophila*. *Science*, 32(812), 120–122.
- Mortusewicz, O., & Leonhardt, H. (2007). XRCC1 and PCNA are loading platforms with distinct kinetic properties and different capacities to respond to multiple DNA lesions. *BMC*

- Molecular Biology* 2007 8:1, 8(1), 1–8. <https://doi.org/10.1186/1471-2199-8-81>
- Mortusewicz, O., Rothbauer, U., Cardoso, M. C., & Leonhardt, H. (2006). Differential recruitment of DNA Ligase I and III to DNA repair sites. *Nucleic Acids Research*, 34(12), 3523–3532. <https://doi.org/10.1093/NAR/GKL492>
- Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959), 1501. <https://doi.org/10.1126/SCIENCE.1178817>
- Muller, H. J. (1927). Artificial Transmutation of the Gene. *Science*, 66(1699), 84–87. <https://doi.org/10.1126/SCIENCE.66.1699.84>
- Musunuru, K., Chadwick, A. C., Mizoguchi, T., Garcia, S. P., DeNizio, J. E., Reiss, C. W., Wang, K., Iyer, S., Dutta, C., Clendaniel, V., Amaonye, M., Beach, A., Berth, K., Biswas, S., Braun, M. C., Chen, H.-M., Colace, T. V., Ganey, J. D., Gangopadhyay, S. A., ... Kathiresan, S. (2021). In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature*, 593(7859), 429–434. <https://doi.org/10.1038/s41586-021-03534-y>
- Nambiar, T. S., Billon, P., Diedenhofen, G., Hayward, S. B., Taglialatela, A., Cai, K., Huang, J.-W., Leuzzi, G., Cuella-Martin, R., Palacios, A., Gupta, A., Egli, D., & Ciccia, A. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nature Communications* 2019 10:1, 10(1), 1–13. <https://doi.org/10.1038/s41467-019-11105-z>
- Nazarkina, Z. K., Khodyreva, S. N., Marsin, S., Lavrik, O. I., & Radicella, J. P. (2007). XRCC1 interactions with base excision repair DNA intermediates. *DNA Repair*, 6(2), 254–264. <https://doi.org/10.1016/J.DNAREP.2006.10.002>

- Nelson, J. R., Lawrence, C. W., & Hinkle, D. C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. *Nature*, 382(6593), 729–731. <https://doi.org/10.1038/382729a0>
- Nelson, J. W., Randolph, P. B., Shen, S. P., Everette, K. A., Chen, P. J., Anzalone, A. V., An, M., Newby, G. A., Chen, J. C., Hsu, A., & Liu, D. R. (2021). Engineered pegRNAs improve prime editing efficiency. *Nature Biotechnology*, 1–9. <https://doi.org/10.1038/s41587-021-01039-7>
- Nickoloff, J. A., Chen, E. Y., & Heffron, F. (1986). A 24-base-pair DNA sequence from the MAT locus stimulates intergenic recombination in yeast. *Proceedings of the National Academy of Sciences*, 83(20), 7831–7835. <https://doi.org/10.1073/PNAS.83.20.7831>
- Nicolas, E., Golemis, E. A., & Arora, S. (2016). POLD1: Central mediator of DNA replication and repair, and implication in cancer and other pathologies. *Gene*, 590(1), 128–141. <https://doi.org/10.1016/J.GENE.2016.06.031>
- Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*. <https://doi.org/10.1126/science.aaf8729>
- Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., Oura, S., Holmes, B., Tanaka, M., Seki, M., Hirano, H., Aburatani, H., Ishitani, R., Ikawa, M., Yachie, N., ... Nureki, O. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, 361(6408), 1259–1262. <https://doi.org/10.1126/science.aas9129>

- O'Driscoll, M., & Jeggo, P. A. (2006). The role of double-strand break repair — insights from human genetics. *Nature Reviews Genetics* 2006 7:1, 7(1), 45–54.
<https://doi.org/10.1038/nrg1746>
- Omelina, E. S., Ivankin, A. V., Letiagina, A. E., & Pindyurin, A. V. (2019). Optimized PCR conditions minimizing the formation of chimeric DNA molecules from MPRA plasmid libraries. *BMC Genomics*, 20(7), 1–10. <https://doi.org/10.1186/S12864-019-5847-2>
- Osborn, M. J., Starker, C. G., McElroy, A. N., Webber, B. R., Riddle, M. J., Xia, L., Defeo, A. P., Gabriel, R., Schmidt, M., Von Kalle, C., Carlson, D. F., Maeder, M. L., Joung, J. K., Wagner, J. E., Voytas, D. F., Blazar, B. R., & Tolar, J. (2013). TALEN-based gene correction for epidermolysis bullosa. *Molecular Therapy*, 21(6), 1151–1159.
<https://doi.org/10.1038/mt.2013.56>
- Park, S.-J., Jeong, T. Y., Shin, S. K., Yoon, D. E., Lim, S.-Y., Kim, S. P., Choi, J., Lee, H., Hong, J.-I., Ahn, J., Seong, J. K., & Kim, K. (2021). Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. *Genome Biology*, 22(1), 1–11.
<https://doi.org/10.1186/S13059-021-02389-W>
- Parkinson, N. J., Maslau, S., Ferneyhough, B., Zhang, G., Gregory, L., Buck, D., Ragoussis, J., Ponting, C. P., & Fischer, M. D. (2012). Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Research*, 22(1), 125–133. <https://doi.org/10.1101/GR.124016.111>
- Pattanayak, V., Guilinger, J. P., & Liu, D. R. (2014). Determining the specificities of TALENs, cas9, and other genome-editing enzymes. In *Methods in Enzymology* (Vol. 546, Issue C, pp. 47–78). Academic Press Inc. <https://doi.org/10.1016/B978-0-12-801185-0.00003-9>

- Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, 31(9), 839–843. <https://doi.org/10.1038/nbt.2673>
- Pattanayak, V., Ramirez, C. L., Joung, J. K., & Liu, D. R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature Methods*, 8(9), 765–772. <https://doi.org/10.1038/nmeth.1670>
- Pavletich, N., & Pabo, C. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, 252(5007), 809–817. <https://doi.org/10.1126/SCIENCE.2028256>
- Petri, K., Kim, D. Y., Sasaki, K. E., Canver, M. C., Wang, X., Shah, H., Lee, H., Horng, J. E., Clement, K., Iyer, S., Garcia, S. P., Guo, J. A., Newby, G. A., Pinello, L., Liu, D. R., Aryee, M. J., Musunuru, K., Joung, J. K., & Pattanayak, V. (2021). Global-scale CRISPR gene editor specificity profiling by ONE-seq identifies population-specific, variant off-target effects. *BioRxiv*, 2021.04.05.438458. <https://doi.org/10.1101/2021.04.05.438458>
- Petri, K., Zhang, W., Ma, J., Schmidts, A., Lee, H., Horng, J. E., Kim, D. Y., Kurt, I. C., Clement, K., Hsu, J. Y., Pinello, L., Maus, M. V., Joung, J. K., & Yeh, J.-R. J. (2021). CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nature Biotechnology*, 1–5. <https://doi.org/10.1038/s41587-021-00901-y>
- Picelli, S., Björklund, A. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*, 24(12), 2033–2040. <https://doi.org/10.1101/gr.177881.114>

- Potapov, V., & Ong, J. L. (2017). Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLOS ONE*, 12(1), e0169774.
<https://doi.org/10.1371/JOURNAL.PONE.0169774>
- Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151(3), 653–663. <https://doi.org/10.1099/MIC.0.27437-0>
- Prasad, R., Poltoratsky, V., Hou, E. W., & Wilson, S. H. (2016). Rev1 is a base excision repair enzyme with 5'-deoxyribose phosphate lyase activity. *Nucleic Acids Research*, 44(22), 10824–10833. <https://doi.org/10.1093/NAR/GKW869>
- Pruett-Miller, S. M., Connelly, J. P., Maeder, M. L., Joung, J. K., & Porteus, M. H. (2008). Comparison of Zinc Finger Nucleases for Use in Gene Targeting in Mammalian Cells. *Molecular Therapy*, 16(4), 707–717. <https://doi.org/10.1038/MT.2008.20>
- Puebla-Osorio, N., Lacey, D. B., Alt, F. W., & Zhu, C. (2006). Early Embryonic Lethality Due to Targeted Inactivation of DNA Ligase III. *Molecular and Cellular Biology*, 26(10), 3935–3941. <https://doi.org/10.1128/MCB.26.10.3935-3941.2006>
- Qiu, X., Wu, L., Huang, H., McDonel, P. E., Palumbo, A. V., Tiedje, J. M., & Zhou, J. (2001). Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology*, 67(2), 880–887.
<https://doi.org/10.1128/AEM.67.2.880-887.2001>
- Rees, H. A., Komor, A. C., Yeh, W.-H., Caetano-Lopes, J., Warman, M., Edge, A. S. B., & Liu,

- D. R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nature Communications*, 8, 15790.
<https://doi.org/10.1038/ncomms15790>
- Rees, H. A., & Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics*, 19(12), 770–788.
<https://doi.org/10.1038/s41576-018-0059-1>
- Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D., & Joung, J. K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology* 2012 30:5, 30(5), 460–465. <https://doi.org/10.1038/nbt.2170>
- Richter, M. F., Zhao, K. T., Eton, E., Lapinaite, A., Newby, G. A., Thuronyi, B. W., Wilson, C., Koblan, L. W., Zeng, J., Bauer, D. E., Doudna, J. A., & Liu, D. R. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology*, 38(7), 883–891. <https://doi.org/10.1038/s41587-020-0453-z>
- Roberts, R. J. (2005). How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences*, 102(17), 5905–5908.
<https://doi.org/10.1073/PNAS.0500923102>
- Römer, P., Hahn, S., Jordan, T., Strauß, T., Bonas, U., & Lahaye, T. (2007). Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science*, 318(5850), 645–648. <https://doi.org/10.1126/SCIENCE.1144958>
- Rouet, P., Smih, F., & Jasin, M. (1994a). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of*

Sciences, 91(13), 6064–6068. <https://doi.org/10.1073/PNAS.91.13.6064>

Rouet, P., Smih, F., & Jasin, M. (1994b). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology*, 14(12), 8096–8106. <https://doi.org/10.1128/MCB.14.12.8096-8106.1994>

Rudin, N., Sugarman, E., & Haber, J. E. (1989). Genetic and Physical Analysis of Double-Strand Break Repair and Recombination in *Saccharomyces Cerevisiae*. *Genetics*, 122(3), 519. [/pmc/articles/PMC1203726/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/1203726/)

Ryu, S.-M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.-T., Kim, H. S., Kim, D., Lee, H., Chung, E., & Kim, J.-S. (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nature Biotechnology* 2018 36:6, 36(6), 536–539. <https://doi.org/10.1038/NBT.4148>

Saksela, M., & Raivio, K. O. (1996). Cloning and expression in vitro of human xanthine dehydrogenase/oxidase. *Biochemical Journal*, 315(1), 235–239. <https://doi.org/10.1042/BJ3150235>

Sander, J. D., Cade, L., Khayter, C., Reyon, D., Peterson, R. T., Joung, J. K., & Yeh, J.-R. J. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature Biotechnology* 2011 29:8, 29(8), 697–698. <https://doi.org/10.1038/nbt.1934>

Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., Blackburn, J. S., Thibodeau-Beganny, S., Qi, Y., Pierick, C. J., Hoffman, E., Maeder, M. L., Khayter, C., Reyon, D., Dobbs, D., Langenau, D. M., Stupar, R. M., Giraldez, A. J., ... Joung, J. K. (2010). Selection-free zinc-finger-nuclease engineering by context-dependent

assembly (CoDA). *Nature Methods* 2010 8:1, 8(1), 67–69.

<https://doi.org/10.1038/NMETH.1542>

Sansbury, B. M., Hewes, A. M., & Kmiec, E. B. (2019). Understanding the diversity of genetic outcomes from CRISPR-Cas generated homology-directed repair. *Communications Biology* 2019 2:1, 2(1), 1–10. <https://doi.org/10.1038/s42003-019-0705-y>

Schene, I. F., Joore, I. P., Oka, R., Mokry, M., van Vugt, A. H. M., van Boxtel, R., van der Doef, H. P. J., van der Laan, L. J. W., Verstegen, M. M. A., van Hasselt, P. M., Nieuwenhuis, E. E. S., & Fuchs, S. A. (2020). Prime editing for functional repair in patient-derived disease models. *Nature Communications* 2020 11:1, 11(1), 1–8. <https://doi.org/10.1038/s41467-020-19136-7>

Schmid-Burgk, J. L., Gao, L., Li, D., Gardner, Z., Strecker, J., Lash, B., & Zhang, F. (2020). Highly Parallel Profiling of Cas9 Variant Specificity. *Molecular Cell*, 78(4), 794-800.e8. <https://doi.org/10.1016/J.MOLCEL.2020.02.023>

Schröder, A. R. W., Shinn, P., Chen, H., Berry, C., Ecker, J. R., & Bushman, F. (2002). HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell*, 110(4), 521–529. [https://doi.org/10.1016/S0092-8674\(02\)00864-4](https://doi.org/10.1016/S0092-8674(02)00864-4)

Seibert, E., Ross, J. B. A., & Osman, R. (2002). Role of DNA flexibility in sequence-dependent activity of uracil DNA glycosylase. *Biochemistry*, 41(36), 10976–10984. <https://doi.org/10.1021/BI026121O>

Shen, M. W., Arbab, M., Hsu, J. Y., Worstell, D., Culbertson, S. J., Krabbe, O., Cassa, C. A., Liu, D. R., Gifford, D. K., & Sherwood, R. I. (2018). Predictable and precise template-free

CRISPR editing of pathogenic variants. *Nature*, 563(7733), 646–651.

<https://doi.org/10.1038/s41586-018-0686-x>

Shimoni, L., & Glusker, J. P. (1995). Hydrogen bonding motifs of protein side chains:

Descriptions of binding of arginine and amide groups. *Protein Science*, 4(1), 65–74.

<https://doi.org/10.1002/PRO.5560040109>

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351(6268), 84–88.

<https://doi.org/10.1126/SCIENCE.AAD5227>

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.-C., Chames, P., Prieto, J., Redondo, P.,

Blanco, F. J., Bravo, J., Montoya, G., Pâques, F., & Duchateau, P. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic*

Acids Research, 34(22), e149–e149. <https://doi.org/10.1093/NAR/GKL720>

Song, M., Kim, H. K., Lee, S., Kim, Y., Seo, S.-Y., Park, J., Choi, J. W., Jang, H., Shin, J. H.,

Min, S., Quan, Z., Kim, J. H., Kang, H. C., Yoon, S., & Kim, H. H. (2020). Sequence-specific prediction of the efficiencies of adenine and cytosine base editors. *Nature*

Biotechnology, 38(9), 1037–1043. <https://doi.org/10.1038/S41587-020-0573-5>

Sonoda, E., Hohegger, H., Saberi, A., Taniguchi, Y., & Takeda, S. (2006). Differential usage of non-homologous end-joining and homologous recombination in double strand break repair.

DNA Repair, 5(9–10), 1021–1029. <https://doi.org/10.1016/J.DNAREP.2006.05.022>

Spencer, J. M., & Zhang, X. (2017). Deep mutational scanning of *S. pyogenes* Cas9 reveals important functional domains. *Scientific Reports*, 7(1), 1–14.

<https://doi.org/10.1038/s41598-017-17081-y>

Srinivasan, S., Torres, A. G., & Pouplana, L. R. de. (2021). Inosine in Biology and Disease.

Genes 2021, Vol. 12, Page 600, 12(4), 600. <https://doi.org/10.3390/GENES12040600>

Stark, J. M., Pierce, A. J., Oh, J., Pastink, A., & Jasin, M. (2004). Genetic Steps of Mammalian

Homologous Repair with Distinct Mutagenic Consequences. *Molecular and Cellular*

Biology, 24(21), 9305–9316. <https://doi.org/10.1128/MCB.24.21.9305-9316.2004>

Stenson, P. D., Mort, M., Ball, E. V., Shaw, K., Phillips, A. D., & Cooper, D. N. (2013). The

Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human*

Genetics, 133(1), 1–9. <https://doi.org/10.1007/S00439-013-1358-4>

Sutton, W. S. (1902). On the morphology of the chromoso group in *Brachystola magna*. *The*

Biological Bulletin, 4(1), 24–39.

Szczepek, M., Brondani, V., Büchel, J., Serrano, L., Segal, D. J., & Cathomen, T. (2007).

Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnology* 2007 25:7, 25(7), 786–793.

<https://doi.org/10.1038/nbt1317>

Tahara, Y., Auld, D., Ji, D., Beharry, A. A., Kietrys, A. M., Wilson, D. L., Jimenez, M., King,

D., Nguyen, Z., & Kool, E. T. (2018). Potent and Selective Inhibitors of 8-Oxoguanine DNA Glycosylase. *Journal of the American Chemical Society*, 140(6), 2105–2114.

<https://doi.org/10.1021/JACS.7B09316>

Tahara, Y., Kietrys, A. M., Hebenbrock, M., Lee, Y., Wilson, D. L., & Kool, E. T. (2019). Dual

- Inhibitors of 8-Oxoguanine Surveillance by OGG1 and NUDT1. *ACS Chemical Biology*, 14(12), 2606–2615. <https://doi.org/10.1021/ACSCHEMBIO.9B00490>
- Tajai, P., Fedeles, B. I., Suriyo, T., Navasumrit, P., Kanitwithayanun, J., Essigmann, J. M., & Satayavivad, J. (2018). An engineered cell line lacking OGG1 and MUTYH glycosylases implicates the accumulation of genomic 8-oxoguanine as the basis for paraquat mutagenicity. *Free Radical Biology and Medicine*, 116, 64–72. <https://doi.org/10.1016/J.FREERADBIOMED.2017.12.035>
- Tanaka, S., Yoshioka, S., Nishida, K., Hosokawa, H., Kakizuka, A., & Maegawa, S. (2018). In vivo targeted single-nucleotide editing in zebrafish. *Scientific Reports* 2018 8:1, 8(1), 1–11. <https://doi.org/10.1038/S41598-018-29794-9>
- Taylor, G. K., Petrucci, L. H., Lambert, A. R., Baxter, S. K., Jarjour, J., & Stoddard, B. L. (2012). LAHEDES: the LAGLIDADG homing endonuclease database and engineering server. *Nucleic Acids Research*, 40(W1), W110–W116. <https://doi.org/10.1093/NAR/GKS365>
- Tesson, L., Usal, C., Ménoret, S., Leung, E., Niles, B. J., Remy, S., Santiago, Y., Vincent, A. I., Meng, X., Zhang, L., Gregory, P. D., Anegón, I., & Cost, G. J. (2011). Knockout rats generated by embryo microinjection of TALENs. *Nature Biotechnology* 2011 29:8, 29(8), 695–696. <https://doi.org/10.1038/nbt.1940>
- Thomas, K. R., Folger, K. R., & Capecchi, M. R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell*, 44(3), 419–428. [https://doi.org/10.1016/0092-8674\(86\)90463-0](https://doi.org/10.1016/0092-8674(86)90463-0)

- Thuronyi, B. W., Koblan, L. W., Levy, J. M., Yeh, W.-H., Zheng, C., Newby, G. A., Wilson, C., Bhaumik, M., Shubina-Oleinik, O., Holt, J. R., & Liu, D. R. (2019). Continuous evolution of base editors with expanded target compatibility and improved activity. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-019-0193-0>
- Tsai, S. Q., & Joung, J. K. (2016). Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nature Reviews Genetics*, *17*(5), 300–312. <https://doi.org/10.1038/NRG.2016.28>
- Tsai, S. Q., Nguyen, N. T., Malagon-Lopez, J., Topkar, V. V., Aryee, M. J., & Joung, J. K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nature Methods*, *14*(6), 607–614. <https://doi.org/10.1038/nmeth.4278>
- Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, a J., Le, L. P., Aryee, M. J., & Joung, J. K. (2014). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, *33*(2), 187–198. <https://doi.org/10.1038/nbt.3117>
- Tycko, J., Barrera, L. A., Huston, N. C., Friedland, A. E., Wu, X., Gootenberg, J. S., Abudayyeh, O. O., Myer, V. E., Wilson, C. J., & Hsu, P. D. (2018). Pairwise library screen systematically interrogates *Staphylococcus aureus* Cas9 specificity in human cells. *Nature Communications*, *9*(1), 2962. <https://doi.org/10.1038/s41467-018-05391-2>
- Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., & Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* *2005* *435*:7042, *435*(7042), 646–651. <https://doi.org/10.1038/nature03556>

- Van Dyck, E., Stasiak, A. Z., Stasiak, A., & West, S. C. (1999). Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* 1999 401:6751, 401(6751), 403–403.
<https://doi.org/10.1038/43942>
- Van Loon, B., Markkanen, E., & Hübscher, U. (2010). Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA Repair*, 9(6), 604–616.
<https://doi.org/10.1016/J.DNAREP.2010.03.004>
- Vidigal, J. A., & Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nature Communications*, 6(1), 8083.
<https://doi.org/10.1038/ncomms9083>
- Visnes, T., Cázares-Körner, A., Hao, W., Wallner, O., Masuyer, G., Loseva, O., Mortusewicz, O., Wiita, E., Sarno, A., Manoilov, A., Astorga-Wells, J., Jemth, A.-S., Pan, L., Sanjiv, K., Karsten, S., Gokturk, C., Grube, M., Homan, E. J., Hanna, B. M. F., ... Helleday, T. (2018). Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation. *Science*, 362(6416), 834–839. <https://doi.org/10.1126/SCIENCE.AAR8048>
- Vonesch, S. C., Li, S., Szu Tu, C., Hennig, B. P., Dobrev, N., & Steinmetz, and L. M. (2021). Fast and inexpensive whole-genome sequencing library preparation from intact yeast cells. *G3 Genes|Genomes|Genetics*, 11(1). <https://doi.org/10.1093/G3JOURNAL/JKAA009>
- Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, 368(6488), 290–296. <https://doi.org/10.1126/SCIENCE.ABA8853>
- Wang, H., Zhang, S., Song, L., Qu, M., & Zou, Z. (2020). Synergistic lethality between PARP-

- trapping and alantolactone-induced oxidative DNA damage in homologous recombination-proficient cancer cells. *Oncogene*, 39(14), 2905–2920. <https://doi.org/10.1038/s41388-020-1191-x>
- Wang, L., Xue, W., Yan, L., Li, X., Wei, J., Chen, M., Wu, J., Yang, B., Yang, L., & Chen, J. (2017). Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Research* 2017 27:10, 27(10), 1289–1292. <https://doi.org/10.1038/cr.2017.111>
- Wang, L., Xue, W., Zhang, H., Gao, R., Qiu, H., Wei, J., Zhou, L., Lei, Y.-N., Wu, X., Li, X., Liu, C., Wu, J., Chen, Q., Ma, H., Huang, X., Cai, C., Zhang, Y., Yang, B., Yin, H., ... Chen, J. (2021). Eliminating base-editor-induced genome-wide and transcriptome-wide off-target mutations. *Nature Cell Biology*, 23(5), 552–563. <https://doi.org/10.1038/S41556-021-00671-4>
- Wang, Yanhong, Zhou, L., Tao, R., Liu, N., Long, J., Qin, F., Tang, W., Yang, Y., Chen, Q., & Yao, S. (2020). sgBE: a structure-guided design of sgRNA architecture specifies base editing window and enables simultaneous conversion of cytosine and adenosine. *Genome Biology*, 21(1), 1–16. <https://doi.org/10.1186/S13059-020-02137-6>
- Wang, Yanli, & Schlick, T. (2007). Distinct energetics and closing pathways for DNA polymerase β with 8-oxoG template and different incoming nucleotides. *BMC Structural Biology*, 7(1), 1–14. <https://doi.org/10.1186/1472-6807-7-7>
- Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y., & Axel, R. (1977). Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. *Cell*, 11(1), 223–232. [https://doi.org/10.1016/0092-8674\(77\)90333-6](https://doi.org/10.1016/0092-8674(77)90333-6)

- Wolfe, S. A., Greisman, H. A., Ramm, E. I., & Pabo, C. O. (1999). Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *Journal of Molecular Biology*, 285(5), 1917–1934. <https://doi.org/10.1006/JMBI.1998.2421>
- Wood, A. J., Lo, T. W., Zeitler, B., Pickle, C. S., Ralston, E. J., Lee, A. H., Amora, R., Miller, J. C., Leung, E., Meng, X., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., & Meyer, B. J. (2011). Targeted genome editing across species using ZFNs and TALENs. *Science*, 333(6040), 307. <https://doi.org/10.1126/SCIENCE.1207773>
- Wu, H., Yang, W. P., & Barbas, C. F. (1995). Building zinc fingers by selection: toward a therapeutic application. *Proceedings of the National Academy of Sciences*, 92(2), 344–348. <https://doi.org/10.1073/PNAS.92.2.344>
- Xiang, X., Qu, K., Liang, X., Pan, X., Wang, J. J. J., Han, P., Dong, Z., Liu, L., Zhong, J., Ma, T., Wang, Y., Yu, J., Zhao, X., Li, S., Xu, Z., Wang, J. J. J., Zhang, X., Jiang, H., Xu, F., ... Luo, Y. (2020). Massively parallel quantification of CRISPR editing in cells by TRAP-seq enables better design of Cas9, ABE, CBE gRNAs of high efficiency and accuracy. *BioRxiv*, 2020.05.20.103614. <https://doi.org/10.1101/2020.05.20.103614>
- Xie, Y., Yang, H., Cunanan, C., Okamoto, K., Shibata, D., Pan, J., Barnes, D. E., Lindahl, T., McIlhatton, M., Fishel, R., & Miller, J. H. (2004). Deficiencies in Mouse Myh and Ogg1 Result in Tumor Predisposition and G to T Mutations in Codon 12 of the K-Ras Oncogene in Lung Tumors. *Cancer Research*, 64(9), 3096–3102. <https://doi.org/10.1158/0008-5472.CAN-03-3834>
- Yan, F., Kuang, Y., Ren, B., Wang, J., Zhang, D., Lin, H., Yang, B., Zhou, X., & Zhou, H. (2018). Highly Efficient A·T to G·C Base Editing by Cas9n-Guided tRNA Adenosine

Deaminase in Rice. *Molecular Plant*, 11(4), 631–634.

<https://doi.org/10.1016/J.MOLP.2018.02.008>

- Yang, B., Zhu, W., Johnson, L. B., & White, F. F. (2000). The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. *Proceedings of the National Academy of Sciences*, 97(17), 9807–9812. <https://doi.org/10.1073/PNAS.170286897>
- Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J., & Xia, N. (2020). Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *International Journal of Molecular Sciences* 2020, Vol. 21, Page 6461, 21(18), 6461. <https://doi.org/10.3390/IJMS21186461>
- Yang, M., Peng, S., Sun, R., Lin, J., Wang, N., & Chen, C. (2018). The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Reports*, 22(2), 372–382. <https://doi.org/10.1016/J.CELREP.2017.12.048>
- Yang, Y., Gao, Y., Zlatanou, A., Tateishi, S., Yurchenko, V., Rogozin, I. B., & Vaziri, C. (2018). Diverse roles of RAD18 and Y-family DNA polymerases in tumorigenesis. *Cell Cycle*, 17(7), 833. <https://doi.org/10.1080/15384101.2018.1456296>
- Yasui, M., Kanemaru, Y., Kamoshita, N., Suzuki, T., Arakawa, T., & Honma, M. (2014). Tracing the fates of site-specifically introduced DNA adducts in the human genome. *DNA Repair*, 15(1), 11–20. <https://doi.org/10.1016/J.DNAREP.2014.01.003>
- Yeh, W.-H., Chiang, H., Rees, H. A., Edge, A. S. B., & Liu, D. R. (2018). In vivo base editing of post-mitotic sensory cells. *Nature Communications* 2018 9:1, 9(1), 1–10.

<https://doi.org/10.1038/S41467-018-04580-3>

- Yoshida, J., Akagi, K., Misawa, R., Kokubu, C., Takeda, J., & Horie, K. (2017). Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Scientific Reports*, 7(1), 1–18. <https://doi.org/10.1038/srep43613>
- Yu, Y., Leete, T. C., Born, D. A., Young, L., Barrera, L. A., Lee, S.-J., Rees, H. A., Ciaramella, G., & Gaudelli, N. M. (2020). Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nature Communications*, 11(1), 1–10. <https://doi.org/10.1038/s41467-020-15887-5>
- Zaman, Q. U., Li, C., Cheng, H., & Hu, Q. (2019). Genome editing opens a new era of genetic improvement in polyploid crops. *The Crop Journal*, 7(2), 141–150. <https://doi.org/10.1016/J.CJ.2018.07.004>
- Zhang, L., Richards, A., Barrasa, M. I., Hughes, S. H., Young, R. A., & Jaenisch, R. (2021). Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proceedings of the National Academy of Sciences*, 118(21). <https://doi.org/10.1073/PNAS.2105968118>
- Zhang, Y., Qin, W., Lu, X., Xu, J., Huang, H., Bai, H., Li, S., & Lin, S. (2017). Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nature Communications* 2017 8:1, 8(1), 1–5. <https://doi.org/10.1038/S41467-017-00175-6>
- Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., Rosser, S. J., Bi, C., & Zhang, X. (2020). Glycosylase base editors enable C-to-A and C-to-G base changes. *Nature Biotechnology*, 39(1), 35–40. <https://doi.org/10.1038/S41587-020-0592-2>

Zhao, Y., Gregory, M. T., Biertümpfel, C., Hua, Y.-J., Hanaoka, F., & Yang, W. (2013).

Mechanism of somatic hypermutation at the WA motif by human DNA polymerase η .

Proceedings of the National Academy of Sciences, 110(20), 8146–8151.

<https://doi.org/10.1073/PNAS.1303126110>

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L. M., Li, Y., &

Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*, 364(6437), 289–292.

<https://doi.org/10.1126/SCIENCE.AAV9973>

Zuo, Z., & Liu, J. (2016). Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Scientific Reports*, 6(1), 1–9.

<https://doi.org/10.1038/srep37584>