UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Functional Characterization of DNA Repair Gene Variants in Live Cells Enabled Through Precision Genome Editing, Chemical Biology, and Biochemical Tools

Permalink

https://escholarship.org/uc/item/2v40t8qm

Author

Vasquez, Carlos Anthony

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Functional Characterization of DNA Repair Gene Variants in Live Cells Enabled Through Precision Genome Editing, Chemical Biology, and Biochemical Tools

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Carlos Anthony Vasquez

Committee in charge:

Professor Alexis Komor, Chair
Professor Thomas Bussey
Professor Galia Debelouchina
Professor Daniel Donoghue
Professor Valerie Schmidt
Professor JoAnn Trejo

2024

Copyright

Carlos Anthony Vasquez, 2024

All rights reserved

The Dissertation of Carlos Anthony Vasquez is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

In honor of my grandmother, Maria Luz Vasquez.

The most courageous and sweetest woman I have ever known. Thank you for sacrificing so much for the slight possibility of giving your family a better future. Here we are. I love you very much.

“En honor a mi abuelita, Maria Luz Vasquez.

Eras la mujer más valiente y la mas agradable persona en el mundo. Gracias por sacrificar tanto por la mínima posibilidad de darle un futuro mejor a tu familia. Hemos llegado. Te quiero mucho, mi luz brillante.”

TABLE OF CONTENTS

Dissertation Approval Page……. iii

Dedication……... iv

Table of Contents……... v

List of Abbreviations …….ix

List of Figures…….... xi

List of Tables……..…….... xii

Acknowledgements……..... xiii

Vita…… xvi

Abstract of the Dissertation……. xvii

Chapter 1 Introduction……..……..…….1

1.1. Introduction to Genome Editing……..……1

1.2. CRISPR-based Genome Editing Technologies…….……. 3

1.3. CRISPR-Cas9……. 4

1.4. Base Editing……. 8

1.5. Dissertation Overview …….…… 11

1.6. Acknowledgments…….. 11

Chapter 2 Optimized Protocols on Using Base Editing Technology in Mammalian Cell to Produce Isogenic Cell Lines..…..……13

2.1. Introduction……..…..……..13

2.2. Strategic Planning …….……..14

2.3 Protocol 1: Design and Production of Plasmids for Base Editing Experiments……..15

2.3.A Protocol Steps……..18

2.4 Protocol 2: Transfecting Adherent Cells and Harvesting Genomic DNA……26

2.4.A Protocol Steps……...27

2..5 Protocol 3: Genotyping Harvested Cells using Sanger Sequencing……...…….32

2.5.A Protocol Steps……...……...…….32

2.5.B. Using EditR to Quantify Base Editing Efficiencies …… 35

2.6. Alternate Protocol 1: Next-Generation Sequencing to Quantify Base Editing……37

2.6.A Protocol Steps…….……. 38

2.7. Protocol 4: Single Cell Isolation of Base Edited Cells using FACS……41

2.7.A Protocol Steps…….……. 42

2.8 Alternate Protocol 2: Single Cell Isolation of Cells using Dilution Plating…….45

2.8.A Protocol Steps…….……..45

2.9 Protocol 5: Genotyping of Clonally-Expanded Isogenic Cells……47

2.9.A Protocol Steps…….……. 47

2.10. Critical Parameters:.…… 50

2.10.A. gRNA design considerations…….……...…… 50

2.10.B. Bystander Editing …….……51

2.10.C. Base Editor Selection Considerations …….…….52

2.10.D. Inclusion of Proper Controls ……..…..…….53

2.10.E. Quantification of Base-Editing Efficiency in Bulk Cells……..54

2.10.F. Cell Line Considerations…….……. 55

2.11 Understanding Results…….……. 55

2.12 Acknowledgments…….…….56

Chapter 3 Functional Characterization of *MUTYH* Variants in Live Cells Enabled Through Precision Genome Editing, Chemical Biology, and Biochemical Tools…… 58

3.1. Introduction……58

3.2. Results…….…….……. 63

3.2.1. Engineering Isogenic Cell Lines with MUTYH SNVs……. 63

3.2.2. Analysis of Protein Expression Levels in Isogenic Cell Lines……. 67

3.2.3. Constructing Fluorescent Reporter Plasmids to Measure Repair of 8-oxoG•A, 8-oxoG•C, and 8-oxoG•[O] Within Live Cells……. 68

3.2.4. Evaluation of 8-oxoG•A Repair Activities of MUTYH Mutants Within Live Cells using a Fluorescent Reporter…… 70

3.2.5. Reductions in 8-oxoG•[O] Repair Efficiency for the W131* and D271G MUTYH Mutants Suggest Defective Interactions with Downstream BER Proteins……74

3.2.6. Defective MUTYH-APE1 Interactions Identified by Co-Immunoprecipitation…….. 75

3.3. Discussion……. 78

3.4. Methods……. 80

3.4.1. Molecular cloning……. 80

3.4.2. Cell culture and transfections…….82

3.4.3. Sanger sequencing of the MUTYH locus. ……..……. 83

3.4.4. Fluorescence activated cell sorting (FACS)……...……. 83

3.4.5. Next-generation sequencing (NGS) of isogenic cell lines……...……84

3.4.6. Transfections for flow cytometry. …….……..……. 85

3.4.7. Flow cytometry analysis of DNA repair with EGFP reporter vectors…..…..……. 85

3.4.8. Preparation of cell extracts and Western blotting……. 86

3.4.9. Co-immunoprecipitation experiments ……. 87

3.4.10. Data analysis & statistics……..…….89

3.5. Acknowledgments……..90

Chapter 4 Non-Thesis-Related University Service ……..……91

4.1. Development of a Research-Practice Partnership with Local High Schools in San Diego: The Genome Editing Technologies Program……..91

4.1.1. Implementing the Genome Editing Technologies Program……...93

4.1.2. Discussion on Student Experiences and Evaluation of the Program…….95

4.2. Teaching at UCSD Through the Summer Graduate Teaching Scholars Program ….……. 98

4.3. Acknowledgments……. 99

References……...……107

LIST OF ABBREVIATIONS

ABE	Adenine base editor
BEs	Base editors
CBE	Cytosine base editor
CRISPR	Clustered regularly interspaced short palindromic repeats
dCas9	Catalytically-dead or inactive Cas9
nCas9	nickase Cas9
DSB	Double strand break
FACS	Fluorescence activated cell sorting
gDNA	Genomic DNA
gRNA	Guide RNA or spacer
HDR	Homology-directed repair
HEK293T	Human embryonic kidney immortalized cell line
HTS	High-throughput sequencing
Indels	Insertions and deletions
MMR	Mismatch repair pathway
NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
PAM	Protospacer adjacent motif
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SSB	Single strand break
ssDNA	Single-stranded DNA

TALENs	Transcription-activator-like effector nucleases
UGI	Uracil glycosylase inhibitor
UNG	Uracil-N-glycosylase enzyme
ZFN	Zinc-finger nuclease

LIST OF FIGURES

Figure 1	Illustrations of genome editing modifications	3
Figure 2	Overview of CRISPR genome editing techniques	7
Figure 3	Schematic of base editing technologies for genome editing	12
Figure 4	Steps of generation of isogenic cell lines	17
Figure 5	Overview of protocol 1	26
Figure 6	Images of successful and unsuccessful transfection experiments	31
Figure 7	Sanger sequencing traces from genotyping of cells	36
Figure 8	Images of the dilution plating method to generate single cell colonies	44
Figure 9	Potential pitfalls of Base Editing	52
Figure 10	Overview of MUTYH structure and function	62
Figure 11	Sanger sequencing of on-target locus for *MUTYH* isogenic cell lines	66
Figure 12	Generation of isogenic cell lines harboring clinically-relevant *MUTYH* single nucleotide variants (SNVs)	68
Figure 13	Repair of 8-oxoG•A using a MUTYH lesion-specific plasmid reporter	73
Figure 14	Defective MUTYH-APE1 Interactions Identified by 8-oxoG•[O] repair and co-immunoprecipitation	77
Figure 15	Illustration of constructs used in the “GFP-itis” activity	94
Figure 16	Student survey responses for Sage Creek High School	97

LIST OF TABLES

Table 1 List of commonly used CRISPR-based genome editing tools……..……. 101

Table 2 CRISPR-based genome editing software.…… 102

Table 3 Suggested base editors designed to maximize on-target editing or minimize off-target (OT) editing……. 103

Table 4 Primers and additional sequences used to generate custom gRNAs or amplify the target locus…… 104

Table 5 List of primers used for PCR amplification to produce gRNA plasmids…….. 105

Table 6 List of primers used for Sanger sequencing and Next-generation sequencing (NGS) for *MUTYH* genomic DNA…… 106

ACKNOWLEDGEMENTS

As I stand at the threshold of completing this monumental accomplishment for my family and I, I want to first give thanks to God. I have been blessed with an unforgettable journey full of ups and downs, full of wounds and healing, full of failures and tremendous successes – all for the purpose of life-learning and to help me connect purposefully with others. I thank You for your blessings. I thank You for speaking what I need when I need it and that when You speak, there is nothing the enemy can do. I am open to Your word. I am open to the possibility that You are leading me like never before. To a place that you have prepared for us before this world began. Thank You for giving us a slight glimpse of the world you have created through the field of chemistry.

To my grandmother, Maria, and parents, Ana, Rosie, Carlos, and Jessie, you are my pillars of strength and motivation. I am grateful that you have always believed in me and stood by my side. Your sacrifices, guidance, and love have made it possible for me to reach milestones like these; ones I could never have imagined accomplishing without you. I am forever grateful for everything you have done for me. This achievement is a reflection of your unconditional love. To my siblings, Jason, Jasmine, Gianna and Sophia, I love you very much. You all mean the world to me.

I am also deeply grateful for the guidance, mentorship, and support provided by the mentors who have played pivotal roles in shaping my overall academic growth. Their expertise, encouragement, and invaluable insights have been instrumental to shaping my career and me as a person. And although it is impossible to acknowledge everyone who has helped me throughout the years, I do want to acknowledge a few by name: Dr. Anita Casavantes Bradford, Dr. Luis Mota-Bravo, Dr. Thomas Bussey, Dr. Mark Richardson, Dr. Marlene De la Cruz, and Professor Walker-

Waugh. You all spent countless hours to support me, provide me with feedback, and encouraged me throughout the years. Your kindness will not been wasted and I will forever be grateful for your incredible support and care.

To Alexis, you are a rare gift to this world, embodying not only remarkable scientific achievements but also exceptional leadership, grace, and boundless patience in mentoring others. Each day spent in your lab was a testament to your ability to instill confidence and provide the fortitude needed to overcome challenges. I owe everything to you and am profoundly grateful for your unwavering belief in my capabilities. Thank you for being my beacon of inspiration.

Why use real names when you can just dazzle everyone with a string of initials? To my lab mates — CAB, SG, ME, BLR, QTC, ZSB, EMP, KLR, BRP, KAD, ZM, RAA, MJH, SM, NMZ, ZK, NO, KS, (including my mentees during my graduate school tenure, Marc (MUZ) and Dominika (DKS) — thank you for putting up with me and most of all, for being the caring and beautiful people you are. Behind those initials lie real-life superheroes.

To my best friends, Yoshio, Alex, Oscar, CAB (Cameron) and Asia. I love you all very much – especially Asia :).

Chapter 1 is reproduced, in part, with permission, from: **Vasquez, C. A**. and Komor, A. (2023) Textbook: Advanced Chemical Biology: Chemical Dissection and Reprogramming of Biological Systems. Chapter 6: Chemical Biology of Genome Engineering. *Wiley-VCH.* 99-134. ISBN: 978-3-527-34733-9

Chapter 2 is reproduced, in full, with permission, from: **Vasquez, C.A.,** Cowan, Q.T., Komor, A.C. (2020). Base Editing in Human Cells to Produce Single Nucleotide Variant Clonal Cell Lines. *Curr. Protoc. Mol. Biol.* **133***,* e129*.*

Chapter 3 is reproduced, in full, with permission, from: **Vasquez, C.A.**, Zepeda, M.U., Sandel, D.K., Cowan, Q.T., Peiris, M., Donoghue, D.J., Komor, A.K. (2024) Precision Genome Editing and In-Cell Measurements of Oxidative DNA Damage Repair Enable Functional Characterization of Cancer-Associated MUTYH Variants. *Submitted. Nucleic Acids Research.*

Chapter 4 is reproduced, in full, with permission, from: **Vasquez, C.A.,** Evanoff, M., Ranzau, B., Gu, S., Deters, Emma; Komor, A.C. (2023). Curing "GFP-itis" in Bacteria with Base Editors: Development of a Genome Editing Science Program Implemented with High School Biology Students. *CRISPR J.* **6**, 3:186-195.

VITA

2016 Associate in Arts, General Science, Santa Monica College, Santa Monica, CA

2018 Bachelor of Science, Chemistry, University of California, Irvine, Irvine, CA

2020 Master of Science, Chemistry, University of California San Diego, La Jolla, CA

2024 Doctorate of Philosophy, Chemistry, University of California San Diego, La Jolla, CA

PUBLICATIONS

Journals

1. Richardson, M.B., Brown, D.B., **Vasquez, C.A**., Ziller, J.W., Johnston, K.M., Weiss, G.A. (2018). Synthesis and explosion hazards of 4-azido-L-phenylalanine. *J. Org. Chem.* **83***,* 8: 4525-4536. Reported in *Chem.& Eng. News* by Carmen Drahl on April 9, 2018. Commentary by Derek Lowe's "In the Pipeline" blog on April 27, 2018.

2. **Vasquez, C.A.,** Cowan, Q.T., Komor, A.C. (2020). Base Editing in Human Cells to Produce Single Nucleotide Variant Clonal Cell Lines. *Curr. Protoc. Mol. Biol.* **133***,* e129.

3. Burnett, C.A., Wong, A.T., **Vasquez, C.A.**, McHugh, C.A. Yeo, G.W., Komor, A.C. (2022). Examination of the Cell Cycle Dependence of Cytosine and Adenine Base Editors. *Front. Genome Ed.* **4.** 923718.

4. **Vasquez, C.A.,** Evanoff, M., Ranzau, B., Gu, S., Deters, Emma; Komor, A.C. (2023). Curing "GFP-itis" in Bacteria with Base Editors: Development of a Genome Editing Science Program Implemented with High School Biology Students. *CRISPR J.* **6**, 3:186-195.

5. Lawrence, E., et al., (2024). Functional EPAS1/HIF2A missense variant is associated with hematocrit in Andean highlanders. *Science Advances.* **824.** 111779.

6. **Vasquez, C.A.**, Zepeda, M.U., Sandel, D.K., Cowan, Q.T., Peiris, M., Donoghue, D.J., Komor, A.K. (2024) Precision Genome Editing and In-Cell Measurements of Oxidative DNA Damage Repair Enable Functional Characterization of Cancer-Associated MUTYH Variants. *Submitted. Nucleic Acids Research.*

Textbook Contributions

7. **Vasquez, C.A.**, & Komor, A.C. (2023) Advanced Chemical Biology: Chemical Dissection and Reprogramming of Biological Systems. Chapter 6: Chemical Biology of Genome Engineering. *Wiley-VCH.* 99-134. ISBN: 978-3-527-34733-9

ABSTRACT OF THE DISSERTATION

Functional Characterization of DNA Repair Gene Variants in Live Cells Enabled Through Precision Genome Editing, Chemical Biology, and Biochemical Tools

by

Carlos Anthony Vasquez

Doctor of Philosophy in Chemistry

University of California San Diego, 2024

Professor Alexis Komor, Chair

Progress in next-generation sequencing (NGS) technologies has streamlined the detection of human genetic variations, and the identification of clinically actionable genes and pathogenic mutations has transformed precision medicine. However, only a small fraction of identified human genetic variants have been assigned a clinical classification or functionally characterized. This

highlights the importance of investigating genetic variants and obtaining mechanistic insight into disease etiology and progression.

Base editing is a new precision genome editing methodology that utilizes native DNA repair pathways within living cells to either fix or install genetic variants. In Chapter 2, we detail our findings which aim to provide an optimized protocol in utilizing base editing technology. Then in Chapter 3, we detail how we harness base editing to overcome many of the limitations of traditional genome editing to generate both homozygous and heterozygous isogenic cell lines of four *MUTYH* variants. To date, these were the first reports of successful generation of isogenic cell lines containing *MUTYH* variants and also the first to functionally characterize them within living cells.

Finally, in Chapter 4, non-thesis-related university service garnered throughout the primary author and researcher's tenure is briefly discussed as this work has also led to both personal and professional advances.

Chapter 1

Introduction

1.1 Introduction to Genome Editing

One of the most important questions in science is a simple one. What makes us human? A chemical biologist interested in genomics and human health would likely approach this question from the perspective of DNA. Afterall, this simple biomolecule affects every aspect of biology, including protein function, cellular operations, and ultimately human health.

In the early 1940's, it was established that all the information needed for reproduction, growth, and survival was stored in our DNA, and it was our DNA that defined the traits (or genes) that were passed on from generation to generation.[1] What was not understood was how the primary sequence of the approximately 3 billion base pairs within our genomes led to vastly different biological phenotypes. Since then, chemical biologists have been building tools to manipulate the genome within live cells to better understand this phenomenon. This ability, called genetic engineering, has made the understanding of the genome possible, and to a larger extent, gives us the opportunity to answer questions that are foundational to us as human beings.

In 2017, the genome editing field made headlines when a 44-year old man named Brian Madeux became the world's first human patient to be treated for a genetic disease by "in body" (*in vivo*) genome editing.[2] People that suffer with Madeux's condition, Hunter syndrome, lack a gene required to break down carbohydrates. Without this ability, carbohydrates will eventually build up inside of cells and wreak havoc in organs throughout the body. Scientists administered a Zinc Finger Nuclease[3] into Madeux's liver cells to cut the region of the defective gene and allow a corrected copy to be inserted in its place. This was only the first instance of an "in body" genome editing therapeutic. Since this key milestone, there has been an explosion of genome editing technology-based therapeutics, the effect of which can be directly seen in the surge of IND (Investigational New Drug) filings and clinical trials. It is therefore an exciting and opportune time to learn about this burgeoning field.

Genetic engineering as the process of manipulating, or altering, the primary sequence of the genome of a living cell or organism. **Genome editing** is a type of genetic engineering in which the alterations are made at a specific, pre-defined location within the genome (or genomic locus). Although there has been many breakthroughs in the field of mammalian cell genome engineering, starting with gene targeting methods and followed by double-stranded break (DSB)-reliant genome editing tools, this chapter will focus on DSB-free genome editing tools, referred to as based editing. I will focus on the benefits and its current limitations with an emphasis on precision, specificity, efficiency, and programmability and then how I used this tool to study genetic variants within the *MUTYH* gene.

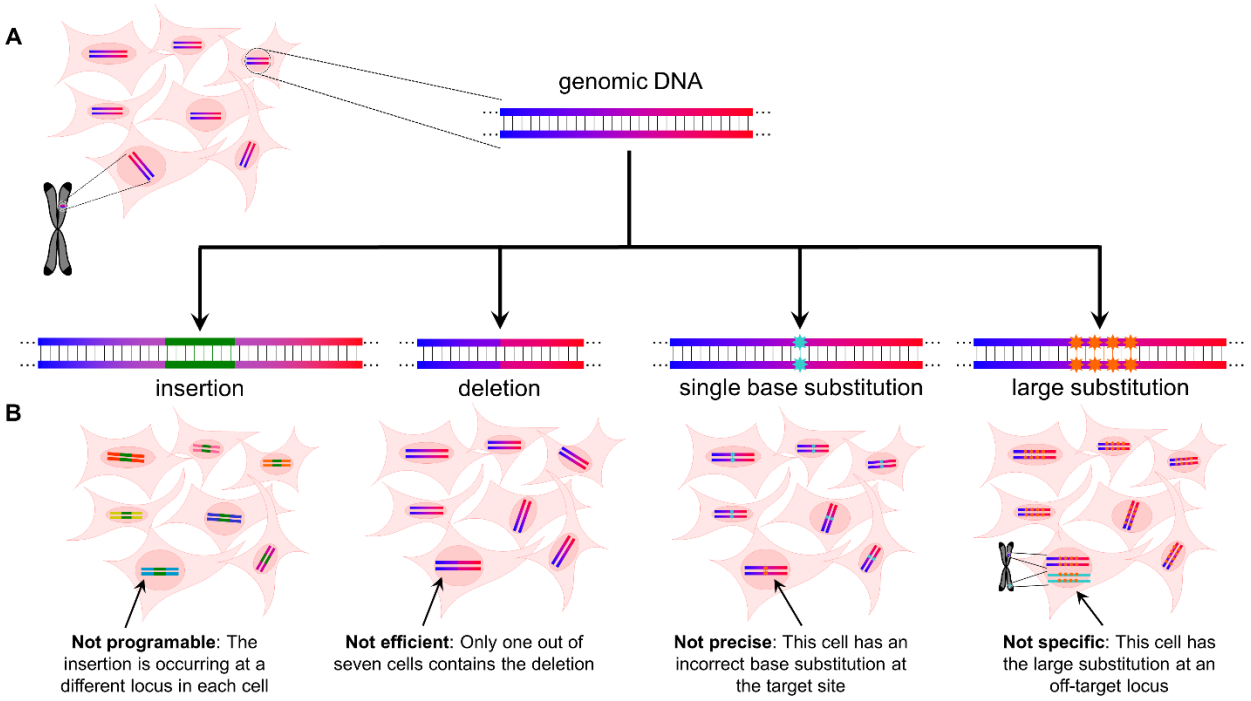

Figure 1. Schematic overview of different types of modifications that can be done using genome editing. **A**) Magnification of genomic locus of interest to be modified (top). The target genomic DNA sequence is shown with a color gradient, blue to red (left to right), that represents the primary DNA sequence. Below the target genomic DNA are four different examples of modifications that can be introduced using genome editing tools. Left to right: An insertion of a new DNA sequence (colored green), a deletion of a specific portion of the genomic DNA sequence (represented by the loss in the middle section of the color gradient), a single base substitution (cyan colored star; the original genomic DNA sequence is preserved except for the identity of a single base pair), and a multi-base substitution (orange colored stars; this can be thought of as a combination of a deletion and insertion). **B**) Examples of the four key genome editing tool characteristics that are important to consider for effective genome editing experiments. Left to right: programmability, efficiency, precision, and specificity.

1.2 CRISPR-based Genome Editing Technologies

For his dissertation thesis in 1992, Francisco Mojica started sequencing the genome of halophiles (bacterial and archaeal species that can grow in saline conditions) to better understand how their genes help them adapt to changes in salt concentrations. Through these experiments, he discovered that these organisms contained unique regularly spaced tandem repeats of genetic sequences in their genome (the sequences between the repeats, called spacers, were different from one another, **Figure 2A**), but their function was a mystery.[4] Compared to the complexity of the human genome, halophiles are primitive organisms; every portion of the genome has to serve a

purpose. Since these repeats comprised a decent portion of their genome, he hypothesized (later proven to be correct) that they must somehow be beneficial for their survival. In 2005, he reported that the spacer portion of these CRISPR (a term he coined through a correspondence with Ruud Jansen[5]) sequences were identical to sequences from the genomes of bacteriophage, and they were part of an adaptive immune system that the host used to degrade the DNA of invading viruses. [6,7] It took many years of painstaking work by an assortment of scientists to fully understand all the components of CRISPR, and then to repurpose this system for mammalian cell genome editing.

1.3 CRISPR-Cas9

In 2012, Jennifer Doudna and Emmanuelle Charpentier [8], as well as independent work by Virginijus Siksnys[9], biochemically characterized the mechanism by which the CRISPR-associated protein Cas9 binds and cleaves DNA, in effect ushering in the current age of genome editing (**Figure 2B**). They discovered that Cas9 forms a ribonucleoprotein (RNP) complex with a piece of RNA called the guide RNA (gRNA). The 5' end of the gRNA, called the spacer region, is ~20 to 30 nt in length (depending on the organism that the CRISPR system comes from), and must be complimentary to the target DNA sequence (this corresponding sequence in the genomic DNA, which is identical in sequence to the spacer, is called the "protospacer"). Additionally, the protospacer must be directly next to a protospacer adjacent motif (PAM) to enable RNP binding. Each Cas9 protein (from different organisms) has its own unique PAM recognition sequence. For the *Streptococcus pyogenes* (Sp) Cas9 homolog that Doudna and Charpentier characterized (and which most genome engineers continue to use to this day), this sequence is NGG (where N = adenine/cytosine/guanine/thymine). Upon recognition of the target site, the Cas9:gRNA complex unwinds the dsDNA and the gRNA anneals to the protospacer region to form an R-loop (**Figure 2B**). By convention, the strand that is base-paired with the gRNA is called the "target strand", and

the strand that lacks a base-pairing partner (the protospacer, which has the same sequence as the spacer) is called the "non-target strand". Cas9 then introduces a blunt-end DSB 3 bp upstream of the PAM sequence using two separate endonuclease domains (**Figure 2B**) [9,10]. This *in vitro* characterization of how certain CRISPR systems are able to protect their hosts from invading phage was of great interest to both the prokaryotic immunity field as well as genome engineers. Within a year, publications by Feng Zhang, George Church, Jin-Soo Kim, Jennifer Doudna, and colleagues had demonstrated the ability of Cas9 to perform genome editing in human cells with high efficiency and an ease of programmability that was lacking with traditional DSB-reliant technology such as Zinc Finger Nucleases (ZFN) and Transcription Activator-Like Effector Nucleases (TALENs) [11–14].

Although the CRISPR-Cas9 system introduces a blunt-ended DSB rather than 4 bp overhangs as with ZFNs and TALENs, the downstream DNA repair events for Cas9-mediated genome editing are largely the same: indels (random insertion and deletion of nucleotides at the DSB site, can be introduced via Non-Homology End Joining (NHEJ), or Homology-Directed Repair (HDR) can be used to incorporate a new sequence via an appropriately designed donor template (**Figure 2B**). However, the crucial difference between Cas9 and ZFNs/TALENs is that the use of a gRNA eliminates the requirement of designing, developing, and validating a custom DNA-binding protein for each new genomic locus target. Instead, reprogramming the endonuclease to bind and cleave at a new DNA sequence can be as simple as copying and pasting the target sequence. Now, the design and validation process can be done in a matter of days rather than months. The field's largest bottleneck was essentially eliminated, providing researchers with a drastic increase in time and resources to spend on previously sidelined challenges in the field;

the explosion of innovative new tools and strategies that followed closely on the heels of the mechanistic elucidation of CRISPR-Cas9 should, therefore, be no surprise.

In particular, the speed with which CRISPR-based genome editing agents have been applied therapeutically is quite striking. As of early 2024, there are currently 42 completed or actively recruiting clinical trials involving CRISPR on www.clinicaltrials.gov, with more to surely follow. These CRISPR-based clinical trials are intended to treat cancers, eye disease, and blood disorders. Specifically, the first *in vivo* (within a living patient) clinical trial to use CRISPR-Cas9, called *BRILLIANCE*, is aimed to treat Leber congenital amaurosis 10 (LCA10), an inherited form of blindness. This clinical trial will also assess the safety, tolerability and efficacy of CRISPR-based therapeutics.

However, therapeutics are just a small fraction of how CRISPR systems are revolutionizing the genome editing field. Because of the potential widespread applicability of genome editing to many different scientific disciplines, Virginijus Siksnys, Emmanuelle Charpentier, and Jennifer Doudna were awarded the Kavli Prize in Nanoscience in 2018, and Charpentier and Doudna were also awarded the Nobel Prize in Chemistry in 2020 for their groundbreaking work, which prompted an outburst of scientific innovation,

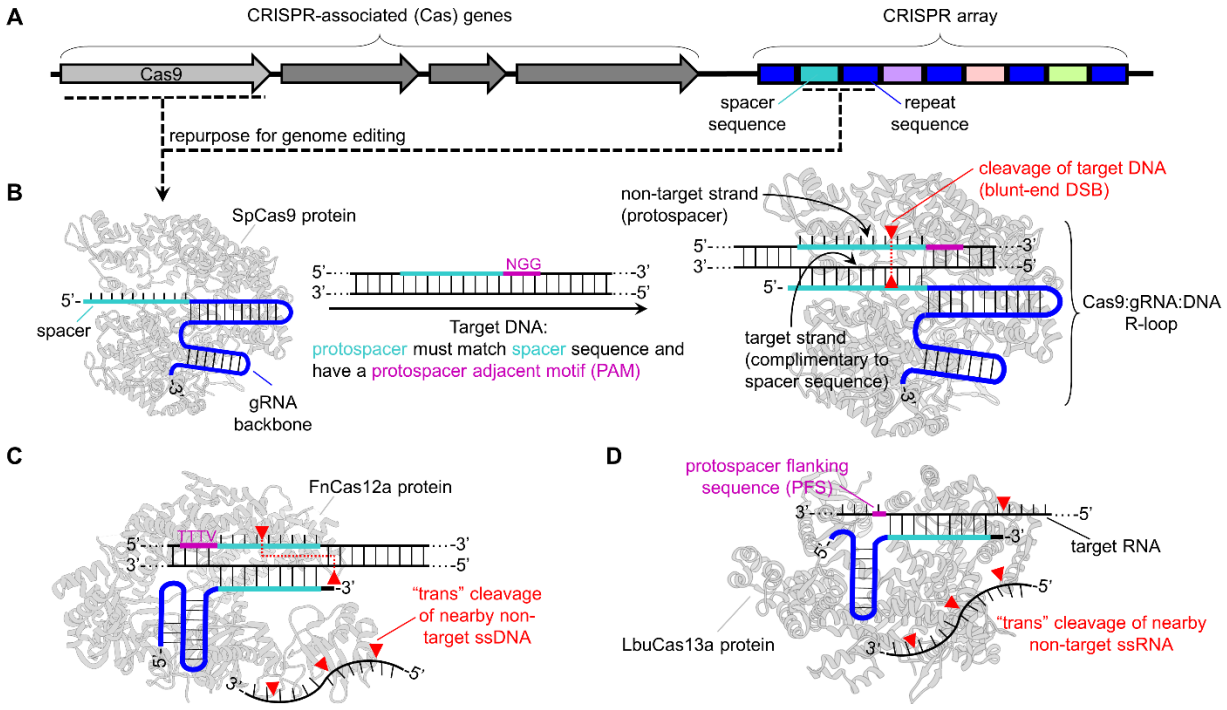

Figure 2. Overview of CRISPR genome editing technologies. A) CRISPR locus of the *S. pyogenes* (Sp) bacterial CRISPR system. The locus is comprised of a CRISPR array and CRISPR-associated (Cas) genes. The CRISPR array consists of multiple variable spacer regions (represented by different colors) separated by constant repeat sequences (shown in dark blue). The SpCas9 gene, along with other Cas genes, are located in the vicinity of the CRISPR array. **B**) Repurposing of this CRISPR system for genome editing requires expression of the Cas9 gene and transcription of one of the spacer-repeat gRNA sequences. Once SpCas9 is expressed (colored light gray, PDB: 5Y36), it will form a ribonucleoprotein (RNP) complex with the gRNA (colored in blue and cyan). The 5' end of the gRNA is called the spacer region (colored in cyan) and can be modified by researchers to be complimentary to any target DNA sequence within the genome. This corresponding sequence in the target genomic DNA is called the protospacer (colored in cyan as well). An additional requirement for SpCas9 RNP target recognition and binding is the presence of a protospacer adjacent motif (PAM; colored in purple) sequence next to the protospacer (note: For SpCas9, the PAM is an NGG sequence. For other Cas9 homologs, the PAM requirement is different, see **Table 1**). Upon recognition of the target site, the Cas9:gRNA complex unwinds the dsDNA and the gRNA anneals to the protospacer region to form an R-loop. By convention, the strand that is base-paired with the gRNA is called the "target strand", and the strand the lacks a base-pairing partner (the protospacer) is called the "non-target strand". SpCas9 then introduces a blunt-end DSB 3 bp upstream of the PAM sequence (represented by the red triangles and red dashed lines). **C**) Repurposing the *F. novicida* (Fn) bacterial CRISPR system, FnCas12a (originally known as Cpf1) for genome editing (PDB: 5NFV; color coded to match part **B**). FnCas12a is also an RNA-guided DNA endonuclease but has a different PAM requirement (TTTV where V = A, C, or G), binds to DNA in the opposite orientation compared to Cas9 systems, has a spacer sequence on the 3' end of the gRNA, and introduces staggered DSBs. Additionally, Cas12 enzymes cleave nearby nontarget ssDNA after binding to their target sequences. This activity is referred to as "trans cleavage" activity. **D)** Repurposing the *L. buccalis* (Lbu) bacterial CRISPR system, LbuCas13a, for genome editing (PDB: 5XWP; color coded to match part **B**). LbuCas13a, and other Cas13 enzymes, are also RNA-guided, but instead of targeting dsDNA, they target and cleave ssRNA. In this case, the protospacer flanking sequence (PFS), which is similar to the PAM sequences for SpCas9 and FnCas12a, is located at the 3' end of the spacer sequence and consists of a single U, C, or A bp. Additionally, Cas13 enzymes perform trans cleavage on nearby nontarget ssRNA after binding to their target RNA sequences.

1.4 Base Editing

The explosion in genome editing tool development was largely facilitated by the enormous reduction in time and resources required to re-program the Cas9 protein[15] as compared to traditional DSB-reliant technology such as TALENs and ZFNs. However, another unique aspect of Cas9, its use of an R-loop to bind to DNA, enabled the development of a new class of genome editing agents: those that utilize non-DSB DNA damage intermediates to introduce genomic modifications. Together with her postdoctoral mentor David Liu, my advisor, Dr. Alexis Komor, developed the first such method, base editing (**Figure 3**). In their initial study, they demonstrated how to introduce point mutations into the genome of live cells without first introducing a DSB, but instead through the chemical modification of a DNA base.[16] The first base editors, which are now called cytosine base editors (CBEs), consisted of a ssDNA-specific cytidine deaminase enzyme fused to either a catalytically inactivated or impaired Cas9 (either dCas9 or nCas9, respectively). The Cas9 portion of the BE will unwind the target dsDNA and allow the gRNA to anneal to the protospacer region. Once annealed, a small window of ~5 nucleotides of ssDNA (narrowing or broadening of this window can be accomplished with different BE variants) on the complimentary strand is exposed and readily accessible to the deaminase.[17] The cytidine deaminase enzyme will then catalyze the deamination of any cytidines within this window into uracils (which have the base pairing properties of thymine). Subsequent DNA replication or repair using the resulting uracils as a template results in the permanent introduction of a C•G bp to T•A (**Figure 3**). When using the ncas9-derived CBE, the target DNA strand (the one that is base-paired with the gRNA) is cleaved, to promote the cell to use the uracil-containing strand as a template during DNA repair. Additionally, due to the high efficiency of uracil excision by the cell's endogenous Uracil-N-glycosylase enzyme (UNG)[18], a short peptide that binds and inactivates

UNG (called UGI) is fused to the end of the CBE construct. This initial paper described the development of the first, second, and third generation base editors (BE1, BE2, and BE3), and a fourth-generation editor (BE4, which was a more optimized version of BE3) was reported shortly thereafter.[19] A plethora of CBE variants that employ different deaminase enzymes, alternate construct architectures, and different Cas enzymes have since been developed; the choice of which construct to use will depend on the experiment being conducted. [20–30]

A year after the development of cytosine base editing came the development of adenine base editors (ABEs) by Nicole Gaudelli and David Liu. [31] The premise of ABEs was similar: use a ssDNA adenosine deaminase enzyme to convert a target A•T bp to G•C, through an inosine (the deamination product of adenine, which has the base pairing properties of guanine) intermediate. However, unlike cytidine deaminases, there were no naturally occurring adenosine deaminase enzymes that could accept ssDNA as a substrate. Seven rounds of directed evolution were used to convert a tRNA-specific adenosine deaminase enzyme into a ssDNA adenosine deaminase, giving rise to ABE7.10 (**Figure 3**). ABE and CBEs introduce their respective point mutations with remarkably high efficiency and precision compared to DSB-reliant tools – particularly in non-dividing cells. Also due to their avoidance of DSBs, large-scale chromosomal rearrangements (which can result when using DSB-reliant technologies) are not observed when using BEs, even when multiplexing (performing genome editing at multiple genomic loci at once). Protocols with detailed descriptions of how researchers can harness base editing technologies to introduce mutations in live cells have recently been published [32,33], as well as software to aid in BE gRNA design and predictions of base editing efficiency and precision (**Table 2**).

However, like all current genome-editing technologies, base editing has limitations. BEs cannot introduce transversion mutations, insertions, or deletions. Recently, the basic architecture

of CBEs was repurposed to produce a new class of BE that facilitates C•G to G•C point mutations (CGBEs).[34–36] As mentioned earlier, uracil excision by UNG is quite efficient, and in fact can result in nontrivial levels of C•G to non-T•A outcomes when using CBEs, even when UGI is included in the CBE construct (**Figure 3**). CGBEs take advantage of this and replace the UGI component of CBEs with various DNA repair proteins to enhance abasic site formation following uracil introduction. The most common C•G to non-T•A outcome is to G•C, thus CGBEs produce a mixture of outcomes, with C•G to G•C being the most common.[34] These C•G to non-T•A outcomes by CBE need to be better understood from a DNA repair perspective to allow for additional strategies that improve the precision of CBEs (**Figure 3C**). It is important to note that, because of the inefficient repair of inosine by the cell, ABEs do not exhibit this same phenomenon.

The strict requirement for a PAM to be positioned 12-16 nt from the target base makes certain targets inaccessible. However, the use of expanded/relaxed PAM Cas variants has greatly alleviated this issue. [20,37–39] Another key limitation of BEs is "bystander editing," which occurs when other C or A nucleobases fall within the same editing window as the target base. In this case, these "bystander" bases can be unintentionally modified alongside the target base. However, bystander editing can be minimized through the use of deaminase mutants (such as less processive enzymes, or those with strict sequence preferences), or through careful protospacer design to "push" bystanders outside of the editing window [20,26]. Finally, BEs also suffer from off-target editing. BEs have "gRNA-dependent" off-targets, which occurs when the Cas component of the BE binds at an off-target location and the deaminase component of the BE edits C's or A's within the editing window of this new locus. High-fidelity Cas variants have been incorporated into BE architectures to combat these off-targets. [22] CBEs, but not ABEs, have also been found to have "gRNA-independent" off-target DNA editing activity. [40,41] This occurs when DNA replication or

transcription exposes ssDNA and the cytidine deaminase component of the CBE deaminates Cs within this exposed DNA. These off-targets are potentially due to an inherent difference in binding affinity between the respective deaminases. CBE variants with alternate deaminase domains have been developed that combat these off-targets.[29,30] Finally, both CBEs and ABEs have "gRNA-independent" RNA off-target editing. [42,43] This occurs when the deaminase component of the BE binds to and edits RNA transcripts within the cell. Due to the short lifetime of RNA, these off-target edits are transient; nevertheless, mutated deaminases have been developed to combat these off-targets as well. [43–45] BEs hold great potential as therapeutics, and in fact proof-of-concept studies have already demonstrated their potential in cell therapies, as well as to treat progeria, sickle cell disease, and liver diseases. [46–50]

1.5 Dissertation overview

The primary objectives of this dissertation is to contribute to the use of using the DSB-free genome editing technology called base editing in the field of functional genomics, especially towards generating *HIF2a* variants and *MUTYH* variants for functional characterization during my PhD work. Chapter 2 will address specific protocols and considerations for scientists to use this technology, then Chapter 3 will have a detailed functional characterization of several *MUTYH* variants. Finally, Chapter 4 will briefly cover some individual growth and development work, as well as mentorship work during my graduate studies.

1.6 Acknowledgments

Chapter 1 is reproduced, in part, with permission, from: **Vasquez, C. A**. and Komor, A. (2023) Textbook: Advanced Chemical Biology: Chemical Dissection and Reprogramming of Biological Systems. Chapter 6: Chemical Biology of Genome Engineering. *Wiley-VCH.* 99-134.

ISBN: 978-3-527-34733-9. The dissertation author was the primary author on this reprinted materials.

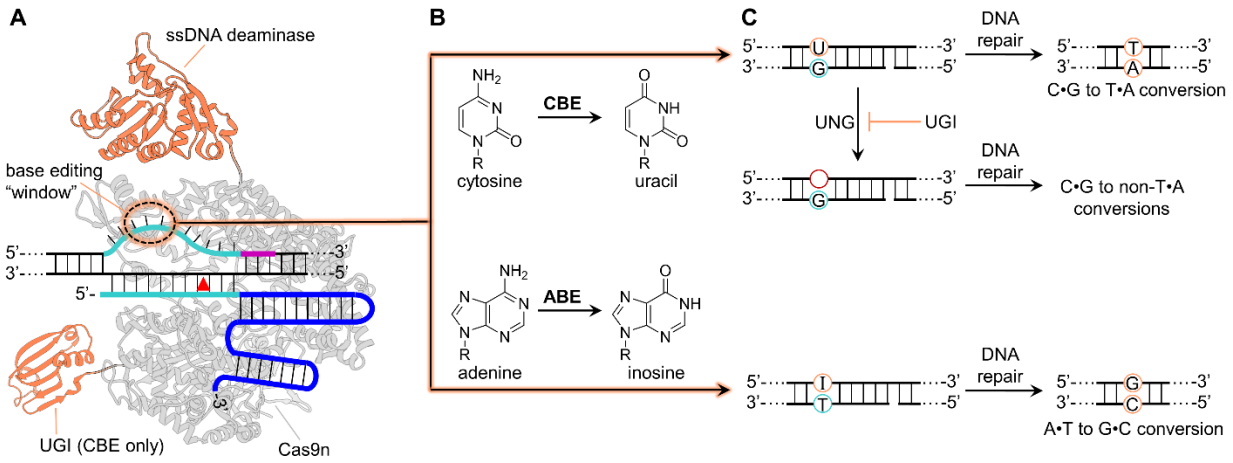

Figure 3. Schematic of base editing technologies for genome editing. **A**) General base editor (BE) architecture. nCas9 (colored light gray, PDB: 5Y36) forms an RNP complex with the gRNA. The 5' end of the gRNA (the spacer, colored cyan), can be altered by researchers to be complimentary to any target DNA sequence within the genome that is next to a PAM sequence (purple). Upon R-loop formation, a small window (referred to as the base editing window) of ~5 nucleotides on the ssDNA are exposed and readily accessible for modification by a ssDNA-specific deaminase enzyme (orange, PDB: 6X91) that is directly tethered to nCas9. **B**) The nucleobase deamination chemistries that are catalyzed by the two types of BEs. In cytosine base editing, ssDNA cytidine deaminase enzymes will catalyze the deamination of any cytosine nucleobases within the base editing window into uracils (which have the base pairing properties of thymine). Analogously, in adenine base editing, ssDNA adenosine deaminase enzymes will catalyze the deamination of any adenine nucleobases into inosines (which have the base pairing properties of guanine). **C**) During cytosine base editing, subsequent DNA replication or repair of the resulting U•G intermediate results in the permanent introduction of a C•G bp to T•A. Due to the high efficiency of uracil excision by the cell's endogenous uracil-N-glycosylase (UNG) enzymes, C•G to non-T•A conversions can occur from an abasic site intermediate. To combat this, a short peptide that binds and inactivates UNG (called UGI; shown in light orange below the nCas9 in part **A,** PDB:3WDG) is fused to the end of nCas9 in CBEs. During adenine base editing, the resulting I•T intermediate is permanently converted into a G•C base pair after DNA replication or repair, resulting in a permanent introduction of a A•T bp to G•C. **Note**, with both BEs, the DNA strand opposite from the nucleobase modification is nicked by nCas9 to promote use of the intermediate-containing strands (U- and I-containing strands) as templates during DNA repair.

Chapter 2

Optimized Protocols on Using Base Editing Technology in Mammalian Cells To Produce Isogenic Cell Lines

2.1 Introduction

Progress in next-generation sequencing (NGS) technologies has streamlined the detection of human genetic variations,[51] and the identification of clinically actionable genes and pathogenic mutations has transformed precision medicine. However, only a small fraction of identified human genetic variants have been assigned a clinical classification. The Genome Aggregation Database has documented 787 million genetic variants (96% of which are single nucleotide variants, or SNVs),[52–54] but less than 0.5% possess clinical interpretations in the ClinVar database.[55] Furthermore, 99% of these identified variants are rare or unique to specific ethnic populations, making it particularly challenging to predict their functional impact through genome-wide association studies (GWAS) or computational methods.[56–58] In particular, mutations in genes encoding for DNA repair proteins are linked to such diseases as cancer, premature aging, immune deficiencies, and neurodegenerative disorders,[59–61] highlighting the importance of functionally

characterizing genetic variants and obtaining mechanistic insight into disease etiology and progression.

Base editing technologies enable the introduction of point mutations at targeted genomic sites in mammalian cells, with higher efficiency and precision than traditional genome-editing methods that use DNA double-strand breaks, such as Double-Stranded DNA Break (DSB)-reliant technology such as, zinc finger nucleases (ZFNs), transcription-activator-like effector nucleases (TALENs), and the clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (CRISPR-Cas9) system. This allows the generation of single-nucleotide-variant isogenic cell lines (i.e., cell lines whose genomic sequences differ from each other only at a single, edited nucleotide) in a more time- and resource-effective manner. These single-nucleotide-variant clonal cell lines represent a powerful tool with which to assess the functional role of genetic variants in a native cellular context. Base editing can therefore facilitate genotype-to-phenotype studies in a controlled laboratory setting, with applications in both basic research and clinical applications. Here, I provide optimized protocols (including experimental design, methods, and analyses, with the steps provided in **Figure 4**) to design base-editing constructs, transfect adherent cells, quantify base-editing efficiencies in bulk, and generate single-nucleotide-variant clonal cell lines. For reading consideration, the materials for each protocol, which are excluding in this thesis for the sake of simplicity, can be found in the full publication: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpmb.129.

2.2 Strategic Planning

The optimal BE:gRNA combination will change based on the experimental goal and sequence surrounding the target base. While selecting a base editor, gRNA spacer, and suitable PAM, there are several considerations that should be reviewed before starting. These include

sequence context surrounding the target base, PAM and Cas variant selection, and acceptable levels of indels or off-target editing.

A particular challenge that many researchers encounter is bystander editing, as described in chapter 1. However, bystander editing can be minimized by using a particular deaminase variant (such as one with a shifted editing window or strict sequence preference) and careful PAM selection. The nucleotides flanking the target base—especially the 5' nucleotide for CBEs—will also influence deaminase selection (see reference [62] for a detailed analysis of deaminase sequence preferences). Finally, while certain experiments require no detectable indels and off-target editing (where high-fidelity Cas and deaminase variants can be used), others will benefit from using BEs with the highest on-target efficiency. As base editing systems are modular, we recommend testing multiple BE:gRNA options initially (through Protocol 3), and then selecting the combination that produces the best editing profile for isogenic cell line generation or downstream assays.

An additional matter to consider before beginning is the fact that immortalized cell lines may contain genetic variation or mutations in the target gene compared to the human reference genome; we thus recommend sequencing the locus of interest in your cell line before beginning. Within this protocol, we have included details on how to do this: see Protocol 1 step 6A for primer design, Protocol 2 for "Harvesting Genomic DNA", and Protocol 3 for sequencing the target site. Additionally, be sure to plan proper negative controls when generating SNV-containing isogenic cell lines.

2.3 Protocol 1: Design and Production of Plasmids for Base Editing Experiments

Although there are many delivery systems to introduce base editors into cells, during my Ph.D. work, we optimized the use of plasmid-based base editor delivery. These experiments rely on the preparation of two high-quality plasmids: one expresses gRNA that designates the target

genomic location, and the other expresses the BE protein component (Cas9-deaminase fusion). Accurate and efficient base editing relies on careful design of the most appropriate BE:gRNA combination. Central to this, is the identification of a PAM that enables the Cas protein to bind in a manner that only positions the target C or A within the base editing window. We recommend the use of BEs that incorporate the engineered SpCas9-NG for minimal targeting restrictions with the smallest reduction in on-target activity in mammalian cells.[63] BEs that incorporate other engineered variants, Cas9 homologs, and Cas12a orthologs, such as SpRY, Sau/Sauri/Spy-mac, and Lb/As/Aa offer additional PAM options but may have reduced on-target efficiencies at certain genomic loci or broadened editing windows, leading to bystander mutations (see the Critical Parameters section). The observed editing efficiency will depend on the context of the genomic site and base editor used. We have listed recommended CBE and ABE variants in **Table 3** and discuss alternatives in Critical Parameters. Of note, optimized "BE-max" plasmids contain GenScript human codon optimization for maximal expression levels, bipartite NLSs on both termini for enhanced nuclear import, and an optional bicistronic enhanced green fluorescent protein (EGFP) for assessment of transfection efficiency and BE expression.[24,25] The gRNA expression vector should contain a U6 promoter driving expression of an *S. pyogenes* Cas9 gRNA (e.g. Addgene #47511). The initial spacer sequence is irrelevant because site-directed mutagenesis will be used in this protocol to replace the gRNA spacer using a 5'overhang on the forward primer.

This protocol outlines the steps to properly design gRNAs for BE experiments, replace the spacer sequence in a gRNA expression vector, and prepare the appropriate plasmids for mammalian cell transfection. Additionally, Protocol 1 details primer design for the amplification

of a target genomic locus to quantify targeted SNV introduction by Sanger or next-generation sequencing.

NOTE: If multiple BE:gRNA options are available, we recommend experimentally validating all of them to identify the optimal combination that maximizes target C or A editing and minimizes bystander editing.

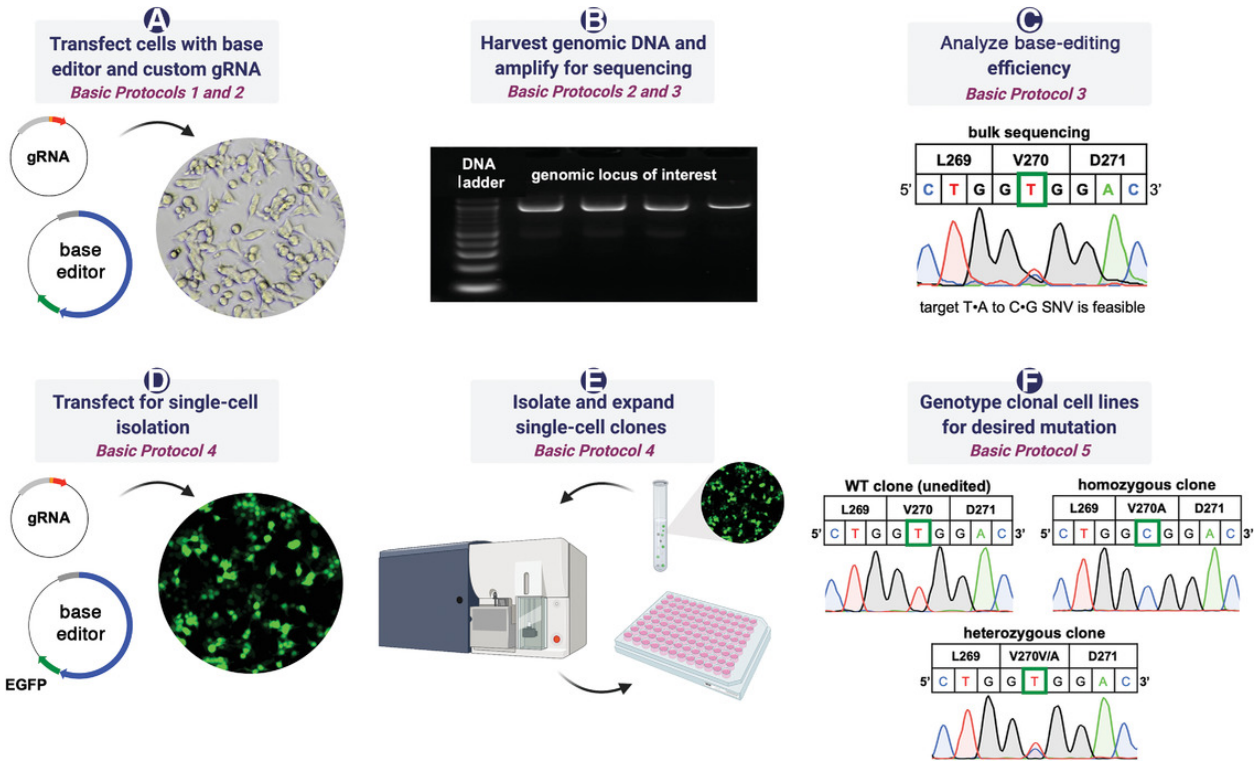

Figure 4. Generating isogenic cell lines harboring a target SNV, using the MUTYH V270A variant as an example. (**A**) Transfection-quality plasmids encoding a custom-designed gRNA and base editor (BE) are transfected into mammalian cells. (**B**) 3 days post-transfection, genomic DNA is harvested, and the genomic locus of interest is amplified and then prepared for sequencing. (**C**) The resulting PCR products are sequenced to assess the feasibility of generating an isogenic cell harboring the target SNV. Shown is bulk sequencing of the total cell population transfected with a BE:gRNA combination to introduce the V270A variant in MUTYH. The sequencing chromatogram shows base editing activity on the target base. (**D**) Following confirmation of >10% editing efficiency, step A is repeated. The BE plasmid also encodes for an EGFP fluorescent marker, allowing for the use of FACS to isolate individual cells expressing a base editor. (**E**) FACS is utilized to sort single, EGFP-positive cells into individual wells of a 96-well plate containing culture media. (**F**) Single cell-derived colonies clonally expand for 1-2 weeks. Sequencing the resulting clonal cell lines confirms the generation of an isogenic cell lines harboring the target SNV. Data from three individual isogenic cell lines of the V270A MUTYH variant are shown, containing wild-type, homozygous, and heterozygous genotypes. The target nucleotide is outlined in green.

2.3.A Protocol Steps

Design gRNA

1. Import genomic locus of interest into sequence viewing software of choice. We frequently use Benchling (https://www.benchling.com).

   Be sure to import in the sense orientation using the reference genome ('build') and isoform that matches the SNV annotation.

2. Identify the target nucleotide (C or A) and find an NG PAM exactly 12 to 16 nucleotides downstream (i.e. towards the 3' end of the strand containing the target nucleotide, see **Figure 5A** for an example). The target can be on either strand, as long as there is a suitable PAM on the same strand. Annotate the protospacer. Alternately, free software programs such as the Benchling wizard , BE-designer , and BE-Hive can be used for automated gRNA design. As most automated programs lack flexibility in deaminase and Cas variants, we recommend manually designing your protospacers (**Figure 5A**) and checking predicted editing efficiencies in BE-Hive .

   If bystander edits are unavoidable, choose the protospacer with bystander nucleotides farthest from the center of the window (position 6) or one that will incorporate silent mutations (as long as it is not a splice site).

3. If the protospacer does not start with guanine, add a 5'G to create a 21-nucleotide spacer in the gRNA. For protospacers that already start with a 5'G (such as the example V270 protospacer), move to the next step.

   Adding a guanine drastically increases gRNA transcription from the U6 promoter, yielding high editing efficiencies, despite a possible 5' mismatch.

4. Design a custom forward primer with the sequence 5'-[N_{20-21}]GTTTTAGAGCTAGAAATAGCA-3', replacing the [N_{20-21}] portion with the protospacer sequence from the previous step (e.g. **Table 4**, #2 for MUTYH V270). This will be used with the universal reverse primer (**Table 4**, #1). Analyze this primer pair for homo- and hetero-dimers using the IDT oligo analyzer (https://www.idtdna.com/pages/tools/oligoanalyzer). Note that amplification may be difficult using primers with a $\Delta G < -10$ kcal/mol.

    *The custom primer sequence should only include the protospacer sequence without the PAM. In base editing experiments, the spacer sequence will always be identical to the protospacer and contain the target A or C because direct nucleobase modification occurs on ssDNA of the canonical "non-target strand" that is not duplexed with gRNA. See **Figure 5A** for an example. It is also recommended to include a non-targeting gRNA as a control (see **Table 4** primer #6).*

5. Order the primers from IDT or another manufacturer of choice for custom gRNA construction via site-directed mutagenesis of the spacer or "around-the-horn" cloning.

Design primers for amplification of genomic locus

6. A. Use a primer design tool (e.g. Primer3 or other) to generate a pair of primers that amplify a ~1 kilobase (kb) region containing the protospacer, for Sanger sequencing. See **Figure 5A** and **Table 4**, primers #11-12 for V270 example using the Benchling Primer3 design wizard.

    These primers should ideally be 18-24 base pairs (bp) long, and have 40-60% GC content, and melting temperatures of 60-65˚C. To ensure that at least one of

these primers is appropriately positioned to yield maximum quality sequencing data of the protospacer region of the amplicon, the target sequencing region should be at least 50 bp away from the primer used for sequencing.

B. Alternatively, design primers for Illumina next-generation sequencing (NGS). This requires specific adapter sequences be added to both the forward and reverse primers, and a much shorter amplicon. Design a primer pair that amplifies 200-250 bp of DNA (for a 300-cycle NGS run; if a 150-cycle NGS run will be used, the amplicon should be 100-125bp long) with the same parameters as above. Then, add the adapter sequences #13 and #14 from **Table 4** onto the 5'end of the forward and reverse primers, respectively. For the MUTYH V270 example, the Primer3 Wizard on Benchling was used to identify 22 nt PCR primers that amplify a 249 bp region around the protospacer and the Illumina NGS adapters were added to create primers #15 and 16 in **Table 4.** Order these as the round 1 PCR primers. A second round of amplification is required to provide the samples with P5/P7 tails and unique barcodes. Order round 2 PCR primers based on the Illumina barcoding system to be used during library preparation. See Alternate Protocol 1 for more information.

Clone new spacer into gRNA expression plasmid via site-directed mutagenesis

7. In 0.2 mL PCR tubes, 5' phosphorylate each primer (forward and reverse, from step #5 above) individually in separate 20 µL reactions by combining the following (in the order stated):
   - 15 µL nuclease-free water
   - 2 µL primer (100 µM stock)
   - 2 µL T4 DNA ligase buffer

- 1 µL T4 polynucleotide kinase

Mix well and quickly spin to collect sample. The 20 µL of phosphorylated primer is enough for 8 PCR reactions—scale up if the primer is needed for more than 8 gRNAs (i.e. the universal reverse).

8. Place the reaction tube in a thermocycler and run the following program:
   - 37°C for 20 min
   - 95°C for 5 min
   - Hold at 12°C
9. In a PCR tube on ice, combine the following reagents for a 50 µL reaction (in the order stated):
   - X µL nuclease-free water (fill to a total volume of 50 µL)
   - 10 µL Phusion HF buffer
   - 1 µL dNTP mix
   - 2.5 µL FWD phosphorylated primer (10 µM from previous step)
   - 2.5 µL REV phosphorylated primer (10 µM from previous step)
   - ~1 ng DNA template: gRNA expression plasmid with different spacer sequence, (Addgene # 47511, noted above)
   - 0.5 µL Phusion polymerase
10. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

1 cycle:	30 sec	98°C	(initial denaturation)

~35 cycles:	10 sec	98°C	(denaturation)
	20 sec	65°C	(annealing)
	35-45 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
	hold	12°C	(hold)

11. Run a 1 uL aliquot out on an agarose gel (1-1.5%) with a 1kb DNA ladder to check for a PCR product that is ~2.3 kb long (see **Figure 5B** for an example).

    If there is no PCR product, use the GC buffer, or additives such as DMSO or formamide (5%). If there are unwanted products, increasing the annealing temperature or decreasing the primer concentration can improve specificity. Gel extracting the band of interest is also possible during step 13.

12. Add 1 µL DpnI to the PCR product, mix, and spin. Incubate at 37°C for 1 hr.
13. Purify PCR products using the Qiagen PCR clean-up protocol.

    Using a vacuum manifold followed by thoroughly drying the column in a centrifuge is suggested.

14. Quantify the concentration of the PCR product using a Nanodrop. Then, prepare a 20 uL ligation reaction to circularize the linear PCR product, which contains the new spacer sequence and 5' phosphate groups. In a PCR tube, combine:

    ~50 µg PCR product after clean-up (from previous step)

    X µL nuclease-free water (fill to a total volume of 20 µL)

    10 µL 2X Quick Ligase buffer

    1 µL Quick Ligase

15. Mix, quickly spin, and incubate at room temperature for 5-10 min.
16. Using sterile technique, transform ~5 μL of the ligated product into competent bacterial cells of choice. Plate 10-50 uL cells on an agar plate containing 100 mg/mL ampicillin or carbenicillin in a manner that will yield single colonies, and grow at 37°C overnight.

    The BE and gRNA plasmids we recommend confer ampicillin/carbenicillin resistance. Expect ~20-200 colonies, with the majority containing the desired spacer sequence replaced.
17. Pick 1-2 colonies per clone and grow in liquid LB or 2xYT media with 100 mg/mL ampicillin/carbenicillin until saturated.

    Using cells that double rapidly, such as Mach1, allows for short growth periods (~6 hrs). Inoculating colonies in 1-3 mL of media allows for sequencing on the same day. Save the plates airtight at 4˚C (using Parafilm) and mark the colonies screened.
18. Mini-prep saturated cultures.

    We recommend saving the original culture tubes with un-lysed bacteria at 4˚C to create starter cultures for correct clones
19. Quantify the prep using a Nanodrop, and then run an aliquot on an agarose gel (1.5%) to check the quality and size of the plasmid (**Figure 5C** for an example).

    The supercoiled gRNA plasmid runs around ~1.5 kb on a linearized ladder.
20. Send in preps for Sanger sequencing using gRNA sequencing primer #7 in **Table 4** (U6), to confirm the presence of the new spacer sequence.

*Sequence the entire backbone using primers #8-10 in **Table 4** (at this step or after step 26, the midi-prep) as mutations may occur during PCR. Note that the universal U6 promoter primer that many sequencing companies use anneals very close to the spacer region and may not provide high-quality sequencing data of this region of the plasmid.*

Prepare High Quality, Endotoxin-free gRNA and BE plasmid DNA

21. Once the base editor plasmids are obtained from Addgene, streak the surface of the bacterial stab onto an LB or 2xYT plate containing 100 mg/mL ampicillin or carbenicillin. Use sterile technique and streak for single colonies. Incubate at 37°C overnight (see https://www.addgene.org/recipient-instructions/myplasmid/).
22. In the morning, grow 0.5 mL starter cultures of LB or 2xYT media with 100 mg/mL ampicillin or carbenicillin for each BE and gRNA plasmid required, at 37°C for ~6 hr or until saturated. For BEs, inoculate a single colony from plate in the previous step. For gRNAs, add antibiotic-containing media to culture tube saved at step 18 or inoculate from the original plate to avoid re-transforming.
23. Dilute into 50 mL for midi-prep (or 200 mL for maxi-prep) cultures for overnight growth.
24. Once saturated, remove 0.5 mL of culture and combine with 0.5 mL 50% glycerol in a 2 mL cryogenic vial. Freeze stock at -80˚C.
25. Use the remainder of the culture to prepare endotoxin-free midi- or maxi-preps based on the manufacturer's protocol.

If solution is cloudy after the syringe filter step of the midi- or maxi-prep, centrifuge the 50 mL tube for 10 min at 4,000 rcf and proceed with supernatant. Using a vacuum manifold is suggested. Drying the wash buffer from the column is critical (spin in microcentrifuge for 2-3 min at ≥12,000 rcf). At the elution step, warming the buffer to 55°C and incubating 3-5 min increases DNA yield.

26. Quantify the plasmid sample using a Nanodrop and run an aliquot out on an agarose gel to confirm the quality of midi- or maxi-prep (**Figure 5C**).

    BEs plasmids are high-copy while the gRNA plasmid is medium- to low-copy.

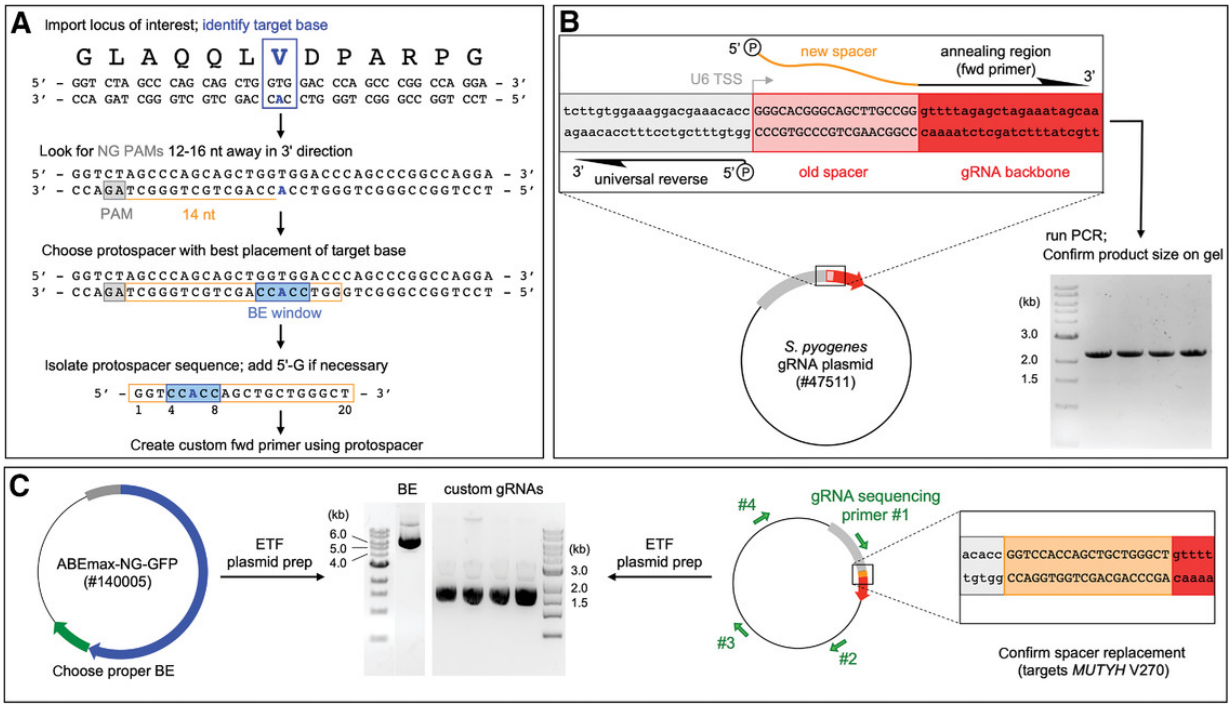

Figure 5. Overview of Protocol 1. (A) To design a custom gRNA (protocol steps 1-5), first import the target locus into a viewing software of choice, being mindful of the isoform and reference genome used to annotate the SNV of interest (shown is the MUTYH gene sequence surrounding residue 270, using the hg38 reference genome). Identify the target amino acid and nucleotide (blue, V270 codon; target adenine is on the template strand), and identify potential PAMs that are positioned 12-16 nt downstream, in the 3′ direction, on the same DNA strand as the target nucleotide. Select the protospacer that positions the target nucleotide closest to the center of the editing window (position 6) or that positions potential bystander edits outside of the editing window. The gray "AG" PAM optimally positions the target base, while simultaneously pushing a potential bystander base to position 9, outside the canonical editing window. Order universal reverse and custom forward primers to replace the spacer sequence on a compatible gRNA expression plasmid (S. pyogenes; for example, Addgene cat. no. 47511). (B) Schematic showing the site-directed mutagenesis or around-the-horn cloning strategy for gRNA spacer replacement, and agarose gel showing confirmation of a PCR product band with the correct size (steps 9-11). (C) Preparation of endotoxin-free (ETF) midi- or maxipreps of gRNA and BE plasmids (steps 21-26), with a representative agarose gel showing proper size and quality of supercoiled ETF plasmids. Additionally, in step 20, Sanger sequencing is used to verify the sequence of the full gRNA plasmid.

2.4 Protocol 2: Transfecting Adherent Cells and Harvesting Genomic DNA

Before proceeding to generate isogenic cell lines, it is highly recommended to validate the BE:gRNA combination(s) from Protocol 1. The most time- and resource-effective way to accomplish this is with an easy-to-transfect cell line, such as Human Embryonic Kidney 293 (HEK293) cells. While many different methods are available to deliver the BE and gRNA into

cells (such as viral transduction, mRNA transfection or electroporation, and ribonucleoprotein transfection or electroporation), the protocol outlined here utilizes commercial cationic lipid reagents to transfect cells with plasmids encoding the BE and gRNA. This protocol details how to deliver transfection-quality plasmids into HEK293 cells and extract the genomic DNA (gDNA) for downstream sequencing.

We utilize HEK293T cells for base editing experiments. HEK293T cells contain an endogenous copy of the SV40 large T antigen. As such, plasmids containing an SV40 origin of replication will be replicated in HEK293T cells, resulting in longer-term expression of recombinant proteins, which can cause higher levels of off-target editing with genome editing agents. Because of that, the plasmids used by the authors do not contain an SV40 origin of replication.

NOTE: All culture incubations should be performed in a humidified 37°C, 5% CO_2 (g) incubator. Use 10% FBS in DMEM with or without added pen/strep for all steps as indicated. All experimental work in this section must be done using sterile reagents and proper aseptic techniques.[64] Failure to do so, will result in possible contamination, which will disrupt downstream workflow and cause wasting of valuable laboratory resources.

2.4.A Protocol Steps

Transfecting HEK293T Cells with BE and gRNA

1. Plate HEK293T cells such that they will be 70-85% confluent upon transfection in medium without antibiotics (see **Figure 6A** for example).

    For a 48-well plate, 100,000 cells/well (in a well consisting of 250 µL of 10% FBS in DMEM culture medium , or a cell solution of 4.0 x 10^5 cell/mL) is usually ready to transfect 4 hours after plating. 50,000 cells/well (in 250 µL of medium per well,

or a cell solution of 2.0 x 10^5 cell/mL) is usually ready to transfect 16 hours after plating.

2. Warm Opti-MEM to room temperature prior to adding Lipofectamine 2000.
3. For each transfection, 200 ng of gRNA plasmid and 800 ng of BE plasmid is needed if done in a 48-well plate format. Include a negative control sample that contains a non-targeting gRNA or no gRNA. For different transfection formats, the total amount of plasmid and Lipofectamine will need to be scaled up or down. Prepare two different tubes per transfection:
    a. Tube A: add 200 ng of gRNA plasmid and 800 ng of BE plasmid. Dilute the DNA with Opti-MEM to a total volume of 12.5 µL.
    b. Tube B: add 1.5 µL of Lipofectamine 2000 and 11 µL of Opti-MEM.

    In some instances, you may be required to generate multiple edits in parallel. In this case, follow step 3a and 3b for each BE:gRNA combination.
4. Mix the contents of tube A and tube B and incubate the mixture for 15 minutes at room temperature prior to transfection.
5. Carefully add the 25 µL mixture dropwise to the cells and place back into the incubator.

    Be careful not to touch the bottom of the well with the pipet tip to avoid disturbing the cell monolayer.
6. Monitor the cells under a fluorescence microscope after 24 hr. EGFP-positive cells should be observed evenly across the surface of the plate (see **Figure 6A**). Transfection efficiency should be at least 70% (ideally over 90%). If not, please refer to the Troubleshooting section.

Lipofectamine can be toxic to the cells, so we recommend checking on the health status of the cells on the microscope every day. If cells appear healthy, changing medium after 48 hr is also sufficient. If high cell toxicity is observed, as seen in ***Figure 6B****, change to fresh DMEM medium prior to 24 hr.*

7. After 24 hr, aspirate the old medium and gently rinse cells with 150 µL of PBS *Be careful, as roughly adding the PBS will result in dislodging adherent cells.*
8. Gently aspirate off the PBS and add 250 µL of fresh pre-warmed culture medium (10% FBS in DMEM with 1% Pen/Strep).
9. Return to incubator for two additional days after media is replaced.

Harvesting Genomic DNA

10. Prepare cell lysis solution prior to taking cells from step 9 out of the incubator. 100 µL of cell lysis solution will be required per well when done in a 48-well plate format. As such, prepare a master mix solution containing the following: 10 mM Tris (pH 7.5), 0.05% SDS, 25 µg/mL proteinase K.

    For example, a 1-mL master mix cell lysis solution will contain the following: 983.75 µL of nuclease-free water, 10 µL of 1M Tris HCL stock solution (pH 8.0), 5 µL of 10% SDS stock solution, and 1.25 µL of Proteinase K enzyme. The Tris and SDS may be prepared in advance, but the Proteinase K must be added immediately prior to use.

11. Aspirate 250 µL of the old medium and gently rinse the cells with 150 µL of PBS. Carefully aspirate off the PBS.
12. Add 100 µL of cell lysis solution prepared in step 10 to each well and wait for 3-5 minutes, or until cells are completely dislodged from the bottom of the well.

13. Collect the cell lysate into PCR tubes.

    Slightly tilt the 48-well plate. Prior to contacting the cell lysate with the pipette tip, displace 100 µL of air. Collect 100 µL of cell lysate on the first suction and transfer to PCR tubes. Pipetting the cell lysate up and down, especially vigorously, will result in air bubbles that will reduce the amount of harvested genomic DNA. Be advised—the viscosity of genomic DNA increases the chance for cross-contamination of samples to occur.

14. Incubate the cell lysis reaction in the thermocycler:

    - 37°C for 1 hr 30 min
    - 80°C for 30 min
    - Hold at 4°C

    Keep cell lysate at 4˚C and avoid freeze/thaw cycles. Amplification works best on fresh gDNA, which can be used directly or diluted (1:5-1:50) depending on the cell density before harvesting. After incubation in the thermocycler, cell lysate normally appears slightly cloudy and will be more viscous when higher amounts of gDNA are present. Optionally, a Nanodrop can be used to estimate gDNA concentration (blank with lysis solution), though impurities in the lysate reduce the reading's accuracy.

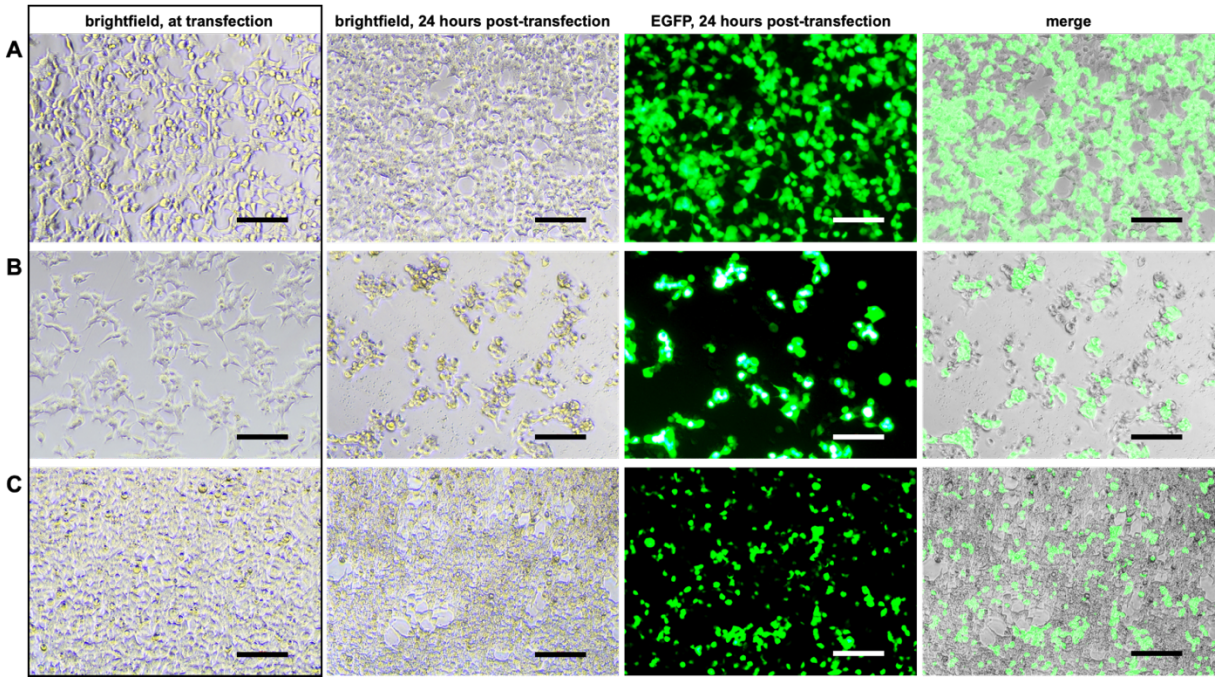

Figure 6. Representative images of successful and unsuccessful transfection experiments (Protocol 2, steps 1-9). Shown are brightfield images of HEK293T cell confluency at the time of transfection (left most column, in box), as well as brightfield (middle left column), GFP fluorescence (middle right column), and merged (right most column) images of the same cells 24 hours post-transfection with gRNA and BE plasmids to generate the V270A MUTYH variant. Shown are cells that were transfected at **row (A) –** ideal confluency (70-85%), resulting in high transfection efficiency and high cell viability across the surface of the plate, **row (B) –** under confluency (<70%), resulting in high transfection efficiency but low cell viability, and **row (C) –** over confluency (>85%), resulting in low transfection efficiency but high cell viability. The latter two conditions are illustrative of poor base editing experiments. Scale bars = 166 µm.

2.5 Protocol 3: Genotyping Harvested Cells using Sanger Sequencing

After transfection with the BE:gRNA combination(s), base editing efficiency needs to be assessed to determine the feasibility of generating an isogenic cell line harboring the target SNV. Generally, an editing efficiency >10% is sufficient to proceed (**Figures 4C; 7A**). However, base editing efficiencies of >30% greatly increase the chances of generating all genotypes.

Sanger sequencing is a cost-effective alternative to next-generation sequencing (Alternate Protocol 1) and is usually sufficient to determine the feasibility of generating cell lines with the SNV of interest.

2.5.A Protocol Steps

PCR of genomic DNA

1. For each sample, combine the following reagents (in the order stated) for a 50 µL PCR reaction to amplify the genomic locus of interest:
    - X µL Nuclease-free water (fill to a total volume of 50 µL)
    - 10 µL GC Buffer
    - 1 µL dNTP mix (10 mM each)
    - 0.5 µL 10µM forward primer (in this example, #11 **Table 4**)
    - 0.5 µL 10µM reverse primer (in this example, #12 **Table 4**)
    - 1.5 µL 100% DMSO
    - 0.5 µL cell lysate (gDNA) from step 14 of Protocol 2
    - 0.5 µL of Phusion DNA polymerase

It is imperative to include a negative control sample every time amplification from gDNA is performed. In this sample, 0.5 µL of water is added instead of the cell lysis solution (gDNA) to control for reagent contamination. We suggest making a master mix, then adding the cell lysate and polymerase to individual aliquots.

2. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

1 cycle:	30 sec	98°C	(initial denaturation)
25 to 34 cycles*:	10 sec	98°C	(denaturation)
	25 sec	65°C**	(annealing)
	10 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
1 cycle:	∞	12°C	(hold)

***The annealing temperature is specific to the primers chosen. In this protocol, it is recommended that researchers use an automated primer design tool (Step 6A of Protocol 1). Using the Benchling Primer3 design wizard, the default optimal annealing temperature is 65°C. If the primer pair's annealing temperature is not 65°C, please adjust accordingly.*

3. Run a 5 µL aliquot of each PCR reaction on a 2-3% agarose gel using a 100 bp ladder for size comparison (see **Figure 5B** for an example).

   If you do not get the correct-size product bands, please refer to the Troubleshooting section.

4. After confirming the presence of a correctly-sized PCR product, prepare the samples for sequencing. Genewiz (https://www.genewiz.com/en) offers two Sanger sequencing options, but this may depend on the sequencing vendor of choice:

    a. Sending purified PCR product: Purify the product with QIAquick PCR purification kit following the manufacturer's instructions. Measure the DNA concentration after PCR clean-up of each sample using a Nanodrop. Prepare the appropriate amount of DNA along with the sequencing primer (we recommend using either the forward or reverse primer from the PCR reaction that is positioned at least 50-75 bp away from the target base of interest) in the same tube according to company protocol.

    b. Sending unpurified PCR product: unpurified PCR samples sent for Sanger sequencing will undergo an enzymatic PCR purification protocol. Provide the appropriate amount of the successful PCR reaction in one tube according to sequencing company guidelines. Provide the appropriate amount of sequencing primer (we recommend using either the forward or reverse primer from the PCR reaction, whichever one is at least 75-bp away from the target base of interest) in a separate tube. We recommend submitting a picture of the labelled gel as well to the sequencing company as this can help with quantification and troubleshooting, if necessary.

    Sending unpurified PCR product (step 4b.) usually results in better sequencing coverage quality and thus, an easier to interpret chromatogram, but costs extra.

2.5.B Using EditR to Quantify Base Editing Efficiencies

Quantifying base editing efficiency of the target nucleotide is needed either (a) to determine the feasibility of generating a cell line with a given SNV as seen in Protocol 3, or (b) to confirm the introduction of either a homozygous edit or heterozygous edit into mammalian cells as seen in Protocol 5. In both instances, we recommend measuring the base editing efficiency of the total harvested cells in bulk using a simple and publicly available program, termed "EditR." (https://moriaritylab.shinyapps.io/editr_v10/; . EditR is a free online tool or desktop application which requires an .ab1 Sanger sequencing file of the potentially edited region (~300-700bp) and the gRNA protospacer sequence (~20bp) to predict where a base edit occurred. Once the .ab1 file and DNA protospacer sequence are correctly uploaded, EditR generates a plot displaying editing efficiencies at each base within the protospacer (**Figure 7D and 7E**).

NOTE: If we state, "an editing efficiency of ~10%", this means that ~10% of the cells of the total harvested cell population have a successful edit at the target base. This is will be indicated by overlapping chromatogram peaks at the targeted nucleotide. In this hypothetical example, it would show ~90% unedited base and ~10% edited base in the sequencing chromatogram. Using MUTYH V270 as an example, ***Figure 7A*** *and* ***7D*** *shows a chromatogram where the base edit efficiency is reported as 38% (P = .01) by EditR. It is important to note that a poor-quality Sanger sequencing chromatograph (such as that shown in* ***Figure 7C****) may produce an EditR output file that falsely shows editing at the target base (see* ***Figure 7E****). It is, therefore, important to confirm that non-A/C bases in the protospacer display less than ~7% editing.*

The following steps have been adapted from the online protocol found at https://moriaritylab.shinyapps.io/editr_v10/:

5. Upload your .ab1 file of the sequenced region.

6. Enter the gRNA sequence protospacer sequence.

    - If your gRNA is antisense to the .ab1 file, check the “Guide sequence is reverse complement” box.

7. Click the “Predicted Editing” tab.

8. Examine the gRNA protospacer chromatogram and underlying tile plot to determine if base editing occurred. Observing >10% editing efficiency at the target base (as indicated by a double peak at the targeted nucleotide showing <90% unedited base and >10% edited base in the sequencing chromatogram) is sufficient to proceed (**Figure 7A**).

    - All colored tiles represent base calls that are deemed significant, i.e. if there are multiple colored tiles under a single base call, base editing likely occurred.

9. If you wish to download a report of the operations performed on your data, click the "Download Report" tab on the top of the page.

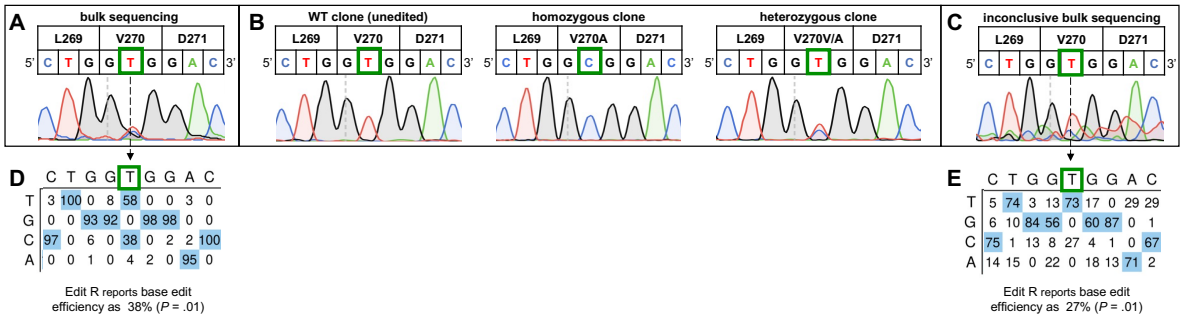

Figure 7. Representative Sanger sequencing traces from genotyping of cells in bulk and after clonal expansion (Protocols 3 and 5). (**A**) Sanger sequencing trace of the bulk population of total cells after transfecting with a BE:gRNA combination to introduce the MUTYH V270A mutation. (**B**) Following re-transfection and clonal expansion, single cell-derived colonies potentially harboring the MUTYH V270A variant were then sequenced. Sanger sequencing chromatograms show the generation of three individual isogenic cell lines, in this instance, with genotypes of wild-type (left), homozygous (middle), and heterozygous (right). (**C**) A low-quality chromatogram with a high degree of background noise obtained after the bulk population of total cells were sequenced via Sanger sequencing. (**D**) Analysis of the .ab1 file of the bulk population of total cells, as described in panel **A**. EditR quantifies a base editing efficiency of 38% (P = .01) The observed base editing efficiency at the target nucleotide indicates the feasibility of generating the targeted V270A variants, since the editing efficiency is >10%. (**E**) Analysis of the .ab1 file of the bulk population of total cells with a degree of background noise, as described in panel **C.** EditR quantifies a base editing efficiency of 27% (P = .01), which may mislead the researcher into incorrectly concluding a successful initial base editing experiment. Please refer to Basic Protocol 3 for information on EditR.

2.6 Alternate Protocol 1: Next-Generation Sequencing to Quantify Base Editing

Some base editing experiments require rigorous quantification to accurately determine absolute editing efficiencies or to deconvolute editing patterns in bulk. Next-generation sequencing enables the researcher to quantify individual edited alleles, which can be helpful to establish the frequency of bystander mutations in bulk before isolating single cells. NGS is also desirable when genotyping edited cell lines at target sites with increased copy number (due to gene duplication) or if robust quantification of indels is needed. In these cases, the harvested genomic DNA samples from Protocol 2 or Protocol 5 should be amplified and prepared for Illumina NGS. Targeted amplicon sequencing for genome editing experiments has been previously described in depth by Veeranagouda and coworkers, as well as by Yang and coworkers. [65,66] We recommend quantifying base editing using a 300-cycle, paired-end NGS run with a 200-250 bp amplicon. However, the amplicon length can be easily altered for other types of sequencing runs. Primers for the initial amplification (round 1) should be designed with the proper adapter sequence and distance from the target site, as described in Protocol 1, step 6B and **Figure 5A**. Then, sample barcodes (for de-multiplexing) and Illumina-specific P5/P7 tail sequences will be added during the round 2 PCR. Once the data is acquired, we suggest performing data analysis with the free CRISPResso2 software using the batch mode and base editor output.[67]

2.6.AProtocol Steps

Next-Generation Sequencing

1. To amplify the genomic locus of interest, create a 25 µL round 1 (rd1) PCR reaction for each gDNA sample (BE:gRNA combination) and a negative control containing water. For each sample, combine the following reagents (in the order stated):
    - X µL Nuclease-free water (fill to a total volume of 25 µL)
    - 5 µL GC Buffer
    - 0.5 µL 10 mM dNTPs
    - 0.5 µL 10µM forward NGS rd1 primer (for the MUTYH V270 example, primer #15 **Table 4**)
    - 0.5 µL 10µM reverse NGS rd1 primer (for the MUTYH V270 example, primer #16 **Table 4**)
    - 0.75 µL 100% DMSO
    - 0.5 µL cell lysate (gDNA)
    - 0.25 µL of Phusion DNA polymerase

    It is imperative to include a negative control sample every time amplification from gDNA is performed. In this sample, 0.5 µL of water is added instead of the cell lysis solution, to control for gDNA contamination. We suggest making a master mix, then adding the cell lysate and polymerase to individual aliquots.
2. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

1 cycle:	60 sec	98°C	(initial denaturation)

22 to 28 cycles*:	10 sec	98°C	(denaturation)
	20 sec	65°C	(annealing)
	10 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
1 cycle:	∞	12°C	(hold)

**To avoid PCR bias, use the minimum number of cycles that provides robust amplification. This may require optimization for each target locus (see also "Quantification of base editing efficiency in bulk cells" in the Critical Parameters section).*

3. Run a 1-2 uL aliquot of each PCR reaction on a 2% agarose gel using a 100 bp ladder.

   NGS round 1 adapter sequences add 66 bp to the length of your amplicon. If you do not get the correct-size product bands, please refer to the Troubleshooting section.

4. Round 2 (rd2) PCR adds a unique barcode designated by an 8 nt sequence in each of the primers: A1 in this example (Fwd-A/Rev-1). For each rd1 sample, create a 25 µL rd2 PCR reaction by combining the following reagents (in the order stated):

   - X µL Nuclease-free water (fill to a total volume of 25 µL)
   - 5 µL GC Buffer
   - 0.5 µL 10 mM dNTPs
   - 0.5 µL 10µM forward NGS rd2 primer (custom barcode; **Table 5**)
   - 0.5 µL 10µM reverse NGS rd2 primer (custom barcode; **Table 5**)
   - 0.75 µL 100% DMSO

- 0.5 µL rd1 PCR product
- 0.25 µL of Phusion DNA polymerase

The negative control sample in this PCR should contain 0.5 µL of the round 1 negative control PCR product. Barcodes can be custom-generated and input into the sequencer, such as those provided in ***Table 5****, or ordered in a kit through Illumina (TG Nextera® XT Index Kit v2).*

5. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

1 cycle:	60 sec	98°C	(initial denaturation)
8 to 16 cycles*:	10 sec	98°C	(denaturation)
	20 sec	65°C	(annealing)
	10 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
1 cycle:	∞	12°C	(hold)

**Use the minimum number of cycles that provides robust amplification*.

6. Run a 1-2 uL aliquot of each PCR reaction on a 2% agarose gel using a 100 bp ladder (see **Figure 5C** for an example).

   NGS round 2 adapter sequences add 74 bp to the length of your rd1 PCR product (or a total of 140 bp to your original amplicon).

7. Pool rd2 PCR products from all samples together (e.g. no gRNA control and each BE:gRNA combination at the locus of interest). Perform a gel extraction on the pooled samples to eliminate lower molecular weight products that would decrease the quality of the NGS data.

We recommend doing a second PCR clean-up on the elution to remove any agarose (peak absorbance at 230nm) that remains after gel purification. A Nanodrop can be used to determine purity and roughly quantify concentration. Then, dilute into the quantification range for the next step.

8. Accurately quantify the DNA concentration of the libraries following the Qubit dsDNA HS assay kit protocol.

   Alternatively, the NEBNext Library Quant Kit for Illumina protocol describes quantification by qPCR.

9. Set up the NGS run according to Illumina protocols or submit to a sequencing core facility.

2.7 Protocol 4: Single Cell Isolation of Base Edited Cells using FACS

Following confirmation of >10% editing efficiency, the next step is to isolate single cells and clonally expand, to obtain isogenic cell lines harboring the SNV of interest. In this section, we describe two different methods for this: one that utilizes fluorescence-activated cell sorting (FACS) (this protocol), and one that utilizes dilution plating (Alternate Protocol 2). Using FACS followed by clonal expansion is preferred over dilution plating for two key reasons. Firstly, FACS allows for the discrimination between single cells and multiplets, which eliminates the chances of obtaining doublet cell-derived colonies (a frequent observation seen in dilution plating, **Figure 8E**). Secondly, FACS allows for the discrimination between untransfected and transfected cells when using plasmids with fluorescent markers, which is particularly important when working with cell lines with low transfection efficiencies. Dilution plating, however, is more cost-effective, gentler on the cells, and does not require specialized instrumentation. Dilution plating, therefore, may be preferred over FACS when working with sensitive cells or if the researcher does not have access to FACS instrumentation (or simply when working with

plasmids which do not possess fluorescent markers). This Protocol describes the process for transfecting and isolating single cells for clonal expansion.

NOTE: We used ABEmax-NG-GFP (Addgene #140005) to generate the V270A MUTYH variant used in this example. In this plasmid, although EGFP and ABE are both transcribed in a single mRNA transcript, they are translated into unique and separate proteins via the "self-cleaving" P2A linker. Therefore, FACS can be used to sort individual EGFP-positive cells (**Figure 4E**), which selects cells that are actively expressing the BE. There are also methods to select for cells with high BE activity (rather than simply expression).

2.7.AProtocol Steps

Transfect for Single Cell Isolation

1. Transfect the HEK293T cells with the BE:gRNA combination(s) that show desired editing activity in bulk, as described in Protocol 2 steps 1 to 8. Incubate the cells for three days post transfection.

   It is important to include a proper negative control sample (see Critical Parameters), such as a non-targeting gRNA or no gRNA.

2. 1-2 days after the transfections, prepare 96-well plates which contain 100 µL of culture medium in each well (50% FBS, 1% Pen/Strep). Generally, 2 plates per well in step 1 is sufficient to obtain cells containing the SNVs of interest. Place the plates into the tissue-culture incubator overnight.

   Single transfected cells from step 1 will be sorted into individual wells of the 96-well plate. For generating single nucleotide variant clonal cell lines, two 96-well plates per BE:gRNA combination containing DMEM with 50% FBS, 1% Pen/Strep

is used but other sorting culture medium is also commercially available. Preparing extra collection plates with medium is optional.

Preparation of the Cells for FACS

3. Three days post-transfection, aspirate 250 µL of the old medium from the transfected cells and gently rinse cells with 150 µL of PBS. Carefully aspirate off the PBS.
4. Add 300 µL of Accumax to each well and incubate at 37° for 5 to 10 min, or until cells are completely dislodged from the bottom of the well.

   The use of trypsin is not advised to prepare cells for sorting as it is too harsh and will result in fewer cells surviving the clonal expansion process.
5. Create a suspension by gently pipetting up and down several times.
6. Transfer the cell suspension to a 15-mL conical tube, then centrifuge for 5 min at 100 RCF at room temperature. Aspirate supernatant.
7. Resuspend each sample with 1 ml PBS supplemented with 0.5 µl of the viability dye PI. Take an aliquot and count the cells using a hemocytometer.

   This target concentration should be met if the well surface is near confluency. Although not a requirement, PI will improve efficiency of obtaining single cell clones when sorting.
8. Filter cells through a sterile 35µm cell strainer to collect the uniform suspension in a 5 mL polystyrene round-bottom tube. Place on ice until sort.
9. Using the FACSAria II system (or equivalent), and under sterile conditions, sort single EGFP-positive cells into individual wells of the 96-well plates containing 100 µL of culture media containing DMEM with 50% FBS, 1% Pen/Strep.
10. Place plates into tissue-culture incubator after sorting as soon as possible.

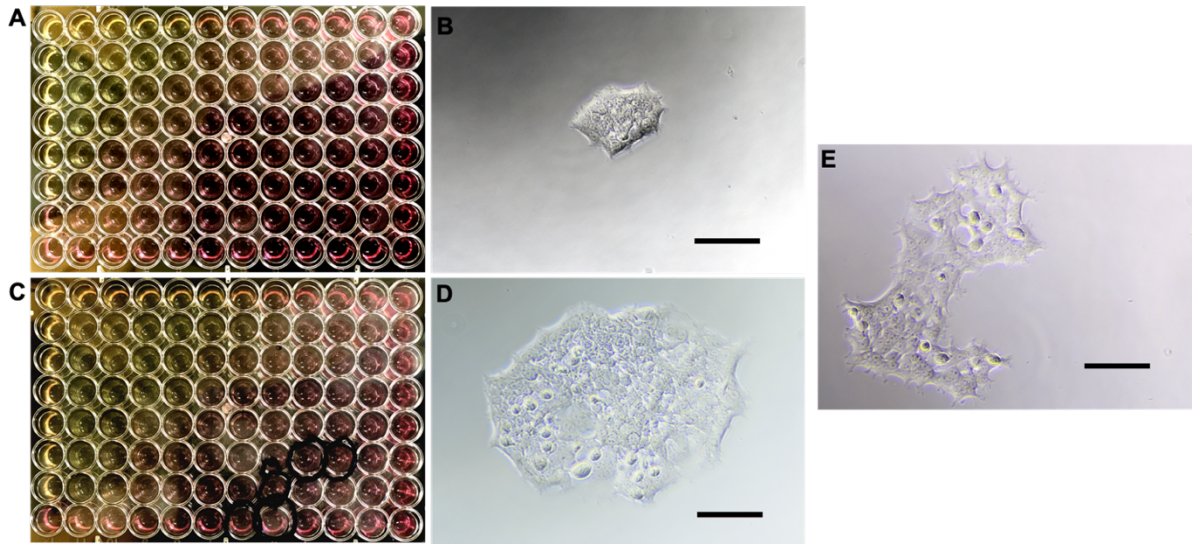

Figure 8. Representative images from dilution plating method (Alternate Protocol 2). (**A**) Image of a 96-well plate taken four days post-dilution plating. The top left wells have a higher concentration of cells as indicated by the low pH level (yellow-colored culture medium). (**B**) 20X magnification image of a well from (**A**) with a single cell-derived colony, taken four days post-dilution plating. (**C**) Image of the 96-well plate from (**A**), taken seven days post-dilution plating. Wells which are harboring single cell-derived clones are marked. (**D**) 20X magnification image of the same colony from (**B**), taken seven days post-dilution plating. (**E**) 20X magnification image of a doublet cell-derived colony, taken seven days post-dilution plating. Scale bars = 166 µm.

2.8 Alternate Protocol 2: Single Cell Isolation of Base Edited Cells using Dilution Plating

Dilution plating offers a cost-effective alternative when researchers do not have access to a sorting machine or when they are working with plasmids which do not possess fluorescent markers. Furthermore, dilution plating might be preferred when working with sensitive cell lines that are less likely to survive the sorting conditions. However, there are two major limitations to dilution plating. Firstly, the chances of obtaining doublet cell-derived colonies increases. We have found, however, that using a 35µm sterile cell strainer prior to dilution plating reduces the chances of obtaining doublet-derived colonies. Secondly, dilution plating does not allow for the discrimination between untransfected and transfected cells, something that is possible when using FACS. Therefore, the rates of successfully obtaining a cell line harboring the SNV of interest are reduced. Therefore, screening more single cell-derived colonies per transfection may be necessary to obtain the cell line of interest. This can be accomplished by using more than two plates per transfection. This protocol uses an adaptation of methods previously described.[68]

2.8.A Protocol Steps

Single Cell Isolation Using Dilution Plating

1. Add 100 µL of pre-warmed DMEM medium (10% FBS, 1% Pen/Strep) to all the wells in a 96-well plate, except to well A1.

    In general, we recommend doing four plates per SNV when the editing efficiency from Protocol 3 is > 20 %. More plates may be needed per SNV when editing efficiency is < 20%.

2. After the cells from Protocol 2 step 6 have been transfected and incubated for three days, pipette off 250 µL of the old medium from the cells and gently rinse cells with 150 µL of PBS. Carefully aspirate off the PBS.
3. Add 50 µL of TrypLE to each well and incubate at 37°C for 5 to 10 min, or until cells are completely dislodged from the bottom of the well.
4. Resuspend the transfected cells with culture medium.
5. Add 200 µL of the cell suspension from step 4 to well A1 through a 35µm sterile cell strainer.

    The use of the strainer is optional, but when used, we have observed a higher frequency of isolating single cell-derived colonies as opposed to doublet cell-derived colonies.

4. Using a single channel pipette, make 1:2 dilutions by transferring 100 µL of the cell suspension from A1 down the first column (B1 to H1) using the same tip. Discard 100 µL of cells from the last well.

    Mix gently before each transfer.

5. Add an additional 100 µL of medium to each well in column 1.
6. Using an 8-channel pipette, make 1:2 dilutions by transferring 100 µL of the cell suspension across each column of the plate starting from column 1 and ending at column 12. Discarding 100 µL of cell suspension from the last column is optional.
7. Place plates into tissue-culture incubator undisturbed after dilutions are made. Observe cells 4-7 days later (Figure 8).

2.9 Protocol 5*:* Clonal Expansion to Generate Isogenic Cell Lines and Genotyping of Clones

Regardless of which method the researcher uses to isolate single cell-derived colonies (FACS - Protocol 4- or dilution plating -Alternate Protocol 2-), colony formation of HEK293T cells should be apparent after 4-7 days and they should be ready to subculture 10-14 days after the isolation protocol. **Figures 8B** and **8D** provide examples of single cell-derived colonies. This protocol describes the process for clonally expanding the isolated cells from Protocol 4 into isogenic cell lines, genotyping the resultant lines, and subculturing the appropriate colonies into larger flasks for storage or downstream experiments. For the SNV of interest, aim to obtain at least three different isogenic cell lines, one harboring a wild-type genotype, one harboring a heterozygous genotype, and one harboring a homozygous genotype. All these cell lines can be used for comparison and use in downstream experiments.

2.9.A Protocol Steps

Subculturing and Genotyping Single Cells

1. Carefully inspect each well of the plates from either Protocol 4 or Alternate Protocol 2 under the microscope and circle the wells that are harboring single cell-derived clones (usually, it takes around 3-4 days to notice distinguishable colony formation, but it can also be longer depending on clonal expansion rate). Be wary of potential doublet cell-derived colonies (see **Figure 8** for examples).

We have found that each 96-well plate will yield around 8 colonies per plate when prepared according to the FACS method, and around 11 single cell-derived colonies per plate when prepared according to the dilution method.

2. Once the colony covers at least 30% of the well's surface area, pipette off 100 µL of the old medium and gently rinse the cells with 50 µL of PBS.

   Since clonal expansion rate may differ from clone to clone, we recommend checking the colonies daily after they have been identified in step 1 of Protocol 5. This step usually takes 1-2 weeks but can vary.

3. Add 30 µL of TrypLE Express. Wait 3-5 minutes, or until cells are completely dislodged from the bottom of the well.
4. While cells are being trypsinized, add 225 µL of pre-warmed (37°C) culture medium (10% FBS, 1% Pen/Strep) to each well of a separate 48-well plate. Allocate two wells for each single cell-derived colony.

   Half of the cells will be clonally expanded while the other half will be used for genotyping.

5. Resuspend the trypsinized cells with 120 µL of pre-warmed DMEM medium (10% FBS, 1% Pen/Strep).
6. Passage two individual 75 µL aliquots of the clonal cell suspension into each well from step 4.
7. Allow the cells to reach 80-90% confluency before proceeding to step 8. This will usually take 3 days, but it is recommended to check the cells under the microscope every day until they reach the appropriate confluency.

8. Two simultaneous steps must be followed. Thus, proper labelling of wells for identification of which clone it corresponds to is crucially important:
    a. Once both wells have reached the appropriate confluency, take one of the wells, and harvest the genomic DNA and genotype, as previously described in Protocols 2 and 3 or Alternate Protocol 1.
    b. For the other well containing the same clone, continue passaging the cells into larger wells or flasks before storage and until the sequencing data from step 8a is obtained. We recommend moving from a 48-well plate (250 µL of DMEM medium with 10% FBS, 1% Pen/Strep), to a 6-well plate (1 mL of DMEM medium with 10% FBS, 1% Pen/Strep), then to a T25 flask (5 mL of DMEM medium with 10% FBS, 1% Pen/Strep), and finally to a T75 flask (10 mL of DMEM medium with 10% FBS, 1% Pen/Strep. Passage the cells when they are at 85-90% confluent, therefore, it is recommended to monitor the cells under the microscope every day. We recommend keeping at least three clones, three cell lines containing wild-type, three clones containing heterozygous, and three clones containing homozygous genotypes (**Figure 7B**).
9. Continue to clonally expand the sequenced-validated cell lines by passaging the cells into larger wells or flasks as indicated in step 8b. We additionally recommend preparing a fourth cell line per each genotype that is a mixture of each of the clones with that specific genotype.
10. Cryopreserve cell lines.

There are many protocols that show researchers how to cryopreserve mammalian cell lines. [69,70] *We also recommend using https://www.abcam.com/protocols/cryopreservation-of-mammalian-cell-lines-video-protocol as a resource.*

2.10 Critical Parameters:

2.10.A gRNA design considerations

Programs including the Benchling wizard (https://benchling.com/pub/liu-base-editor; limited to CBEs), BE-designer (http://www.rgenome.net/be-designer/), DeepHF (http://www.deephf.com/), and BE-Hive (https://www.crisprbehive.design/) can be used for automated gRNA design or scoring.[16,62,71,72] BE-Hive incorporates ABEs and alternate PAMs, making it the most comprehensive software. This machine learning algorithm enables predictions of both editing outcome and efficiency for a given BE:gRNA combination. Other programs are available for designing and scoring gRNAs for traditional Cas9 genome editing, and have been previously reviewed.[73] These may be helpful for predicting potential off-targets or gauging how effectively Cas9 will bind to a given target site.

The human U6 promoter drastically increases expression of the gRNA if guanine is the first nucleotide transcribed. Any decrease in Cas9 binding due to a mismatch at the first position of the protospacer that this may cause is more than compensated for by the increased expression levels. The mouse U6 promoter exhibits high expression with A or G as the first nucleotide, and can be used as an alternative.[72] Additionally, the *S. pyogenes* gRNA backbone contains an early transcription termination signal (UUUU) that may decrease expression levels, and can be mutated

to increase cellular gRNA levels.[74] While we have found that low base editing efficiency is usually caused by other factors, this may be an issue in certain cell types.

2.10.B Bystander Editing

Bystander edits occur when C's or A's beyond the desired target are also present in the editing window and become mutated concurrently with the target base. Bystander editing can be avoided by designing a protospacer that pushes potential bystander bases out of the editing window, as shown with the gRNA example targeting V270 with an adenine in position 9 (**Figure 7A**). Bystander edits may also be mitigated by using deaminase enzymes that are modified to be less processive (such as the YE1, YE2, and YEE CBEs)[20] or sequence-specific (such as the eA3A CBE)[26], or by alternative BE architectures, such as rigid linkers.[28] In some cases, bystander edits are acceptable, such as in gene knockout experiments (**Figure 9C**) or if the bystander creates a silent mutation or intronic mutation (as long as it is not a splice site). When attempting to knock out genes via premature stop codon introduction or splice site disruption, disruptions targeted to the first and last few exons can be inefficient due to translation reinitiating and/or alternative splicing.[75] We suggest base editing splice donor sites in the middle of the gene body to achieve the highest level of protein loss.[76]

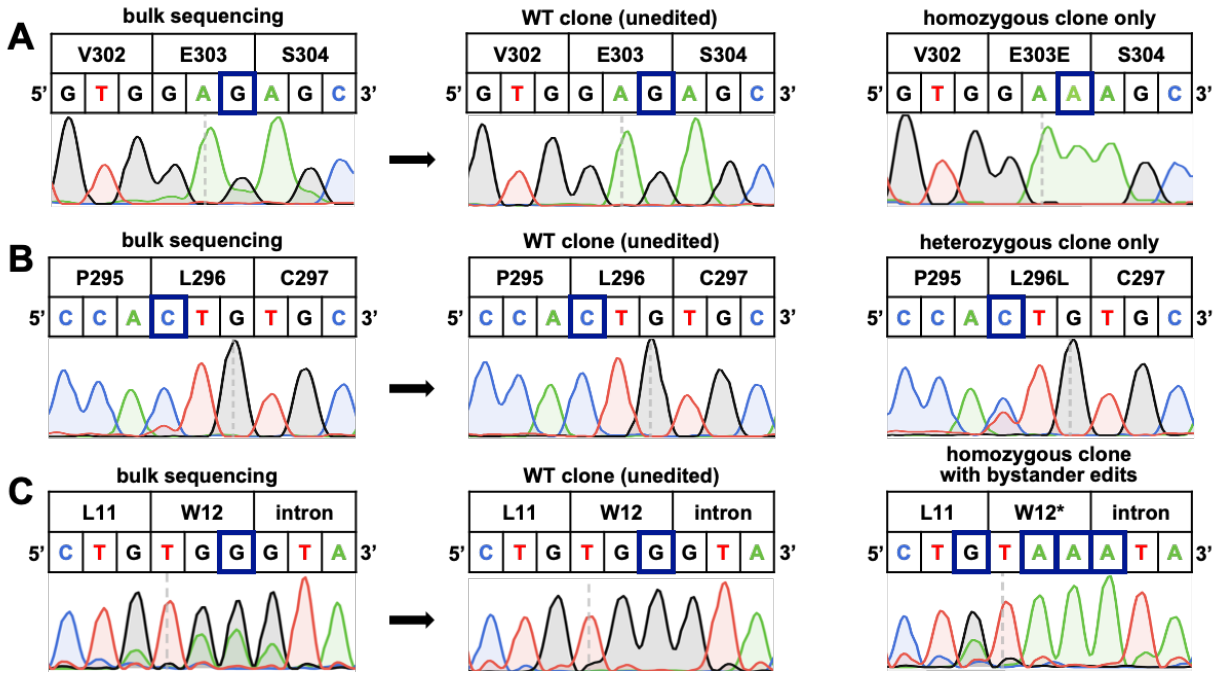

Figure 9. Potential pitfalls in using Base Editing to generate isogenic cell lines. (**A**) Attempted introduction of the MUTYH E303E mutation resulted in only WT and homozygous cell lines, with no heterozygous clones observed. Experiment can be repeated with a less active BE variant, or cells with lower GFP signal isolated. (**B**) Attempted introduction of the MUTYH L296L mutation resulted in only a WT and heterozygous cell lines, with no homozygous clones observed. Check gene and mutation for lethality, and repeat experiment using base editor activity selection scheme. (**C**) Attempted introduction of the MUTYH W12* mutation resulted in bystander edits of non-target bases in the protospacer region. Because a premature stop codon is being introduced, bystander edits are acceptable.

2.10.C Base Editor Selection Considerations

There is a large selection of base editor constructs from which to choose, and the most suitable choice depends on your experimental goal. The two editors that we recommend (BE4max-NG-P2A-EGFP and ABEmax-NG-P2A-EGFP) are good options with which to start, and modifications can be made according to specific experimental requirements. In most cases, nCas9 (D10A) should be utilized, which will direct DNA repair machinery to use the modified base as a template, but dCas9 can be employed if indels must be avoided at all costs (we note that indel formation with CBEs is target site dependent, and ABEs generally do not introduce indels even when using nCas9). If an NG PAM is not available, we suggest using SaCas9-KKH—which recognizes an NNNRRT PAM, but will result in a widened window that can cause bystander

editing[20]—or the SpRY variant [77] instead of SpCas9-NG. A major consideration is the ability of the deaminase enzyme portion of the BE to deaminate other free ssDNA or RNA in the cell, [40–42] causing off-target editing. Engineering efforts have rationally designed deaminase variants that substantially decrease off-target editing but often do so at the expense of sequence specificity, causing decreased on-target activity at certain sites. Specifically, rAPOBEC1-YE1 and the next-generation CBEs listed in **Table 3** are examples of CBEs with greatly reduced off-target DNA editing, and ABEs containing the mutations V106W, F148A, and V82G show almost no off-target RNA editing.[43] If small amounts of off-target editing are acceptable, utilizing the most efficient base editors will aid in isogenic cell line generation. ABE8s significantly increase editing efficiencies but also have a wider editing window, increasing the chance of bystander mutations.

2.10.D Inclusion of Proper Controls

The process of clonal expansion represents an enormous genetic bottleneck and puts the cells under extreme selective pressure. This can result in genetic, epigenetic, and/or phenotypic variation of single-cell-derived lines.[68,78] Additionally, some CBEs can increase the inherent mutation rate twofold during isogenic cell line generation.[29] It is, therefore, of the utmost importance to include proper controls to confirm that observed differences in phenotypes or protein activity are due to the mutation of interest. When generating isogenic lines, transfections should contain a negative control sample that lacks a gRNA or has a nontargeting gRNA sequence that is absent from the human genome (**Table 4**, primer 6). For each set of transfections to generate isogenic cell lines (Protocol 4, single-cell clones should be isolated from this negative control sample alongside the on-target gRNA samples. Using these clones as the "wild-type" cell lines for phenotyping and SNV characterization can control for the process of clonal expansion and potential effects due to DNA damage introduction by the BEs. Additionally, for each genotype of

interest (wild-type, heterozygous, and homozygous for each SNV of interest), it is crucial to generate at least three lines derived from different clones. We also recommend generating an additional "control" line for each genotype that is a mixture of each of the individual clones with that specific genotype. Whole-genome or whole-exome sequencing of the resultant cell lines is highly encouraged to determine whether other genomic modifications might have occurred during the process of clonal expansion.

2.10.E Quantification of Base-Editing Efficiency in Bulk Cells

After base editing has had 3-5 days to proceed, select the appropriate endpoint analysis for your experiment. For bulk Sanger sequencing, EditR can be used to reduce background signal and generate editing percentages based on a *P*-value with a detection limit for base editing of ~7%.[79] However, the reliability of this method is highly dependent on the quality of the Sanger sequencing read. Low-quality Sanger sequencing reads may mislead the researcher into incorrectly concluding that their base-editing efficiency is high enough to proceed to cell line generation (see **Figure 7C** and **7E** for an example). Genomic DNA samples can also be prepared for Illumina NGS to robustly quantitate editing efficiencies. Targeted amplicon sequencing is the most common method with which to do this, and has been previously described.[31,65,66] We recommend the use of the CRISPResso2 open-access software to quantify base-editing efficiencies from fastq files. [67] However, it is important to note that over-amplification during either round of PCR can create PCR bias, which will result in inaccurate quantification; care should be taken to use the fewest possible PCR cycles during gDNA amplification and barcoding.

2.10.F Cell Line Considerations

HEK293T cells are easily transfected and robust enough to tolerate clonal expansion. This provides researchers with a relatively "well-behaved" cell line to optimize experimental conditions and assess the feasibility of generating a cell line with a given SNV before moving forward with other mammalian cell lines. The ultimate cell line to use will be entirely dependent on downstream experimental goals and should be thoroughly investigated before embarking on isogenic cell line generation. Furthermore, immortalized cell lines, such as HEK293T cells, often harbor gene duplications, chromosomal rearrangements, and mutations that allow them to effectively propagate in tissue culture. The exact genomic modifications may even vary from laboratory to laboratory for a given cell line. Additionally, these modifications can differ greatly from the reference genome. As such, it is prudent to first sequence the target locus of interest for any mutations or variation before designing gRNA sequences. Additionally, if the data is available for that specific cell line, check the ploidy at each locus to determine the copy number of the gene of interest.[80]

2.11 Understanding Results

Having the capability to generate both homozygous and heterozygous clones harboring SNVs highlights the importance of using base-editing technologies; generating matched wild-type, heterozygous, and homozygous knock-in clones with traditional genome editing methods is typically quite inefficient or impossible without the use of "blocking mutations" [81] or multiple clonal expansion steps.[82] Using this protocol, we have found that base editors typically introduce their respective SNVs with >10-fold higher efficiency and >100-fold higher precision than traditional genome-editing methods without the use of potentially undesired blocking mutations,

as quantified by NGS, and without the need to physically manipulate cells into phase-enriched populations (physical fractionization) or add exogenous chemicals to block cells into specific phases of the cell cycle (chemical blockade). This is in direct contrast to traditional genome-editing methods, where, typically, accurate SNV introduction occurs at a frequency of <1%. [81,83–86] When using FACS to isolate and clonally expand single cells, we have obtained homozygous edits with a success rate of ~25% and heterozygous edits with a success rate of ~22%. This is measured from an average 8 clones that we typically obtain per 96-well plate. Additionally, in these cases, initial base-editing efficiencies (when measured in bulk, as in Protocol 3) were estimated at ~23%. When using the dilution method, we have obtained homozygous edits with a success rate of ~5%, and heterozygous edits with a success rate of ~14%. This is measured from an average of 11 clones that we typically obtain per 96-well plate. Additionally, in these cases, initial base-editing efficiencies (when measured in bulk, as in Protocol 3) were estimated at ~38%. In some cases, however, it is not possible to generate a complete set of WT, heterozygous and homozygous genotype (**Figure 9**).

2.12 Acknowledgements

This work was supported by the Ruth L. Kirschstein National Research Service Award, US National Institutes of Health (NIH) Grant T32 CA009523. I am also grateful to Jesus Olvera, Cody Fine, and Vu Nguyen of the UCSD Human Embryonic Stem Cell Core Facility for technical assistance of flow cytometry experiments which was made possible in part by the CIRM Major Facilities grant (FA1-00607) to the Sanfond Consortium for Regenerative Medicine. Chapter 2 is reproduced, in full, with permission, from: **Vasquez, C.A.**, Cowan, Q.T., and Komor, A.C (2020). Base Editing in Human Cells to Produce Single Nucleotide Variant Clonal Cell Lines. *Curr.*

Protoc. Mol. Biol. **133***,* e129. The dissertation author was the primary author on all reprinted materials.

Chapter 3

Functional Characterization of *MUTYH* Variants in Live Cells Enabled Through Precision Genome Editing, Chemical Biology, and Biochemical Tools

Introduction

3.1 Introduction to Functional Genomics

Continuing from Chapter 2, I sought to use the protocols written by my colleague, Dr. Cowan, and I, to produce *MUTYH* variant cell lines and functionally characterize them using chemical biology and biochemical biology experimental techniques. This process is becoming increasingly more important because, as discussed in Chapter 2, only a small fraction of identified human genetic variants have been assigned a clinical classification. Therefore, functional characterization of genetic variants has the potential to advance the field of precision medicine by enhancing the efficacy of current therapies and accelerating the development of new approaches to combat genetic diseases. This is particularly crucial when clinical information is limited, such as for the human *MUTYH* gene.

The human MUTYH gene (which encodes for a DNA repair protein) exemplifies the variant interpretation problem. Of the 2,230 total SNVs currently listed in ClinVar, 58.4% (as of early 2024) are listed as a variants of uncertain significance (VUS) or have conflicting reports. However, mutations in this gene are associated with multiple cancer types, the most common of which being MUTYH-associated polyposis (MAP), an increased risk of developing colorectal cancer.[87] Furthermore, while most clinical data suggest that MAP is caused by autosomal recessive (homozygous) point mutations,[88] there are still conflicting reports concerning carrier (heterozygous) genotypes.[89–94] Overall, the clinical classification of VUS in the MUTYH gene using computational methods has been a major challenge, underscoring the importance of evaluating the functional consequences of MUTYH variants using cellular models, particularly in instances of rare variants where clinical information is limited.

MUTYH encodes for the MUTYH protein, an enzyme that contributes to the protection of the genome from damage caused by reactive oxygen species (ROS), which are readily present after a cell undergoes oxidative stress. The most readily oxidized DNA base is guanine (G), which is converted to 8-oxoguanine (8-oxoG) by ROS at an estimated frequency of 2,400 times per cell per day (**Figure 10A**).[95,96] In mammals, the glycosylases OGG1 and MUTYH initiate the base excision repair (BER) process of recognizing, excising, and replacing 8-oxoG lesions (**Figure 10A**).[94,97–102] If 8-oxoG is still base-paired with cytosine (8-oxoG•C), OGG1 will recognize the lesion and excise the 8-oxoG, resulting in an apurinic/apyrimidinic site (abasic site, AP site; [O]) across from C ([O]•C intermediate), which is then further repaired by downstream BER proteins (**Figure 10A**). However, if the 8-oxoG•C is not repaired before DNA replication, 8-oxoG will preferentially base-pair with adenine (A) through Hoogsteen interactions to form an 8-oxoG•A mismatch (Figure 10A).[96,103,104] MUTYH recognizes 8-oxoG•A lesions and excises A to create an abasic site (8-

oxoG•[O] lesion intermediate), which is then further processed by downstream BER proteins into 8-oxoG•C (**Figure 10A**). If, however, MUTYH does not recognize or cannot repair 8-oxoG•A prior to an additional round of DNA replication, thymine (T) will be incorporated opposite the A, resulting in a permanent G•C to T•A mutation.

Notably, the DNA repair process for 8-oxoG (and nearly all DNA lesions) is a tightly coordinated process involving an intertwined network of protein-protein interactions that work together to detect damage and regulate repair before mutations can be generated. [105–111] Several of its interactions with other proteins (including APE1, SIRT6, and Hus1) are via the interdomain connector (IDC, amino acids 309-364), which links the N- and C-terminal MUTYH domains (the catalytic adenine glycosylase and 8-oxoG•A substrate recognition domains, respectively, **Figure 10B-C**). APE1 (apurinic/apyrimidinic endodeoxyribonuclease 1), a multifunctional enzyme that possesses nuclease activity within the BER pathway, is one such interaction partner of MUTYH and performs its enzymatic activity directly after A excision. Studies investigating the physical interaction between the two proteins have suggested the binding site for APE1 resides within the IDC of MUTYH (amino acids 309 to 331).[106,108]

While thousands of *MUTYH* variants have been observed in humans, most are understudied from a mechanistic standpoint. Previous work to functionally characterize *MUTYH* variants in live cells with identical genetic backgrounds has involved the introduction of an exogenous expression construct of *MUTYH* variants (usually using *MUTYH* cDNA) into cells that may (or may not) have the endogenous *MUTYH* gene knocked out.[102,112–121,117,122] Notably, the use of exogenously supplied MUTYH ignores its natural expression levels, which can alter the biology of the protein and how it interacts with its fellow BER proteins. Further, due to multiple transcription initiation sties and alternative splicing, more than nine different isoforms of MUTYH

exist in human cells, which are not faithfully recapitulated with cDNA overexpression experiments.[104] Finally, these types of experiments cannot properly model heterozygous genotypes. Because of the intricacies of the BER pathway, it is of the utmost importance to study MUTYH in its native environment within a living cell and without altering its native expression level or those of its interacting partners, as this can alter the stoichiometry with which they interact. Introducing mutations into the endogenous *MUTYH* gene of living cells would provide the most accurate model to study such variants mechanistically, enabling clinical classification and elucidating disease mechanisms.

We adapted a fluorescent reporter for 8-oxoG•A repair activity for our system and directly measured the MUTYH repair efficiencies of each mutant compared to wild-type in our cell lines. We are able to define thresholds in the signal from this assay using known pathogenic (L111P and W131*) and benign (L296L) variants (according to the guidelines established by the Clinical Genome [ClinGen] Resource),[123–126] which enable the putative classification of a VUS (D271G) as pathogenic. We further found only homozygous mutations to have reduced 8-oxoG•A repair capacity. We expand on this fluorescent reporter to measure 8-oxoG•[O] repair activity in live cells, which reports on the ability of MUTYH to coordinate downstream BER of its substrate, and find the D271G mutant to be defective at coordinating repair of this intermediate. We then used co-immunoprecipitation to find that the D271G mutation disrupts the MUTYH-APE1 interaction. This work described a general strategy for studying genetic variants in DNA repair protein genes using a combination of precision genome editing, chemical biology, and biochemical techniques to provide mechanistic information on how each defective mutant cannot perform its function.

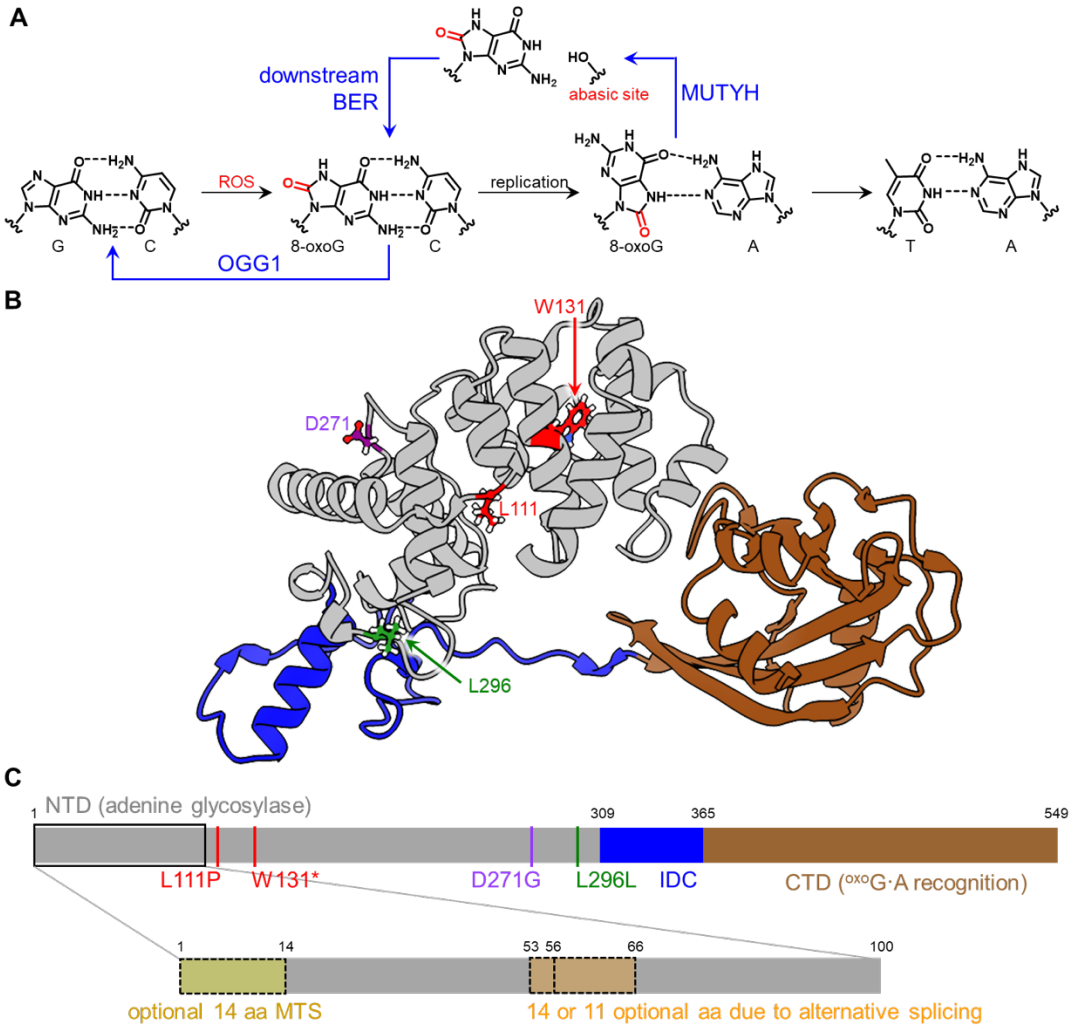

Figure 10. Overview of MUTYH structure and function. (A) Reactive oxygen species (ROS) readily oxidize guanine (G) bases in DNA into 8-oxoguanine (8-oxoG). In mammals, the glycosylases OGG1 and MUTYH initiate the base excision repair (BER) process of recognizing, excising, and replacing 8-oxoG lesions. If 8-oxoG is still base-paired with cytosine (8-oxoG•C), OGG1 will recognize the lesion and initiate BER to convert it back to a canonical G•C base-pair. However, if the 8-oxoG•C is not repaired before DNA replication, 8-oxoG will preferentially base-pair with adenine (A) through Hoogsteen interactions to form an 8-oxoG•A mismatch. MUTYH recognizes 8-oxoG•A lesions and excises A to create an apurinic/apyrimidinic site (abasic site, AP site; [O]) across from 8-oxoG (8-oxoG•[O]), which is then further processed by downstream BER proteins back into 8-oxoG•C. If, however, MUTYH does not recognize or cannot repair 8-oxoG•A prior to an additional round of DNA replication, thymine (T) will be incorporated opposite the A, resulting in a permanent G•C to T•A mutation. (B) Structural model of MUTYH (AlphaFold model AF-Q9UIF7-F1), colored by domain (N-terminal catalytic domain in grey, interdomain connector in blue, and C-terminal substrate recognition domain in brown), and with amino acid residues mutated in this study labelled. Amino acids are colored according to ClinVar classification (red is pathogenic or likely pathogenic, purple is variant of uncertain significance, and green is benign or likely benign). There are a Zn center and FeS cluster missing from the AlphaFold model, and the first 86 and last 51 amino acids are omitted for clarity, as the model confidence of these regions is low. (C) Domain map of isoform 5 of MUTYH is shown, colored according to (B) and with mutations studied in this work indicated. Inset is showing a zoom-in of the first 100 amino acids and the optional regions due to alternative splicing. NTD: N-terminal domain; IDC: Interdomain connector; CTD: C-terminal domain; MTS: mitochondrial targeting signal.

3.2 Results

3.2.1 Engineering Isogenic Cell Lines with MUTYH SNVs

To study *MUTYH* variants in their native context and with the BER pathway intact, we chose to generate isogenic cell lines with endogenously mutated *MUTYH* loci. The generation of heterozygous and homozygous isogenic cell lines requires high genome editing efficiency and precision, and we expected this could be more easily achieved using base editing rather than traditional, double-strand break (DSB)-mediated methods. We initially designed BE:gRNA combinations to install 18 clinically relevant *MUTYH* SNVs, guided primarily by three considerations:

1. We sought to include both pathogenic and benign variants to serve as positive and negative controls and VUS so our studies could provide new clinical insights. Clinical classifications were obtained from the ClinVar database.
2. We focused on mutations that correspond to amino acids in the N-terminal domain of the MUTYH protein (amino acid residues 1-350), as this domain of human MUTYH has been structurally determined.[127] Therefore, we could use structural information to guide experimental interpretations of such variants.
3. Technical considerations were taken into account, such as predicted base editing efficiencies and potential bystander mutations (bystander editing occurs when the deaminase enzyme inadvertently edits additional Cs or As near the target base).

We chose to use HEK293T cells as they possess a wild-type *MUTYH* gene,[80,111] are easy to maintain and transfect, and have been used previously for MUTYH activity studies.[117–119,128,129] We transfected HEK293T cells with plasmids expressing either ancBE4max-NG (hereafter referred to as BE4)[16,19] or ABE7.10max-NG (hereafter referred to as ABE7.10)[31] and a custom-

designed gRNA (**Figure 3B**), allowed three days for editing to occur, lysed the cells, amplified the *MUTYH* loci of interest, and sequenced the resulting amplicons with Sanger sequencing to assess bulk editing efficiencies and precision.

We proceeded with the four *MUTYH* variants listed in **Figure 11**: L296L (c.802C>T), W131* (c.309G>A) , L111P (c.248T>C), and D271G (c.728A>G). Note that we are using the MUTYH amino acid numbering scheme of isoform 5, as this isoform includes all possible amino acids.[104] This numbering system differs from that of isoform 4 (the most abundant nuclear isoform, which is used in ClinVar) by 28 for the majority of the protein (**Figure 10C**). The L296L variant has been detected eight times in patients, with seven likely benign interpretations and one VUS interpretation. The D271G variant has been observed in one patient and is a VUS. This mutation occurs at the interface of the catalytic and IDC domains (**Figures 10B and 11E**). The W131* variant has been detected six times and has four pathogenic and two likely pathogenic classifications. We also chose to include this variant as a possible knock-out control and proceeded with the isogenic cell lines shown despite a bystander mutation that caused a conservative V132L substitution. Finally, the L111P variant has been detected twice with both being classified as likely pathogenic. This mutation occurs in the middle of the catalytic domain. **Figures 10B** and **11E** demonstrate the location of these N-terminal amino acid changes within the MUTYH protein.

We proceeded with these variants as we successfully generated three wild-type (null) genotype clones, three heterozygous genotype clones, and three homozygous genotype clones for each. Null clones were those subjected to the base editing workflow and clonal expansion but resulted in no editing at the *MUTYH* locus, thus serving as wild-type controls (**Figure 11**). Genotyping of the heterozygous clones through NGS revealed three *MUTYH* alleles (chromosome 1 p34.3–p32.1). In our interpretations, we define cells as heterozygous if at least one allele copy

of *MUTYH* was unedited while one allele was edited. Therefore, for the W131* variant, clone 1 possesses two edited alleles and one unedited allele, whereas clones 2 and 3 possess only one edited allele and two unedited alleles. For the D271G heterozygous variant, all three clones contain two edited alleles and one unedited allele. For the L296L heterozygous variant, all three clones contain two unedited alleles and one edited allele. All homozygous clones possess three edited *MUTYH* alleles.

To control for "background" mutations that may have resulted from the clonal expansion processes (in which case, mutations would occur in different locations in different clones), we established three independent cell lines for each genotype. Each cell line also served as biological replicates for all experiments. We additionally analyzed our cell lines for potential off-target base editing, in which case mutations would occur in in the same sites (those with homology to the gRNA spacer sequence) across multiple clones and may convolute interpretation of mechanistic results. To do this, we identified the top two coding and two non-coding predicted off-target sites per gRNA sequenced these loci in each of our 36 isogenic cell lines.[22,130] Notably, we observed no off-target mutations at any of these sites. Overall, these data demonstrate our ability to use base editing to obtain both homozygous and heterozygous isogenic models of *MUTYH* SNVs with no significant off-target mutations.

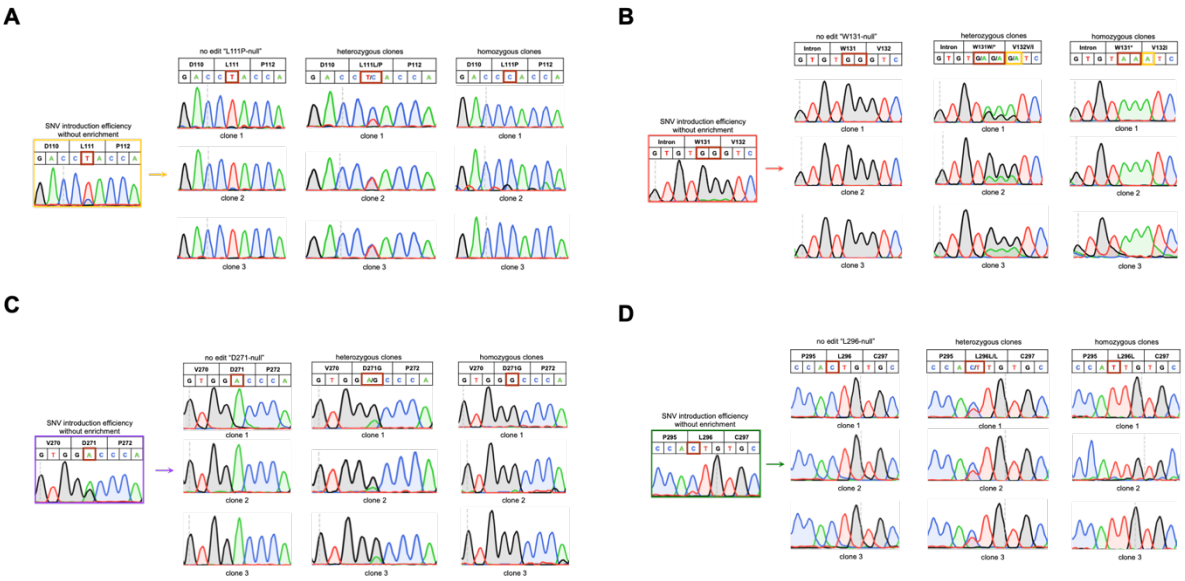

Figure 11. Sanger sequencing of on-target locus for *MUTYH* isogenic cell lines. Isogenic cell lines harboring the **A)** L111P, **B)** W131*, **C)** D271G, and **D)** L296L mutations were generated. The *MUTYH* locus were sequenced with Sanger sequencing. Shown are the Sanger sequencing traces, zoomed in on codons, of bulk cells prior to FACS (left) and all nine isogenic cell lines (right), labeled with their genotype, respectively, for each variant.

3.2.2 Analysis of Protein Expression Levels in Isogenic Cell Lines

We then evaluated the expression levels of the MUTYH protein in the isogenic cell lines by performing western blot analyses. All cell lines except the homozygous W131* mutant showed MUTYH expression levels within error or slightly higher than unedited HEK293T cells (**Figure 12D**). In the homozygous W131* mutant cell lines, we observed an average 24.1 ± 4.3% reduction in MUTYH protein levels. We expected to observe reduced protein levels in the heterozygous cell lines and a complete knock-out of protein levels in the homozygous cell lines due to nonsense-mediated mRNA decay (NMD) from the premature termination codon. Indeed, the ClinVar database proposes that this variant is pathogenic due to NMD of the *MUTYH* transcript,[131] which is based on studies of other frameshifting variants in *MUTYH*.[114,132] However, not only did we detect MUTYH protein in these lines, but the molecular weight of the detected protein is consistent with full-length MUTYH. This suggests that the mechanism of pathogenicity of the W131* *MUTYH* variant may be independent of protein truncation or knock out, although functional investigation in additional cell types would be necessary to confirm this. With these isogenic cell lines in hand, we next sought to evaluate the DNA repair capacity of each MUTYH mutant.

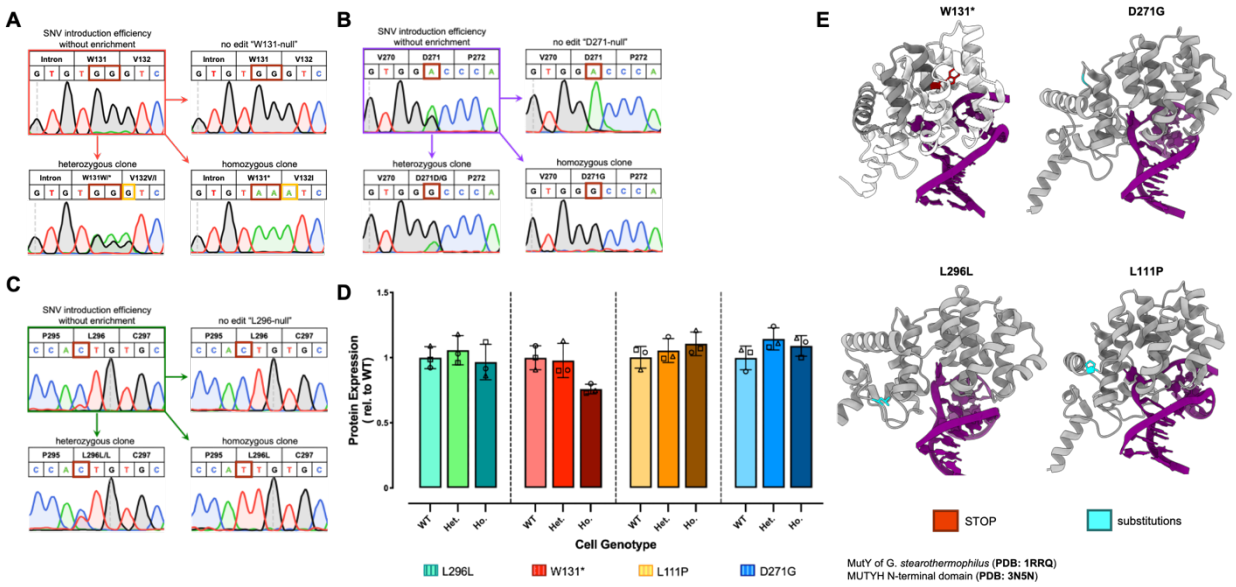

Figure 12. Generation of isogenic cell lines harboring clinically-relevant *MUTYH* single nucleotide variants (SNVs). (A-C). Shown are Sanger sequencing traces of the SNV loci after bulk editing (top left) and after isogenic cell line generation, with one representative isogenic cell line per genotype (null/unedited, heterozygous, and homozygous). (D) Quantification of Western blot data of isogenic cell line lysate for MUTYH. Bars represent the grand average of n=2 technical replicates of n=3 biological replicates (each data point represents the average of two technical replicates for each clone, with circles showing clone 1, triangles showing clone 2, and squares showing clone 3). Error bars represent the standard deviation of the grand averages of the three biological replicates repeated in duplicate experiments. (E) Mutants of interest in this study and their location within the N-terminal domain. The structures were generated by superimposing the crystallized N-terminal domain of MUTYH (PDB: 3N5N) and the MutY bacterial homologue structure (PDB: 1RRQ). The DNA harboring an 8-oxoG•A substrate is shown in purple. The N-terminal domain is shown in grey. SNVs leading to amino acid changes are shown in turquoise and the premature stop codon for W131* is shown in red. For the W131*, amino acids after the STOP codon have been colored in white.

3.2.3 Constructing Fluorescent Reporter Plasmids to Measure Repair of 8-oxoG•A, 8-oxoG•C, and 8-oxoG•[O] Within Live Cells

Next, we sought to develop a method to quantify the enzymatic activities of the MUTYH mutants within their native cellular environments. As mentioned, WT MUTYH recognizes the 8-oxoG•A lesion and excises the A to produce an abasic site (8-oxoG•[O]). It then coordinates with downstream BER proteins (most notably APE1) to convert this lesion into 8-oxoG•C (**Figure 10A**). To quantify 8-oxoG•A repair by MUTYH in our isogenic cell lines, we adapted a previously described reporter system to measure the DNA repair activity of overexpressed MUTYH using

flow cytometry.[117] We generated an mCherry-P2A-EGFP construct in which mCherry and EGFP (enhanced green fluorescent protein) are transcribed on the same mRNA transcript but translated into separate proteins due to ribosomal skipping of the P2A linker during translation. Within the EGFP gene, we incorporated dual Type IIS restriction enzyme (BsaI) sites (which we call a Golden Gate site, or GG site) which allow restriction digestion of the plasmid to produce custom sticky-ends. This construct enabled us to ligate various inserts with compatible sticky ends into the digested backbone and produce an intact mCherry-P2A-EGFP plasmid with custom base-pairs at codon 34. In particular, an 8-oxoG•A mismatch could be incorporated at codon 34 to generate a non-fluorescent EGFP protein with a premature stop codon at position 34 (E34*, **Figure 13A**). Notably, the 8-oxoG lesion would be on the coding strand, with the mispaired A base on the template strand (**Figure 13A**). Therefore, repair of the 8-oxoG•A to 8-oxoG•C by MUTYH and downstream BER proteins (or back to G•C by MUTYH, OGG1, and downstream BER proteins) would restore EGFP fluorescence, with mCherry fluorescence reporting on transfection efficiency (**Figure 12A**). Additionally, the plasmid does not contain a mammalian-compatible origin of replication, so DNA replication does not interfere with the assay.

As positive and negative controls, we cloned mCherry-P2A-EGFP constructs with either Glu34 (wild-type EGFP; pCAV033) or Stop at codon 34 (pCAV034). We transfected either plasmid into wild-type HEK293T cells, waited 24 hours, and analyzed the cells by flow cytometry. As expected, 99.56% of transfected cells (as determined by cells with mCherry fluorescence above background levels) were EGFP+ when using the Glu34 plasmid, and only 0.22% of transfected cells were EGFP+ when using the Stop34 plasmid, suggesting our strategy for using EGFP fluorescence to differentiate C (repaired) versus A (unrepaired) on the template strand was viable.

Finally, we used this same strategy to additionally site-specifically incorporate an 8-oxoG•C DNA lesion (serving as a positive control since C is on the template strand or an 8-oxoG•[O] lesion (where [O] represents an abasic site). Notably, the 8-oxoG•[O] substrate mimics the intermediate that MUTYH passes off to APE1, and wild-type MUTYH has an affinity for this lesion in addition to its canonical 8-oxoG•A substrate.[122] We therefore reasoned that EGFP fluorescence levels in cells transfected with the 8-oxoG•[O] substrate might report on how well MUTYH coordinates downstream repair of 8-oxoG•A after it enzymatically processes the lesion to excise A. Transfection of both lesion-containing DNA constructs into wild-type HEK293T cells followed by flow cytometry analyses showed high levels of EGFP fluorescence; 96.5 ± 2.2% of transfected cells were EGFP+ when using the 8-oxoG•C-containing construct, while 91.8 ± 4.3% of transfected cells were EGFP+ when using the 8-oxoG•[O]-containing construct. Notably, while this assay has been used with 8-oxoG•A and several 8-oxoG analogs,[133] it has never been used to measure repair of the 8-oxoG•[O] intermediate before. Having developed and validated these reporters for MUTYH activity, we next sought to use them to directly evaluate MUTYH-mediated repair in in our isogenic cell lines.

3.2.4 Evaluation of 8-oxoG•A Repair Activities of MUTYH Mutants Within Live Cells using a Fluorescent Reporter

We first transfected the 36 *MUTYH* variant isogenic cell lines with the 8-oxoG•C repair reporter, which we expected should show similarly high levels of EGFP fluorescence across all cell lines. After 24 hours, cells were analyzed by flow cytometry. As expected, all cell lines had similar levels of transfected cells with EGFP fluorescence to the parental, unedited HEK293T cells (96.5 ± 2.2%).

The 36 individual isogenic cell lines were then transfected with the 8-oxoG•A fluorescent reporter. After 24 hours, cells were imaged by fluorescence microscopy and 8-oxoG•A repair was quantified by flow cytometry (**Figure 13B**). The L296L cell lines behaved similarly to the parental, unedited HEK293T cells, as expected given the L296L clinical significance of "benign". Specifically, we observed an 8-oxoG•A repair efficiency (which we define as the percent of transfected, or mCherry+, cells with EGFP fluorescence) of 97.1 ± 1.2% for the WT (null) clones, 95.1 ± 0.12% for the heterozygous clones, and 98.1 ± 0.4% for the homozygous clones (**Figure 13B**). In contrast, we observed a noticeable decrease in 8-oxoG•A repair activity for the homozygous W131* (pathogenic) clones. Specifically, we observed an 8-oxoG•A repair efficiency of 98.7 ± 0.3% for the WT (null) clones, 95.3 ± 1.2% for the heterozygous clones, and 59.2 ± 4.1% for the homozygous clones (**Figure 13B**), representing a 38% decrease compared to the parental, unedited HEK293T cells. This decrease in repair activity is also greater than the decrease in protein expression levels (24.1 ± 4.3%) for these cell lines, suggesting a mechanism of repair deficiency involving more than just decreased protein expression levels for this variant. Interestingly, the heterozygous genotype behaved similarly to the parental, unedited HEK293T cells and the WT (null) line, which supports reports that MUTYH-associated cancers are autosomal recessive. The L111P (pathogenic) clones had a similar phenotype to the W131* clones, with an 8-oxoG•A repair efficiency of 93.3 ± 1.4% for the WT (null) clones, 88.4 ± 4.1% for the heterozygous clones, and 63.4 ± 9.9% for the homozygous clones (**Figure 13B**), representing a 34% decrease compared to the parental, unedited HEK293T cells. Again, the WT (null) and heterozygous clones behaved similarly to each other. Finally, when measuring 8-oxoG•A repair for the D271G (VUS) clones, we observed similar 8-oxoG•A repair efficiencies to both pathogenic variants. Specifically, we observed 8-oxoG•A repair efficiencies of 95.9 ± 0.6% for the WT (null) clones, 94.2 ± 3.8% for

the heterozygous clones, and 61.1 ± 8.9% for the homozygous clones (**Figure 13B**), representing a 36% decrease compared to the parental, unedited HEK293T cells. Once again, the heterozygous clones had a similar phenotype to their WT (null) counterparts. In all cases where we observed large reductions in 8-oxoG•A repair efficiencies, as defined as the percent of transfected cells with EGFP fluorescence. These data demonstrate the utility of this fluorescent reporter strategy for quantifying the repair capacity of MUTYH variants in live cells. Furthermore, while the sample size is small, we observed reduced MUTYH repair capacity to correlate with clinical pathogenicity.

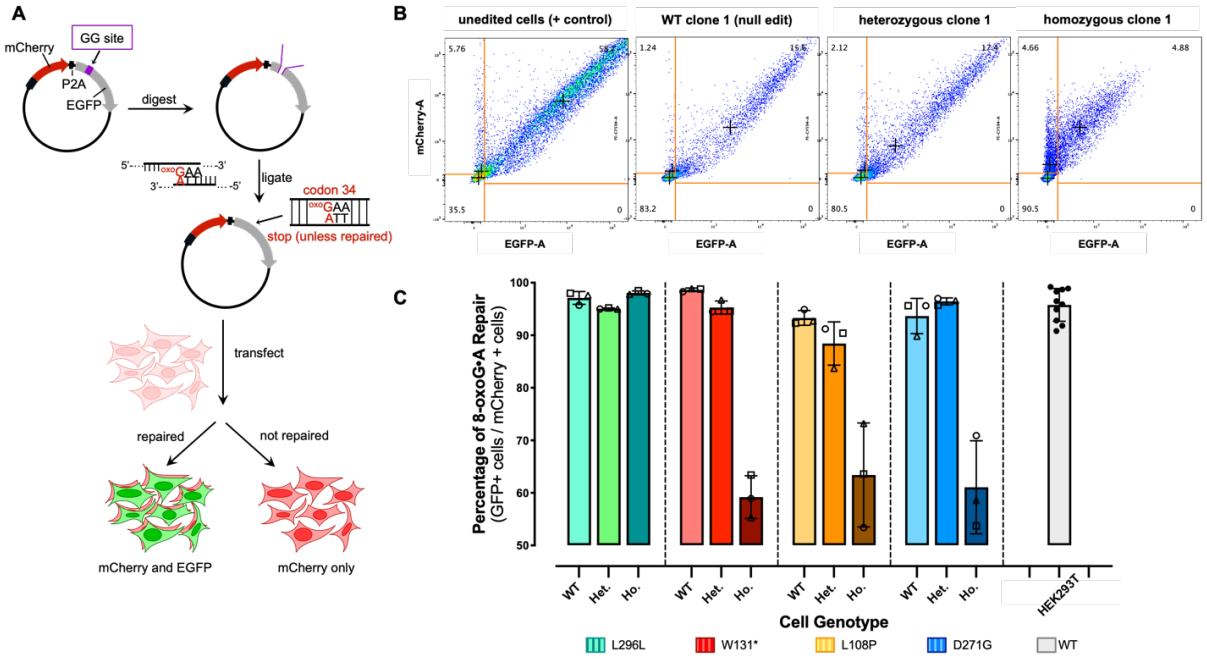

Figure 13. Repair of 8-oxoG•A using a MUTYH lesion-specific plasmid reporter. (A) Schematic diagram of generation and use of the 8-oxoG•A-containing plasmid. The plasmid-based fluorescent reporter contains both a red fluorescent protein, mCherry, and an inactive enhanced green fluorescent protein (EGFP). EGFP is inactive due to a sequence of DNA that frameshifts the EGFP and harbors an incorporated dual Type IIS restriction enzyme (BsaI) recognition site (referred to as a Golden Gate site, or GG site), which allows restriction digestion to produce custom sticky-ends. This construct enables various inserts with compatible sticky ends to be ligated into the digested plasmid backbone and produce an intact mCherry-P2A-EGFP plasmid with custom base-pairs at codon 34 of EGFP. To evaluate MUTYH DNA repair activity, an 8-oxoG•A lesion is incorporated into codon 34. Once transfected into living cells harboring either WT or mutated *MUTYH*, both genes are transcribed in a single mRNA transcript, but are translated into unique and separate fluorescent proteins via the self-cleaving P2A linker. mCherry fluorescence acts as a transfection marker, and EGFP fluorescence occurs only there is repair of 8-oxoG•A to 8-oxoG•C or G•C by MUTYH and downstream base excision repair proteins. (B) Representative flow cytometry plots and gating schemes to quantify 8-oxoG•A repair in live cells using matching L111P MUTYH mutant cell lines as an example. The plots show compensated red fluorescence intensity (*y*-axis) versus compensated EGFP fluorescence intensity (*x*-axis) for four representative samples, left to right: unedited HEK293T cells, HEK293T cells that were transfected with the L111P gRNA, but produced no editing at the target site (null clone), heterozygous L111P MUTYH clone 1, and homozygous L111P MUTYH clone 1. 8-oxoG•A repair is quantified by calculating the percent of EGFP+ cells divided by the transfected, or mCherry+, cells. Scatter gates were applied to remove nonviable cells and doublets. Quadrant boundaries for analysis were set by using unedited HEK293T cells that were transfected with mCherry only or EGFP only plasmids. The numbers in each quadrant represents the percentage of cells within that population. "+"'s in the quadrants indicate the median EGFP fluorescence intensity of EGFP-positive cells. (C) Percentage of 8-oxoG•A repair in living cells harboring various MUTYH mutants. Values calculated as described in (B). Bars represent the average of n=3 biological replicates (circles show clone 1, triangles show clone 2, and squares show clone 3). Error bars represent the standard deviation of the three biological replicates.

3.2.5 Reductions in 8-oxoG•[O] Repair Efficiency for the W131* and D271G MUTYH Mutants Suggest Defective Interactions with Downstream BER Proteins

To further characterize the DNA repair capacities of our *MUTYH* variants, we next transfected all 36 isogenic lines with the 8-oxoG•[O] fluorescent reporter (**Figure 14A**). Again, after 24 hours, cells were imaged by fluorescence microscopy, and 8-oxoG•[O] repair was quantified by flow cytometry. MUTYH is a monofunctional DNA glycosylase and thus only catalyzes the excision of the adenine opposite the damaged 8-oxoG lesion to produce the 8-oxoG•[O] intermediate. APE1 is then required to cleave the DNA backbone prior to gap filling. We reasoned that any pathogenic MUTYH mutants that are deficient in adenine excision but still able to interact with APE1 for downstream processing would facilitate repair of the 8-oxoG•[O] lesion at levels similar to that of WT. However, mutants that are defective at interacting with APE1 (or any other MUTYH binding partners) would have reduced abilities to repair the 8-oxoG•[O] lesion. The L296L cell lines again behaved similarly to the parental, unedited HEK293T cells. Specifically, we observed 8-oxoG•[O] repair efficiencies (which we again define as the percent of transfected, or mCherry+, cells with EGFP fluorescence) of 98.5 ± 0.2% for the WT (null) clones, 98.1 ± 1.3% for the heterozygous clones, and 93.3 ± 0.5% for the homozygous clones (**Figure 14B**) In the W131* cell lines, we observed 8-oxoG•[O] repair efficiencies of 92.9 ± 0.3% for the WT (null) clones, 94.8 ± 1.9% for the heterozygous clones, and 72.8 ± 8.3% for the homozygous clones (**Figure 14B**) indicating a deficiency in coordination of downstream BER for this variant. The L111P mutation within *MUTYH* is near the enzyme's active site and is predicted by *in silico* methods to be defective in adenine excision.[134] Correspondingly, for the L111P cell lines we

observed 8-oxoG•[O] repair efficiencies of 93.0 ± 0.7% for the WT (null) clones, 90.2 ± 4.3% for the heterozygous clones, and 87.9 ± 6.9% for the homozygous clones (**Figure 14B**). This indicates that while this variant is overall deficient in 8-oxoG•A repair, the deficiency is likely at the adenine excision step, as it is able to coordinate repair of the 8-oxoG•[O] intermediate. Finally, for the D271G cell lines, we observed 8-oxoG•[O] repair efficiencies of 93.3 ± 1.0% for the WT (null) clones, 97.0 ± 1.8% for the heterozygous clones, and 78.6 ± 2.3% for the homozygous clones (**Figure 14B**), indicating a deficiency in coordinating 8-oxoG•[O] repair by this variant. This suggested to us that the mechanism of 8-oxoG•A repair deficiency of the D271G mutant involves its inability to interact with downstream BER proteins, while the L111P mutant is proficient in coordinating downstream repair of the 8-oxoG•[O] intermediate. Notably, these data are consistent with the respective locations of each mutation (D271G near the IDC, and L111P in the catalytic domain, **Figure 10A**). Furthermore, these data demonstrate that this assay is able to report on the capacity of MUTYH to coordinate downstream repair of its substrate following adenine excision.

3.2.6 Defective MUTYH-APE1 Interactions Identified by Co-Immunoprecipitation

The W131* and D271G MUTYH mutants were defective at repair of both 8-oxoG•A and 8-oxoG•[O] lesions, which we hypothesized to be due to deficiencies in interacting with APE1. We therefore sought to probe the interaction of MUTYH with APE1 in our cell lines. We examined whether each of our four mutations in MUTYH impacted the corresponding protein's ability to interact with APE1 by co-immunoprecipitation (co-IP) experiments. Since co-IP requires robust and durable protein association, we temporarily induced elevated (but still physiologically relevant) expression levels of MUTYH by incubating cells (wild-type and homozygous mutant

lines) with hydrogen peroxide for an hour to induce oxidative stress.[135] We then performed a nuclear extraction, immunoprecipitated MUTYH from the resulting nuclear lysate, and blotted for APE1. The co-IP experiments were consistent with our data from the 8-oxoG•[O] repair assay (**Figure 14C**). Specifically, we detected high levels of APE1 in the immunoprecipitants from all wild-type, L296L, and L111P homozygous cell lines (mutants that displayed 8-oxoG•[O] repair activities similar to wild-type). However, this interaction was greatly reduced in the immunoprecipitants from the W131* and D271G homozygous cell lines (mutants that displayed greatly reduced 8-oxoG•[O] repair activities compared to wild-type). Quantification of APE1 signal relative to that of MUTYH, normalized to unedited clones, is shown in **Figure 14C** and supports these observations. We observed APE1 levels similar to wild-type in the L296L (0.80 ± 0.13 relative to wild-type levels; not significantly different, $p = 0.180$) and L111P (0.82 ± 0.19 relative to wild-type levels; not significantly different, $p = 0.085$) homozygous lines, but this was greatly reduced in the W131* (0.21 ± 0.10 relative to wild-type levels, $p = 0.001$) and D271G (0.29 ± 0.07 relative to wild-type levels, $p = 0.010$) homozygous lines. These data support the notion that the 8-oxoG•[O] repair assay reflects upon MUTYH's ability to coordinate downstream BER of this intermediate. This also demonstrates that the mechanism of pathogenicity of the D271G mutant involves its failure to interact with APE1.

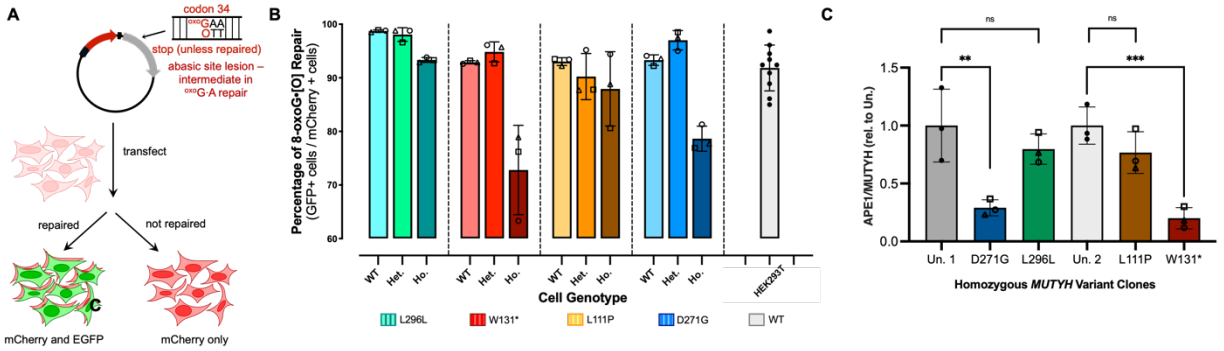

Figure 14. Defective MUTYH-APE1 Interactions Identified by 8-oxoG•[O] repair and co-Immunoprecipitation. (A) Schematic diagram of the fluorescent reporter for 8-oxoG•[O] repair. The backbone of the reporter is the same as that in **Figure 13A**. However, an 8-oxoG•[O] lesion is incorporated at codon 34 instead of 8-oxoG•A. (B) Percentage of 8-oxoG•[O] repair in living cells harboring various MUTYH mutants. Values calculated as described in **Figure 13B**. Bars represent the average of n=3 biological replicates (circles show clone 1, triangles show clone 2, and squares show clone 3). Error bars represent the standard deviation of the three biological replicates. (C) Plotted are the ratios of the relative amount of APE1 protein to total MUTYH protein in MUTYH immunoprecipitated samples, normalized to that of untreated HEK293T cells (based on the co-immunoprecipitation experiments). Plotted for each experimental condition are n=3 biological replicates containing homozygous genotypes (circles show clone 1, triangles show clone 2, and squares show clone 3). The D271G and L296L homozygous clones were normalized to the average of untreated HEK293T cells from the same blot (Un.1) and L111P and W131* homozygous clones were normalized to the average of untreated HEK293T cells on the same blot (Un.2). Error bars represent the standard deviation of the three biological replicates. Data were analyzed with unpaired, one-tailed, parametric t-tests, and p values are marked as follows: ns=$p \geq 0.05$ not significant, and *$p \leq 0.05$, **$p \leq 0.01$, and ***$p \leq 0.001$ are significant.

3.3 Discussion

Overall, we leveraged base editing to generate cellular models of clinically-relevant *MUTYH* variants, allowing for the mechanistic study of the corresponding mutant proteins in their native cellular environment. Notably, studying these mutants by introducing their corresponding mutations into the endogenous genomic locus does not impact their native expression levels, resulting in conditions that more closely resemble their natural environment. To evaluate MUTYH-mediated repair of 8-oxoG•A in these cell lines, we adapted a fluorescence-based adenine glycosylase assay, in which excision of the mispaired adenine opposite the 8-oxoG followed by the installation of C restores EGFP expression. This fluorescent reporter allowed us to directly measure the MUTYH repair efficiencies of each mutant compared to wild-type in our cell lines. Furthermore, we modified this assay to also introduce an 8-oxoG•[O] (abasic site-containing) substrate at the lesion site, enabling us to observe MUTYH's ability to coordinate downstream repair of this intermediate with other BER factors. Together, these assays allowed us to assess overall 8-oxoG•A repair capacity (which correlated with clinical pathogenicity) as well as provided us with mechanistic information regarding each mutant's repair deficiency. The results from the 8-oxoG•[O] repair assay were then complemented with co-IP studies of MUTYH and APE1. Notably, we found that only homozygous (and not heterozygous) cell lines demonstrated reduced DNA repair capacity, which supports reports that MAP is autosomal recessive.

We used the L296L and W131* mutants as positive and negative controls, respectively. The L296L mutant, which is clinically classified as likely benign, is caused by a silent C>T mutation, while W131*, which is clinically classified as likely pathogenic, is due to a G>A nonsense mutation. Correspondingly, we found that all L296L cell lines (heterozygous and homozygous) behaved as wild-type across all assays, while the homozygous W131* lines were

defective at both 8-oxoG•A and 8-oxoG•[O] repair. While we expected this phenotype to be due to protein knock-out, we observed only a partial (average 24 ± 4.3%) reduction in MUTYH expression levels in these cell lines. The MUTYH protein expressed in these cell lines appeared to be full-length as well. While additional mechanistic investigation of this phenomenon was beyond the scope of this work, this observation suggests a pathogenic mechanism for this variant that may be distinct from NMD of the *MUTYH* mRNA.

We additionally studied two missense variants, L111P (clinically classified as likely pathogenic) and D271G (clinically classified as VUS). The homozygous cell lines of both mutants had MUTYH protein expression levels within error of wild-type controls, but were found to be defective at 8-oxoG•A and repair. Interestingly, the L111P homozygous cell lines were proficient at 8-oxoG•[O] repair, suggesting that this mutant is proficient at binding to this substrate and interacting with downstream BER proteins. Furthermore, co-IP experiments on the L111P homozygous cell lines showed the MUTYH-APE1 interaction to be intact and comparable to that in wild-type and L296L cell lines. Taken together, these data suggest that the L111P mutant is likely pathogenic due to defective adenine excision of the 8-oxoG•A substrate.

In contrast, the D271G mutant was defective at both 8-oxoG•A and 8-oxoG•[O] repair, suggesting a deficiency of this variant to interact properly with other BER proteins necessary to complete the repair back to 8-oxoG•C. Consistent with these results, co-IP experiments on the D271G homozygous cell lines confirmed that this mutant no longer interacts with APE1. Studies investigating the physical interaction between the two proteins have suggested the binding site for APE1 on MUTYH to be residues 293-318.[106,108] While the D271G mutation is outside this region, the highly unconservative nature of this mutation may be responsible for impacting this interaction. Overall, these data suggest a faulty “baton handoff” (miscoordination) between the D271G

MUTYH mutant and WT APE1 necessary to process 8-oxoG•A repair within BER. Since the catalytic activity of the D271G mutation may also be impaired, future studies could consider studying this variant *in vitro*, which is beyond the scope of this work.

Currently, over 1,000 MAP-associated missense variants are VUS, which will surely rise with the increasing use of sequencing technologies. This limited understanding of *MUTYH* variant dysfunction can be combated by generating relevant human-derived cell models and complementary assays for elucidating their pathogenicity. We developed here a framework for engineering *MUTYH* variant cell lines using base editing, which, prior to this study, had not been reported, and assessed their ability to repair DNA damage in live cells. Additionally, we modified a previously developed fluorescent reporter for 8-oxoG•A repair activity to also report on repair of the MUTYH intermediate 8-oxoG•[O]. Importantly, we characterized the DNA repair capacity of the D271G mutant, which is currently classified as a VUS and found that this mutation disrupts crucial protein-protein interactions with APE1. These findings underscore the importance of studying potentially pathogenic variants in relevant cell lines in which the full BER pathway is intact. Future applications of this genome editing framework include the clinical and mechanistic characterization of *MUTYH* variants in high-throughput.

3.4 Methods

3.4.1 Molecular cloning

All primers in this study were ordered through Integrated DNA technologies (IDT). All PCR reactions were performed with Phusion DNA Green High-Fidelity Polymerase (ThermoFisher Scientific, #F534L) or Phusion U (Thermo Fisher, #F556L) where appropriate. gRNA plasmids were cloned by site directed mutagenesis using a 5' tail in the forward primer to replace the 20nt spacer region (Protocol 1[32]; *S. pyogenes* Cas9 gRNA vector Addgene plasmid

#47511). Primer sequences can be found in **Table 5** and **6**. Base editor plasmids were generated by taking either AncBE4max-P2A-GFP (Addgene plasmid #112100) or ABE7.10max-P2A-GFP (Addgene plasmid #112101) and replacing the C-terminal 300 amino acids of Cas9n with that from SpCas9-NG (Addgene plasmid #117919) using USER (Uracil-Specific Excision Reagent) cloning[136] to produce CBE and ABE variants that recognize a relaxed PAM of NG.

The intact mCherry-P2A-EGFP reporter plasmids were generated with USER cloning following New England Biolabs (NEB) protocols, by replacing the ABE gene in the ABE-P2A-EGFP plasmid (Addgene plasmid #112101) with the mCherry gene from the pBAD-mCherry plasmid (Addgene plasmid #54630). The sequence of the mCherry-P2A-EGFP open reading frame used in this work can be found in the Supplementary Sequences section. All variations (i. e., point mutations) on this plasmid were cloned using site-directed mutagenesis.[137] The pCAV035 parental DNA Repair reporter plasmid (Addgene plasmid #219807) with a custom Golden Gate site was generated via USER cloning from the pCAV033 plasmid (mCherry_P2A_EGFP construct) to insert a unique DNA sequence containing two BsaI recognition sites at codon 34 of EGFP. To generate reporter constructs with custom base-pairs at codon 34, pCAV035 was digested with BsaI-HFv2 (NEB, #R3733S) according to the manufacturer's protocol and purified using a gel extraction kit (Qiagen, #28704). Next, pre-annealed oligos containing synthetic DNA lesions (sequences of which can be found in **Table 5**) and complementary sticky ends matching the digested overhangs were annealed into the digested site using T4 DNA ligase (NEB, #M0202) overnight at 12 °C. The ligation product was then purified using columns from the QIAquick PCR purification kit (Qiagen, #28106). The abasic site-containing oligo (for the for the 8-oxoG•[O] reporter) was produced by digestion of the oligo listed in **Table 5** with Uracil-DNA Glycosylase (UDG; NEB # M0280S) according to the manufacturer's protocol prior to ligation into the digested

pCAV035 plasmid. All plasmids used in our transfections were endotoxin-free and prepared with the ZymoPURE II Midiprep Kit (Zymo Research, #D4201). Sequences of the pCAV033 and pCAV035 constructs are available. The 8-oxoG-containing oligos was both obtained and quality controlled, using capillary electrophoresis and electrospray ionization mass spectrometry, from IDT.

3.4.2 Cell culture and transfections

HEK293T cells (ATCC CRL-3216) were cultured in high glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with GlutaMAX (ThermoFisher Scientific, #10566-016) and 10 % (v/v) fetal bovine serum (ThermoFisher Scientific, #10437-028) at 37 °C with 5 % CO2. Cells were passaged every 3 days using TrypLE (ThermoFisher Scientific, #12605028).

HEK293T cells were seeded at 50,000 cells/well in 250 µL media in a 48-well plate and transfected after 16 hours, when they were at ~70% confluency.[32] Mixtures of plasmids encoding gRNA and appropriate BE were created in a total volume of 12.5 µL with Opti-MEM reduced-serum medium (ThermoFisher Scientific, #31985-070) using 200 ng gRNA plasmid and 800 ng of CBE or ABE plasmid. These DNA mixtures were then combined with a mixture comprised of 1.5 µL Lipofectamine 2000 regent (ThermoFisher Scientific, #11668030) and 11 µL of Opti-MEM, incubated at room temperature for 15 minutes, and added to the plated cells. To measure bulk base editing efficiencies, cells were incubated for 72 hours, washed with 150 µL phosphate buffered saline (PBS, Gibco #10010-023), and genomic DNA (gDNA) was extracted by adding 100 µL of freshly prepared buffer containing 10 mM Tris-HCl (pH 7.0), 0.05% SDS, and 25 µg/ml of Proteinase K (NEB, #P8107S,). Digestion was done at 37°C for 1 hour, followed by an 80°C heat treatment for 30 minutes. The gDNA was then used as a template for Sanger or next generation sequencing

3.4.3 Sanger sequencing of the *MUTYH* locus.

To obtain Sanger sequencing of each *MUTYH* variant, a 50-µl PCR reaction was run, using the primers specified in **Table 6**, to amplify the genomic loci of interest using the Phusion high-fidelity DNA polymerase protocol (NEB, #M0530) along with the thermocycler parameters previously described (Protocol 3).[32] Once the presence of the predicted PCR product had been confirmed via gel electrophoresis, the samples were sent to the Genewiz (from Azenta Life Sciences) using their unpurified PCR-product Sanger sequencing service. In instances in which Sanger sequencing results from the unpurified PCR-product service failed, the purified PCR-product service was used. For this service, PCR products were purified using the QIAquick PCR purification kit (Qiagen, #28106), and then the appropriate amount of PCR product (according to the Genewiz sample submission guidelines) was pre-mixed with 5 µL of either the forward or reverse primer (5 µM).

3.4.4 Fluorescence activated cell sorting (FACS)

For the generation of isogenic cell lines, transfections were repeated as described above. After 72 hours, cells were washed with 150 µL of PBS and detached with 30 µL of Accumax (Innovative Cell Technologies, #AM-105) at room temperature. After 1 minute, cells were resuspended with 170 µL of cold PBS, passed through a FACS tube containing a 35 µm cell strainer (Falcon, #352235), and kept on ice. EGFP-positive cells were sorted into either individual wells of a 96-well plate containing 100 µl DMEM with 50% FBS and 1% Pen/Strep or sorted in bulk (at least 5,000 cells) for dilution plating as previously described.[32] FACS was performed on a FACSAria II. Cells were expanded for 1-2 weeks and then genotyped with Sanger sequencing as described above or next-generation sequencing as described below.

3.4.5 Next-generation sequencing (NGS) of isogenic cell lines

After isogenic cell line generation, on-target and potential off-target genome editing loci of isogenic cell lines were sequenced via targeted amplicon NGS.[32,138] Briefly, gDNA was extracted from ~70,000 cells after washing the cells with 150 µL of PBS and lysing in 100 µL digestion buffer as described above in "*Transfections for base editing experiments"*. Genomic loci of interest were amplified via two rounds of PCR. Round 1 PCRs were completed using a 25 µL PCR reaction with Phusion polymerase (ThermoFisher Scientific, #F534L) comprised of 1 µL of genomic DNA, GC buffer, 3% DMSO, and 25% of the recommended primer amount (to reduce the amount of primer dimers; NGS primers are listed in **Table 6**; these locus-specific primers were designed to contain an adapter sequence, allowing for sample barcoding with a second round of PCR) and amplified for 24-27 cycles (minimal amount to avoid PCR bias) using an annealing temperature of 62 °C and an extension time of 25 seconds. After confirmation of the accurately-sized product on a 2% agarose gel, round 2 PCR was performed to barcode samples in a 10 µL total reaction volume, comprised of 0.10 µL Phusion polymerase, 5.95 µL water, 2 µL 5X HF buffer, 0.2. µL dNTPs, 1.25 µL primer (0.0625 µM), and 0.5 µL round one PCR product, and amplified for 8-12 cycles using an annealing temperature of 65 °C and an extension time of 25 seconds. Second round PCR products were pooled together based on amplicon size and purified from a 2% agarose gel using a gel extraction kit (Qiagen, #28704). The resulting purified libraries were then quantified with the Qubit dsDNA high sensitivity kit (ThermoFisher Scientific, #Q32854). Samples were then diluted to 1.4 pM following Illumina's sample preparation guidelines. The final library was mixed with 1.4 pM PhiX sequencing control (10% v/v) and then sequenced on an Illumina MiniSeq via paired end sequencing (2x151 paired end reads).

3.4.6 Transfections for flow cytometry

HEK293T cells were seeded at 100,000 cells/well in 250 µL media in a 48-well plate and transfected 16 hours after plating, when they were at ~70% confluency. 500 ng of mCherry-P2A-EGFP reporter plasmid (intact plasmids or with custom inserts) was diluted to a total volume of 12.5 µL with Opti-MEM reduced-serum medium. The DNA mixture was then combined with a solution of 1.5 µL of Lipofectamine 2000 regent in 11 µL of Opti-MEM and added to the plated cells, Cells were then incubated for 24 hours before harvesting for flow cytometry.

3.4.7 Flow cytometry analysis of DNA repair with EGFP reporter vectors

For all DNA reporter fluorescence measurements, the medium was removed from each well, and each well was washed with 150 µL of PBS. To detach cells, 30 µL of Accumax was added to each well. Cells were counted and diluted to a concentration of 1×10^6 cells/mL in PBS, then strained into 5 mL tubes through a cell strainer cap (Corning, #352235) and kept on ice. Flow cytometry data was collected using a Bio-Rad S3e cell sorter equipped with 488 nm, 561 nm and 640 nm lasers, and analyzed using FlowJo v10.8.1 Software (BD Life Sciences). Scatter gates were applied to remove nonviable cells and doublets. For reporter experiments, gates were applied based on cells transfected with mCherry only or EGFP only plasmids. mCherry fluorescence was detected using FL3 (602–627 nm) and a PMT voltage of 360. EGFP fluorescence was detected using FL1 (510–540 nm). ~10,000 cells (after scatter gating) were collected for each sample.

3.4.8 Preparation of cell extracts and Western blotting

Roughly 20 x 10^6 cells were harvested from a 150 mm dish by first removing the medium and washing with 2 mL of cold PBS. To detach cells, 1 mL of RIPA buffer (1% Triton X-100, 0.5% DOC, 0.1% SDS, and 50 mM of Tris, pH 7.4) plus Halt Protease and Phosphatase Inhibitor Cocktail (ThermoFisher Scientific, #78440) was used. The resulting suspension was then transferred into a pre-cooled microcentrifuge tube. The cells were then maintained at constant agitation for 30 min at 4°C. The cell suspension was then centrifuged for 10 minutes at 16,000 r.c.f at 4°C. After centrifugation, the tubes were placed on ice and the supernatant was aspirated and placed in a fresh tube kept on ice. Protein concentrations were determined using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific, #23225). Equal amounts of total protein from clarified cell lysate solutions (40 µg of total protein) were mixed with NuPAGE LDS Sample Buffer (Invitrogen #NP0007) plus 10 mM dithiothreitol (DTT, ThermoFisher Scientific, #R0861) to a total volume of 80 µL and heated for 10 minutes at 95 °C, then 40 µL of each sample was loaded and electrophoresed on a 7.5% Criterion™ TGX™ Precast Midi Protein Gel (Bio-Rad, #5671024). The protein was then transferred to a 0.45-µm PVDF transfer membrane (ThermoFisher Scientific, #88585) using the Mini Trans-Blot® Cell (Bio-Rad, # 1703930; 50 V, 60 minutes). The membrane was then incubated with Revert Total Protein Stain (LI-COR #926-11010) and washed according to the manufacturer's instructions. The membrane was then blocked with 5 mL of 5% milk in tris-buffered saline with 0.1% Tween 20 (TBST; 20 mM Tris, 150 mM NaCl, and 0.1% Tween 20 from ThermoFisher Scientific, #85114) for 1 hour and then incubated overnight at 4 °C with MUTYH (C-6) primary antibody (Santa Cruz Biotechnology #sc-374571) diluted 1:500. The membrane was washed three times with 5 mL TBST for each wash, then incubated with HRP-anti-mouse IgG (Cell Signaling #7076) diluted 1:2,000 in 10 mL TBST for 1 hour at room temperature.

The membrane was then washed again three times with 5 mL TBST, then soaked in 200 µL chemiluminescent substrate (Bio-Rad #1705061) for 5 minutes and imaged with a Syngene G:Box Chemi XX6 imager.

3.4.9 Co-immunoprecipitation experiments

The homozygous cell lines and wild-type HEK293T cells were seeded into six-well plates at a density of 5×10^5 per well in 2 mL of media and incubated overnight. Prior to harvesting, the cells were treated with 0.5 mM hydrogen peroxide (H_2O_2) for 1 hour. The cells were then harvested by first removing the medium and washing with 2 mL of cold PBS. To detach cells, 200 µL of TrypLE (ThermoFisher Scientific, #12605028) was used. The cells were then transferred to a pre-cooled microcentrifuge tube and then centrifuged at 500 × r.c.f. for 5 minutes at 4 °C. The cells were then washed by suspending the cell pellet with 1 mL of cold PBS. The cells were then transferred to another pre-cooled microcentrifuge tube and then centrifuged again at 500 × r.c.f. for 3 minutes at 4 °C. The supernatant was then carefully removed and discarded. NE-PER Nuclear and Cytoplasmic Extraction Reagent (ThermoFisher Scientific #78833) plus Halt Protease and Phosphatase Inhibitor Cocktail was used to lyse the cell pellets. Briefly, 500 µL of cytoplasmic extraction reagent I (CER I) was added to each tube and vortexed vigorously for 15 seconds. After a 10 minute incubation on ice, 27.5 µL cytoplasmic extraction reagent II (CER II) was added, vortexed for 5 seconds, then incubated on ice for 1 minute. The tube was then vortexed again for 5 seconds and centrifuged for 5 minutes at 16,000 × g. The supernatant (cytoplasmic extract) was then immediately transferred to a clean pre-chilled tube and stored at -80 °C. The pellet faction, which contained the nuclei, was then resuspended with 250 µL of ice-cold nuclear extraction reagent (NER) and vortexed vigorously for 15 seconds. The tube was then placed on ice and vortexed for 15 seconds in 10 minutes increments for a total of 40 minutes. The tube was then

centrifuged at 16,000 × g at 4 °C for 10 minutes. The supernatant (containing the nuclear extract) was then transferred to a clean pre-chilled tube and placed on ice. Meanwhile, 50 µL of Dynabeads™ Protein G (ThermoFisher Scientific #10004D) was washed in a microcentrifuge tube twice with 200 µL lysis buffer (4000g, 1 minute, 4 °C). 10 µg of either MUTYH (C-6) (Santa Cruz Biotechnology, sc-374571,) or Normal Mouse IgG_3 (Cell Signaling, 75952) antibody diluted in 200 µL of PBS (Gibco #10010-023) mixed with 0.1% Tween 20 (ThermoFisher Scientific, #85114; to make PBST) was then added to beads and incubated at 10 minutes at room temperature. The supernatant was removed, and the beads were gently washed with 200 µL of PBST. 250 µL of the clarified nuclear extracted cell lysate and 750 µL of PBST was then added to the prewashed beads and incubated for 1 hour on a rotation wheel at 4 °C. After incubation, the supernatant was removed using a magnet, and the beads were gently washed three times with 500 µL of PBST after each wash. After the last wash step, 50 µl of 2x Laemmli sample buffer (Bio-Rad, #1610737) was added to the beads, and the bound complexes were eluted by boiling for 5 minutes at 95 °C. The boiled samples were then loaded and electrophoresed on a 7.5% Criterion™ TGX™ Precast Midi Protein Gel (Bio-Rad #5671024). The protein was then transferred to a 0.45-µm PVDF transfer membrane (ThermoFisher Scientific, #88585) using the Mini Trans-Blot® Cell (Bio-Rad, # 1703930; 50 V, 60 minutes). The membrane was then washed with ultrapure water for two minutes with shaking and then treated with SuperSignal Western Blot Enhancer (ThermoFisher Scientific #46640). Briefly, the membrane was immersed with 10 mL of Antigen Pretreatment Solution and incubated at room temperature for 10 minutes with shaking. The solution was then discarded, and the membrane was then rinsed five times with 5 mL of ultrapure water. The membrane was then blocked with 5 mL of 5% nonfat dried milk in tris-buffered saline with 0.1% Tween 20 (TBST; 20 mM Tris, 150 mM NaCl, and 0.1% Tween 20 from ThermoFisher Scientific, #85114) for 1

hour, and then incubated overnight at 4 °C with the following primary antibodies diluted in 10 mL of TBST: MUTYH (C-6) (Santa Cruz Biotechnology #sc-374571, diluted 1:2,500); APE1 (Thermo Fisher Scientific #PA5-29157, diluted 1:2,500); Normal Mouse IgG_3 (Cell Signaling# 75952, diluted 1:3,000). The membrane was washed three times with 5 mL TBST for each wash, then incubated with VeriBlot for IP Detection Reagent (Abcam, #ab131366) diluted 1:100,00 in 10 mL TBST for 1 hour at room temperature. The membrane was then washed again three times with 5 mL TBST and soaked in 200 µL of chemiluminescent SuperSignal West Atto Ultimate Sensitivity Substrate (ThermoFisher Scientific, #A38554) and imaged with a Syngene G:Box Chemi XX6 imager.

3.4.10 Data analysis and statistics

Next-generation sequencing data were collected, demultiplexed, and trimmed with Illumina Local Run Manager Generate FASTQ analysis module (v2.0) and MiniSeq control software (v2.2.1). The FASTQ files were analyzed with CRISPResso2 (version 2.0.20b) on batch mode (parameters: --base_edit -wc -10 -w 10 -q 30) to assess genomic base editing efficiencies.[139] All editing efficiencies values were reported as nucleotide percentages around gRNA (from CRISPResso2). Sanger sequencing was analyzed by aligning the .ab1 trace file to a reference amplicon on Benchling (RRID:SCR_013955). To analyze data obtained from flow cytometers, the .fcs files were analyzed and quadrant plots were made using the FlowJo™ Software Version 10.8.1. by Becton, Dickinson and Company; 2021, 2022. Unpaired, one-tailed, parametric t-tests was performed when comparing unedited cells with homozygous *MUTYH* mutations using GraphPad Prism Version 10.2.1, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com. Plots were also made using GraphPad Prism.

3.5 Acknowledgments

This research was supported by the University of California, San Diego and the Research Corporation for Science Advancement through Cottrell Scholar Award no. 27502 (to A.C.K). C.A.V. was supported by the Howard Hughes Medical Institute Gilliam Fellowship Program and the National Academies of Sciences, Engineering, and Medicine Ford Foundation Predoctoral Fellowship Program. Q.T.C. was supported by the Molecular Biophysics Training Grant, NIH Grant T32 GM008326. A.C.K.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies.

Chapter 3 is reproduced, in full, with permission, from: Vasquez, C.A., Zepeda, M.U., Sandel, D.K., Cowan, Q.T., Peiris, M., Donoghue, D.J., Komor, A.K. (2024) Precision Genome Editing and In-Cell Measurements of Oxidative DNA Damage Repair Enable Functional Characterization of Cancer-Associated MUTYH Variants. *Submitted. Nucleic Acids Research.* The dissertation author was the primary author on all reprinted material

Chapter 4

Non-Thesis-Related University Service

4.1 Development of a Research-Practice Partnership with Local High Schools in San Diego: The Genome Editing Technologies Program

Outreach activities for broadening the early engagement of diverse groups is important to me. I knew going into graduate school that I wanted to figure out a way to leverage the experiences I have lived through, the good and the bad, as a way to connect with young students pursing science. As I explained to Michelle Franklin, who is the Director of Communications for the Department of School of Physical Sciences at UC San Diego, for the interview she did on the eventual program we developed, “It’s important for us to reach students who may not have even considered a career in STEM or medicine. To look in their eyes and instill confidence, to show we believe in them — having someone like that when I was in high school would have made a world of difference.”

Finding a PI to support me in this goal was vitally important especially given the rarity of this kind of feat being done. And from the very first day I met Alexis, I could tell she genuinely cared about sharing this goal with me. We subsequently discussed possible outreach ideas in out one-on-one's after I joined her lab. But it wasn't until one Saturday morning in late 2018, at Alexis's house, in a meeting with Mallory Evanoff, Alexis and I, that we established the foundation for the "Genome Editing Technologies Program". Initially, we came up with a hands-on activity where students would perform base editing on bacteria and if successful, the bacteria would fluoresce green. However, just doing a simple hands-on activity was not good enough for me. It wasn't until Alexis trusted me enough to ask me to help her with her NSF Career Grant application that this program became a reality. In the "Broader Impacts" section, I laid out the skeleton of a three-day program.

Over several years, Mallory, Brodie and I, fully developed this three-day outreach program into what it is now. The "Genome Editing Technologies Program" is a program centered around a 50-minute hands-on genome editing laboratory experiment with minimal equipment required and brings gene-editing technology to local high school students (**Figure 15**). It has been implemented in four high schools and has reached more than 200 students thus far. Besides genome editing, the program also allows us to introduce students to chemical biology – the ability to use chemistry to manipulate and better understand biological systems, such as human diseases. Most importantly, which is something I value the most about this program is it allows us to expose students to scientists from diverse backgrounds and invite questions about college, professional development, and the everyday life of a graduate student or faculty member within academia. The more we connect with the students throughout the three-day event, the more impact the last day is on the students. My work in leading this effort led to an EDI Excellence Award for the School of Physical

Sciences at UC San Diego in 2023. Furthermore, this outreach program has been highlighted in several news outlets such as UC San Diego Today and CBS8 San Diego.

4.1.1 Implementing the Genome Editing Technologies Program

Our goal was not only to make base editing accessible to high school students but also to have students think critically and reflect on base editing in a social and cultural context. We developed a 3-day program that centered around the following activities:

- Day 1: An interactive lecture on genome editing technologies (time: 50–90 min)
- Day 2 (**Figure 15**): A hands-on base editing experiment and discussion of ethics (time: 50–90 min)
- Day 3: Reviewing experimental results and an open forum panel discussion (time: 50–90 min). *Activity lengths can be adjusted according to the high school's classroom schedules.

To implement the "Genome Editing Technologies Program", we initially reached out to a local public high school with an advanced, elective biology curriculum. In our area, this was the Project Lead the Way (PLTW)'s biomedical sciences program. PLTW is a nonprofit educational organization that develops K-12 STEM curricula. Researchers can use PLTWs "school locator" to identify potential partners for implementing the Genome Editing Technologies Program. However, this program is not necessary for implementing the program; we suggest working with AP Biology, IB coursework students, or similar programs.

Our first iteration included three intermediate-level biology classes consisting of 20–25 students each. In total, 61 high school seniors participated in the first iteration of the Genome Editing Technologies Program. After our first implementation of the course, we were able to leverage connections the high school teacher had and reach out to two other high schools with

similar advanced biology elective courses. We proceeded to iterate and improve on the material with three classes at two new schools, reaching 67 additional students.

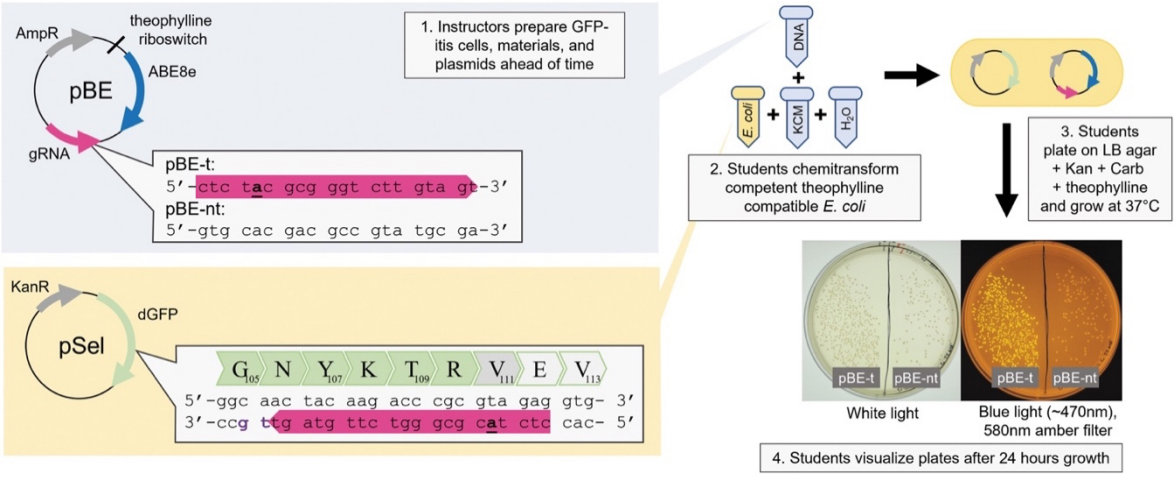

Figure 15. **Illustration of constructs used in the "GFP-itis" activity.** The inactivated *GFP* (*dGFP*) gene is in the pSel construct (bottom left, yellow background). Shown is the sequence of the *dGFP* gene, zoomed in on the inactivating A111V mutation that is corrected by editing of the target A to G. The protospacer is shown with a pink arrow, and the NG-PAM sequence is indicated with purple letters. The pBE plasmids (top left, blue background) contain the ABE8e editor under control of a theophylline-responsive riboswitch, and one of two gRNA sequences. In pBE-t, the gRNA matches the pSel *dGFP* mutation site (pink arrow) and will lead to correction of the *GFP* gene and green fluorescence. In pBE-nt, the gRNA has a nontargeting sequence and acts as a negative control. Instructors prepare all plasmids and, before student transformation, incorporate pSel into *Escherichia coli* to create "GFP-itis cells." Base editing activity (GFP fluorescence) can be visualized 24 h post-transformation (shown on the right). dGFP, dead green fluorescent protein.

4.1.2 Discussion on Student Experiences and Evaluation of the Program

As part of the outreach program, and after I worked to get Institutional Review Board (IRB) approval from the University of California San Diego (under IRB Project #200185, this enabled us to publish our survey results), we had students answer survey questions to identify areas where we could improve. The results from our feedback survey at our pilot school are shown in **Figure 16.** We first asked the students to evaluate the accessibility of the various components of the program. The students indicated that most of the components were accessible (e.g., 87% of the students [$n = 60$] indicated that the lecture was accessible, and 85% of the students [$n = 60$] indicated that the ethics discussion was accessible **Figure 16A**). However, student-evaluated accessibility of the worksheet was only 43%. We have since improved the worksheet using the students' feedback by further clarifying some questions and updating instructions for labeling diagrams. We have since improved the worksheet. We were encouraged to see that student-evaluated accessibility of the worksheet improved after using this updated worksheet (88% of students at our second school rated the worksheet as accessible, and 54% of students at our third school rated the worksheet as accessible.

We also evaluated the engagement of the program by asking the students to quantify how much they liked each component (ranging from a one being "not so much" to a five being "it was great!," **Figure 16B**). We also included an "open comments" section for anonymous feedback. The lowest rating was observed for the lecture, which overall scored a 3.6 out of 5 ($n = 59$), with 50 students indicating a 3 or higher. In the open comments section, several students commented favorably on the active learning elements of the lecture, prompting us to consider including more of these strategies in future designs. We have since incorporated more open discussion segments

within the slides that allow students to reflect on the material, including a seven-question Kahoot quiz.

Although the hands-on laboratory experiment was the most challenging element to prepare, it was the highest-rated activity in our program, with 88% of students (n = 60) saying the activity was accessible (**Figure 16A**), and an average rating of 4.61 out of 5 (n = 59; **Figure 16B**). The feedback results highlight the need for programs such as ours; not only were these students introduced to tools and techniques that are used in many areas of biology through the lecture, but also researchers provided students the opportunity to use them in a hands-on laboratory experiment, which has been shown to increase student learning and retention.

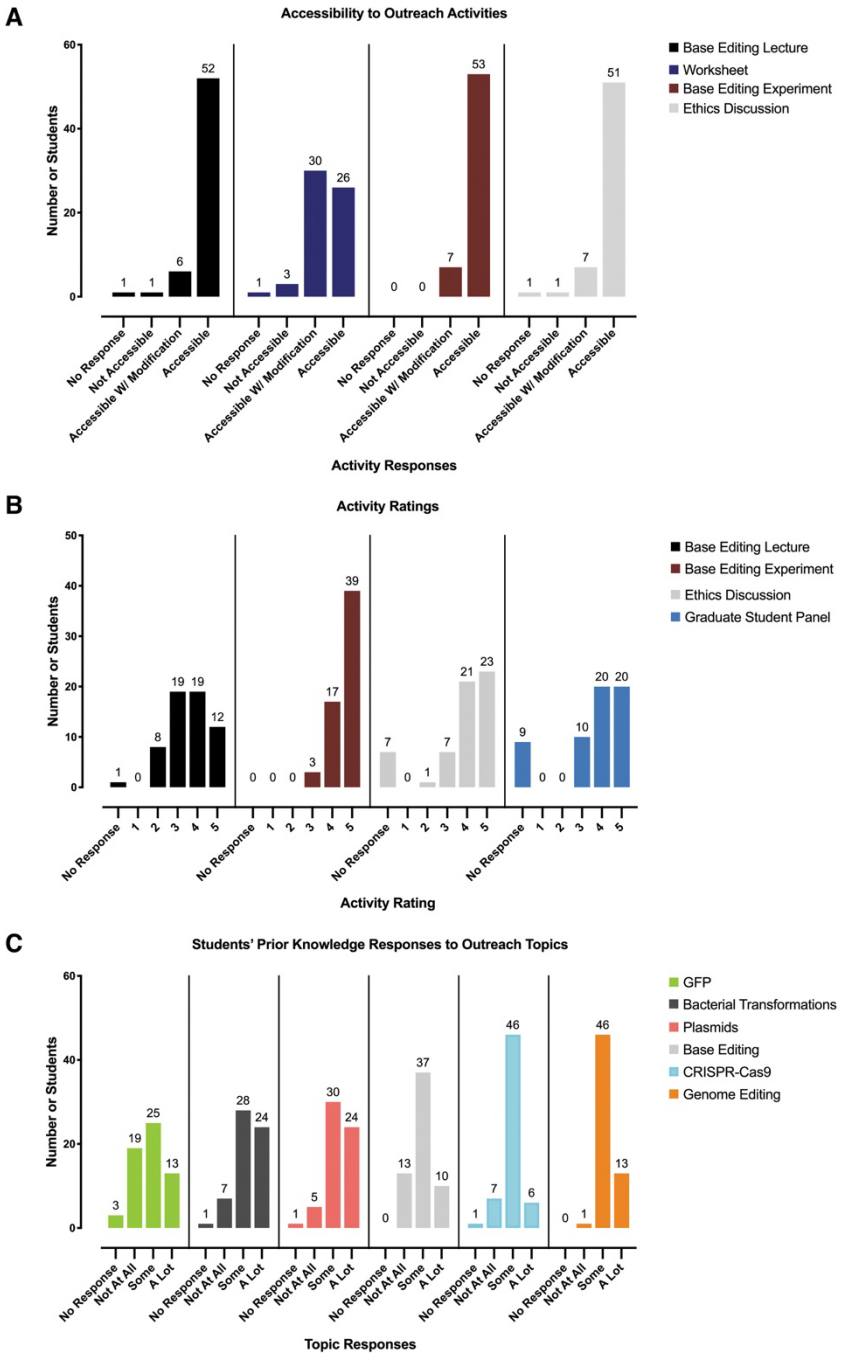

Figure 16. Student Survey Responses for Sage Creek High School. (A) Accessibility rating for each of the four outreach activities: (from left to right) the base editing lecture (87% accessibility rating), the base editing worksheet (47% accessibility rating), the base editing experiment (88% accessibility rating), and the ethics discussion (85% accessibility rating). In total, there were 60 student responses. (B) Activity rating of each of the outreach activities. Answer options ranged from 1 being "not so much" to 5 being "it was great!." Topics measured were (from left to right): the base editing lecture (3.60 activity rating average), the base editing experiment (4.61 activity rating average), the ethics activity (4.20 activity rating average), and the open forum discussion (4.27 activity rating average). In total, there were 59 student responses. (C) Measurements of the students' prior knowledge to various topics covered in the program. Answer options include "no response," "not at all," "some," and "a lot." The topics queried were (from left to right): GFP, bacterial transformations, plasmids, base editing, CRISPR-Cas9, and genome editing. In total, there were 60 student responses.

4.2 Teaching at UCSD Through the Summer Graduate Teaching Scholars Program

My passion for teaching stems from my childhood. I grew up as a native of South Vermont Avenue—a dangerous neighborhood in South Los Angeles known as "Death Alley" by the LA Times newspaper. Learning, teaching, being open-minded, trusting those who want the best for you. Care for you. It goes a long way towards surviving. I've always wanted to provide safety to others and I found a way to do that through mentoring and teaching. I had many valuable experiences at SMC and at UCI. And as a graduate student, I remained true to my aspirations to teach and mentor students. Besides being a TA, I enrolled into the *Introduction to College Teaching* course at the Engaged Teaching Hub, led by Erilynn Heinrichsen, Noel Martin, and Reina Muzrahi in the Fall of 2020. This course provided me with the teaching skills such as, how to build effecting learning outcomes, distinguish the differences between formative and summative assessments, lesson delivery, and how to better implement Growth Mindset teaching (and its variations suggested by Dr. J. Luke Wood, a Distinguished Professor of Education at San Diego State University). Furthermore, to continue my development as a college educator, I also used money I received from the Gilliam Fellowship to attend the Original Lily Conference on College Teaching where I was fortunate to attend many sessions led by international and U.S. teacher-scholars on evidence-based teaching practices to improve as an educator.

My career goal was always to become a professor of chemistry at a 8teaching-intensive institution. This was the optimal career goal for me because it would allow me to continue my passions in mentoring and teaching. However, getting a full-time tenure-track position right out of a PhD is an almost impossible task. It is very common in academia for PhD graduates who are interested in teaching at the college level to either pursue post-doctorate work or to teach part-

time. I did not want to pursue either of those options. The participation in Summer Graduate Teaching Scholars Program was critical to both my advancement as an educator and career aspirations. I applied, and was accepted to lead a General Chemistry course over the summer of 2022. I excelled in this role (according to student evaluations) in large part to my mentor, Dr. Thomas Bussey, who provided me not only with the teaching materials, but, the guidance from a world-class educator. The opportunity to practice leading my own class and the responsibilities that came with it would help me to further develop my teaching and leadership skills. Because of my success, the department invited me to teach the course again in the summer of 2023, adding valuable teaching experiences to my resume. These experiences all played a large role in securing my position at Butte College.

4.3 Acknowledgments

For the Genome Editing Technologies Program, we gratefully acknowledge the mentorship, guidance, and support from Dr. Thomas Bussey, Dr. Stacey Brydges (from the Department of Chemistry and Biochemistry at UC San Diego), and graduate student Monica Molgaard (in the Education Studies Program at UC San Diego). We also acknowledge support from laboratory members of the Komor Group, Zulfiqar Mohamedshah, Rachel Anderson, Sam Mawson, Michael Hollander, and Annika So. Thank you to Valerie Park and her students at Sage Creek High School, Dara Rosen and her students at Mission Vista High School, and Jessica Bosch and her students at University City High School. This study was approved by the University of California San Diego Institutional Review Board (IRB) under IRB Project #200185. C.A.V. was supported by the Howard Hughes Medical Institute Gilliam Fellowship Program and the National Academies of Sciences, Engineering, and Medicine Ford Foundation Predoctoral Fellowship

Program. A.C.K.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies.

Chapter 4 is reproduced, in full, with permission, from: Vasquez, C.A., Evanoff, M., Ranzau, B., Gu, S., Deters, Emma; Komor, A.C. (2023). Curing "GFP-itis" in Bacteria with Base Editors: Development of a Genome Editing Science Program Implemented with High School Biology Students. *CRISPR J.* **6**, 3:186-195. The dissertation author was the primary author on all reprinted material.

Table 1. List of commonly used CRISPR-based genome editing tools.

Improvement	Tool	Notes
Programmability	wild-type *Streptococcus pyogenes* (Sp) Cas9	NGG PAM on 3' end; 1368 aa in length
	wild-type *Staphylococcus aureus* (Sa) Cas9	NNGRRT PAM on 3' end; 1053 aa in length
	wild-type *Neisseria meningitidis* (Nm) Cas9	NNNNGATT PAM on 3' end; 1082 aa in length;
	wild-type *Lachnospiraceae bacterium* (Lb) Cas12a	TTTV PAM on 5' end; 1228 aa in length
	Bacillus hisashii (Bh) Cas12b v4 mutant (K846R/S893R/E837G)	ATTN PAM on 5' end; 1120 aa in length
	VQR SpCas9 (D1135V/R1335Q/T1337R)	NGA PAM on 3' end
	VRER SpCas9 (D1135V/G1218R/R1335E/T1337R)	NGCG PAM on 3' end
	SpCas9-NG (R1335V/L1111R/D1135V/ G1218R/E1219F/A1322R/T1337R)	NG PAM on 3' end
	SpRY SpCas9 (D1135L/S1136W/G1218K/ E1219Q/R1335Q/T1337R)	NRN (and to lesser extent NYN) PAM on 3' end
	KKH SaCas9 (E782K/N968K/R1015H)	NNNRRT PAM on 3' end
Specificity	eSpCas9-1.1	K848A/K1060A/K1003A mutations in SpCas9
	Cas9-HF	N497A/R661A/Q926A/Q695A mutations in SpCas9
	HypaCas9	N692A/M694A/Q695A/H698A mutations in SpCas9
	Sniper-Cas9	M763I/K890N/F539S mutations in SpCas9
	Evo-Cas9	M495V/Y515N/R661Q/K526E mutations in SpCas9
	SpCas9n (D10A or H840A)	Coordinate two nicking events
	fCas9	FokI-dCas9 fusion; coordinate two binding events

Table 2. CRISPR-based genome editing software.

Software	Use
CHOPCHOP https://chopchop.cbu.uib.no/	Cas9, Cas12, and Cas13 gRNA design and evaluation (on-target efficiency prediction and off-target prediction)
Cas-OFFinder http://www.rgenome.net/cas-offinder/	Cas9 and Cas12 gRNA off-target prediction
CRISPOR http://crispor.tefor.net/	Cas9 and Cas12 gRNA design and evaluation
Horizon CRISPR Design Tool https://horizondiscovery.com/en/ordering-and-calculation-tools/crispr-design-tool	Cas9 gRNA design for gene KOs
IDT Design Tool https://www.idtdna.com/site/order/designtool/index/CRISPR_CUSTOM	Cas9 gRNA design and evaluation
Off-Spotter https://cm.jefferson.edu/Off-Spotter/	Cas9 gRNA off-target prediction
Synthego CRISPR Design Tool https://design.synthego.com/#/	Cas9 gRNA design for gene KOs
BE-Designer http://www.rgenome.net/be-designer/	automated BE gRNA design
inDelphi https://indelphi.giffordlab.mit.edu/	prediction of indel sequences for a given protospacer
BE-Hive https://www.crisprbehive.design/	prediction of CBE and ABE efficiencies
PrimeDesign https://drugthatgene.pinellolab.partners.org/	pegRNA automated design

Table 3. Suggested base editors designed to maximize on-target editing or minimize off-target (OT) editing.

	High-efficiency	Addgene #	Reduced off-target editing	Notes	Addgene #
CBE	BE4max-NG-P2A-EGFP	125616, 140001	BE4max-NG (YE1)	Less processive, narrowed window	138159
	RrAPOBEC3F	138340	RrAPOBEC3F (F130L)	Retains high on-target activity	138341
	PpAPOBEC1	138349	PpAPOBEC1 (H122A)	Minimal OTs, slightly reduced on-target activity	138345, 138338
	SsAPOBEC3B	138343	SsAPOBEC3B (R54Q)	Minimal OTs, "BC" preference	138344
	AncBE4max-P2A-EGFP	112100	eA3A-BE3 (N57G)	"(A)UC" preference	131315
ABE	ABEmax-NG-P2A-EGFP	140005	evo-TadA (V106W)	Inactivated or deleted wt-TadA	125647, 138495
	NG-ABE8e	138491	evo-TadA (F148A)		131313
	ABE8.20-m	136300	evo-TadA (V82G)		

Table 4. Primers and additional sequences used to generate custom gRNAs or amplify the target locus.

Underlined sequences anneal to the gRNA plasmid backbone; Bold sequences are protospacers (replace this with your custom protospacer).

Usage	#	Primer	Sequence	Notes
gRNA cloning	1	universal reverse	GGTGTTTCGTCCTTTCCACAAG	Site-directed mutagenesis (of spacer region)
	2	V270 gRNA forward	GGTCCACCAGCTGCTGGGCTGTTTTAGAGCTAGAAATAGCA	
	3	E303 gRNA forward (adds 5'G)	GACAGGCTCTCCACAGGGCACGTTTTAGAGCTAGAAATAGCA	
	4	L296 gRNA forward (adds 5'G)	**GCCACTGTGCAGCCAGTGCCC**GTTTTAGAGCTAGAAATAGCA	
	5	W12 (exon 1 pmSTOP) gRNA reverse	**GTACCCACAGACGACTCAGG**GTTTTAGAGCTAGAAATAGCA	
	6	non-targeting gRNA spacer	GGTATTACTGATATTGGTGGG	Spacer sequence only
Sanger sequencing	7	gRNA sequencing 1 (U6)	TACGTGACGTAGAAAGTAAT	To verify gRNA clones
	8	gRNA sequencing 2 (ColE1)	TTTCTCCCTTCGGGAAGCGT	
	9	gRNA sequencing 3 (backbone)	GAGATTATCAAAAAGGATCT	
	10	gRNA sequencing 4 (AmpR)	TTGTGCAAAAAAGCGGTTAG	
	11	Sanger forward* *MUTYH* codons 193–332	AGGAGGTGAATCAACTCTGGGC	*Used to sequence V270, L296, E303
	12	Sanger reverse *MUTYH* codons 193–332	CCGAACCCTACTCAAGCCAAGA	
NGS	13	NGS rd1 adapter forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN	Add sequence to 5' end of target-specific primers
	14	NGS rd1 adapter reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCT	
	15	NGS rd1 *MUTYH* codons 264–291 forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCAGCTCTGGTAGGATGTTGG	Rd1 primers for V270 example
	16	NGS rd1 *MUTYH* codons 264–291 reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGTAGGCTTACTCTCTGGC	

Table 5. List of primers used for PCR amplification to produce gRNA plasmids

Use	Number	Primer	Sequence
gRNA Cloning	0	Universal FWD	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
	1	W12* gRNA REV	CCTGAGTCGTCTGTGGGTACGGTGTTTCGTCCTTTCCACAAG
	2	P18L and CRISPRi gRNA REV	ACGGCTGCTCGTGGCTTCCTCGGTGTTTCGTCCTTTCCACAAG
	3	L111P gRNA REV	AGAGAAACGGGACCTACCATCGGTGTTTCGTCCTTTCCACAAG
	4	Y128H gRNA REV	GGACAGGCGGGCATATGCTGCGGTGTTTCGTCCTTTCCACAAG
	5	W131* gRNA REV	TTTCCCCCAGTGTGGGTCTCGGTGTTTCGTCCTTTCCACAAG
	6	Y179C gRNA REV	GGCCACGAGAATAGTAGCCCGGTGTTTCGTCCTTTCCACAAG
	7	R182C gRNA REV	GCAGCCGCCGGCCACGAGAACGGTGTTTCGTCCTTTCCACAAG
	8	G189E gRNA REV	GCGGCTGCAGGAGGGAGCTCGGTGTTTCGTCCTTTCCACAAG
	9	I223V gRNA REV	AAGGCGATAGAGGCAATGGCGGTGTTTCGTCCTTTCCACAAG
	10	R241W gRNA REV	GACACGGCACAGCACCCGTGCGGTGTTTCGTCCTTTCCACAAG
	11	R245C gRNA REV	CCAATGGCTCGGACACGGCACGGTGTTTCGTCCTTTCCACAAG
	12	V246I gRNA REV	CGGGTGCTGTGCCGTGTCCGCGGTGTTTCGTCCTTTCCACAAG
	13	Q260* gRNA REV	TCCTACCAGAGCTGCTGGGACGGTGTTTCGTCCTTTCCACAAG
	14	D271G gRNA REV	CCTGGCCGGGCTGGGTCCACGGTGTTTCGTCCTTTCCACAAG
	15	P295L gRNA REV	CTGGCTGCACAGTGGGCGCTCGGTGTTTCGTCCTTTCCACAAG
	16	L296L gRNA REV	GGGCACTGGCTGCACAGTGGCGGTGTTTCGTCCTTTCCACAAG
	17	E303E gRNA REV	GTGCCCTGTGGAGAGCCTGTCGGTGTTTCGTCCTTTCCACAAG
	18	S304N gRNA REV	GTGCCCTGTGGAGAGCCTGTCGGTGTTTCGTCCTTTCCACAAG

Table 6. List of primers used for Sanger sequencing and Next-generation sequencing (NGS) for *MUTYH* genomic DNA

Use	Number	Primer	Sequence
Sanger Sequencing	19	gRNA plasmid sequencing (U6)	TACGTGACGTAGAAAGTAAT
	20	Sanger FWD: MUTYH codons 1-12	TCTCCCAGAGCGCAGAGGCTTT
	21	Sanger REV: MUTYH codons 1-12	CTCCTAGTCTAACTCCTGGGCGTGC
	22	Sanger FWD: MUTYH codons 16-55	GCAGAGAAACCGCCTACCCCCA
	23	Sanger REV: MUTYH codons 16-55	CTACAGACGCTCACCACCACGC
	24	Sanger FWD: MUTYH codons 56-168	AGCCAGTAGTACCACCCTGAGA
	25	Sanger REV: MUTYH codons 56-168	GCCCAGAGTTGATTCACCTCCT
	26	Sanger FWD: MUTYH codons 176-332	AGGAGGTGAATCAACTCTGGGC
	27	Sanger REV: MUTYH codons 176-332	CCGAACCCTACTCAAGCCAAGA
NGS	28	NGS rd1 adapter FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN
	29	NGS rd1 adapter REV	TGGAGTTCAGACGTGTGCTCTTCCGATCT
	30	NGS rd1 *MUTYH* D271G and L296L FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCAGCTCTGGTAGGATGTTGG
	31	NGS rd1 *MUTYH* D271G and L296L REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGTAGGCTTACTCTCTGGC
	32	NGS rd1 *MUTYH* W131* FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGGCGGGCATATGCTGGTCAG
	33	NGS rd1 *MUTYH* W131* REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCCTGGAGTCACCTGCATCCA

REFERENCES

1. Avery OT, MacLeod CM, McCarty M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *J Exp Med*. 1944;79(2):137-158.

2. KaiserSep. 5 J, 2018, Am 10:45. New gene-editing treatment might help treat a rare disorder, hints first human test. Science | AAAS. Published September 5, 2018. Accessed May 8, 2021. https://www.sciencemag.org/news/2018/09/new-gene-editing-treatment-might-help-treat-rare-disorder-hints-first-human-test

3. Choo Y, Klug A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A*. 1994;91(23):11168-11172. doi:10.1073/pnas.91.23.11168

4. Mojica FJM, Ferrer C, Juez G, Rodríguez-Valera F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular Microbiology*. 1995;17(1):85-93. doi:https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x

5. Jansen R, Embden JDA van, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. 2002;43(6):1565-1575. doi:10.1046/j.1365-2958.2002.02839.x

6. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol*. 2005;60(2):174-182. doi:10.1007/s00239-004-0046-3

7. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology,*. 2005;151(3):653-663. doi:10.1099/mic.0.27437-0

8. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012;337(6096):816-821. doi:10.1126/science.1225829

9. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS*. 2012;109(39):E2579-E2586. doi:10.1073/pnas.1208507109

10. Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 2010;468(7320):67-71. doi:10.1038/nature09523

11. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-Guided Human Genome Engineering via Cas9. *Science*. 2013;339(6121):823-826. doi:10.1126/science.1232033

12. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819-823. doi:10.1126/science.1231143

13. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology*. 2013;31(3):230-232. doi:10.1038/nbt.2507

14. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Weigel D, ed. *eLife*. 2013;2:e00471. doi:10.7554/eLife.00471

15. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012;337(August):816-822. doi:10.1126/science.1225829

16. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533(7603):420-424. doi:10.1038/nature17946

17. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*. 2016;351(6275):867-871. doi:10.1126/science.aad8282

18. Wood RD. Dna repair in eukaryotes. *Annu Rev Biochem*. 1996;65(1):135-167. doi:10.1146/annurev.bi.65.070196.001031

19. Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH, Liu DR. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv*. 2017;3(8):eaao4774. doi:10.1126/sciadv.aao4774

20. Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol*. 2017;35(4):371-376. doi:10.1038/nbt.3803

21. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z, Kondo A. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*. 2016;353(6305). doi:10.1126/science.aaf8729

22. Rees HA, Komor AC, Yeh WH, Caetano-Lopes J, Warman M, Edge ASB, Liu DR. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nature Communications*. 2017;8(1):15790. doi:10.1038/ncomms15790

23. Li X, Wang Y, Liu Y, Yang B, Wang X, Wei J, Lu Z, Zhang Y, Wu J, Huang X, Yang L, Chen J. Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol*. 2018;36(4):324-327. doi:10.1038/nbt.4102

24. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, Maianti JP, Raguram A, Liu DR. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol*. 2018;36(9):843-846. doi:10.1038/nbt.4172

25. Zafra MP, Schatoff EM, Katti A, Foronda M, Breinig M, Schweitzer AY, Simon A, Han T, Goswami S, Montgomery E, Thibado J, Kastenhuber ER, Sánchez-Rivera FJ, Shi J, Vakoc CR, Lowe SW, Tschaharganeh DF, Dow LE. Optimized base editors enable efficient editing in cells, organoids and mice. *Nature Biotechnology*. 2018;36(9):888-893. doi:10.1038/nbt.4194

26. Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, Pinello L, Joung JK. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol*. 2018;36(10):977-982. doi:10.1038/nbt.4199

27. Wu Y, Xu W, Wang F, Zhao S, Feng F, Song J, Zhang C, Yang J. Increasing Cytosine Base Editing Scope and Efficiency With Engineered Cas9-PmCDA1 Fusions and the Modified sgRNA in Rice. *Front Genet*. 2019;10:379. doi:10.3389/fgene.2019.00379

28. Tan J, Zhang F, Karcher D, Bock R. Expanding the genome-targeting scope and the site selectivity of high-precision base editors. *Nature Communications*. 2020;11(1):629. doi:10.1038/s41467-020-14465-z

29. Yu Y, Leete TC, Born DA, Young L, Barrera LA, Lee SJ, Rees HA, Ciaramella G, Gaudelli NM. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nature Communications*. 2020;11(1):2052. doi:10.1038/s41467-020-15887-5

30. Doman JL, Raguram A, Newby GA, Liu DR. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat Biotechnol*. 2020;38(5):620-628. doi:10.1038/s41587-020-0414-6

31. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*. 2017;551(7681):464-471. doi:10.1038/nature24644

32. Vasquez CA, Cowan QT, Komor AC. Base Editing in Human Cells to Produce Single-Nucleotide-Variant Clonal Cell Lines. *Curr Protoc Mol Biol*. 2020;133(1):e129. doi:10.1002/cpmb.129

33. Huang TP, Newby GA, Liu DR. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nature Protocols*. 2021;16(2):1089-1128. doi:10.1038/s41596-020-00450-9

34. Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, Grünewald J, Joung JK. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol*. 2021;39(1):41-46. doi:10.1038/s41587-020-0609-x

35. Zhao D, Li J, Li S, Xin X, Hu M, Price MA, Rosser SJ, Bi C, Zhang X. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nature Biotechnology*. 2021;39(1):35-40. doi:10.1038/s41587-020-0592-2

36. Chen L, Park JE, Paa P, Rajakumar PD, Prekop HT, Chew YT, Manivannan SN, Chew WL. Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nature Communications*. 2021;12(1):1384. doi:10.1038/s41467-021-21559-9

37. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*. 2020;368(6488):290-296. doi:10.1126/science.aba8853

38. Gaudelli NM, Lam DK, Rees HA, Solá-Esteves NM, Barrera LA, Born DA, Edwards A, Gehrke JM, Lee SJ, Liquori AJ, Murray R, Packer MS, Rinaldi C, Slaymaker IM, Yen J, Young LE, Ciaramella G. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol*. Published online April 13, 2020. doi:10.1038/s41587-020-0491-6

39. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, Doudna JA, Liu DR. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol*. Published online March 16, 2020. doi:10.1038/s41587-020-0453-z

40. Lee HK, Willi M, Miller SM, Kim S, Liu C, Liu DR, Hennighausen L. Targeting fidelity of adenine and cytosine base editors in mouse embryos. *Nature Communications*. 2018;9(1):4804. doi:10.1038/s41467-018-07322-7

41. Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu JL, Zhang F, Gao C. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*. 2019;364(6437):292-295. doi:10.1126/science.aaw7166

42. Grünewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, Joung JK. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*. 2019;569(7756):433-437. doi:10.1038/s41586-019-1161-z

43. Zhou C, Sun Y, Yan R, Liu Y, Zuo E, Gu C, Han L, Wei Y, Hu X, Zeng R, Li Y, Zhou H, Guo F, Yang H. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature*. 2019;571(7764):275-278. doi:10.1038/s41586-019-1314-0

44. Rees HA, Wilson C, Doman JL, Liu DR. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Science Advances*. 2019;5(5):eaax5717. doi:10.1126/sciadv.aax5717

45. Grünewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, Joung JK. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol*. 2019;37(9):1041-1048. doi:10.1038/s41587-019-0236-6

46. Chadwick AC, Wang X, Musunuru K. In Vivo Base Editing of PCSK9 as a Therapeutic Alternative to Genome Editing. *Arterioscler Thromb Vasc Biol*. 2017;37(9):1741-1747. doi:10.1161/ATVBAHA.117.309881

47. Villiger L, Grisch-Chan HM, Lindsay H, Ringnalda F, Pogliano CB, Allegri G, Fingerhut R, Häberle J, Matos J, Robinson MD, Thöny B, Schwank G. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nature Medicine*. 2018;24(10):1519-1525. doi:10.1038/s41591-018-0209-1

48. Zeng J, Wu Y, Ren C, Bonanno J, Shen AH, Shea D, Gehrke JM, Clement K, Luk K, Yao Q, Kim R, Wolfe SA, Manis JP, Pinello L, Joung JK, Bauer DE. Therapeutic base editing of human hematopoietic stem cells. *Nat Med*. 2020;26(4):535-541. doi:10.1038/s41591-020-0790-y

49. Wang L, Li L, Ma Y, Hu H, Li Q, Yang Y, Liu W, Yin S, Li W, Fu B, Kurita R, Nakamura Y, Liu M, Lai Y, Li D. Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. *Cell Research*. 2020;30(3):276-278. doi:10.1038/s41422-019-0267-z

50. Koblan LW, Erdos MR, Wilson C, Cabral WA, Levy JM, Xiong ZM, Tavarez UL, Davison LM, Gete YG, Mao X, Newby GA, Doherty SP, Narisu N, Sheng Q, Krilow C, Lin CY, Gordon LB, Cao K, Collins FS, Brown JD, Liu DR. In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature*. 2021;589(7843):608-614. doi:10.1038/s41586-020-03086-7

51. Shendure J, Akey JM. The origins, determinants, and consequences of human mutations. *Science*. 2015;349(6255):1478-1483. doi:10.1126/science.aaa9119

52. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM, Broad GO, Seattle GO, NHLBI Exome Sequencing Project. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337(6090):64-69. doi:10.1126/science.1219240

53. Ipe J, Swart M, Burgess K, Skaar T. High-Throughput Assays to Assess the Functional Impact of Genetic Variants: A Road Towards Genomic-Driven Medicine. *Clin Transl Sci*. 2017;10(2):67-77. doi:10.1111/cts.12440

54. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, Consortium GAD, Rehm HL, MacArthur DG, O'Donnell-Luria A. Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation*. 2022;43(8):1012-1030. doi:10.1002/humu.24309

55. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067. doi:10.1093/nar/gkx1153

56. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature*. 2011;475(7355):163-165. doi:10.1038/475163a

57. Gouveia MH, Bentley AR, Leal TP, Tarazona-Santos E, Bustamante CD, Adeyemo AA, Rotimi CN, Shriner D. Unappreciated subcontinental admixture in Europeans and European Americans and implications for genetic epidemiology studies. *Nat Commun*. 2023;14(1):6802. doi:10.1038/s41467-023-42491-0

58. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. Genome-wide association studies. *Nat Rev Methods Primers*. 2021;1(1):1-21. doi:10.1038/s43586-021-00056-9

59. Out A, Tops C, Nielsen M, Weiss M, van Minderhout I, Fokkema I, Buisine M, Claes K, Colas C, Fodde R, Fostira F, Franken P, Gaustadnes M, Heinimann K, Hodgson S, Hogervorst F, Holinski-Feder E, Lagerstedt-Robinson K, Olschwang S, van den Ouweland A, Redeker E, Scott R, Vankeirsbilck B, Gronlund R, Wijnen J, Wikman F, Aretz S, Sampson J, Devilee P, Dunnen J, Hes F. Leiden Open Variation Database of the MUTYH Gene. *Human Mutation*. 2010;31(11):1205-1215. doi:10.1002/humu.21343

60. Slatter MA, Gennery AR. Primary immunodeficiencies associated with DNA-repair disorders. *Expert Rev Mol Med*. 2010;12:e9. doi:10.1017/S1462399410001419

61. Maynard S, Fang EF, Scheibye-Knudsen M, Croteau DL, Bohr VA. DNA Damage, DNA Repair, Aging, and Neurodegeneration. *Cold Spring Harb Perspect Med*. 2015;5(10). doi:10.1101/cshperspect.a025130

62. Arbab M, Shen MW, Mok B, Wilson C, Matuszek Ż, Cassa CA, Liu DR. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell*. 2020;182(2):463-480.e30. doi:10.1016/j.cell.2020.05.037

63. Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, Noda T, Abudayyeh OO, Gootenberg JS, Mori H, Oura S, Holmes B, Tanaka M, Seki M, Hirano H, Aburatani H, Ishitani R, Ikawa M, Yachie N, Zhang F, Nureki O. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*. 2018;361(6408):1259-1262. doi:10.1126/science.aas9129

64. Phelan MC. Techniques for mammalian cell tissue culture. *Curr Protoc Protein Sci*. 2006;Appendix 3:Appendix 3C. doi:10.1002/0471140864.psa03cs46

65. Veeranagouda Y, Debono-Lagneaux D, Fournet H, Thill G, Didier M. CRISPR-Cas9-Edited Site Sequencing (CRES-Seq): An Efficient and High-Throughput Method for the

Selection of CRISPR-Cas9-Edited Clones. *Current Protocols in Molecular Biology*. 2018;121(1):31.14.1-31.14.11. doi:10.1002/cpmb.53

66. Yang L, Yang JL, Byrne S, Pan J, Church GM. CRISPR/Cas9-Directed Genome Editing of Cultured Cells. *Current Protocols in Molecular Biology*. 2014;107(1):31.1.1-31.1.17. doi:10.1002/0471142727.mb3101s107

67. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol*. 2019;37(3):224-226. doi:10.1038/s41587-019-0032-3

68. Giuliano CJ, Lin A, Girish V, Sheltzer JM. Generating Single Cell–Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. *Current Protocols in Molecular Biology*. 2019;128(1):e100. doi:10.1002/cpmb.100

69. Stacey GN, Masters JR. Cryopreservation and banking of mammalian cell lines. *Nat Protoc*. 2008;3(12):1981-1989. doi:10.1038/nprot.2008.190

70. Yokoyama WM, Thompson ML, Ehrhardt RO. Cryopreservation and Thawing of Cells. *Current Protocols in Immunology*. 2012;99(1):A.3G.1-A.3G.5. doi:10.1002/0471142735.ima03gs99

71. Hwang GH, Park J, Lim K, Kim S, Yu J, Yu E, Kim ST, Eils R, Kim JS, Bae S. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics*. 2018;19(1):542. doi:10.1186/s12859-018-2585-4

72. Wang D, Zhang C, Wang B, Li B, Wang Q, Liu D, Wang H, Zhou Y, Shi L, Lan F, Wang Y. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nature Communications*. 2019;10(1):4284. doi:10.1038/s41467-019-12281-8

73. Tycko J, Myer VE, Hsu PD. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular Cell*. 2016;63(3):355-370. doi:10.1016/j.molcel.2016.07.004

74. Dang Y, Jia G, Choi J, Ma H, Anaya E, Ye C, Shankar P, Wu H. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology*. 2015;16(1):280. doi:10.1186/s13059-015-0846-3

75. Smits AH, Ziebell F, Joberty G, Zinn N, Mueller WF, Clauder-Münster S, Eberhard D, Fälth Savitski M, Grandi P, Jakob P, Michon AM, Sun H, Tessmer K, Bürckstümmer T, Bantscheff M, Steinmetz LM, Drewes G, Huber W. Biological plasticity rescues target activity in CRISPR knock outs. *Nature Methods*. 2019;16(11):1087-1093. doi:10.1038/s41592-019-0614-5

76. Webber BR, Lonetree CL, Kluesner MG, Johnson MJ, Pomeroy EJ, Diers MD, Lahr WS, Draper GM, Slipek NJ, Smeester BA, Lovendahl KN, McElroy AN, Gordon WR, Osborn MJ, Moriarity BS. Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nat Commun*. 2019;10(1):5222. doi:10.1038/s41467-019-13007-6

77. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*. 2020;368(6488):290-296. doi:10.1126/science.aba8853

78. Grav LM, Sergeeva D, Lee JS, Marin de Mas I, Lewis NE, Andersen MR, Nielsen LK, Lee GM, Kildegaard HF. Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. *ACS Synth Biol*. 2018;7(9):2148-2159. doi:10.1021/acssynbio.8b00140

79. Kluesner MG, Nedveck DA, Lahr WS, Garbe JR, Abrahante JE, Webber BR, Moriarity BS. EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal*. 2018;1(3):239-250. doi:10.1089/crispr.2018.0014

80. Lin YC, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, Reumers J, Moisse M, Plaisance S, Drmanac R, Chen J, Speleman F, Lambrechts D, Van de Peer Y, Tavernier J, Callewaert N. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun*. 2014;5. doi:10.1038/ncomms5767

81. Paquet D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, Olsen KM, Gregg A, Noggle S, Tessier-Lavigne M. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*. 2016;533(7601):125-129. doi:10.1038/nature17664

82. Kim SI, Matsumoto T, Kagawa H, Nakamura M, Hirohata R, Ueno A, Ohishi M, Sakuma T, Soga T, Yamamoto T, Woltjen K. Microhomology-assisted scarless genome editing in human iPSCs. *Nat Commun*. 2018;9. doi:10.1038/s41467-018-03044-y

83. Ousterout DG, Kabadi AM, Thakore PI, Majoros WH, Reddy TE, Gersbach CA. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat Commun*. 2015;6:6244. doi:10.1038/ncomms7244

84. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308. doi:10.1038/nprot.2013.143

85. Riesenberg S, Chintalapati M, Macak D, Kanis P, Maricic T, Pääbo S. Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Res*. 2019;47(19):e116. doi:10.1093/nar/gkz669

86. Yeh CD, Richardson CD, Corn JE. Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol*. 2019;21(12):1468-1478. doi:10.1038/s41556-019-0425-z

87. Poulsen MLM, Bisgaard ML. MUTYH Associated Polyposis (MAP). *Curr Genomics*. 2008;9(6):420-435. doi:10.2174/138920208785699562

88. Georgeson P, Harrison TA, Pope BJ, Zaidi SH, Qu C, Steinfelder RS, Lin Y, Joo JE, Mahmood K, Clendenning M, Walker R, Amitay EL, Berndt SI, Brenner H, Campbell PT, Cao Y, Chan AT, Chang-Claude J, Doheny KF, Drew DA, Figueiredo JC, French AJ, Gallinger S, Giannakis M, Giles GG, Gsur A, Gunter MJ, Hoffmeister M, Hsu L, Huang

WY, Limburg P, Manson JE, Moreno V, Nassir R, Nowak JA, Obón-Santacana M, Ogino S, Phipps AI, Potter JD, Schoen RE, Sun W, Toland AE, Trinh QM, Ugai T, Macrae FA, Rosty C, Hudson TJ, Jenkins MA, Thibodeau SN, Winship IM, Peters U, Buchanan DD. Identifying colorectal cancer caused by biallelic MUTYH pathogenic variants using tumor mutational signatures. *Nat Commun*. 2022;13(1):3254. doi:10.1038/s41467-022-30916-1

89. Farrington SM, Tenesa A, Barnetson R, Wiltshire A, Prendergast J, Porteous M, Campbell H, Dunlop MG. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet*. 2005;77(1):112-119. doi:10.1086/431213

90. Webb EL, Rudd MF, Houlston RS. Colorectal Cancer Risk in Monoallelic Carriers of MYH Variants. *Am J Hum Genet*. 2006;79(4):768-771.

91. Peterlongo P, Mitra N, Sanchez de Abajo A, de la Hoya M, Bassi C, Bertario L, Radice P, Glogowski E, Nafa K, Caldes T, Offit K, Ellis NA. Increased frequency of disease-causing MYH mutations in colon cancer families. *Carcinogenesis*. 2006;27(11):2243-2249. doi:10.1093/carcin/bgl093

92. Venesio T, Balsamo A, D'Agostino VG, Ranzani GN. MUTYH-associated polyposis (MAP), the syndrome implicating base excision repair in inherited predisposition to colorectal tumors. *Front Oncol*. 2012;2:83. doi:10.3389/fonc.2012.00083

93. Win AK, Dowty JG, Cleary SP, Kim H, Buchanan DD, Young JP, Clendenning M, Rosty C, MacInnis RJ, Giles GG, Boussioutas A, Macrae FA, Parry S, Goldblatt J, Baron JA, Burnett T, Le Marchand L, Newcomb PA, Haile RW, Hopper JL, Cotterchio M, Gallinger S, Lindor NM, Tucker KM, Winship IM, Jenkins MA. Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. *Gastroenterology*. 2014;146(5):1208-1211.e1-5. doi:10.1053/j.gastro.2014.01.022

94. Curia MC, Catalano T, Aceto GM. MUTYH: Not just polyposis. *World J Clin Oncol*. 2020;11(7):428-449. doi:10.5306/wjco.v11.i7.428

95. Ohno M, Miura T, Furuichi M, Tominaga Y, Tsuchimoto D, Sakumi K, Nakabeppu Y. A genome-wide distribution of 8-oxoguanine correlates with the preferred regions for recombination and single nucleotide polymorphism in the human genome. *Genome Res*. 2006;16(5):567-575. doi:10.1101/gr.4769606

96. Nakabeppu Y. Cellular Levels of 8-Oxoguanine in either DNA or the Nucleotide Pool Play Pivotal Roles in Carcinogenesis and Survival of Cancer Cells. *Int J Mol Sci*. 2014;15(7):12543-12557. doi:10.3390/ijms150712543

97. Shinmura K, Yokota J. The OGG1 Gene Encodes a Repair Enzyme for Oxidatively Damaged DNA and Is Involved in Human Carcinogenesis. *Antioxidants & Redox Signaling*. 2001;3(4):597-609. doi:10.1089/15230860152542952

98. Kairupan C, Scott RJ. Base excision repair and the role of MUTYH. *Hered Cancer Clin Pract*. 2007;5(4):199-209. doi:10.1186/1897-4287-5-4-199

99. Krokan HE, Bjørås M. Base Excision Repair. *Cold Spring Harb Perspect Biol*. 2013;5(4):a012583. doi:10.1101/cshperspect.a012583

100. Mazzei F, Viel A, Bignami M. Role of MUTYH in human cancer. *Mutat Res*. 2013;743-744:33-43. doi:10.1016/j.mrfmmm.2013.03.003

101. Oka S, Leon J, Tsuchimoto D, Sakumi K, Nakabeppu Y. MUTYH, an adenine DNA glycosylase, mediates p53 tumor suppression via PARP-dependent cell death. *Oncogenesis*. 2014;3(10):e121. doi:10.1038/oncsis.2014.35

102. Raetz AG, David SS. When you're strange: Unusual features of the MUTYH glycosylase and implications in cancer. *DNA Repair (Amst)*. 2019;80:16-25. doi:10.1016/j.dnarep.2019.05.005

103. Kasai H, Nishimura S. Hydroxylation of deoxyguanosine at the C-8 position by ascorbic acid and other reducing agents. *Nucleic Acids Res*. 1984;12(4):2137-2145. doi:10.1093/nar/12.4.2137

104. Oka S, Nakabeppu Y. DNA glycosylase encoded by MUTYH functions as a molecular switch for programmed cell death under oxidative stress to suppress tumorigenesis. *Cancer Sci*. 2011;102(4):677-682. doi:10.1111/j.1349-7006.2011.01869.x

105. Giglia-Mari G, Zotter A, Vermeulen W. DNA damage response. *Cold Spring Harb Perspect Biol*. 2011;3(1):a000745. doi:10.1101/cshperspect.a000745

106. Parker A, Gu Y, Mahoney W, Lee SH, Singh KK, Lu AL. Human homolog of the MutY repair protein (hMYH) physically interacts with proteins involved in long patch DNA base excision repair. *J Biol Chem*. 2001;276(8):5547-5555. doi:10.1074/jbc.M008463200

107. Shi G, Chang DY, Cheng CC, Guan X, Venclovas Č, Lu AL. Physical and functional interactions between MutY glycosylase homologue (MYH) and checkpoint proteins Rad9–Rad1–Hus1. *Biochem J*. 2006;400(Pt 1):53-62. doi:10.1042/BJ20060774

108. Luncsford PJ, Manvilla BA, Patterson DN, Malik SS, Jin J, Hwang BJ, Gunther R, Kalvakolanu S, Lipinski LJ, Yuan W, Lu W, Drohat AC, Lu-Chang AL, Toth EA. Coordination of MYH DNA glycosylase and APE1 endonuclease activities via physical interactions. *DNA Repair (Amst)*. 2013;12(12). doi:10.1016/j.dnarep.2013.09.007

109. Brinkmeyer MK, David SS. Distinct functional consequences of MUTYH variants associated with colorectal cancer: Damaged DNA affinity, glycosylase activity and interaction with PCNA and Hus1. *DNA Repair (Amst)*. 2015;34:39-51. doi:10.1016/j.dnarep.2015.08.001

110. Hwang BJ, Jin J, Gao Y, Shi G, Madabushi A, Yan A, Guan X, Zalzman M, Nakajima S, Lan L, Lu AL. SIRT6 protein deacetylase interacts with MYH DNA glycosylase, APE1 endonuclease, and Rad9-Rad1-Hus1 checkpoint clamp. *BMC Mol Biol*. 2015;16:12. doi:10.1186/s12867-015-0041-9

111. Malm M, Saghaleyni R, Lundqvist M, Giudici M, Chotteau V, Field R, Varley PG, Hatton D, Grassi L, Svensson T, Nielsen J, Rockberg J. Evolution from adherent to suspension: systems biology of HEK293 cell line development. *Sci Rep*. 2020;10(1):18996. doi:10.1038/s41598-020-76137-8

112. Wooden SH, Bassett HM, Wood TG, McCullough AK. Identification of critical residues required for the mutation avoidance function of human MutY (hMYH) and implications in colorectal cancer. *Cancer Letters*. 2004;205(1):89-95. doi:10.1016/j.canlet.2003.10.006

113. Bai H, Jones S, Guan X, Wilson TM, Sampson JR, Cheadle JP, Lu AL. Functional characterization of two human MutY homolog (hMYH) missense mutations (R227W and V232F) that lie within the putative hMSH6 binding domain and are associated with hMYH polyposis. *Nucleic Acids Res*. 2005;33(2):597-604. doi:10.1093/nar/gki209

114. Ali M, Kim H, Cleary S, Cupples C, Gallinger S, Bristow R. Characterization of mutant MUTYH proteins associated with familial colorectal cancer. *Gastroenterology*. 2008;135(2):499-507. doi:10.1053/j.gastro.2008.04.035

115. Molatore S, Russo MT, D'Agostino VG, Barone F, Matsumoto Y, Albertini AM, Minoprio A, Degan P, Mazzei F, Bignami M, Ranzani GN. MUTYH mutations associated with familial adenomatous polyposis: functional characterization by a mammalian cell-based assay. *Human Mutation*. 2010;31(2):159-166. doi:10.1002/humu.21158

116. Komine K, Shimodaira H, Takao M, Soeda H, Zhang X, Takahashi M, Ishioka C. Functional Complementation Assay for 47 MUTYH Variants in a MutY-Disrupted Escherichia Coli Strain. *Hum Mutat*. 2015;36(7):704-711. doi:10.1002/humu.22794

117. Raetz AG, Xie Y, Kundu S, Brinkmeyer MK, Chang C, David SS. Cancer-associated variants and a common polymorphism of MUTYH exhibit reduced repair of oxidative DNA damage using a GFP-based assay in mammalian cells. *Carcinogenesis*. 2012;33(11):2301-2309. doi:10.1093/carcin/bgs270

118. HAHM SH, CHUNG JH, AGUSTINA L, HAN SH, YOON IS, PARK JH, KANG LW, PARK JW, NA JJ, HAN YS. Human MutY homolog induces apoptosis in etoposide-treated HEK293 cells. *Oncol Lett*. 2012;4(6):1203-1208. doi:10.3892/ol.2012.921

119. Han SH, Hahm SH, Tran AHV, Chung JH, Hong MK, Paik HD, Kim KS, Han YS. A physical association between the human mutY homolog (hMYH) and DNA topoisomerase II-binding protein 1 (hTopBP1) regulates Chk1-induced cell cycle arrest in HEK293 cells. *Cell & Bioscience*. 2015;5(1):50. doi:10.1186/s13578-015-0042-x

120. Nuñez NN, Khuu C, Babu CS, Bertolani SJ, Rajavel AN, Spear JE, Armas JA, Wright JD, Siegel JB, Lim C, David SS. The Zinc Linchpin Motif in the DNA Repair Glycosylase MUTYH: Identifying the Zn2+ Ligands and Roles in Damage Recognition and Repair. *J Am Chem Soc*. 2018;140(41):13260-13271. doi:10.1021/jacs.8b06923

121. Turco E, Ventura I, Minoprio A, Russo MT, Torreri P, Degan P, Molatore S, Ranzani GN, Bignami M, Mazzei F. Understanding the role of the Q338H MUTYH variant in oxidative damage repair. *Nucleic Acids Res*. 2013;41(7):4093-4103. doi:10.1093/nar/gkt130

122. Banda DM, Nuñez NN, Burnside MA, Bradshaw KM, David SS. Repair of 8-oxoG:A mismatches by the MUTYH glycosylase: Mechanism, metals and medicine. *Free Radic Biol Med*. 2017;107:202-215. doi:10.1016/j.freeradbiomed.2017.01.008

123. Harrison SM, Dolinsky JS, Johnson AEK, Pesaran T, Azzariti DR, Bale S, Chao EC, Das S, Vincent L, Rehm HL. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genetics in Medicine*. 2017;19(10):1096-1104. doi:10.1038/gim.2017.14

124. Rivera-Muñoz EA, Milko LV, Harrison SM, Azzariti DR, Kurtz CL, Lee K, Mester JL, Weaver MA, Currey E, Craigen W, Eng C, Funke B, Hegde M, Hershberger RE, Mao R, Steiner RD, Vincent LM, Martin CL, Plon SE, Ramos E, Rehm HL, Watson M, Berg JS. ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. *Human Mutation*. 2018;39(11):1614-1622. doi:10.1002/humu.23645

125. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, Kanavy DM, Luo X, McNulty SM, Starita LM, Tavtigian SV, Wright MW, Harrison SM, Biesecker LG, Berg JS, Abou Tayoun AN, Berg JS, Biesecker LG, Brenner SE, Cutting GR, Ellard S, Greenblatt MS, Harrison SM, Karbassi I, Karchin R, Mester JL, O'Donnell-Luria A, Pesaran T, Plon SE, Rehm H, Tavtigian SV, Topper S, On behalf of the Clinical Genome Resource Sequence Variant Interpretation Working Group. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Medicine*. 2019;12(1):3. doi:10.1186/s13073-019-0690-2

126. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG, ClinGen Sequence Variant Interpretation Working Group. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum Mutat*. 2018;39(11):1525-1530. doi:10.1002/humu.23642

127. Luncsford PJ, Chang DY, Shi G, Bernstein J, Madabushi A, Patterson DN, Lu AL, Toth EA. A Structural Hinge in Eukaryotic MutY Homologues Mediates Catalytic Activity and Rad9–Rad1–Hus1 Checkpoint Complex Interactions. *Journal of Molecular Biology*. 2010;403(3):351-370. doi:10.1016/j.jmb.2010.08.045

128. Kim DV, Kulishova LM, Torgasheva NA, Melentyev VS, Dianov GL, Medvedev SP, Zakian SM, Zharkov DO. Mild phenotype of knockouts of the major apurinic/apyrimidinic endonuclease APEX1 in a non-cancer human cell line. *PLoS One*. 2021;16(9):e0257473. doi:10.1371/journal.pone.0257473

129. Gupta A, Hwang BJ, Benyamien-Roufaeil D, Jain S, Liu S, Gonzales R, Brown RA, Zalzman M, Lu AL. Mammalian MutY Homolog (MYH or MUTYH) is Critical for Telomere Integrity under Oxidative Stress. *OBM Geriat*. 2022;6(2):196. doi:10.21926/obm.geriatr.2202196

130. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 2016;34(2):184-191. doi:10.1038/nbt.3437

131. LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, Chen E, Gau CL, Palmaer E, Shoaepour K, Shah D, Speare V, Gandomi S, Chao E. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med*. 2014;16(11):830-837. doi:10.1038/gim.2014.40

132. Nielsen M, Morreau H, Vasen HFA, Hes FJ. MUTYH-associated polyposis (MAP). *Crit Rev Oncol Hematol*. 2011;79(1):1-16. doi:10.1016/j.critrevonc.2010.05.011

133. Conlon SG, Khuu C, Trasviña-Arenas CH, Xia T, Hamm ML, Raetz AG, David SS. Cellular Repair of Synthetic Analogs of Oxidative DNA Damage Reveals a Key Structure–Activity Relationship of the Cancer-Associated MUTYH DNA Repair Glycosylase. *ACS Cent Sci*. 2024;10(2):291-301. doi:10.1021/acscentsci.3c00784

134. Yanus GA, Akhapkina TA, Ivantsov AO, Preobrazhenskaya EV, Aleksakhina SN, Bizin IV, Sokolenko AP, Mitiushkina NV, Kuligina ES, Suspitsin EN, Venina AR, Holmatov MM, Zaitseva OA, Yatsuk OS, Pashkov DV, Belyaev AM, Togo AV, Imyanitov EN, Iyevleva AG. Spectrum of APC and MUTYH germ-line mutations in Russian patients with colorectal malignancies. *Clin Genet*. 2018;93(5):1015-1021. doi:10.1111/cge.13228

135. Lu J, Li X, Zhang M, Chen Z, Wang Y, Zeng C, Liu Z, Chen H. Regulation of MUTYH, a DNA Repair Enzyme, in Renal Proximal Tubular Epithelial Cells. *Oxidative Medicine and Cellular Longevity*. 2015;2015:e682861. doi:10.1155/2015/682861

136. Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res*. 2007;35(7):e55. doi:10.1093/nar/gkm106

137. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol*. 2008;8:91. doi:10.1186/1472-6750-8-91

138. Bodai Z, Bishop AL, Gantz VM, Komor AC. Targeting double-strand break indel byproducts with secondary guide RNAs improves Cas9 HDR-mediated genome editing efficiencies. *Nat Commun*. 2022;13(1):2351. doi:10.1038/s41467-022-29989-9

139. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol*. 2019;37(3):224-226. doi:10.1038/s41587-019-0032-3