UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
The Use of Base Editing Technology for Characterization of Single Nucleotide Variants

Permalink
https://escholarship.org/uc/item/9868k81s

Author
McDaniel, Sophia Mika

Publication Date
2021

Peer reviewed|Thesis/dissertation

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

UNIVERSITY OF CALIFORNIA SAN DIEGO

The Use of Base Editing Technology for Characterization of Single Nucleotide Variants

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

In

Biology

By

Sophia M. McDaniel

Committee in charge:

Professor Alon Goren, Chair
Professor Xin Sun, Co-Chair
Professor Lorraine Pillus

2021

The Thesis of Sophia M. McDaniel is approved, and it is acceptable in quality and form for publication on microfilm and electronically

University of California San Diego

2021

TABLE OF CONTENTS

Thesis Approval Page....................iii

Table of Contents....................iv

List of Abbreviations....................v

List of Figures....................vii

List of Tables....................vii

Acknowledgements....................viii

Abstract of the Thesis....................ix

Chapter 1: Introduction....................1

Chapter 2: CRISPR/Cas9 as a gene knockout tool....................5

Chapter 3: CRISPR Derivatives Enable Further Manipulation of the Genome....................9

Chapter 4: Genome-wide pooled CRISPR screens increase the throughput of studying variants....................13

Chapter 5: CRISPR Screens with scRNAseq for Higher Resolution Read-out....................15

Chapter 6: Example Studies....................17

Chapter 7: Future Directions....................20

Bibliography....................22

LIST OF ABBREVIATIONS

ABE	Adenine Base Editor
BE	Base Editor
CBC	Cell barcode
CBE	Cytosine Base Editor
CRISPR	Clustered Regularly Interspersed Short Palindromic Repeat
CRISPRa	CRISPR activation
CRISPRi	CRISPR inhibition
CRISPRko	CRISPR knock-out
DDR	DNA Damage Repair
DSB	Double-stranded break
GBC	Guide barcode
PAM	Protospacer adjacent motif
sc-RNAseq	Single-cell RNA Sequencing
sgRNA	single-guide RNA
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
TTD	Trichothiodystrophy
UGI	Uracil Glycosylase Inhibitor

UMI	Unique Molecular Identifier
VEP	Variant Effect Predictor
VUS	Variant of Uncertain Significance
XP	Xeroderma Pigmentosum
XP/ CS	Xeroderma Pigmentosum/ Cockayne's Syndrome

LIST OF FIGURES

Figure 1: ClinVar Distribution of SNV Effects in Humans…………………………………………...…3

Figure 2: TFIIH complex structure……………………………………………………………….…...4

Figure 3: Schematic of the structural makeup and mechanism of cytosine base editors (CBEs) and adenine base editors (ABEs)……………………………………………………………………………..……11

Figure 4: Schematic of single cell RNA sequencing (scRNAseq) workflow……………………….…….15

LIST OF TABLES

Table 1: Phenotypic, genotypic, and clinical comparison of ERCC2 pathogenic variants……………….4

Table 2: Summary of example studies………………………………………………………………..19

ACKNOWLEDGMENTS

I would like to acknowledge Dr. Alon Goren for his mentorship throughout the course of this challenging year. His unwavering support and commitment to his lab even in the face of the COVID-19 pandemic was invaluable.

All parts of this paper are currently being prepared for submission for publication of the material. McDaniel, Sophia. The thesis author was the primary investigator and author of this material.

ABSTRACT OF THE THESIS

The Use of Base Editing Technology for Characterization of Single Nucleotide Variants

by

Sophia M. McDaniel

Master of Science in Biology

University of California San Diego, 2021

Professor Alon Goren, Chair
Professor Xin Sun, Co-Chair

Single nucleotide variants represent the most common type of polymorphism in the human genome. However, the phenotypic impacts of these mutations are not well understood in many cases. Intriguingly, while some SNVs cause debilitating diseases, other variants in the same protein may have no or limited effect. The mechanisms underlying these complex interactions are difficult to study at a high throughput scale. In this review we discuss base editing technologies and their potential to accelerate progress in this field, particularly in combination with single-cell RNA sequencing. Using the XPD protein as an example, we explore how base editing screens can help link SNVs to distinct disease

phenotypes. We then highlight several studies that take advantage of single-cell RNA sequencing and CRISPR screens to emphasize the current limitations and future potential of this technique.

Chapter 1: Introduction

Single nucleotide variants (SNVs) are the most common type of polymorphism in the human genome. Recent studies suggest that there are approximately 4 million in the average individual and 8.6 billion distinct SNVs possible[1]. Of these, only 720,000 have been identified and studied according to the Clinvar database, of which approximately 12% are pathogenic or likely pathogenic, 38% are benign or likely benign, and 45% are variants of uncertain significance (VUS)[2] (*Figure 1*). A point mutation can have varying degrees of impact depending on the location ranging from synonymous mutations, which sometimes have virtually no effect, to nonsense mutations, which can functionally knock out a gene. Point mutations in the active sites or binding domains of enzymes can be particularly damaging and cause a plethora of downstream effects that may manifest as a genetic disease. For example, SNVs in the DNA damage response (DDR) pathways have been shown to cause a wide variety of pathogenic diseases such as many forms of cancer, neurodegenerative diseases, immune disorders, metabolic dysfunction, and more[3].

Even SNVs in various locations within a single gene can cause vastly different phenotypic effects. This can be exemplified by the ERCC2 gene, encoding XPD which is a component of the transcription factor IIH (TFIIH) complex. TFIIH is composed of 10 protein subunits and is highly conserved across eukaryotes, with roles in various biological functions such as transcription, cell cycle regulation, and DNA repair. When DNA lesions are detected the XPD subunit of TFIIH acts as a helicase to catalyze the unwinding of the mutated DNA site, making it accessible to DNA repair enzymes. Structurally, XPD is in direct contact with p44 which anchors it to the rest of the protein. Both components are necessary to maintain the structural integrity of TFIIH[4] (*Figure 2*). There are over 8,000 identified SNVs in the ERCC2 gene alone, most of which have not been characterized according to the NCBI Variation Viewer database (https://www.ncbi.nlm.nih.gov/variation/view/). SNVs in ERCC2 were

shown to cause three distinct diseases depending on the precise location of the edit— xeroderma pigmentosum (XP), trichothiodystrophy (TTD), and xeroderma pigmentosum/ Cockayne's syndrome (XP/CS)[5] (*Table 1*).

XP is characterized by sensitivity to UV light, resulting in extreme pigmentation of the skin and an up to 10,000-fold increase in the likelihood of developing skin cancer before the age of 20. Individuals afflicted with this disease often suffer from neurological disorders as well, presumed to be caused by apoptosis of neurons after the buildup of gene damage[6]. Mutations that cause this disease are commonly located in areas of the protein that impact helicase activity but not TFIIH structure, namely the catalytic site or DNA- and ATP-binding domains[5]. Our lab has been studying the preliminary involvement of the Y542C mutation in XPD. While Y542C is known to cause XP, it has not yet been fully characterized. XP/CS presents many similar symptoms to XP, but includes the addition of more severe neurological disorders that are the result of an impaired TFIIH transcriptional initiation activity[7]. These SNVs have been mapped to areas of XPD that are associated with binding the p44 protein, ATP, or DNA, which ultimately destabilizes the TFIIH structure and reduces both the ability to recover from DNA damage and to initiate transcription[5]. Finally, TTD is distinct from XP and XP/CS, and it results in brittle hair and nails in conjunction with some levels of photosensitivity and neurological disorders. Interestingly, TTD is not known to cause an increased likelihood to develop skin cancer despite being caused by mutations in the ERCC2 gene. Moreover, the neurological disorders associated with TTD are not caused by diminished DNA repair function, but rather by developmental defects, indicating a dysfunction in the transcriptional initiation activity of the TFIIH complex rather than the DNA repair activity[8]. This is supported by X-ray crystallography data which shows that TTD-causing SNVs in ERCC2 cluster in areas that destabilize XPD structure while conserving its helicase activity, consequentially compromising TFIIH structure as well and leading to decreased transcription[5]. Though progress has been made in identifying environmental factors that affect the severity of these diseases, these can only explain a portion of the variation seen in the disease phenotypes of XP, XP/CS, and TTD. Thus, analysis of the transcriptomic effects of each

mutation could potentially provide improved understanding of such diseases and may bring us a step closer to treatment with drugs or genetic therapies.

Previous studies have linked disease phenotypes to SNVs through a combination of sequencing, analysis of crystal structures, and induction of SNVs into model organisms. However, these techniques are highly labor intensive, low throughput, and are limited in their ability to determine the biochemical mechanism underlying disease phenotypes. Recently, CRISPR-based technologies including CRISPR knock-out (CRISPRko), CRISPR inhibition and activation (CRISPRi/a), and prime editing have provided a transformative ability to mutate or regulate genes with higher fidelity and flexibility than ever before. Particularly, base editing has exciting potential to study SNVs with minimal disruption to the natural state of a mammalian cell[9]. In this review, we discuss the use of base editing (BE) screens in combination with single cell RNA sequencing (scRNA-seq) to overcome some of the drawbacks of other approaches and understand how SNVs cause disease.

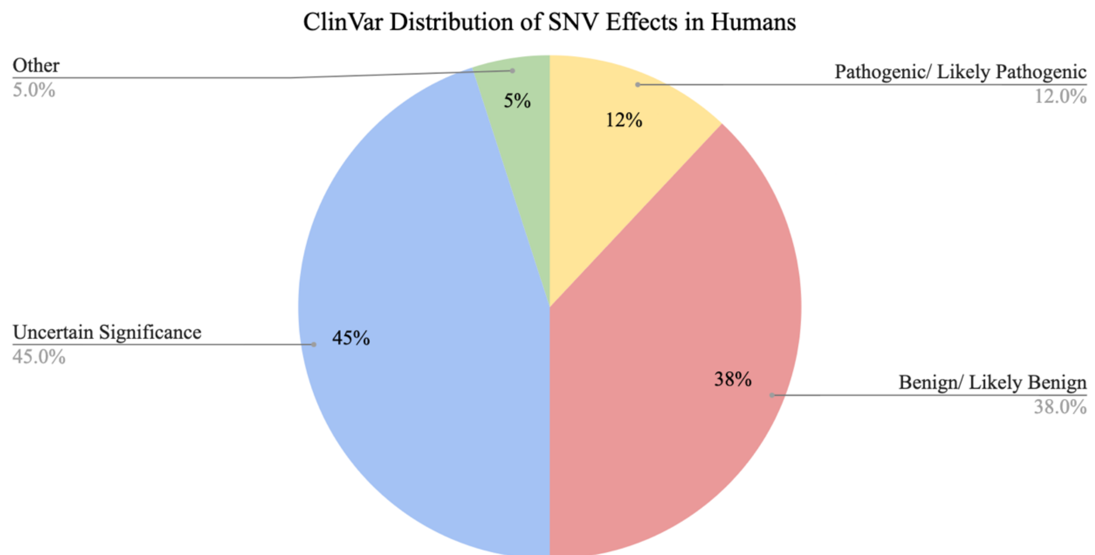

Figure 1: ClinVar Distribution of SNV Effects in Humans. Charts the approximate distribution of SNVs according to their phenotypic effect in humans, as listed in the ClinVar database[2].

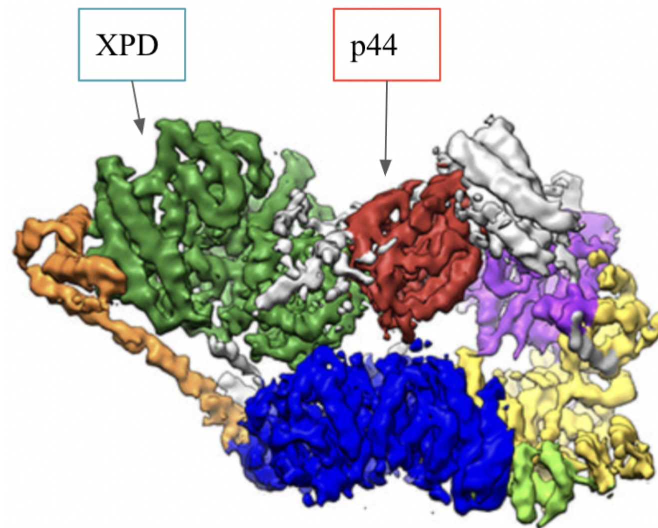

Figure 2: TFIIH complex structure. The XPD protein subunit is necessary for catalytic helicase activity during nucleotide excision repair, and to maintain the structural integrity of the TFIIH complex. It interacts directly with p44 which attaches XPD to the rest of the protein complex[4]. Adapted from "The cyro-electron microscopy structure of human transcription factor IIH," by Author B. J. Greber *et. al,* 2017, *Nature, 549*, 414. Copyright [2017] by Macmillan Publishers Limited. Adapted with Permission.

Table 1: Phenotypic, genotypic, and clinical comparisons of ERCC2 pathogenic variants[5-8].

Phenotypic, genotypic, and clinical comparison of ERCC2 pathogenic variants			
Disease	**XP**	**XP/CS**	**TTD**
Clinical phenotype	Severe photosensitivity, skin cancer, neurological disorders	Photosensitivity, developmental disorders, neurological disorders	Mild photosensitivity, developmental disorders, neurological disorders, brittle hair and nails
Molecular mechanism of pathogenesis	Impairs helicase function while maintaining XPD structure.	Affects ATP/ssDNA binding sites or p44 binding, impacting TFIIH structure.	Affects XPD structure and destabilize TFIIH.
Incidence (Western Europe)	2.3 per million	0.26 per million	1.2 per million

Chapter 2: CRISPR/Cas9 as a gene knockout tool

Sequence-specific gene knockouts in eukaryotes have in the past been achieved with zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) which are composed of naturally occurring DNA binding domains linked to endonucleases. These technologies are limited by strict DNA sequence requirements, low efficiency, and large size of the required components. In contrast, the CRISPR/Cas9 system is simple and easy to implement, allowing it to become a widespread gene editing method and encouraging high-throughput approaches to gene interrogation[10].

The CRISPR/Cas9 system originates from the adaptive immune system of bacteria and archaea. It was modified for use as a gene knock-out tool by replacing the two components of the guide RNA— the chromosomal RNA and trans-chromosomal RNA— with a single-guide RNA (sgRNA) which can be designed to target a DNA sequence of interest. The sgRNA consists of 3 components: a 20-nucleotide complementary region which binds target DNA, a 42-nucleotide Cas9 handle which binds to the Cas9 protein, and a 40-nucleotide transcriptional termination sequence. The sgRNA, once transcribed in a target cell, directs the Cas9 protein and binds its complementary DNA strand where the Cas9 can induce a double-stranded break (DSB) upon recognition of a protospacer-adjacent motif (PAM). The protospacer, or 20bp sequence bound by the sgRNA, is commonly numbered from 1-20 for reference purposes with the first base being the furthest upstream. Cas proteins consists of two nucleolytic domains, HNH and RuvC, which cleave the complementary and non-complementary strands, respectively. PAM is a 2-6bp motif that is located just downstream of the target DNA and is a requirement for the Cas protein to activate. The PAM sequence depends on the bacterial species from which the Cas is taken and simultaneously allows for higher specificity of the targeted gene sequence while restricting possible cut sites[10, 11].

The DSB induced by Cas cleavage spontaneously recruits the target cell's DNA repair mechanisms and results in homologous repair or, more often, non-homologous end joining which is error-prone and commonly causes random insertions or deletions (indels). These indels can result in frameshift mutations and early stop codons in the gene, achieving the desired knockout. For simplicity, this approach of using CRISPR for inducing gene knockouts will hereafter be referred to as CRISPRko. An exogenous template DNA strand may be introduced in the cleaved cell, which, during mitosis, can be used as a template to introduce specific mutations of interest via homologous repair[10, 11]. However, this method has limitations due to its extremely low efficiency as well as its tendency to cause unintended indels[12] and off-target effects[13].

Limitations and Improvements of CRISPR

The introduction of CRISPR was a powerful step forward in the field of mammalian gene editing, yet there are many drawbacks to this technology which have been incrementally improved and worked around within the last decade.

A major limitation to the utility of CRISPR is that low edit efficiency makes it difficult to study recessive genes whose effects only become apparent with a double knock-out. The efficiency of inducing a single edit is low enough as is, so the probability of achieving a double knock-out is even lower. Some research groups have overcome this issue by exclusively studying genes with a dominant effect or using near-haploid cell lines such as KBM7[14] and HAP1[15]. However, these cell lines are not representative of the multitude of cell types present within the body and have chromosome structures that are vastly different from human cells, which calls into question the viability of translating these findings directly to clinical applications.

One reason for limited efficiency in CRISPR-mediated gene editing is that DSBs tend to cause edited cells to enter apoptosis through the p53-mediated DNA damage response. Death of edited cells can

be reduced by inhibiting the p53 gene[16], but this in turn limits the cell's ability to repair other DNA damage, thereby making it difficult to decipher whether observed phenotypes are due to the edit or other uncontrolled factors[17]. p53-mediated apoptosis can be substantially reduced by using CRISPR derivative techniques, discussed below, which edit the genome without causing DSBs. Optimizing sgRNA design has been a major focus for improving binding specificity and efficiency of cleavage. Some studies have identified factors that influence edit efficiency such as targeting the coding region proximal to the N-terminus, the presence of a G as the last spacer nucleotide before the PAM sequence, and local heterochromatin structure[18, 19]. These factors and others have been taken into consideration in the development of sgRNA designing algorithms that can predict the relative efficiency of cleavage and presence of off-target locations[20–24]. The presence of nucleosomes can also impede the Cas protein's ability to bind to certain gene sequences[25,26], thus some groups have taken nucleosome location into account when designing sgRNA libraries to maximize edit efficiency[27]. These sgRNA algorithms also take the presence of a PAM sequence into consideration. PAM sequences severely limit the possible edit locations in a cell, but many groups are working to use various Cas proteins such as Cas12 which have alternative PAM requirements. Others have mutated pre-existing Cas proteins to make the PAM requirements less restrictive[28], though the merit of expanded edit locations must be weighed against the fact that overly lax PAM requirements increase the likelihood of non-specific binding and cleavage. The various techniques to overcome the limitations of CRISPRko must be used selectively on a case-by-case basis to maximize efficiency and possible edit sites while minimizing byproducts.

Some drawbacks of the CRISPRko system stem from the very nature of the technology itself. Indels induced by DSBs are unpredictable and can cause mutations in locations outside of the target gene[12, 29], and even successful edits in the gene of interest may only truncate the protein without knocking it out. Alternatively, the function of some genes can only be studied by modulating gene expression rather than knocking it out completely, such as with cell-essential genes or genes that show different phenotypic

effects in a manner that depends on their expression levels. CRISPR-derived technologies mitigated many of these CRISPRko limitations and expanded the applicability of the CRISPR toolbox.

Chapter 3: CRISPR Derivatives Enable Further Manipulation of the Genome

CRISPR technology has been modified for a multitude of purposes by inactivating the nucleolytic property of Cas9 and attaching new catalytic groups to allow studying different facets of gene function or to improve CRISPR performance. CRISPRi[30] and CRISPRa[31,32] replace RNA interference[33] and expression of DNA vectors[34], respectively, and enables detecting the effects of modulated gene function without the need to genetically remove that gene. Prime editing improved the predictability and range of possible edits[35], while BE optimized induction of SNVs. CRISPRi, CRISPRa, and prime editing are reviewed here[34, 36]. Here, we will mainly focus on base editing as a method to interrogate SNVs.

Base Editing Improves the Induction of SNVs

The range of possible edits inducible with CRISPR technologies greatly expanded with the implementation of BE which uses the Cas/sgRNA targeting abilities and the tethering of a deaminase protein to specifically modify nucleotides. The original base editor was a cytidine base editor (CBE) which takes advantage of the rat APOBEC1 protein, a naturally occurring component of the pyrimidine salvage pathway. In the most commonly used CBEs, APOBEC1 is tethered to the N- terminus of a partially inactivated Cas9[9]. A uracil DNA glycosylase inhibitor (UGI) adapted from *Bacillus subtilis* is linked to the C-terminus to prevent DNA repair pathways from initiating base excision repair[37] (*Figure 3)*. The Cas9 protein here has an inactive RuvC domain but retains nucleolytic activity in the HNH domain, causing only the non-edited strand to be cleaved. APOBEC1 converts cytidine residues to uridine residues, then the cleavage by the HNH domain stimulates DNA repair pathways which preferentially modifies the cleaved strand to match the strand with the new base[9].

The other transition mutation, A–> G, is achieved with TadA, an RNA adenine deaminase that was taken from *E. Coli* and was artificially evolved to bind and cleave DNA with high efficiency.

Adenine is deaminated to generate inosine, which is recognized as a guanine by DNA binding proteins, functionally leading to an A–>G edit after completion of DNA repair[38]. Both CBEs and adenine base editors (ABEs) can only act on single stranded DNA, restricting their function to the target DNA strand and reducing undesired off-target editing. Nucleotides are edited within a small editing window of ~5 nucleotides located between bases 4-8 in the protospacer where the bound Cas9 protein separates the double stranded DNA. If multiple bases (cytosines for CBEs and adenines for ABEs) are located within the edit window, they all have the potential to be edited albeit with varying edit efficiencies[9, 38].

Base editing has been implemented to cause gene knockouts while minimizing the impact to the target cell by inducing stop codons[39,40] or disrupting splice sites[41]. SNVs induced to cause stop codons achieve gene knockouts with high predictability and high efficiency compared to non-homologous end joining or homology-directed repair mediated knock-outs[39], making BEs desirable as a substitute for CRISPRko.

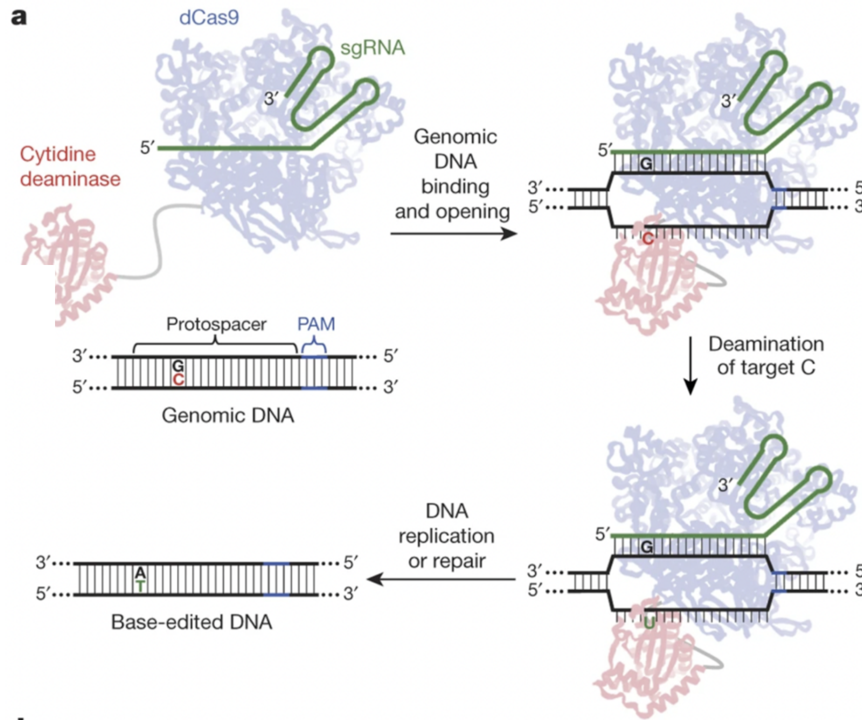

Figure 3: Schematic of the structural makeup and mechanism of cytosine base editors (CBEs). CBEs are composed of a catalytically dead Cas9 protein (dCas9) with a cytidine deaminase fused to the N-terminus and an uracil glycosylase inhibitor fused to the C-terminus. After the sgRNA binds to the target DNA, the cytosine deaminase can convert any cytosine residues within the edit window to a uridine residue[9]. Adapted from "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage," by A. C. Komor *et. al*, 2016, *Nature, 533(7603),* 420. Copyright [2016] by Macmillan Publishers Limited. Adapted with Permission.

Modified Base editors

Despite their differences in mechanism, BEs retain many of the same limitations that are seen in CRISPRko. These include low editing efficiency in a manner that depends on the location in the gene, off target DNA editing[42, 43], and limited editable loci. In addition, the APOBEC1 enzyme used in the original BE was shown to have not only off-target DNA editing but also spontaneous RNA deamination in substantial numbers of both protein-coding and non-coding RNA[44]. Results from BE experiments are further complicated by bystander edits within the edit window and the inability to validate results in the same way as CRISPRko. CRISPRko results are commonly validated by targeting multiple sgRNAs to the same gene. Having multiple sgRNAs that yield the same result ensures that outcomes are due to desired

edits rather than other factors. In contrast, BEs induce SNPs in single bases, so there is often only 1 sgRNA, if any, that can target the desired site on the gene[45].

Multiple approaches have been taken to alleviate these hurdles. For example, mutation of the APOBEC gene has resulted in improved proteins with reduced off-target RNA editing and high specificity for certain bases within an edit window[46]. Other engineered APOBEC enzymes have reduced the induction of bystander edits by making the edit window smaller (from ~5 nucleotides to ~2 nucleotides)[47, 48] or biasing the specificity towards a single cytosine within the window[46]. The addition of a nuclear localization signal to the base also improved editing efficiency by increasing expression of BEs in transfected cells[49]. SgRNA prediction tools specifically for BEs have been developed by testing a large library of sgRNAs and using deep learning approaches to understand which factors contribute to a more successful BE[50–52]. New BEs with relaxed PAM requirements have been developed to expand the possible edit locations in the genome[46, 49, 53].

An impressive amount of research has been conducted to improve BEs since the technology first emerged in 2016. Although these are valuable contributions to the gene editing field, there are drawbacks to moving forward at such a fast pace; individual methods cannot be intensively implemented or tested before the next generation emerges. Published studies often utilize outdated versions of base editors as new implementations are rapidly developed, and it is challenging to decide which base editors to use for new research because the reliability of each version has not been rigorously tested over a long period of time. This disadvantage of "bleeding edge" technology is important when considering the significance and impact of BE studies.

Chapter 4: Genome-wide pooled CRISPR screens increase the throughput of studying variants

The power and efficiency of studying gene function and interactions increased dramatically with the introduction of CRISPR screens. A successful screen requires two components: a small nucleotide that can be easily transduced using a lentiviral or retroviral vector, and the ability to select for and identify edited cells in a population. CRISPR screens are conducted by producing a library of sgRNAs (which are small compared to previously popular platforms) along with genetic "barcodes" (used to link cells to the sgRNA they received downstream) and transducing them into target cells with a lentiviral vector. The sgRNA library is first inserted into a plasmid that contains markers for downstream selection, then transferred into a cell line that are easily for transfected (e.g. HEK 293FT) along with other plasmids containing structural components of the virus. After virus production, the viruses are isolated and used to transduce the target cells at a low multiplicity of infection (MOI) to optimize the likelihood that each target cell will only get one sgRNA inserted. Analysis occurs after challenging the cell population to identify genes which cause a selective advantage or disadvantage upon perturbation[54].

Positive and Negative selection screens

CRISPR screens are carried out by perturbing a population of cells with a library of sgRNAs, then using next generation sequencing (NGS) to measure the relative abundance of cells by profiling the sgRNAs after allowing time for cell growth and competition. Positive selection screens measure the most abundant sgRNAs after challenging cells with a drug or toxin, allowing identification of genes conferring resistance to the given substance. Negative selective screens, as the name suggests, compare a non-perturbed cell population to a perturbed cell population to identify genes that, when knocked out, confer a growth disadvantage, causing them to be selected out from the population. Early implementations at a genome-wide level screened for resistance to 6-thioguanine[55], Vemurafenib[14], *Clostridium septicum*

alpha-toxin[56] anthrax and diphtheria toxins[57]. These studies identified genes essential to the DNA repair pathway and genes whose loss led to resistance to a cancer drug and provided a better understanding of pathways that result in cell death by microbial toxins, respectively. These early proof-of-concept efforts obtained expected results for previously well-studied genes and served to establish genome-wide libraries for future work. Subsequent studies used optimized sgRNAs to improve the composition of libraries[58] and established libraries for CRISPRi/a[27,58,59] and BE screens. However, these low-resolution screens which only test for a crude phenotype are limited in their capacity for several reasons. For one, the selection phenotype must confer a growth advantage or disadvantage, limiting the possible phenotypes that can be screened for. There is a limited number of phenotypes that can be tested at once, and the underlying mechanism causing each phenotype cannot be elucidated without further study. Moreover, specific phenotypes caused by cell cycle effects or cell subpopulations may be masked because of the low-resolution readout[54].

Chapter 5: CRISPR screens with scRNAseq for higher resolution read-out

CRISPR screens allowed the unprecedented scale-up of studying genetic perturbations, however positive and negative selection screens are limited because of their low-resolution read-out. Understanding the transcriptomic readout of a perturbed cell provides valuable information that goes beyond a simple phenotype (survival, proliferation, or drug resistance), and can uncover varying pathways that result in specific phenotypes. This has been previously achieved by selecting single cells from a perturbed cell population and performing RNA seq as an array; however, this method has limited scalability. Single-cell RNA sequencing (scRNAseq) allows this process to be scaled up by multiple magnitudes. The current most common approaches include isolation of single cells into a nanoliter droplet along with a unique bead. The bead's surface is coated with oligonucleotides with 4 components: a constant region, a cell barcode (CBC), a unique molecular identifier (UMI), and a poly T region. The cells are lysed inside of the droplets so that the mRNA can be captured, reverse transcribed, and PCR amplified. The CBCs are used to trace mRNA transcripts to the cell from which they originated, and the UMIs can correct for amplification bias[60].

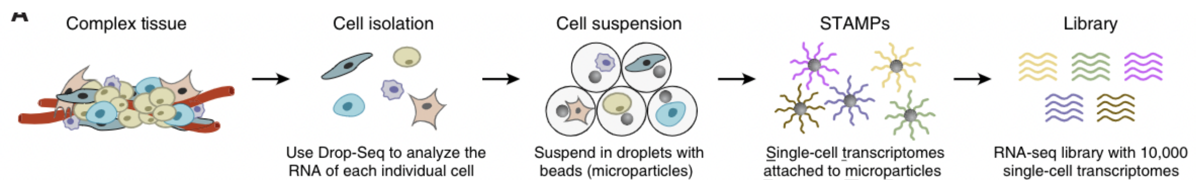

Figure 4: Schematic of single cell RNA sequencing (scRNAseq) workflow. Cells are separated and isolated into droplets, each containing a unique bead, via a microfluidics device that creates a water in oil suspension. The cells are lysed and mRNA is captured by oligonucleotides on the beads. The emulsion is broken and the oligonucleotides are reverse transcribed into a library which is then sequenced and analyzed[60]. Adapted from "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets," by Author E. Z. Macosko *et. al*, 2015, *Cell, 161(5)*, 1202. Copyright [2015] by Elsevier Inc. Adapted with Permission.

scRNAseq for CRISPR screens

CRISPR screens were modified to be compatible with scRNAseq by including a poly A tail at the end of the sgRNA transcript so that it can be captured by the beads and traced to corresponding transcriptomes[61]. Alternatively, unique polyadenylated guide barcodes (GBCs) can be included in the sgRNA plasmids and be similarly captured on the bead oligos to determine the associated perturbation[62–64]. These methods were found to be problematic because barcodes could be uncoupled from the sgRNAs due to lentiviral recombination, as often as 50% of the time depending on the distance of the barcode from the sgRNA[65,66]. The first method only captured guides in 40-60% of the cell, preventing false results due to barcode swapping but still losing substantial amounts of transcriptomic data[61]. Targeted amplification of the guide RNA[66] improved the efficiency, but the large cassette size of this technique restricts the use for delivery of multiple cassettes to one target cell. Currently, the most used platform of scRNAseq for CRISPR-based screens is the 10x Genomics system[67] which operates using direct-capture Perturb-seq[68]. Other platforms and their comparative performance are reviewed here[69].

Chapter 6: Example Studies

To date, little research has been conducted using scRNAseq in combination with BEs, hereafter referred to as scBE screens. Here, we focus on one scBE screen as well as other CRISPR-based screens to discuss limitations and emphasize strategies that can be incorporated into scBE screens as they become more commonly used in the gene editing space.

One of the first scBE screens, conducted by Jun *et. al*, took advantage of thid-generation cytosine base editor BE3[9] to screen all possible sRNAs across the MAP2K1, NRAS, and KRAS genes for conferral of Vemurafenib resistance, a selection agent that had been well studied previously in CRISPRko screens. Notably, the use of transcriptomic sequencing revealed immune response gene upregulation in response to the E203K mutation in MAP2K1, which agreed to findings from previous Vemurafenib positive selection screens. The team was also able to identify cell subpopulations that would be masked if the cells were pooled rather than being sequenced at a single cell level. This study established scBE screens as a viable and cost-effective method for robust SNV interrogation. Though this study successfully validated the success of identifying clinically relevant SNVs, the authors cite low efficiency of BE3 (5-20%) and comparatively low throughput of the implemented scRNAseq method (CROP-seq) as areas for improvement. Substituting these methods with BE4(max) and 10x Genomics Chromium respectively could improve the readout for future studies[70].

While scBE screens are still in preliminary stages of use, phenotypic selection-based BE screens provide valuable insights that could be applied when designing and implementing scBE screens. One such study by contributed to the scBE screen space by validating the use of BE screens to discover clinically relevant SNVs causing gain-of-function or loss-of-function phenotypes. In a recent work, Hanna *et. al* screened a library of 68,500 sgRNAs targeting 3,584 genes and conducted both positive and negative selection with Cisplatin, a common drug in cancer treatment, and Hygromycin, a protein synthesis

inhibitor. To validate their findings, they compared the scRNA-seq data generated to expected phenotypic outcomes based on the Ensembl Variant Effect Predictor (VEP). VEP is an open-source online tool which predicts the effect severity of an SNV based on the mutation type (synonymous, nonsense, etc.). SgRNAs targeting essential genes that had a high impact according to VEP were significantly depleted in the negative selection screens, as expected. Similarly, sgRNAs targeting the tp53 gene were significantly enriched in the positive selection screens. In this study, sgRNAs that were significantly enriched in BE screens but not in CRISPRko screens mapped well to known pathogenic variants in the ClinVar database, establishing the utility of this method to identify significant mutations for further study. The authors recommend this method as a new tool to identify drug targets or identify alleles conferring drug resistance, cumulatively better understanding the mechanism of drugs[45].

Another phenotypic selection-based BE screen conducted by Kweon *et. al* applies BE3 to HAP1 cells with a library of every sgRNAs targeting all BRCA1 exons in the human genome. Cells challenged with the PARP inhibitor Olaparib were clustered to identify deficient sgRNAs, resulting in 13 hits that corresponded to known pathogenic mutations according to the ClinVar database as well as multiple other VUSs. The VUSs were shown to be pathogenic as well after downstream analysis. This study was an important proof-of-concept study that established BE screens as a method to study DDR genes, but it was conducted on a relatively small scale (745 gRNAs)[15]. Another highly relevant study by Cuella-Martin *et. al* utilizes a BE screen to assess DDR used BE3 to target sgRNAs with the NGG PAM sequence across 86 DDR associated genes. Perturbed cells were challenged with four DNA damaging agents previously used in CRISPRko screens (Cisplatin, Olaparib, Doxorubicin, and Camptothecin) and enrichment and depletion phenotypes. Importantly, this work correctly differentiated between known pathogenic and benign SNVs from the ClinVar database and accurately predicted the clinical relevance of SUVs in DDR genes (compiled into the following database: https://www.ciccialab-database.com/ddr-variants/#/)[71].

The final study that we will discuss here is a CRISPR knock-out screen in RPE1 cells, conducted by Olivieri *et. al*. Although this does not incorporate scRNAseq, the authors conduct a series of

CRISPR/Cas9 screens against 25 DNA damaging agents, representing an array of DNA damage types. By mining the resulting data set to find hits across screens, genes were linked to specific DNA repair pathways based on their depletion. Genes were clustered by their impact across all screens when perturbed, indicating the function of many uncharacterized DNA damage repair genes by their clustering with well-studied genes. Nearly 5% of protein coding genes in the human genome caused reduced viability in response to genotoxic agents when knocked out, reiterating the essentiality of gene repair pathways. This method of conducting a multitude of screens in parallel to map DNA damage repair is applicable to scBE screens as well and could provide valuable insight into a pathway highly relevant to a multitude of clinically relevant SNPs[16].

Table 2: Summary of example studies[15,16,45,70,71].

Summary of Example Studies						
Study	**Investigated Genes**	**Selection type**	**Selection Agent**	**Editor Type**	**scRNAseq method**	**Cell line**
Jun *et. al*	MAP2K1, KRAS, NRAS	Positive	Vemurafenib	BE3	CROP-seq	A375
Hanna *et. al*	Many	Positive/ Negative	Cisplatin, Hygromycin	BE3, BE4	N/A	HT29, MELJUSO
Kweon *et. al*	BRCA1	Negative	Olaparib	BE3	N/A	HAP1
Cuella-martin *et. al*	DDR Pathway	Positive/ Negative	Cisplatin, Olaparib, Doxorubicin, Camptothecin	BE3	N/A	MFC10A, MFC7, HAP1
Olivieri *et. al*	DDR Pathway	Positive/ Negative	DNA damaging agents	Cas9	N/A	RPE1

Chapter 7: Future Directions

The DDR pathways are an ideal target to study using sc-BE screens due to the complex interactions between DDR proteins and the relevance to diseases. As mentioned previously, our lab has preliminary indications that the Y542C mutation in ERCC2 is deleterious, but conclusive results are challenging to obtain due to the low survival rate of mutated cells and the work intensive process of producing single cell clones. sc-BE screens will accelerate this process significantly by testing many mutations in parallel and getting a transcriptomic readout, which can provide insight into the mechanisms that cause TTD, XP, and XP/CS. Conflicting interpretations of these SNVs may be attributed to differences in varying cell type, cell cycle stages, cell subpopulations, or cell-cell interactions that are not detectable when analyzing cells in a population. Transcriptomic data at a single cell level will give further insight into where these differences stem from and can help us understand how these SNVs impact humans. As screening technologies improve, it will become more reasonable to read out not only the transcriptome, but protein expression, chromatin accessibility and more at a single cell level[72].

Conclusion

To conclude, scBE screens are an increasingly expanding toolset that have untapped potential for inducing knockouts with minimal perturbation, elucidating the mechanism-of-action of pharmaceuticals, and most importantly understanding the phenotypic effect of clinically relevant genetic variants. We expect that BEs will become increasingly more efficient and widely applicable with optimized sgRNAs, modified Cas enzymes that enable flexibility in PAM sequences, narrowed editing windows and improved computational platforms that can predict base editing outcomes. The volume of information that we can gather from these perturbations is also increasing as single cell technology improves, allowing higher readout and larger scale screens. As we move forward, it is important to conduct scBE screens

with cells that are as relevant as possible to normal cells in the human body, such as hESCs, iPSCs, and cells differentiated from them to improve the applicability of downstream results to clinical applications. sc-BE screens will open pathways to investigate genes with complex interactions such as chromatin regulators and transcription factors, which have previously been restrictively difficult to study.

Acknowledgements

All parts of this paper are currently being prepared for submission for publication of the material. McDaniel, Sophia. The thesis author was the primary investigator and author of this material.

Bibliography:

1. Wu, J., Wu, M., Li, L., Liu, Z., Zeng, W., Jiang, R. DbWGFP: a database and web server of human whole-genome single nucleotide variants and their functional predictions. *Database* **2016,** (2016).

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

3. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* vol. 461 1071–1078 (2009).

4. Rimel, J. K. & Taatjes, D. J. The essential and multifunctional TFIIH complex. *Protein Science* **27,** 1018–1037 (2018).

5. Liu, H. Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., McRobbie, A., Brown, S. E., Naismith, J. H., White, M. F. Structure of the DNA Repair Helicase XPD. *Cell* **133**, 801–812 (2008).

6. Lehmann, A. R., McGibbon, D. & Stefanini, M. Xeroderma pigmentosum. *Orphanet Journal of Rare Diseases* **6,** (2011).

7. Natale, V. A comprehensive description of the severity groups in Cockayne syndrome. *American Journal of Medical Genetics, Part A* **155,** 1081–1095 (2011).

8. Hashimoto, S. & Egly, J. M. Trichothiodystrophy view from the molecular basis of DNA repair/transcription factor TFIIH. *Human Molecular Genetics* **18,** (2009).

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

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

11. Feng Zhang, F., Wen, Y., Guo, X., CRISPR/Cas9 for genome editing: progress, implications and challenges. *Human Molecular Genetics* **23**, R40–R46 (2014).

12. Lin, Y., Cradick, T. J., Brown, M. T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B. m., Vertino, P. M., Stewart, F. J., Bao, G. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research* **42**, 7473–7485 (2014).

13. Fu, Y. Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., Sander, J. D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology* **31**, 822–826 (2013).

14. Wang, C., Wei, J. J., Sabatini, D. M. & Lander, E. S. ". Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343**, 80–84 (2013).

15. Kweon, J. Jang, A., Shin, H. R., See, J., Lee, W., Lee, J. W., Chang, S., Kim, K., Kim, Y. A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. *Oncogene* **39**, 30–35 (2020).

16. Olivieri, M. Cho, T., Alvarez-Quilon, A., Li, K., Schellenberg, M. J., Zimmermann, M., Hustedt, N., Rossi, S. E., Adam, S., Melo, H., Heijink, A. M., Sastre-Moreno, G., Moatti, N., Szilard, R. K., McEwan, A., Ling, A. K., Serrano-Benitez, A., Ubhi, J., Feng, S., Pawling, J., Delgado-Sainz, I., Ferguson, M. W., Dennis, James, W., Brown, G. W., Cortes-Ledesma, F., Williams, R. S., Martin, A., Xu, D., Durocher, D. A genetic map of the response to DNA damage in human cells. *bioRxiv* 845446 (2019) doi:10.1101/845446.

17. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine* **24**, 927–930 (2018).

18. Doench, J. G., Hartenian, E., Graham, D. B., Tothova, M. H., Smith, I., Sullender, M., Elbert, B. J., Xavier, R. J., Root, D. E. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology* **32**, 1262–1267 (2014).

19. Doench, J. G., Fusi, Nicolo, Sullender, M., Hegde, M., Vaimberg, . W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., Root, D. E. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* **34**, 184–191 (2016).

20. Labun, K., Montague, T. G., Krause, M., Cleuren, Y. N. T., Tjeldnes, H., Valen, E. CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research* **47**, W171–W174 (2019).

21. Chow, R. D., Chen, J. S., Shen, J. & Chen, S. A web tool for the design of prime-editing guide RNAs. *Nature Biomedical Engineering* **5**, 190–194 (2021).

22. Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J., Fernandez, J. P., Mis, E. K., Khokha, M. K., Giraldez, A. J. CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature Methods* **12**, 982–988 (2015).

23. Meier, J. A., Zhang, F. & Sanjana, N. E. GUIDES: SgRNA design for loss-of-function screens. *Nature Methods* vol. 14 831–832 (2017).

24. Listgarten, J., Weinstein, M., Kleinstiver, B. P., Sousa, A. A., Joung, J. K., Crawford, J., Gao, K., Hoang, L., Elibol, M., Doench, J. G., Fusi, N. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature Biomedical Engineering* **2**, 38–47 (2018).

25. Hinz, J. M., Laughery, M. F. & Wyrick, J. J. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry* **54,** 7063–7066 (2015).

26. Horlbeck, M. A., Witkowsky, L. B., Guglielmi, B., Replogle, J. M., Gilbert, L. A., Villalta, J. E., Torigoe, S. E., Tjian, R., Weissman, J. S. Nucleosomes impede cas9 access to DNA in vivo and in vitro. *eLife* **5,** (2016).

27. Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M., Weissman, J. S. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, (2016).

28. Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., Liu, D. R. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

29. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* **36**, 765 (2018).

30. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., Lim, W. A. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

31. Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W., Guilak, F., Crawford, G. E., Reddy, T. E., Gersbach, C. A. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10**, 973–976 (2013).

32. Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., Juong, J. K. CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10**, 977–979 (2013).

33. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).

34. Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chemical Biology* **13**, 406–416 (2018).

35. Anzalone, A. V., Randoplh, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., Liu, D. R. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

36. Yan, J., Cirincione, A. & Adamson, B. Prime Editing: Precision Genome Editing by Reverse Transcription. *Molecular Cell* **77,** 210–212 (2020).

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

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

39. Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., Mammadov, R., Adli, M. CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. *Nature Methods* **14**, 710–712 (2017).

40. Billon, P., Bryant, E. E., Joseph, S. A., Nambiar, T. S., Hayward, S. B., Rothstein, R., Ciccia, A. CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Molecular Cell* **67**, 1068-1079.e4 (2017).

41. Kluesner, M. G., Lahr, W. S., Lonetree, C., Smeester, B. A., Qiu, X., Slipek, N. J. Vasquez, P. N. C., Pitzen, S. P., Pomeroy, E. J., Vignes, M. J., Lee, S. C., Bingea, S. P., Andrew, A. A., Webber, B. R., Moriarity, B. S. CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary cells. *bioRxiv* 2020.04.16.045336 (2020) doi:10.1101/2020.04.16.045336.

42. Kim, D., Lim, K., Kim, S., Yoon, S., Kim, K., Ryu, S., Kim, J. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nature Biotechnology* **35**, 475–480 (2017).

43. Zuo, E., Sun, Y., Wie, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L. M., Li, Y., Yang, H. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).

44. Grünewald, J., Zhou, R., Garcia, S. P., Iyer, S., Lareau, C. A., Aryee, M. J., Joung, J. K. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).

45. Hanna, R., Hegde, M., Fagre, C. R., DeWeirdt, P. C., Sangree, A. K., Szegletes, Z., Griffith, A., Feeley, M. N., Sanson, K. R., Baidi, Y., Koblan, L. W., Liu, D. R. Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064-1080.e20 (2021).

46. Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., Lai, L., Li, Z. Precise base editing with CC context-specificity using engineered human APOBEC3G-nCas9 fusions. *BMC Biology* **18**, (2020).

47. Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., Joung, J. K.. An apobec3a-cas9 base editor with minimized bystander and off-target activities. *Nature Biotechnology* **36**, 977 (2018).

48. Kim, Y. B., Komor, A. C, Levy, J. M., Packer, M. S., Zhao, K. T., Liiu, D. R. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology* **35**, 371–376 (2017).

49. Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., Maianti, J. P., Raguram, A., Liu, D. R. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology* **36**, 843–848 (2018).

50. Song, M., Kim, H. K., Lee, S., Kim, Y., Seo, S., Park, J., Choi, J. W., Jang, H., Shin, J. H., Min, S., Quan, Z., Kim, J. H., Kang, H. C., Yoon, S., Kim, H. H. Sequence-specific prediction of the efficiencies of adenine and cytosine base editors. *Nature Biotechnology* **38**, 1037–1043 (2020).

51. Arbab, M., Shen, M. W., Mok, B., Wilson, C., Matuszek, Z., Cassa, C. A., Liu, D. R. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* **182**, 463-480.e30 (2020).

52. Marquart, K., Allam, A., Janjuha, S., Sintsova, A., Villiger, L., Frey, N., Krauthammer, M., Schwank, G.. Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. *bioRxiv* 2020.07.05.186544 (2020) doi:10.1101/2020.07.05.186544.

53. Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., Welch, M. M., Horng, J. E., Malagon-Lopez, J., Pinello, L., Aryee, M. J., Joung, J. K. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nature Biotechnology* **37**, 276–282 (2019).

54. Sanjana, Neville E. "Genome-scale CRISPR pooled screens." *Analytical biochemistry* vol. 532 (2017): 95-99. doi:10.1016/j.ab.2016.05.014

55. Shalem, O., Sanjana, N. E., Nartenian, E., Shi, X., Scott, D. A., Mikkelson, T., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., Zhang, F. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. doi:10.1126/science.1247005.

56. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology* **32**, 267–273 (2014).

57. Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., Wei, W. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).

58. Sanson, K. R., Hanna, R. E., Hegde, M., Donovan, K. F., Strand, C., Sullender, M. E., Vaimburd, E. W., Goodale, A., Root, D. E., Piccioni, F., Doench, J. G. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature Communications* **9**, (2018).

59. Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. ., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., Weissman, J. S. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661 (2014).

60. Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., McCarroll, S. A. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

61. Datlinger, P., Rendeiro, A. F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L. C., Kuchler, A., Alpar, D., Bock, C. Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* **14**, 297–301 (2017).

62. Dixit, A., Parna, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., Adamson, B., Norman, T. M., Lander, E. S., Weissman, J. S., Friedman, N., Regev, A. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).

63. Jaitin, D. A, Weiner, A., Yofe, I., Lara-Astiaso, D., Shaul, Keren-Shaul, H., David, E., Salame, T. M., Tanay, A., Oudenaarden, A. V., Amit, I. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896.e15 (2016).

64. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Molecular Cell* **66**, 285-299.e5 (2017).

65. Xie, S., Cooley, A., Armendariz, D., Zhou, P. & Hon, G. C. Frequent sgRNA-barcode recombination in single-cell perturbation assays. *PLoS ONE* **13**, (2018).

66. Hill, A. J., McFaline-Figueroa, J. L., Starita, L. M., Gasperini, M. J., Matreyek, K. A., Packer, j., Jackson, D., Shendura, J., Trapnell, C. On the design of CRISPR-based single-cell molecular screens. *Nature Methods* **15**, 271–274 (2018).

67. Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Zilardo, S. B., Wheeler, T. D. McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W.,Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J., Bielas, J. H. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, (2017).

68. Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, A. J., Terry, J. M., Riordan, D. P., Srinivas, N., Fiddes, I. T., Arthur, J. G., Alvarado, L. J., Pfeiffer, K. A., Mikkelsen, T. S., Weissman, J. S., Adamson, B. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nature Biotechnology* **38**, 954–961 (2020).

69. Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., Enard, W. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell* **65**, 631-643.e4 (2017).

70. Jun, S., Lim, H., Chun, H., Lee, J. H. & Bang, D. Single-cell analysis of a mutant library generated using CRISPR-guided deaminase in human melanoma cells. *Communications Biology* **3**, 1–12 (2020).

71. Cuella-Martin, R., Hayward, S. B., Fan, X., Chen, X., Huang, J., Taglialatela, A., Leuzzi, G., Zhao, J., Rabadan, R., Lu, C., Shen, Y., Ciccia, A. Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**, 1081-1097.e19 (2021).

72. Mimitou, E. P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalexi, E., Ouyang, Z., Satija, R., Sanjana, N. E., Koralov, S. B., Smibert, P. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nature Methods* **16**, 409–412 (2019).