



journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)



Review

# The use of base editing technology to characterize single nucleotide variants



Sophia McDaniel <sup>a</sup>, Alexis Komor <sup>b,\*</sup>, Alon Goren <sup>a,\*</sup>

<sup>a</sup> Department of Medicine, University of California San Diego, La Jolla, CA, United States

<sup>b</sup> Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, United States

ARTICLE INFO

Article history:

Received 8 December 2021  
 Received in revised form 23 March 2022  
 Accepted 27 March 2022  
 Available online 31 March 2022

Keywords:

Single nucleotide variants  
 Base editing  
 CRISPR  
 Screens  
 Single-cell RNA sequencing

ABSTRACT

Single nucleotide variants (SNVs) represent the most common type of polymorphism in the human genome. However, in many cases the phenotypic impacts of such variants are not well understood. Intriguingly, while some SNVs cause debilitating diseases, other variants in the same gene may have no, or limited, impact. The mechanisms underlying these complex patterns are difficult to study at scale. Additionally, current data and research is mainly focused on European populations, and the mechanisms underlying genetic traits in other populations are poorly studied. Novel technologies may be able to mitigate this disparity and improve the applicability of personalized healthcare to underserved populations. 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. We further 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. Lastly, we consider the use of such approaches to potentially accelerate the study of genetic mechanisms in non-European populations.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	1671
2. CRISPR/Cas9 as a Gene Knockout Tool	1671
2.1. Limitations of and improvements to CRISPR	1672
3. CRISPR derivatives enable further manipulation of the genome	1672
3.1. Base editing improves the efficiency SNV introduction	1673
3.2. Improvements to the base editor systems	1675
4. Genome-wide pooled CRISPR screens increase the throughput of studying variants	1675
4.1. Positive and negative selection screens	1675
5. CRISPR screens using scRNAseq for higher resolution read-out	1676
5.1. scRNAseq-based CRISPR screens	1676
6. Combining CRISPR screens with BEs	1677
7. Conclusions	1678
8. Competing interests	1678
Declaration of Competing Interest	1678
Acknowledgments	1678
References	1678

From our discussions with Dr. Panagiotou, our manuscript should be in the celebration issue for the 10th anniversary of Computational and Structural Biotechnology Journal.

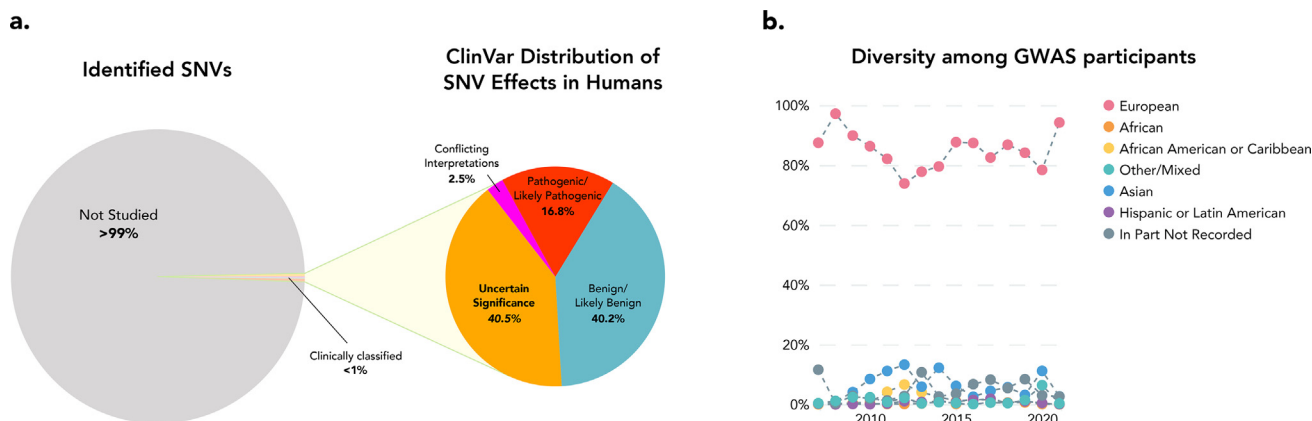
\* Corresponding authors.

E-mail addresses: [akomor@ucsd.edu](mailto:akomor@ucsd.edu) (A. Komor), [agoren@ucsd.edu](mailto:agoren@ucsd.edu) (A. Goren).

<https://doi.org/10.1016/j.csbj.2022.03.031>

2001-0370/© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



**Fig. 1.** ClinVar distribution of SNV effects in humans and demonstration of the GWAS population biases. a. Graphs demonstrating the approximate percent of identified (by sequencing) human genetic variants that have been clinically classified (left), and the distribution of phenotypic effects, as listed in the ClinVar database [53] (right). b. Estimated ratios of the GWAS in different populations during the last 15 years (the colors on the right represent the different populations). The plot was obtained from <https://gwasdiversitymonitor.com/> [67].

## 1. Introduction

Single nucleotide variants (SNVs) are the most common type of polymorphism in the human genome. Recent studies suggest that there are approximately 3–4 million SNVs in the average individual and the recent dbSNP (build 155; [81]) documented over a billion possible distinct SNVs in humans. Of these, 229 million were identified by sequencing according to the gnomAD database [35], and only ~1 million are clinically classified in the Clinvar database [53]. Of the clinically classified SNVs, approximately 16.8% are pathogenic or likely pathogenic, 40.2% are benign or likely benign, and the largest fraction (40.5%) are of uncertain significance [26,53] (Fig. 1).

Point mutations can have varying degrees of impact on protein expression and function in a manner that depends on the exact location of the variant within the gene, as well as the nature of the mutation. Such variants include synonymous mutations (which sometimes have virtually no effect), nonsense mutations (which can knockout a gene), and missense mutations (which can result in a loss- or gain-of-function). Point mutations in the active sites or binding domains of enzymes can be particularly damaging to protein function and cause a plethora of downstream effects that may manifest as a genetic disease [37]. Mutations in noncoding regions (such as enhancers and intronic regions) can have detrimental impacts by modulating expression levels or causing improper mRNA splicing [2,41,56].

Previous studies have linked disease phenotypes to SNVs through a combination of genome-wide association studies (GWAS), analyses of crystal structures, and generation of model organisms harboring SNVs of interest. However, these techniques are highly labor intensive and are mostly low throughput, making it practically impossible to systematically study the SNVs without a priori selection or prioritization. Additionally, most of the collected datasets and research efforts have focused on European individuals, thus limiting the understanding of mechanisms underlying the genetics of non-European populations. In recent years, GWAS data have been used to develop increasingly accurate predictors of disease on the basis of genetic risk factors, allowing for preventative treatment and a personalized approach to healthcare [36]. However, nearly 80% of genetic studies in the past 20 years have been conducted in European populations. Multiple studies have consistently observed the limited applicability of models built on euro-centric data to non-European patients. This discrepancy limits the generalizability of models based on GWAS data to non-European populations, further contributing to the fail-

ure of the current healthcare system to aid underserved populations [64]. Novel techniques are critical to overcome this ethical challenge and fully take advantage of the ever-growing treasure trove of genomic data.

The recent development of CRISPR-based technologies for mammalian cell genome editing has been transformative for researchers' abilities 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 [47]. Importantly, the relative simplicity of using CRISPR-based tools to study genetic mechanisms in various cell types and genetic backgrounds has the potential to greatly expand functional genomics research, enabling researchers to characterize variants from previously understudied populations.

In this review, we first detail the CRISPR/Cas9 system as background for a discussion of base editors (BEs). We then consider how the use of BE screens in combination with single cell RNA sequencing (scRNA-seq) can overcome several current challenges in the field and improve the ability to functionally characterize SNVs. Lastly, we highlight the potential of such approaches to overcome the current population bias.

## 2. CRISPR/Cas9 as a Gene Knockout Tool

To better understand the technological underpinnings of base editing we will first discuss the modified CRISPR/Cas9 system (CRISPR/Cas9 hereafter). Note, while the use of CRISPR/Cas9 and similar approaches as well as their applications have been reviewed elsewhere (e.g., [3,46,82,86]), we provide here a short overview for simplicity and coherence. The CRISPR/Cas9 system, as repurposed for genome editing, is composed of a single-guide RNA (sgRNA), which is designed to target a DNA sequence of interest (called the protospacer), and a Cas9 protein. The sgRNA consists of 2 components: a 20-nucleotide spacer sequence which binds to a target DNA protospacer via sequence complementarity, and a "handle" which folds into a specific three-dimensional structure and is bound by the Cas9 protein. The sgRNA, once transcribed and bound by Cas9 in a target cell, directs the Cas9:sgRNA ribonucleoprotein complex to its complementary protospacer, which must also be directly next to a protospacer adjacent motif (PAM). The PAM sequence is a 2–6 bp sequence motif recognized by the Cas protein. While the canonical PAM for the most commonly used *Streptococcus pyogenes* (Sp) Cas9 is NGG (with N being any of the

nucleotides A, C, G, or T), the specific sequence of the PAM depends on the bacterial or archaeal species from which the Cas protein was taken. While the PAM requirement allows for higher specificity of the targeted gene sequence, it restricts possible cut sites [33,93]. Following DNA binding, SpCas9 will induce a double-strand break (DSB) three base pairs away from the PAM. The protospacer, or ~20 bp sequence complementary to the spacer portion of the sgRNA, is commonly numbered from 1 to 20 for reference purposes with the first base being the furthest away from the PAM. DSB introduction by Cas9 is facilitated by two nuclease domains, the HNH and RuvC domains, which cleave the target and non-target strands, respectively.

The DSBs induced by Cas9 are then processed by the target cell's DNA repair mechanisms, either homology-directed repair (HDR) or, more often, end-joining pathways, such as non-homologous end joining (NHEJ). Notably, NHEJ is error-prone under genome editing conditions, and commonly causes insertions or deletions (indels) of bases at the site of the DSB. These indels can result in frameshift mutations and early stop codons, in effect knocking out the gene. For simplicity, this approach of using CRISPR for inducing gene knockouts will hereafter be referred to as CRISPRko (CRISPR knockout). Researchers may also leverage the HDR repair mechanism by introducing into the cell an exogenous DNA template in parallel, which will be used as a repair template in actively dividing cells. This allows for the introduction of specific mutations of interest, as dictated by the sequence of the DNA template [33,93]. However, NHEJ still occurs in dividing cells, resulting in a mixture of genome editing products due to competition between the two DNA repair pathways. Integration of the repair template by HDR usually happens at a low efficiency (0.5 – 20%) compared to NHEJ (20 – 60%) [18,57].

### 2.1. Limitations of and improvements to CRISPR

While the application of CRISPR for mammalian genome editing was a powerful step forward, there are several limitations to this technology as discussed below. We additionally consider the current efforts aimed at mitigating such drawbacks.

A major limitation to the utility of using CRISPR/Cas to specifically modify the genome with HDR is the low efficiency of this process. This is especially exacerbated when attempting to study recessive mutations, whose effects only become apparent after achieving a knock-in of the mutation in both alleles. Some research groups have overcome this challenge by exclusively studying mutations with a dominant effect or using near-haploid cell lines such as KBM7 [87] and HAP1 [51]. However, these cell lines might not be able to accurately represent the physiology of variants occurring in diploid cells. For instance, these near-haploid cell lines have chromosome structures that differ from those present in normal cells [91]. Further, the haploid state is unstable, and can spontaneously convert to a diploid state [17]. These and similar limitations of these cell lines may call into question the viability of translating findings in haploid cells directly to clinical applications.

One reason for the limited efficiency of CRISPR-mediated gene editing is that DSBs are toxic and can cause cells to enter apoptosis through the p53-mediated DNA damage response before editing has a chance to occur. This cytotoxicity can be reduced by inhibiting p53 expression [70], but this in turn limits the cell's ability to repair other DNA damage and thus increases global mutagenesis rates. Mutations elsewhere in the genome caused by this elevated mutagenesis rate can make it difficult to decipher whether observed phenotypes are truly due to the edit of interest or other uncontrolled factors [24]. The use of DSB-free genome editing techniques, discussed below, can substantially reduce p53-mediated apoptosis.

Editing efficiency can also be low due to inefficient DSB introduction. Optimizing sgRNA design has been a major focus for improving binding specificity and efficiency of cleavage (Table 1). Some studies have identified factors that influence editing efficiency such as GC content and local heterochromatin structure [15,16]. These factors and others were taken into consideration in the development of sgRNA designing algorithms that predict the relative efficiency of cleavage [10,52,58,65,68]. Specifically, it is important for the GC content of the spacer to be neither too low (in which case binding affinity of the Cas9:sgRNA complex for the genomic DNA may, for example, not be high enough for efficient binding to occur) nor too high (in which case the spacer can for instance have unwanted secondary structure that interferes with DNA binding). Additionally, the presence of nucleosomes can impede the Cas9:sgRNA complex's ability to bind to certain genomic loci [28,30]. Therefore, many sgRNA design algorithms will account for chromatin accessibility [29] (Table 1).

An additional challenge with knock-in experiments is that the HDR efficiency drops precipitously as the distance between the DSB and the intended edit site increases [71]. Therefore, finding an optimal protospacer for knock-in experiments may be difficult given the PAM requirement. To overcome this limitation, many groups are optimizing the use of Cas proteins which have alternative PAM requirements such as Cas12. Others have mutated pre-existing Cas proteins to relax the PAM sequence requirement [31,69,85].

While CRISPRko experiments are less restrictive in terms of DSB introduction location requirements, additional factors should be taken into consideration when designing these protospacers. Specifically, only certain indel sequences (when the number of bases inserted or deleted is not a factor of three) will facilitate gene knockout. Indel sequence prediction tools (such as Microhomology-Predictor [7] and inDelphi [80]) can be used to help identify protospacers that will be most effective in attaining a functional indel (Table 1). Additionally, it was shown that the introduction of indels in the coding region proximal to the C-terminus can be ineffective, resulting in truncated proteins that retain some function rather than a full knockout [16]. Conversely, alternative start sites can be used by the translation machinery when indels are introduced at the very N-terminus of the gene [16]. Altogether, there are a multitude of factors that must each be scrutinized 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. The cellular process for correcting the DSBs may induce unwanted genomic modifications at the site of the DSB, including large deletions or genomic rearrangements [49,57]. Additionally certain genes, such as essential genes or genes that show different phenotypic effects in an expression level-dependent manner, can only be studied by modulating expression levels rather than a complete knockout. Next generation CRISPR-derived technologies (such as base editors, discussed below) have mitigated many of these CRISPRko limitations and expanded the applicability of the CRISPR toolbox.

### 3. CRISPR derivatives enable further manipulation of the genome

CRISPR technologies have been modified for a multitude of purposes by inactivating the nucleolytic activity of Cas and appending to the protein new functional groups (Table 2). These CRISPR derivatives have enabled the study of different facets of gene function as well as the ability to perform genome editing with enhanced efficiency and/or precision. Here, we discuss base editors (BEs) as tools to introduce SNVs (Fig. 2).

**Table 1**  
Summary of sgRNA design and microhomology prediction tools (non-exhaustive).

sgRNA Design Tools							
Design Tool	CRISPR Tool	Cas Types Included	Input	Output	Cell Types	Weblink	Study
pegFinder	Prime editing	Cas9	~200 bp sequence context and cleavage site, desired PAM sequence, and desired edited sequence	Top sgRNAs, on-target score, off-target score (for Cas9-NGG only), secondary nicking sgRNAs	Not specified	<a href="https://pegfinder.sidichenlab.org/">https://pegfinder.sidichenlab.org/</a>	[10]
CHOPCHOP	CRISPRko, CRISPRa, CRISPRi	Cas9; Cas9n; Cas12a; Cas13	Target gene, reference genome of interest	A list of scored sgRNAs as well as a visual representation	Not specified	<a href="https://chopchop.cbu.uib.no">https://chopchop.cbu.uib.no</a>	[52]
Elevation	CRISPRko	Cas9	Gene/transcript ID	Top sgRNAs, on target score, off target score	Not specified	<a href="https://crispr.ml/">https://crispr.ml/</a>	[58]
GUIDES	CRISPRko	Cas9	Genome of interest, gene of interest	Top sgRNAs, on target score, off target score	Not specified	<a href="https://github.com/sanjanalab/GUIDES">https://github.com/sanjanalab/GUIDES</a>	[65]
CRISPRscan	CRISPRko	Cas9; Cas12a	Genome of interest, gene or transcript of interest	CRISPRscan score	Not specified	<a href="https://www.crisprscan.org/">https://www.crisprscan.org/</a>	[68]
BE-Hive	Base Editing	ABE; BE4	Target genomic DNA sequence, desired protospacer sequence	Editing efficiency score for each editable base within edit window	HEK293T; mES	<a href="https://www.crisprbehave.design/guide">https://www.crisprbehave.design/guide</a>	[5]
BE-DICT Bystander	Base Editing	ABE; BE4	Target genomic DNA sequence	Top sgRNAs, editing efficiency score for each editable base within the edit window	Not specified	<a href="https://bedict.forone.red/#/bystander">https://bedict.forone.red/#/bystander</a>	[63]
DeepBaseEditor	Base Editing	ABE; CBE	Target genomic DNA sequence	Top sgRNAs, DeepCas9 score based on edit efficiency and probability of bystander edits	Not specified	<a href="https://deepcrispr.info/DeepBaseEditor/">https://deepcrispr.info/DeepBaseEditor/</a>	[83]
Microhomology Prediction Tools							
CRISPR RGEN Tools: Microhomology-Predictor	All engineered nucleases	Cas9	60–80 bp seq flanking the cleavage site Up to 5 kb sequence	Score predicting likelihood of out-of-frame deletions Scores for all possible Cas9 targets in the inquiry	Not specified	<a href="https://www.rgenome.net/mich-calculator/">https://www.rgenome.net/mich-calculator/</a>	[7]
inDelphi	CRISPRko	Cas9	Sequence context and cleavage site, cell type of interest, and PAM sequence of interest	Score predicting likelihood of out-of-frame deletions	mESC; HEK293; U2Os; K563; HCT116	<a href="https://www.crisprindelphi.design/">https://www.crisprindelphi.design/</a>	[80]

**Table 2**  
Summary of CRISPR-based tools (non-exhaustive).

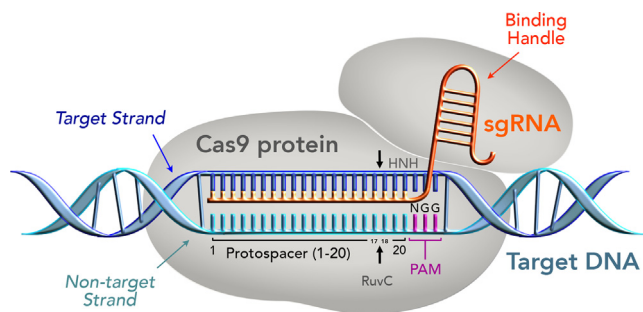
CRISPR Derivative	Cas Protein Type	Edit Type	Appended Enzyme	Modification	Study
CRISPR	CRISPR	Indels	N/A	N/A	[33]
CRISPRi	dCas9	Inhibition	N/A	Both nucleases deactivated	[75]
CRISPRa	dCas9	Activation	VP64	Both nucleases deactivated and VP64 transcriptional activator domain appended	[62]
CRISPRa	dCas9	Activation	VP64	Both nucleases deactivated and VP64 transcriptional activator domain appended	[74]
ABE	nCas9	Transition SNP	Modified TadA adenosine deaminase	Both nucleases deactivated and evolved TadA adenosine deaminase appended	[20]
BE1	dCas9	Transition SNP	APOBEC	Both nucleases deactivated and APOBEC cytosine deaminase appended	[47]
BE2	dCas9	Transition SNP	APOBEC	UGI appended	
BE3	nCas9	Transition SNP	APOBEC	HNH domain nucleolytic activity restored	
BE4	nCas9	Transition SNP	APOBEC	Linker lengths extended and additional UGI appended	[48]
BE4max	nCas9	Transition SNP	APOBEC	NLS appended	[44]
PE1	nCas9	Small indels, Transition SNPs, Transversion SNPs	M-MLV Reverse Transcriptase	Reverse transcriptase appended; pegRNA appended	[4]
PE2	nCas9	Small indels, Transition SNPs, Transversion SNPs	M-MLV Reverse Transcriptase	Reverse transcriptase modified	
PE3	nCas9	Small indels, Transition SNPs, Transversion SNPs	M-MLV Reverse Transcriptase	Additional Cas9 nickase appended; simple sgRNA targeting complementary strand appended	

### 3.1. Base editing improves the efficiency SNV introduction

The efficiency and precision of inducing single-nucleotide edits with CRISPR technologies has greatly improved with the implementation of base editing. Base editing uses the targeting abilities

of the Cas9:sgRNA complex, but tethers to it a deaminase enzyme that specifically modifies target nucleotides. The original base editor, a cytosine base editor (CBE), uses the rat APOBEC1 protein as its deaminase component. In typical CBE constructs, APOBEC1 is tethered to the N-terminus of a partially inactivated Cas9, Cas9n.





**Fig. 2.** Structure and mechanism of the CRISPR/Cas system. The CRISPR/Cas system (exemplified by Cas9 above) is composed of an sgRNA (shown in orange), which is complementary to the target DNA and binds to the target DNA, and a Cas protein (shown in grey), which helps bind and cleave the DNA through two nucleolytic domains. The HNH nuclease domain cleaves the complementary (i.e., target) strand while the RuvC nuclease cleaves the non-target strand. By definition, the non-target strand is called the protospacer, and has the same sequence as the sgRNA. The protospacer adjacent motif (PAM) is located directly downstream of the protospacer and is required for the Cas protein to initiate DNA binding [33]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

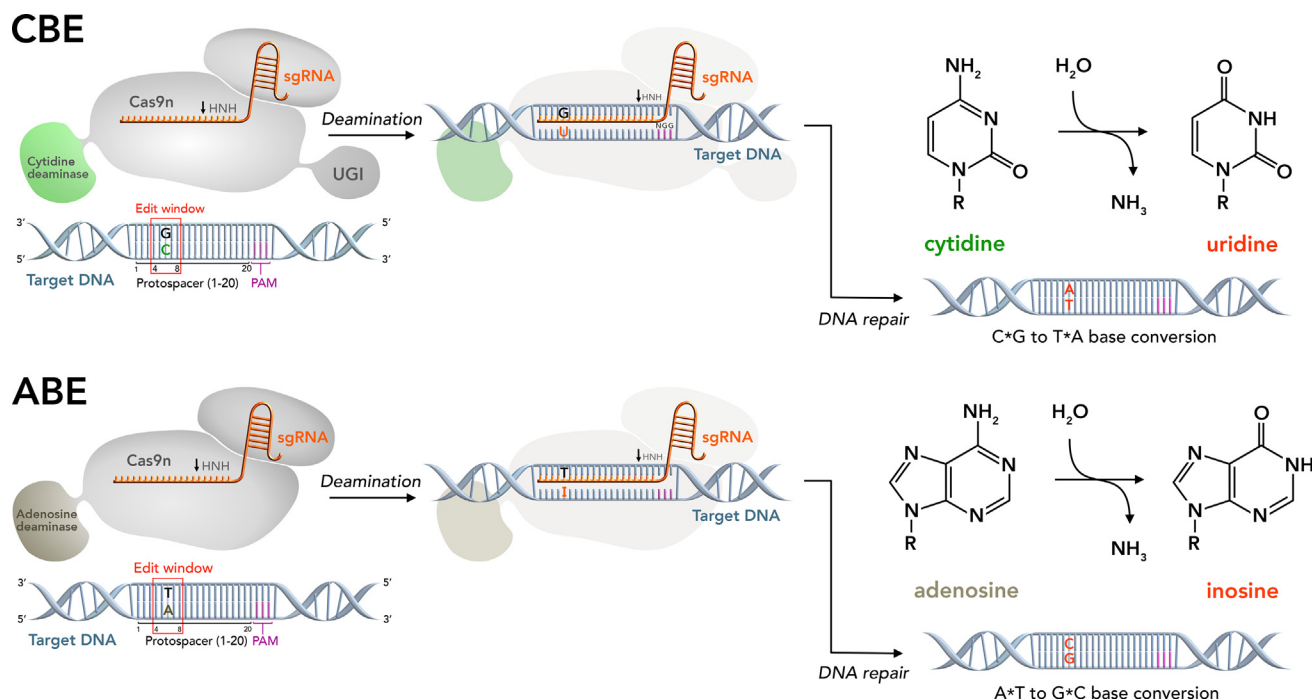
This mutated Cas protein has an inactive RuvC domain but retains nucleolytic activity in the HNH domain, causing only one strand to be cleaved [47] (Fig. 3). A uracil DNA glycosylase inhibitor (UGI) adapted from *Bacillus subtilis* is linked to the C-terminus of the editor to prevent DNA repair pathways from excising the base editing intermediate [48] (Fig. 3). Upon DNA binding by the Cas9:sgRNA component of the CBE, APOBEC1 may convert any cytidine residues within a section of the protospacer (called the “base editing window”, described in more detail below) to uridine residues, which have the base pairing properties of thymidine. DNA backbone cleavage by the HNH domain stimulates DNA repair to preferen-

tially replace the cleaved strand, using the uracil-containing strand as a template [47]. Overall, this catalyzes a C•G to T•A base pair conversion which is an example of a transition mutation – namely a pyrimidine being exchanged for a pyrimidine, or a purine for a purine.

The other type of transition mutation, A•T to G•C, is achieved using an adenine base editor (ABE; Fig. 3). ABEs work similarly to CBEs but use as their deaminase component a mutated RNA adenosine deaminase from *E. Coli* (TadA) that was artificially evolved to bind and modify DNA with high efficiency. ABEs do not require a UGI or other DNA repair manipulation components. ABEs may deaminate adenosines within the base editing window to inosines, which are recognized as guanosines by the DNA replication machinery. This leads to an A•T to G•C edit after processing of the intermediate by DNA repair [20].

Importantly, the deaminase components of both CBEs and ABEs can only act on single-strand DNA. This confines their activity to only accessible nucleotides on the non-target DNA strand and therefore restricts editing to a small window of ~5 nucleotides located between bases 4–8 in the protospacer (Fig. 3). If multiple target bases (cytidines for CBEs, adenosines for ABEs) are located within the editing window, “bystander editing” may occur, in which multiple bases are edited, albeit with varying efficiencies [20,47].

Base editing has been implemented to generate gene knockouts by introducing premature stop codons [9,50] or disrupting splice sites [43]. SNVs induced to cause premature stop codons can achieve gene knockouts with higher predictability and efficiency compared to NHEJ or HDR methods for gene knockouts [50]. Additionally, base editing intermediates are less toxic than DSBs and do not cause large chromosomal rearrangements when multiplexed, making gene knockout with BEs (particularly when knocking out multiple genes) a viable substitute for CRISPRko.



**Fig. 3.** Schematic of the structural makeup and mechanism of base editors. Top: Cytosine base editors (CBEs) are composed of a catalytically impaired Cas9 protein (Cas9n) with a cytidine deaminase fused to the N-terminus and a uracil glycosylase inhibitor fused to the C-terminus. After the Cas9n:sgRNA complex binds to the target DNA, the cytidine deaminase may convert any cytidine bases within the edit window to uridines. DNA repair pathways then preferentially replace the non-edited strand and incorporate an adenine base across from the uridine. Overall, CBEs catalyze a C•G to T•A base pair conversion. Bottom: Adenine base editors (ABEs) are similar to CBEs but have an adenine deaminase as the DNA base modifying enzyme and have no DNA repair inhibitor on the C-terminus. After the Cas9n:sgRNA complex binds, adenosine residues in the edit window may be converted to inosines, which have the base-pairing properties of guanosine. Overall, ABEs catalyze an A•T to G•C base pair conversion.

### 3.2. Improvements to the base editor systems

Despite their differences in mechanism, BEs present similar limitations to their parental CRISPR systems. These include low editing efficiencies with certain gRNAs, off-target DNA editing [38,98], and protospacer design limitations due to PAM requirements, which ultimately restrict the number of editable bases. In addition, the deaminase components of both CBEs and ABEs may deaminate large numbers of cytidines and adenosines, respectively, in both protein-coding and non-coding RNAs [23]. Further, certain CBEs can cause elevated mutations in genomic DNA due to DNA replication and transcription “bubbles” that expose single stranded DNA to the cytidine deaminase component. Additionally, results obtained by base editing experiments may be complicated by bystander editing which make it difficult to tease apart the extent to which individual SNVs cause a phenotypic effect [25].

As with CRISPRko, multiple approaches have been taken to further improve the functionality of BEs and overcome these limitations (Table 2). For example, certain mutations in the deaminase component of both CBEs and ABEs have resulted in variants with reduced off-target RNA editing (and reduced off-target DNA editing for CBE [95]). Further, engineered APOBEC enzymes have reduced bystander editing by making the editing window narrower (from ~5 nucleotides to ~2 nucleotides [21,40], or imparting a sequence motif preference upon the deaminase [21,60]). Construct improvement, such as modulating the number of nuclear localization signals included in the vector and codon optimization, can also enhance the editing efficiency by increasing import of BEs into the nucleus [44]. Additionally, the development of sgRNA prediction tools specific to BEs generated through machine learning from large datasets of BE:sgRNA libraries have enabled better prediction of the factors that facilitate higher base editing efficiencies [5,63,83] (Table 1). Finally, modified BEs with relaxed PAM requirements have expanded the number of targetable bases in the genome [42,44,60,85].

Further, as with every technology, BEs have drawbacks inherent to the technology that somewhat limit their utility in understanding genetic variation as a whole. Base editing by nature is restricted to introducing transition mutations, as discussed above. Other mutation types, such as transversions, deletions, and duplications cannot be implemented using this technology. Prime editing [4], reviewed elsewhere [3], is an additional iteration of CRISPR/Cas9 that has the flexibility to induce any type of transition or transversion mutation, as well as small insertions and deletions into the target gene *in vitro* and *in vivo* [19,59,72]. Though prime editing technology will likely require similar iterations and improvements as discussed with CRISPR and BEs, prime editors may contribute to our understanding of genetic variation by further empowering biologists to induce mutations of interest and study the downstream effects. Still, base editing has much to offer, given that SNVs make up the vast majority of variation in the human genome.

Altogether, there have been a multitude of efforts to improve BEs and similar approaches 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. New techniques are developed at such a rapid pace that studies often utilize BEs that are already outdated by the time they are published. It is challenging to identify the ideal BE to use for new research experiments because the reliability of each version could not be rigorously tested over a long period of time. This disadvantage of “bleeding edge” technology is important when considering the significance and impact of genome editing based studies.

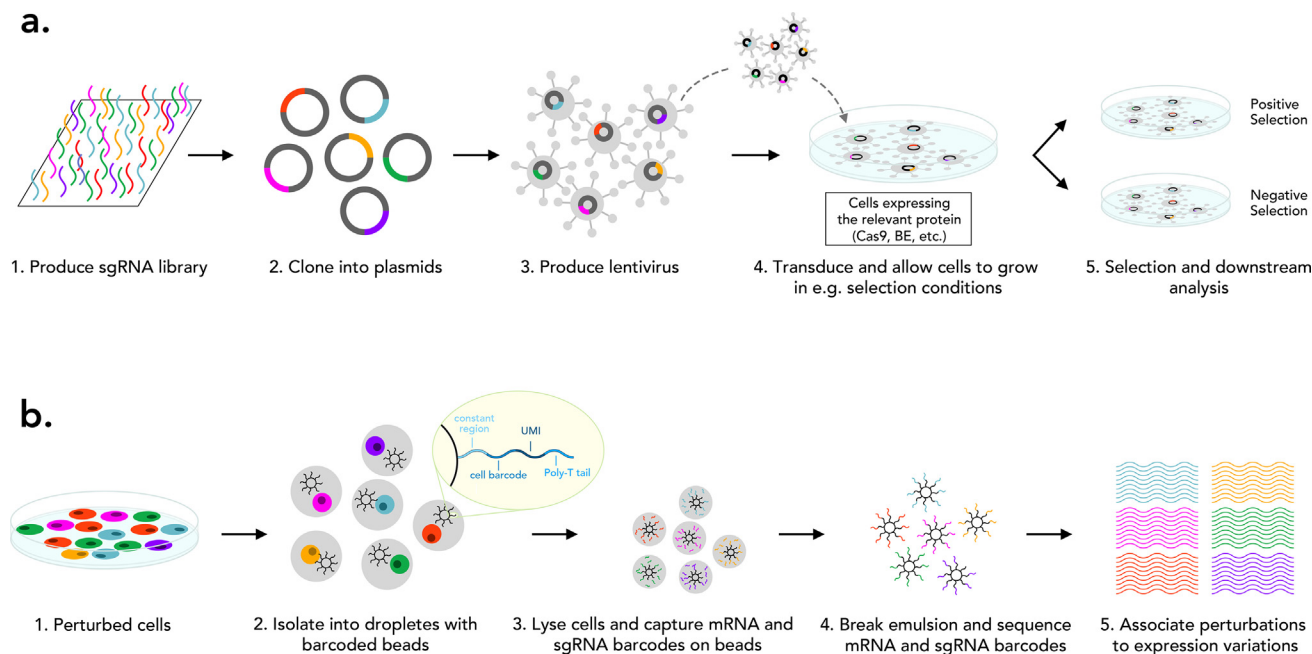
## 4. Genome-wide pooled CRISPR screens increase the throughput of studying variants

The power and efficiency of studying gene function and interactions increased dramatically with the introduction of CRISPR screens. CRISPR systems are ideal for screens because their variable component, the sgRNA, is small and straight-forward to design, making them conducive for library design, generation, and transduction into cells. One way to conduct CRISPRko screens is by producing a library of sgRNAs along with genetic “barcodes” (used to link cells to the sgRNA they received downstream) and transducing them into target cells using a lentiviral vector at a low multiplicity of infection, such that each cell receives only one sgRNA. The target cells may already contain the Cas protein, or it may be transduced along with the sgRNA. The transduced cells will receive their respective edits (in effect knocking out the corresponding gene) after the sgRNA and Cas protein are expressed, resulting in a library of cells (each cell is considered a “library member”), in which each cell has a particular gene knocked out. The cells are then subjected to a perturbation/challenge and given time (typically a few weeks) to grow and compete with one another in response to the given perturbation/challenge. The cells are then analyzed to identify genes whose knockout cause selective advantages or disadvantages in response to the perturbation/challenge [76].

### 4.1. Positive and negative selection screens

Positive and negative selections represent a category of CRISPR screens in which the impact of mutations is assessed based on relative growth within a population of perturbed cells (Fig. 4). In positive selection screens, the perturbation (such as treatment with a drug or toxin) results in a growth advantage for a subset of the perturbed cells, which then overtake the population. The sgRNAs that increase in abundance after the perturbation correspond to target genes involved in resistance to the given challenge (because their knockout results in cells no longer susceptible to the drug or toxin). In negative selection screens, a non-perturbed cell population is compared to a perturbed cell population to identify sgRNAs that decreased in abundance due to the perturbation. These sgRNAs correspond to target genes that are required for cell proliferation in response to the perturbation (because their knockout results in a growth disadvantage when exposed to the perturbation). One limitation of CRISPRko screens is that knockout of essential genes is inherently disadvantageous for cell growth, and it is therefore difficult to obtain reliable data for these genes. This challenge was mitigated by CRISPR interference (CRISPRi) screens, using a dCas9-transcriptional repressor fusion construct to knockdown the genes of interest. In a parallel manner, CRISPR activation (CRISPRa) screens employ a dCas9-transcriptional activator construct to induce gene expression.

Early implementations of CRISPR screens at a genome-wide level were focused on selecting for resistance to 6-thioguanine [78], Vemurafenib [87], *Clostridium septicum* alpha-toxin [45], anthrax, and diphtheria toxins [96]. These studies identified genes essential to the DNA damage response to 6-thioguanine, genes whose loss leads to resistance to Vemurafenib (a cancer treatment drug) and provided a better understanding of pathways that result in cell death by microbial toxins, respectively. These early proof-of-concept efforts verified expected results for previously well-studied genes and served to establish genome-wide sgRNA libraries for future work. Subsequent studies led to the identification of optimized sgRNA libraries [77] for use with CRISPRi/a screens [22,30,77].



**Fig. 4.** A general schematic of BE screens and single cell RNA sequencing (scRNAseq) workflows. Note, as there are multiple approaches that can be used for linking scRNAseq with BEs (or CRISPRko/a/i) screens we provide here a general overview of the main steps. a. A library of sgRNA spacer sequences is designed and generated, then assembled into a viral vector. Lentiviruses are then produced, which are transduced into a population of target cells that express a BE. Expression of both the BE and a sgRNA will result in the introduction of an SNV of interest. The resulting cell population (which harbors a library of SNVs) is then subjected to a challenge to induce growth competition, followed by investigation of the effect of each SNV. b. After subjecting a pool of cells (that harbor a library of mutations) to a perturbation, the cells are passed through a microfluidics device to isolate individual cells into droplets containing a barcoded bead. The cells are lysed, and mRNA is captured by oligonucleotides on the beads. The oligonucleotides are reverse transcribed into a library, the droplets are recombined, then the library is sequenced and analyzed [61].

However, these initial low-resolution CRISPR 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. Moreover, specific phenotypes caused by cell cycle effects or cell subpopulations may be masked because of the low-resolution readout [76]. Recent screens have expanded beyond cell survival or growth-based assays, including techniques that rely on fluorescence-activated cell sorting (FACS) to physically separate different populations of cells based on differences in cell morphology, gene expression levels, and virus infectivity [8,12,73,79,88,92]. While these advancements have enabled researchers to use CRISPR screens to study more complex phenotypes, only a single phenotype can be studied at once. Furthermore, detailed mechanistic information regarding the phenotype being studied can only be elucidated with additional studies. The coupling of CRISPR screens with single cell sequencing technologies, however, can greatly expand the ability to gain mechanistic information associated with a specific phenotype.

## 5. CRISPR screens using scRNAseq for higher resolution read-out

Studying the transcriptomic readout of a cell following a specific perturbation (such as knockout, knockdown, or activation of a specific gene) provides valuable information that has the potential to uncover mechanistic details behind specific phenotypes. This has been previously achieved by selecting single cells from a population of modulated cells and performing RNA-seq as an assay; yet this approach is limited in its scalability. single-cell RNA sequencing (scRNAseq), however, has provided a key opportunity to scale up this process.

The comparative performance of various scRNAseq platforms was evaluated several years ago [97]. Currently, one of the most

commonly used methods include isolation of single cells into nanoliter droplets that each contain a unique bead. The bead's surface is coated with oligonucleotides that have four 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 on the poly T region of the oligonucleotide, reverse transcribed, amplified, and sequenced. The CBCs are used to trace each sequenced mRNA transcript back to the cell from which it originated, and the UMIs can correct for amplification bias [94].

### 5.1. scRNAseq-based CRISPR screens

To enable the use of scRNAseq together with CRISPR screens, the sgRNA needs to be captured and sequenced with the transcriptome (Fig. 4). This was demonstrated by several groups and can be done via two general approaches: (i) A unique polyadenylated guide barcode (GBC) can be included in the sgRNA viral vector construct. The poly T region of the bead will then capture the sgRNA construct, in the process appending it to the CBC. The GBC is sequenced and connected back to its corresponding sgRNA [1,14,32,90]; (ii) In an alternative approach, named CROP-seq, a poly A tail is simply added to the end of the sgRNA transcript to allow for capture by the poly T region of the bead. Here, the sgRNA spacer sequence is directly sequenced [13]. Later analysis has shown that these two general approaches are susceptible to challenges associated with the use of lentiviruses. In particular, it was found that in the first approach, the GBCs could be uncoupled from their respective sgRNAs due to lentiviral recombination. This recombination can happen as often as 50% of the time, depending on the distance between the barcode and the sgRNA [27,89]. On the other hand, while CROP-seq was not impacted by barcode swapping, it could only capture guides in 40–60% of the cells, and thus lost a substantial amount of transcriptomic data [13]. Targeted amplification of the guide RNA [27] improved this efficiency.



Optimizing the signal-to-noise ratio is an additional intrinsic challenge that must be considered when working with scRNAseq-based technology. Single-cell technology must by nature rely on an extremely small amount of starting material that is amplified to the level necessary for sequencing of the resulting library. This process involves reverse transcription and subsequent PCR amplification, both of which are imperfect *in vitro* molecular biology reactions that have the potential to introduce errors and biases [39]. The ability to differentiate between the technical noise and signal that arises from biological variation between cells is crucial to maximize the value of scRNAseq data. Multiple approaches have been employed to address such issues, including the appendment of UMIs to the barcoded beads to retroactively correct for amplification bias, as discussed above. Additionally, several studies have created computational models that use spike-in molecules to correct for technical and biological noise within a sample [6]. Further work is still necessary, and it is important to understand the drawbacks of this technology to know the utility and limitations of single-cell sequencing data as the field continues to progress.

## 6. Combining CRISPR screens with BEs

Just as CRISPRko, CRISPRi, and CRISPRa screens have been used to study the impact of removing, reducing, or enhancing the expression of a library of genes, respectively, BE screens can be used to systematically study the effect of a library of SNVs. In these experiments, cells expressing a BE are transduced with a sgRNA library to produce a pool of cells harboring a library of SNVs. The library is then subjected to a perturbation, and the relative abundance of each sgRNA or SNV in the resulting population is used to relate that SNV to the phenotype of interest. To date, few studies have used BE screens, and an even smaller number have used BE screens in combination with scRNAseq. Though BEs are limited in terms of the number of possible SNVs that they can introduce, the sheer volume of uncharacterized SNVs (Fig. 1) is considerable enough such that even a highly reduced SNV library would be an excellent contribution to the variant interpretation challenge. Here, we discuss the recent efforts in performing BE screens (Table 3) and in particular focus on the use of scRNA-seq in conjunction with BE (scBE screens, hereafter).

While scBE screens are still in their preliminary stages of implementation, phenotypic selection-based BE screens have provided valuable insights that can be applied when designing and implementing scBE screens. One such study demonstrated the utility of BE screens in discovering clinically relevant SNVs causing gain-of-function or loss-of-function phenotypes. This work studied over 50,000 C•G to T•A SNVs across ~3,500 genes [25]. The sgRNA library included ~70,000 members and was coupled with CBE-expressing cells. The resulting SNV library was screened via both positive and negative selections. The authors also performed the same screens using CRISPRko machinery with the sgRNA library to directly compare the impact of each SNV to a corresponding

indel as a control. In this study, sgRNAs that were significantly enriched in a BE screen but not an analogous CRISPRko screen mapped well to known pathogenic variants in the ClinVar database [54], establishing the utility of BE screens in clinically classifying SNVs [25].

Another phenotypic selection-based BE screen used a CBE in HAP1 cells with a library of sgRNAs that were tiled across all *BRCA1* exons [51]. Following SNV library generation, cells were challenged with the PARP inhibitor Olaparib, a chemotherapeutic agent frequently used to treat patients with *BRCA1* mutant cancers. The sgRNAs from the cells that survived Olaparib treatment were then sequenced, revealing 13 sgRNAs that corresponded to SNVs that were known pathogenic mutations according to the ClinVar database [54], as well as multiple other variants of uncertain significance (VUS). The VUS were then shown to be pathogenic via a downstream analysis. This work was an important proof-of-concept study that established BE screens as a method to functionally interrogate SNVs in DNA repair genes, but it was conducted on a relatively small scale (only 745 sgRNAs), and in the near-haploid HAP1 cell line.

Another BE screen focused on investigating the impact of mutations in DNA damage response (DDR) genes by using sgRNAs to target 37,000 SNVs across 86 DDR genes [11]. Cells harboring the SNV library were then separately challenged with four DNA damaging agents (Cisplatin, Olaparib, Doxorubicin, and Camptothecin) and analyzed to determine enrichment and depletion of sgRNAs. In this study, the expectation was that sgRNAs that became enriched would represent SNVs that provided resistance to these DNA damaging agents by blocking checkpoint regulations, and thus allowing the cells to proliferate. Importantly, this work correctly differentiated between known pathogenic and benign SNVs from the ClinVar database and predicted the clinical relevance of a variety of VUS in DDR genes [11].

One of the first scBE screens focused on *MAP2K1*, *KRAS*, and *NRAS*, mutations which have been shown to be associated with Vemurafenib resistance in melanoma [34]. This work employed a CBE (BE3) in combination with all possible sgRNAs across the target genes and screened for conferral of Vemurafenib resistance. The surviving cells (which harbored Vemurafenib resistance) were then subjected to scRNA-seq. Notably, the use of transcriptomic data allowed identification of cell subpopulations that would have been masked if the cells were sequenced in bulk rather than at a single-cell level. This revealed two distinct clusters with different mechanisms of acquired Vemurafenib resistance. The first cluster, composed primarily of mutations in *MAP2K1*, resulted in an upregulation of immune response genes. The second subpopulation identified, which had mainly *KRAS* mutations, was enriched in the chemokine signaling pathway. These differences are informative regarding the distinct mechanisms by which cancer cells acquire Vemurafenib resistance. While this study successfully classified SNVs, the authors cited low efficiency of SNV introduction (5–20%) by BE3, and comparatively low-throughput of the implemented scRNAseq method (CROP-seq) as areas for improvement.

**Table 3**  
Summary of example BE screen studies.

Investigated Genes	Selection Type	Selection Agent	Editor Type	scRNAseq Capture Method	Cell Line(s)	Study
MAP2K1, KRAS, NRAS	Positive	Vemurafenib	BE3	CROP-seq	A375	[34]
Multiple	Positive/Negative	Cisplatin, Hygromycin	BE3, BE4	N/A	HT29, MELJUSO	[25]
BRCA1	Negative	Olaparib	BE3	N/A	HAP1	[51]
DDR Pathway	Positive/Negative	Cisplatin, Olaparib, Doxorubicin, Camptothecin	BE3	N/A	MFC10A, MFC7, HAP1	[11]



Substituting these methods with newer generation BEs and advanced scRNAseq approaches could potentially improve the readout for future studies.

## 7. Conclusions

To conclude, scBE screens 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 scBE screens 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 with improvements in single cell technologies, allowing incorporation of additional genomic measurements such as chromatin accessibility or protein expression [55,66].

Moving forward, we expect that scBE screens will start to be conducted using cell types that may provide added physiological relevance; for instance, cells differentiated from human pluripotent stem cells or organoids. Such relevant cell model systems would have the potential to improve the significance of these experimental, laboratory-based results to clinical applications. Further, we expect that scBE screens will further improve the ability to systematically investigate SNVs in a variety of genes, from DDRs to trans-acting factors such as chromatin regulators or transcription and splicing factors. Though GWAS studies have progressed the field of genetics significantly in terms of picking apart genotype-phenotype associations, they often lack the granularity of identifying causal SNVs and translating findings to clinical applications [84]. We expect data garnered through scBE screens to supplement GWAS data and create a bridge between genetic sequencing data and medical advancements. Lastly, we see scBE screens being used in the future to perform comparative studies of SNVs between cells derived from individuals from various populations. Thus, for instance, we envision that performing such screens in human induced pluripotent stem cells (and relevant differentiated cells derived from these cells) may help elucidate the diverse impacts of certain SNVs when introduced into different genomic backgrounds.

## 8. Competing interests

A.C.K. is a member of the SAB and a consultant of Pairwise Plants and is an equity holder for Pairwise Plants and Beam Therapeutics. 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.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank Rahel Wachs for her help with the graphic illustrations, and Drs. Sharona Shleizer-Burko, Amit Majithia, and Xin Sun for critical comments. We acknowledge the NIH support with grant no. 1R35GM138317-01 to A.C.K. and 1RM1HG011558-01 to A.G.

## References

- [1] Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 2016;167(1867–1882):e1821.
- [2] Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet* 2018;59:253–68.
- [3] Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* 2020;38:824–44.
- [4] Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;576:149–57.
- [5] Arbab M, Shen MW, Mok B, Wilson C, Matuszek Z, Cassa CA, et al. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* 2020;182(463–480):e430.
- [6] Bacher R, Chu LF, Argus C, Bolin JM, Knight P, Thomson JA, et al. Enhancing biological signals and detection rates in single-cell RNA-seq experiments with cDNA library equalization. *Nucleic Acids Res* 2022;50:e12.
- [7] Bae S, Kweon J, Kim HS, Kim JS. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods* 2014;11:705–6.
- [8] Baggen J, Persoons L, Vanstreels E, Jansen S, Van Looveren D, Boeckx B, et al. Genome-wide CRISPR screening identifies TMEM106B as a proviral host factor for SARS-CoV-2. *Nat Genet* 2021;53:435–44.
- [9] Billon P, Bryant EE, Joseph SA, Nambiar TS, Hayward SB, Rothstein R, et al. CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Mol Cell* 2017;67(1068–1079):e1064.
- [10] Chow RD, Chen JS, Shen J, Chen S. A web tool for the design of prime-editing guide RNAs. *Nat Biomed Eng* 2021;5:190–4.
- [11] Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Tagliatala A, et al. Functional interrogation of DNA damage response variants with base editing screens. *Cell* 2021;184(1081–1097):e1019.
- [12] Daniloski Z, Jordan TX, Wessels HH, Hoagland DA, Kasela S, Legut M, et al. Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells. *Cell* 2021;184(92–105):e116.
- [13] Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* 2017;14:297–301.
- [14] Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 2016;167(1853–1866):e1817.
- [15] Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 2016;34:184–91.
- [16] Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 2014;32:1262–7.
- [17] Essletzbichler P, Konopka T, Santoro F, Chen D, Gapp BV, Kralovics R, et al. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res* 2014;24:2059–65.
- [18] Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013;31:822–6.
- [19] Gao P, Lyu Q, Ghanam AR, Lazzarotto CR, Newby GA, Zhang W, et al. Prime editing in mice reveals the essentiality of a single base in driving tissue-specific gene expression. *Genome Biol* 2021;22:83.
- [20] Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A>T to G>C in genomic DNA without DNA cleavage. *Nature* 2017;551:464–71.
- [21] Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol* 2018;36:977–82.
- [22] Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 2014;159:647–61.
- [23] Grunewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* 2019;569:433–7.
- [24] Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* 2018;24:927–30.
- [25] Hanna RE, Hegde M, Fagre CR, DeWeirdt PC, Sangree AK, Szegletes Z, et al. Massively parallel assessment of human variants with base editor screens. *Cell* 2021;184(1064–1080):e1020.
- [26] Henrie A, Hemphill SE, Ruiz-Schultz N, Cushman B, DiStefano MT, Azzariti D, et al. ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data. *Hum Mutat* 2018;39:1051–60.
- [27] Hill AJ, McFaline-Figueroa JL, Starita LM, Gasperini MJ, Matreyek KA, Packer J, et al. On the design of CRISPR-based single-cell molecular screens. *Nat Methods* 2018;15:271–4.
- [28] Hinz JM, Laughery MF, Wyrick JJ. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry* 2015;54:7063–6.
- [29] Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* 2016;5.

- [30] Horlbeck MA, Witkowski LB, Guglielmi B, Plogle JM, Gilbert LA, Villalta JE, et al. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife* 2016;5.
- [31] Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 2018;556:57–63.
- [32] Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* 2016;167(1883–1896):e1815.
- [33] 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;337:816–21.
- [34] Jun S, Lim H, Chun H, Lee JH, Bang D. Single-cell analysis of a mutant library generated using CRISPR-guided deaminase in human melanoma cells. *Commun Biol* 2020;3:154.
- [35] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
- [36] Karczewski KJ, Snyder MP. Integrative omics for health and disease. *Nat Rev Genet* 2018;19:299–310.
- [37] Katsonis P, Koire A, Wilson SJ, Hsu TK, Lua RC, Wilkins AD, et al. Single nucleotide variations: biological impact and theoretical interpretation. *Protein Sci* 2014;23:1650–66.
- [38] Kim D, Lim K, Kim ST, Yoon SH, Kim K, Ryu SM, et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol* 2017;35:475–80.
- [39] Kim JK, Kolodziejczyk AA, Illicic T, Teichmann SA, Marioni JC. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat Commun* 2015;6:8687.
- [40] 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;35:371–6.
- [41] Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
- [42] Kleinstiver BP, Sousa AA, Walton RT, Tak YE, Hsu JY, Clement K, et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* 2019;37:276–82.
- [43] Kluesner MG, Lahr WS, Lonetree CL, Smeester BA, Qiu X, Slipek NJ, et al. CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary and immortalized cells. *Nat Commun* 2021;12:2437.
- [44] Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* 2018;36:843–6.
- [45] Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Md C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 2014;32:267–73.
- [46] Komor AC, Badran AH, Liu DR. Editing the Genome Without Double-Stranded DNA Breaks. *ACS Chem Biol* 2018;13:383–8.
- [47] 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;533:420–4.
- [48] Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, et al. 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:eaa04774.
- [49] Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* 2018;36:765–71.
- [50] Kucus C, Parlak M, Tufan T, Yang J, Szlachta K, Wei X, et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat Methods* 2017;14:710–2.
- [51] Kweon J, Jang AH, Shin HR, See JE, Lee W, Lee JW, et al. A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. *Oncogene* 2020;39:30–5.
- [52] Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucl Acids Res* 2019;47:W171–4.
- [53] Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucl Acids Res* 2016;44:D862–868.
- [54] Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucl Acids Res* 2018;46:D1062–7.
- [55] Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med* 2020;52:1428–42.
- [56] Leslie R, O’Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* 2014;30:i185–194.
- [57] Lin Y, Cradick TJ, Brown MT, Deshmukh H, Ranjan P, Sarode N, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucl Acids Res* 2014;42:7473–85.
- [58] Listgarten J, Weinstein M, Kleinstiver BP, Sousa AA, Joung JK, Crawford J, et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat Biomed Eng* 2018;2:38–47.
- [59] Liu Y, Li X, He S, Huang S, Li C, Chen Y, et al. Efficient generation of mouse models with the prime editing system. *Cell Discov* 2020;6:27.
- [60] Liu Z, Chen S, Shan H, Jia Y, Chen M, Song Y, et al. Precise base editing with CC context-specificity using engineered human APOBEC3G-nCas9 fusions. *BMC Biol* 2020;18:111.
- [61] Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 2015;161:1202–14.
- [62] Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 2013;10:977–9.
- [63] Marquart KF, Allam A, Janjuha S, Sintsova A, Villiger L, Frey N, et al. Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. *bioRxiv* 2020.2020.2007.2005.186544.
- [64] Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019;51:584–91.
- [65] Meier JA, Zhang F, Sanjana NE. GUIDES: sgRNA design for loss-of-function screens. *Nat Methods* 2017;14:831–2.
- [66] Miao Z, Humphreys BD, McMahon AP, Kim J. Multi-omics integration in the age of million single-cell data. *Nat Rev Nephrol* 2021;17:710–24.
- [67] Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat Genet* 2020;52:242–3.
- [68] Moreno-Mateos MA, Vejnár CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods* 2015;12:982–8.
- [69] Nishimatsu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 2018;361:1259–62.
- [70] Olivieri M, Cho T, Alvarez-Quilon A, Li K, Schellenberg MJ, Zimmermann M, et al. A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* 2020;182(481–496):e421.
- [71] Paquet D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* 2016;533:125–9.
- [72] Park SJ, Jeong TY, Shin SK, Yoon DE, Lim SY, Kim SP, et al. Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. *Genome Biol* 2021;22:170.
- [73] Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* 2015;162:675–86.
- [74] Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 2013;10:973–6.
- [75] Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013;152:1173–83.
- [76] Sanjana NE. Genome-scale CRISPR pooled screens. *Anal Biochem* 2017;532:95–9.
- [77] Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* 2018;9:5416.
- [78] Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014;343:84–7.
- [79] Sharon DM, Nesdoly S, Yang HJ, Gelinas JF, Xia Y, Ansoorge S, et al. A pooled genome-wide screening strategy to identify and rank influenza host restriction factors in cell-based vaccine production platforms. *Sci Rep* 2020;10:12166.
- [80] Shen MW, Arbab M, Hsu JY, Worstell D, Culbertson SJ, Krabbe O, et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 2018;563:646–51.
- [81] Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999;9:677–9.
- [82] Shivram H, Cress BF, Knott GJ, Doudna JA. Controlling and enhancing CRISPR systems. *Nat Chem Biol* 2021;17:10–9.
- [83] Song M, Kim HK, Lee S, Kim Y, Seo SY, Park J, et al. Sequence-specific prediction of the efficiencies of adenine and cytosine base editors. *Nat Biotechnol* 2020;38:1037–43.
- [84] Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet* 2019;20:467–84.
- [85] Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 2020;368:290–6.
- [86] Wang D, Zhang F, Gao G. CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. *Cell* 2020;181:136–50.
- [87] Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 2014;343:80–4.
- [88] Wei J, Alfajaro MM, DeWeirdt PC, Hanna RE, Lu-Culligan WJ, Cai WL, et al. Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection. *Cell* 2021;184(76–91):e13.

- [89] Xie S, Cooley A, Armendariz D, Zhou P, Hon GC. Frequent sgRNA-barcode recombination in single-cell perturbation assays. *PLoS ONE* 2018;13:e0198635.
- [90] Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol Cell* 2017;66(285–299):e285.
- [91] Yaguchi K, Yamamoto T, Matsui R, Tsukada Y, Shibamura A, Kamimura K, et al. Uncoordinated centrosome cycle underlies the instability of non-diploid somatic cells in mammals. *J Cell Biol* 2018;217:2463–83.
- [92] Yan X, Stuurman N, Ribeiro SA, Tanenbaum ME, Horlbeck MA, Liem CR, et al. High-content imaging-based pooled CRISPR screens in mammalian cells. *J Cell Biol* 2021;220.
- [93] Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum Mol Genet* 2014;23:R40–46.
- [94] Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;8:14049.
- [95] Zhou C, Sun Y, Yan R, Liu Y, Zuo E, Gu C, et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* 2019;571:275–8.
- [96] Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 2014;509:487–91.
- [97] Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell* 2017;65(631–643):e634.
- [98] Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* 2019;364:289–92.