

## REVIEW ARTICLE OPEN



# Recent advances in CRISPR-based functional genomics for the study of disease-associated genetic variants

Heon Seok Kim <sup>1,2,3</sup>, Jiyeon Kweon <sup>4</sup> and Yongsub Kim <sup>4,5</sup>✉

© The Author(s) 2024

Advances in sequencing technology have greatly increased our ability to gather genomic data, yet understanding the impact of genetic mutations, particularly variants of uncertain significance (VUSs), remains a challenge in precision medicine. The CRISPR–Cas system has emerged as a pivotal tool for genome engineering, enabling the precise incorporation of specific genetic variations, including VUSs, into DNA to facilitate their functional characterization. Additionally, the integration of CRISPR–Cas technology with sequencing tools allows the high-throughput evaluation of mutations, transforming uncertain genetic data into actionable insights. This allows researchers to comprehensively study the functional consequences of point mutations, paving the way for enhanced understanding and increasing application to precision medicine. This review summarizes the current genome editing tools utilizing CRISPR–Cas systems and their combination with sequencing tools for functional genomics, with a focus on point mutations.

*Experimental & Molecular Medicine* (2024) 56:861–869; <https://doi.org/10.1038/s12276-024-01212-3>

## INTRODUCTION

Many human diseases are attributed to genetic mutations, but comprehending how these mutations impact cellular phenotypes is limited by limited knowledge. Consequently, a significant number of mutations are classified as variants of unknown significance (VUSs), highlighting the crucial importance of understanding them for successful translational precision medicine<sup>1,2</sup>. Previously, researchers have relied on naturally occurring mutations found in existing biological samples, often through the use of genome-wide association studies (GWASs), to study their effects on phenotypes<sup>3,4</sup>. However, this approach was restricted to mutations present in specific samples. Genome editing methods, such as the CRISPR–Cas system, facilitate the study of genetic variants associated with diseases. These methods are applicable to both protein-coding and noncoding regions of the genome, offering a comprehensive approach to understanding the genetic influences on disease<sup>5</sup>. Over time, various CRISPR–Cas system orthologs have been identified, enabling successful genome editing in a wide range of organisms<sup>6–14</sup>. One of the key advantages of CRISPR technology lies in the ease of designing and synthesizing multiple guide RNAs (gRNAs), facilitating its application in high-throughput assays<sup>15,16</sup>. CRISPR-based high-throughput screens allow the simultaneous analysis of the functions of numerous genetic mutations. Furthermore, the concurrent development of diverse sequencing technologies, such as Illumina, Oxford Nanopore Technology (ONT)<sup>17–21</sup>, Pacific Bioscience (PacBio)<sup>22</sup>, and single-cell sequencing<sup>23,24</sup>, has opened up new avenues for the high-throughput evaluation of mutations. The integration of these technologies with CRISPR-based approaches allows researchers to comprehensively study the

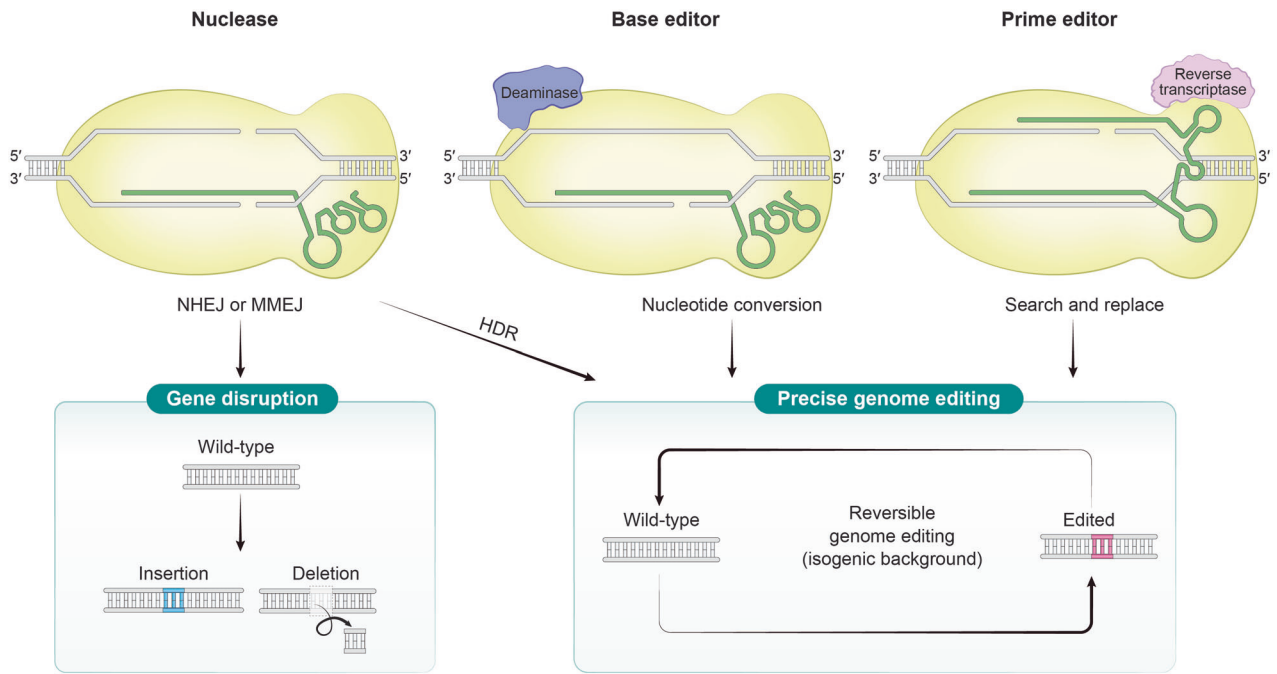
functional consequences of genetic mutations, paving the way for increased understanding and application of precision medicine.

Point mutations are the most common feature in human mutation databases, accounting for more than 50% of the mutations<sup>1,2</sup>. This review provides a summary of current genome editing tools utilizing the CRISPR–Cas system and their combination with sequencing tools for functional genomics, with a focus on disease-associated point mutations.

## Tools for precise genome editing for functional genomics research

The CRISPR–Cas system utilizes the Watson–Crick base pairing code to recognize specific DNA sequences. Its target specificity is determined by two factors: the presence of a protospacer adjacent motif (PAM) sequence in the target DNA and the presence of a protospacer sequence in the gRNAs. PAM sequences are crucial for allowing Cas proteins to identify their target DNA, and assorted Cas variants have been engineered to recognize diverse PAM sequences, expanding the versatility of the system<sup>25–29</sup>. By simply altering the protospacer sequences in gRNAs, the target site of CRISPR–Cas can be easily modified, making this system highly adaptable for functional genomics assessments. Furthermore, CRISPR–Cas9 techniques enable the generation of isogenic cell models that are genetically identical to wild-type cells except for the specific mutation of interest, providing valuable insights into the molecular mechanisms of genetic variations. Three primary classes of genome editing tools have been developed to date for precise functional genomics: nucleases, base editors, and prime editors.

<sup>1</sup>Department of Life Science, College of Natural Sciences, Hanyang University, Seoul, Republic of Korea. <sup>2</sup>Hanyang Institute of Bioscience and Biotechnology, Hanyang University, Seoul, Republic of Korea. <sup>3</sup>Hanyang Institute of Advanced BioConvergence, Hanyang University, Seongdong-gu, Seoul, Republic of Korea. <sup>4</sup>Department of Cell and Genetic Engineering, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. <sup>5</sup>Stem Cell Immunomodulation Research Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. ✉email: yongsub1.kim@gmail.com



**Fig. 1 A schematic overview of genome engineering strategies for the functional study of disease-associated genetic variants.** CRISPR–Cas toolkits are classified into three categories on the basis of the use of nucleases, base editors, and prime editors. Red arrows indicate DNA breaks induced by CRISPR–Cas toolkits. Nucleases can induce gene disruption by NHEJ-mediated mechanisms in gene knockout studies. Precise genome editing, which encompasses the introduction of disease-associated genetic variants and their correction, can be accomplished using nuclease-mediated HDR mechanisms, as well as through base editor-mediated and prime editor-mediated genome engineering.

**Nucleases.** Cas nucleases, exemplified by Cas9 and Cas12, are directed to specific DNA target sites in the genome through gRNAs and trigger DNA double-strand breaks (DSBs)<sup>30</sup>. Cellular DSB repair pathways, such as nonhomologous end joining (NHEJ) and homology-directed repair (HDR), then repair these breaks. In mammalian cells, nuclease-induced DSBs are predominantly repaired by the error-prone NHEJ pathway, leading to a mix of small insertion and deletion (indel) mutations at the target sites<sup>31</sup>. These indel mutations often cause frameshifts in coding sequences (CDSs), which disrupt genetic functions. Alternatively, DSBs can be accurately repaired by the error-free HDR pathway during specific cell cycle phases, namely, S and G2, allowing precise gene corrections with DNA donor templates<sup>32,33</sup>. While NHEJ-mediated gene disruption is highly efficient and is widely used to generate isogenic models that disrupt the gene of interest, facilitating comparisons between wild-type and knockout cells, HDR-mediated gene correction is commonly employed to understand the functional mechanisms of specific point mutations (Fig. 1). However, HDR has limitations for broader applications, including the need for additional DNA templates<sup>30</sup>, the potential for DSBs to induce undesired genomic alterations and activate the p53 response<sup>34–36</sup>, and variations in correction efficiency among different mammalian cell types<sup>37</sup>. Despite these challenges, the combination of Cas nucleases with multiple gRNAs enables multiplex gene knockout and the creation of large-scale structural variations in the genome, offering versatile tools for genomic manipulation and functional genomics research<sup>38–40</sup>.

**Base editors.** Base editors (BEs) are groundbreaking tools developed for precise nucleotide conversion, which is achieved by combining catalytically modified Cas proteins (such as Cas9-D10A nickases) with deaminases<sup>41</sup>. When a Cas protein and its gRNA recognize target DNA sequences, they form a single-stranded DNA (ssDNA) R-loop through gRNA hybridization with the target DNA strand. The deaminase domain then accesses the

ssDNA R-loop and induces nucleotide conversion without causing DSBs. Two primary types of BEs have been developed: cytosine base editors (CBEs), which convert C:G to T:A base pairs, and adenine base editors (ABEs), which convert A:T to G:C base pairs<sup>42–44</sup>. Recently, engineered BEs, such as C:G to G:C base editors (CGBEs) and A:T to C:G base editors (ACBEs), have been further developed by fusing additional DNA repair factors to conventional BEs, significantly expanding their capabilities and application scope<sup>45–48</sup>. BEs can introduce and correct point mutations, which account for the majority of human somatic mutations associated with genetic diseases<sup>49</sup> (Fig. 1). They can also be employed for targeted gene disruptions, e.g., the introduction of premature stop codons in coding sequences (CDS) or alterations in RNA splice-site motifs, all without causing DNA DSBs<sup>50–53</sup>. However, BEs have limited ability to induce specific types of nucleotide conversions and may also introduce undesired nucleotide changes within the base editing window<sup>41</sup>. Despite these challenges, BEs represent a promising tool for precise genome editing with significant potential for advancing our understanding and treatment of genetic diseases.

**Prime editors.** Prime editors (PEs) represent a significant advance in genome editing achieved by fusing a reverse transcriptase to catalytically impaired Cas9 proteins (Cas9-H840A nickases)<sup>54</sup>. These PEs are guided to target sites by a prime editing guide RNA (pegRNA), which contains template sequences for reverse transcription and protospacer sequences. The Cas9-H840A nickases cleave the target DNA strand, allowing the reverse transcriptase to synthesize a template DNA strand, thereby manipulating the target site<sup>54</sup>. This unique approach enables PEs to generate targeted genome modifications, including all types of substitutions and small indel mutations, to harness cellular DNA repair mechanisms (Fig. 1). Compared to Cas nucleases or base editors, PEs have a distinctive advantage: they can directly rewrite a target DNA without inducing DSBs or

requiring donor DNA. Consequently, PEs offer high editing purity and target specificity. Recent studies have further increased prime editing efficiency through the engineering of PEs and the optimization of pegRNAs to increase expression, nuclear localization, and degradation resistance<sup>55–60</sup>. PEs are remarkably versatile methods for precise genome editing, facilitating functional genomics investigations of nearly all types of genetic variation with exceptional specificity. Notably, while Cas nucleases and base editors induce mutations mainly within protospacer regions, PEs can introduce modifications both in the 3' regions of protospacers and within the protospacer sequences themselves. This unique feature enables PEs to correct multiple genetic variations, such as *KRAS* mutational hotspots, using a single pegRNA in a novel 'one-to-many' approach<sup>61</sup>. However, despite their distinct advantages, PEs currently exhibit lower efficiency than other genome editing technologies. Addressing this limitation will be crucial for the broader application of PEs in various fields and for unlocking their full potential in precision genome editing.

#### Advantages of CRISPR-based genome editing for the functional study of genetic variants

CRISPR-based genome editing is unique in its ability to precisely target and modify specific locations within the genome, surpassing traditional methods that rely on the integration of external cDNA encoding genetic variants. This precision facilitates the generation of isogenic disease models, allowing nuanced and accurate analysis of phenotypic changes resulting from specific genetic mutations (Fig. 2). This targeted approach reduces artifacts from gene overexpression and advances the understanding of intricate gene regulatory mechanisms influenced by cellular signaling pathways<sup>62</sup>. Additionally, CRISPR technology is applicable beyond protein-coding sequences, enabling research on noncoding regions such as splicing junctions, untranslated regions (UTRs), promoters, enhancers, and other regulatory elements, as well as mutations in cellular noncoding RNAs and microRNAs<sup>63–66</sup>. This broadens the scope of genetic research and opens up new avenues for therapeutic interventions targeting genetic disorders at their root.

#### Bulk CRISPR KO screens increase the throughput of functional genomics research

CRISPR-mediated multiplexed genome engineering has revolutionized gene function studies by enabling increased complexity and efficiency. In comparison to other genome engineering tools, such as zinc finger nucleases and TALENs, CRISPR offers distinct advantages for multiplexed engineering<sup>30</sup>. First, the design of gRNA, which determines the target gene, is remarkably straightforward. Second, gRNA molecules are compact in size, facilitating high-throughput synthesis. These features allow the implementation of high-throughput CRISPR screens. A widely utilized approach for conducting a CRISPR screen entails the delivery of a comprehensive genome-wide gRNA library into a large cell population via lentiviral vectors<sup>67–70</sup>. A distinct gRNA expression cassette is integrated into each cell to serve as both a barcode sequence and a specific knockout inducer. As a consequence, a pooled population of knockout cells is generated. These cells are subsequently exposed to diverse selective pressures, including cellular stressors, drugs, and toxins, which affect their overall fitness. After the selection process, the gRNA sequences within the cells are meticulously analyzed to identify genes that demonstrate growth advantages or disadvantages under the specific selection conditions. This detailed analysis enables the identification and characterization of genes that play a role in cellular responses and adaptations under the given selective pressures.

However, the scope of initial CRISPR screens is limited to phenotypes that exhibit a clear growth advantage or disadvantage. To overcome this limitation, researchers have incorporated

various assays, including image-based assays, to expand the range of phenotypes that can be assessed<sup>71,72</sup>.

#### Single-cell CRISPR KO screens enhance the granularity of functional genomics studies

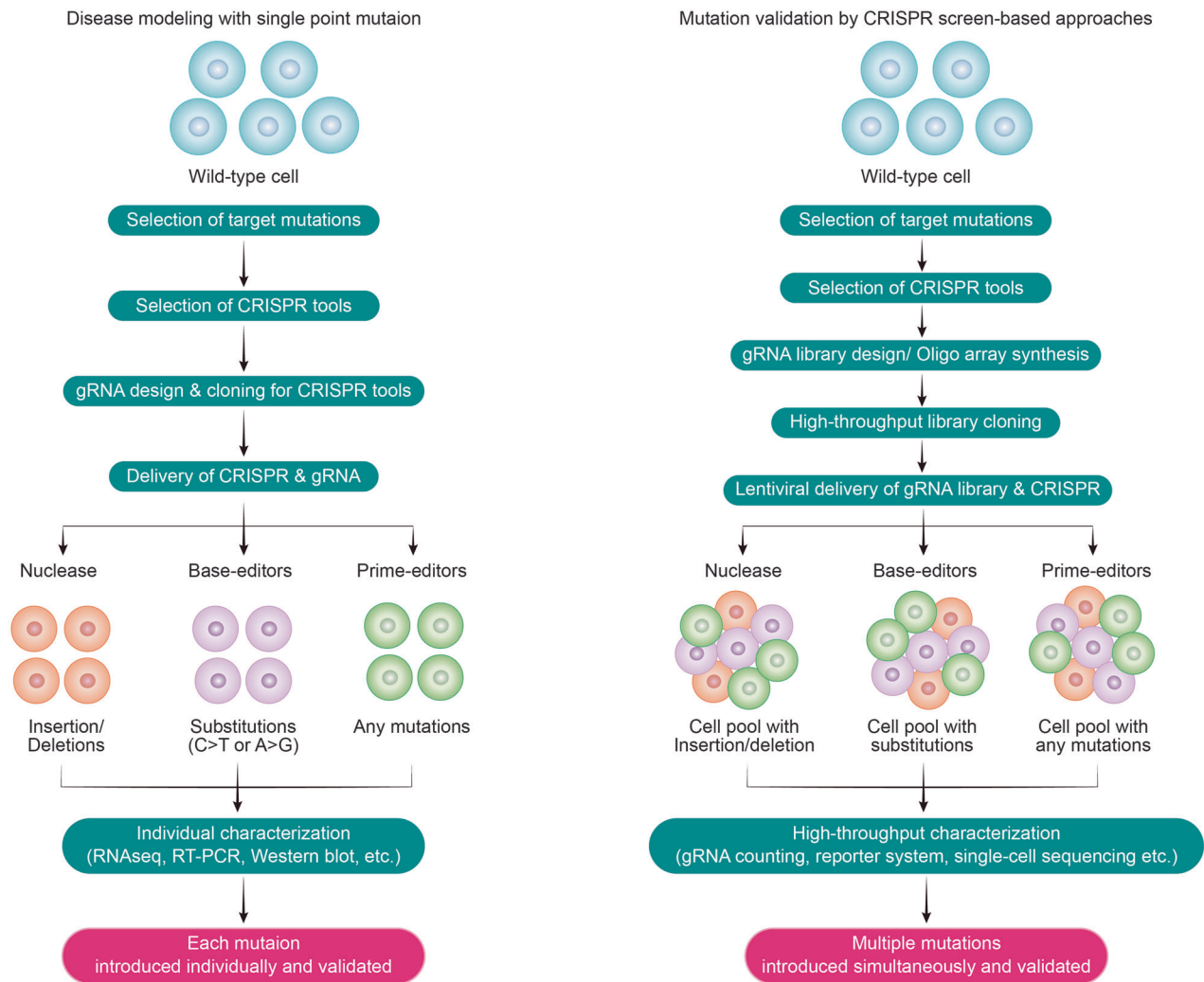
Single-cell sequencing offers significant advantages and synergies when combined with CRISPR KO screens<sup>73–75</sup>. First, this approach allows individual analysis of each knockout cell within a pooled population. This capability provides valuable insights into the specific effects of each KO event. Second, this approach enables comprehensive analysis of the entire transcriptome of each knockout cell, shedding light on the global changes in gene expression resulting from the KO event. Previously, obtaining transcriptome information from KO cells required isolating individual KO cells and performing bulk RNA sequencing. With single-cell sequencing, it becomes possible to perform the same analysis on a pooled population of cells. This development enables researchers to obtain transcriptome information from a vast number of individual cells simultaneously, providing a comprehensive view of gene expression patterns across the pooled population. Consequently, by applying single-cell sequencing to pooled populations of knockout cells, researchers can achieve high-throughput analysis and gain deeper insights into the outcomes of each KO event, even in the absence of selective pressures or screening conditions. This combined approach significantly increases the resolution and understanding of the impact of CRISPR-mediated knockout on cellular processes.

Single-cell CRISPR screens involve the sequencing of both the transcriptome and gRNA from individual cells using droplet-based single-cell sequencing platforms. Unlike most coding genes that have a poly-A tail, gRNA lacks this feature. To address this, researchers have employed specific gRNA-encoding lentiviral vectors or incorporated specific reverse transcription primers to facilitate single-cell cDNA generation<sup>20,75,76</sup>. Subsequently, these single-cell cDNAs are sequenced using general short-read sequencing platforms. By comparing groups of cells with different gRNAs, the gene expression phenotype resulting from each KO event can be thoroughly analyzed. This integrated approach enables comprehensive elucidation of the functional consequences of gene knockouts at single-cell resolution. The adoption of long-read sequencing has further facilitated the analysis of differential transcript isoform usage resulting from gene knockout<sup>20</sup>.

#### Bulk-level CRISPR screen for SNV characterization

Although Cas9 nuclease-based screens have significantly expanded the scalability of genetic studies, their applications have been focused on gene KO studies<sup>20,67–77</sup>. This limitation is noteworthy considering that more than half of human somatic mutations associated with genetic diseases are point mutations. Additionally, a considerable number of human single nucleotide variants (SNVs) have not been thoroughly studied and are classified as VUSs<sup>1,2</sup>. Therefore, it is crucial to identify the phenotypic effects of multiple human SNVs and address this gap. The development of strategies and technologies that enable the investigation of the functional consequences of SNVs will provide valuable insights into their effects on cellular processes and disease development (Fig. 2). By understanding the phenotypes associated with specific SNVs, researchers can better assess their clinical significance, inform diagnostics, and guide therapeutic interventions.

Prior to the advent of CRISPR base editors, Findlay et al.<sup>78</sup> developed a CRISPR-based saturation genome editing method for characterizing single nucleotide variants (SNVs) (Table 1). They employed Cas9 nuclease-mediated multiplex homology-directed repair (HDR) to generate all possible single-nucleotide variants (SNVs) at the target locus and then analyzed their functional impact. The Cas9-guide RNA complex introduces double-strand



**Fig. 2 Comparison between single and multiplexed SNV characterization.** This figure depicts a comparative analysis of two distinct approaches to SNV characterization. In the traditional method, individual clonal cells, each containing a specific SNV, are generated and separately characterized. In contrast, the high-throughput CRISPR screen-based multiplexed approach introduces multiple SNVs simultaneously and enables their collective analysis, greatly expanding the scale of SNV characterization.

breaks (DSBs) at the target site, and these DSBs are repaired through HDR using a complex library of donor templates containing all possible SNVs. This process generates a pooled library of cells harboring diverse SNVs, which can be utilized for functional screening. This pioneering method was successfully applied to the accurate classification of nearly 4000 *BRCA1* variants<sup>79</sup>. Radford et al.<sup>80</sup> applied this approach to characterize 12,776 *DDX3X* variants and identified 3432 functionally abnormal variants, demonstrating its potential for large-scale variant characterization (Table 1).

Base editor-based approaches enable more efficient and straightforward analysis of SNVs than HDR-based approaches<sup>79,80</sup>. Unlike HDR, which necessitates a repair template library in addition to a gRNA library, base editor-based methods streamline the analysis process. Hence, leveraging CRISPR base editors for SNV characterization will advance our comprehension of human genetic variation, offering insights into its implications for health and disease. A reliable gRNA design tool and efficient base editor constructs are indeed crucial for conducting high-throughput CRISPR base editor assays. In particular, the multiple-reporter-based gRNA efficiency prediction system has demonstrated its utility in increasing the success of CRISPR genome engineering<sup>81–87</sup>.

Additionally, base editors with higher editing efficiency and broader PAM usage are valuable<sup>28</sup>.

CRISPR base editors have been adapted for the characterization of multiple SNVs, particularly nonsense mutations that introduce premature termination codons (PTCs). Researchers have conducted CRISPR screening studies utilizing cytosine base editors, which enable C-to-T substitutions and can introduce PTCs. Genome-wide analyses have demonstrated that cytosine base editors with the NGG PAM can potentially introduce PTCs into approximately 17,000 human genes<sup>50,51</sup> (Table 1). While these studies have successfully introduced multiple expected PTCs at specific sites, it is important to note that the cellular consequences of these mutations are still similar to gene KO effects, similar to those in conventional CRISPR KO screens. The introduction of PTCs leads to the loss of functional protein products. While this approach provides valuable insights into the consequences of PTCs in specific genes, it is crucial to address a broader range of SNVs beyond nonsense mutations.

CRISPR base editor screens have been extended to target and investigate various missense mutations, which are commonly associated with cancer<sup>88–91</sup>. These studies have aimed to evaluate the functional consequences of multiple SNVs beyond nonsense mutations, particularly in cancer-related genes such as *BRCA1* and

**Table 1.** Comparison of multiplexed SNV characterization studies.

Investigated Genes	SNV introduction	Investigated SNV types	scRNAseq Method	Sequencing platform	SNV detection method	Analyzed features	Data analysis	Number of hits	Study
Genome-wide (~17,000 genes)	BE3	Premature termination codon	N/A	Illumina short-read	Indirect, sgRNA detection	Cellular fitness on 6-TG treatment	gRNA counting by custom script	1 gene	Kuscu et al. <sup>50</sup>
<i>BRCA1</i> (~4000 SNVs)	Cas9 mediated HDR	Point mutations	N/A	Illumina short-read	Direct, mRNA sequencing	Cellular fitness	Custom python script	400 mis-sense SNVs	Findlay et al. <sup>79</sup>
<i>DDX3X</i> (12,776 nucleotide variants)	Cas9 mediated HDR	Point mutations	N/A	Illumina short-read	Direct, gDNA sequencing	Cellular fitness	Custom R script	3432 functionally abnormal variants	Radford et al. <sup>80</sup>
<i>BRCA1</i> (745 gRNAs)	BE3	Point mutations	N/A	Illumina short-read	Indirect, sgRNA detection	Cellular fitness on olaparib	MAGECK	27 known mutations 4 VUSs	Kweon et al. <sup>88</sup>
86 DNA damage response genes (~37,000 gRNAs)	BE3	Point mutations	N/A	Illumina short-read	Indirect, sgRNA detection	Cellular fitness on DNA damage	MAGECK	~2000 gRNAs affecting DNA damage response	Cuella-Martin et al. <sup>89</sup>
3584 ClinVar variant genes (~68,526 gRNAs)	BE3.9max	Point mutations	N/A	Illumina short-read	Indirect, sgRNA detection	Cellular fitness on various condition	gRNA counting using PoolQ tool and custom statistics	75 hit genes in the cisplatin screens	Hanna et al. <sup>90</sup>
298 genes (~200,000 gRNAs)	Various CBEs	Point mutations	N/A	Illumina short-read	Indirect, surrogate target	Cellular fitness	Custom script	~5795 gRNAs affecting fitness	Sanchez-rivera et al. <sup>93</sup>
29,060 cancer-related transition mutations	AncBE4max, ABEmax	Point mutations	N/A	Illumina short-read	Indirect, surrogate target	Cellular fitness	Custom script	18 outgrowing 157 likely outgrowing mutations	Kim et al. <sup>92</sup>
<i>MAP2K1</i> , <i>KRAS</i> , <i>NRAS</i> (420 gRNAs)	BE3	Point mutations	CROP-seq	Illumina short-read	Indirect, sgRNA detection	Single-cell transcriptome	DROP-seq tools, CROP-seq software	Individual characterization of each mutations	Jun et al. <sup>95</sup>
<i>TP53</i> , <i>KRAS</i> (100 variants per gene)	ORF library	Point mutations	Perturb-seq	Illumina short-read	Indirect, ORF barcode	Single-cell transcriptome	Cellranger, Seurat, Custom scripts	Individual characterization of each mutations	Ursu et al. <sup>94</sup>
<i>TP53</i> , <i>SF3B1</i> (162 gRNAs)	AncBE4, ABEmax, CBE4max-SpG, ABEmax-SpG	Point mutations	Regular 5'	Illumina short-read, ONT long-read	Direct, full-length mRNA sequencing	Single-cell genotype Single-cell transcriptome	Cellranger, Seurat, Guppy, minimap, Custom scripts	Individual characterization of each mutations	Kim et al. <sup>21</sup>

*BRCA2* (Table 1). By utilizing CRISPR cytosine base editors, researchers have been able to introduce specific nucleotide changes corresponding to known missense mutations found in cancer patients. This enable the examination of the resulting cellular phenotypes and assessment of the functional impact of these mutations on cancer-related pathways and processes. These expanded CRISPR base editor screens offer valuable insights into the functional consequences of missense mutations that have been categorized as VUSs and their potential associations with cancer development and progression. By characterizing the effects of specific SNVs through CRISPR base editing, researchers can elucidate the impact of these mutations on cellular processes and pathways involved in cancer. Moreover, experimentally subjecting cells with specific SNVs to anticancer chemical treatments can illuminate how the presence of certain mutations influences drug sensitivity. This knowledge is crucial in personalized medicine, as it helps identify patient-specific mutations that may affect the efficacy of targeted therapies or other interventions.

The interpretation of data obtained from CRISPR base editor screens poses challenges compared to that obtained from conventional CRISPR KO screens owing to the differences between CRISPR nucleases and base editors. First, in CRISPR KO screens, multiple gRNAs can be designed to knock out a gene, allowing a more robust interpretation of the cellular consequences of gene KO by individual gRNAs. However, when a CRISPR base editor is used to introduce specific SNVs, the design process is more constrained. To introduce a desired SNV, precise targeting of a specific site is necessary, which limits the ability to design multiple gRNAs that induce the same SNV. As a result, the interpretation of the results from CRISPR-based base editing screens often relies on a limited number of gRNAs, typically only one. Second, base editors can introduce multiple SNVs for each gRNA. When the target sequence contains multiple target bases, different substitution patterns can produce distinct amino acid changes. Consequently, analyzing the gRNA sequence alone does not fully capture the resulting SNVs introduced by base editors. Third, the efficiency of gene KO achieved by base editors is generally lower than the efficiency of introducing SNVs. Therefore the presence of a gRNA in a cell does not guarantee the successful introduction of the intended SNVs.

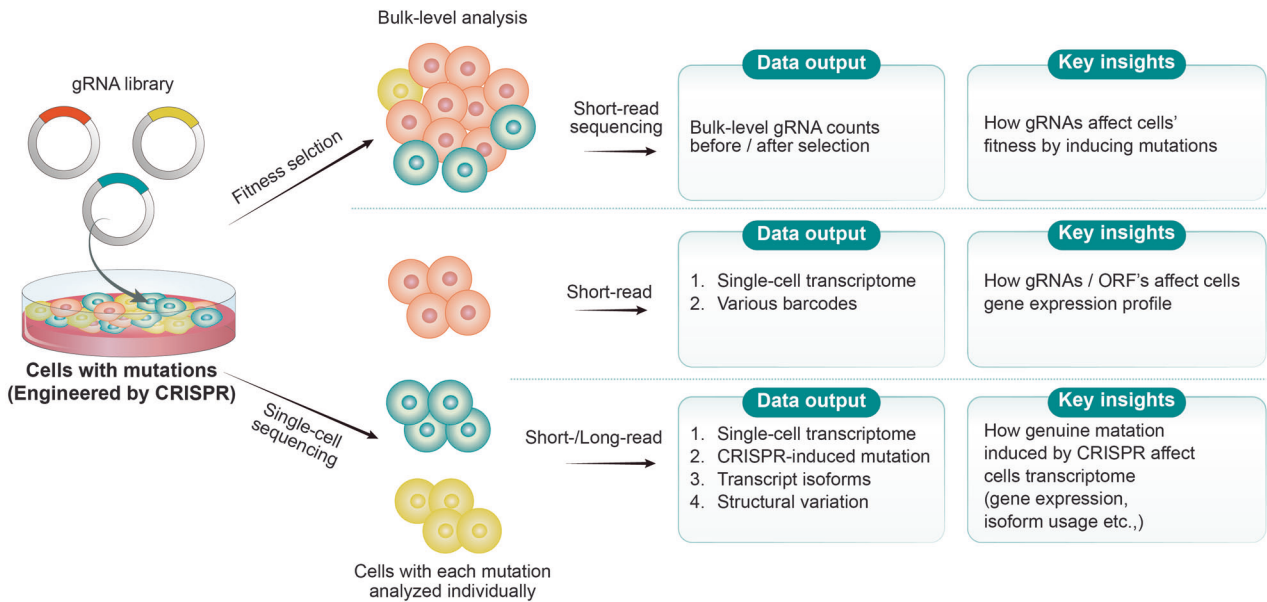
To overcome the challenges associated with interpreting CRISPR base editor screen data, two research groups recently introduced a reporter-assisted base editor screen method<sup>92,93</sup>. Instead of delivering only gRNAs to cells, they included both the gRNA and its corresponding target sites, which function as reporters of base editing events. By incorporating these target sites as reporters, the researchers were able to predict and quantify the base editing events that occurred at endogenous locations more precisely. This approach provides a more comprehensive assessment of editing efficiency and allows the precise identification and analysis of the introduced SNVs. For instance, Kim et al. identified 175 mutations in 160 genes as crucial drivers of cancer proliferation through a bulk screen using ABE and CBE<sup>92</sup> (Table 1). These findings underscore the importance of functional studies in uncovering specific genetic mutations linked to diseases. Despite the advances enabled by the reporter-assisted base editor screen method, there are still certain limitations to consider. First, it is important to note that the base editing events observed in the reporter may not always be concordant with the endogenous homologous sites, as these events are independent of each other. Consequently, the reporter system might not accurately predict heterozygous mutations or fully reflect the complexity of the editing outcomes. Second, similar to other bulk CRISPR screening approaches, the scope of these methods is generally limited to phenotypes that can be assessed through simple growth advantages or disadvantages. This means that more intricate phenotypic effects resulting from SNVs, such as subtle changes in gene regulation or complex

cellular responses, might not be captured by these screening methods alone.

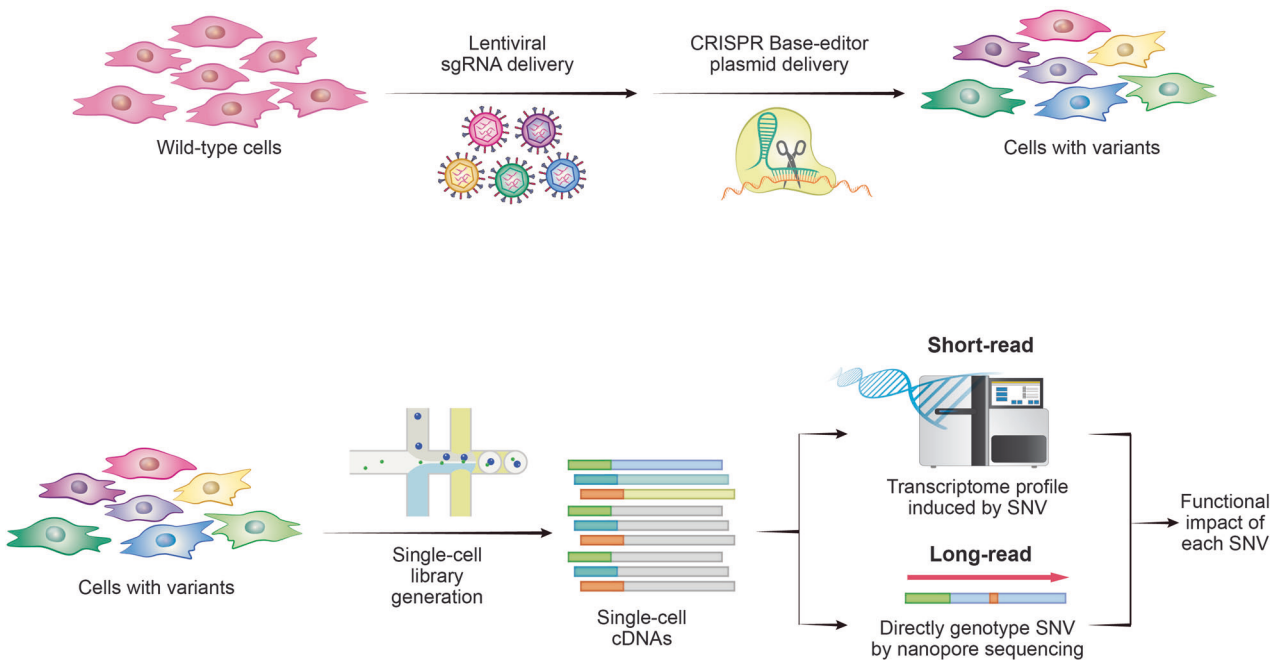
### Single-cell-level approaches for SNV characterization

Single-cell RNA sequencing is a powerful technique that provides a more comprehensive understanding of the complex cellular responses induced by SNVs. By employing single-cell RNA sequencing to analyze pooled cells with multiple SNVs, researchers can gain insights into the impact of each SNV on the cellular transcriptome (Fig. 3). Ursu et al. reported a Perturb-seq-based technique that involves the overexpression of an open reading frame (ORF) library containing various SNVs (e.g., 100 for *TP53* and *KRAS*) in individual cells<sup>94</sup> (Table 1). The authors evaluated the effects of these SNVs on the transcriptome using single-cell RNA sequencing. However, their study had several limitations. First, the overexpression of ORFs with SNVs differs from the introduction of endogenous mutations: these SNV-containing ORFs are not regulated by the endogenous promoter, and their expression levels can therefore differ from those of endogenous genes. This discrepancy might impact the interpretability of the results. Second, the endogenous wild-type genes continue to be expressed, potentially masking the effects of the SNVs. Finally, the barcode matching approach used in Perturb-seq, which employs DNA barcodes for each SNV, is known to be associated with a high frequency of barcode swapping, which can introduce errors during data analysis and lead to inaccurate interpretations. In a study conducted by Jun et al., a combination of base editor screening and single-cell RNA sequencing was employed<sup>95</sup> (Table 1). The researchers utilized a cytosine base editor along with 420 gRNAs to introduce multiple endogenous SNVs and then performed single-cell RNA sequencing using the CROP-seq method. However, this study had limitations in terms of accurately determining the genuine SNVs and their effects. Although the researchers successfully introduced endogenous SNVs, they were unable to directly identify the specific SNVs that were introduced. Instead, they relied on detecting the gRNA present in each single cell, which allowed them to make educated guesses regarding which codon might have been edited. Consequently, they could not precisely interrogate the effects of each SNV. The inability to directly identify SNVs hindered the ability to establish a direct link between the introduced SNVs and the observed transcriptional changes. As a result, the conclusions drawn from this approach might lack the specificity and accuracy necessary to understand the true impact of SNVs on cellular responses. Moreover, both short-read-based assays lacked direct detection of SNVs at the single-cell level due to their limited read length, which was not able to cover both the cell barcode and SNVs simultaneously.

The new technique called transcript-informed single-cell CRISPR sequencing (TISCC-seq) overcomes the limitations of short-read-based assays by the adaptation of long-read nanopore sequencing to single-cell base editor screens<sup>21</sup> (Table 1). TISCC-seq utilizes various CRISPR base editor series and gRNA libraries to introduce multiple SNVs (Fig. 4). To comprehensively analyze the cellular landscape, TISCC-seq employs both short-read and long-read sequencing platforms. Short-read single-cell RNA sequencing provides transcriptome profiles for each cell, following the conventional approach. TISCC-seq further incorporates single-cell long-read sequencing, which enables the direct detection of each SNV introduced by the CRISPR base editor. This direct detection eliminates the need to rely on guesses based on gRNA or barcodes, increasing the accuracy and specificity of SNV identification. By simultaneously capturing the transcriptomic profile and genotypic information from the same single cell, TISCC-seq facilitates the high-throughput evaluation of genuine endogenous SNVs. In this study, the authors successfully obtained complete transcriptome data from cells harboring 169 mutations. They systematically categorized these mutations into 5 with phenotypes resembling the wild-type phenotype and 69 that



**Fig. 3 Strategies for multiplexed SNV characterization.** This figure illustrates two main strategies for characterizing SNVs after high-throughput introduction. First, bulk analysis aided by gRNA tracking identifies SNVs with significant effects on cellular fitness. Second, single-cell analysis, including transcriptome analysis, provides deeper insights into the effects of individual SNVs, and the results can be extended by long-read sequencing to identify transcript isoforms and detect SNVs directly. These strategies enable comprehensive elucidation of SNV behavior, from overall fitness effects to precise gene expression patterns, thereby significantly advancing genetic research.



**Fig. 4 Schematic of TISCC-seq.** This figure illustrates the detailed procedure of TISCC-seq. This approach seamlessly captures transcriptome profiles and genotypic information to enable the high-throughput evaluation of endogenous SNVs, surpassing the constraints of short-read methodologies.

exhibited statistically significant functional alterations. This novel approach significantly expands the ability to study the functional impact of SNVs on cellular responses and provides a valuable tool for comprehensive single-cell analysis.

**CONCLUSION**

The exploration of human genetic mutations and their impact on cellular phenotypes, particularly the understanding of variants of

unknown significance (VUSs), is crucial for advancing translational precision medicine. The integration of CRISPR-Cas genome editing tools with advanced sequencing technologies represents a substantial advance that has significantly broadened our understanding of the genetic influences of diseases on both protein-coding and noncoding regions. The versatility of the CRISPR-Cas system, characterized by the ease of guide RNA (gRNA) design and its suitability for high-throughput assays, has revolutionized the functional study of genetic mutations. This system enables both

low-throughput assays for the analysis of individual variants and high-throughput assays for the massive parallel analysis of genetic variants. These developments have not only increased our ability to analyze numerous genetic mutations simultaneously but also facilitated a deeper understanding of their functional consequences. As we continue to elucidate the roles of these genetic factors, come closer to realizing the full potential of translational precision medicine to offer more personalized and effective treatment strategies for various human diseases. The development of CRISPR–Cas systems and their integration with state-of-the-art sequencing tools stand as a testament to the relentless pursuit of scientific innovation in understanding and combating genetic diseases.

## REFERENCES

- Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862–D868 (2016).
- Tate, J. G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941–D947 (2019).
- Visscher, P. M. et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet* **101**, 5–22 (2017).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet* **88**, 76–82 (2011).
- Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).
- Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **110**, 15644–15649 (2013).
- Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
- Hirano, H. et al. Structure and Engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961 (2016).
- Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500 (2017).
- Agudelo, D. et al. Versatile and robust genome editing with *Streptococcus thermophilus* CRISPR1-Cas9. *Genome Res* **30**, 107–117 (2020).
- Hu, Z. et al. A compact Cas9 ortholog from *Staphylococcus Auricularis* (SauriCas9) expands the DNA targeting scope. *PLoS Biol.* **18**, e3000686 (2020).
- Wu, Z. et al. Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat. Chem. Biol.* **17**, 1132–1138 (2021).
- Kim, D. Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **40**, 94–102 (2022).
- Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet* **16**, 299–311 (2015).
- Kweon, J. & Kim, Y. High-throughput genetic screens using CRISPR-Cas9 system. *Arch. Pharm. Res* **41**, 875–884 (2018).
- Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet* **19**, R227–R240 (2010).
- Deamer, D., Akesson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
- Singh, M. et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* **10**, 3120 (2019).
- Kim, H. S., Grimes, S. M., Hooker, A. C., Lau, B. T. & Ji, H. P. Single-cell characterization of CRISPR-modified transcript isoforms with nanopore sequencing. *Genome Biol.* **22**, 331 (2021).
- Kim H. S. et al. Direct measurement of engineered cancer mutations and their transcriptional phenotypes in single cells. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01949-8> (2023).
- Al'Khafaji A. M. et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01815-7> (2023).
- Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
- Macosko, E. Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
- Kleinstiver, B. P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
- Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
- Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
- Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
- Gao, L. et al. Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**, 789–792 (2017).
- Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet* **15**, 321–334 (2014).
- Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21**, 1468–1478 (2019).
- Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu Rev. Genet* **44**, 113–139 (2010).
- Moynahan, M. E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **11**, 196–207 (2010).
- Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
- Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med* **24**, 927–930 (2018).
- lhry, R. J. et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med* **24**, 939–946 (2018).
- Lieber, M. R. The mechanism of double-strand DNA break repair by the non-homologous DNA end-joining pathway. *Annu Rev. Biochem* **79**, 181–211 (2010).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Torres, R. et al. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* **5**, 3964 (2014).
- Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**, 423–427 (2014).
- Huang, T. P., Newby, G. A. & Liu, D. R. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* **16**, 1089–1128 (2021).
- 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).
- Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
- Gaudelli, N. M. et al. Programmable base editing of A\* to G\* in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
- Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).
- Zhao, D. et al. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2021).
- Tong, H. et al. Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nat. Biotechnol.* **41**, 1080–1084 (2023).
- Chen L. et al. Adenine transversion editors enable precise, efficient A\* to C\* G base editing in mammalian cells and embryos. *Nat Biotechnol.* <https://doi.org/10.1038/s41587-023-01821-9> (2023).
- Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet* **19**, 770–788 (2018).
- Kuscu, C. et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
- Billon, P. et al. CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Mol. Cell* **67**, 1068–1079 e1064 (2017).
- Gapinske, M. et al. CRISPR-SKIP: programmable gene splicing with single base editors. *Genome Biol.* **19**, 107 (2018).
- Yuan, J. et al. Genetic Modulation of RNA Splicing with a CRISPR-Guided Cytidine Deaminase. *Mol. Cell* **72**, 380–394 e387 (2018).
- Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
- Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652 e5629 (2021).
- Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).
- Ferreira da Silva, J. et al. Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nat. Commun.* **13**, 760 (2022).
- Liu, P. et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun.* **12**, 2121 (2021).



59. Velimirovic, M. et al. Peptide fusion improves prime editing efficiency. *Nat. Commun.* **13**, 3512 (2022).
60. Song, M. et al. Generation of a more efficient prime editor 2 by addition of the Rad51 DNA-binding domain. *Nat. Commun.* **12**, 5617 (2021).
61. Jang, G., Kweon, J. & Kim, Y. CRISPR prime editing for unconstrained correction of oncogenic KRAS variants. *Commun. Biol.* **6**, 681 (2023).
62. Gibson, T. J., Seiler, M. & Veitia, R. A. The transience of transient overexpression. *Nat. Methods* **10**, 715–721 (2013).
63. Liu, S. J. et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, aah7111 (2017).
64. Frangoul, H. et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
65. Korkmaz, G. et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
66. Chavez, M., Chen, X., Finn, P. B. & Qi, L. S. Advances in CRISPR therapeutics. *Nat. Rev. Nephrol.* **19**, 9–22 (2023).
67. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
68. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
69. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
70. Zhou, Y. et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
71. Kim, H. S. et al. Arrayed CRISPR screen with image-based assay reliably uncovers host genes required for coxsackievirus infection. *Genome Res* **28**, 859–868 (2018).
72. Wang, C., Lu, T., Emanuel, G., Babcock, H. P. & Zhuang, X. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc. Natl Acad. Sci. USA* **116**, 10842–10851 (2019).
73. Dixit, A. et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866.e1817 (2016).
74. Jaitin, D. A. et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883–1896.e1815 (2016).
75. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome read-out. *Nat. Methods* **14**, 297–301 (2017).
76. Replogle, J. M. et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).
77. Kim, H. S. et al. CRISPR/Cas9-mediated gene knockout screens and target identification via whole-genome sequencing uncover host genes required for picornavirus infection. *J. Biol. Chem.* **292**, 10664–10671 (2017).
78. Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123 (2014).
79. Findlay, G. M. et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
80. Radford, E. J. et al. Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation. *Nat. Commun.* **14**, 7702 (2023).
81. Kim, H. K. et al. SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. *Sci. Adv.* **5**, eaax9249 (2019).
82. Kim, H. K. et al. High-throughput analysis of the activities of xCas9, SpCas9-NG and SpCas9 at matched and mismatched target sequences in human cells. *Nat. Biomed. Eng.* **4**, 111–124 (2020).
83. Kim, H. K. et al. Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.* **39**, 198–206 (2021).
84. Kim, N. et al. Deep learning models to predict the editing efficiencies and outcomes of diverse base editors. *Nat. Biotechnol.* **42**, 484–497 (2024).
85. Kim, N. et al. Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat. Biotechnol.* **38**, 1328–1336 (2020).
86. Song, M. et al. Sequence-specific prediction of the efficiencies of adenine and cytosine base editors. *Nat. Biotechnol.* **38**, 1037–1043 (2020).
87. Yu, G. et al. Prediction of efficiencies for diverse prime editing systems in multiple cell types. *Cell* **186**, 2256–2272.e2223 (2023).
88. Kweon, J. et al. A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. *Oncogene* **39**, 30–35 (2020).
89. Cuella-Martin, R. et al. Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**, 1081–1097.e1019 (2021).
90. Hanna, R. E. et al. Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064–1080.e1020 (2021).
91. Huang, C., Li, G., Wu, J., Liang, J. & Wang, X. Identification of pathogenic variants in cancer genes using base editing screens with editing efficiency correction. *Genome Biol.* **22**, 80 (2021).
92. Kim, Y. et al. High-throughput functional evaluation of human cancer-associated mutations using base editors. *Nat. Biotechnol.* **40**, 874–884 (2022).
93. Sanchez-Rivera, F. J. et al. Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. *Nat. Biotechnol.* **40**, 862–873 (2022).
94. Ursu, O. et al. Massively parallel phenotyping of coding variants in cancer with Perturb-seq. *Nat. Biotechnol.* **40**, 896–905 (2022).
95. 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. *Commun. Biol.* **3**, 154 (2020).

## ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea [2021R1C1C1007162, 2018R1A5A2020732, RS-2023-00260462 to Y.K. and RS-2023-00243993, RS-2023-00261114 to H.S.K.] and by the Research Fund of Hanyang University (HY-202300000001145 to H.S.K.).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Yongsu Kim.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024