Nucleic Acids Research, 2024, **52**, W121–W125
https://doi.org/10.1093/nar/gkae304
Advance access publication date: 29 April 2024
Web Server issue

SynDesign: web-based prime editing guide RNA design and evaluation tool for saturation genome editing

Jinman Park[1,2], Goosang Yu[1,2], Sang-Yeon Seo[1,2], Jinyeong Yang[1,2] and Hyongbum Henry Kim[1,2,3,4,5,6,7,8,*]

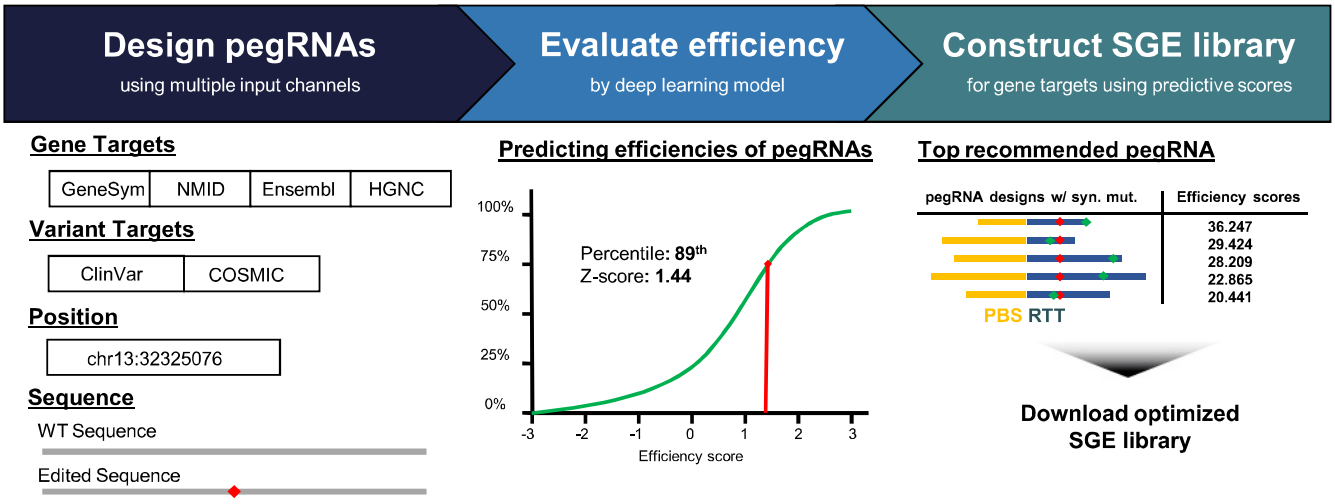
[1]Department of Pharmacology, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[2]Graduate School of Medical Science, Brain Korea 21 Plus Project for Medical Sciences, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[3]Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[4]Center for Nanomedicine, Institute for Basic Science (IBS), Seoul 03722, Republic of Korea
[5]Yonsei-IBS Institute, Yonsei University, Seoul 03722, Republic of Korea
[6]Institute for Immunology and Immunological Diseases, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[7]Woo Choo Lee Institute for Precision Drug Development, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[8]Won-Sang Lee Institute for Hearing Loss, Yonsei University College of Medicine, Seoul 03722, Republic of Korea

*To whom correspondence should be addressed. Tel: +82 2 2228 0879; Fax: +82 2 313 1894; Email: hkim1@yuhs.ac

Abstract

Saturation genome editing (SGE) enables in-depth functional evaluation of disease-associated genes and variants by generating all possible single nucleotide variants (SNVs) within a given coding region. Although prime editing can be employed for inducing these SNVs, designing efficient prime editing guide RNAs (pegRNAs) can be challenging and time-consuming. Here, we present SynDesign, an easy-to-use webtool for the design, evaluation, and construction precision pegRNA libraries for SGE with synonymous mutation markers. SynDesign offers a simple yet powerful interface that automates the generation of all feasible pegRNA designs for a target gene or variant of interest. The pegRNAs are selected using the state-of-the-art models to predict prime editing efficiencies for various prime editors and cell types. Top-scoring pegRNA designs are further enhanced using synonymous mutation markers which improve pegRNA efficiency by diffusing the cellular mismatch repair mechanism and serve as sequence markers for improved identification of intended edits following deep sequencing. SynDesign is expected to facilitate future research using SGE to investigate genes or variants of interest associated with human diseases. SynDesign is freely available at https://deepcrispr.info/SynDesign without a login process.

Graphical abstract



Introduction

When first introduced, prime editing presented a novel and powerful utility to genome editing by introducing prime editing guide RNAs (pegRNAs) that could be designed to install all possible point mutations a well as small insertions and deletions (indels), or any combinations of these DNA alter-

Received: February 4, 2024. **Revised:** March 22, 2024. **Editorial Decision:** April 2, 2024. **Accepted:** April 24, 2024

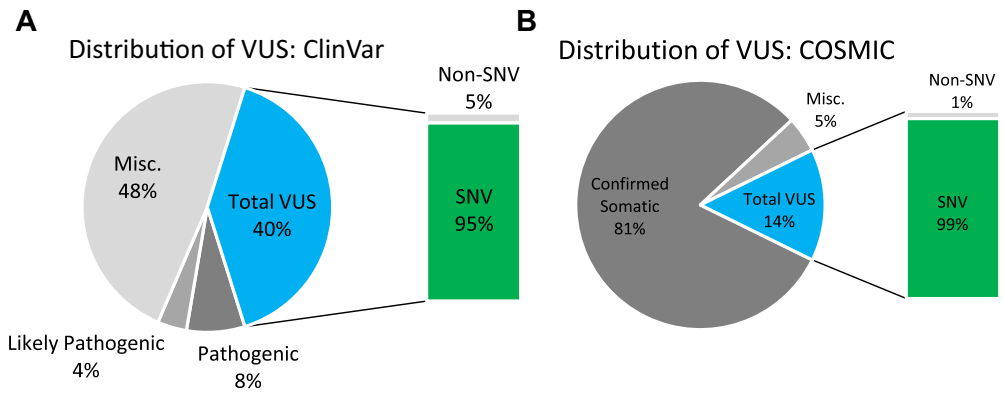


Figure 1. Distribution of VUSs in ClinVar and COSMIC databases. Of the variants categorized as uncertain significance or unknow origin, ClinVar and COSMIC databases respectively (VUS or simplicity), a major portion of them are single nucleotide variants and easily targetable by SynDesign. Investigating these VUSs remains an important task for understanding their role in human disease.

ations to a target DNA sequence (1). However, the design and evaluation of pegRNAs for efficient prime editing have been a challenge as various sequence and biochemical factors can attribute to efficient pegRNAs (1–3). Furthermore, cell conditions exhibiting differential levels of mismatch repair (MMR) systems can also play a role in determining editing efficiencies. Notably, introducing a synonymous mutation adjacent to the intended edit on the target DNA have been shown to improve on-target prime editing efficiency by diffusing the local cellular MMR activity (3). Additionally, the synonymous mutation can serve as an indicator of properly edited targets following deep sequencing (4). Taken together, the application of prime editing for saturation genome editing (SGE) can be a powerful tool in conducting systematic functional screening of variants for target genes of interest, especially for investigating variants of uncertain significance (VUS) that represent a major portion of the single nucleotide variants (SNVs) on the ClinVar and COSMIC databases, 95% and 99%, respectively (Figure 1). However, the designing and evaluation efficient pegRNAs for generating SNVs at every nucleotide position within a given coding sequence of a target gene present technical challenges that may hinder its wider application in research. Despite recent computational models based on deep learning greatly aiding our understanding of pegRNA designs and facilitating prime editing applications in diverse cell types and PEs (2,5,6), reliable and easy-to-use tools for SGE library design and evaluation have been limited.

Here, we present SynDesign, a web-based platform developed for the systematic design and assessment of pegRNAs, specifically tailored for the construction of precision pegRNA libraries employed in SGE. SynDesign provides an efficient and user-friendly means to easily target genes or variants of interest, streamlining the automation of pegRNA design and evaluation processes. Moreover, it seamlessly incorporates synonymous mutation markers into pegRNA designs, generating a curated list of top-performing pegRNAs for experimental applications. SynDesign is expected to alleviate the labor-intensive procedures associated with pegRNA design and facilitate the widespread adoption of SGE across diverse research applications.

Materials and methods

SynDesign webtool

SynDesign was developed using Python 3.9.13 on CodeIgniter 5 based on PHP 5.2.4. The web interface was developed using Bootstrap 5 and made accessible using Apache 2.4.18.

Input processing

To facilitate the systematic targeting of genes or variants of interest, the major databases for obtaining reference gene information on the human genome and genetic variants were collected, cleaned and indexed for fast and resource-efficient manipulation. Databases included the NCBI's RefSeq gene information (7), Ensembl database for vertebrate genomes (8), the HUGO Gene Nomenclature Committee, HGNC, database (9), where the gene name and symbol, with corresponding identification and accession numbers for representative gene forms were cross-referenced and verified using an in-house Python script. In doing so, a reliable reference list of >18 000 human genes and their collective IDs from each database was established. In addition, we collected and indexed the latest version of ClinVar (10) and COSMIC (11) variant archives for quickly identifying variant ID queries and generating a library of pegRNAs targeting the variant location for installing or editing the mutation.

PegRNA design and scoring

Genic information for target genes including transcription start and end positions, the coding sequence (CDS) start and end positions, and the exon count are retrieved using the in-house reference gene list. Using these coordinates, the CDS sequences are fetched from genome version GRCh38/hg38. Within the given CDS sequence, all possible pegRNA designs are determined for each position in the CDS sequence to elicit all three SNV alterations. Components such as the reverse transcription template (RTT) (max 40 bp) and primer-binding site (PBS) (max 17 bp) sequences are determined according to DeepPrime input parameters. Variant inputs from the ClinVar and COSMIC archives or single position input require considerably less computational resources as only a single position is targeted.

All pegRNAs are evaluated using DeepPrime and DeepPrime-FT and ranked according to predicted efficiency scores. DeepPrime and DeepPrime-FT are CNN + GRU-based deep learning model developed on pytorch 1.9.1 that predicts prime editing efficiencies for introducing 1–3-bp substitutions, insertions, or deletions at any given target sites in the human genome (2). It was developed and fine-tuned using more than 300 000 pegRNAs for a diverse group of prime editors and cell types and validated against multiple independent datasets from previous studies (1,3,12,13).

ID	GenicPos	DNA Change	AA Change	Guide Sequence	PBS-RTT Sequence	PBS-RTT wSynMut Sequence	DeepPrime Score	Z-score	Score Percentile
5.7675174_0	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGAGTTGATTCCACACCCCCGCCC	CCCTGTGCAGCTGTCAGTTGATTCCACACCCCCGCCC	73.90944	4.162683113221741	100th
5.7675174_1	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGAGTTGATTCCACACCCCC	CCCTGTGCAGCTGTCAGTTGATTCCACACCCCC	73.90001	4.161958551374201	100th
5.7675174_2	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGAGTTGATTCCACACCCCCGCC	CCCTGTGCAGCTGTCAGTTGATTCCACACCCCCGCC	73.71944	4.14808430611332	100th
5.7675174_3	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGAGTTGATTCCACACCCCCGC	CCCTGTGCAGCTGTCAGTTGATTCCACACCCCCGC	73.48082	4.129749741101682	100th
5.7675174_4	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGAGTTGATTCCACACCCC	CCCTGTGCAGCTGTCAGTTGATTCCACACCCC	73.436615	4.126353213374168	100th
5.7675174_5	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCTGTGCAGCTGTGAGTTGATTCCACACCCCCGCCC	CCTGTGCAGCTGTCAGTTGATTCCACACCCCCGCCC	73.134186	4.103115831084733	100th
5.7675174_6	chr17:7675174	C/A	W146C	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGTGTTGATTCCACACCCCCGCCC	CCCTGTGCAGCTGTATGTTGATTCCACACCCCCGCCC	73.116936	4.101790413070942	100th
5.7675174_7	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCTGTGCAGCTGTGAGTTGATTCCACACCCCC	CCTGTGCAGCTGTCAGTTGATTCCACACCCCC	73.06306	4.097650806061609	100th
5.7675174_8	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCTGTGCAGCTGTGAGTTGATTCCACACCCCCGCC	CCTGTGCAGCTGTCAGTTGATTCCACACCCCCGCC	73.0022	4.092974577637302	100th
5.7675174_9	chr17:7675174	C/A	W146C	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGTGTTGATTCCACACCCCCGCC	CCCTGTGCAGCTGTATGTTGATTCCACACCCCCGCC	72.87422	4.083141128512377	100th
5.7675174_10	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCTGTGCAGCTGTGAGTTGATTCCACACCCCCGC	CCTGTGCAGCTGTCAGTTGATTCCACACCCCCGC	72.81052	4.078246686339712	100th
5.7675174_11	chr17:7675174	C/A	W146C	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGTGTTGATTCCACACCCCC	CCCTGTGCAGCTGTATGTTGATTCCACACCCCC	72.8	4.077438373440867	100th
5.7675174_12	chr17:7675174	C/T	W146_	GACCTGCCCTGTGCAGCTGTGG	CCTGTGCAGCTGTGAGTTGATTCCACACCCC	CCTGTGCAGCTGTCAGTTGATTCCACACCCC	72.46921	4.052021850265107	100th
5 7675174 13	h 17 7675174	C/A	W146C	GACCTGCCCTGTGCAGCTGTGG	CCCTGTGCAGCTGTGTGTTGATTCCACACCCCCGC	CCCTGTGCAGCTGTATGTTGATTCCACACCCCCGC	72 451	4 050622669857505	100th

Figure 2. Sample Result Page from SynDesign. Interactive results page from *TP53* exon 5 run shows pertinent information regarding the pegRNAs ranked according to predicted efficiency scores from DeepPrime. The *z*-score plot depicts the score of the selected pegRNA compared to those from model training under the same prime editor and cellular conditions. Top 200 pegRNAs are shown as default but a link to download the full results is available at the bottom of the page.

Output processing and visualization

As co-editing of the PAM sequence has been shown to increase editing efficiency (5), the incorporation of a synonymous mutation marker is prioritized to the PAM sequence but can be user-specified to other region of the pegRNA design. Visualization of the results include important features of the top pegRNAs ranked according to the predicted efficiency scores included genic coordinates, nucleotide and amino acid changes, and *Z*-score graph of the top pegRNA designs against those from model testing for the corresponding PE and cell type (Figure 2). This contextual information provides a more complete understanding of their efficiency in relation to a more general performance of pegRNAs under similar experimental conditions.

Results

Analysis: gene target

For GeneSym, NMID, Ensembl and HGNC, the target gene information is automatically filled within a dropdown menu that includes the corresponding target gene's gene symbol, chromosome number and strand, and the exon numbers that include the CDS for quick reference by the user. Exons can be targeted by the number with single exon being the best for current available resources. Recently, we used SynDesign to develop a saturation resistance profiling library of the *EGFR* gene against various the chemotherapy drugs, afatinib and osimertinib. A SGE library was designed and evaluated to profile over 2476 SNVs in the EGFR gene that included 95% (1726/1817) of the possible protein variants in the tyrosine kinase domain (exons 18–21) which represents an important functional hotspot for cancer-related mutations (4,14). SynDesign was able to compile a list of all feasible pegRNAs targeting every position in the *EGFR* gene to systematically evaluate this important oncogene. SynDesign determines the top pegRNAs for each position according to our DeepPrime/DeepPrime-FT prediction models and automatically incorporates a synonymous mutation marker that further enhances pegRNA efficiency and sequenced product identification.

Analysis: variant targets

For specific variant targets from ClinVar or COSMIC archives, the mutation ID from each database can be used without the lettering prefix fillers, VCV- and COSM-, respectively. The model or therapy setting switches the wild-type (unedited) and edited sequences in the input processing step to construct a library for disease modelling (installing the mutation) or disease therapy (correcting the mutation), respectively. There are more than 600K and 750K variants that are categorized as VUS on the ClinVar and COSMIC databases respectively. Of these VUSs, 95% and 99% are SNVs (Figure 1). SynDesign can quickly retrieve these VUSs to generate all possible pegRNA designs targeting the VUS and score each pegRNA using DeepPrime/DeepPrime-FT. This secondary function to SGE that retrieves target variants directly from the ClinVar and COSMIC archives can further enhance the systematic survey of VUSs associated with human disease such as cancers. SynDesign is expected to serve as a useful platform that can aid in addressing this critical need in the investigation of VUSs.

Limitations of SynDesign

The basic sequence and position inputs are intended for quick test runs as they do not involve gene annotation or additional annotation using the variant archives. The sequence input tests single unedited and edited sequences of 121-bp in length that follows the DeepPrime protocol, https://deepcrispr.info/DeepPrime. The computational requirements for conducting full gene or multiple exon runs have proven to be prohibitive, surpassing reasonable runtimes and exceeding the VRAM limits of our existing webserver infrastructure.

Discussion

This paper introduces a user-friendly platform designed for precision targeting of genes or variants of interest. The platform seamlessly integrates systematic pegRNA design with robust prediction models, DeepPrime and DeepPrime-FT. SynDesign generates results that include synonymous mutation markers and important features for the top predicted pegRNAs, facilitating precision SGE for target genes and constructing robust pegRNA libraries for variants of interest. Thorough documentation including all available parameters and sample inputs and expected interactive outcomes, are accessible through the Support section of the webpage.

Data availability

SynDesign is freely available at https://deepcrispr.info/SynDesign. Webtool version of the custom Python code for guide design and feature extraction is available on https://github.com/josephjinpark/SynDesign and https://doi.org/10.6084/m9.figshare.25565994.

Acknowledgements

We would like to thank Younghye Kim and Seonmi Park for assisting with the experiments..

Funding

National Research Foundation of Korea grant funded by the Korean government (MSIT) [2022R1A3B1078084 to H.H.K., 2018R1A5A2025079 to H.H.K.]; Bio and Medical Technology Development Program of the NRF funded by the Korean government (MSIT) [2022M3A9E4017127 to H.H.K., 2022M3A9F3017506 to H.H.K., RS-2023-00260968 to H.H.K.]; Yonsei Signature Research Cluster Program of 2023-22- 0012 (to H.H.K.); Brain Korea 21 FOUR Project for Medical Science (Yonsei University College of Medicine); SNUH Kun-hee Lee Child Cancer and Rare Disease Project, Republic of Korea [22B-000-0101 to H.H.K.]; Yonsei Fellow Program, funded by Lee Youn Jae. Funding for open access charge: National Research Foundation of Korea grant funded by the Korean government (MSIT) [2022R1A3B1078084 to H.H.K., 2018R1A5A2025079 to H.H.K.]; Bio and Medical Technology Development Program of the NRF funded by the Korean government (MSIT) [2022M3A9E4017127 to H.H.K., 2022M3A9F3017506 to H.H.K., RS-2023-00260968 to H.H.K.]; Yonsei Signature Research Cluster Program of 2023-22- 0012 (to H.H.K.); Brain Korea 21 FOUR Project for Medical Science (Yonsei University College of Medicine); SNUH Kun-hee Lee Child Cancer and Rare Disease Project, Republic of Korea [22B-000-0101 to H.H.K.]; Yonsei Fellow Program, funded by Lee Youn Jae.

Conflict of interest statement

Authors declare no conflicts of interest.

References

1. Anzalone,A.V., Randolph,P.B., Davis,J.R., Sousa,A.A., Koblan,L.W., Levy,J.M., Chen,P.J., Wilson,C., Newby,G.A., Raguram,A., *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.
2. Yu,G., Kim,H.K., Park,J., Kwak,H., Cheong,Y., Kim,D., Kim,J., Kim,J. and Kim,H.H. (2023) Prediction of efficiencies for diverse prime editing systems in multiple cell types. *Cell*, **186**, 2256–2272.
3. Chen,P.J., Hussmann,J.A., Yan,J., Knipping,F., Ravisankar,P., Chen,P.F., Chen,C., Nelson,J.W., Newby,G.A., Sahin,M., *et al.* (2021) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*, **184**, 5635–5652.
4. Kim,Y., Oh,H.-C., Lee,S. and Kim,H.H. (2023) Saturation resistance profiling of EGFR variants against tyrosine kinase inhibitors using prime editing. bioRxiv doi: https://doi.org/10.1101/2023.12.03.569825, 05 December 2023, preprint: not peer reviewed.
5. Kim,H.K., Yu,G., Park,J., Min,S., Lee,S., Yoon,S. and Kim,H.H. (2021) Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.*, **39**, 198–206.
6. Mathis,N., Allam,A., Kissling,L., Marquart,K.F., Schmidheini,L., Solari,C., Balazs,Z., Krauthammer,M. and Schwank,G. (2023) Predicting prime editing efficiency and product purity by deep learning. *Nat. Biotechnol.*, **41**, 1151–1159.
7. O'Leary,N.A., Wright,M.W., Brister,J.R., Ciufo,S., Haddad,D., McVeigh,R., Rajput,B., Robbertse,B., Smith-White,B., Ako-Adjei,D., *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
8. Harrison,P.W., Amode,M.R., Austine-Orimoloye,O., Azov,A.G., Barba,M., Barnes,I., Becker,A., Bennett,R., Berry,A., Bhai,J., *et al.* (2024) Ensembl 2024. *Nucleic Acids Res.*, **52**, D891–D899.
9. Seal,R.L., Braschi,B., Gray,K., Jones,T.E.M., Tweedie,S., Haim-Vilmovsky,L. and Bruford,E.A. (2023) Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res.*, **51**, D1003–D1009.
10. Landrum,M.J., Lee,J.M., Benson,M., Brown,G., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Hoover,J., *et al.* (2016)

ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*., **44**, D862–D868.
11. Sondka,Z., Dhir,N.B., Carvalho-Silva,D., Jupe,S.,,Madhumita, McLaren,K., Starkey,M., Ward,S., Wilding,J., Ahmed,M., *et al.* (2024) COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res*., **52**, D1210–D1217.
12. Jang,H., Jo,D.H., Cho,C.S., Shin,J.H., Seo,J.H., Yu,G., Gopalappa,R., Kim,D., Cho,S.R., Kim,J.H., *et al.* (2022) Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. *Nat. Biomed. Eng*., **6**, 181–194.
13. Erwood,S., Bily,T.M.I., Lequyer,J., Yan,J., Gulati,N., Brewer,R.A., Zhou,L., Pelletier,L., Ivakine,E.A. and Cohn,R.D. (2022) Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol*., **40**, 885–895.
14. Sharma,S.V., Bell,D.W., Settleman,J. and Haber,D.A. (2007) Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer*, **7**, 169–181.

Received: February 4, 2024. **Revised:** March 22, 2024. **Editorial Decision:** April 2, 2024. **Accepted:** April 24, 2024