nature portfolio

Corresponding author(s): Hyongbum Henry Kim

Last updated by author(s): Nov 19, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina Novaseq 6000 and Miniseq were used to collect targeted deep sequencing data.
Data analysis	Screening data were analyzed with in-house custom python scripts. For the deep learning model, the data preprocessing and training script is available on https://github.com/yumin-c/esm_ar. Pymol (version 2.5.5) was used to figure out structure.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Final classification and prime editing outcomes are available at Supplementary Table 2 and 3.

We have submitted the deep sequencing data from this study to the NCBI Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA1242881.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We designed pegRNA libraries to include up to three pegRNAs for all possible SNV in AR ligand binding domain from exon4 to exon8. We transduced cells such that each pegRNA was represented in about 10000 cells, on average in each replicate.
Data exclusions	To achieve a high frequency of prime editing, we calculated the DeepPrime-FT score (Fine-tuned model for PEmax with optimized scaffold in HEK293T cell) for pegRNAs designed to generate all possible 1-bp substitution in the targeted region. We excluded all pegRNAs with left homology arm (LHA) lengths shorter than 5bp and chose the top three pegRNA with the highest DeepPrime-FT score for each SNV. In addition, we introduced 1-bp additional silent mutations in codons different from those harboring the intended mutations in the LHA of the RT template of pegRNAs.
Replication	All high throughput experiments were performed in duplicate. In each replicate, cells were transduced with different batch of lentiviral libraries on different day. The adjusted LFCs were well correlated between replicates. For conventional evaluation of resistance, all experiments were performed in triplicate (Figure 5). All replication attempts yielded consistent results, confirming reproducibility of the findings
Randomization	Selection of candidate pegRNA was not random and details are described in the methods.
Blinding	Functional classification for pegRNAs and nucleotide, protein variants were performed without prior knowledge of the function of the variants.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	The source of the HEK293T cell line is American Type Culture Collection (ATCC). The source of the LNCaP cell line is Korean Cell Line Bank. The source of the PC-3 cell line is Korean Cell Line Bank. The source of the CWR-R1-AD1 cell line is Kerafast.
Authentication	HEK293T, LNCaP, PC-3, and CWR-R1-AD1 cells were authenticated by the supplier.
Mycoplasma contamination	All cells were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Live cells, mixed cell populations of mCherry and eGFP positive, negative cells.
Instrument	BD FACS Verse I (3-Laser, 8-color analyzer)
Software	FlowJo (version 10.7.1)
Cell population abundance	Over 30,000 single cells with normal shape were analyzed for the percentage of mCherry and eGFP positive cells.
Gating strategy	To gate single cells with normal shape, FSC-A and SSC-A (P1) were used to gate cells with normal shape. Then, SSC-W and SSC-H (P2) following FSC-W and FSC-H (P3) were used to gate single cells. An example for flow cytometry gating strategy is provided as Extended Data Fig. 3d and 8e.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.