nature research

Corresponding author(s): Kevin Tianmeng Zhao

Last updated by author(s): Mar 18, 2025

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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 Miseq and BGI G99 platforms were used to collect the amplicon deep sequencing data. RNA samples were sequenced using an MGI T7 (2 × 150 PE) platform. Cells were sorted on a FACSAria III (BD Biosciences) cytometer.

Data analysis

Amplicon sequencing data was analyzed using the published code as previously described and noted in the methods. Graphpad prism 9 was used to analyze the data.

Flow cytometry data was analyzed using the BD FACSDiva™ Software (Version 9.7).

Raw RNA reads were cleaned by fastp. The reads were mapped to the human reference genome (hg38) by Hisat2 software (Version 2.2.1). After removing duplication, variants were identified using strelka29 (Version 2.9.10).

Initial protein structure models of mini-Sdd9-BE4max and mini-Sdd9-QBEmax ternary complex were predicted using AlphaFold3, with the optimal model_0, selected as the starting structure for MD simulations. MD simulations were conducted using GROMACS (version 2023.1, CUDA) with the Amber99BSC1 force field and the TIP3P water model. In each simulation system, the complex was positioned at the center of a cubic box with a distance of 15 Å maintained between the surface and the edges of the box. Na^+ and Cl^- ions were added to the system to achieve charge neutrality with an ion concentration set at 0.15 M. During the simulation, the Particle Mesh Ewald (PME) method was employed to calculate electrostatic interactions and a cutoff distance of 14 Å was utilized for analysis of short-range electrostatic and van der Waals forces. The LINCS algorithm was applied to constrain bonds involving hydrogen atoms, and periodic boundary conditions (PBC) were specified in all three dimensions. To optimize the system, energy minimization was performed using the steepest descent method, continuing until a maximum of 2500 steps was reached or the maximum force 1000 $kJ{\cdot}mol^{-1}{\cdot}nm^{-1}$. Next, 100 ps of NVT equilibration and 100 ps of NPT equilibration were conducted. During the equilibration process, the V-rescale temperature and Parrinell-Rahman pressure coupling methods were employed, with the system's temperature and pressure set to 310 K and 1.0 bar, respectively. After system equilibration, a 300 ns MD simulation was performed.

RMSD and RMSF data were generated using the gmx rmsd and gmx rmsf commands in GROMACS (version 2023.1, CUDA). The SASA data

were obtained through the gmx sasa command, with a probe diameter of 1.0 nm, and snapshots of the SASA state were collected every 1 ns within the stable region. All data were statistically analyzed and visualized using R packages ggplot2 (version 3.5.1) and tidyverse (version 2.0.0). The free energy landscape data were generated using the gmx sham command with the parameter -tsham set to 310 K. The data were plotted along the two projection components of RMSD and Rg. Visualization of this data was performed using the matplotlib (version 3.8.2) package in Python.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

The human reference genome hg38 was used for data analysis in this study. The authors declare that all data supporting the findings of this study are available in the article or its Supplementary Information files. The deep amplicon sequencing data is deposited in NCBI under BioProject Accession Code PRJNA1147008.

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	The experiments in HEK293T, A549, HeLa , HCT116, and Jurkat cells were performed with at least three independent biological replicates for all data with statistical analyses. Cell transfections were performed as described in the methods.
Data exclusions	No data exclusion.
Replication	All attempts for replication were successful. For the experiments in HEK293T cells, a minimum of three independent experiments were included.
Randomization	All cells were separately passaged and transfected to ensure biological randomization.
Blinding	Not applicable. As samples were processed identically through standard and in some cases automated procedures (DNA sequencing, transfection, DNA isolation) that should not bias outcomes.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The HEK293T, A549, HeLa, and HCT116 cells were obtained from ATCC. The Jurkat, Clone E6-1 cells were kindly provided by Cell Bank, Chinese Academy of Sciences.

Authentication	The cells with validated by the company or academy and authenticated by STR profiling.
Mycoplasma contamination	Validated by the company and tested in our studies to ensure no mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

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	All cells were cultured and transfected as described in the methods.
Instrument	Cells were sorted on a FACSAria III (BD Biosciences) cytometer.
Software	BD FACSDiva™ Software (Version 9.7)
Cell population abundance	For A549, HeLa and HCT116 cells, 50k, 200k or 400k cells were seeded and cultured for two days. For Jurkat cells, 400k cells were used for electroporation and cultured for two days.
Gating strategy	Negative control (untreated) and fluorophore-positive cells were used to establish gates for each cell type. Gates were drawn to collect cells expressing either fluorophore.

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