nature portfolio

Corresponding author(s): Dacheng Ma, Yong Yi, Ying Cai, Chunhua Shen

Last updated by author(s): Nov 3, 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<br>*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<br>*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 | We obtained the translated CDS fasta file of the reference genome from the NCBI genome dataset (https://www.ncbi.nlm.nih.gov/datasets/genome/). The operons were predicted by Operon Mapper (https://biocomputo.ibt.unam.mx/operon_mapper/). For each phylum of Archaea and Bacteria, a representative sequence was selected from the GTDB database (https://gtdb.ecogenomic.org/tree). We utilized ColabFold v1.5.5 (https://github.com/YoshitakaMo/localcolabfold) to predict protein complexes, employing databases including UniRef30 (uniref30_2302), ColabFoldDB (colabfold_envdb_202108), and the template database (pdb100_230517) for the inference process. ColabFold operates through two stages including Multiple Sequence Alignment (MSA) generation and protein structure prediction. BIO-RAD ChemiDoc XRS+ was used to western blot data collection. Leica TCS SP5II was used to immunofluorescence image data collection. |
| Data analysis | The code for this manuscript is provided in github repository: https://github.com/wensm77/Protein-Complex-Atlas and and , and on Zenodo: https://doi.org/10.5281/zenodo.18630539.<br>We visualized all the 3D structures of proteins using ChimeraX (https://www.cgl.ucsf.edu/chimerax/).<br>We utilized Operon Mapper (https://biocomputo.ibt.unam.mx/operon_mapper/) to predict the operons of 36 pathogenic bacteria, as a step in the pipeline illustrated in Fig. 1a.<br>We conducted phylogenetic tree analysis using OrthoFinder 2.5.5 (https://github.com/davidemms/OrthoFinder) on the genomic proteins of 36 pathogenic bacteria, 21 representative species of Archaea, and 167 representative species of Bacteria. We visualized the resulting phylogenetic trees in Fig. 1g and Supplementary Fig.7 with iTOL v6 (https://itol.embl.de/).<br>We constructed an atlas comprising 1.1 million protein-protein interaction structures using ColabFold v1.5.5 (https://github.com/YoshitakaMo/localcolabfold) and several databases, including UniRef30 (uniref30_2302), ColabFoldDB (colabfold_envdb_202108), and the template database (pdb100_230517).<br>We utilized MMseqs2 (https://github.com/soedinglab/MMseqs2) version 15.6f452 to perform homologous sequence alignments against the NCBI database. |

We employed ProteinMPNN (https://github.com/dauparas/ProteinMPNN) was fine-tuned on high-quality predicted heterologous complexes and MPBind(https://github.com/jianlin-cheng/MPBind/).
To assemble large protein complexes containing homo-oligomers, we used QSproteome (https://github.com/HugoSchweke/QSproteome_protocol) to truncate the flexible regions of homodimers, and then input it in AlphaFold2-bigbang to predict the full-length symmetric complex structure. Finally, we used CombFold (https://github.com/dina-lab3D/CombFold) as instructed to assemble them into a complete homo-oligomer containing a large protein complex.
For protein structure comparisons, we employed the Foldseek (https://github.com/steineggerlab/foldseek) tools, which use structural-alignment-based clustering algorithms. Specifically, we compared protein monomers and complexes against the AlphaFold Protein Structure Database (AFDB50) to identify protein fusion and fission events. Additionally, we utilized Foldseek to analyze our own protein complex dataset.

For AlphaFold3 comparison, we employed AlphaFold3 server (http://alphafoldserver.com/). Protenix (https://github.com/bytedance/Protenix), Chai-1 (https://github.com/chaidiscovery/chai-lab), Boltz-2(https://github.com/jwohlwend/boltz).

GraphPad 8.0 was used to do the graph figures and statistics.
LAS X (v3.3.0) was used to analyze fluorescent images.
Image Lab Software 5.0. software to analyze western blot data.
Image J. Fiji software was used to determine the PRKCZ co-localization with NPM1.

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

All the 1.1 million predicted protein structures are open for download in https://www.modelscope.cn/collections/protein_complex_atlas-2ae5e7d4f4a343. We have also launched a companion website showcasing the curated subset of PPI structures (https://www.biopredictnavigator.cn). Source data are provided in the source data file.

## 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 | No related with this. |
| Reporting on race, ethnicity, or other socially relevant groupings | No related with this. |
| Population characteristics | No related with this. |
| Recruitment | No related with this. |
| Ethics oversight | No related with this. |

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 | The experimental sample size was set as three biological repeats. |
| Data exclusions | No Data exclusion was used. |
| Replication | Three biological repeats (not technical repeats) were used. |

| | |
|---|---|
| Randomization | Allocation was randomly. |
| Blinding | Research was performed blindingly. |

# 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 |
|---|---|
| [ ] | [x] Antibodies |
| [ ] | [x] Eukaryotic cell lines |
| [x] | [ ] Palaeontology and archaeology |
| [x] | [ ] Animals and other organisms |
| [x] | [ ] Clinical data |
| [x] | [ ] Dual use research of concern |
| [x] | [ ] Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| [x] | [ ] ChIP-seq |
| [x] | [ ] Flow cytometry |
| [x] | [ ] MRI-based neuroimaging |

## Antibodies

Antibodies used

For western blot, antibody for GAPDH (AB0037,1:5000) was purchased from Abways (Shanghai, China). Antibody for SP1 (SC-59,1:1000) were purchased from Santa Cruz (USA). Antibody for PKCζ (340815, 1:1000) was purchased from Zenbio (Chengdu, China). Antibody for KPNA2 (ET1705-61, 1:1000) was purchased from Huabio (Hangzhou, China). Antibody for NPM1 (MA5-12508, 1:1000) was purchased from ThermoFisher Scientific (Wilmington, DE, USA). Antibodies for HA (3724, 1:1000), rabbit IgG (2729, 1:1000), and Flag (14793, 1:1000) were purchased from Cell Signaling Technology (Danvers, MA, USA).

For immunofluorescence staining, the antibody for PKCζ (1:100, 340815) was purchased from Zenbio (Chengdu, China). Antibody for NPM1 (MA5-12508, 1:200) was purchased from ThermoFisher Scientific (Wilmington, DE, USA).

Validation

anti-GAPDH rabbit monoclonal antibody, Abways cat.no.# AB0037,WB 1:5000.
http://www.abways.com/showproduct.asp?cid=AB0037.
Validated by the company and following publication: Zhang F, Tang H, Jiang Y, Mao Z. The transcription factor GATA3 is required for homologous recombination repair by regulating CtIP expression. Oncogene. 2017 Sep 7;36(36):5168-5176. doi: 10.1038/onc.2017.127. Epub 2017 May 8. PMID: 28481869.

anti-SP1 rabbit polyclonal antibody, Santa Cruz cat.no.# sc-59, WB 1:1000.
https://www.scbt.com/zh/p/sp1-antibody-pep-2.
Validated by the company and following publication: Meng Xia, Juan Liu, Xiaohui Wu, Shuxun Liu, Gang Li, Chaofeng Han, Lijun Song, Zhiqing Li, Qingqing Wang, Jianli Wang, Tian Xu, Xuetao Cao. Immunity. 2013 Sep 19;39(3):470-81. doi: 10.1016/j.immuni.2013.08.016. Epub 2013 Sep 5.

anti-PKCζ rabbit monoclonal antibody, Zenbio cat.no.# 340815,WB 1:5000, IF: 1:100.
http://www.zen-bio.cn/prod_view.aspx?TypeId=246&Id=681443&Fld=t3:246:3
Validated by the company.

anti-KPNA2 rabbit monoclonal antibody, Huabio cat.no.# ET1705-61 ,WB 1:1000.
http://www.huabio.cn/products/KPNA2-antibody-ET1705-61
Validated by the company.

anti-NPM1 rabbit monoclonal antibody, ThermoFisher Scientific cat.no.# MA5-12508, WB 1:1000, IF: 1:200
https://www.thermofisher.cn/cn/zh/antibody/product/NPM1-Antibody-clone-NA24-Monoclonal/MA5-12508
Validated by the company and following publication: Peng Wang, Fang Wu, Yupo Ma, Liang Li, Raymond Lai, Leah C Young. J Biol Chem. 2010 Jan 1;285(1):95-103. doi: 10.1074/jbc.M109.059758. Epub 2009 Nov 2.

anti-HA rabbit monoclonal antibody, Cell Signaling Technology. cat.no.#3724, WB 1:1000.
https://www.cellsignal.com/products/primary-antibodies/ha-tag-c29f4-rabbit-mab/3724
Validated by the company and following publication: Zhang L, Geng X, Wang F, Tang J, Ichida Y, Sharma A, Jin S, Chen M, Tang M, Pozo FM, Wang W, Wang J, Wozniak M, Guo X, Miyagi M, Jin F, Xu Y, Yao X, Zhang Y. 53BP1 regulates heterochromatin through liquid phase separation. Nat Commun. 2022 Jan 18;13(1):360. doi: 10.1038/s41467-022-28019-y.

anti-HA rabbit IgG, Cell Signaling Technology. cat.no.# 2729, WB 1:1000.
https://www.cellsignal.cn/products/primary-antibodies/normal-rabbit-igg/2729.
Validated by the company and following publication: Linyong Shen, Xue Bai, Liru Zhao, Jiamei Zhou, Cheng Chang, Xinquan Li, Zhiping Cao, Yumao Li, Peng Luan, Hui Li, Hui Zhang. Nat Commun. 2024 Oct 28;15(1):9274. doi: 10.1038/s41467-024-53692-6.

anti-Flag rabbit monoclonal antibody, Cell Signaling Technology. cat.no.# 14793, WB 1:1000.
https://www.cellsignal.cn/products/primary-antibodies/dykddddk-tag-d6w5b-rabbit-mab-binds-to-same-epitope-as-sigma-aldrich-anti-flag-m2-antibody/14793.
Validated by the company and following publication: Yingwei Ge , Lijie Zhou, Yesheng Fu, Lijuan He, Yi Chen, Dingchang Li, Yuping Xie, Jun Yang, Haitao Wu, Hongmiao Dai, Zhiqiang Peng, Yong Zhang, Shaoqiong Yi, Bo Wu, Xin Zhang, Yangjun Zhang, Wantao Ying, Chun-Ping Cui, Cui Hua Liu, Lingqiang Zhang.

## Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

| | |
|---|---|
| Cell line source(s) | H1299 (CRL-5803) and HEK293T (CRL-3216) were obtained from ATCC |
| Authentication | All cell lines used in this study were were kept at low passages to maintain their identity and were authenticated by morphology check and growth curve analysis. |
| Mycoplasma contamination | All cell lines used in this study were routinely tested to be negative for mycoplasma contamination |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used. |

## Plants

| | |
|---|---|
| Seed stocks | No related with this. |
| Novel plant genotypes | No related with this. |
| Authentication | No related with this. |