nature research

Corresponding author(s): Patrick Cramer, Nature Manuscript 2018-03-03844

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	FEI EPU 1.7.0, ProteomeDiscoverer 1.4, pLink (v. 1.23), MaxQuant86 (version 1.5.2.8)
Data analysis	RELION 2.1; COOT version 0.8.9; vmd Version 1.9.3; NAMD Version 2.12; ROBETTA (online server, version as of September 2017-March 2018); PHENIX 1.13; Phyre2 (online server, version as of September 2017-February 2018); GraphPad Prism version 6, Gctf, MotionCorr2, XiNet, XlinkAnalyzer version 1.1, Chimera version 1.12, PyMol (Schrödinger LLC versions 1.8.2.3 and 1.8.6.0) , ImageJ version 1.47v, Molprobibty plugin (Phenix), Psipred(online version September 2017-December 2017) , Sable (online version September 2017-December 2017), Jalview version 2.10.262, MAFFT (webserver, January -March 2018), , PDB2PQR (online version September 2017-February 2018), APBS pymol plugin, TPRpred (online version August 2017-December 2017)

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 upon request. 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 electron density reconstructions and final EC* model were deposited with the Electron Microscopy Data Base (EMDB) under accession codes EMD-0030 to EMD-0037, and with the Protein Data Bank (PDB) accession code 6GMH. The tSH2 domain model was deposited with the PDB accession code 6GME. Source data for Figures 1a and 5a, Extended Data Figs. 1a, e-k, m, 2a-j, i-e are found in Supplementary Figures 1, 2 and Supplementary Table 8.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. All biochemical experiments were replicated two or more times. Structural data was collected on three independently prepared samples.
Data exclusions	No data were excluded from the analyses.
Replication	All attempts at replication were successful.
Randomization	Samples were not allocated to groups.
Blinding	Investigators were not blinded during data acquisition and analysis because it is not a common procedure for the methods employed.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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

Antibodies

Antibodies used	RPB1 CTD anti phos ser2 3E10, RPB1 CTD anti phos ser5 3E8, RPB1 CTD anti phos ser7 4E12, RPB1 CTD anti phos tyr1 3D12, general RPB1 CTD antibodies F12 and MABI0601, anti mouse (Abcam ab5870), anti rat (Sigma-Aldrich A9037)
Validation	All antibodies were previously validated in the following publications or firms: 3E10, 3E8, 4E12 Chapman et al., Science 2007; 3D12 Mayer et al., Science 2012; MABI0601 Nojima 2015; F12 Santa Cruz Biotechnology; ab5870 AbCam; A9037 Sigma-Aldrich

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Hi5 cells: Expression Systems, Tni Insect Cells in ESF921 media, Item 94-002F Sf9 cells: ThermoFisher, Catalogue Number 12659017, Sf9 cells in Sf-900TM III SFM Sf21 cells: Expression Systems, Sf21 insect cells in ESF921 medium, Item 94-003F
Authentication	Provided by commercial supplier (ThermoFisher and Expression Systems)
Mycoplasma contamination	Mycoplasma test was not required for used cell lines.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.