nature portfolio

Corresponding author(s): Lauren L. Porter

Last updated by author(s): Jun 15, 2022

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	hmmer 2.3.1, blast 2.13.0+
Data analysis	jnet_src_v.2.3.1, mview 1.67, clustal omega 2.1, python3, Geneious Prime 2021.1, Taxonkit 0.12, GREMLIN webserver (http://gremlin.bakerlab.org), evcouplings webserver (https://v2.evcouplings.org), Robetta webserver (https://robetta.bakerlab.org), Phyre2 webserver (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index), AlphaFold 2.0, TALOS-N, NMRfam Sparky, NMRPipe, BestSel webserver (https://bestsel.elte.hu/index.php), custom code

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

Tabular data from all figures are provided in the Source Data File. Data recording all predictions are included in Supplementary Data 1 (bacteria and some archaea and eukaryotes) and Supplementary Data 2 (expanded predictions for archea and eukaryotes). Additional data are available at https://github.com/ncbi/sequence_space. Chemical shift assignments were deposited in the Biomolecular Magnetic Resonance Bank (BMRB) with the following accession codes: 51529 [https://dx.doi.org/10.13018/BMR51429] (Full-length Variant 5 (CTD only)), 51428 [https://dx.doi.org/10.13018/BMR51428] (Variant 5 isolated CTD), 51433

[https://dx.doi.org/10.13018/BMR51433] (Variant 8 CTD). PDB accession codes used in Fig. 1: 6ZTJ [http://doi.org/10.2210/pdb 6ZTJ/pdb] (E. coli 70S-RNAP expressome. Complex in NusG-coupled state (38 nt intervening mRNA), chain CF, 5OND [http://doi.org/10.2210/pdb5OND/pdb] (RfaH from Escherichia coli in complex with ops DNA), chain A, and 6C6S [http://doi.org/10.2210/pdb6C6S/pdb] (CryoEM structure of the E. coli RNA polymerase elongation complex bound with RfaH), chain D and Fig. 3: 5OND [http://doi.org/10.2210/pdb5OND/pdb] (RfaH from Escherichia coli in complex with ops DNA), chain A, 2LCL [http://doi.org/10.2210/pdb2LCL] (Solution Structure of RfaH carboxyterminal domain), chain A. Constructs for protein expression are available upon request.

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	15,516 NusG-like sequences were found after 3 rounds of BLAST on the UniProt Database. Because the
Data exclusions	No Data Excluded
Replication	Circular dichroism measurements were taken in replicates of 5-10 per protein sample. Identical buffer conditions were used for all samples. All replicates were consistent with one another. Two HSQCs of Variant 5 full-length, Variant 5 CTD and Variant 8 CTD were taken and found to be consistent with one another.
Randomization	This is not relevant to the study because we were not performing clinical trials or the affect of drugs/conditions/etc. on model organisms.
Blinding	This is not relevant to the study because we were not performing clinical trials or the affect of drugs/conditions/etc. on model organisms.

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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