nature portfolio

Corresponding author(s): Jill Moore

Last updated by author(s): Oct 10, 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	All protocols are described in the Methods and Supplementary Methods sections of the manuscript and available in GitHub.
Data analysis	All code and pipelines are open source and available via GitHub. Each repository has documentation on how to run code and pipelines. Datasets were processed using uniform ENCODE processing pipelines (https://www.encodeproject.org/pipelines/) and the V4 cCRE pipeline (https://github.com/weng-lab/ENCODE-cCREs/tree/master/Version-4). The following software versions were used for data analysis. BEDTools (v2.19.1) UCSC Tools (v350) DESeq2 (v1.46.0) FIMO (v5.1.1) tensorflow(2.17.0) keras(3.8.0) umap-learn(0.5.7)

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 cCRE annotations (cell type-agnostic and cell type-specific) are available on SCREEN (screen.wenglab.org). All other ENCODE data are available on the ENCODE portal (encodeproject.org) and are referenced by experiment and file accession. Supplementary Data are also available at https://users.moore-lab.org/ENCODE-cCREs/Supplementary-Data/.

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	Each biosample on the ENCODE portal (encodeproject.org) has annotated biological sex information. No sex-specific analyses were performed as part of this work.
Reporting on race, ethnicity, or other socially relevant groupings	Biosamples collected during ENCODE phase 4 have annotated ethnicity information on the ENCODE portal. No race or ethnicity-based analyses were performed as part of this work
Population characteristics	This study used publicly available human biospecimen data generated by the ENCODE consortium. Metadata describing relevant covariates—including donor age, sex, tissue or cell type, and disease status where applicable—are available for all human samples through the ENCODE Portal. These attributes were not directly modeled in all analyses but were incorporated in comparisons where biologically relevant, such as developmental (embryonic versus adult) contrasts and tissue-specific activity patterns. All metadata were curated and quality-controlled according to ENCODE standards and are provided in full through the portal links listed in the Data Availability section.
Recruitment	ENCODE samples were collected from a variety of sources including postmortem and post surgical tissue samples, primary blood cells from healthy donors and legacy cell lines. Details of biosample collection can be found on the ENCODE portal.
Ethics oversight	This study analyzed publicly available human data generated by the ENCODE consortium. All human datasets were collected by ENCODE production laboratories under protocols approved by their respective Institutional Review Boards (IRBs) and in accordance with informed consent and relevant ethical guidelines. No new human data were collected or generated specifically for this study, and all analyses were performed on de-identified datasets accessible through the ENCODE Portal.

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	No formal sample size calculation was performed. This study leveraged all available high-quality ENCODE datasets at the time of analysis, encompassing over 5,000 experiments across 1,679 biosamples representing diverse tissues, primary cells, and cell lines. These sample sizes were determined by the scope of ENCODE data production and are sufficient to provide robust statistical power for genome-wide analyses and cross-biosample comparisons. For the REST transgenic mouse reporter assays, the number of embryos per construct followed standard ENCODE validation protocols, which have been empirically demonstrated to yield reproducible results across multiple independent elements.
Data exclusions	Each ENCODE experiment is subject to assay specific quality control measurements which are available on the ENCODE portal. To create the Registry of cCREs we selected all released DNase experiments with SPOT score > 0.3 and transcription factor ChIP-seq datasets with FRiP > 0.003. To annotate cCREs, we selected one representative experiment per biosample to account for assay redundancy based on QC metrics.
Replication	The majority of ENCODE assays included in this study were performed in biological duplicate, as required by ENCODE production standards. In rare cases where biosample scarcity permitted only a single replicate, these instances are clearly annotated on the ENCODE Portal. All available replicates met ENCODE-defined quality control thresholds, and no failed experiments were included in this analysis. For the REST transgenic mouse reporter assays, each candidate element was tested in multiple independent embryos. A predicted element was scored positive as an enhancer only if at least three embryos exhibited identical β-galactosidase staining in the same tissue. All tested elements and

corresponding results are provided in Supplementary Table 7e and at https://enhancer.lbl.gov/

Randomization

Randomization was not relevant for this study. Most analyses were based on publicly available ENCODE consortium datasets that were uniformly processed and analyzed computationally. For the REST transgenic mouse reporter assays performed specifically for this study, embryos were not randomly assigned to treatment or control groups because all embryos were generated and assayed under identical experimental conditions to evaluate reproducible enhancer or silencer activity of the tested sequences. Thus, no random allocation was necessary.

Blinding

Blinding was not relevant for the computational components of this study, which relied on publicly available ENCODE datasets analyzed through automated pipelines. For the REST transgenic mouse reporter assays, embryos were processed and imaged under identical conditions, and reporter expression was scored without knowledge of which candidate elements were tested to minimize potential bias in interpretation.

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

Antibodies

Antibodies used

The > 3,000 antibodies that were used are listed on the ENCODE portal at https://www.encodeproject.org/search/?type=AntibodyLot&status=released. Each antibody page contains information about the supplier name, catalog number, clone name, lot number and dilution. Each experiment is linked with its corresponding antibody.

Validation

The > 3,000 antibodies that were used are listed on the ENCODE portal at https://www.encodeproject.org/search/?type=AntibodyLot&status=released. Each antibody page contains information about the antibody validation. Antibody characterization guidelines can be found here: https://www.encodeproject.org/documents/4bb40778-387a-47c4-ab24-cebe64ead5ae/@@download/attachment/ENCODE_Approved_Oct_2016_Histone_and_Chromatin_associated_Proteins_Antibody_Characterization_Guidelines.pdf

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

This study used data from 200 human and mouse cell lines. A complete list of all cell lines, including biosample accessions, is provided in Supplementary Table 1ab. Each experiment is linked to a biosample page on the ENCODE Portal, which includes donor, tissue, and derivation information . No new cell lines were generated or cultured for this study.

Authentication

No new cell lines were generated or cultured in this study. All cell lines were obtained by ENCODE data production labs, which performed standard quality control and authentication procedures prior to data release, including verification of species identity as part of ENCODE production guidelines. Information on the provenance and validation of each biosample is available on the corresponding ENCODE Portal biosample pages.

Mycoplasma contamination

Cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

Mice (Mus musculus) used for the transgenic mouse enhancer assays were housed at the LBNL Animal Care Facility, which is fully accredited by AAALAC International. FVB/NJ (Jackson Laboratory; Strain:001800) and CD-1 (Charles River Laboratories; Strain: 022)

mice were housed on a 12-hour light-dark cycle in standard micro-isolator cages on hardwood bedding with enrichment consisting of crinkle cut naturalistic paper strands. Mice were maintained on ad libitum PicoLab Rodent Diet 20 (5053) and water supply with 30-70% environmental humidity and temperature of 20 – 26.2oC. All mice were health checked and monitored daily for food and water intake by trained personnel. Single-cell stage fertilized embryos were harvested from the oviducts of super-ovulated 7–8 week old FVB/NJ females mated to 7–8 week old FVB/NJ males. Following pronuclear injection, embryos were transferred into the uteri of pseudo-pregnant 7–8 week old CD-1 surrogate mothers. Sample size selection strategies and scoring criteria were followed based on experience of performing transgenic mouse assays for ~3000 published enhancer candidates.

Wild animals	No wild animals were used in this study
Reporting on sex	Each biosample on the ENCODE portal (encodeproject.org), including mouse biosamples, has annotated biological sex information. No sex-specific analyses were performed as part of this work.
Field-collected samples	None
Ethics oversight	All animal work was reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee.

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	All data is available on the ENCODE portal (encodeproject.org)
Files in database submission	All accession are available in Supplementary tables
Genome browser session (e.g. UCSC)	Available via ENCODE portal directly from each experiment and via SCREEN

Methodology

Replicates	Experiments were replicated as defined by ENCODE standards https://www.encodeproject.org/data-standards/
Sequencing depth	All experiments were subject to sequence depth QC as defined by https://www.encodeproject.org/data-standards/
Antibodies	All antibodies were subject to validation as defined by https://www.encodeproject.org/about/experiment-guidelines/
Peak calling parameters	All ENCODE data was uniformly processed through uniform processing pipelines: https://www.encodeproject.org/pipelines/
Data quality	All experiments were subject to QC https://www.encodeproject.org/data-standards/
Software	All software for data processing (https://www.encodeproject.org/pipelines/) and cCRE annotations and analyses are available on GitHub: https://github.com/weng-lab/ENCODE-cCREs/tree/master/Version-4