nature portfolio

Corresponding author(s): Barak Cohen

Last updated by author(s): 11/21/2022

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	No software was used for data collection
Data analysis	Python 3.9.6; Numpy 1.21.1; Scipy 1.6.3; Scanpy 1.8.1; R 4.0.2; MAST 4.10.0; DEseq2 3.14; CellRanger 6.01; Custom data pr0cessing available on GitHub (https://github.com/szhao045/scMPRA

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

Sequencing data is available in the Gene Expression Omnibus using accession code GSE188639

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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	For the cell culture experiments we included two biological replicates in order to compute a correlation. A total of 676 cis-regulatory sequences were tested, with each sequence being represented by 10 DNA barcodes. For the retina experiments we included two biological replicates in order to compute a correlation. Each replicate contained three retinas. A total of 115 cis-regulatory sequences were tested, with each sequence being represented by 1 DNA barcode.
Data exclusions	Sequencing reads from l0X single-cell transcriptome data were filtered based on standard l0X cell ranger pipeline. Cells with less than 100 genes and genes with less than 3 cells were removed. Cells presenting transcription signature of both cell types were removed as they were likely duplets. Sequencing reads from cis-regulatory sequence (CRS) library were filtered based on (i) whether they are present in the cells with well-measured transcriptome, (ii) whether they have enough depth (minimum: 1 for mixed cell experiment and 10 for k562 alone experient), (iii) whether each CRS is measured in more than 100 cells. Those criteria are commonly used for single-cell RNA-seq data processing and Massively Parallel Reporter Assay data processing.
Replication	Two replicates were included in the study and both were successful.
Randomization	Randomization was not applicable because the same library was used in each experiment.
Blinding	Blinding was not applicable because the same library was used in each experiment.

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

Methods

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Human K562 cell line was obtained from the Genome Engineering & iPSC Center at Washington University School of Medicine; Human HEK293 cell line was obtained from ATCC (catalog number CRL-1573).
Authentication	Neither cell line was authenticated
Mycoplasma contamination	Both cell lines tested negative for Mycoplasma
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	CD-1 mice from Charles River Laboratory were used in this study. Retinas from P0 mice were used. The sex of the mice were not determined because it is difficult to determine the sex of P0 mice.
Wild animals	No wild animals were used in this study.
Reporting on sex	Because it is difficult to determine the sex of P0 mice the mice were not sexed before harvesting the retinas
Field-collected samples	No field-collected samples were used in this study
Ethics oversight	This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All of the animals were handled according to protocol# A-3381-01 approved by the Institutional Animal Care and Use Committee of Washington University in St. Louis. Euthanasia of mice was performed according to the recommendations of the American Veterinary Medical Association Guidelines on Euthanasia. Appropriate measures are taken to minimize pain and discomfort to the animals during experimental procedures.

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

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	Mice retinas were dissociated and sorted for GFP+/DsRed+ and GFP+/DsRed- populations in a 1:1 ratio
Instrument	Cytoflex SRT
Software	NA
Cell population abundance	NA

Gating strategy | Supplementary Figure 1 now shows the gating strategy

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