nature research

Corresponding author(s): Jonathan S. Weissman, Britt Adamson

Last updated by author(s): Feb 17, 2020

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.

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	Data was collected using commercially available platforms from 10x Genomics and Illumina. Flow cytometry data was collected using BD FACSDiva software (versions 8.0 and 8.0.1). FACSDiva was also used for cell sorting.
Data analysis	For alignment of scRNA-seq reads, collapsing reads to unique molecular identifier (UMI) counts, cell calling, and depth normalization of mRNA libraries, we used Cell Ranger 3.0 software (10x Genomics, https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest). Our previously-published analytic framework for Perturb-seq analysis is available at https://github.com/thomasmaxwellnorman/Perturbseq_GI. Python scripts and Jupyter notebooks for direct capture guide identity assignment are available at https://github.com/josephreplogle/guide_calling. Python Jupyter notebooks for the design of hybridization capture probes are available at https://github.com/josephreplogle/target_enrichment. Downstream analyses were performed in Python, using a combination of Numpy, Pandas, scikit-learn, pomegranate, polo, and seaborn libraries, and are described in the Methods, Supplementary Notes, and/or Figure legends.

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. 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

Sequencing data generated during this study are available from the authors and will be deposited in a public repository (SRA/GEO).

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 single-cell RNA-sequencing experiments, cells were loaded to recover a median coverage of 200 cells per guide. This coverage was selected based on our previous work, which estimated the number of cells per perturbation required to evaluate gene level and signature level effects on transcription (Dixit et al, 2016), and based on the recommended upper limit of cells per sample lane on the Chromium Controller (https://kb.10xgenomics.com/hc/en-us/articles/360001378811-What-is-the-maximum-number-of-cells-that-can-be-profiled-).
Data exclusions	Subpopulations of cells with low UMI counts or that contained apoptotic cells were removed from select analyses as detailed in the Methods section.
Replication	Replication was performed where indicated. Single-cell RNA sequencing datasets are internally controlled and were not replicated.
Randomization	No randomization was performed. All Perturb-seq experiments were run with the same negative controls.
Blinding	No blinding was performed.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	GFP+ K562 dCas9-KRAB cells (Adamson et al, 2016), K562 dCas9-KRAB cells (Gilbert et al. 2014), K562 dCas9-SunTag with scFV–VP64 cells (Gilbert et al, 2014), and iPSC Cas9 cells (Mandegar et al., 2016) were constructed as previously described.
Authentication	K562 dCas9-KRAB, GFP+ K562 dCas9-KRAB, and HEK293T cell lines were STR authenticated on 01/25/2019.
Mycoplasma contamination	K562 dCas9-KRAB, GFP+ K562 dCas9-KRAB, K562 dCas9-SunTag with scFV–VP64 cells, and HEK293T cell lines tested negative for mycoplasma on 01/11/2019.
Commonly misidentified lines (See ICLAC register)	None.

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	K562 cells were analyzed by flow cytometry or purified by fluorescence activated cell sorting after lentiviral transduction and growth in standard culture conditions. BFP expression was used as a marker of successful transduction.
Instrument	Flow cytometry data was collected using one of two LSR II flow cytometers and cell sorting was performed on a FACSAria2 (BD Biosciences).
Software	Flow cytometry data was collected using BD FACSDiva software (versions 8.0 and 8.0.1). FACSDiva was also used for cell sorting.
Cell population abundance	n/a
Gating strategy	SSC/FSC gates were first applied to determine cells. Cells were then gated to separate BFP+ from BFP- cells.

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