nature research

Corresponding author(s): Cole Trapnell

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

The number of cells for single cell RNA-Seq were determined by obtaining a reasonable amount of coverage in terms of minimum number of cells per target (roughly more than 50 cells per genotype).

2. Data exclusions

Describe any data exclusions.

No data exclusions

3. Replication

Describe whether the experimental findings were reliably reproduced.

Similar signatures were observed across initial arrayed and CROP-Seq pooled experiments. CRISPRi knockdown of mCherry was performed on multiple cells lines with multiple controls. GFP/BFP experiments were done sorting for both BFP and GFP to ensure results were symmetrical.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not applicable

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	The single cells analysis packages cellranger (10X Genomics) and Monocle2 were used in this article. Custom analysis software is being prepared for distribution on github and will be available upon request for review.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used in the study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	MCF10A breast epithelium cells were purchased from ATCC (CRL-10317), MCF10A TP53 -/- cells were purchased from Horizon Discovery (HD 101-005) and K562 cells were a gift from the Bassik lab.
b. Describe the method of cell line authentication used.	MCF10A and MCF10A TP53 -/- cells were not authenticated but used within 10 passages of purchase. K562 cells were not authenticated.
c. Report whether the cell lines were tested for mycoplasma contamination.	MCF10A cell lines were tested and confirmed negative for mycoplasma contamination. K562 and TP53 -/- cells were not tested for mycoplasma contamination.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None of the cell lines used in this study are registered in the ICLAC database of commonly misidentified lines.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	No human subjects were used in the study.

nature research

Corresponding author(s): Trapnell and Shendure

☒ Initial submission ☐ Revised version ☐ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. 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).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	The cells are MCF10A with TP53 deleted from Horizon Discovery. The cells were transduced with indicated lentiviruses and cultured for an additional 4 weeks in the recommended media. For flow cytometry, 6 cm plates of cells were removed from the plate with 0.25% trypsin, washed with PBS and resuspended in PBS + 1% heat-inactivated FBS, 1mM EDTA, 25 mM HEPES pH 7.5.
6. Identify the instrument used for data collection.	Becton Dickinson FACS Aria II
7. Describe the software used to collect and analyze the flow cytometry data.	The data was collected using FACSDiva version 8 software. Data was analyzed using FlowJo 10
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Cells numbers and percentages with post-sort fraction are listed in Figure S4.
9. Describe the gating strategy used.	Before analysis of fluorescence, live, single cells were gated using FSC-A and SSC-A (for intact cells) and FSC-A and FSC-H (to ensure that only singlets were analyzed). The green+ and blue+ gates were set after compensating for the overlap between the EGFP and EBFP emission using negative and singly positive cells. Those gates were set to exclude non and double-fluorescent cells.

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