nature portfolio

Corresponding author(s): Ken S. Lau and Robert J. Coffey

Last updated by author(s): Jul 7, 2024

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	scRNA-seq data was processed and aligned using DropEst pipeline with STAR aligner. The filtered gene-barcode matrices were then processed in scanpy v1.9.6 (https://pypi.org/project/scanpy/) as AnnData object and normalized to median library size and log transformed, dimensionality reduction (PCA), and generation of umap plots, which use the number of principal components calculated by elbow method (https://github.com/haotian-zhuang/findPC). Additional code used to align and to process scRNA-seq data can be found at https://github.com/Ken-Lau-Lab/STAR_Protocol. Moreover, an example data processing notebook deposited in GitHub (https://github.com/Ken-Lau-Lab/NSC-seq).
Data analysis	Data was analyzed using using open source and custom softwares. Detailed software version and github link can be found in Supplemental Information file

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

Mouse scRNA-seq and WES data is available in GEO with accession number GSE235119. Single cell lineage tree is available in GitHub: https://github.com/Ken-Lau-Lab/NSC-seq. All figures use raw data generated in this project. E7.0 and E8.0 data is from GSE122187. E8.75 data is from GSE123046. Human data have been deposited to the HTAN Data Coordinating Center Data Portal at the National Cancer Institute: https://data.humantumoratlas.org/ (under the HTAN Vanderbilt Atlas). HTAN dbGaP (phs002371). We used reference genome human-hg38, mouse-mm10. TCGA data from cBioportal. GENIE data from AACR GENIE portal.

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	Patient subjects were deidentified. Sex and gender information was self reported.
Reporting on race, ethnicity, or other socially relevant groupings	Patient subjects were deidentified. Race and ethnicity information was self reported.
Population characteristics	Patient subjects were deidentified. Population information was self reported.
Recruitment	Individuals were recruited from those undergoing colonoscopy. Individuals who gave consent were recruited. No other selection criteria were used. Age (41-75) and other informations can be found in supplemental table 4.
Ethics oversight	TCPS was approved by the VUMC and VA Institutional Review Boards and the VA Research and Development Committee. HTAN study was approved by the VUMC Institutional Review Board. All animal experiments were performed under protocols approved by the Vanderbilt University Animal Care and Use Committee (M1600047) and in accordance with NIH guidelines.

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	In this study, sample size was not calculated, rather the reported number of embryos were dependent on variability of embryos at the experimental time frame known by our group via experience. Each embryo accumulates stochastic mutations through developmental time point and required to analyze independently. Embryo reproducibility was accomplished by analytical approaches like proportion of progenitor field size, normalized mosaic fraction, lineage tree, and distribution of mutation density. For human single-cell studies, sample size was determined previously using power calculations in the Chen et al. study [48]. We targeted the number of tumors to greater than the number from the previous study in all conditions.
Data exclusions	We excluded one adult mouse intestinal epithelium dataset for clonal dynamic analysis. There is insufficient barcodes (hgRNAs) found in this dataset, possible failure in library preparation step.
Replication	We demonstrate the reproducible nature of our findings like asymmetric contribution of early embryonic cells across embryos. However, 1st cell's contribution that we reported at E7.75 embryo is not reproducible in other embryos, as we didn't get any mutation at that early stage of the development. This is because indel mutation accumulation is random and we can't control to have a mutation at 2-cell stage to calculate that contribution from 1st cell generation. However, this doesn't invalidate our general asymmetric contribution conclusion, given that other studies also reported similar conclusion. For HCR-FISH and Ab staining, we performed at least 3 replicates per condition. Note that, not all replication attempts were successful. This is due to the fact that 8 um embryo section may not always contain the right tissue sections (somites or gut epithelium).
Randomization	Our study does not follow a hypothesis driven design, and as such, no groupings of embryos or adult mouse were made therefore randomization was not applicable.

Blinding	Single-cell lineage tree reconstruction and cell state assignments operate with the same parameters independently of the embryo, therefore, no need to blind the investigator to the data being handled. We did minor exception for E7.75 tree due to large cell number, mutation number, and increased processing time that we reported in the method section. Human studies were blinded to initial annotation of tumors, but were subsequently unblinded because that information is not critical to the study examining poly or mono-clonality.

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	Tob2 antibody (Invitrogen, Catalog # PA5-62923)
Validation	This antibody has been validated by the manufacturer (https://www.thermofisher.com/antibody/product/TOB2-Antibody-Polyclonal/PA5-62923).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	The HEK293 and EpH4 cell lines originated from ATCC. Please find details of these cell line in supplemental methods section.
Authentication	None
Mycoplasma contamination	Cell lines tested negatively for mycoplasma
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study

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	MARC1 mouse from MMRRC (https://mmrrc.ucdavis.edu/featured-strains-marc1-the-barcoding-lines/). Cas9 mouse (Gt(ROSA)26Sortm1.1(CAG=cas9*,EGFP)Fezh/J strain mouse) from Jackson labs. The age of the barcoded adult mice (MARC1;Cas9) was mentioned in the supplemental methods and corresponding figure legends. ApcMi/+ mouse was 4 months old. Mouse data in Extended Data Fig. 11i is from 18 months old. Mouse data from Extended Data Fig. 2i-j, 10m, and 11a was <3 months old.
Wild animals	None
Reporting on sex	Both male and female mice were used in this study. Sex is not relevant to results shown.
Field-collected samples	None
Ethics oversight	All animal experiments were performed under protocols approved by the Vanderbilt University Animal Care and Use Committee (M1600047) and in accordance with NIH guidelines. Animals were humanely euthanized at the end of experiments according to approved guidelines.

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.