nature portfolio

Corresponding author(s): Paul A. Khavari

Last updated by author(s): 9/12/2024

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	MPRA samples were sequenced on the Illumina NextSeq. CRISPR knockout screen samples were sequenced on the Illumina MiSeq.
Data analysis	Code to generate MPRA library is available on GitHub (https://github.com/zhaozhangyangzi/MPRA_SNP_List/tree/main/SNP_List_pipeline). Code to analyze sequencing data and MPRA results is available at https://github.com/lkellman124/Cancer_MPRA_Project and https://github.com/khavarilab. The following software packages were used as described in the methods: FASTX toolkit (version 0.0.14), UMI-tools (version 1.0.0), Samtools (version 1.7), Bowtie2 (version 2.3.3), MAGeCK (v. 0.5.9.4), bwa (v0.7.17), Picard Tools (v2.21.2), gatk (v4.1.4.1), StringDB (11.0b), MPRAnalyze (v1.8.0), clusterProfiler (v3.12.0), MotifBreakR, tidyverse (v1.3.0), ggplot2 (v3.3.2), pheatmap (v1.0.12), viridis (0.5.1), eulerr (v6.1.0), ggraph (2.0.3) , org.Hs.eg.db (v3.8.2), RColorBrewer (1.1-2), ggpubr (v0.4.0), MASHr (v0.2.69), ComplexHeatmap (v2.10.0), HaploR (4.0.2), TFBS tools (1.32.0), Cytoscape (v3.10), ReactomePA (1.38.0), CHOPCHOP (3.0.0), bedtools (2.30.0), PRECOG, CRISPICK (accessed 12/2020), CRISPOR (accessed 12/2020) and Rmarkdown scripts in R 3.6.1 in RStudio (version 1.3.1093).

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

Raw and processed MPRA sequencing data from this study are available through the Gene Expression Omnibus under accession number GSE183301. CRISPR knock out screen sequencing data are available under accession number GSE183348. Processed MPRA data and ptGene assignments are also available at https://arvid-data.shinyapps.io/cancer/.

Additional data was used from the GWAS catalog (https://www.ebi.ac.uk/gwas/, version 1.0); H3K27ac HiChIP, ATAC-seq, and RNA-seq from Donohue et al, Cell Genomics, 2022; HaploReg (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php, version 4.1); DNAse-seq narrowpeak files from Roadmap (http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/; E027, E028, E119, E057, E058, E127, E059, E061, E097, E098, E125) and ENCODE (https://www.encodeproject.org/; ENCFF514PVZ, ENCFF835QIK, ENCFF934DSX, ENCFF063KYR, ENCFF238YWY, ENCFF439GRJ, ENCFF655IBQ, ENCFF196OHE, ENCFF218MZJ, ENCFF231GLQ, ENCFF239LDT, ENCFF358CBP, ENCFF423OSL, ENCFF563UCE, ENCFF583YED, ENCFF613SZK, ENCFF694GPP, ENCFF752VHK, ENCFF927KUT, ENCFF993TCA, ENCFF229BEY, ENCFF173DAV, ENCFF732RPT,ENCFF263ZFE ENCFF922SBV, ENCFF884CGH, ENCFF021XXM, ENCFF461SHV, ENCFF271AGL); HOCOMOCO (https://hocomoco11.autosome.org/ , version 11); TCGA ATAC peaks (Corces and Granja et al. 2018, accessed April, 2021); ChromHMM annotations from Roadmap (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html; accessed Sept. 2021); GTEx (https://www.gtexportal.org/home/; version 8); eQTLgen (https://eqtlgen.org/cis-eqtls.html; version 2019-12-11); OMIM (https://www.omim.org/; accessed Dec. 2020); StringDB (https://string-db.org/; version 11.0b); COSMIC cancer gene census (https://cancer.sanger.ac.uk/census; v92); co-essential genes from Wainberg et al. 2019; BioGRID (version 4.4.203); BioPlex (version 3.0); DepMap (https://depmap.org/portal/; Public 23Q2) and processed in custom Rmarkdown scripts in R 3.6.1 in RStudio (version 1.3.1093).

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	Melanocytes and keratinocytes were isolated from tissue from male infants circumcised at Stanford Hospital.
Reporting on race, ethnicity, or other socially relevant groupings	Melanocytes and keratinocytes were isolated from tissue from infants circumcised at Stanford Hospital, race and ethnicity were not recorded.
Population characteristics	Participants were male infants whose parents elected circumcision at Stanford hospital. Race, diagnostic and genotypic information were not collected, as specified in our IRB.
Recruitment	Parents of patients scheduled for circumcisions were asked to participate. The population is impacted by who goes to Stanford hospital, who chooses circumcision, and who is interested in participating in research.
Ethics oversight	Stanford IRB

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. Sample size for MPRA was determined based on initial test data in our lab. Sample sizes for the CRISPR knockout screen, luciferase assays, CRISPRi, HDR editing and RhoA activity assay were based on rough estimation of the expected effect size and previous experience with these assays in the lab.
Data exclusions	GM12878 and H9 MPRA data was excluded from the study due to poor sequence representation. SW948 CRISPR knockout screen in 2D culture was excluded due to a high false negative rate. Exclusion criteria were not pre-established.
Replication	For MPRA, 3 biological replicates were performed in each cell type. For the CRISPR KO screen, 10 tumors in 5 mice (one on each flank) were

Replication	used for each of three colon cancer cell lines, and 2 replicates per line were used for the 2D cell culture screen. All replication attempts were successful.
Randomization	Randomization is not relevant to this study. For MPRA experiments, controls are included within the library and measurements are compared within the same donor, not between different control and treatment donors, so there is no randomization of donors into treatment and control groups. For CRISPRi and HDR editing experiments, the same donors were split into both controls and treatments (for HDR, T vs. C allele), so randomization is not applicable. For the CRISPR knock out tumorigenesis experiments, mice of the same strain and age were randomly assigned to each colon cancer cell line, but control comparisons in this experiment were internal to the library, not between cell lines.
Blinding	Blinding was not used for this study. Samples were from cell culture or mouse tumorigenesis experiments, and assessed through qPCR, sequencing and in one case, western blot, and were uniformly processed. For the MPRA and CRISPR knockout screens, controls are internal to the library and so can't be separately influenced by the experimenter.

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	Anti-RhoA mouse IgM Mab, Cytoskeleton Inc, Cat #ARH03, used at 1:500. Clone 54D6.1.16, RRID: AB_2884965. Part of RhoA pull-down activation assay biochem kit - Cytoskeleton BK036).
Validation	Validation statements were provided on the manufacturer's website (https://www.cytoskeleton.com/arh05).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	LentiX 293T cells (Takara, female) MCF7 cells (ATCC, female) HCT116 (ATCC, male) SW948 (ATCC, female) DLD1 (ATCC, male)
Authentication	Cells were not authenticated, but were each bought directly from the company for these experiments.
Mycoplasma contamination	All cell lines tested negative for mycoplasma using MycoAlert (Lonza) before use.
Commonly misidentified lines (See ICLAC register)	None 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, NOD scid gamma (Jackson Laboratory), female, 7-12 weeks
Wild animals	No wild aniimals were used in this study.
Reporting on sex	The CRISPR knockout tumorigenesis experiments were performed in female mice only.
Field-collected samples	No field-collected samples were used in this study.

Ethics oversight	Stanford University Administrative Panel of Laboratory Animal Care (Protocol #9863)

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.*