nature portfolio

Corresponding author(s): Tuomas Tammela

Last updated by author(s): Sep 15, 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

Histological, immunofluorescent, immunohistochemistry and in situ hybridization images were acquired using Mirax scanner (Carl Zeiss). Organoid images were acquired using EVOS M5000 microscope. Quantitative RT-PCR data was acquired using QuantStudio 7 Flex Real-Time PCR System (Life Technologies). Flow cytometry analysis was acquired using LSRFortessa (BD Biosciences) and FACS Aria III (BD Biosciences) instruments with FACSDiva software (v 9.0). Cell sorting was performed using FACS Aria III cytometers (BD Biosciences) with FACSDiva software (v 9.0). Sequencing data was collected from NovaSeq 6000 platform (Illumina). Pyrosequencing primers were designed by PyroMark Assay Design Software (Qiagen, v 2.0.2) and data was acquired by CFX 96 Touch Real-time PCR Detection System (Bio-Rad) and analyzed by CFX Manager (Bio-Rad, v 3.1). Cellular iron measurement was performed induction-coupled plasma mass spectrometry (ICP-MS, Agilent 7900).

Data analysis

Statistical analysis: GraphPad Prism v 9.0.0
Image processing and analysis: Fiji (ImageJ v 2.0.0-rc-65-1.51w)
Flow cytometry analysis: FlowJo v 10.8.1
Mouse LUAD grading analysis: antomated deep neural network by Aiforia Technologies, NSCLC_v25.
scRNA-seq and bulk mRNA sequencing: scanpy (v 1.6.0)/ AnnData (v 0.7.5) or Seurat (v 3.0.2) ecosystem, CellRanger pipeline (v 5.0.0), scrublet (v 0.2.1), leiden (v 0.8.7), bbknn (v 1.3.9), SEACells (v 0.2.0), harmony (v 0.0.5), MAST (1.16.0), EnrichR and PreRank in gseapy package (v 0.10.3), MSigDB Hallmark (repositories), KEGG mouse (repositories), Reactome (repositories), GO Biological Process (repositories), STAR (v 2.7.5a), HTSeq-count(v 0.13.5), DESeq2 (1.26.0),and cor.test and filterByExpr () command in edgeR pipeline (v 3.34.1)
DNA methylation data analysis: Bismark pipeline (v 0.19.0), Trim Galore (v 0.6.4), Bowtie 2 (v 2.3.4), DSS R (v 4.0.4) package
Module score analysis for human LUAD data: cor.test and lm function in R (version 4.0.3)
Schematics: Adobe Illustrator (v 28.5)
Analysis scripts for TCGA analysis are available at: https://github.com/bioinfoDZ/AgingLUAD. Scripts for methylome data pre-processing are

available at: https://github.com/zyqfrog10/DNAmethylAnalysis. Scripts for single-cell/bulk mRNA sequencing and DMC analyses are available at: https://github.com/ewonglab/LUAD-aging-project.

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 lung AT2 cell chromatin immunoprecipitation data was obtained from the Gene Expression Omnibus (GEO) under accession code GSE15820541. Single-cell mRNA sequencing data of aged vs. young normal AT2 and LUAD cells is available from GSE277408). Single-cell mRNA sequencing of AT2 cells harboring Nupr1 and Dnmt1 knockdown by shRNAs are available from GSE277203 and GSE277210, respectively. Bulk mRNA sequencing of aged vs young lung tumor cells with Nupr1 knockout by sgRNAs (GSE277076), ex vivo transformed tumor cells (GSE277241) and DNMT1-i-treated AT2 cells (GSE277061) are available from GEO. DNA methylome sequencing data is available from GSE276548. Pyrosequencing data is available from Figshare (https://figshare.com/s/9e382098b24c29b04fa6).

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	Sex of human patient samples is defined by the biological sex of the individual. Sex of human patients is reported in Extended Data Table 14 and 15. No sex-specific differences were identified. Thus, no additional sex-based analysis was included in the manuscript.
Reporting on race, ethnicity, or other socially relevant groupings	No race, ethnicity or other socially constructed or socially relevant categorization was used in this manuscript.
Population characteristics	Relevant characteristics of all human subjects are listed in Extended Data Tables 14 and 15.
Recruitment	Aged and young participants were recruited from the pool of patients that had undergone treatment at MSKCC without gender, ethnicity or other self-selection bias. This pool is representative of the ethnic diversity in the United States.
Ethics oversight	All human specimens were obtained and approved under MSKCC Institutional Review Board approval (IRB #06-107 and IRB #12-245). All patients provided pre-procedure informed consent.

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 statistical method was used to predetermine the sample size. Samples sizes were chosen based on prior experience and pilot experiments for detecting statistically significant differences between conditions. In compliance with IACUC guidelines, for in vivo experiments, a minimal number of animals with balanced sex for a statistically significant results was used.
Data exclusions	Two mice were excluded in the survival analysis because of early deaths that were not due to cancer, as confirmed by histological evaluation of the lungs post-mortem (see Source Data). One human LUAD case was excluded from the tumor microarray imaging analysis because sections detached during staining.
Replication	All cell and molecular biology expriments were performed at least three times with biological replicates. All reagents are listed in detail in Extended Data Table 17 for ease of replicating the experiments. All attempts at replication were successful.
Randomization	For in vivo treatment experiments with different shRNAs, sgRNAs and DNMT1 inhibitor, mice were randomly distributed into control and experimental groups after balancing for sex. For other experiments, samples were randomly assigned to each group.
Blinding	Analysis of immunofluorescent staining (signal intensity and positivity), immunohistochemistry staining (signal positivity)and in situ

Blinding

hybridization were performed using Fiji in batch. In histological analysis of mouse LUAD tissues, Dr. Carrasco was blinded to the sample group identifiers. For histological analysis of archived human samples, investigators were blinded to the group identifier until the completion of data collection of the image analysis. Investigators were not blinded to treatment groups for other in vivo or in vitro experiments, as such knowledge is essential to conduct the experiments.

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

We report all antibodies used in this manuscript in Extended Data Table 17, including information on manufacturer, catalog number, clone number, and usage.

Validation

Validation statement for each antibody is provided on manufacturers' websites.
Antibody Website
Rabbit anti-mouse Prosurfactant Protein C antibody https://www.emdmillipore.com/US/en/product/Anti-Prosurfactant-Protein-C-proSP-C-Antibody,MM_NF-AB3786?ReferrerURL=https%3A%2F%2Fwww.google.com%2F
Ki67 Monoclonal antibody https://www.thermofisher.com/antibody/product/Ki-67-Antibody-clone-SolA15-Monoclonal/14-5698-82
Anti-Integrin alpha 2 antibody [EPR17338] https://www.abcam.com/en-us/products/primary-antibodies/integrin-alpha-2-antibody-eprTotalSeq™-B anti-mouse Hashtag Antibodies 17338-c-terminal-ab181548
Hnf4a(C11F12) Rabbit mAb https://www.cellsignal.com/products/primary-antibodies/hnf4a-c11f12-rabbit-mab/3113
Human/Mouse/Rat Lipocalin-2/NGAL Antibody https://www.rndsystems.com/products/mouse-lipocalin-2-ngal-antibody_af1857?gad_source=1&gclid=Cj0KCQjw0Oq2BhCCARIsAA5hubVNtQD7V592SegYpx6kBfJS9BEnAqQSos1JE8lT0oR2ZHQkVdjWEFAaAmFVEALw_wcB&gclsrc=aw.ds
Anti-human HTII-280 https://www.terracebiotech.com/product-page/anti-ht2-280-1ml
Rabbit anti-Human Nupr1 antibody https://www.ptglab.com/products/NUPR1-Antibody-15056-1-AP.htm
RFP (tdTomato) Antibody Pre-adsorbed https://www.rockland.com/categories/primary-antibodies/rfp-antibody-pre-adsorbed-600-401-379/
Rabbit anti-mouse p16 INK4A https://www.ptglab.com/products/P16,P19-Antibody-10883-1-AP.htm#product-information
Anti-cytokeratin8, Clone TROMA-1 https://www.sigmaaldrich.com/US/en/product/mm/mabt329m
5-Methylcytosine (5-mC) antibody https://www.activemotif.com/catalog/details/39649/5-methylcytosine-5-mc-antibody-mab-clone-33d3
Anti-mouse CD31_APC https://www.biolegend.com/fr-ch/products/apc-anti-mouse-cd31-antibody-118
Anti-mouse CD45_APC https://www.biolegend.com/fr-ch/products/apc-anti-mouse-cd45-antibody-97
Anti-mouse CD45.2_APC https://www.biolegend.com/fr-ch/products/apc-anti-mouse-cd45-2-antibody-2759
Anti-mouse CD11b_APC https://www.biolegend.com/fr-ch/products/apc-anti-mouse-human-cd11b-antibody-345
Anti-mouse CD11c_APC https://www.biolegend.com/fr-ch/products/apc-anti-mouse-cd11c-antibody-1813
Anti-mouse F4/80_APC https://www.biolegend.com/fr-ch/products/apc-anti-mouse-f4-80-antibody-4071
Anti-mouse TER-119_APC https://www.thermofisher.com/antibody/product/TER-119-Antibody-clone-TER-119-Monoclonal/17-5921-82
Anti-mouse EpCAM_PE https://www.biolegend.com/fr-ch/products/pe-anti-mouse-cd326-ep-cam-antibody-4726
Anti-mouse MHC II_BV650 https://www.biolegend.com/fr-ch/products/brilliant-violet-650-anti-mouse-i-a-i-e-antibody-12085
Anti-mouse Pdpn_PE-Cy7 https://www.thermofisher.com/antibody/product/Podoplanin-Antibody-clone-eBio8-1-1-8-1-1-Monoclonal/25-5381-82
Anti-mouse Sca-1_BV711 https://www.biolegend.com/fr-ch/products/brilliant-violet-711-anti-mouse-ly-6a-e-sca-1-antibody-8632
Anti-human CD45_AF647 https://www.biolegend.com/fr-ch/products/alexa-fluor-647-anti-human-cd45-antibody-14904
Anti-human CD31_APC https://www.biolegend.com/fr-ch/products/apc-anti-human-cd31-antibody-6123
Anti-human EpCAM_PE-Cy7 https://www.biolegend.com/fr-ch/products/pe-cyanine7-anti-human-cd326-epcam-antibody-8107
Anti-mouse IgM_PE https://www.thermofisher.com/antibody/product/IgM-Antibody-clone-II-41-Monoclonal/12-5790-81

TotalSeq™-B anti-mouse Hashtag Antibodies
https://www.biolegend.com/fr-ch/products/totalseq-b0301-anti-mouse-hashtag-1-antibody-17771
https://www.biolegend.com/fr-ch/products/totalseq-b0302-anti-mouse-hashtag-2-antibody-17772
https://www.biolegend.com/fr-ch/products/totalseq-b0303-anti-mouse-hashtag-3-antibody-17773
https://www.biolegend.com/fr-ch/products/totalseq-b0304-anti-mouse-hashtag-4-antibody-17774

https://www.biolegend.com/fr-ch/products/totalseq-b0305-anti-mouse-hashtag-5-antibody-17775
https://www.biolegend.com/fr-ch/products/totalseq-b0306-anti-mouse-hashtag-6-antibody-17776
https://www.biolegend.com/fr-ch/products/totalseq-b0307-anti-mouse-hashtag-7-antibody-17777
https://www.biolegend.com/fr-ch/products/totalseq-b0308-anti-mouse-hashtag-8-antibody-17778
https://www.biolegend.com/fr-ch/products/totalseq-b0309-anti-mouse-hashtag-9-antibody-17779
https://www.biolegend.com/fr-ch/products/totalseq-b0310-anti-mouse-hashtag-10-antibody-18225

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	All cell lines used in this manuscript were derived from primary tissues of genetically engineered mouse models. Three male and 2 female mouse cell lines were used in the serial passage of ex vivo transformed AT2 cells. Two female and one male mouse cell lines were used in the CRISPRi screen of Nupr1 enhancers. Three female and one male mouse cell lines were used in pyrosequencing. HEK293T cells and L-WRN cells used for lentivirus production and conditioned media production was purchased from ATCC (CRL-3216 and CRL-3276, respectively).
Authentication	All cell lines used in this manuscript were derived from genetically engineered mouse models, which have been genotyped at 2 weeks of age using robust, established protocols.
Mycoplasma contamination	All cell lines were subjected to bi-weekly mycoplasma contamination test (PCR) and mycoplasma positive cell lines were excluded from all experiments.
Commonly misidentified lines (See ICLAC register)	All cell lines were derived from primary mouse tissues developed in house. HEK293T and L-WRN are not listed in commonly misidentified lines.

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	All genetically engineered mice were maintained on a mixed C57BL/6 x Sv129 background. Similar numbers of female and male aged and young mice of the following genetically engineered strains were used: KrasLSL-G12D; Trp53flox/flox (KP), with or without reporter alleles including Rosa26-LSL-Tomato (T), Rosa26-LSL-rtTA-IRES-mKate2 (RIK), Rosa26-LSL-Luciferase (L), and Rosa26-LSL-eGFP-Cas9 (Cas9). Wild-type C57BL/6 mice were purchased from Jackson Laboratories (strain #000664) at 10 and 77-78 weeks old and housed at the Sloan Kettering Institute animal facility until they reached 12 or at least 104 weeks to start experiments (young and aged, respectively). Rosa26-tdTomato (T), Rosa26-mTmG (mTmG) were purchased from Jackson Laboratories (strain #007914 and #007676, respectively), expanded and aged at the Sloan Kettering Institute animal facility until they reached 12 or equal to or over 104 weeks to start experiments (young and aged, respectively).
Wild animals	No wild animals were used.
Reporting on sex	No findings applied only to one sex. Similar numbers of female and male aged and young mice have been used in this study
Field-collected samples	No field-collected samples were used.
Ethics oversight	All animal studies were approved by the MSKCC Institutional Animal Care and Use Committee (protocol # 17-11-008). MSKCC guidelines for the proper and humane use of animals in biomedical research were followed.

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

Plants

Seed stocks	No plants were used in this study.
Novel plant genotypes	N/A
Authentication	N/A

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	For isolation of normal AT2 cells and tumor cells derived from 4-, 8- and 12-week time points, lungs were inflated with 3 ml digestion buffer [S-MEM (Thermo Fisher Scientific, #11380037) with 1.7 UI/ml dispase (Corning, #354235), 0.5 UI/µl collagenase IV (Thermo Fisher Scientific, #17104019) and 10 µg/ml DNase I (StemCell Technologies, #07900) through the trachea and then finely minced. For the 17-week time point, tumors were micro-dissected, finely minced, and suspended in ~2-3 volumes of digestion buffer. Tissues were dissociated in a 37 °C oven with gentle agitation for 45-60 minutes. The dissociated tissue cells were filtered using a 100 µm strainer (Corning, #431752) and spun at 1500 rpm for 5 minutes at 4 °C. The supernatant was removed by aspiration and red blood cell lysis was performed using BD PharmLyse (BD Biosciences, #555899) following manufacturer's protocol. Cells were washed with S-MEM (Thermo Fisher Scientific, #11380037) with 2% heat-inactivated fetal bovine serum (HI-FBS, Hyclone, #SH30910.03), filtered through a 40 µm strainer (Corning, #431750), and pelleted by spinning at 1500 rpm for 5 minutes at 4 °C. Cell pellets were resuspended in 2% HI-FBS S-MEM for antibody staining. Human lung tissues were obtained from 4 patients undergoing pulmonary resections (Extended Data Table 14). Normal tissue was obtained from outside the tumor margin, as determined by intraoperative pathology examination. The human lung tissues were dissociated using a similar protocol as for the mice. Briefly, the lung tissue was minced into small cubic pieces and washed several times by pre-cooled PBS. Excessive PBS was drained and tissues were further minced and dissociated in 8-10 volumes of digestion buffer for 1-1.5 hours. Dissociated tissues were processed as described above and cell pellets were resuspended in 2% HI-FBS S-MEM for antibody staining. Single cell solutions of lung cells were incubated with mouse FcR block (BD Biosciences, #553142) on ice for 10 minutes, followed by 30 minutes of staining with desired antibody panels on ice. To sort cells for single-cell sequencing, cell-hashing method was used to maximize the number of biological replicates and control for batch effect in each single-cell sequencing lane. To do this, the cell suspension derived from each mouse was labelled by a different hashtag antibody (2 µl/test, Biolegend, TotalSeq Panel B, #B0301-B0310) together with the desired antibody panel, following the manufacturer's protocol. After the 30-minute staining period, cells were centrifuged at 1500 rpm for 5 minutes at 4 °C and washed twice with 2% HI-FBS S-MEM. Cell pellets were resuspended in PBS with 2% HI-FBS (Hyclone, #SH30910.03) containing 1 µg/mL DAPI (Sigma Aldrich, #D9542) to detect dead cells. Samples were sorted using a BD FACS Aria Sorter using 4-way purity mode. AT2 cells were defined as MHCII+/EpCAM+/lineage- (CD45, CD31, CD11b, CD11c, F4/80 and Ter119)/DAPI- cells. Sca1 and podoplanin were included to maximize the purity of AT2 cells. Early-stage cancer cells (4, 8, and 12 weeks post-tumor initiation) were isolated by selecting tdTomato+ (for KPT mice), mKate2+ (for KP-RIK mice), or GFP+ (for KP-Cas9 mice) and EpCAM+/lineage-/DAPI- cells. For later-stage tumors, EpCAM+/lineage-/DAPI- cancer cells were isolated from micro-dissected tumors. For the isolation of human AT2 cells, single-cell solution of lung cells was incubated with Human TruStain FcX (1:20, Biolegend, #422302) for 10 minutes, followed by staining with anti-human CD45 (Biolegend, #368538), CD31 (Biolegend, #303116), EpCAM (Biolegend, #324222) and HTII-280 (marker for human AT2 cells, 1:40, Terrace Biotech, #TB-27AHT2-280) for 30 minutes on ice. Cells were washed once with 2% HI-FBS S-MEM and further stained with PE-conjugated anti-mouse IgM (Thermo Fisher Scientific, #12-5790-81) for 30 minutes to detect the anti-HTII-280 primary antibody. Cell samples were spun at 1500 rpm for 5 minutes at 4 °C and washed twice with 2% HI-FBS S-MEM. Cell pellets were resuspended in 2% HI-FBS PBS with 1 µg/mL DAPI (Sigma Aldrich, #D9542) to stain dead cells. Samples were sorted using a BD FACS Aria Sorter on purity mode. AT2 cells were defined as HTII-280+/EpCAM+/CD45-/CD31-/DAPI-.
Instrument	BD FACSAria III
Software	FACSDiva v 9.0
Cell population abundance	Cell population abundance was determined by immediate flow cytometry analysis of the post-sort fractions. The purity range was 95% ~ 98%.
Gating strategy	Preliminary FSC/SSC gating were performed for all samples. Mouse AT2 cells were defined by MHCII+/EpCAM+/lineage-(CD45, CD31, CD11b, CD11c, F4/80 and Ter119)/DAPI- cells. Human AT2 cells were defined by HTII-280+/EpCAM+/CD45-/CD31-/DAPI-. Positive and negative staining cells populations were defined by the fluorescence-minus-one (FMO) method.

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