

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> 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)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), 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	<p>For processing of raw data we used the pipeline available at https://github.com/veltenlab/scTAM-seq-scripts based on the Mission Bio Tapestri pipeline version 2 (https://support.missionbio.com/hc/en-us/articles/4411030945815-Tapestri-DNA-DNA-Protein-Pipeline-v2-0-2-19-Aug-2021, v3 for the aging experiment https://support.missionbio.com/hc/en-us/articles/22363469850135-Tapestri-DNA-DNA-Protein-Pipeline-v3-4-27-March-2024). Briefly, barcodes were extracted from the raw sequencing files before alignment to the reference genome subset to the CpG panel. Reads mapping to each of the amplicons were quantified to generate a count matrix and DNA methylation states were determined using a cutoff of one sequencing read as in the original scTAM-seq publication. We used those cellular barcodes that had more than 10 sequencing reads in at least 70% of the control (non-Hhal) amplicons. Doublets were removed using the DoubletDetection tool (version 3.0, https://zenodo.org/record/2678042).</p> <p>To determine the primer combinations that reliably amplify in our panel, we performed a single experiment without the restriction enzyme. For this experiment, wildtype Lin-cKIT+ cells were used and we determined that 453 of the 573 non-control amplicons (79%) amplified in more than 90% of the cells. These amplicons were used for subsequent analysis.</p> <p>For the surface protein data, the Mission Bio pipeline v2 (https://support.missionbio.com/hc/en-us/articles/4411030945815-Tapestri-DNA-DNA-Protein-Pipeline-v2-0-2-19-Aug-2021) was used to extract sequencing reads for a particular cell-barcode/antibody-barcode combination. We restricted analysis of the protein data to those cellular barcodes identified in the DNA methylation library.</p>
Data analysis	<p>Data was analysed mainly using Seurat v4.3.0 in R 4.2.2. All custom scripts, including scripts to generate the figures of the paper, are available at: https://github.com/veltenlab/EPI-clone (release version 2.0). Further R packages used were: ggplot2 v3.4.1</p>

EPI-clonal clustering was performed with EPI-clone version 2.0 (<https://github.com/veltenlab/EPI-clone>). For human samples, EPI-clones uses CHOIR version 0.2.0 (<https://github.com/corceslab/CHOIR/>).

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](#)

The scDNA methylation datasets (M.1-M.3: <https://doi.org/10.6084/m9.figshare.24204750>, M.4: <https://doi.org/10.6084/m9.figshare.25472467>, M.5: <https://doi.org/10.6084/m9.figshare.27917427>, M.6: <https://doi.org/10.6084/m9.figshare.27960771>, M.7: replicate 1: <https://doi.org/10.6084/m9.figshare.25472434>, replicate 2: <https://doi.org/10.6084/m9.figshare.27917454>, M.8: <https://doi.org/10.6084/m9.figshare.27917331>, A.1-A.7: <https://doi.org/10.6084/m9.figshare.25526899>, B.1-B.5: <https://doi.org/10.6084/m9.figshare.28082048>, X.1: <https://doi.org/10.6084/m9.figshare.27991574>, X.2: <https://doi.org/10.6084/m9.figshare.28082066>) and the scRNA-seq dataset (<https://doi.org/10.6084/m9.figshare.24260743>) are available as Seurat objects from Figshare. Source data are provided with this paper. Count matrices are available from GEO under accession number GSE282971. Raw reads for the mouse experiments are available from SRA with BioProject number PRJNA1191391. Raw sequencing data for the human cohort is deposited at EGA (accession number EGAS00001008056). To address ethics board mandates and patient privacy concerns, access is restricted to research projects in hematology and bioinformatics methods development, but excludes ancestry research, surname inference and other research. Requests for access need to be addressed to Lars Velten (lars.velten@crg.eu). For comparison of our endothelial data with published data, we downloaded the following data from the CELLxGENE database: <https://cellxgene.cziscience.com/collections/48d354f5-a5ca-4f35-a3bb-fa3687502252>. The lung EC atlas was downloaded from: https://endotheliomics.shinyapps.io/ec_atlas/.

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	The human cohort (14 individuals) was represents male and female donors, at a ratio of 8:6, see Supplementary Table 1
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Hematologically healthy individuals of different ages. Age range 23-77 years. For details see Supplementary Table 1
Recruitment	The scDNA methylation datasets (M.1-M.3: https://doi.org/10.6084/m9.figshare.24204750 , M.4: https://doi.org/10.6084/m9.figshare.25472467 , M.5: https://doi.org/10.6084/m9.figshare.27917427 , M.6: https://doi.org/10.6084/m9.figshare.27960771 , M.7: replicate 1: https://doi.org/10.6084/m9.figshare.25472434 , replicate 2: https://doi.org/10.6084/m9.figshare.27917454 , M.8: https://doi.org/10.6084/m9.figshare.27917331 , A.1-A.7: https://doi.org/10.6084/m9.figshare.25526899 , B.1-B.5: https://doi.org/10.6084/m9.figshare.28082048 , X.1: https://doi.org/10.6084/m9.figshare.27991574 , X.2: https://doi.org/10.6084/m9.figshare.28082066) and the scRNA-seq dataset (https://doi.org/10.6084/m9.figshare.24260743) are available as Seurat objects from Figshare. Source data are provided with this paper. Count matrices are available from GEO under accession number GSE282971. Raw reads for the mouse experiments are available from SRA with BioProject number PRJNA1191391. Raw sequencing data for the human cohort is deposited at EGA (accession number EGAS00001008056). To address ethics board mandates and patient privacy concerns, access is restricted to research projects in hematology and bioinformatics methods development, but excludes ancestry research, surname inference and other research. Requests for access need to be addressed to Lars Velten (lars.velten@crg.eu). For comparison of our endothelial data with published data, we downloaded the following data from the CELLxGENE database: https://cellxgene.cziscience.com/collections/48d354f5-a5ca-4f35-a3bb-fa3687502252 . The lung EC atlas was downloaded from: https://endotheliomics.shinyapps.io/ec_atlas/ .
Ethics oversight	Samples A.1, A.6 and A.7 : Use of these samples was approved by the Ethics Committee of the Medical Faculty of Heidelberg University (S-480/2011). Samples A2., A.3, and A.5: Use of these samples was approved by the Yorkshire & The Humber - Bradford Leeds Research

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](https://www.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 methods were used to determine optimal sample number for this study. Sample number was limited by the amount of funding available to this study. The samples were selected to cover a broad range of ages (23-77 years). Samples were pre-screened for the existence of mutations in commonly mutated genes in clonal hematopoiesis including DNMT3A and TET2. See Supplementary Table 1 for an overview of the individuals studied. See Supplementary Table 7 for an overview of the number of cells analyzed. Analogously, no statistical methods were used to determine sample numbers for the mouse study. See Supplementary Table 1 for an overview of the mice used in this study (16 in total, age range 10-100 weeks). See Supplementary Table 7 for an overview of the number of cells analyzed.
Data exclusions	As described in the methods part of the paper, cells were excluded as potential doublets based on their DoubletDetection (https://zenodo.org/record/2678042) score. In the context of the human study, a cluster of overstained cells (i.e. positive for all antibodies used) was excluded.
Replication	The main LARRY experiment and the mouse ageing experiment were conducted in two experimental batches to ensure reproducibility. All attempts at replication were successful. No additional attempts for replication were performed. The human study included 13 different bone marrow donors. A scTAM-seq run for a 14th sample resulted in surface antigen data that was for unknown reasons of low technical quality and could not further be analyzed.
Randomization	Not relevant - no treatment groups.
Blinding	Not relevant - no treatment groups.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Sorting antibodies: Pacific Blue™ anti-mouse Lineage Cocktail (Biolegend, cat# 133310) at 1:100 https://www.biolegend.com/en-ie/products/pacific-blue-anti-mouse-lineage-cocktail-7765 CD117 MicroBeads, mouse (Miltenyi biotec; cat# 130-091-224) at 1:100 https://www.miltenyibiotec.com/ES-en/products/cd117-microbeads-mouse.html#130-097-146 APC anti-mouse CD117 (c-kit) Antibody (Biolegend, cat# 135108) at 1:100 https://www.biolegend.com/fr-lu/products/apc-anti-mouse-cd117-c-kit-antibody-6358 PE/Cyanine7 anti-mouse Ly-6A/E (Sca-1) Antibody (Biolegend, cat# 108114) at 1:100 https://www.biolegend.com/fr-lu/products/pe-cyanine7-anti-mouse-ly-6a-e-sca-1-antibody-3137 PE/Cyanine7 anti-human CD3 Antibody; Clone:UCHT1 (BioLegend, cat# 300420) at 1:30 https://www.biolegend.com/en-us/products/pe-cyanine7-anti-human-cd3-antibody-3070
-----------------	---

Alexa Fluor® 488 anti-human CD34 Antibody; Clone: 581 (BioLegend, cat# 343517) at 1:100 <https://www.biolegend.com/en-us/products/alexa-fluor-488-anti-human-cd34-antibody-6201>
 APC anti-human CD38 Antibody; Clone: HIT2 (BioLegend, cat# 303509) at 1:30 <https://www.biolegend.com/en-us/products/apc-anti-human-cd38-antibody-744>

TotalSeq-B antibodies (see also Supplementary Table 2):

CD27 at 1:500 (<https://www.biolegend.com/fr-lu/products/totalseq-b0191-anti-mouse-rat-human-cd27-antibody-19054>)
 CD34 at 1:125 (<https://www.biolegend.com/fr-lu/products/totalseq-b0857-anti-mouse-cd34-20186>)
 CD90 at 1:200 (<https://www.biolegend.com/fr-lu/products/totalseq-b0075-anti-mouse-cd902-thy12-antibody-18955>)
 CD135 at 1:250 (<https://www.biolegend.com/fr-lu/products/totalseq-b0098-anti-mouse-cd135-antibody-18993>)
 CD201 at 1:500 (<https://www.biolegend.com/en-gb/products/totalseq-b0439-anti-mouse-cd201-eprc-antibody-20754>)
 Esam at 1:200 (<https://www.biolegend.com/fr-lu/products/totalseq-b0596-anti-mouse-esam-antibody-21962>)
 CD16_32 at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-b0109-anti-mouse-cd1632-antibody-18458>)
 CD41 at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-b0443-anti-mouse-cd41-antibody-19753>)
 CD115 at 1:200 (<https://www.biolegend.com/fr-lu/products/totalseq-b0105-anti-mouse-cd115-csf-1r-antibody-18913>)
 CD127 at 1:200 (<https://www.biolegend.com/fr-lu/products/totalseq-b0198-anti-mouse-cd127-il-7ra-antibody-19268>)
 CER1A at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-b0115-anti-mouse-fcepsilon1alpha-antibody-19364>)
 CD9 at 1:200 (<https://www.biolegend.com/fr-lu/products/totalseq-b0813-anti-mouse-cd9-antibody-19916>)
 CD61 at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-b0910-anti-mouse-rat-cd61-antibody-20830>)
 CD49f at 1:200 (<https://www.biolegend.com/fr-lu/products/totalseq-b0070-anti-human-mouse-cd49f-antibody-18854>)
 Ter119 at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-b0122-anti-mouse-ter-119erythroid-cells-antibody-18912>)
 MHC_II at 1:200 (<https://www.biolegend.com/fr-lu/products/totalseq-b0117-anti-mouse-i-ai-e-antibody-18916>)
 SCA1 at 1:500 (<https://www.biolegend.com/ja-jp/products/totalseq-b0130-anti-mouse-ly-6a-e-antibody-18949>)
 cKIT at 1:250 (<https://www.biolegend.com/fr-ch/products/totalseq-b0012-anti-mouse-cd117-c-kit-antibody-18323>)
 CD48 at 1:200 (<https://www.biolegend.com/fr-ch/products/totalseq-b0429-anti-mouse-cd48-antibody-19674>)
 CD150 at 1:200 (<https://www.biolegend.com/fr-ch/products/totalseq-b0203-anti-mouse-cd150-slam-antibody-1926>)

TotalSeq-D antibodies (see also Supplementary Table 6).
 Heme Oncology Cocktail, V1.0 (<https://www.biolegend.com/en-gb/products/totalseq-d-human-heme-oncology-cocktail-v10-20465>),
 CD49f at 1:100 (<https://www.biolegend.com/en-gb/products/totalseq-d0070-anti-human-mouse-cd49f-antibody-21430>),
 CD99 at 1:100 (<https://www.biolegend.com/en-gb/products/totalseq-d0845-anti-human-cd99-antibody-21134>),
 CD366 at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-d0169-anti-human-cd366-tim-3-antibody-20814>),
 GPR56 at 1:800 (<https://www.biolegend.com/en-gb/products/totalseq-d0912-anti-human-gpr56-antibody-22747>),
 CD371 at 1:800 (<https://www.biolegend.com/en-gb/products/totalseq-d0853-anti-human-cd371-clec12a-antibody-21577>),
 CD47 at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-d0026-anti-human-cd47-antibody-21052>),
 CD45RA at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-d0063-anti-human-cd45ra-antibody-20992>),
 CD150 at 1:200 (<https://www.biolegend.com/en-gb/products/totalseq-d0870-anti-human-cd150-slam-antibody-23235>),
 CD41 at 1:133 (<https://www.biolegend.com/en-gb/products/totalseq-d0353-anti-human-cd41-antibody-21953>),
 CD61 at 1:133 (<https://www.biolegend.com/en-gb/products/totalseq-d0372-anti-human-cd61-antibody-21954>),
 CD135 at 1:50 (<https://www.biolegend.com/en-gb/products/totalseq-d0351-anti-human-cd135-flt-3-flk-2-antibody-21951>),
 CD96 at 1:50 (<https://www.biolegend.com/en-gb/products/totalseq-d0175-anti-human-cd96-tactile-antibody-22745>).

Validation

All antibodies are standard and well-established monoclonal ABs. Validation is described on <https://www.biolegend.com/fr-fr/bio-bits/highly-specific-validated-antibodies>:

"To ensure they are both specific and sensitive, we validate our antibodies through a variety of methods including:

Testing on multiple cell and tissue types with a variety of known expression levels.

Validation in multiple applications as a cross-check for specificity and to provide additional clarity for researchers.

Comparison to existing antibody clones.

Using cell treatments to modulate target expression, such as phosphatase treatment to ensure phospho-antibody specificity."

Antibody specific information is provided in the web links listed above.

For the Miltenyi CD117 antibody, see <https://www.miltenyibiotec.com/ES-en/products/macs-antibodies/antibody-validation.html> for validation information.

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 mice were carefully monitored by the researchers, facility staff during experiments, and an external veterinary expert responsible for overseeing animal welfare. The mice were housed in a specific-pathogen-free (SPF) facility under a 12-hour light-dark cycle, with regulated temperature (18–23°C) and humidity (40–60%). They had continuous access to a standard diet and water.

Wild animals

No wild animals were used in this study

Reporting on sex

Both male and female mice were used, but sample numbers are too small to assess if findings are specific to one sex.

Field-collected samples

No field collected samples were used in this study

Ethics oversight

All procedures involving animals adhered to the pertinent regulations and guidelines. Approval and oversight for all protocols and

Ethics oversight

strains of mice were granted by the Institutional Review Board and the Institutional Animal Care and Use Committee at Parque Científico de Barcelona under protocol CEEA-PCB-22-001-ARF. The study follows all relevant ethical regulations. Mice were kept under specific pathogen-free conditions for all experiments.

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, mosaicism, off-target gene editing) were examined.

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

Following euthanasia, bone marrow was harvested from the femur, tibia, pelvis, and sternum through mechanical crushing, ensuring the retrieval of most of the cells. The collected bone marrow cells were then sieved through a 40-µm strainer and cleansed with a cold 'Easy Sep' buffer containing PBS, 2% fetal bovine serum (FBS), 1 mM EDTA, and Penicillin/Streptomycin 482 followed by lysis of red blood cells using RBC lysis buffer (Biolegend, Catalog no. 420302). At first, mature lineage cells were selectively depleted through the Lineage Cell Depletion Kit, mouse (Miltenyi Biotec, Catalog no. 130-110-470), while the resulting Lin⁻ (lineage-negative) 485 fraction was then enriched for c-Kit expression using CD117 MicroBeads (Miltenyi Biotec, 486 Catalog no: 130-091-224). These cKit-enriched cells were washed, blocked with FcX and 487 stained with following fluorescently labeled antibodies: APC anti-mouse CD117, clone ACK2 488 (Biolegend catalog no. 105812), PE/Cy7 anti-mouse Ly6a (Sca-1) (Biolegend, catalog no. 489 108114); Pacific Blue anti-mouse Lineage Cocktail (Biolegend, catalog no. 133310); PE anti-490 mouse CD201 (EPCR) (Biolegend, catalog no. 141504); PE/Cy5 anti-mouse CD150 (SLAM) 491 (Biolegend, catalog no. 115912); APC/Cyanine7 anti-mouse CD48 (Biolegend, catalog no. 492 103432). For transplants, EPCR+Lin-Sca-1+c-Kit+ HSCs were sorted via fluorescence-activated cell sorting (FACS) employing a BD FACSAria Fusion with a 70uM nozzle.

Instrument

BD FACSAria Fusion I and BD FACSAria Fusion II

Software

FlowJo™ v10.10

Cell population abundance

Unless otherwise specified we run 50% LK and 50% LSK
See Figure S1 for more details

Gating strategy

For transplant: Lineage⁻, c-Kit⁺, Sca⁺, CD48⁻, CD150⁺, EPCR⁺
For methylation experiment: Lineage⁻, c-Kit⁺, Sca⁺,
For mature cell experiment: CD11b⁺ from WBM and CD45.2⁺ CD11b⁺ from Lung
For more detailed information please refer to the Figure S1

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