nature research

Corresponding author(s): Jennifer E. Adair (jadair@fredhutch.org)
NM18113814

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒ An indication of 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 statistics 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
☐	☒ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

NIS Elements AR 3.00, SP7, Hotfix 10 (Build 555) (Nikon Instruments Inc.)
ZEN Lite software version 2.3 SP1 (Zeiss) (https://www.zeiss.com/microscopy/int/products/microscope-software/zen-lite.html)
Cas-OFFinder (no version information available) (CRISPR RGen Tools) (http://www.rgenome.net/cas-offinder/)
Zetasizer software version 7.13 (Malvern Analytical)
BD FACSDiva software version 8.0 (BD Biosciences)

Data analysis	ImageJ version 1.5i (National Institutes of Health) FlowJo version 10.1 (Tree Star) Cellomics High Content Informatics (HCS) software version 1.6.3.0 (ThermoFisher Scientific) ZEN Lite software version 2.3 SP1 (Zeiss) (https://www.zeiss.com/microscopy/int/products/microscope-software/zen-lite.html) Cas-OFFinder (no version information available) (CRISPR RGen Tools) (http://www.rgenome.net/cas-offinder/) TIDE (no version information available) (Tracking of Idels by Decomposition): https://tide.nki.nl/ PAIR (no version information available) (Paired End Read Merger): https://sco.h-its.org/exelixis/web/software/pear/ Needleman-Wunsch Alignment Tool (no version information available): https://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html GraphPad Prism version 7.03 for Windows (GraphPad Software, USA) Custom Python scripts for analysis of Next Generation Sequence Data previously published by another group in PMID29276718 with all code sources cited and deposited in GitHub (FredHutch_CGT_Gene_Edit_1). Custom code links are now provided.

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data are available for download from NCBI BioProject (ID PRJNA529681). There are no restrictions on data availability. All data are available upon request. Figures with associated raw data include Figure 1, panels c-f; Figure 3, panels a, c, d; Figures 4-6; Supplemental Figures S2, S3 (panels b,c), S5, S6, S7 (panel c), and S8-S10.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study represents a proof-of-concept. No calculation of sample sizes was performed for in vitro studies with primary human cells because there was no data available to estimate differences. Multiple biological replicates for primary human cells are used throughout the study typically with one biological donor per experiment and multiple independent experiments. For each in vitro experiment, two to three independent experiments were sufficient to determine variation among samples most often with statistical significance. This was due to inclusion of appropriate controls to establish noise levels. Sample sizes are designated throughout the manuscript as independent experiments, and technical or biological replicates where applicable. Independent repeats of experiments always produced similar outcomes and are noted where applicable. For mouse studies, we estimated that differences in any of the measured parameters could range between 10%–40%. Biostatistics support determined that, when using minimal requirements for a t-test, accounting for a standard deviation of 5%, and an error probability (alpha-score) of 0.05 with a power of 90%, five to seven biological replicates were required per condition to establish significant differences for each experimental group tested (http://www.biomath.info/power/ttest.htm). Five to ten mice per group were used in these studies. As a negative control for weight measurements, we included four untreated age- and sex-matched mice.
Data exclusions	No data were excluded from analysis.
Replication	This study represents a proof-of-concept. Technical replicates were performed within each experiment. Independent experiments were performed to determine statistical significance between variables. Where representative data are shown, the number of independent experiment repeats and outcomes are described. Biological replicates were performed where possible. Mouse studies were sufficiently powered to determine differences in engraftment levels. Multiple biological donors and genetic loci of interest were evaluated to confirm proof-of-concept. All attempts at replication were successful.
Randomization	For each in vitro experiment, primary cells from each biological donor were equally aliquotted into all experimental groups. When multiple biological donors were used or available, each experimental condition was evaluated across each donor. No information on age, sex or health of the donors of primary cells used in these studies was available at the time of purchase, thus biological variations in these parameters were not controlled. For mouse studies, mice were age- and sex- matched for each experimental group tested.
Blinding	Due to the proof-of-concept nature of this study, true blinding was not possible. Where it was possible, data collection and analysis was done before experimental groups were parsed, such that the collector/analyzer was not aware at the time of data collection and analysis which samples corresponded to which experimental group. In these cases, experimental groups were parsed by a separate individual after data

collection and analysis of individual samples was complete. For example sequence data, mouse study data and colony-forming assay data were collected and analyzed in this manner wherever possible.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

De-identified primary human CD34+ hematopoietic cells used in these studies were obtained from an approved institutional resource, the Fred Hutch Center of Excellence in Hematology (CCEH) Hematopoietic Cell Procurement and Processing Services. These cell products are available for purchase by contacting the director, Dr. Derek Stirewalt (dstirewa@fredhutch.org; 206-667-5309). This contact information will be made available upon request.

Antibodies

Antibodies used

Antibodies used include anti-human CD34 conjugated to brilliant violet 421 (Biolegend; Clone 581; catalog number 343610, lot number B263445), CD38 conjugated to PerCP/Cy5.5 (Biolegend, Clone HIT2, catalog number 303522, lot number B245218), CD45RA conjugated to allophycocyanin (APC) (BD Biosciences, Clone 5H9, catalog number 561212, lot number 7222952), CD90 conjugated to phycoerythrin (PE) (Biolegend, Clone 5El0, catalog number 328110, lot number B234526), CD49f conjugated to PE-Cy7 (Biolegend, Clone GoH3, catalog number 313622, lot number B238675). For mouse studies all antibodies were from BD Biosciences and multiple lot numbers are used over the course of study. Antibodies used include anti-human CD45 conjugated to PerCP (Clone 2D1, catalog number 347464, multiple lot numbers used), anti-human CD3 conjugated to FITC (Clone UCHT1, catalog number 561807, multiple lot numbers used), anti-human CD4 conjugated to V450 (Clone RPA-T4, catalog number 561838, multiple lot numbers were used), anti-human CD20 conjugated to PE (Clone 2H7, catalog number 556633, multiple lot numbers were used), anti-human CD14 conjugated to APC (Clone M5E2, catalog number 555399, multiple lot numbers were used).

Validation

All antibodies are validated in the technical data sheet and product information sheet provided by the manufacturer. For CD34 antibody, human peripheral blood mononuclear cells were stained with CD14 PE, CD45 PerCP/Cy5.5, and CD34 (clone 561) Brilliant Violet 421™ or mouse IgG2a Brilliant Violet 421™ isotype control for validation. For human CD38 antibody, human peripheral blood lymphocytes were stained with CD38 (clone HIT2) PerCP/Cyanine5.5 or mouse IgG1, κ PerCP/Cyanine5.5 isotype control for validation. For anti-human CD45RA antibody rhesus macaque whole blood was stained with APC-H7 Mouse Anti-Human CD45RA antibody (Cat. No. 561212) or with a APC-H7 Mouse IgG1, κIsotype Control (Cat. No. 560167). For human CD90 (Thy1) antibody, the human erythroleukemic cell line HEL was stained with 5E10 PE. For human CD49f antibody, human peripheral blood lymphocytes were stained with anti-human/mouse CD49f (clone GoH3) PE/Cy7 or rat IgG2a, κ PE/Cy7 isotype control. For human CD45 PerCP flow cytometric analysis was performed on whole blood stained with the conjugated antibody. For human CD3 FITC, peripheral blood lymphocytes were stained with the antibody and analyzed on a BD FacsScan instrument. For CD4 V450, whole blood was stained with BD Horizon™ V450 Mouse Anti-Human CD4 and compared to whole blood stained with BD Horizon™ V450 Mouse IgG1, κ Isotype Control (clone MOPC-21, Cat. No. 560373). For CD20 PE, antibody reactivity was tested on peripheral blood ymphocytes of Rhesus macaques. For CD14 APC antibody, human peripheral blood monocytes were stained and analyzed by flow cytometry.
Where possible, we used the same clones originally identified by Notta, F. et al., Sci., 2016 (PMID 26541609). We have previously validated these antibodies for use in our laboratory as referenced in Radtke, S. et al., Sci. Transl. Med., 2017 (PMID 29093179). Validation of antibodies against human and nonhuman primate cell proteins both in our own lab and in publications noted above was performed by sorting primary human hematopoietic cells from mobilized blood or bone marrow and assessing morphology, proliferation capacity, colony-forming ability and co-expression of additional markers, as well as in vitro behavior assays and in vivo repopulation studies in mice.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Mus musculus, strain NOD.Cg-PrkdcscidIl2rgtm1Wjl/Szj (NOD SCID gamma-/-; NSG), 8-12 weeks of age, sex matched across

Laboratory animals	experimental groups.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected in the field.

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 sorting CD34+ cell subsets, bulk CD34+ cells were suspended in a buffer containing sterile Dulbecco's phosphate buffered saline (D-PBS) (Gibco), 1% fetal bovine serum (Atlas Biologics) and 1mM ethylenediaminetetraacetic acid (EDTA) (Sigma) and stained with antibodies for 30 minutes in the dark. Unbound, excess antibodies were washed away with the same buffer used for staining and cells were kept on ice for FACS. For mouse studies, preparation is previously published Haworth, G. et al. Mol. Ther. Meth. Clin. Dev., 2017 (PMID 28649577), which is cited in the methods section. Whole blood was collected by retro-orbital puncture and diluted 1:1 in D-PBS. A portion of this diluted blood mixture was used for antibody staining and flow cytometry. At necropsy, organ samples were harvested and filtered through a 70 micrometer filter and washed with D-PBS. Blood and tissue samples were stained with appropriate antibodies for 15 minutes at room temperature using dilutions suggested by the manufacturer. Red blood cells were removed by ammonium chloride lysis and then diluted out with additional D-PBS prior to flow cytometry analysis.
Instrument	Sony MA900 Mutli-application Cell Sorter and FACS Canto II (BD Biosciences).
Software	Sony standard MA900 cell sorter software was used for human CD34+ cell sorting studies. For mouse studies, BD FACSDiva software version 8.0 was used to acquire data on the FACS Canto II instrument. FlowJo version 10.1 software from FlowJo, LLC/ Tree Star was used to analyze all flow cytometry and FACS data.
Cell population abundance	All sorted samples were confirmed to be >99% pure by flow cytometry analysis following sorting. Mouse engraftment data was calculated as the total percent of human CD45-expressing cells present within the sample analyzed. For specific lineages (CD3, CD4, CD14, CD20) abundance was measured as a function of the proportion of human CD45+ cells only.
Gating strategy	For cell sorting studies: a live cell gate was first established by FSC/SSC plotting of an unstained CD34+ cell population. As multiple markers were used for sorting simultaneously, compensation was accomplished by in-parallel staining of compensation beads (UltraComp eBeads; Invitrogen/Thermo Fisher, catalog number 01-2222-41). One unstained live cells were established and compensation was set, unstained cells were used to set gates for markers in the following order: CD34 and CD38, then CD90 and CD45RA, then CD49f. For mouse studies, gating examples are previously published by Haworth, G. et al. Mol. Ther. Meth. Clin. Dev., 2017 (PMID 28649577). In brief, live cell populations were first gated by FCS/SSC plotting. Gating for each marker was established using full minus one (FMO) controls. For all blood cell lineages other than CD45, a gate was first set for human CD45 expressing cells before gating on lineage-specific cells.

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