nature portfolio

Corresponding author(s): Nikoletta Psatha

Last updated by author(s): Feb 25, 2025

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	No software was used for data collection
Data analysis	Sequence trimming: seqtk v1.3 (https://github.com/lh3/seqtk/releases/tag/v1.3); trimmomatic v0.36 Alignment: STAR v2.4.2; BWA v0.7.12 Post-alignment processing: samtools v1.3 DNase I hotspot detection: Hotspot2 v2.1.1 (https://github.com/Altius/hotspot2/releases/tag/v2.1.1) Genomic interval operations: BEDOPS v2.4.40-typical Differential expression analysis: DESeq2 Motif enrichment: findMotifs.pl (Homer v4.11); hyperMotif v1.0.0 (https://github.com/ggeorgol/hyperMotif/releases/tag/v1.0.0) Chromatin state analysis: ChromHMM v1.2517 FACS data analysis: FlowJo v10.0.8 Statistical analysis and visualization: R 4.0.5; python 3.8

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

Ex vivo CD34+ erythroid differentiation DNase I data generated previously is available from GEO GSE183268 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183268]. DNase I data from CD4+ activated T cells, CD8+ activated T cells, macrophages, CD14+ monocytes, NK cells, CD19+ B cells K562 were obtained from ENCODE (accession numbers: ENCFF751MKF [https://www.encodeproject.org/experiments/ENCSR438USP/], ENCSR294YEJ [https://www.encodeproject.org/experiments/ENCSR294YEJ/], ENCSR721XAP [https://www.encodeproject.org/experiments/ENCSR721XAP/], ENCSR407WGG [https://www.encodeproject.org/experiments/ENCSR407WGG/], ENCSR241BNZ [https://www.encodeproject.org/experiments/ENCSR241BNZ/], ENCSR381PXW [https://www.encodeproject.org/experiments/ENCSR381PXW/], and ENCSR000EKS [https://www.encodeproject.org/experiments/ENCSR000EKS/]). DNase I data from HUDEP-2 cells is deposited in GEO with accession number: GSE252160 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252160]. CUT&RUN data from HUDEP-2 cells is deposited in GEO with accession number: GSE252157 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252157]. MPRA DNAseq data in HUDEP-2 cells is deposited in GEO with accession number GSE252159 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252159].

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	Donor characteristics (age, sex, race) were not requested in the present study for randomization purposes.
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	Human subject recruitment and material collection was carried out by the Cooperative Centers of Excellence in Hematology Core B (Fred Hutch Research Center, Seattle WA) and Leukapheresis Department of Hematology Clinic of "G.Papanikolaou" Hospital.
Ethics oversight	Cells were isolated under the protocol approved by the Fred Hutch Institutional Review Board, The Aristotle University of Thessaloniki and the "G. Papanikolaou" Hospital Institutional Review Board and in accordance with the Declaration of Helsinki.

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	CD34-enriched (CD34+), peripheral blood mononuclear cells from three G-SCF mobilized, healthy donors (purchased from Fred Hutch Cancer Research Center, Seattle, Seattle WA). CD34-enriched (CD34+), peripheral blood mononuclear cells from five, G-CSF and Plerixafor mobilized, patient with beta-Thalassemia major, previously enrolled in mobilization clinical trials (provided by Gene and Cell Therapy Unit, Hematology Clinic, "G. Papanikolaou" Hospital, Thessaloniki, Greece). Each experiment was performed at least three times, independently, to ensure statistical power. Experimental design and sample size for each experiment are detailed in the manuscript.
Data exclusions	No data were excluded.
Replication	All results were reliably reproduced in multiple independent experiments. The exact number of replicates was indicated in figure legends.
Randomization	CD34+ cells donors were randomly selected for age, race and gender.
Blinding	Blinding was not applicable to this study design as it does not involve case/control specimens.

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
[ ]	[x] Antibodies
[ ]	[x] Eukaryotic cell lines
[x]	[ ] Palaeontology and archaeology
[ ]	[x] Animals and other organisms
[x]	[ ] Clinical data
[x]	[ ] Dual use research of concern
[x]	[ ] Plants

Methods

n/a	Involved in the study
[x]	[ ] ChIP-seq
[ ]	[x] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Antibodies

Antibodies used

APC anti-human CD45 1:20 dilution(Cat. No.: 340942, Clone 2D1, BD Biosciences)
PE anti-human CD19, 1:5 dilution (Cat. No.: 555413, Clone HIB19, BD Biosciences)
PE anti-human CD3, 1:5 dilution (Cat. No. : 1P-202-T100, Clone MEM-57, ExBIO)
PerCP anti-human CD33, 1:10 dilution (Cat. No.: PC-662-T100, Clone WM53, ExBIO)
PE anti-human CD235a, 1:10 dilution (Cat. No.: 1P-784-T100, Clone JC159, ExBIO)
APC anti-human CD36, 1:10 dilution (Cat. No.:1A-451-T100, Clone TR9, ExBIO)
APC anti -human CD34 1:20 dilution(Cat. No.: 340667 ,Clone 8G12,BD Biosciences)
PE anti-human CD41, 1:10 dilution (Cat. No.: 555467, Clone HIP8, BD Biosciences)
PE anti- human HBF, 1 µg per 1 x 10E6 cells, (Cat. No.: sc-21756 PE, Clone 51-7, Santa-Cruz)
NucRed™ Live 647 ReadyProbes™ Reagent (Cat. No.: R37106, Invitrogen)
CellROX Deep Red Flow 476 Cytometry Assay kit (Cat. No.: C10491, Invitrogen by Thermo Scientific)

Validation

All antibodies are commercially available and are validated by the respective manufacturer.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

K562 cells and 293FT cells were purchased from Thermo Fischer Scientific and American Type Culture Collection (ATCC), respectively. HUDEP-2 cells were kindly provided Ryo Kurita and Yukio Nakamura, Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki, Japan.

Authentication

K562, 293FT and HUDEP-2 cells were not authenticated.

Mycoplasma contamination

All the cell lines and the primary cell cultures tested negative for mycoplasma.

Commonly misidentified lines
(See ICLAC register)

None of the cell lines mentioned above are listed in the ICLAC Database of Cross-contaminated or Misidentified Cell 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

Immunodeficient NOD.Cg-KitW-41J Tyr+ Prkdcscid Il2rgtm1Wjl/ThomJ (NBSGW) mice were purchased from Jackson Laboratory (Bar Harbor, ME). Mice were housed at a temperature of 22-24°C, under a 12-hour light/dark cycle with free access to water and food. Animal care and handling procedures were in accordance with the Institutional Animal Care and Use Comittee (IACUC) protocol. The animals were 8 weeks old.

Wild animals

No wild animals were used in this study.

Reporting on sex

In study design only female mice were used due to their ability to support human hematopoiesis. Sex is reported in the methods

section.

Field-collected samples: No filed-collected samples were used.

Ethics oversight: All procedures were approved by the institutional animal care and use committee.

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.*

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: Upon media removal, cell samples were incubated with staining solution where antibodies were diluted per manufacturer's recommendations. Details on sample preparation for flow cytometry are available in the Methods section.

Instrument: Samples were acquired on FACS-Calibur (BD Biosciences, San Jose, CA) and MoFlo Astrios Cell Sorter

Software: Flow Cytometry sample acquisition was performed by FlowJo (version 10.0.8, FlowJo, LLC).

Cell population abundance: HUDEP-2 cells, 5 days after transduction with lentiviral enhancer screening library at MOI=0.4, were sorted based on their GFP expression on a MoFlo Astrios Cell Sorter. Three bins were selected represented the 5% of top, medium, and low GFP expressing cells with an average MFI of 15.000,3.000 and 750 respectively (Supplementary Figure 3).

Gating strategy: All Flow Cytometry analyses were performed after gating on FSC-A/SSC-A>FSC-H/FSC-A (SINGLETS).
Gating strategy used to determined the percentage of GFP+ and HBF+ within GFP+ HUDEP2 cells (Fig. 4C): After gating on SC-A/SSC-A>FSC-H/FSC-A (SINGLETS) the GFP population was determined from this gated population. Then the percentage of HbF+ cells was determined within the GFP+ population.

Gating strategy used to determined the percentage of HBF+ cells within the nucred- (enucleated) and nucred+ (nucleated) population (Fig. 5b): After gating on SC-A/SSC-A>FSC-H/FSC-A (SINGLETS) the nucred- and nucred+ populations were determined from this gated population. Then the percentage of HbF+ cells was determined within the two populations.

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