nature research

Corresponding author(s): Miriam Merad

Last updated by author(s): 01/28/2021

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.

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	FACS Diva software version 7(BD), QuPath (version 0.1.2)
Data analysis	FlowJo v10.0.6 (Tree Star Inc.), GraphPad Prism v8.0e (GraphPad Software Inc.), QuPath v0.1.2, ImageJ v1.52, RStudio v1.2, R v3.5.3, R packages: matrixStats, Matrix.utils, mixtools, matrix. tglkmeans (https://tanaylab.bitbucket.io/tglkmeans/).

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. 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

The dataset supporting the findings presented in this study are available from the corresponding author upon reasonable request. Any data that can be shared will be released via a material transfer agreement. Micro-array sequencing data obtained from CD34+ bone marrow cells transduced with BRAFV600E or control NGFR lentiviral vector are listed in Extended Data Table 3. RNA sequencing data obtained from bone marrow purified CD34+ cells from LCH patients and healthy donor are listed in Extended data table 4.

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 methods were used to predetermine sample size. The phenotype had a robust penetrance (100% genotype to phenotype) so each experiment was done with a sample size of 3-8 mice per group.
Data exclusions	No data was excluded
Replication	All experiments were reproduced at least once. All attempts for replication were successfull.
Randomization	Mice were allocated to study groups randomly.
Blinding	Quantification of immune infiltration was blinded by de-identifying samples.

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
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used

Antibodies used for immuno-histochemistry (IHC):
Langerin Human Novocastra 12D6
CD1a Human Dako O10
Ki67 Human Ventana 30-09
CD207 Mouse eBioscience eBioRMUL.2
Ki67 Mouse Bethyl IHC-00375
CDKN2A Human Santa Cruz Biotechnology Sc-759

Antibodies used for flow cytometry and FACS:
CD16/32 mouse 93 eBioscience
Sca1 mouse D7 eBioscience
c-kit mouse 2B8 eBioscience
FLT3L mouse A2F10 eBioscience
CD34 mouse RAM34 eBioscience
CD11b mouse M1/70 eBioscience
Ter119 mouse TER-119 eBioscience
Gr1 mouse RB6-8C5 eBioscience
B220 mouse RA3-6B2 eBioscience
CD115 mouse AFS98 eBioscience
CD45 mouse 30-F11 Biolegend
Ly6G mouse 1A8 Biolegend
MHC II mouse M5/114.15.2 eBioscience
CD11c mouse N418 Invitrogen
Ly6C mouse A2-21 eBioscience
F4/80 mouse BM8 Biolegend

Siglec F mouse E50:2440 BD bioscience
CD24 mouse M1/69 BD bioscience
CD103 mouse 2E7 Biolegend
EpCAM mouse G8.8 eBioscience
CD4 mouse RM4.5 eBioscience
CD8 mouse 5H10 Invitrogen
CD3 mouse 145-2C11 BD bioscience
NK1.1 mouse PK136 eBioscience
CD45 human 2D1 eBioscience
CD34 human 581 Biolegend
CD38 human HIT2 BD bioscience
CD123= IL3R□ 6 human 6H6 Biolegend
CD11c human B-ly6 BD Horizon
CD1a human HI149 Biolegend
CD14 human MHCD1405 Invitrogen
CD33 human
CD207 human IM3577 Beckman Coulter.
All flow cytometry antibody were used at 1/200 dilution.

Validation

All primary anti-mouse and anti-human antibodies were validated for flow cytometry by the manufacturers.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

293T cells were used for lentiviral construct production. They were received from Brian Brown Lab (Icahn School of Medecine at Mount Sinai).

Authentication

The cell line was not authenticated. However, these cells were 100% transfected and were able to produce lentiviral construct, so they can possibly be only 293T cells.

Mycoplasma contamination

293T cell line was tested negative for mycoplasma.

Commonly misidentified lines
(See ICLAC register)

No misidentified cell lines were used in the study.

Animals and other organisms

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

Laboratory animals

Both male and female mice were used, the phenotypes observed were not depended on sex.
C57BL/6 and BALB/c mice were used for experiments. The tamoxifen treatment was initiated between 8 weeks old and 12 weeks old. Mice were housed in the animal facility at Mount Sinai Hospital.

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve field-collected samples.

Ethics oversight

Ethical approval for mouse experiments was obtained by the IACUC at Mount Sinai Hospital.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics

The study involves staining on OCT embedded tissues and on paraffin embedded tissues from LCH patients followed at the Texas Children's Hospital. The study involves also bulk RNAseq of hematopoietic progenitors from bone marrow of LCH patients followed at the Texas Children's Hospital in Houston. Patients clinical characteritics are detailed in Extended Data Table 2.

Recruitment

For the bulk RNAseq of hematopoietic progenitors, patients were selected according to the % of BRAFV600E + cells detected in their bone marrow. We selected the three patients with the highest %. We could obtained bone marrow from one healthy infants. FFPE skin sections and OCT sections from LCH patients were selected according to their availability. There was no other criteria of selection.

Ethics oversight

Human tissues used in these studies were collected, stored, and processed according to protocols approved by the Baylor College of Medicine Institutional Review Board.

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

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	Adult skin (from the ears), lung, liver, spleen, and femurs were dissected from mice at 12-16 weeks old (4 weeks after tamoxifen injections). Adult skin was first incubated overnight at 4° in dispase II solution (Roche) (2.5mg/mL in PBS) with the dermal side facing down. The dermis was separated mechanically from the epidermis and both tissues were digested in a solution of collagenase D (Roche) (1 mg/mL) and DNase I (Roche) (1 mg/mL) in RPMI (Corning) +10% FBS for one hour at 37°C. After digestion, the dermis and epidermis were homogenized using a 18G needle. The cell suspension was filtered through a 70 um cell strainer into a flow cytometry tube. For the spleen and the lung, tissues were enzymatically digested in a solution of collagenase D (Roche) (1 mg/mL) and DNase I (Roche) (1 mg/mL) in HBSS (Corning) for 30 minutes at 37°C, followed by mechanical trituration with an 18G blunt tipped syringe and filtration through a 70 um filter. Single cell suspensions from liver were obtained after digestion with collagenase IV (C5138, Sigma) in HBSS (Corning) at 37°C for 45 minutes. Nonparenchymal cells were enriched by centrifugation in 35% Percoll (17-0891-01, GE Healthcare) for 30 min at 1,300 rcf. BM single cell suspensions were obtained after flushing out the BM from the femur using a 27G needle and then were incubated with ACK lysing buffer (420301, BioLegend) for 3 minutes at room temperature. For flow cytometry, cells were stained in FACS buffer (PBS without Ca2+ and Mg2+ supplemented with 2% heat inactivated FBS and 5mM EDTA) with flow cytometry monoclonal antibodies for 20 minutes at 4°C.
Instrument	BD LSRFortessa
Software	FACS Diva software version 7 (BD).
Cell population abundance	Mostly GFP+ cells from transduced cord blood were FACS-sorted. Purity of sorted populations was ~97% in all samples and was assessed by flow cytometry.
Gating strategy	All gating strategy started by: 1. Gate on fsc-a vs. ssc-a was set to include all cell populations, but excluding debris 2. Gate on fsc-a vs. fsc-w was set to exclude doublets. 3. Gate on ssc-a vs. ssc-w was set to exclude doublets. 4. Gate on fsc-a vs. DAPI was set to exclude dead cells (DAPI+) Gating for HSC (hematopoietic stem cells): 5. Gate on ssc-a vs. lineage to include lineage negative cells 6. Gate on c-kit+ Sca1+ 7. Gate on CD34- Flt3- Gating for MPP (mutipotent progenitors): 5. Gate on ssc-a vs. lineage to include lineage negative cells 6. Gate on c-kit+ Sca1+ 7. Gate on CD34+ Flt3-/+ Gating for MEP (megakaryocyte erythroid progenitors): 5. Gate on ssc-a vs. lineage to include lineage negative cells 6. Gate on c-kit+ Sca1- 7. Gate on CD34- CD16/32- Gating for CMP (common myeloid progenitors): 5. Gate on ssc-a vs. lineage to include lineage negative cells 6. Gate on c-kit+ Sca1- 7. Gate on CD34+ CD16/32int Gating for GMP (granulocyte macrophage progenitors): 5. Gate on ssc-a vs. lineage to include lineage negative cells 6. Gate on c-kit+ Sca1- 7. Gate on CD34+ CD16/32high Gating for bone marrow neutrophils: 5. Gate on CD45 vs. fsc-a was set to exclude CD45- cells 6. Gate on CD11b+ Ly6G+ Gating for bone marrow dendritic cells: 5. Gate on CD45 vs. fsc-a was set to exclude CD45- cells 6. Gate on Ly6G vs CD11b to exclude Ly6G high cells 7. Gate on MHC II+ CD11c+

Gating for bone marrow monocytes
5. Gate on CD45 vs. fsc-a was set to exclude CD45- cells
6. Gate on CD11b vs Ly6G was set to include CD11b+ Ly6G- cells
7. Gate on Ly6C vs. CD115 was set to include Ly6C high CD115 negative as monocytes.
Gating for bone marrow macrophages:
5. Gate on CD45 vs. fsc-a was set to exclude CD45- cells
6. Gate on CD11b vs Ly6G was set to include CD11b+ Ly6G- cells
7. Gate on CD115 vs. LY6C to exclude CD115+ cells
8. Gate on F4/80 vs. CD115 to include F4/80+ cells
9. Gate on ssc-a vs. fsc-a, macrophages are fsc-a high but ssc-a low.

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