nature research

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

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	BD Diva 7.0, Leica Acquisition
Data analysis	R 3.3.1 and general data analysis packages (data.table, ggplot2, dplyr, igraph, MASS), flow cytometry packages (flowDensity, flowClean), and Phenstat_2.8 (available from Bioconductor), UFO vailable from Ryan Brinkman, R studio Version 1.1.36, bespoke R code for reference range analysis and assessment of false positive rate, FlowJo vl.10, Cytoscape 3.6.1, BD Diva 7.0, Graph pad Prism 6, Definiens Developer, bespoke Image J macro and Python scoring script for ANA assay, Excel Microsoft 2010 Code used for initial hit calling and pre-processed per-mouse data for flow cytometry, ear epidermis and DSS assays are available from https://github.com/AnnaLorenc/3i_heatmapping. The ImageJ macro and the Python code used to score ANA positivity, the Definiens Developer code to assess the ear epidermis images and the R code used to assess the false positive rate are available on request.

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 flow cytometry files that support the findings of this study are available from www.flowrepository.org with the identifiers listed below:
3i PBMC panel 1: FR-FCM-ZYPJ
3i PBMC panel 2: FR-FCM-ZYPK
3i T-cell Spleen IMPC: FR-FCM-ZYX9

3i T-cell MLN IMPC: FR-FCM-ZYXB
3i B-cell Spleen IMPC: FR-FCM-ZYXC
3i B-cell MLN IMPC: FR-FCM-ZYXE
3i M-cell Spleen IMPC: FR-FCM-ZYXF
3i M-cell MLN IMPC: FR-FCM-ZYXG
3i P2 SPL IMPC: FR-FCM-ZYXN
3i BM IMPC: FR-FCM-ZYXQ

Vignettes showing gating of affected cell populations in KO mice and controls are available for all phenotypes deemed significant in this study in flow cytometry assays through www.mousephenotype.org (Associated Images on a gene page).
Microscopy image files from ear epidermis assay, ANA assay and DSS histology are available www.mousephenotype.org (Associated Images on a gene page).
All assay results on mouse and strain level that support the findings of this study are available through http://www.immunophenotype.org.
Figures with associated raw data are:
1BCD, 2ABC, 4BC, 5BCDE, 6ABC, 7BCDEF, S1EGHI, S2, S3ABCD.

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	Sample sizes for the different assays were chosen based on experience of previous assay variation and availability of mice for phenotyping for each ko mouse line. Bespoke statistical pipelines were developed to account for the relatively low n (4-8 mice per ko line, depending on the assay). For the observational screen, a reference range approach was used, discussed in White et al 2013, (PMID:23870131), Karp et al 2012 (PMCID: PMC3530558) and Karp et al 2014 (PMCID: PMC4208881). Throughout the project, false positive rates were monitored (figure S3). All potential hits were manually curated by an expert. The hit calling strategy errs on the side of caution, thus limiting the number of false positive calls, but potentially causing false negative calls. Users are encouraged to access the raw data online to identify phentoypes that might have been missed by this stringent approach.
Data exclusions	Data was only excluded if results could be traced back to an obvious technical error, e.g. if CD62L shedding had occurred during the preparation of flow cytometry samples, thus preventing the proper assessment of activation status. For all data excluded, predefined error codes were used to record the reason for exclusion in the Wellcome Sanger Institute Sanger database. All phenotypes were determined by statistical analysis but were manually curated. If, for example, ko mice had a high percentage of NK cells, but all controls on the same day also had a high percentage of NK cells, the expert performing manual cu ration could come to the conclusion that the phenotype was due to temporal variation. However, all raw data is available online and users are encouraged to form their own opinion based on the raw data.
Replication	Due to the nature of the screen, findings in the observational screen were not replicated. However, we performed false discovery rates analyses based on our work flow (supplementary figure 3) to show that the likelihood of false positives was very small. In the challenge screens, some but not all phenotypes were confirmed with additional cohorts, subject to resources.
Randomization	Allocation into experimental groups was driven by KO mice availability and always mice from several genotypes were processed together. For the observational screen, mice from one ko line were analysed on at least two and mostly three to five different days, to minimize batch effects. A lot of thought went into designing the observational screen at the Well come Sanger Institute, see for example Karp et al 2012 (PMCID: PMC3530558) and Karp et al 2014 (PMCID: PMC4208881). For the challenge screens, concomitant wild type controls were used for operational reasons. In sample processing (for example microscopy slides analysis) samples were analysed in batches comprising samples collected over few weeks.
Blinding	Investigators were blinded to group allocation during sample preparation and data acquisition.

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

Marker, Flurochrome, Clone, dilution, Company Cat No.
TCR-d Pe-Cy7 GL3 1:200 Biolegend 118124
CD4 BV786 RM4-5 1:200 BD 563331
CD8 AF700 53-6.7 1:200 BD 557959
CD25 APC PC61 1:100 BD 557192
GITR PE DTA-1 1:800 BD 558119
CD44 FITC IM7 1:200 BD 553133
CD62L PerCP-Cy5.5 MEL-14 1:200 BD 560513
KLRG1 BV421 2F1 1:200 BD 562897
F4/80 PerCP-Cy5.5 BM8 1:800 Biolegend 123128
CD103 PE M290 1:200 BD 557495
CD317 BV650 927 1:400 Biolegend 127019
CD86 PE-Cy7 GL1 1:400 BD 560582
CD3 (lin) BV421 145-2C11 1:100 BD 562600
CD19 (lin) BV421 ID3 1:400 BD 562701
CD161/NK1.1 (lin) BV421 PK136 1:100 Biolegend 108732
CD43 PerCP-Cy5.5 1B11 1:400 Biolegend 121224
GR1 AF700 RB6-8C5 1:800 Biolegend 108422
B220/CD45R Pe-Cy7 RA3-6B2 1:200 Biolegend 103222
CD24 APC M1/69 1:1600 BD 562349
BP1/Ly51 PE BP-1 1:400 BD 553735
CD3 BV785 145-2C11 1:100 Biolegend 100232
IgD AF488 11-26c.2a 1:800 Biolegend 405718
IgM BV421 RMM-1 1:100 Biolegend 406518
CD19 Pe-Cy7 ID3 1:400 BD 552854
CD5 APC 53-7.3 1:300 BD 561895
Ly6G APC 1A8 1:800 BD 560599
Ly6C AF700 AL-21 1:200 BD 561237
MHCII/IA/IE FITC 2G9 1:800 BD 553623
CD11c BV786 HL3 1:200 BD 563735
CD161/NK1.1 BV650 PK136 1:100 Biolegend 108735
CD21/35 PE 7G6 1:3000 BD 552957
CD11b BV510 M1/70 1:800 Biolegend 101245
CD23 BV421 B3B4 1:200 BD 562929
CD45 Qdot 605 30-F11 1:200 eBiosciences 93-0451-42
IgG1 PE A85-1 1:800 BD 550083
B220 (CD45R) AF-700 RA3-6B2 1:200 BD 557957
IgM BV786 R6-60.2 1:100 BD 564028
IgD PerCP-Cy5.5 11-26c.2a 1:200 Biolegend 405710
GL-7 AF647 GL7 1:1000 BD 561529
CD95 PE-Cy7 Jo2 1:400 BD 557653
CD138 BV650 281-2 1:400 Biolegend 142517
CD5 BV510 53-7.3 1:300 BD 563069
CD21/35 FITC 7G6 1:100 BD 553818
CD4 BV510 RM4-5 1:3000 Biolegend 100553
CD8a APC-H7 53-6.7 1:200 BD 560182
KLRG1 PE-Cy7 2F1 1:500 Biolegend 138416
NK1.1 BV421 PK136 1:600 Biolegend 108732
TCRαβ PerCP-Cy5.5 H57-597 1:600 Biolegend 109228
TCRγδ APC GL3 1:600 Biolegend 118116
CD62L PE-CF594 MEL-14 1:2000 BD 562404
CD44 FITC IM7 1:2000 BD 561859
CD25 PE PC61 1:500 Biolegend 102008
CD45 Alexa 700 30-F11 1:600 Biolegend 103128
Ly6C PerCP-Cy5.5 HK1.4 1:5000 Biolegend 128012
CD11b PE-Cy7 M1/70 1:2000 Biolegend 101216
Ly6G V450 1A8 1:600 BD 560603
IgD BV510 11-26c.2a 1:2000 BD 563110

CD115 APC AF598 1:500 Biolegend 135510
CD19 PE-CF594 ID3 1:2000 BD 562291
Ly6B FITC 7/4 1:1000 AbD Serotec MCA771FB
I-A/I-E PE M5/114.15.2 1:4000 Biolegend 107608
CD45 eFluor-450 30-F11 1:200 eBioscience 48-0451-82
Vg3 TCR FITC 536 1:300 BD 553229
MHC I-A/I-E AlexaFluor-647 M5/114.15.2 1:500 Biolegend 107618
CD4 PE RM4-5 1:400 Biolegend 116006
CD8 APC 53-6.7 1:400 Biolegend 100712

Validation

Commercial antibodies are validated by the manufacturer:
-Biolegend: https://www.biolegend.com/reproducibility
-BD :https://www.bdbiosciences.com
-Ebioscince: https://www.thermofisher.com

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

State the source of each cell line used.

Authentication

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

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

Laboratory animals

Age and sex of mice used in this study are stated in supplementary data accompanying each figure.

Wild animals

NA

Field-collected samples

NA

Ethics oversight

All experiments were performed according to protocols approved by the UK Home Office regulations, UK Animals (Scientific Procedures) Act of 1986 and were approved by the Wellcome Sanger Institute Animal Welfare and Ethical Review Board.

Animals were housed at a density of 2 to 5 mice per cage in polysulfone individually ventilated cages (Tecniplast) with sterilised Aspen bedding substrate and standard environmental enrichment, in a specific pathogen-free unit. The light cycle was maintained at 12h light/12h dark with lights off at 19:30 hours and no twilight period. Room temperature was 21 ± 2 °C and humidity was regulated at 55 ± 10%. Mice received sterilized (vacuum packed and irradiated) diet (Lab Diets, 5021–3) from weaning and had ad libitum access to autoclaved water and food.

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

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, 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	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

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	Blood:

Sample preparation	Blood was collected into EDTA coated tubes (peripheral blood leukocytes assay) via the retro-orbital sinus. Whole blood for peripheral blood leukocyte assays was stained with two titrated cocktails of antibodies. Spleen and MLN: After removing the fat, MLN were transferred into 1.7 ml microfuge tubes with 200 µl of buffer (PBS Ca2+/Mg2+, 2% FCS (v/v), 10 mM HEPES) and ruptured using small plastic pestles. 400 µl enzyme buffer (PBS Ca2+/Mg2+, 2% FCS (v/v), 10 mM HEPES), Collagenase (1 mg/ml), and DNAse (0.1 mg/ml) were and samples were incubated for 15 minutes at 37°C. The reaction was stopped by adding 60 µl stop buffer (PBS 0.1 M EDTA) and samples filtered through 50 µm filters and centrifuged for 5 minutes at 400 x g at 8°C. The cell pellet was resuspended in FACS buffer (PBS Ca2-/Mg2-, 0.5% FCS, EDTA 2 mM) and transferred to 96-well V bottom plates for antibody staining Bone Marrow: The tibia was cleared of muscle tissue and cut below the knee and above the ankle. The open bone was placed into a cut pipette tip placed into a microfuge tube, thereby keeping the bone away from the bottom of the tube and allowing the bone marrow to be centrifuged out of the bone at 1000 g for 1 minute. The bone was discarded and the pellet resuspended in 50 µl RBC lysis buffer at room temperature for 1 minute. Cells were washed in 200 µl FACS buffer and again centrifuged at 1000 g for 1 minute. Cells were resuspended in 400 µl FACS buffer and transferred to 96-well V bottom plates for antibody staining.
Instrument	4 laser BD Fortessa X20 (Spleen, MLN, Bone marrow) 4 laser BD LSR II (blood) Full configuration of both instruments is available on the 3i website (www.immunophenotype.org)
Software	Flow cytometry data was analysed using the following software: -DIVA (BD) -FlowJo v10 (Treestar) -Bespoke supervised automated analysis. (for details see: Materials and methods and Rahim, A. et al. High throughput automated analysis of big flow cytometry data. Methods 134-135, 164–176 (2018).)
Cell population abundance	Cell subset frequencies are reported as %of appropriate parent population and as % of total CD45. In addition, analysis of blood cell absolute numbers was facilitated through back calculation of percentages relative to absolute cell counts as determined by haematological analysis.
Gating strategy	Gating strategies are outlined in Supplementary Figure 6.

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

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis