nature research

Corresponding author(s): Lori Broderick

Last updated by author(s): Jul 10, 2019

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	BioRad CFX Manager v3.0 Image Pro-Plus (Media Cybernetics) ImageJ v2.0.0
Data analysis	Sequence analysis was performed using available packages as described in Methods, pages 18-20. Data contained in ChIP- seq and End-seq BED files (reference 48, available from GEO, accession GSE99197, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99197) were visualized using the Integrated Genomics Viewer (IGV 2.4.8) Broad Institute. Statistical analyses and graphing were performed in Microsoft Excel and Graphpad Prism (version 5.03; Graph Pad, Graph Pad Software Inc., CA). Page 31.

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 authors declare that the data supporting the findings of this study are available within the paper and its supplementary information. Data contained in ChIP- seq and End-seq BED files (Reference 48, available from GEO, accession GSE99197, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99197) were visualized using the Integrated Genomics Viewer (IGV 2.4.8) Broad Institute. Genome and exome data for individual patients cannot be made publicly available for reasons of patient confidentiality. Qualified researchers may apply for access to these data, pending University of California, or CHOP institutional review board approval. Source data for Figures 2D and 6I is included in the Source Data file.

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 size was not predetermined with any statistical test. Sufficient sample sizes were estimated based on previous experiments with cre breedings, and strains.
Data exclusions	No murine data were excluded. For S. cerevisiae sporulation, spore clones retaining a second copy of chromosome XIV due to chromosomal nondisjunction were not included in the analysis of haploid spores (p. 22).
Replication	All experiments were successfully reproduced with biological replicates or independent technical replicates as described in methods and figure legends. All figures shown are representative of the replicates performed.
Randomization	Mice were identified by genotype, with age-matching within experiments.
Blinding	Mice and yeast strains were identified by numbers only during the data collection and analysis. All histological and image analyses were performed by blinded investigator.

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	CD3 (eBioscience, 17A2, 17-0032-80, 1:200), CD19 (eBioscience, 1D3, 12-0193-81, 1:400), B220 (eBioscience, RA3-682, 11-0452-82, 1:200), B220 (BioLegend, RA3-682, 130-102-357, 1:50), IgM (eBioscience, 121-15F9, 48-5890-82, 1:100), IgD (eBioscience, 11-26, 17-5993-80, 1:200), Flt3 (eBioscience, A2F10, 46-1351-82, 1:100), GL-7 (eBioscience, GL-7, 53-5902-80, 1:200), CD23 (eBioscience, B3B4, 12-0232-81, 1:200), CD38 (eBioscience, 90, 12-0381-81, 1:200), CD21/35 (eBioscience, 8D9, 47-0211-80, 1:100), MHC class II (eBioscience, M5/114.15.2, 12-5321-81, 1:100), CXCR4 (eBioscience, 2B11, 147-9991-80, 1:200), CD19 (eBioscience, 1D3, 17-0193-80, 1:400), Flt3 (eBioscience, A2F10, 46-1351-82, 1:100), CD43 (eBioscience, R2/60, 11-0431-81, 1:200), BP-1 (eBioscience, 6C3, 12-5891-81, 1:200), CD25 (eBioscience, PC61.5, 61-0251-80, 1:200), CD24 (eBioscience, M1/69, 25-0242-80, 1:200), CD3 (Alexa Fluor 488, clone 17A2, Biolegend, 100210, 10 µg/mL), anti-B220 (eFluor® 570 CD45R, clone RA3-6B2, eBioscience, Inc., 41-0452-80, 1 µg/mL), anti-CD169 (Alexa Fluor 647, clone 3D6.112, BioLegend, Inc., 142408, 5 µg/mL), Alexa Fluor 488 Rat IgG2b, kappa (BioLegend, Inc., 400625, 10 µg/mL), (eFluor 570 Rat IgG2a, kappa (ThermoFisher Scientific, Inc., 41-4321-82, 1 µg/mL), Alexa Fluor 647 Rat IgG2a, kappa (BioLegend, Inc., 400526, 50 µg/mL), anti-mouse IgM-HRP (1:1000, ThermoScientific, 62-6820), goat anti-mouse IgG3 (1:1000, ThermoScientific, M32707), goat anti-mouse IgG2b (1:1000, ThermoScientific, M32407), anti-mouse Top2b (1ug/mL, ThermoFisher Scientific, MA5-24310), anti-mouse-HRP (1:3000, 7076S, Cell Signaling Technology, Inc.)
Validation	Only commercial antibodies were used that were validated by the vendor, per data available on the manufacturer's website. For immunocytochemistry, samples were stained with only secondary antibodies to determine specificity of the primary antibody.

Animals and other organisms

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

Laboratory animals	The following murine lines were used (all bred to C57/Bl6): Top2bflox2, Cd79atm1(cre)Reth, cd19-cre, Top2b+/delAGA. All mice were used between 12 and 22 weeks, with age and gender matching within experiments.
Wild animals	Study did not include samples collected from the wild.
Field-collected samples	Study did not include samples collected from the field.
Ethics oversight	Institutional Animal Care and Use Committee at the University of California, San Diego

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	This was not a population study. Subjects were previously identified as having features consistent with Hoffman syndrome.
Recruitment	Subjects were known as having features of Hoffman syndrome.
Ethics oversight	UCSD Institutional review board and CHOP 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	Peripheral whole mouse blood was obtained by cheek venipuncture, and collected in EDTA-treated collection tubes. Red cells were lysed using ACK Lysing Buffer (ThermoFisher Scientific) per the manufacturer's instructions. 10^6 cells were stained with conjugated monoclonal antibodies. Murine spleens were mechanically dissociated and erythrocytes removed by hypotonic lysis. 1x10^6 cells were stained with conjugated monoclonal antibodies. Bone marrow cells were harvested from femurs and tibiae by flushing with PBS, and 1x10^6 cells stained with conjugated monoclonal antibodies.
Instrument	BD Biosciences LSR II cytometer
Software	Data was collected with FACSDiva software, and analyzed with FlowJo V10 software
Cell population abundance	Cells were not sorted by flow cytometry.
Gating strategy	Supplementary Figure 9. Flow cytometry gating strategy. Bone marrow (A,B) and spleen (C) cells were isolated from adult mice and analyzed by 8-color flow cytometry for B cell subsets. In the bone marrow, FSC vs SSC was used to distinguish lymphocytes and their precursors. Cells were subsequently gated on B220-V500 and CD43-PE, and then by IgD-APC, IgM-e450 and BP-1-PE and CD24-APC-e780. To further interrogate the Pro to PreB cell transition, FSC vs SSC was used to distinguish lymphocytes and their precursors. Cells were subsequently gated on B220-V500 and CD19-APC, and then by IgM-e450, BP-1-PE, CD25-PE-e610 and CD24-PE-Cyanine7.72 This strategy applies to Figures 3K and 5E. In the spleen, lymphocytes were gated by FSC vs SSC, and then on CD19-PE and B220-V500, followed by IgM-e450 and IgD-APC. CD38, Flt3 and CD21/35 positivity was also assessed. This stragey applies to Figures 3L and 5F. In all cases, 500,000 events were collected per sample. (D) For CFSE and propidium iodide staining, single lymphocytes were identified by FSC vs SSC. Live cells were distinguished by exclusion of propidium iodide and CFSE gates were established using the biology function in Flow-Jo (Figures 7E).

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