nature research

Corresponding author(s): Ari M. Melnick

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	BD FACS Canto II, Panoramic Viewer Software
Data analysis	FlowJo v10, ImageJ 1.44, Graphpad Prism v7.0, R 3.3.2, DESeq2, STAR, BWA, GSEA, MACS2, deepTools 3.0.2, Picard Tools 2.18.11, SAMtools v1.8

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/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

Data availability
Sequence data that support the findings of this study have been deposited in GEO under SuperSeries GSE118783 with the accession codes GSE106417 (RNA-seq for

GC B cell), GSE106416 (RNA-seq for lymphoma) and GSE107920 (ChIP-seq). Other data that support these findings are available from the corresponding author upon reasonable request.

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	For animal studies, the number of animals was chosen to ensure 90% power with 5% error based on observed SD from previous studies. No statistical analyses were performed to predetermine sample sizes for in vitro experiments, but our sample sizes are similar to those generally employed in the field.
Data exclusions	No data were excluded from any analyses with the exception of one LSD1fl/fl and one Cy1-Cre; Lsd1fl/fl ATAC-seq sample that were excluded due to very low sequencing coverage. In addition, in our survival analysis we censored 3 mice that died less that one month after bone marrow transplantation lacking distinct pathological features due to failed BM engraftment, one animal that was euthanized due to tail necrosis and one euthanized due to dermatitis based on Research Animal Resource Center recommendations.
Replication	All experiments were performed independently more than once using biologically independent replicates. All results reported were successfully replicated.
Randomization	Randomization of experimental groups were not applicable in most experiments in this study. For treatment with LSD1 inhibitor GSK-LSD1 (Figure 5a-b), C57Bl6 mice were randomized before treating with active compound or PBS control.
Blinding	Blinded review and lymphoma diagnosis was performed for reporting pathology results in Figure 6.

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

Antibodies

Antibodies used	Flow cytometry antibodies: PE-Cy7 anti-B220 (eBioscience, 25-0452, dilution 1:500), APC anti-CD38 (eBioscience,17-0381, dilution 1:500), APC anti-B220 (BD Biosciences, 553092, dilution 1:500), PE anti-FAS (BD Biosciences, 554258, dilution 1:500), FITC anti-GL7 (BD Biosciences, 553666, dilution 1:500), BV421 anti-CD138 (BioLegend, 142507, 1:500), FITC anti-IgK (BioLegend, 409509, lot#B195860, 1:500) or APC-Cy7 anti-IgK (BioLegend, 409503, lot#B232635, 1:500), PE-Cy7 anti-CD11b (BioLegend, 101216, dilution 1:300, lot#B268267), APCCy7 anti-Gr1 (108424, dilution1:300, lot#B244266) and V450 anti-BrdU (BD Pharmingen, V450, dilution: 1:50). ChIP-seq antibodies: LSD1 (abcam 17721, lot#GR3193508-2), BCL6 (Santa Cruz N3, sc-858, lot#I2607) and H3K4me1 (abcam, ab8895, lot#GR193882-2). Western Blot: LSD1(Santa Cruz, 1B2E5, sc-53875, lot#I2508, dilution 1:500), FLAG (F1804, M2, Sigma, dilution 1:1000, lot#SLBN5629V), GAPDH (Santa Cruz, FL-335, sc-25778, dilution 1:500), Actin (Santa Cruz, sc-81760, dilution 1:500, lot#C2008), H3K4me1(abcam, ab8895, dilution 1:5000,lot#GR193882-2), H3(Millipore, 07-690, dilution 1:10,000, lot#DAM1832538).
Validation	All antibodies are validated upon arrival to the lab, no matter what is claimed on the data sheets. In this study, antibodies were

tested using western blot, immunoprecipitation and immunohistochemistry, either following RNAi depletion of the antigen from cells or in gene knockout tissues to confirm that the antibody reacts against the corresponding antigen. In the case of histone antibodies, we performed screening on histone arrays to assess specificity. For ChIP-seq antibodies were additionally validated by qCHIP after protein knockdown. We have purchased entire lots or large quantities of antibodies to ensure reproducibility over time, given the well-known problems with lot-to-lot variation.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	OCI-LY1: Ontario Cancer Institute (OCI) TMD8 and HBL-1: Kindly provided by Dr. Jose Angel Martinez-Climent, Centre for Applied Medical Research (CIMA), Pamplona, Spain SUDHL4: Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ)
Authentication	GC derived B-lymphoma cell lines were obtained from the cell lines repositories, American Type Culture Collection (ATCC) and the Ontario Cancer Institute (OCI). Cell lines are kept in our institutional bio-repository and we conduct semi-annual genomic identification by matching short tandem repeat DNA profiles with those deposited in databases from repositories and monthly testing for mycoplasma and other potential contaminants. Serum is also purchased in bulk and is validated on a number of cell lines before use. All CRISPR-generated and newly derived cell lines were characterized and validated by surveyor assay.
Mycoplasma contamination	Biannual testing for mycoplasma was performed. All cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No cell line used is part of ICLAC.

Animals and other organisms

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

Laboratory animals	The Research Animal Resource Center of Weill Cornell Medicine approved all mouse procedures. Conditional Lsd1 knockout mice (loxP-flanked Lsd1 allele, Lsd1fl/fl) were purchased from the Jackson laboratory (023969). By crossing Lsd1fl/fl with the transgenic Cγ1-Cre strain (The Jackson Laboratory, 010611) we generated heterozygous Cγ1-Cre Lsd1fl/+ mice, which were crossed to yield Cγ1-Cre Lsd1fl/fl mice. As control group, we used Cγ1-Cre negative Lsd1fl/fl littermates. Age- and sex-matched C57Bl6 mice were immunized intraperitoneally at 8 to 12 weeks of age were used for assessment of the germinal center formation, which were induced by intra-peritoneal injection of 2% SRBC solution or affinity maturation by intra-peritoneal injection of NP-CGG28-30 in alum. We also used Iμ-Bcl6 mice (obtained from Dr. Ricardo Dalla-Favera, Columbia University to generate Iμ-Bcl6 Cγ1-Cre Lsd1fl/fl by crossing with Cγ1-Cre Lsd1fl/fl. These mice were used for germinal center formation assays induced by SRBC IP injections or harvested bone marrow to perform transplantation to C57Bl6 recipients. Finally, B cell conditional LSD1 model was generated by crossing Lsd1fl/fl mice with CD19-Cre mice (Jackson Laboratory, 006785) where Cre is expressed from the pre-B cell stage. For treatment with LSD1 inhibitor GSK-LSD1, mice were injected intraperitoneally with drug or vehicle drug or PBS vehicle starting the following day after induction of GC by SRBC and administered daily at a concentration of 50 mg/kg per day for 9 consecutive days after which the mice were sacrificed (day 10). All studies were conducted in accordance with the GSK Policy on the Care, Welfare and Treatment of Laboratory Animals and were reviewed the Institutional Animal Care and Use Committee either at GSK or by the ethical review process at the Research Animal Resource Center of Weill Cornell Medicine. For murine bone marrow transplantation assays, bone marrow cells from 6-8 week old male donors were harvested. One million bone marrow cells of each genotype (IμBcl6 Cγ1-Cre Lsd1+/+, IμBcl6 Cγ1-Cre Lsd1fl/fl and Cγ1-Cre Lsd1fl/fl mice or Lsd1fl/fl) were injected into the tail veins of lethally irradiated female C57Bl6 mice. All mice were followed until any one of several criteria for euthanizing were met, including severe lethargy, more than 10% body weight loss, and palpable splenomegaly that extended across the midline, in accordance with our Weill Cornell Medicine and Memorial Sloan-Kettering Cancer Center Institutional Animal Care and Use Committee–approved animal protocols. Animal care was in strict compliance with institutional guidelines established by Weill Cornell Medicine, the Memorial Sloan-Kettering Cancer Center, the Guide for the Care and Use of Laboratory Animals (National Academy of Sciences 1996) (Silverman et al., 2006), and the Association for Assessment and Accreditation of Laboratory Animal Care International.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve field collected samples.

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.*	Sequence data that support the findings of this study have been deposited in GEO under SuperSeries GSE118783 with accession code GSE107920.

Files in database submission

LY1_LSD1 baseline FASTQ
LY1_LSD1 NT FASTQ
LY1_LSD1 si FASTQ
LY1_BCL6 NT FASTQ
LY1_BCL6 si FASTQ
LY1_H3K4me1 NT FASTQ
LY1_H3K4me1 si FASTQ
LY1_Input_NT FASTQ
LY1_Input_si FASTQ
LY1_LSD1 baseline BEDGRAPH
LY1_LSD1 NT BEDGRAPH
LY1_LSD1 si BEDGRAPH
LY1_BCL6 NT BEDGRAPH
LY1_BCL6 si BEDGRAPH
LY1_H3K4me1 NT BEDGRAPH
LY1_H3K4me1 si BEDGRAPH
LY1_LSD1 r1/NT overlap BED
LY1_LSD1peaks (depleted by siBCL6) BED
SUDHL4_BCL6_fc.bw bedgraph
SUDHL4_LSD1_fc.bw bedgraph
SUDHL4_H3K4me1_NT_norm.bw bedgraph
SUDHL4_H3K4me1_si_norm.bw bedgraph
SUDHL4_LSD1_peaks_fc.5.p.10.txt bed
SUDHL4_Bcl6_peaks_fc.8.p.10.txt bed
SUDHL4_BCL6_fastq fastq
SUDHL4_LSD1_fastq fastq
SUDHL4_H3K4me1_NT_fastq fastq
SUDHL4_H3K4me1_si_fastq fastq
SUDHL4_Input_NT_fastq fastq
SUDHL4_Input_si_fastq fastq

Genome browser session
(e.g. UCSC)

Trackhub url is:
https://s3-us-west-2.amazonaws.com/kac209/hatziLSD1/hatziLSD1.hub.txt

Methodology

Replicates

For ChIP-seq experiments with cells treated with siRNAs, OCI-Ly1 or SUDHL4 cells were transfected using Nucleofector 96-well Shuttle system (Lonza) with 1uM siRNA against BCL6 (HSS100968) or non- targeted siRNA (46-2001) (Stealth RNAi, Invitrogen) for 24h. Briefly cells were fixed, lysed, and sonicated to generate fragments less than 400 bp. Sonicated lysates were incubated with antibodies against LSD1 (abcam 17721), BCL6 (Santa Cruz N3, sc-858), H3K4me1 (abcam, ab8895) or IgG control (abcam, ab171870) and after increasing stringency washes immunocomplexes were recovered and DNA was isolated. ChIP enrichment was confirmed by qPCR. H3K4me1 ChIPs in SUDHL4 also contained Drosophila spike-in chromatin control (750ng per ChIP, Active Motif, 53083) added during precipitation along with Drosophila antibody (2ug, Active Motif, 61686) per manufacturer's recommendation for proper signal normalization. ChIP-seq libraries were prepared using the Illumina ChIP-seq Library preparation Kit following the manufacture's instructions with minor modifications. Briefly 10ng of purified ChIP DNA (quantified using Qubit 2.0 fluorometer, Invitrogen) was end repaired by conversion of overhangs to phosphorylated blunt ends. A' bases were added to the 3' ends of the DNA fragments and Illumina adapters (1:30 dilution) were ligated to the ends of ChIP fragments. After adaptor ligation DNA was separated by electrophoresis and size selected by isolating a gel band of 250 ± 25bp. Size selected fragments were PCR amplified for 15cycles using Illumina genomic DNA primers 1.1 and 1.2 with the following program (30 s at 98oC, 15cycles of 10 at 98oC, 30 s at 65oC, 30 s at 72oC and 5min extension at 75oC). Q-PCR was repeated to confirm retention of relative enrichment. For ChIP-seq, raw images generated went through primary image analysis and basecalling (RTA v1.6) that was followed by Illumina Genome Analyzer Off-Line Basecaller (OLB v1.6) analysis. Raw ChIP-seq reads were trimmed and filtered for quality using Trim Galore and FASTQC. Reads were aligned using BWA-mem against human reference genome hg38, and non-uniquely mapping reads and PCR duplicates were removed using samtools and Picard, respectively only sequences mapped uniquely to the genome with not more than 2 mismatches were used for downstream analysis. Several reads mapping to the same exact location (clonal reads) were considered amplification artifacts and were excluded from the analysis. The determine the quality of individual ChIP-seq experiments we used ENCODE consortium quality metrics (Landt, Marinov et al. 2012). Strand cross-correlation analysis was performed to asses signal-to-noise ratios, and samples were retained for analysis that passed a normalized strand coefficient (NSC) threshold >1.05 and a relative strand correlation (RSC) threshold >0.8. Read density tracks were visualized using the UCSC browser. Peak calling was performed using MACS2[41]. Each ChIP-seq dataset was normalized to its corresponding input. H3K4me1 signals from SUDHL4 cells were normalized based on the scaling factor calculated from the ratio of human and drosophila mapped reads as previously described. Peaks were annotated based on the RefSeq database (hg38). Peaks localized +/2kb of the TSS were defined as promoter peaks, peaks localized +/2kb of the TES were defined as 30 end peaks, and peaks > 2kb away from genes were defined intergenic.

Sequencing depth

The sequencing depth of our ChIP-seq results is roughly between 26-58 million reads.
OCI-Ly1 LSD1 baseline total: 32,609,339 uniquely mapped: 26,324,718 (single end)
OCI-Ly1 LSD1 NT total: 54,122,363 uniquely mapped: 40,127,234 (single end)
OCI-Ly1 LSD1 si total: 51,843,752 uniquely mapped: 36,676,237 (single end)
OCI-Ly1 BCL6 NT total: 43,932,942 uniquely mapped: 32,334,558 (single end)
OCI-Ly1 BCL6 si total: 56,680,286 uniquely mapped: 42,924,695 (single end)
OCI-Ly1 H3K4me1 NT total: 41,902,231 uniquely mapped: 34,859,650 (single end)
OCI-Ly1 H3K4me1 si total: 47,129,185 uniquely mapped: 38,791,112 (single end)

	OCI-Ly1 Input_NT total: 42,943,750 uniquely mapped: 32,968,653 (single end) OCI-Ly1 Input_si total: 42,214,823 uniquely mapped: 32,889,628 (single end) SUDHL4 LSD1 total: 20,709,196 uniquely mapped: 19,598,046 (single end) SUDHL4 BCL6 total: 61,010,362 uniquely mapped: 55,939,001 (single end) SUDHL4 H3K4me1 NT total: 68,690,586 uniquely mapped: 61,844,189 (single end) SUDHL4 H3K4me1 si total: 50,686,624 uniquely mapped: 45,978,833(single end) SUDHL4 Input_NT total: 53,543,411 uniquely mapped: 50,756,036 (single end) SUDHL4 Input_si total: 46,291,950 uniquely mapped: 43,428,408 (single end)
Antibodies	LSD1 (abcam 17721), BCL6 (Santa Cruz N3, sc-858) and H3K4me1 (abcam, ab8895)
Peak calling parameters	Peak calling was performed using MACS version 2.1.1.20160309 with the following parameters: OCI-Ly1_LSD1 baseline > FC=5, p val < 10^-10 25,922 peaks OCI-Ly1_LSD1 NT > FC=5, p val < 10^-10, 10,536 peaks OCI-Ly1_LSD1 si > FC=5, p val < 10^-10 4,878 peaks OCI-Ly1_BCL6 NT > FC=8, p val < 10^-10, 13,458 peaks OCI-Ly1_BCL6 si > FC=8, p val < 10^-10 3,745 peaks SUDHL4_BCL6 > FC=8, p val < 10^-10 13,056 peaks SUDHL4_LSD1 > FC=5, p val < 10^-10, 13,119 peaks
Data quality	Raw ChIP-seq reads were trimmed and filtered for quality using Trim Galore and FASTQC. Reads were aligned using BWAmem against human reference genome hg38, and non-uniquely mapping reads and PCR duplicates were removed using samtools and Picard, respectively. An input control was used as internal control for peak calling using MACS2. Peaks were identified using MACS2. The determine the quality of individual ChIP-seq experiments we used ENCODE consortium quality metrics (Landt, Marinov et al. 2012). Strand cross-correlation analysis was performed to asses signal-to-noise ratios, and samples were retained for analysis that passed a normalized strand coefficient (NSC) threshold >1.05 and a relative strand correlation (RSC) threshold >0.8.
Software	For ChIP-seq, raw images generated went through primary image analysis and basecalling (RTA v1.6) that was followed by Illumina Genome Analyzer Off-Line Basecaller (OLB v1.6) analysis, where reads were aligned to the human genome (UCSC hg38) using BWA 0.6.2. Read density tracks were visualized using the UCSC browser. Peak calling was performed using MACS version 2.1.1.20160309. Each ChIP-seq dataset was normalized to its corresponding input. Peaks were annotated based on the RefSeq database (hg38). Peaks localized +/2kb of the TSS were defined as promoter peaks, peaks localized +/2kb of the TES were defined as 30 end peaks, and peaks > 2kb away from genes were defined intergenic.

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	Single-cell suspensions from mouse spleens were ficoll separated and stained using fluorescent-labeled anti-mouse antibodies, incubated on ice in the dark for 30min, then washed 2x with PBS with 0.5% BSA and 5mM EDTA and resuspended in 200ul washing buffer for acquisition.
Instrument	FACS Canto II
Software	FlowJo v10.2
Cell population abundance	Sorted human and murine GC B cells were confirmed to be >90% following each sorting.
Gating strategy	For selection of single cells, cells were first gated based on SSC-A/FSC-A, then FSC-H/FSC-W followed by SSC-H/SSC-W gating. For Sorting: B cells were gated on APC-B220 positive population on a SSC-A/APC-A dotplot. GC B cells were gated on GL7+FAS+ populations on a BV421-A/PE-Cy7 boxplot (Log axes). For regular flow: live B cells were selected as B220APC+DAPI- cells on APC-A/DAPI-A dotplot followed by gating GC B cells either as GL7-FITC+/FAS-PE+ or CD38-APC negative /FAS-PE positive on dotplot with logarithmic axes. All gating strategies are shown in Figure S6.

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