Corresponding author(s): Michael E. Vinyard, Cindy Su, Allison P. Siegenfeld, Amanda L. Waterbury, Allyson M. Freedy, Pallavi M. Gosavi, Yongho Park, Eugene E. Kwan, Benjamin D. Senzer, John G. Doench, Daniel E. Bauer, Luca Pinello, & Brian B. Liau

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	ChIP-seq and RNA-seq analysis was performed on the Illumina NextSeq 500 instrument and targeted exon sequencing was performed on the Illumina MiSeq instrument. Fluorescence polarization, cell viability and enzymatic activity assays were measured on the SpectraMax i3x platereader (Molecular Devices) with SoftMax Pro (version 6.5.1).
Data analysis	Anaconda (version 1.8.7) with python (version 3.0) was used to analyze and visualize the CRISPR-suppressor scanning data. The following packages were used in conjunction with the python base language to perform the analysis: Pandas (version 0.23.3), NumPy (1.15.0), SciPy (version 1.1.0), Matplotlib (version 2.2.2), Seaborn (version 0.9.0), Scikit-learn (version 0.19.1). R (version > 3.4.2) with the ggplot2 (version 3.0.0) package was also used for visualizing some of the CRISPR-suppressor scanning data. Exon sequencing data was analyzed using CRISPResso (version 1.0.13). The algorithms and equations used in the sgRNA clustering analysis are specified in the Methods section. Rstudio (version 1.1.453) with R (version > 3.4.2) and BioConductor (version 3.8) was used for the RNA-seq processing. The following packages were used in conjunction with R to analyze and visualize the RNA-seq data: DESeq2 (version 1.20.0), pheatmap (version 1.0.10), biomaRt (version 2.36.1), gplots (version 3.0.1), reshape (version 0.8.7), ggpubr (version 0.1.8), and ggplot2 (version 3.0.0).

The following software was used for RNA-seq and ChIP-seq read processing and analysis: STAR (version 2.5.0), HTseq (version 0.9.1), GSEA (version 3.0), BWA mem (version 0.7.15), Picard tools (version 2.18.5), SAMtools (version 0.1.12), HOMER (version 4.10), and ngs.plot (version 2.6.1).

Homology models of wild type LSD1 and LSD1 mutants were generated using MODELLER (version 9.21) and Rosetta 2018 (release 7111c54), and a subset of structures were further optimized in Gaussian 16. GFI1B wild type and GFI1B F5A peptides docked to LSD1 were modeled and assessed in Chimera (version 1.10.2) and Maestro Schrödinger (version 11.4.011) using Prime homology modeling. PyMOL (version 1.8.4.2) was used for visualization. Extinction coefficients for determining protein concentration were calculated using ExPASy ProtParam.

Nonlinear regression and plotting for fluorescence polarization, cell viability, ***Thermo***FAD, and enzymatic activity assays were performed in GraphPad Prism (version 7.0b).

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

ChIP-seq and RNA-seq data reported in this paper have been deposited to NCBI GEO and are available under the accession number GSE121426. $Log_2 + 1$ transformed CRISPR-suppressor scanning reads used for Fig 1–3 and Fig S1–S3 are supplied in Supplementary Tables 2–4.

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 all genetic screens, sample sizes were chosen such that sufficient sequencing depth was achieved for all timepoints, as outlined in the criteria defined in Joung et. al, 2017 (cited in the methods section of this manuscript).
Data exclusions	For the CRISPR-suppressor scanning experiments, by pre-established exclusion criteria, libraries that did not receive adequate sequencing depth (>1000X coverage) and sgRNAs with no counts at week 0 were excluded from the analysis.
Replication	Where indicated in the paper, experiments were performed in replicate (duplicate or triplicate). Replicate type is specified in the text. All attempts at replication were successful.
Randomization	Randomization was not relevant to this study as all experiments were performed in vitro or in cells.
Blinding	None of the experiments performed in this study involved the blinding of the investigators, although the genetic screen experiments were performed in an unbiased, pooled fashion. Investigators were not blinded for the practical purposes of running the experiments.

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials: There are no known restrictions on any of the biological materials used in this study. All materials are available from standard commercial sources or our laboratory.

Antibodies

Antibodies used: AB_2614985 (H3K4me2-human antibody, Active Motif 39141, LOT# 01008001), AB_2561016 (H3K27ac-mouse antibody, Active Motif 39133, LOT# 31814008), AB_242533 (Rabbit Anti-BHC110/KIAA0601 Polyclonal Antibody, Unconjugated, Bethyl Laboratories A300-215A, LOT# 2), AB_449000 (GFI1B antibody, Abcam ab26132, LOT# GR318193-8), AB_2616335 (GFI1B-human antibody, Santa Cruz Biotechnology sc-28356X, LOT# D1615), AB_262044 (Monoclonal ANTI-FLAG® M2 antibody, Sigma-Aldrich F1804, LOT# SLBW3851), GAPDH (Santa Cruz Biotechnology sc-47724, LOT# B0210).

Validation: For AB_242533, external validation for lot# 2 is available under ENCODE ID: ENCAB753DKR. For AB_2561016, external validation for lot #16612005 is available under ENCODE ID: ENCAB943WPC. For AB_2616335, DOI: 10.18632/oncotarget.24774. For AB_2614985, DOI: 10.1016/j.cell.2016.10.039. For AB_449000, DOI: 10.1073/pnas.1322570111. For AB_262044, DOI: 10.1210/en.2014-1264. For GAPDH, DOI: 10.1038/srep23372.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s): SET-2 (Matthew D. Shair, Harvard University), MV4;11 (ATCC), CMK-86 (ATCC), HEK 293T (Bradley E. Bernstein, Massachusetts General Hospital, K562 (ATCC)

Authentication: SET-2 and MV4;11 were authenticated using short tandem repeat (STR) profiling through Genetica Cell Line Testing. CMK-86 and K562 were authenticated using short tandem repeat (STR) profiling through Sigma-Aldrich. STR data was processed by the investigator and determined to exceed the threshold for authenticity. HEK293T cells were not authenticated.

Mycoplasma contamination: Cell lines were tested routinely throughout the duration of the study for mycoplasma. In each case, all cell lines tested negative.

Commonly misidentified lines (See ICLAC register): No commonly misidentified lines were used in this study.

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.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121426

Files in database submission

WT_DMSO_Input.bw
WT_DMSO_LSD1.bw
WT_DMSO_GFI1B.bw
WT_DMSO_H3K4ME2.bw
WT_DMSO_H3K27AC.bw
WT_GSK_Input.bw
WT_GSK_LSD1.bw
WT_GSK_GFI1B.bw
WT_GSK_H3K4ME2.bw
WT_GSK_H3K27AC.bw
insR_DMSO_Input.bw
insR_DMSO_LSD1.bw

insR_DMSO_GFI1B.bw
insR_DMSO_H3K4ME2.bw
insR_DMSO_H3K27AC.bw
insR_GSK_Input.bw
insR_GSK_LSD1.bw
insR_GSK_GFI1B.bw
insR_GSK_H3K4ME2.bw
insR_GSK_H3K27AC.bw
WT_DMSO_Input_R1.fastq.gz
WT_DMSO_Input_R2.fastq.gz
WT_DMSO_LSD1_R1.fastq.gz
WT_DMSO_LSD1_R2.fastq.gz
WT_DMSO_GFI1B_R1.fastq.gz
WT_DMSO_GFI1B_R2.fastq.gz
WT_DMSO_H3K4ME2_R1.fastq.gz
WT_DMSO_H3K4ME2_R2.fastq.gz
WT_DMSO_H3K27AC_R1.fastq.gz
WT_DMSO_H3K27AC_R2.fastq.gz
WT_GSK_Input_R1.fastq.gz
WT_GSK_Input_R2.fastq.gz
SET2_DMSOandGSK_DESeqNormalizedCounts.txt
WT_DMSO_rep1.count
WT_DMSO_rep2.count
WT_DMSO_rep3.count
WT_GSK-LSD1_rep1.count
WT_GSK-LSD1_rep2.count
WT_GSK-LSD1_rep3.count
insR_DMSO_rep1.count
insR_DMSO_rep2.count
insR_DMSO_rep3.count
insR_GSK-LSD1_rep1.count
insR_GSK-LSD1_rep2.count
insR_GSK-LSD1_rep3.count
WT_GSK_LSD1_R1.fastq.gz
WT_GSK_LSD1_R2.fastq.gz
WT_GSK_GFI1B_R1.fastq.gz
WT_GSK_GFI1B_R2.fastq.gz
WT_GSK_H3K4ME2_R1.fastq.gz
WT_GSK_H3K4ME2_R2.fastq.gz
WT_GSK_H3K27AC_R1.fastq.gz
WT_GSK_H3K27AC_R2.fastq.gz
insR_DMSO_Input_R1.fastq.gz
insR_DMSO_Input_R2.fastq.gz
insR_DMSO_LSD1_R1.fastq.gz
insR_DMSO_LSD1_R2.fastq.gz
insR_DMSO_GFI1B_R1.fastq.gz
insR_DMSO_GFI1B_R2.fastq.gz
insR_DMSO_H3K4ME2_R1.fastq.gz
insR_DMSO_H3K4ME2_R2.fastq.gz
insR_DMSO_H3K27AC_R1.fastq.gz
insR_DMSO_H3K27AC_R2.fastq.gz
insR_GSK_Input_R1.fastq.gz
insR_GSK_Input_R2.fastq.gz
insR_GSK_LSD1_R1.fastq.gz
insR_GSK_LSD1_R2.fastq.gz
insR_GSK_GFI1B_R1.fastq.gz
insR_GSK_GFI1B_R2.fastq.gz
insR_GSK_H3K4ME2_R1.fastq.gz
insR_GSK_H3K4ME2_R2.fastq.gz
insR_GSK_H3K27AC_R1.fastq.gz
insR_GSK_H3K27AC_R2.fastq.gz
WT_DMSO_rep1.fastq.gz
WT_DMSO_rep2.fastq.gz
WT_DMSO_rep3.fastq.gz
WT_GSK-LSD1_rep1.fastq.gz
WT_GSK-LSD1_rep2.fastq.gz
WT_GSK-LSD1_rep3.fastq.gz
insR_DMSO_rep1.fastq.gz
insR_DMSO_rep2.fastq.gz
insR_DMSO_rep3.fastq.gz
insR_GSK-LSD1_rep1.fastq.gz
insR_GSK-LSD1_rep2.fastq.gz
insR_GSK-LSD1_rep3.fastq.gz

Genome browser session (e.g. UCSC)	https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr21%3A33031597-33041570&hgsid=716459247_GcSJP0LdPKc3mOMzaS6KsH0vAxkM

Methodology

Replicates	Experiment was performed in singlicate.
Sequencing depth	About 20,000,000 paired-end reads of 40 bp in length were obtained per sample. Samples routinely had a mapping efficiency of uniquely mapped reads of > 95%.
Antibodies	AB_2614985 (H3K4me2-human antibody , Active Motif 39141, LOT# 01008001), AB_2561016 (H3K27ac-mouse antibody, Active Motif 39133, LOT# 31814008), AB_242533 (Rabbit Anti-BHC110/KIAA0601 Polyclonal Antibody, Unconjugated, Bethyl Laboratories A300-215A, LOT# 2), AB_449000 (GFI1B antibody, Abcam ab26132, LOT# GR318193-8).
Peak calling parameters	Peaks were called with HOMER (version 4.10) using matched inputs with the following parameters: H3K4me2, –histone –tagThreshold 30; H3K27ac, –histone –tagThreshold 50; LSD1 -style factor -tagThreshold 10 -P 0.000001 -L 5 -F 10; GFI1B -style factor -tagThreshold 10 -P 0.000001.
Data quality	The peak calling parameters are specified above with fold enrichment and P values for each particular modification/factor specified.
Software	Freely available software was used for the ChIP-seq alignment and processing. This includes BWA mem, SAMtools (version 0.1.12) and Picard tools (2.18.5). Homer (version 4.10) was used for the peak calling, peak overlap and motif analysis. Integrated Genome Browser (version 2.4.14) was used for viewing the ChIP-seq tracks. The metaprofile plots for visualizing the ChIP-seq data were generated using ngsplot (version 2.63).

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	For all flow cytometry measurements (viability, fluorescence), 100 µl cells were resuspended in their wells/flasks and sampled directly. To this suspension, when used, Sytox Red Dead Cell Stain was added to a 1X working concentration. For FACS sorting of cells, ~1x10^6 cells were spun and resuspended in cold PBS containing 5% Fetal Bovine Serum (FBS). This suspension was then filtered through FACS tube and kept on ice until sorting. After single cell sorting, plates containing cells were spun to bring all contents to the bottom of the wells and stored in the incubator conditions specified in the manuscript.
Instrument	NovoCyte 2000 flow cytometer
Software	NovoExpress Software
Cell population abundance	Single cells were sorted based on fluorescence (via FITC as described below) and viability (via Styox as indicated above and in the methods section of this manuscript). Purity was determined by follow-up flow cytometry measurements as well as genotyping of the mutagenized gene (described in the methods section of this manuscript).
Gating strategy	First, debris was gated out of our analysis by using FSC-H and SSC-H parameters. Second, viability was determined using FSC-H and APC-A parameters. Third, doublets were excluded using FSC-H and FSC-A parameters. Finally, relative eGFP fluorescence was determined using FITC-A and visualized in a histogram. In cases in which we were studying mutagenized or drug-treated cell lines to, gates were based on wild type controls or controls infected with non-targeting sgRNAs.

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