nature research

Corresponding author(s): Hao Zhu

Last updated by author(s): Jun 12, 2020

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	ChIP-seq and mRNA-seq data were collected using an Illumina NextSeq 500 machine.
Data analysis	Statistical analysis: GraphPad 7.0 ChIP-seq and mRNA-seq analysis: Trimgalore (version 0.4.1), HISAT (version 2.1.0), FeatureCounts (version 1.6.2), DEseq (version 1.36.0), Bowtie2 aligner (version 2.3.4.3), HOMER (version 4.9), edgeR (version 3.22.5), ngs.plot (version 2.63)

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

All sequencing data have been deposited in the GEO with the accession number GSE147664 for mRNA-seq and GSE140183 for ChIP-seq.

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	The data in most panels reflect multiple experiments performed on different days using mice derived from different litters. In all experiments, no mice were excluded from analysis after the experiment was initiated. For histopathological micrographic images, at least five mice were subjected to analysis with similar results. For mRNA seq, four mice were used for each group. For ChIP, at least two independent experiments were performed with similar results. All the western blot experiments were performed at least twice independently with similar results.
Data exclusions	No data were excluded from the analyses.
Replication	All experiments were performed by at least 2 biologically independent replicates with similar results. No data were excluded.
Randomization	Animals were first separated into groups by genotype. Animals having the same genotype were randomly assigned to experimental and control groups when applicable.
Blinding	Image analysis for the quantification was blinded.

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

ARID1A (Sigma-Aldrich #HPA005456, western blot 1:2000 and IHC 1:1000),
ARID1A (Santa Cruz #sc-32761, western blot 1:1000),
ARID1B (Abcam #AB57461, western blot 1:2000 and IHC 1:500),
ARID2 (Abiocode #R2380-1, western blot 1:2000),
Brg1 (Santa Cruz #sc-10768, western blot 1:1000),
BAF170 (Santa Cruz #sc-17838, western blot 1:1000),
BAF155 (Santa Cruz #sc-9746, western blot 1:1000),
BAF60b (Santa Cruz #sc-101162, western blot 1:1000),
BAF57 (Bethyl #A300-810A, western blot 1:2000),
BAF53 (Santa Cruz #sc-137063, western blot 1:1000),
BAF47 (Santa Cruz #sc-166165, western blot 1:1000),
BAF47 (CST #91735, western blot 1:2000),
BAF45d (Santa Cruz #sc-101106, western blot 1:1000),
Ki67 (Abcam #AB15580, IHC 1:200),
HNF4A (CST #3113, IHC 1:200),
CK19 (Abcam #AB15463, IHC 1:200),
EpCAM (CST #14452, IHC 1:200),
Flag M2 Affinity Gel (Sigma-Aldrich #A2220, IP),
DYKDDDDK Tag (Flag) (CST #2368, western blot 1:2000),
Ty1 (Diagenode #C15200054, western blot 1:5000 and ChIP 5ug/sample)

Validation	All the antibodies are commercial antibodies. Statements regarding validation can be found at the manufactures website.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK-293T, ATCC H2.35, ATCC MFE-296, ECACC
Authentication	All cell lines were authenticated by Short Tandem Repeat (STR) profiling.
Mycoplasma contamination	All cell lines were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used in this study are in the misidentified cell line list.

Animals and other organisms

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

Laboratory animals	All ARID1A f/f; ARID1B f/f mice (with/without Alb-Cre or UBC-CreER) were C57BL/6 background and both male and female mice were used in the study with indistinguishable tumor formation ability. For mRNA-seq analysis and western blot on liver lysates, 2-month old female mice were used. For transplantation models, female NSG mice at 3-month old were used. Mice were housed under 12-hour light/dark cycle with controlled temperature of 74-75 Fahrenheit and 51%-59% humidity.
Wild animals	No wild animals were used in the study.
Field-collected samples	No field collected samples were used in the study.
Ethics oversight	All mice were handled in accordance with the guidelines of the Institutional Animal Care and Use Committee at University of Texas, Southwestern Medical Center.

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

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=GSE140183
Files in database submission	GSM4156259 ChIP-seq of ARID1A in WT H2.35 cells [WT-ARID1A-1] GSM4156260 ChIP-seq of ARID1A in WT H2.35 cells [WT-ARID1A-2] GSM4156261 ChIP-seq of ARID1B in WT H2.35 cells [WT-ARID1B-1] GSM4156262 ChIP-seq of ARID1B in WT H2.35 cells [WT-ARID1B-2] GSM4156263 ChIP-seq of Brg1 in WT H2.35 cells [WT-Brg1-1] GSM4156264 ChIP-seq of Brg1 in WT H2.35 cells [WT-Brg1-2] GSM4156265 ChIP-seq of BAF45d in WT H2.35 cells [WT-BAF45d-1] GSM4156266 ChIP-seq of BAF45d in WT H2.35 cells [WT-BAF45d-2] GSM4156267 ChIP-seq of ARID2 in WT H2.35 cells [WT-ARID2-1] GSM4156268 ChIP-seq of ARID2 in WT H2.35 cells [WT-ARID2-2] GSM4156269 ChIP-seq of Brd9 in WT H2.35 cells [WT-Brd9-1] GSM4156270 ChIP-seq of Brd9 in WT H2.35 cells [WT-Brd9-2] GSM4156271 ChIP-seq of BAF45d in ARID1-less H2.35 cells [ARID1-less-BAF45d-1] GSM4156272 ChIP-seq of BAF45d in ARID1-less H2.35 cells [ARID1-less-BAF45d-2] GSM4156273 ChIP-seq of ARID2 in ARID1-less H2.35 cells [ARID1-less-ARID2-1] GSM4156274 ChIP-seq of ARID2 in ARID1-less H2.35 cells [ARID1-less-ARID2-2] GSM4156275 ChIP-seq of ARID2 in ARID1-less H2.35 cells [ARID1-less-ARID2-3] GSM4156276 ChIP-seq of ARID2 in ARID1-less H2.35 cells [ARID1-less-ARID2-4] GSM4156277 ChIP-seq of Brd9 in ARID1-less H2.35 cells [ARID1-less-Brd9-1] GSM4156278 ChIP-seq of Brd9 in ARID1-less H2.35 cells [ARID1-less-Brd9-2] GSM4156279 Input of WT H2.35 cells GSM4156280 Input of ARID1-less H2.35 cells
Genome browser session (e.g. UCSC)	n/a

Methodology

Replicates	All ChIP-seq samples were performed by 2 replicates. ChIP-seq samples of Ty1 tagged ARID2 from ARID1-less H2.35 cells were performed by 4 replicates.
Sequencing depth	All ChIP-Seq samples were single-end sequenced, the raw and mapped read number for each are below: Sample Name Raw Reads Mapped.Read ARID1-less-BAF45d-1 11580476 10084278 ARID1-less-BAF45d-2 11785025 11452687 ARID1-less-ARID2-1 13393829 13024159 ARID1-less-ARID2-2 14530931 14215609 ARID1-less-ARID2-3 14058780 13679192 ARID1-less-ARID2-4 14829568 14411374 ARID1-less-Brd9-1 9819002 9516576 ARID1-less-Brd9-2 10731979 10421824 WT-ARID1A-1 12550861 12195671 WT-ARID1A-2 12043864 11591014 WT-ARID1B-1 13601268 13225873 WT-ARID1B-2 14050779 13713560 WT-BAF45d-1 12412654 12091166 WT-BAF45d-2 11841114 11522588 WT-ARID2-1 14045067 13581579 WT-ARID2-2 15492821 14947473 WT-Brd9-1 10385354 9992787 WT-Brd9-2 11270514 10853504 WT-Brg1-1 14356402 13942937 WT-Brg1-2 14500300 14127642 WT-input 11410521 11226811 ARID1-less-input 11450889 11265384
Antibodies	Ty1 (Diagenode #C15200054), 5ug antibody for each ChIP sample. 1 x 10^7 cells were used for each sample prepared at a final volume of 0.5 ml.
Peak calling parameters	Single-end sequencing reads were aligned to the mouse reference genome GRCm38 (mm10) using bowtie2 aligner (v 2.3.4.3) with default parameters. Peak calling was performed using findpeaks command from HOMER software package version 4.9 with parameter –style histone. Peaks were called when enriched > 2 fold over input controls and > 4 fold over local tag counts, with FDR < 10e-5. ChIP-seq peaks within a 1000 bp range were stitched together to form ChIP-seq regions. To restrict analysis on stringent peaks, peaks were further filtered and peaks with the read tag numbers > ~20 within peaks were kept for analysis.
Data quality	Excepting particular experimental conditions (such as the decreased binding of cBAF and pBAF complexes in ARID1-less cells), at threshold FDR < 1e-3, all samples > 6000 peaks.
Software	Single-end sequencing reads were aligned to the mouse reference genome using bowtie2 aligner (v 2.3.4.3). Peak calling was performed using findpeaks command from HOMER software package (v 4.9). Peaks were annotated by mapping to nearest TSS using annotatePeaks.pl. Shared and unique peaks were identified by mergePeaks. Heatmaps of ChIP-seq signal over input sample were plot using ngs.plot.