nature portfolio

Corresponding author(s): Hao Jiang

Last updated by author(s): Jul 26, 2021

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	UTX mutations in human cancer were analyzed by cBioPortal (http://www.cbioportal.org/).
Data analysis	Disordered regions were identified using IUPred and IUPred3 (http://iupred.elte.hu/). Amino acid composition was analyzed by Composition Profiler (http://www.cprofiler.org/cgi-bin/profiler.cgi). Net Charge Per Residue was analyzed by CIDER (http://pappulab.wustl.edu/CIDER/analysis/). Protein concentration was determined by Nanodrop measurement for OD280 and calculation using extinction coefficient provided by ExPASy ProtParam (https://web.expasy.org/protparam/). RNA-seq analyses used STAR v2.6.0a, RSEM v1.3.1, DESeq2 v1.22.2, DAVID v6.8, Metascape v3.5. PRO-seq analyses used Trim Galore v0.5.0, BWA v0.7.12-r1044, Samtools v1.7, MACS2 v2.1.1.20160309, BEDtools v2.26.0. ChIP-seq analyses used Trim Galore v0.5.0, BWA v0.7.12-r1044, Samtools v1.7, MACS2 v2.1.1.20160309, SICER v2.0. HiChIP analyses used MAPS v1.1.0, HiC-Pro v2.10.0, BART3D v1.1. Details for the genomic data analyses can be found in Methods. Imaging analyses involved: ImageJ 1.53e, ZEN v2.6, SimFCS v4.0.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The high-throughput sequencing data, including RNA-seq, ChIP-seq, PRO-seq, and HiChIP, have been deposited in Gene Expression Omnibus database with the accession number GSE149420. The scripts used to analyze the data from this study are freely available at: https://github.com/zanglab/utx_code. Cancer mutation data are available from cBioPortal (http://www.cbioportal.org/) and TCGA database.

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 estimated by robustness of phenotype based on our preliminary experiments and previous literature reporting [Nature Cell Biology 22, 960-972 (2020)] number that are most optimal to generate statistically significant results. No statistical method was applied to predetermine sample size.
Data exclusions	No data exclusions occurred in this study.
Replication	All in vivo experiments were performed on at least 3 mice. All key in vitro experiments were perfomed at least 3 times. All replication attempts were successful.
Randomization	Animals and cells were randomly allocated into control and experimental groups. All samples used in each set of experiments were equal, except the experimental condition being tested.
Blinding	Investigators were blinded to sample labeling and group allocation during data collections of most experiments until final data analyses.

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 and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Rabbit monoclonal anti-UTX (D3Q1I), Cell Signaling 33510, RRID:AB_2721244. Rabbit polyclonal anti-MLL4, Sigma HPA035977, RRID:AB_10670673 (immunoblotting); Santa Cruz Biotec. sc-293217, RRID:AB_2687515 (ChIP); Rabbit polyclonal anti-MLL3, a gift from Kai Ge; Mouse monoclonal anti-GAPDH clone 6C5, Millipore MAB374, RRID:AB_2107445; Rabbit monoclonal anti-H3K4me3 clone 15-10C-E4, Millipore 05-745R, RRID:AB_1587134; Rabbit polyclonal anti-H3K27me3, Millipore 07-449, RRID:AB_310624; anti-H3K27ac, Abcam ab4729, RRID:AB_2118291; anti-H3K4me2, Millipore 07-030, RRID:AB_310342; anti-H3K4me1, Abcam ab8895, RRID:AB_306847; Rabbit polyclonal anti-RBBP5 and -WDR5, Bethyl Laboratory A300-109A (RRID:AB_210551) and A302-429A (RRID:AB_1944302), respectively; Mouse monoclonal anti-c-Myc (9E10), ThermoFisher Scientific, 132500, RRID:AB_2533008; Mouse

monoclonal anti-FLAG M2 Affinity agarose, Sigma, A2220, RRID: AB_10063035; Alexa Fluor 555 conjugated goat anti-rabbit IgG, ThermoFisher Scientific A-21428, RRID:AB_141784; Alexa Fluor 488 conjugated goat anti-mouse IgG, ThermoFisher Scientific A-11001, RRID:AB_2534069.
All antibodies were diluted 1:1000 in immunoblotting. 5 ug antibody for each reaction was used for UTX ChIP and 3 ug antibody for other ChIPs.

Validation

All these antibodies (except anti-MLL3) were commercially obtained and validated by vendors and multiple published studies, see manufacture's website for references. In addition, for key antibodies such as UTX, the specificity is validated by the lack of signal in our KO cell samples. We are also listing their Research Resource Identifiers here for authentication.
Rabbit monoclonal anti-UTX, Cell Signaling 33510, RRID:AB_2721244. Rabbit polyclonal anti-MLL4, Sigma HPA035977, RRID:AB_10670673 (immunoblotting); Santa Cruz Biotec. sc-293217, RRID:AB_2687515 (ChIP); Rabbit polyclonal anti-MLL3, a gift from Kai Ge; Mouse monoclonal anti-GAPDH, Millipore MAB374, RRID:AB_2107445; Rabbit monoclonal anti-H3K4me3, Millipore 05-745R, RRID:AB_1587134; Rabbit polyclonal anti-H3K27me3, Millipore 07-449, RRID:AB_310624; anti-H3K27ac, Abcam ab4729, RRID:AB_2118291; anti-H3K4me2, Millipore 07-030, RRID:AB_310342; anti-H3K4me1, Abcam ab8895, RRID:AB_306847; Rabbit polyclonal anti-RBBP5 and -WDR5, Bethyl Laboratory A300-109A (RRID:AB_210551) and A302-429A (RRID:AB_1944302), respectively; Mouse monoclonal anti-c-Myc (9E10), ThermoFisher Scientific, 132500, RRID:AB_2533008; Mouse monoclonal anti-FLAG M2 Affinity agarose, Sigma, A2220, RRID: AB_10063035;
Rabbit monoclonal anti-UTX, Cell Signaling 33510, vendor information indicates that the immunogen is a recombinant protein surrounding Ala490 of human UTX, and our data confirmed that it reacts with the UTX region between 419 and 548. Anti-MLL3, a gift from Kai Ge, was validated by multiple publications with assays including pulldown by MLL3/4 core subunits and MLL3 KO cells in PMID: 17500065, PMID: 30335158, and PMID: 28013028.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

THP-1 cells were a gift from Jianjun Chen at Universit of Chicago then. MiaPaca2, COS-7, and LacO array-containing U2OS (U2OS-LacO-TRE) cells were kind gifts from Kimberly Kelly, Jung-Bun Shin, Todd Stukenburg, respectively, all at the University of Virginia. The mouse ES cell E14TG2A was from ATCC. Flp-In T-REx 293 cells were from ThermoFisher Scientific. Other cells were from ATCC: HeLa (ATCC CCL-2), HEK 293 (ATCC CRL-1573), 239T (ATCC CRL-3216).

Authentication

Certificate of Analysis of these cell lines were provided by the vendors. Cells were cultured in media recommended by the vendors. We froze down stocks upon receiving the cell lines, and all experiments will be conducted on cells that have been passaged no more than 10 times. For the genome edited ES cells, we have performed 3 different levels of identifications: genotyping by PCR and sequencing on the extracted genomic DNA and immunoblotting, and the edited cells were indicated by an unambiguous distinct PCR band, correct sequencing results, and unique signal in immunoblotting.

Mycoplasma contamination

Cells were examined every half-1 year for possible mycoplasma contamination using the Universal Mycoplasma Detection Kit from ATCC (ATCC® 30-1012K™) and were found negative.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used.

Animals and other organisms

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

Laboratory animals

8-week-old female immune-deficient NSG (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ) mice were purchased from The Jackson Laboratory.

Wild animals

No involved.

Field-collected samples

No involved.

Ethics oversight

All animal procedures were approved by the Institutional Animal Care and Use Committee at the University of Virginia.

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=GSE149420

Files in database submission

GSM5379642 RNA-seq_Vector_rep1
GSM5379643 RNA-seq_Vector_rep2
GSM5379644 RNA-seq_WT_rep1
GSM5379645 RNA-seq_WT_rep2
GSM5379646 RNA-seq_del-cIDR_rep1
GSM5379647 RNA-seq_del-cIDR_rep2

GSM5379648 RNA-seq_eIFIDR_rep1
GSM5379649 RNA-seq_eIFIDR_rep2
GSM5379650 RNA-seq_del-TRP_rep1
GSM5379651 RNA-seq_del-TRP_rep2
GSM5379652 RNA-seq_FUSIDR_rep1
GSM5379653 RNA-seq_FUSIDR_rep2
GSM5379654 PRO-seq_Vector_rep1
GSM5379655 PRO-seq_Vector_rep2
GSM5379656 PRO-seq_WT_rep1
GSM5379657 PRO-seq_WT_rep2
GSM5379658 PRO-seq_del-cIDR_rep1
GSM5379659 PRO-seq_del-cIDR_rep2
GSM5379660 ChIP-seq_UTX_Vector
GSM5379661 ChIP-seq_UTX_WT
GSM5379662 ChIP-seq_UTX_del-cIDR
GSM5379663 ChIP-seq_H3K4me1_Vector
GSM5379664 ChIP-seq_H3K4me1_WT
GSM5379665 ChIP-seq_H3K4me1_del-cIDR
GSM5379666 ChIP-seq_H3K4me1_MT2
GSM5379667 ChIP-seq_H3K4me2_Vector
GSM5379668 ChIP-seq_H3K4me2_WT
GSM5379669 ChIP-seq_H3K4me2_del-cIDR
GSM5379670 ChIP-seq_H3K4me3_Vector
GSM5379671 ChIP-seq_H3K4me3_WT
GSM5379672 ChIP-seq_H3K4me3_del-cIDR
GSM5379673 ChIP-seq_H3K27ac_Vector
GSM5379674 ChIP-seq_H3K27ac_WT
GSM5379675 ChIP-seq_H3K27ac_del-cIDR
GSM5379676 ChIP-seq_H3K27ac_MT2
GSM5379677 ChIP-seq_H3K27me3_Vector
GSM5379678 ChIP-seq_H3K27me3_WT
GSM5379679 ChIP-seq_H3K27me3_del-cIDR
GSM5379680 ChIP-seq_H3K27me3_MT2
GSM5379681 ChIP-seq_MLL4_Vector
GSM5379682 ChIP-seq_MLL4_WT
GSM5379683 ChIP-seq_MLL4_del-cIDR
GSM5379684 HiChIP_H3K4me3_batch1_Vector_rep1
GSM5379685 HiChIP_H3K4me3_batch1_Vector_rep2
GSM5379686 HiChIP_H3K4me3_batch1_WT_rep1
GSM5379687 HiChIP_H3K4me3_batch1_WT_rep2
GSM5379688 HiChIP_H3K4me3_batch1_del-cIDR_rep1
GSM5379689 HiChIP_H3K4me3_batch1_del-cIDR_rep2
GSM5379690 HiChIP_H3K4me3_batch1_eIFIDR_rep1
GSM5379691 HiChIP_H3K4me3_batch1_eIFIDR_rep2
GSM5379692 HiChIP_H3K4me3_batch2_Vector
GSM5379693 HiChIP_H3K4me3_batch2_WT
GSM5379694 HiChIP_H3K4me3_batch2_del-cIDR
GSM5379695 HiChIP_H3K4me3_batch2_eIFIDR
GSM5379696 HiChIP_H3K4me3_batch2_del-TRP
GSM5379697 HiChIP_H3K27ac_Vector
GSM5379698 HiChIP_H3K27ac_WT
GSM5379699 HiChIP_H3K27ac_del-cIDR
GSM5379700 HiChIP_H3K27ac_eIFIDR
GSM5379701 HiChIP_H3K27ac_del-TRP
GSM5379702 RNA-seq_batch2_Vector_rep1
GSM5379703 RNA-seq_batch2_Vector_rep2
GSM5379704 RNA-seq_batch2_WT_rep1
GSM5379705 RNA-seq_batch2_WT_rep2
GSM5379706 RNA-seq_batch2_MT2_rep1
GSM5379707 RNA-seq_batch2_MT2_rep2
GSM5379708 RNA-seq_mouse_ESCs_WT
GSM5379709 RNA-seq_mouse_ESCs_UTX-KO
GSM5379710 RNA-seq_mouse_ESCs_del-cIDR_clone3
GSM5379711 RNA-seq_mouse_ESCs_del-cIDR_clone72
GSM5379712 RNA-seq_mouse_EBs_WT_rep1
GSM5379713 RNA-seq_mouse_EBs_WT_rep2
GSM5379714 RNA-seq_mouse_EBs_UTX-KO_rep1
GSM5379715 RNA-seq_mouse_EBs_UTX-KO_rep2
GSM5379716 RNA-seq_mouse_EBs_del-cIDR_clone3_rep1
GSM5379717 RNA-seq_mouse_EBs_del-cIDR_clone3_rep2
GSM5379718 RNA-seq_mouse_EBs_del-cIDR_clone72_rep1
GSM5379719 RNA-seq_mouse_EBs_del-cIDR_clone72_rep2

Genome browser session
(e.g. UCSC)

No longer applicable

Methodology

Replicates	Two or more (up to 5) biological replicates for all next-generation sequencing assays. Most replication attempts were successful.
Sequencing depth	Experiment total reads mapped reads non-redundant reads paired or single ChIP-seq_UTX_Vector 60,039,780 51,667,370 50,443,652 single-end ChIP-seq_UTX_WT 32,869,684 27,105,840 26,059,754 single-end ChIP-seq_UTX_del-cIDR 24,704,604 20,210,909 19,496,843 single-end ChIP-seq_H3K4me1_Vector 22,126,371 17,151,882 16,415,263 single-end ChIP-seq_H3K4me1_WT 24,847,170 19,542,004 18,825,638 single-end ChIP-seq_H3K4me1_del-cIDR 20,566,453 16,517,751 15,959,996 single-end ChIP-seq_H3K4me1_MT2 24,611,924 19,676,378 18,888,479 single-end ChIP-seq_H3K4me2_Vector 54,725,568 49,406,192 47,567,395 single-end ChIP-seq_H3K4me2_WT 36,160,877 32,711,838 31,766,711 single-end ChIP-seq_H3K4me2_del-cIDR 41,979,976 37,620,311 34,762,229 single-end ChIP-seq_H3K4me3_Vector 20,388,084 15,750,172 14,361,567 single-end ChIP-seq_H3K4me3_WT 22,131,736 16,807,584 15,314,593 single-end ChIP-seq_H3K4me3_del-cIDR 20,577,428 15,826,777 14,577,146 single-end ChIP-seq_H3K27ac_Vector 30,386,722 24,727,030 23,895,111 single-end ChIP-seq_H3K27ac_WT 25,135,281 19,508,741 18,719,219 single-end ChIP-seq_H3K27ac_del-cIDR 28,464,969 22,593,986 21,732,386 single-end ChIP-seq_H3K27ac_MT2 22,885,124 17,910,567 17,145,787 single-end ChIP-seq_H3K27me3_Vector 28,543,343 21,650,388 20,266,162 single-end ChIP-seq_H3K27me3_WT 24,384,642 18,648,852 17,419,081 single-end ChIP-seq_H3K27me3_del-cIDR 19,920,536 14,842,151 13,875,471 single-end ChIP-seq_H3K27me3_MT2 18,804,994 14,280,221 13,291,247 single-end ChIP-seq_MLL4_Vector 22,105,116 16,393,822 15,102,111 single-end ChIP-seq_MLL4_WT 19,852,527 13,631,751 12,455,895 single-end ChIP-seq_MLL4_del-cIDR 20,728,861 16,054,217 14,908,056 single-end
Antibodies	Rabbit monoclonal anti-UTX, Cell Signaling 33510, RRID:AB_2721244, Rabbit polyclonal anti-MLL4, Santa Cruz Biotec. sc-293217, RRID:AB_2687515, Rabbit monoclonal anti-H3K4me3, Millipore 05-745R, RRID:AB_1587134; Rabbit polyclonal anti-H3K27me3, Millipore 07-449, RRID:AB_310624; anti-H3K27ac, Abcam ab4729, RRID:AB_2118291; anti-H3K4me2, Millipore 07-030, RRID:AB_310342; anti-H3K4me1, Abcam ab8895, RRID:AB_306847.
Peak calling parameters	macs2 callpeak --SPMR -B -q 0.1 --keep-dup 1 -g hs
Data quality	mapping rate>65%, number of peaks>2000, FRiP>0.01
Software	Trim Galore, BWA, Samtools, MACS2. Also see Methods.