### nature portfolio

Corresponding author(s): Bingtao Hao and Shixiu Liao

Last updated by author(s): Aug 23, 2022

#### **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 <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

#### Statistics

For	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			
	X	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.		
X		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. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted Give <i>P</i> values as exact values whenever suitable.		
×		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings		
X		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 <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated		
		Our web collection on statistics for biologists contains articles on many of the points above.		

#### Software and code

Data collection	ggplot2 (3.3.5) (https://github.com/tidyverse/ggplot2) pheatmap (v1.0.12) (https://github.com/raivokolde/pheatmap) ngsplot (v2.41) (https://code.google.com/p/ngsplot/) CellQuest Pro software (BD Biosciences)
Data analysis	Call Pangar (v2.0.2) /https://github.com/10VConomics/callranger)
Data analysis	Secret (v/, 0, 5) (https://github.com/ toxogenomics/cellianger)
	Cutadant (v1.18) (https://github.com/marcelm/cutadant)
	Hisat2 (v2.2.1) (https://github.com/DaehwanKimLab/hisat2)
	deeptools (v3.5.0) (https://deeptools.readthedocs.io/en/develop).
	DESeq2 (v1.30.0) (https://github.com/mikelove/DESeq2)
	Bowtie2 (v2.3.5.1) (https://github.com/BenLangmead/bowtie2)
	Picard (v1.118) (http://broadinstitute.github.io/picard).
	Homer (v4.10.4) (https://github.com/bastienwirtz/homer).
	bedtools (v2.29.2) (https://github.com/arq5x/bedtools).
	deepTools (v3.5.0) (https://github.com/deeptools/deepTools)
	DESeq2 (v1.30.0) (https://github.com/mikelove/DESeq2)
	ChIPSeeker(v1.26.0) (https://github.com/YuLab-SMU/ChIPseeker).
	the rank ordering of super-enhancers (ROSE) algorithm (https://github.com/stjude/ROSE).
	fastp (v0.20.0) (https://github.com/OpenGene/fastp).
	HiC-Pro (v2.11.1) (https://github.com/nservant/HiC-Pro).
	Juicerbox (v1.5.1) (https://github.com/aidenlab/Juicebox)

HiTC (v1.34.0) (https://github.com/bioinfo-pf-curie/HiTC).
matrix2compartment.pl (https://github.com/dekkerlab/cworld-dekker).
multilHiCcompare (v1.8.0) (https://github.com/dozmorovlab/multiHiCcompare)
Fit-Hi-C (v2.0.7) (https://github.com/ay-lab/fithic)
pgltools (v2.2.0) (https://github.com/billgreenwald/pgltools)
hic2cool (v0.8.3) (https://github.com/4dn-dcic/hic2cool).
coolpup.py (v0.9.5) (https://github.com/open2c/coolpuppy).
Pear (v0.9.6) (https://github.com/tseemann/PEAR)
MiXCR (v3.0.11) (https://github.com/milaboratory/mixcr)

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 <u>data availability statement</u>. 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 raw sequence data of single-cell RNA-Seq, bulk RNA-seq, ChIP-seq, Hi-C, 4C-seq, 3C-HTGTS, and 5'RACE reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database under the accession number: GSE182995 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE182995]. GEO accession codes (or SRA accession number) of the published data used in this study are as follows: H3K4me3 ChIP-seq of CD4+ CD8+ DP thymocytes GSE21207 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE21207]; Rad21 and Nipbl ChIP-seq, GSE48763 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE48763]; CTCF ChIP-seq, GSE141223 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE141223]; SATB1 ChIP-seq, GSE90635 [https:// www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE109125]; ATAC-seq of T cell development, GSE109125 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE109125]; ATAC-seq of T cell development, GSE109125 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE109125]; ATAC-seq of T cell development, GSE109125 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE109125]; H3K27ac and H3K4me1 ChIP-seq of Satb1WT and Satb1cKO DP thymocytes, DRP003376 [https://trace.ncbi.nlm.nih.gov/Traces/index.html?view=study&acc=DRP003376]; 3e Hi-C data, GSE79422 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE79422].

#### 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

ences 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	No sample size calculations were performed and number of replicates were based on standard practices in the field. Ji, Y. et al. Cell 141, 419-431 (2010). Guo, J. et al. Nat Immunol 3, 469-476 (2002). Kitagawa, Y. et al. Nat Immunol 18, 173-183 (2017).
Data exclusions	No data was excluded
Replication	Bulk RNA-seq data was acquired in biological triplicate experiments. Independent biological duplicate experiments were performed for Hi-C, 3C-HTGTs, 4C assays. When we acquired scRNA-seq, these were not done with replicates because scRNA-seq contains several thousand RNA-seq assays, it provides enough information, on the other hand, the cost is too high, and it is necessary to balance cost and performance. Ets2-SE 5'RACE has two biological replicates. Bcl6-SE 5'RACE has no biological replicates because we noticed that the usages of Va and Ja in Bcl6-SE-/- mice showed continuous changes that could not be due to individual differences, so one biological experiment was sufficient. For ChIP-seq analysis, we used published sequencing data that had not been acquired in replicates.
Randomization	Animal experiments were based on genotypes.
Blinding	Investigators were not blinded during experiments, because no subjective process is included in all the analyses of the experimental data in this study. Analyses of temporal profiles of mouse tissues were performed blindly.

#### Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the

Non-participation State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.

rationale behind them, indicating whether exclusion criteria were pre-established.

Randomization If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

#### Ecological, evolutionary & environmental sciences study design

#### All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.
Research sample	Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? Yes X No

#### Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access & import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

# nature portfolio | reporting summary

#### 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		Methods	
n/a	Involved in the study	n/a	Involved in the study
	X Antibodies		X ChIP-seq
×	Eukaryotic cell lines		<b>X</b> Flow cytometry
×	Palaeontology and archaeology	×	MRI-based neuroimaging
	X Animals and other organisms		
×	Human research participants		
×	Clinical data		
×	Dual use research of concern		

#### Antibodies

Antibodies used	Antibodies used for FACS: anti-CD3-EFLUOR 450 (eBioscience 48003280, Clone 17A2), anti-CD25-PE (Biolegend 102008, Clone PC61), anti-CD8a-FITC (Biolegend 100706, Clone 53-6.7), anti-CD4-APC (Biolegend 100516, Clone RM4-5), anti-CD44- PE/Cy5 (Biolegend 103010, Clone IM7). Antibodies for Western blot: anti-SATB1 (Rabbit monoclonal [EPR3951], abcam, ab109122), anti-Bcl6 (Mouse monoclonal [Gl191E], abcam, ab241549), anti-ETS2 (Rabbit monoclonal [EPR22419], abcam, ab219948), anti-beta-actin (Mouse polyclonal, FD0060, Fdbio science, China) and anti- GAPDH (Mouse polyclonal, FD0063, Fdbio science, China).
Validation	All the antibodies used were validated by the manufacturers and checked by us by FACS or Western blot. Validation statements on themanufacturer's website for all primary antibodies used in this study were as follows: CD3- EFLUOR 450 (eBioscience 48003280, Clone: 17A2): mouse(+); CD25-PE (Biolegend 102008, Clone: PC61): mouse(+); CD8a-FITC (Biolegend 100706, Clone: 53-6.7): mouse(+); CD4-APC ( Biolegend 100516, Clone:RM4-5): mouse(+); CD44-PE/Cy5 (Biolegend 103010, Clone:IM7): mouse(+). anti-SATB1 (Rabbit monoclonal [EPR3951], abcam, ab109122): Mouse(+), Rat(+), Human(+) anti-Bcl6 (Mouse monoclonal [GI191E], abcam, ab241549): Mouse(+), Rat(+), Human(+) anti-ETS2 (Rabbit monoclonal [EPR22419], abcam, ab219948):Mouse(+), Rat(+), Human(+) anti-beta-actin (Mouse polyclonal, FD0060, Fdbio science, China): Mouse(+), Rat(+), Human(+) anti-GAPDH (Mouse polyclonal, FD0063, Fdbio science, China): Mouse(+), Rat(+), Human(+)

#### Eukaryotic cell lines

Policy information about <u>cell lines</u>		
Cell line source(s)	State the source of each cell line used.	
Authentication	Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.	
Mycoplasma contamination	Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.	
Commonly misidentified lines (See <u>ICLAC</u> register)	Name any commonly misidentified cell lines used in the study and provide a rationale for their use.	

#### Palaeontology and Archaeology

Specimen provenance	Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.
Specimen deposition	Indicate where the specimens have been deposited to permit free access by other researchers.
Dating methods	If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

] Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

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

#### Animals and other organisms

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

Laboratory animals	Male and female Satb1fl/fl vav-cre+ mice; Bcl6 SE-/-, Ets2 SE-/- mice were used in this study. Sex-matched littermates were used for all experiments. All experiments involving mice were performed using protocols approved by Southern Medical University Animal Studies Committee. Animals were housed and bred in a specific pathogen-free animal facility. The housing condition is a light-tight chamber at a constant temperature (23±1°C), humidity (55±10%), and 12-hr light/12-hr dark (LD) cycles.
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 experiments involving mice were performed using protocols approved by Southern Medical University Animal Studies Committee. Animals were housed and bred in a specific pathogen-free animal facility.

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 Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above." Recruitment Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results. Ethics oversight Identify the organization(s) that approved the study protocol.

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

#### Clinical data

Policy information about clinical studies All manuscripts should comply with the ICMJEguidelines for publication of clinical research and a completed<u>CONSORT checklist</u> must be included with all submissions.

Clinical trial registration	Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.
Study protocol	Note where the full trial protocol can be accessed OR if not available, explain why.
Data collection	Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.
Outcomes	Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

#### Dual use research of concern

Policy information about dual use research of concern

#### Hazards

L

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
	Public health
	National security
	Crops and/or livestock
	Ecosystems
	Any other significant area

#### Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
	Demonstrate how to render a vaccine ineffective
	Confer resistance to therapeutically useful antibiotics or antiviral agents
	Enhance the virulence of a pathogen or render a nonpathogen virulent
	Increase transmissibility of a pathogen
	Alter the host range of a pathogen
	Enable evasion of diagnostic/detection modalities
	Enable the weaponization of a biological agent or toxin
	Any other potentially harmful combination of experiments and agents

#### ChIP-seq

#### Data deposition

	Confirm that both raw and final	processed data ha	ive 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 public	For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.	
Files in database submissi	Provide a list of all files available in the database submission.	
Genome browser session (e.g. <u>UCSC</u> )	Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.	
Methodology		
Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.	
Sequencing depth		
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and number.	
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.	
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.	
Software	Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community	

#### Flow Cytometry

#### Plots

Confirm that:

**X** The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

**x** 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).

**X** All plots are contour plots with outliers or pseudocolor plots.

**x** A numerical value for number of cells or percentage (with statistics) is provided.

repository, provide accession details.

#### Methodology

Sample preparation

Thymus, spleen, mesenteric lymph nodes, inguinal lymph nodes, and auxiliary lymph nodes from 6-8 weeks mice were ground in MACS buffer (1×PBS, 0.5% BSA, 2mM EDTA) and filtered with 40 um nylon mesh. Red blood cells were lysed in RBC buffer (Biolegend, USA) for 10 minutes at room temperature.

Instrument	Flow cytometry was carried out on a BD FACS LSR Fortessa, or a FACS Aria3.
Software	CellQuest
Cell population abundance	Purity of sorted CD4+CD8+ DP cell populations was over 95% as checked by flow cytometry
Gating strategy	For all experiments, cells were first gated by FSC/SSC to exclude debris, followed by gating FSC-A and FSC-H to eliminate nonsinglets. Then, target cell population for further analysis were gated by cell surface marker.

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

#### Magnetic resonance imaging

#### Experimental design Indicate task or resting state; event-related or block design. Design type Design specifications Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials. State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used Behavioral performance measures to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects). Acquisition Specify: functional, structural, diffusion, perfusion. Imaging type(s) Field strength Specify in Tesla Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, Sequence & imaging parameters slice thickness, orientation and TE/TR/flip angle. State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined. Area of acquisition Diffusion MRI Used Not used Preprocessing Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, Preprocessing software segmentation, smoothing kernel size, etc.). Normalization If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization. Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. Normalization template original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized. Noise and artifact removal Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration). Volume censoring Define your software and/or method and criteria for volume censoring, and state the extent of such censoring. Statistical modeling & inference Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and Model type and settings second levels (e.g. fixed, random or mixed effects; drift or auto-correlation). Effect(s) tested Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used. Specify type of analysis: ROI-based Both Whole brain Statistic type for inference Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods. (See Eklund et al. 2016) Correction Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

#### Models & analysis

n/a       Involved in the study         Image: State of the study       Functional and/or effective connectivity         Image: State of the study       Graph analysis         Image: State of the study       Multivariate modeling or predictive analysis				
Functional and/or effective connectivity	Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).			
Graph analysis	Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).			
Multivariate modeling and predictive analysis	Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.			