nature research

Corresponding author(s): Stuart H. Orkin

Last updated by author(s): Jan 14, 2021

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 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	Raw data of high throughput sequencing were collect on Illumina Nextseq or Novaseq platforms with the default factory softwares.
Data analysis	For dCas9 and Cas9 dense perturbation analysis, these softwares were used: DESeq2, R 4.0.1, CRISPR-SURF 1.0. For ChIP-seq, CUT&RUN, ATAC-seq analysis, FastQC 0.11.9, CUT&RUNTools, MAnorm 1.1.4 and deeptools 2.0 were used. CUT&RUNTools integrates the following softwares: Trimmomatic 0.36, Bowtie 2.2.9, Samtools 1.3.1, Picard 2.8.0, MACS 2.1.1, MEME 4.12.0, Bedops 2.4.30, Bedtools 2.26.0. Statistic analysis of qPCR was carried out with GraphPad Prism 8. For PRO-seq analysis, Cutadapt 1.14, Bowtie 1.2.2, Samtools 1.3.1 were used. Custom scripts were used to quantify signals on each gene (https://github.com/yao-qiuming/Nan_NG2020). For quantification of Cas9 editing or base editing efficiency, TIDE 3.2.0 and TIDER 1.0.2 were used. For flow cytometry analysis, FlowJo V10 was used. Western blot was quantified with ImageJ 2.0 Genome tracks were viewed using IGV 2.7.0 Protein Structure was visualized with PyMOL 1.8.6.1

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 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 raw and processed CRISPR screen, CUT&RUN, ChIP-seq, PRO-seq and ATAC-seq data have been deposited in the NCBI Gene Expression Omnibus under accession number GSE150530.
All unprocessed western blot gels for figure 4a, Extended Data Figures 2a, 4a, c, 5f can be found in Source Data.

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	High throughput experiments including PRO-seq, CUT&RUN were conducted in duplicates as indicated. No additional sample size calculation was performed. Two replicates are commonly used for high throughput sequencing experiments to ensure reproducibility while controlling cost.
Data exclusions	Base edited cells that showed insertion or deletions at the edited sites were considered not qualified and excluded from the downstream analysis in Figures 3c, g. No other data were excluded from analysis.
Replication	Biological replicates were performed and indicated in Figure legends to ensure reproducibility of the key data.
Randomization	Samples were allocated based on genotypes only. There are no covariates and no randomization was required.
Blinding	All the data were generate by machines and do not involve human intervention of values. Therefore no blinding were performed.

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	BCL11A, ab191401, Abcam; NFYA, sc-17753, Santa Cruz Biotechnology; NFYA, ab139402, Abcam; TBP, ab220788, Abcam; α-Flag M2 antibody, F1804, Sigma-Aldrich;C/EBPB, NBP1-46179, NovusBio; C/EBPG, sc-517003, Santa Cruz Biotechnology; CDP, sc-514008, Santa Cruz Biotechnology; NFIA, HPA008884, Sigma-Aldrich; NFIC, A303-123A-T, Bethyl Laboratories; Histone H3, ab24834, Abcam; HbF, MHFH01, Life Technologies.
Validation	Flag M2 antibody was validated by the manufacturer through immunostaining. HbF antibody was validated by flow cytometry (Extended Data Figure 3g).

All other antibodies were validated by the manufacturers through western blot to confirm specific recognition of the human antigens. All WB images were shown in manufacturers' websites.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HUDEP-2 cells, HUDEP-1 cells were established and shared by Nakamura group at RIKEN, Japan. 293T cells were purchased from ATCC.
Authentication	These cells were not authenticated.
Mycoplasma contamination	Cells were confirmed to be negative for Mycoplasma as determined by PCR.
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=GSE150530 token: crmpygswtvipred
Files in database submission	please see below in Methodology-Sequencing depth
Genome browser session (e.g. UCSC)	https://genome.ucsc.edu/s/nanliu/NFY_CUTRUN_for_NG

Methodology

Replicates	two replicates of CUT&RUN were performed for NFYA CUT&RUN or TBP CUT&RUN in CD34+ cells as indicated in the file names by "rep1" and "rep2". One experiment was done for those experiments without labeling "rep".
Sequencing depth	All experiments were sequenced as paired end, with read length of 42 bp each strand. The numbers of total reads and concordantly aligned reads were listed below. (Uniquely mapped reads were not applicable to this project due to the fact that the HBG1 and HBG2 genes are duplicated.) NFYA_ChIP_BCLKOHDP2_ctr,37164569,34771514 NFYA_ChIP_BCLKOHDP2_shNFYA,36723066,34500130 NFYA_ChIP_HDP1_ctr,42005143,39225567 NFYA_ChIP_HDP1_shNFYA,50820724,47620992 NFYA_ChIP_HDP2_ctr,34634556,32575748 NFYA_ChIP_HDP2_shNFYA,49777590,46855899 NFYA_CUTRUN_BCLKOdCas9_102_rep1,33495349,30674512 NFYA_CUTRUN_BCLKOdCas9_102_rep2,15000000,11920433 NFYA_CUTRUN_BCLKOdCas9_115_rep1,36819206,33553876 NFYA_CUTRUN_BCLKOdCas9_115_rep2,15000000,12138125 NFYA_CUTRUN_BCLKOdCas9_124_rep1,38921171,35940594 NFYA_CUTRUN_BCLKOdCas9_124_rep2,15000000,12697117 NFYA_CUTRUN_BCLKOdCas9_139_rep1,33681061,30854506 NFYA_CUTRUN_BCLKOdCas9_139_rep2,15000000,11807286 NFYA_CUTRUN_BCLKOdCas9_158_rep1,39688300,35082998 NFYA_CUTRUN_BCLKOdCas9_158_rep2,15000000,9213025 NFYA_CUTRUN_BCLKOdCas9_197_rep1,41422415,37376825 NFYA_CUTRUN_BCLKOdCas9_197_rep2,15000000,8202314 NFYA_CUTRUN_BCLKOdCas9_208_rep1,46242216,42301794 NFYA_CUTRUN_BCLKOdCas9_208_rep2,15000000,11279120 NFYA_CUTRUN_BCLKOdCas9_62_rep1,32979815,30333424 NFYA_CUTRUN_BCLKOdCas9_62_rep2,15000000,12421347 NFYA_CUTRUN_BCLKOdCas9_AAVS_rep1,30639854,27773935 NFYA_CUTRUN_BCLKOdCas9_AAVS_rep2,15000000,10369371 NFYA_CUTRUN_BCLKOHDP2_rep1,43920791,40823178 NFYA_CUTRUN_BCLKOHDP2_rep2,34270965,29805299 NFYA_CUTRUN_CD34_AAVS32_rep1,19057148,15307722 NFYA_CUTRUN_CD34_AAVS32_rep2,14637450,13197549 NFYA_CUTRUN_CD34_AAVS72_rep2,25988936,22767333 NFYA_CUTRUN_CD34_KO32_rep1,17456600,15834599 NFYA_CUTRUN_CD34_KO32_rep2,20485966,19147012 NFYA_CUTRUN_CD34_KO72_rep2,9283584,8254157 NFYA_CUTRUN_CD34_shBCL,39674277,34823019

NFYA_CUTRUN_CD34_shCtr,33222760,28880730
NFYA_CUTRUN_CD34_shNFYA,47578352,42263166
NFYA_CUTRUN_cloneA9d,23624493,15919597
NFYA_CUTRUN_cloneB2p,18004344,9288219
NFYA_CUTRUN_cloneB6p,14853815,3893985
NFYA_CUTRUN_cloneB7d,16272000,9493221
NFYA_CUTRUN_cloneB7p,29818667,16787151
NFYA_CUTRUN_cloneB8p,24889493,10095503
NFYA_CUTRUN_cloneC4p,15589029,9744071
NFYA_CUTRUN_cloneD10d,19563736,11344188
NFYA_CUTRUN_cloneE11p,10817294,3269153
NFYA_CUTRUN_cloneE1d,27306601,20014302
NFYA_CUTRUN_cloneE6p,18850494,13870224
NFYA_CUTRUN_cloneF11d,19154878,13213573
NFYA_CUTRUN_cloneG10p,10822717,3922287
NFYA_CUTRUN_cloneG8d,15036288,7312922
NFYA_CUTRUN_Cord_CD34,18609228,16381146
NFYA_CUTRUN_HDP1_rep1,28560486,26192514
NFYA_CUTRUN_HDP1_rep2,25445084,19593285
NFYA_CUTRUN_HDP2_rep1,31546932,29406956
NFYA_CUTRUN_HDP2_rep2,32928500,29206382
TBP_CUTRUN_CD34_AAVS32_rep1,20255744,17996530
TBP_CUTRUN_CD34_AAVS32_rep2,16770308,15517167
TBP_CUTRUN_CD34_AAVS72_rep2,28013959,24829072
TBP_CUTRUN_CD34_KO32_rep1,21978816,20072549
TBP_CUTRUN_CD34_KO32_rep2,24492508,22990688
TBP_CUTRUN_CD34_KO72_rep2,20770535,18348659

Antibodies	BCL11A, ab191401, Abcam, for CUT&RUN; NFYA, sc-17753, Santa Cruz Biotechnology, for CUT&RUN; NFYA, ab139402, Abcam, for ChIP-seq; TBP, ab220788, Abcam, for CUT&RUN
Peak calling parameters	Bowtie2 was used to align sequences to hg19, and MACS2 was used to call peaks using -f BAMPE -q 0.01 -B -SPMR. Duplicates were retained for CUT&RUN analysis due to the fact that pA-MN digestion (like Tn5 transposition) frequently results in the same DNA fragments.
Data quality	Raw data was quality checked with FastQC. Low quality reads and unpaired reads were removed by trimmomatic. Additional read through adapters were trimmed with kseq script in CUT&RUNTools.
Software	CUT&RUNTools was used to process ChIP-seq and CUT&RUN data. CUT&RUNTools integrates the following softwares: Trimmomatic 0.36, Bowtie 2.2.9, Samtools 1.3.1, Picard 2.8.0, MACS 2.1.1, MEME, Bedops 2.4.30, Bedtools 2.26.0.

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	intracellular staining was performed by fixing cells with 0.05% glutaraldehyde (grade II) (Sigma) for 10 min at room temperature. Cells were centrifuged for 5 min at 600x g and then resuspended in 0.1% Triton X-100 (Life Technologies) for 5 min at room temperature for permeabilization. Triton X-100 was diluted with phosphate buffered saline (PBS) with 0.1% BSA and then centrifuged at 600x g for 15 min. Cells were stained with antibodies for HbF (clone HbF-1 with FITC or APC conjugation; Life Technologies) for 20 min in the dark. Cells were washed to remove unbound antibody before flow cytometry.
Instrument	BD Accuri™ C6 Flow Cytometer
Software	Data were collected using built in software in BD Accuri™ C6 Flow Cytometer and analyzed with FlowJo V10
Cell population abundance	10000-50000 cells were analyzed. Dead cells or doublets were excluded with the gating strategy described below.

Gating strategy	FSC/SSC was used to exclude dead cells. FSC-A/FSC-H were used to select single cells. FL1-A channel recorded the FITC fluorescence. Positive and negative populations were defined according to 1) a non-stained sample as negative control. 2) an HbF expressing cell as positive control. 3) an obvious separation of HbF+ and HbF- populations.

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