nature portfolio

Corresponding author(s): Douglas R. Higgs, Richard J. Gibbons

Last updated by author(s): Jan 13, 2026

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

Single cell RNA-seq and ChIP-seq libraries were sequenced on an Illumina NextSeq platform. FASTQ files were generated using bcl2fastq (2.20).

Flow cytometry data was collected using Attune flow cytometer software, BD FACS Diva software or SONY (MA900) software.

Fluorescence images were acquired using a Leica DMi8 Thunder microscope.

Schematics shown in Figure 1a, b, and 3d were created using Adobe Illustrator.

Data analysis

The Sanger sequencing data and relevant genome sequences used in this study were visualised and analysed using SnapGene (v8.0).

The CRISPR knockout efficiency was analysed using the Synthego ICE CRISPR analysis tool (https://ice.editco.bio/#/).

Gene expression analysis of the BFU-E colonies was performed using the Biomark HD Fluidigm built-in software, and data visualisation was done in R Studio (R 4.0.5) using packages ggplot2 (3.4.2), reshape2 (1.4.4), plotly (4.10.0), viridis (0.6.2).

Flow cytometry data was analysed using FlowJo software (10.5.3 to 10.10.0).

Flurorescence images were analysed in Fiji software (2.14.0).

Bioinformatic analyses were performed for single cell RNA-seq and ChIP-seq data, within linux and R (4.0.5). Linux packages include:

TrimGalore (0.4.1), FASTQC (0.11.9), samtools (v19), STAR (2.6.1d), FeatureCounts (1.4.5-p1), Deeptools (2.29.2), Bowtie (2.3.2). R packages include: Rtsne (0.15), ggplot2 (3.4.2), ClusterProfiler (3.18.1).

Statistical analyses and plots generation were finished in GraphPad Prism software (v9) or R (4.0.5).

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 scRNA-seq and ChIP-seq data generated in this study have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) under accession codes GSE291201 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE291201] and GSE291197 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE291197] respectively. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Healthy donor cones were purchased from NHSBT without gender specification. After receiving the samples, gender was confirmed by checking the SRY gene on chromosome Y. Given that ATRX is located on the chromosome X and ATR-X syndrome only affects males, only male donor cells were selected for ATRX knockout experiments.
Reporting on race, ethnicity, or other socially relevant groupings	Not taken into account.
Population characteristics	Not taken into account.
Recruitment	Healthy donor cones were collected by NHSBT without bias.
Ethics oversight	Wales Research Ethics Committee

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

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 determined based on prior experience with the experimental systems used, consistency with sample sizes reported in comparable published studies in the field, and practical considerations such as cell availability, experimental feasibility, and reproducibility.
Data exclusions	For single cell RNA-seq and BFU-E gene expression data, samples that did not meet quality control criteria (as detailed in the Methods section) were excluded from the analysis. For all the other experiments, no data was excluded.
Replication	Key experiments were performed at least three times, with the exact number of replicates specified in the manuscript. All attempts at replication were successful.
Randomization	Experimental groups were separated by genotype, but within each experimental group, samples were allocated and analysed randomly.
Blinding	Blinding was not applicable to this study, as samples were distinguished by genotype. However, all samples from the compared groups were analysed using the same pipelines and scripts to ensure consistency.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used

CD34-PE/Cy7 (Biolegend, 343616)
CD36-APC/Cy7 (Biolegend, 336213)
CD71-PerCP/Cy5.5 (Biolegend, 334114)
CD235a-PE (BD Bioscience, 555570)
CD49D-APC (BD Bioscience, 561892)
CD233 (Band3)-FITC (IBGRL, 9439FI)
Hoechst 33258-violet (Invitrogen, H3569)
CD34-PE/Cy7 (Biolegend, 343616)
Lineage cocktail (CD2/3/56/14/16/19/235a)-APC (Thermo Fisher Scientific, 22-7776-72)
CD36-PerCP/Cy5.5 (Biolegend, 336223)
CD71-FITC (eBioscience, 11-0719-41)
CD41a-EF450 (eBioscience, 48-0419-41)
CD105-PE (Biolegend, 323205)
CD38-AlexaFluor 700 (Biolegend, 303523)
CD45RA-APC-eFluor 780 (eBioscience, 47-0458-41)
CD123-BV605 (BD Bioscience, 564197)
CD90-BV711 (BD Bioscience, 740786)
CD10-APC (eBioscience, 17-0106-42)
Fixable Viability-BV510 (BD Horizon, 564406)
Hoechst 33342-violet (Invitrogen, H3570)
γH2AX antibody (Merck, 05-636)
ubiquityl-histone H2A antibody (Cell Signaling Technology, 8240)
ATRX antibody (Abcam, ab97508)

Validation

All the antibodies used in this study were already validated by the manufacturer or previous literature.

CD34-PE/Cy7 (Biolegend, 343616) https://www.biolegend.com/de-at/products/pe-cyanine7-anti-human-cd34-antibody-13588
CD36-APC/Cy7 (Biolegend, 336213) https://www.biolegend.com/de-at/products/apc-cyanine7-anti-human-cd36-antibody-5457
CD71-PerCP/Cy5.5 (Biolegend, 334114) https://www.biolegend.com/de-at/products/percp-cyanine5-5-anti-human-cd71-antibody-9387
CD235a-PE (BD Bioscience, 555570) https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/pe-mouse-anti-human-cd235a-b.555570
CD49D-APC (BD Bioscience, 561892) https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/apc-mouse-anti-human-cd49d.561892?tab=product_details
CD233 (Band3)-FITC (IBGRL, 9439FI) https://haematologica.org/article/view/haematol.2020.260158
Hoechst 33258-violet (Invitrogen, H3569) https://www.thermofisher.com/order/catalog/product/H3569
Lineage cocktail (CD2/3/56/14/16/19/235a)-APC (Thermo Fisher Scientific, 22-7776-72) https://www.thermofisher.com/antibody/product/Human-Hematopoietic-Lineage-Antibody-clone-RPA-2-10-OKT3-61D3-CB16-HIB19-TULY56-HIR2-Cocktail/22-7776-72
CD36-PerCP/Cy5.5 (Biolegend, 336223) https://www.biolegend.com/de-at/products/percp-cyanine5-5-anti-human-cd36-antibody-14914
CD71-FITC (eBioscience, 11-0719-41) https://www.thermofisher.com/antibody/product/CD71-Transferrin-Receptor-Antibody-clone-OKT9-OKT-9-Monoclonal/11-0719-42
CD41a-EF450 (eBioscience, 48-0419-41) https://www.thermofisher.com/antibody/product/CD41a-Antibody-clone-HIP8-Monoclonal/48-0419-41
CD105-PE (Biolegend, 323205) https://www.biolegend.com/de-at/products/pe-anti-human-cd105-antibody-3711
CD38-AlexaFluor 700 (Biolegend, 303523) https://www.biolegend.com/de-at/products/alexa-fluor-700-anti-human-cd38-antibody-6050
CD45RA-APC-eFluor 780 (eBioscience, 47-0458-41) https://www.thermofisher.com/antibody/product/47-0458-41.html
CD123-BV605 (BD Bioscience, 564197) https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/bv605-mouse-anti-human-cd123.564197?tab=product_details

CD90-BV711 (BD Bioscience, 740786) https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/bv711-mouse-anti-human-cd90.740786?tab=product_details
CD10-APC (eBioscience, 17-0106-42) https://www.thermofisher.com/antibody/product/CD10-Antibody-clone-eBioCB-CALLA-CB-CALLA-Monoclonal/17-0106-42
Fixable Viability-BV510 (BD Horizon, 564406) https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/fixable-viability-stain-510.564406?tab=product_details
Hoechst 33342-violet (Invitrogen, H3570) https://www.thermofisher.com/order/catalog/product/H3570
γH2AX antibody (Merck, 05-636) https://www.sigmaaldrich.com/GB/en/product/mm/05636?msockid=2c1ce964cf6e687e175df831ce9569e9
ubiquityl-histone H2A antibody (Cell Signaling Technology, 8240) https://www.cellsignal.com/products/primary-antibodies/ubiquityl-histone-h2a-lys119-d27c4-rabbit-monoclonal-antibody/8240
ATRX antibody (Abcam, ab97508) https://www.abcam.com/en-us/products/primary-antibodies/atrx-antibody-ab97508

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	CD34+ cells were isolated from healthy donor-derived leucocyte cones purchased from the National Health Service Blood and Transplant (NHSBT) (sex was not disclosed by the supplier). These cells were differentiated into erythroid lineage for cellular and molecular analyses. Wildtype HUDEP-2 cells were obtained from the Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Japan (Kurita et al, 2013).
Authentication	Cells were analysed via cytospins and FACS immunophenotyping during erythroid differentiation. Genotype was confirmed through targeted PCR and Sanger sequencing.
Mycoplasma contamination	All cell lines used in this study were routinely tested for mycoplasma contamination and were confirmed to be mycoplasma-negative at the time of experimentation.
Commonly misidentified lines (See ICLAC register)	None.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

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=GSE291197
Files in database submission	ATRX-ChIP-seq_HUDEP-2_undiff_d0 Input_ChIP-seq_HUDEP-2_undiff_d0 ATRX-ChIP-seq_HUDEP-2_d2 Input_ChIP-seq_HUDEP-2_d2 ATRX-ChIP-seq_HUDEP-2_d4 Input_ChIP-seq_HUDEP-2_d4 ATRX-ChIP-seq_HUDEP-2_d6 Input_ChIP-seq_HUDEP-2_d6 ATRX-ChIP-seq_HUDEP-2_d8 Input_ChIP-seq_HUDEP-2_d8
Genome browser session (e.g. UCSC)	

https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hubUrl=http://datashare.molbiol.ox.ac.uk/public/project/higgslab/yshen/ATRXChIP_HUDEP2degron/hubDirectory/hub.txt&genome=hg19&position=lastDbPos

Methodology

Replicates	70 million differentiated HUDEP-2 cells were collected at different stages (days 0, 2, 4, 6, and 8) from the same differentiation batch for ATRX ChIP-seq analysis. The differentiation process was rigorously validated using cytospin, flow cytometry, and gene expression analysis to confirm proper and expected differentiation. An input control sample was set aside before antibody incubation as a negative control. The aim was to track ATRX binding trends throughout erythroid differentiation and given the substantial cell numbers required, a single well-characterised biological replicate was generated for each time point. Each dataset was referenced against its corresponding input control and across all other time points.
Sequencing depth	All ATRX ChIP-seq samples have a sequencing depth > 20M reads (paired-end).
Antibodies	anti-ATRX antibody (Abcam, Cat No ab97508)
Peak calling parameters	Peak calling was performed on LanceOtron using the respective input controls (https://lanceotron.molbiol.ox.ac.uk/).
Data quality	Sequencing quality was verified via FASTQC.
Software	ChIP-seq libraries were sequenced on an Illumina NextSeq platform. FASTQ files were generated using bcl2fastq (2.20). ChIP-seq data was analysed within linux, using packages FASTQC (0.11.9), samtools (v19), Deeptools (2.29.2), Bowtie (2.3.2), following a customised in-house pipeline (script: https://github.com/Hughes-Genome-Group/NGseqBasic/releases). The bigwig files were visualised in UCSC genome browser.

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 the erythroid differentiation flow panel, 1x10^5 cells on different differentiation days were washed and resuspended in 100 µL of FACS buffer (1x PBS supplemented with 2% FBS and 2 mM EDTA) prior to staining with a panel of fluorophore-conjugated antibodies at 4°C for 30 min. Following antibody staining, the cells were washed and resuspended in FACS buffer containing Hoechst-33258 for live/dead cell analysis. For the HSPC FACS panel, all CD34+ cells after nucleofection were re-suspended in human FcR blocking reagent and incubated for 5 mins at room temperature. They were then stained with 11 fluorophore-conjugated monoclonal antibodies for 20 minutes at 4°C before washing. Samples were analysed to define different progenitor types, and BFU-E/EEP (early erythroid progenitors) were sorted into 1.5 mL eppendorf microcentrifuge tubes containing 100 uL of IMDM with 2% FBS medium (STEMCELL Technologie, Cat No 07700) using a cell sorter for later CFU assay. For single cell sorting for scRNA-seq, 5x10^5 AAVS1 or ATRX KO cells at different differentiation day 13 were collected in their culture media (after confirming viability > 95%). Hoechst 33342 was then added to the media (fianl concentration 1 ug/mL) and the cells were incubated at 37°C for 5 min before washing with FACS buffer (PBS + 2% FBS + 2mM EDTA).
Instrument	Attune NxT analyser (ThermoFisher) for regular differentiation flow analysis, BD FACSAria III cell sorter (Becton Dickinson) for HSPC analysis and sorting, SONY (MA900) cell sorter for single cell sorting.
Software	Flow cytometry data was collected using Attune flow cytometer software, BD FACS Diva software or SONY MA900 software. And data was analysed using FlowJo software (10.5.3 to 10.10.0).
Cell population abundance	All flow cytometry experiments were conducted for analysis, except for BFU-E/EEP (early erythroid progenitors) population sorting and single cell sorting for RNA-seq. HSPC populations, including HSC, MPP, BFU-E/EEP (early erythroid progenitors) and CFU-E/LEP (late erythroid progenitors), were analysed, with EEP sorted in purity mode. The sorted EEP was then seeded into Methocult for a colony forming unit assay, where over 90% of the resulting colonies were erythroid, confiming a purity of more than 90%. For scRNA-seq sorting, single-cell sorting mode was used to sort live nucleated cells. These cells were highly abundant, and the nucleated and enucleated popualtions were clearly separated, expecting a post-sort purify of over 99%.
Gating strategy	Cells were first assessed using FSC vs SSC to exclude debris, followed by gating singlets (FSC-A vs FSC-H) and live cells (viability

Gating strategy

dye). The the populations were then defined based on their cell surface marker expression, as outlined in the Extended Figures. The gates were set based on the FMO controls.

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