nature portfolio

Corresponding author(s): Ulrich Steidl

Last updated by author(s): 07/29/2024

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	Data collection for Flow cytometry occured with FACSDlva 8 software on a BD LSRII, and data was analyzed using FlowJo v10.6.1.
Data analysis	CUT&Tag generated fastq files were mapped to the human genome (hg38) using bowtie2 (Version 2.2.3) with options: --end-to-end --very-sensitive --no-mixed --no-discordant --phred33 -I 10 -X 700. Normalized bedgraph files were generated by using bedtools genomecov with a normalization factor of 1,000,000/No. total human reads. Peak calling was performed using bdgpeakcall from MACS2 (Version 2.1.0) with options -l 100 and -c 2. Bigwig files were generated using bedGraphToBigWig for visualization in IGV_2.4.15. Motif analysis and peak annotation was performed using the Homer package (v4.11.1). Differential peak analysis was performed in R studio with either the Diffbind package (v3.12.0) utilizing DESEQ2 (v1.30.1) if the n>2, or a custom script we termed "GoodpeaksScript" (https://github.com/steidl-lab/rePU.1sitioning) if n=1. Classical ChIP generated fastq files were mapped to the human or drosophila genome (hg38 or dm3) using bowtie2 (Version 2.2.3). Duplicates were then removed with Picard (v2.20.1) and normalization was performed by subsampling .bam files by a drosophila reads ratio of vehicle over DB2115. The MACS2 package (v2.1.0) and callpeaks function was used to identify peaks in normalized files, and bigwig files were generated for IGV_2.4.15 visualization. For RNA sequencing datasets, quality control was performed on the basis of error distribution along the length or reads, GC distribution, N content, base quality, and adapter content. Reads were mapped to the hg38 transcriptome using STAR aligner (v2.7.7a). Raw counts were subsequently normalized and analyzed for differential expression in R using the Bioconductor package DESeq2 (v1.30.1). An enrichment score was generated using the negative logarithm of the adjusted p-value multiplied by the sign of the fold-change for each gene and input into Fast Gene Set Enrichment Analysis (FGSEA v1.19.2). Pre-ranked gene lists were queried against standard c1-8 and hallmark MSigDB gene lists (v7.4, Broad Institute). Additional gene list enrichments were conducted with Enrichr v3.0. For Pro-seq data analysis, adaptor trimming was performed with Cutadapt (v1.18), then the sequences were aligned and mapped using

bowtie2 (v2.5.1). Samtools (v1.9) was used for the file format conversion before using the Nascent RNA Sequencing Analysis (NRSA) pipeline to determine the gene body changes.
Statistical tests were performed in R studio (2023.12.1+402) or GraphPad Prism (v9.5).

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

Raw and processed sequencing data (CUT&Tag, ChIP, ATAC, RNA seq & PRO-seq) has been uploaded to the GEO online database under Super Series GSE267389 publicly available July 30th 2024). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE267389

Hg38 genomic data is available from https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/.

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	Adult AML/MDS BM, and PB samples were obtained for proof-of-concept studies, but no analyses were made based on gender or sex. The gender of patient samples were listed in S. Table 1F.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable (only small proof-of-concept sample size was used in this study)
Population characteristics	The covariate-relevant patient characteristics include age, gender and cytogenetics of the leukemia and have been reported in S.Table 1F)
Recruitment	Patient samples with MDS or AML were obtained after written informed consent, from Montefiore Medical Center / Albert Einstein Cancer Center. These samples were obtained from a tissue repository, they were deidentified, and not specifically collected for this study.
Ethics oversight	Adult AML/MDS BM, and PB samples were obtained for proof-of-concept studies after written informed consent and following Albert Einstein College of Medicine institutional review board approval (CCI 2008-842 and 2006-536).

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 sizes were not based on formal power calculations to detect pre-specified effect sizes for this proof-of-concept study. Sample size was based on availability of primary AML samples, or reproducibility from at least 2 replicates of Ch IP/CUT& Tag sequencing.
Data exclusions	No data was excluded. All data that passed quality control parameters using standard sequencing and analysis algorithms (described in detail in the methods) were included in the analysis.
Replication	All attempts of replication (biological and technical) were successful and have been presented in the manuscript. Multiple approaches were employed to validate the observations/phenomenon described in this paper, with both CUT&Tag and ChIP seq. in multiple different cellular systems showing the same phenomenon of PU.1 redistribution. In summary, all MOLM13 experiments (CUT&Tag, Flow cytometry, Differentiation) were conducted with 2-5 experimental replicates. Primary sample studies were conducted with 2 technical replicates over 7 different biological samples (from different patients). Other cell line and drug CUT&Tag studies were typically from a n=1, but were providing consistent observations over the different contexts/pharmacologica used.
Randomization	This is not relevant for the present study (no group allocation was performed).

Blinding	This is not relevant for the present study (no group allocation was 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
☒	☐ 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	CD14 FITC (#32S603) Biolegend, 1:100 CD15 APCCy7 (#323048) Biolegend, 1:100 CD86 PECy7 (#374210) Biolegend, 1:100 CD34 PE (#343506) Biolegend, 1:100 CD209 AF647 (#330112) Biolegend, 1:100 CD11b PerCPCy5.5 (#101228) Biolegend, 1:100 CSF1R APC (#347306) Biolegend, 1:100 IL4R BV421 (#355014) Biolegend, 1:100 PU.1 (Santa Cruz, sc-352), 1:50 for CnT, 1:1000 for WB rabbit IgG (Santa Cruz; sc-3888), 1:100 Cas9(Cell signaling, 14697S), 1:50 Actin (Sigma-Aldrich, A2066), 1:5000 for WB GABPA (Invitrogen, PA5-27735), 1:50 FLI1 (Invitrogen, PA5-295977), 1:50 RUNX1 (Cell signaling, 4334S), 1:50 GATA3 (Cell signaling, 5852T), 1:50 ELF1 (Proteintech, 22565-1-AP), 1:50 Guinea Pig anti-Rabbit secondary (Antibodies Online, ABIN101961) 1:100 HRP-conjugated anti-rabbit secondary (Cell signaling, 7074S), 1:5000 pSTAT6 AF647 (Biolegend, 686012), 1:100 phospho-S6 ribosomal protein PE (Biolegend, 608604), 1:100
Validation	Standard FACS antibodies obtained from widely used commercial providers were used in this study. All FACS antibodies have been validated by the company through quality control in-house Flow cytometry, and also by comparisons with isotype control antibodies within our laboratory. Western Blot and CUT&Tag antibodies are all commercially available and have been documented by the company to be specific and valid (via Western blot, immunofluorescence or flow cytometry).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	MOLM13-ATCC TH PI -ATCC HL60-ATCC MV411-ATCC JURKAT - ATCC 293T - ATCC
Authentication	ATCC authenticated these cell lines at purchase, which had been authenticated by short tandem repeat (STR) profiling. They were not independently authenticated upon receipt.
Mycoplasma contamination	All cell lines are confirmed negative for Mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	None used

Plants

Seed stocks: N/A

Novel plant genotypes: N/A

Authentication: N/A

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.

Raw and processed sequencing data (CUT&Tag, ChIP, ATAC, RNA seq & PRO-seq) has been uploaded to the GEO online database under Super Series GSE267389 publicly available July 31st 2024).

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE267389

Files in database submission

GSM6262719
MOLM13, VEH, 12hr [MOLM_VEH1]
GSM6262720
MOLM13, VEH, 12hr [MOLM_VEH2]
GSM6262721
MOLM13, VEH, 12hr [MOLM_VEH3]
GSM6262722
MOLM13, 5uM DB2115, 12hr [MOLM_DB1]
GSM6262723
MOLM13, 5uM DB2115, 12hr [MOLM_DB2]
GSM6262724
MOLM13, 5uM DB2115, 12hr [MOLM_DB3]
GSM6262725
MOLM13, VEH, 12hr [MOLM_IgGV1]
GSM6262726
MOLM13, 5uM DB2115, 12hr [MOLM_IgGDB1]
GSM6262727
THP1, VEH, 12hr [THP1_VEH1]
GSM6262728
THP1, VEH, 12hr [THP1_VEH2]
GSM6262729
THP1, 5uM DB2115, 12hr [THP1_DB1]
GSM6262730
THP1, 5uM DB2115, 12hr [THP1_DB2]
GSM6262731
HL60, VEH, 12hr [HL60_VEH1]
GSM6262732
HL60, VEH, 12hr [HL60_VEH2]
GSM6262733
HL60, 5uM DB2115, 12hr [HL60_DB1]
GSM6262734
HL60, 5uM DB2115, 12hr [HL60_DB2]
GSM6262735
MV411, VEH, 12hr [MV411_VEH1]
GSM6262736
MV411, VEH, 12hr [MV411_VEH2]
GSM6262737
MV411, 5uM DB2115, 12hr [MV411_DB1]
GSM6262738
MV411, 5uM DB2115, 12hr [MV411_DB2]
GSM6262739
MV411, VEH, 12hr [MV411_IgG]
GSM6262740
THP1, VEH, 12hr [THP1_IgG]

GSM6262741
HL60, VEH, 12hr [HL60_IgG]
GSM6262742
MOLM13, 5uM DB2373, 12hr [MOLM_DB2373]
GSM6262743
MOLM13, 5uM DB2313, 12hr [MOLM_DB2313]
GSM6262744
MOLM13, 5uM DB2826, 12hr [MOLM_DB2826]
GSM6262745
MOLM13, 5uM DB2750, 12hr [MOLM_DB2750]
GSM6262746
MOLM13, 200nM Daunorubicin, 12hr [MOLM_Daun]
GSM6262747
MOLM13, 400nM Cytarabine, 12hr [MOLM_AraC]
GSM6262748
MOLM13, 200nM Daun, 12hr [MOLM_Daun_IgG]
GSM6262749
MOLM13, VEH, 12hr [MOLM_VEH_DaunAraC]
GSM6262750
MOLM13, 5uM DB2115, 1hr [MOLM_1hr_DB_1]
GSM6262751
MOLM13, 5uM DB2115, 1hr [MOLM_1hr_DB_2]
GSM6262752
MOLM13, 5uM DB2115, 4hr [MOLM_4hr_DB_1]
GSM6262753
MOLM13, 5uM DB2115, 4hr [MOLM_4hr_DB_2]
GSM6262754
MOLM13, 5uM DB2115, 12hr [MOLM_12hr_DB]
GSM6262755
MOLM13, VEH, 12hr [MOLM_VEH_timecourse]
GSM6262756
MOLM13, CRISPRd sgNon-targeting [sgNT_PU1_1]
GSM6262757
MOLM13, CRISPRd sgNon-targeting [sgNT_PU1_2]
GSM6262758
MOLM13, CRISPRd sgNon-targeting [sgNT_dCas9]
GSM6262759
MOLM13, CRISPRd targeting STRAP upstream region [sgSTRAP_dCas9]
GSM6262760
MOLM13, CRISPRd targeting STRAP upstream region [sgSTRAP_PU1_1]
GSM6262761
MOLM13, CRISPRd targeting STRAP upstream region [sgSTRAP_PU1_2]
GSM6262762
MOLM13, CLICKonCUT&TAG with DB2750 [CLICK1]
GSM6262763
MOLM13, CLICKonCUT&TAG with DB2750 [CLICK2]
GSM6262764
MOLM13, CLICKonCUT&TAG with DB2750 [CLICK3]
GSM8264925
MOLM13, VEH, ATAC-seq Rep1
GSM8264926
MOLM13, VEH, ATAC-seq Rep2
GSM8264927
MOLM13, VEH, ATAC-seq Rep3
GSM8264928
MOLM13, 1hr DB2115, ATAC-seq Rep1
GSM8264929
MOLM13, 1hr DB2115, ATAC-seq Rep2
GSM8264930
MOLM13, 4hr DB2115, ATAC-seq Rep1
GSM8264931
MOLM13, 4hr DB2115, ATAC-seq Rep2
GSM8264932
MOLM13, 12hr DB2115, ATAC-seq Rep1
GSM8264933
MOLM13, 12hr DB2115, ATAC-seq Rep2
GSM8264934
MOLM13, VEH, 12hr, ChIP-seq
GSM8264935
MOLM13, DB2115, 12hr, ChIP-seq
GSM8264936
MOLM13, VEH, 12hr, RUNX1, Rep1
GSM8264937
MOLM13, VEH, 12hr, RUNX1, Rep2
GSM8264938
MOLM13, 5uM DB2115, 12hr, RUNX1 Rep1

GSM8264939
MOLM13, 5uM DB2115, 12hr, RUNX1, Rep2
GSM8264940
JURKAT, VEH, 12hr, RUNX1
GSM8264941
JURKAT, 5uM DB2115, 12hr, RUNX1
GSM8264942
JURKAT, VEH, 12hr, ELF1
GSM8264943
JURKAT, 5uM DB2115, 12hr, ELF1
GSM8264944
JURKAT, VEH, 12hr, GATA3
GSM8264945
JURKAT, 5uM DB2115, 12hr, GATA3
GSM8264946
JURKAT, VEH, 12hr, FLI1
GSM8264947
JURKAT, 5uM DB2115, 12hr, FLI1
GSM8264948
MOLM13, VEH, 12hr, ELF1
GSM8264949
MOLM13, 5uM DB2115, 12hr, ELF1
GSM8264950
MOLM13, VEH, 12hr, GABPA
GSM8264951
MOLM13, 5uM DB2115, 12hr, GABPA
GSM8264952
AML sample1, VEH, 12hr
GSM8264953
AML sample1, 5uM DB2115, 12hr
GSM8264954
AML sample2, VEH, 12hr
GSM8264955
AML sample2, 5uM DB2115, 12hr
GSM8264956
MOLM13, VEH, 12hr PRO-seq rep1
GSM8264957
MOLM13, VEH, 12hr PRO-seq rep2
GSM8264958
MOLM13, 5uM DB2115, 1hr PRO-seq rep1
GSM8264959
MOLM13, 5uM DB2115, 1hr PRO-seq rep2
GSM8264960
MOLM13, 5uM DB2115, 4hr PRO-seq rep1
GSM8264961
MOLM13, 5uM DB2115, 4hr PRO-seq rep2
GSM8264962
MOLM13, 5uM DB2115, 12hr PRO-seq rep1
GSM8264963
MOLM13, 5uM DB2115, 12hr PRO-seq rep2
GSM8264964
MOLM13, VEH, 20hr, RNA-seq rep1
GSM8264965
MOLM13, VEH, 20hr, RNA-seq rep2
GSM8264966
MOLM13, VEH, 20hr, RNA-seq rep3
GSM8264967
MOLM13, DB2115, 20hr, RNA-seq rep1
GSM8264968
MOLM13, DB2115, 20hr, RNA-seq rep2
GSM8264969
MOLM13, DB2115, 20hr, RNA-seq rep3

Genome browser session
(e.g. UCSC)

No longer applicable

Methodology

Replicates

The central datasets were repeated multiple times for consistency, minimum of 2 independent Chip or CUT&Tag replicates (MOLM13, THPI, HL60, MV411, dCas9-STRAP, RUNXI, MOLM13 timecourse). For exploratory or primary datasets with low cell material available, n=1 (Daunorubicin, Ara-C, DB2373, DB2826, DB2313, GATA3, FLII, GABPA, ELFI, primary patient samples (multiple)).

Sequencing depth

Sequencing for CUT&Tag was conducted with Paired-end sequencing (35-75bp reads) with an average coverage of 10-15million reads

Sequencing depth	per sample. This is well above the recommended sequencing depth of~5million for CUT&Tag.
Antibodies	Antibodies used for ChIP and CUT&Tag: PU.l (Santa Cruz, sc-352, 1:50) rabbit lgG (Santa Cruz; sc-3888, 1:100) Cas9(Cell signaling, 14697S, 1:50) GABPA (Invitrogen, PA5-27735, 1:50) FLll (Invitrogen, PA5-295977, 1:50) RUNX1 (Cell signaling, 4334S, 1:50) GATA3 (Cell signaling, 5852T, 1:50) ELF1 (Proteintech, 22565-1-AP, 1:50)
Peak calling parameters	CUT&Tag generated fastq files were mapped to the human genome (hg38) using bowtie2 (Version 2.2.3) with options: --end-to-end --very-sensitive --no-mixed --no-discordant --phred33 -I 10-X 700. Normalized bedgraph files were generated by using bedtools genomecov with a normalization factor of 1,000,000/No. total human reads. Peak calling was performed using bdgpeakcall from MACS2 (Version 2.1.0) with options -1100 and -c 2. Bigwig files were generated using bedGraphToBigWig for visualization in IGV 2.4.15. Motif analysis and peak annotation was performed using the Homer package. For CLICKonCUT& Tag, the average MACS2 peak scores across 3 replicates were calculated and compared to input vehicle-treated PU.l CUT& Tag peak scores to generate a log2fold change CLICK score. Peaks with CLICK scores above 0.5 were considered enriched for drug binding, whereas scores below 0.5 were considered non-drug binding. Classical ChIP generated fastq files were mapped to the human or drosophila genome (hg38 or dm3) using bowtie2 (Version 2.2.3). Duplicates were then removed with Picard and normalization was performed by subsampling .barn files by a drosophila reads ratio of vehicle over DB2115. The MACS2 package and call peaks function was used to identify peaks in normalized files, and bigwig files were generated for IGV visualization. Differential peak analysis was performed in R studio with "GoodpeakAnalysis".
Data quality	Initial Fastq files were confirmed passable via multi-FASTQC analyses. After peak calling peaks were confirmed for significant changes using Diffbind package, in particular: Differential peak analysis was performed in R studio with either the Diffbind package utilizing DESEQ2 analyses if the n>2, or a custom script we termed "GoodpeaksScript" (https:/ /github.com/steidl-lab/rePU.lsitioning). if n=l. In Diffbind, FDR cutoff was set to 0.1 for determining significantly changed peaks. For GoodpeakAnalysis, three stringent filters were used for the differential peak analysis of the average peak intensity; min. intensity> 7.5, min. fold change >4, and min. summit >3.
Software	We used the following software for ChIP-seq data analyses: Bowtie2 (Version 2.2.3), Picard (v2.20.1), bedtools genomecov, MACS2 (bdgpeakcall, Version 2.1.0), bedGraphToBigWig, IGV_2.4.15., Homer v4.11.1, Diffbind package utilizing DESEQ2 v3.12.0, STAR aligner (2.7.7a), FGSEA (1.19.2), Bioconductor, DESeq2 v1.30.1. Custom scripts from R studio were used and are described at: https://github.com/steidl-lab/rePU.1sitioning

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	Antibody staining was performed for 30 min prior to flow cytometry analyses, employing the following Biolegend antibodies; CD14 FITC (#325603), CD15 APCCy7 (#323048), CD86 PECy7 (#374210), CD34 PE (#343506), CD209 AF647 (#330112), CDllb PerCPCy5.5 (#101228), CSFIR APC (#347306), IL4R BV421 (#355014) and DAPI for alive cell gating.
Instrument	Flow cytometry analysis was performed on a BD LSRII containing a yellow laser (BD LSR II-Y).
Software	Flow cytometric data was acquired using FACSDiva software v 8.0 and analyzed using FlowJo Software (version 10.6.1)
Cell population abundance	Sorting was not performed
Gating strategy	The gating strategy used followed established markers and schemes for the identification of viable cultured AML cells or cell lines. Gating scheme for primary AML cells is displayed in Supplemental Figure 10. Negativity for any marker was defined as the threshold defined by staining with the respective isotype control antibodies.

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