nature portfolio

Corresponding author(s): Seishi Ogawa, Sören Lehmann, Yotaro Ochi

Last updated by author(s): Apr 20, 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	No software was used to collect additional data.
Data analysis	Softwares used in this study are as follows: ANANSE (v0.5.1),https://github.com/vanheeringen-lab/ANANSE ArchR (v1.0.2), https://www.archrproject.com/ ATACseqQC (v1.18.0),https://bioconductor.org/packages/release/bioc/html/ATACseqQC.html ataqv (v1.2.1), https://parkerlab.github.io/ataqv/ BoneMarrowMap (v0.1.0), https://github.com/andygxzeng/BoneMarrowMap Bowtie (v1.3.1), http://bowtie-bio.sourceforge.net/index.shtml Bowtie2 (v2.5.2), http://bowtie-bio.sourceforge.net/bowtie2/index.shtml BPcells (v0.2.0), https://github.com/bnprks/BPCells Cell Ranger ARC (v2.0.1), https://www.10xgenomics.com/jp/support/software/cell-ranger-arc/latest ChAMP (v2.29.1), https://www.bioconductor.org/packages/release/bioc/html/ChAMP.html CIBERSORTx (v1.0), https://cibersortx.stanford.edu/ ClaNC (v1.0), https://github.com/naikai/sake/blob/master/R/clanc.R clusterProfiler (v4.2.2), https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html CNACS (v1.0), https://github.com/papaemmelab/toil_cnacs Cutadapt (v4.9), https://github.com/marcelm/cutadapt/blob/master/doc/recipes.rst deeptools (v3.3.1), https://github.com/deeptools/deepTools edgeR (v3.36.0), https://bioconductor.org/packages/release/bioc/html/edgeR.html

GATK (v4.1.0.0), https://gatk.broadinstitute.org/hc/en-us
Genomon (v2.5.0), https://github.com/Genomon-Project
Genomon SV (v0.8.0), https://github.com/Genomon-Project/GenomonSV
glmnet (v4.1.7), https://cran.r-project.org/web/packages/glmnet/index.html
GRIDSS (v2.12.0), https://github.com/PapenfussLab/gridss
HMMRATAC (v1.2.10), https://github.com/LiuLabUB/HMMRATAC
HOMER (v4.11.1), http://homer.ucsd.edu/homer/
igraph (v1.5.1), https://igraph.org/
Integrative Genomics Viewer (v2.14.1), https://www.broadinstitute.org/igv/
limma (v3.50.3), https://bioconductor.org/packages/release/bioc/html/limma.html
MACS (v2.1.1), https://github.com/taoliu/MACS
parabricks (v3.5.0), https://docs.nvidia.com/clara/parabricks/latest/index.html
ROSE (v1.0), https://bitbucket.org/young_computation/rose
rpart (v4.1.23), https://cran.r-project.org/web/packages/rpart/index.html
SCENIC+ (v1.0a1), https://github.com/aertslab/scenicplus
scran (v1.22.1), https://rdrr.io/bioc/scran/
Seurat (v5.0.3), https://satijalab.org/seurat/
Signac (v1.12.0), https://stuartlab.org/signac/
skewer (v0.2.2), https://github.com/relipmoc/skewer
STAR (v2.5.3), https://github.com/alexdobin/STAR
Subread (v1.5.3), http://subread.sourceforge.net/
uwot (v0.1.16), https://github.com/jlmelville/uwot

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 raw sequencing data generated in this study, including ATAC-seq, RNA-seq, ChIP-seq, scRNA/ATAC-seq, DNA methylation, and WGS data, have been deposited in controlled-access repositories. WGS datasets are available through the G-CARD controlled-access database (accession number G-CARDS000002), while the remaining datasets are available from the European Genome-phenome Archive (EGA) under accession number EGAS00001008315. No sequencing datasets are duplicated between the repositories. Access to these data is restricted due to ethical, legal, and privacy considerations related to human subject data. Applications are open to qualified researchers and require institutional ethical approval and a research proposal consistent with participant consent. Requests for EGA datasets should be submitted via the EGA Data Access Committee (https://ega-archive.org), whereas requests for G-CARD datasets should be directed to the G-CARD Data Access Committee via the metadata portal (http://vanilla.hgc.jp) or by email at g-card_support@hgc.jp. The review process is typically completed within approximately two weeks after receipt of a complete application. Data use is governed by data use agreements that prohibit any attempt to identify individual participants, restrict use to the approved research purposes, and do not permit redistribution.
Processed data generated in this study are available on Zenodo (doi:10.5281/zenodo.17470436). ATAC-seq data for sorted normal blood and AML cells were obtained from Gene Expression Omnibus (GEO) database under the accession number GSE7491221. Normalized gene count data for external AML cohorts were obtained from the previous paper10. Reference genomes for H.sapiens (hg19 and hg38) are available at https://support.illumina.com/sequencing/sequencing_software/igenome.html. Public dataset used in this study is summarized in STable 5. The data generated and/or analyzed during the current study are available from the corresponding authors on reasonable request.

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	All patients in this study were male or female according to biological sex, and no information on gender identity was obtained.
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	Adult AML patients were included in this study. While no preselection criteria were applied with respect to clinical characteristics, sample inclusion was contingent upon the availability of viable cryopreserved tumor cells suitable for ATAC-seq for consecutively diagnosed AML patients. Diagnoses were made at each participating institution and hospital in accordance with the WHO classification in use at the time. In Sweden, patients were registered through the national AML registry, which prospectively collects clinical and genomic data on all newly diagnosed cases. In Japan, patients were diagnosed and treated at participating hospitals, with biospecimens and clinical information sent to and managed by the Kyoto University biobank, where clinical annotations were updated annually. Treatment was administered according to institutional standards of care and, for Swedish patients, according to the National Swedish Treatment Guidelines for AML, with a subset of patients participating in clinical trials. All samples analyzed in this study were collected at the time of initial diagnosis before treatment. Detailed clinical annotations, including diagnosis, demographic and laboratory data, treatment

	regimens, and outcomes, were extracted from electronic medical records and the Swedish AML Registry. Patient characteristics and diagnoses are summarized in STables 1-2.
Recruitment	Patients were consecutively enrolled from the participating institutes and hospitals between 1997 and 2022 in Sweden and 2011 and 2022 in Japan.
Ethics oversight	This study was reviewed and approved by the Regional Ethics Review Board in Stockholm (nos. 2008/1330-31/2 2017/2085-31/2) and the institutional ethics committees of the participating institutions (Kyoto University, Hyogo Prefectural Amagasaki General Medical Center, Chugoku Central Hospital, Dokkyo Medical University Saitama Medical Center, Gifu University Hospital, Gifu Municipal Hospital, Hyogo Medical University, Hokkaido University, Japanese Red Cross Kyoto Daini Hospital, Kobe City Medical Center General Hospital, Kurashiki Central Hospital, Kitano Hospital, Kyoto Medical Center, Kyoto City Hospital, Matsushita Memorial Hospital, Japanese Red Cross Nagano Hospital, National Cancer Center Hospital, NTT Medical Center Tokyo, Osaka International Cancer Institute, Japanese Red Cross Osaka Hospital, The University of Osaka, Otsu Red Cross Hospital, Shizuoka City Shizuoka Hospital, Shinko Hospital, Shiga General Hospital, Sumitomo Hospital, Takeda General Hospital, Takatsuki Red Cross Hospital, Uji Tokushukai Medical Center, and Japanese Red Cross Society Wakayama Medical Center in Japan) (no. G608) and was performed in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants at participating institutions.

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	No statistical methods were used to predetermine sample size, as this was a retrospective study. Sample sizes were determined by the availability of eligible patient samples with sufficient material. The cohort size (n = 1,563) is substantially large and sufficient to enable robust and reproducible analyses. The large sample size allowed the identification of consistent epigenomic subgroups and associated molecular and clinical features, which were supported by concordant results across independent data modalities and by validation in independent cohorts.
Data exclusions	No data were excluded from the analysis if they met the minimum data quality requirements.
Replication	Experimental replication was not performed. The robustness of the findings is supported by the large cohort size (n=1,563) and the consistency of results across multiple independent analytical approaches, data types, and independent cohorts.
Randomization	Randomization was not applicable to this study, as this was a retrospective analysis of patient samples. Samples were collected independently of study design. Group assignment (ATAC-based clustering) was performed after data collection using unsupervised computational methods, thereby avoiding investigator-driven allocation.
Blinding	Blinding was not applicable to this study. Group assignment was not predefined but instead determined after data collection using unsupervised computational clustering, and therefore investigators were not involved in assigning samples to groups during data collection. For downstream analyses, data processing and statistical analyses were performed using standardized computational pipelines with predefined parameters, minimizing potential bias. No manual intervention based on group identity was applied during data analysis.

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	K562, KG-1, THP-1, SKM-1, KY821, and NOMO-1 cells were obtained from the RIKEN BioResource Center Cell Bank (Tsukuba, Japan); Kasumi-1, HL-60, MOLM-13 and TF-1 cells from the American Type Culture Collection (ATCC); and OCI-AML3 cells from the German Collection of Microorganisms and Cell Cultures (DSMZ).
Authentication	These cell lines were authenticated by short tandem repeat profiling by the providing cell banks.
Mycoplasma contamination	These cell lines were tested for mycoplasma by the providing cell banks.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

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.*	The sequencing datasets are available from the European Genome-phenome Archive (EGA) under accession number EGAS00001008315.
Files in database submission	ChIP-seq data for H3K27ac, H3K27me3, SMC1, CTCF, and Pol II from 240 patients are available. All available files are summarized in ST3.
Genome browser session (e.g. UCSC)	No longer applicable.

Methodology

Replicates	ChIP-seq experiments were conducted on patient samples with input controls but without replicates.
Sequencing depth	Each sample was sequenced in paired-end mode, generating approximately 35 million reads on average.
Antibodies	The antibodies used for ChIP were as follows: SMC1 (Abcam, ab9262), CTCF (Cell Signaling Technology, D31H2), Pol II (CST, D8L4Y), H3K27ac (Cell Signaling Technology, D5E4), and H3K27me3 (Cell Signaling Technology, C36B11).
Peak calling parameters	Peaks were called using MACS and a P-value threshold of 1 x 10–3 with an input control for each sample.
Data quality	The quality of sequencing data was assessed using 'plotFingerprint' in deeptools.
Software	The sequencing reads were aligned to the reference human hg19 genome using bowtie following trimming of adapters by skewer and read tails to a total length of 50 bp using cutadapt. The quality of sequencing data was assessed using 'plotFingerprint' in deeptools. Samples were excluded from the analysis if any of the following criteria were met: 'X-intercept' (deeptools) >0.85 or 'Synthetic JS Distance' (deeptools) <0.225. After removing duplicates and reads on blacklisted regions (ENCODE), peaks were called using MACS and a P-value threshold of 1 x 10–3 with an input control for each sample. For each ChIP-seq (SMC1, CTCF, Pol II, H3K27ac, and H3K27me3), we merged peaks from all AML samples, and recurrently identified peaks were regarded as a consensus peak set.

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	Cryopreserved cells were thawed, stained with DAPI to exclude dead cells, and live mononuclear cells were sorted using the FACS Aria III (BD Biosciences).
Instrument	FACS Aria III (BD Biosciences)
Software	BD FACSDiva
Cell population abundance	Typically >90% viable cells, as determined by DAPI exclusion.
Gating strategy	Cells were first gated based on forward and side scatter (FSC/SSC) to exclude debris, followed by doublet discrimination using FSC-A versus FSC-H. Live cells were defined as DAPI-negative and were subsequently sorted.

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