nature research

Corresponding author(s): Lisa M. Mcginnis, Sandy Klemm, William J. Greenleaf,

Last updated by author(s): 10/22/2019

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 Authors & Referees 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	Cell Ranger ATAC 1.0.0 – Barcode Identification, Alignment, Filter, Deduplication

Data analysis	macs2 2.1.1.20160309 – Peak Calling R version 3.5.1 – R environment for all custom code Irlba 2.3.2 – Running PCA/SVD on large matrices. Reticulate 1.10 – Used for running Python UMAP implementation within R. Rcpp 1.0.0 – Used for writing helpful C++ code to speed up operations. matrixStats 0.54.0 – Used for mathematical operations on large matrices. Cicero 1.0.13 – Used for calculating gene activity scores. chromVAR_1.2.0 – Calculating TF deviation scores which can be associated with TF activity. SingleCellExperiment 1.2.0 – R Data Class Environment used throughout analyses. Motifmatchr 1.2.0 – Matching TF Motifs within peak regions. Seurat 2.3.4 – SNN Graph Clustering Implementation. GenomicFeatures 1.32.2 – Genomic Ranges Operations used for overlap analyses. GenomicRanges 1.32.7 - Genomic Ranges Operations used for overlap analyses. Matrix 1.2-14 – Sparse Matrix math implementations. BSgenome 1.48.0 – Toolkit used for getting Genomic DNA sequences for motif matching and footprinting. Rsamtools 1.32.3 – For manipulating BAM files within R. edgeR 3.24.3 - Toolkit for analyzing differential RNA-seq. FNN 1.1.3 - Identifying Nearest Neighbors. uwot 0.1.4 - Creation of UMAP Embeddings in R. BWA 0.7.17 - Alignment of Fastqs to genomic reference DNA. stats (R 3.5.1) - R base statistical software. survival 2.42-3 - Survival Analysis software used for TCGA analysis. clusterProfiler 3.10.1- enrichment of GO/KEGG terms in gene sets.

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/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

CODE AVAILABILITY
Code used in this study will be posted on GitHub for main analyses (https://github.com/GreenleafLab/MPAL-Single-Cell-2019).

DATA AVAILABILITY
Sequencing data will be deposited in the Gene Expression Omnibus (GEO). There are no restrictions on data availability or use. The main single-cell matrices are stored as Bioconductor Summarized Experiment's (scRNA, scATAC, chromVAR, Log Gene Scores) for each of the main analyses will be accessible from GitHub.

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	Our sample size was determined by the number of MPAL patients diagnosed at Stanford since 2011 (n = 5). We performed multi-omic analyses on these samples in technical duplicate, profiling over 50,000 single cells.
Data exclusions	No inclusion or exclusion criteria were used for human studies. No data were excluded from the manuscript.
Replication	All results presented in manuscript were reliably reproduced across technical duplicate. Additionally we compared differential results to previously published data sets.
Randomization	No randomization of human participants was used.
Blinding	No blinding was used.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used

FACS for MPAL Patients:

BD Biosciences Catalog - 340662 - CD3 (clone SK7) PE
BD Biosciences Catalog - 340738- CD7 (M-T701) FITC
BD Biosciences Catalog - 340952 - CD45 (clone 2D1) PerCP-Cy™5.5
BD Biosciences Catalog - 555824 - CD34 (clone 581) APC
BD Biosciences Catalog - 340722- CD19 (clone SJ25C1) APC
BD Biosciences Catalog - 340673- CD20 (clone L27) FITC
BD Biosciences Catalog - 340654 - CD79a (clone HM47) PE
BD Biosciences Catalog - 561708 - CD14 (M5E2) APC
BD Biosciences Catalog - 644386 - CD64 (clone 10.1) PE
Agilent - F071401-1- MPO (clone MPO-7) FITC
Beckman Coulter - IM0775U - CD38 (clone LS198-4-3) FITC
Beckman Coulter - IM3524U - TdT (HT1 + HT4 + HT* + HT9) FITC

Antibody Derived Tags (ADT)-sequencing: all antibodies are from biolegend:

Antibody list for PBMC & BMMC
Category Specificity Name Clone Reactivity Barcode Sequence
TotalSeq™ Catalog - 300477 -B CD3 CD3_TotalB UCHT1 Human AACAAGACCCTTGAG
TotalSeq™ Catalog - 300565 -B CD4 CD4_TotalB RPA-T4 Human TACCCGTAATAGCGT
TotalSeq™ Catalog - 301069 -B CD8a CD8a_TotalB RPA-T8 Human ATTGGCACTCAGATG
TotalSeq™ Catalog - 301857 -B CD14 CD14_TotalB M5E2 Human GAAAGTCAAAGCACT
TotalSeq™ Catalog - 323051 -B CD15 CD15_TotalB W6D3 Human ACGAATCAATCTGTG
TotalSeq™ Catalog - 302063 -B CD16 CD16_TotalB 3G8 Human GTCTTTGTCAGTGCA
TotalSeq™ Catalog - 302647 -B CD25 CD25_TotalB BC96 Human GTGCATTCAACAGTA
TotalSeq™ Catalog - 302263 -B CD19 CD19_TotalB HIB19 Human TCAACGCTTGGCTAG
TotalSeq™ Catalog - 304161 -B CD45RA CD45RA_TotalB HI100 Human GATGAGAACAGGTTT
TotalSeq™ Catalog - 304257 -B CD45RO CD45RO_TotalB UCHL1 Human TGCATGTCATCGGTG
TotalSeq™ Catalog - 392423 -B CD56 (NCAM) Recombinant CD56_TotalB QA17A16 Human GTTGTCCGACAATAC
TotalSeq™ Catalog - 329961 -B PD-1 PD-1_TotalB EH12.2H7 Human AAGTCGTGAGGCATG
TotalSeq™ Catalog - 372727 -B TIGIT TIGIT_TotalB A15153G Human TGAAGGCTCATTTGT
TotalSeq™ Catalog - 351354 -B CD127 CD127_TotalB A019D5 Human ACATTGACGCAACTA
TotalSeq™ Catalog - 400291 -B Mouse_IgG2a MouseIgG2a_TotalB MOPC-173 Mouse CTCTATTCAGACCAG
TotalSeq™ Catalog - 400185 -B Mouse_IgG1 MouseIgG1_TotalB MOPC-21 Mouse ACTCACTGGAGTCTC
TotalSeq™ Catalog - 400379 -B Mouse_IgG2b MouseIgG2b_TotalB MPC-11 Mouse ATCACATCGTTGCCA

Antibody list for CD34+ BM & MPAL samples

Category Barcode Specificity Clone Reactivity Barcode Sequence
TotalSeq™ Catalog - 100251 -A 0049 CD3 SK7 Human TATCCCTTGGGATGG
TotalSeq™ Catalog - 317451 -A 0045 CD4 SK3 Human GAGGTTAGTGATGGA
TotalSeq™ Catalog - 343123 -A 0066 CD7 CD7-6B7 Human TGGATTCCCGGACTT
TotalSeq™ Catalog - 301067 -A 0080 CD8a RPA-T8 Human GCTGCGCTTTCCATT
TotalSeq™ Catalog - 312231 -A 0062 CD10 HI10a Human CAGCCATTCATTAGG
TotalSeq™ Catalog - 367131 -A 0081 CD14 M5E2 Human TCTCAGACCTCCGTA
TotalSeq™ Catalog - 302259 -A 0050 CD19 HIB19 Human CTGGGCAATTACTCG
TotalSeq™ Catalog - 366629 -A 0052 CD33 P67.6 Human TAACTCAGGGCCTAT
TotalSeq™ Catalog - 304157 -A 0063 CD45RA HI100 Human TCAATCCTTCCGCTT
TotalSeq™ Catalog - 328135 -A 0060 CD90 (Thy1) 5E10 Human GCATTGTACGATTCA
TotalSeq™ Catalog - 306037 -A 0064 CD123 6H6 Human CTTCACTCTGTCAGG

Validation	Highly optimized flow cytometric analysis was used for this study that has been shown to be specific in previous studies. For the flow cytometry analysis of MPALs, we used lymphocytes as a control for validation for the antibodies. Additionally, BD Biosciences, Agilent and Beckman Coulter have validated these antibodies directly by testing vs an isotype control. For the antibody derived tags (ADT) we first performed this analysis on human peripheral blood and bone marrow segregating known subpopulations. Additionally, Biolegend tests for every antibody by staining multiple target cells (with positive and negative controls) and with serial dilutions to make sure the titer is appropriate.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Healthy human subjects were male and female, ages 20-50. MPAL patients were male and female, ages 20-75. MPAL patients were diagnosed with mixed phenotype acute leukemia according to WHO 2016 classification.
Recruitment	No selective recruitment of healthy subjects was performed. MPAL samples were obtained from patients at Stanford Hospital and Clinics with acute leukemia. MPAL is an extremely rare leukemia and all patients with a diagnosis of MPAL were included in the study, regardless of age. This study did not attempt to exclude healthy volunteers or patients based on sex, race, or ethnicity.
Ethics oversight	This study was approved by the Stanford University Administrative Panels on Human Subjects in Medical Research, and written informed consent was obtained from all participants.

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

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	Peripheral blood and bone marrow aspirate samples were processed by Lymphoprep (STEMCELL Technologies) gradient centrifugation and fresh frozen in Bambanker media.
Instrument	Flow cytometry was performed on a FACSCalibur or FACSCanto II (Becton Dickinson, San Jose, Ca, USA) cytometer using commercially available antibodies.
Software	Data was analyzed using FlowJo v10 software.
Cell population abundance	No sorting was performed on various samples. 30,000 events were collected and analyzed on all MPAL samples.
Gating strategy	Cells were first Live/Dead selected (using 7-AAD). Lymphocytes were identified by low side-scatter and bright CD45 expression. The gate was validated by backgating on CD3-positive or CD19-positive events. Blasts were identified by low side-scatter and dim CD45 expression. The gate was further assessed by backgating on CD34-positive events. Gates were drawn by additionally using isotype controls and internal positive and negative controls.

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