blood advances

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodadvances@hematology.org

Progenitor like cell type of an MLL-EDC4 fusion in acute myeloid leukemia

Tracking no: ADV-2022-009096R1

Linda Schuster (German Cancer Research Center (DKFZ), Germany) Afzal Syed (German Cancer Research Center, Germany) Stephan Tirier (German Cancer Research Center (DKFZ), Germany) Simon Steiger (German Cancer Research Center (DKFZ), Germany) Isabelle Seufert (German Cancer Research Center (DKFZ), Germany) Heiko Becker (University Freiburg Medical Center, Germany) Jesus Duque-Afonso (University of Freiburg Medical Center, Germany) Tobias Ma (University of Freiburg Medical Center, Germany) Seishi Ogawa (Kyoto University, Japan) Jan-Philipp Mallm (Heidelberg University, Germany) Michael Lübbert (Medical Center - University of Freiburg, Germany) Karsten Rippe (German Cancer Research Center (DKFZ), Germany)

Abstract:

Conflict of interest: No COI declared

COI notes: The authors declare no competing financial interests.

Preprint server: No;

Author contributions and disclosures: Study design and coordination: ML, KR. Acquisition of patient samples: HB, TM, JDA, SO. Acquisition of data: LCS, JPM. Analysis of data and data curation: LCS, APS, SMT, SS, IS, JPM, KR. Drafting of manuscript: LCS, KR. Review and editing of manuscript: all authors. Supervision: KR, ML.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: The scRNA-seq data are available as read count matrices at the Zenodo open repository at http://www.doi.org/10.5281/zenodo.7832876

Clinical trial registration information (if any):

Progenitor like cell type of an *MLL-EDC4* fusion in acute myeloid leukemia

Linda C. Schuster[1,2,3], Afzal P. Syed[1,2,4], Stephan M. Tirier[1], Simon Steiger[1,2,3], Isabelle Seufert[1,2,3], Heiko Becker[5], Jesus Duque-Afonso[5], Tobias Ma[5], Seishi Ogawa[6], Jan-Philipp Mallm[2,4], Michael Lübbert[5,*] and Karsten Rippe[1,2,*]

[1] German Cancer Research Center (DKFZ) Heidelberg, Division of Chromatin Networks, Germany

[2] Center for Quantitative Analysis of Molecular and Cellular Biosystems (BioQuant), Heidelberg University, Germany

[3] Faculty of Biosciences, Heidelberg University, Germany

[4] Single Cell Open Lab, German Cancer Research Center (DKFZ), Heidelberg, Germany

[5] Department of Medicine I, University Freiburg Medical Center, Freiburg, Germany

[6] Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan

* Corresponding authors:

Karsten Rippe, German Cancer Research Center (DKFZ), Division Chromatin Networks, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany, e-mail: karsten.rippe@dkfz.de

Michael Lübbert, Department of Medicine I, University Freiburg Medical Center, Hugstetter Straße 55, 79106 Freiburg, Germany, e-mail: michael.luebbert@uniklinik-freiburg.de

Keywords: acute myeloid leukemia (AML), mixed-lineage leukemia (MLL) gene, MLL-rearranged, enhancer of mRNA decapping 4 (*EDC4*) gene, single cell RNA sequencing

Supplemental data: The article is accompanied by supplemental data.

Data availability: The scRNA-seq data are available via the Zenodo open repository at http://www.doi.org/10.5281/zenodo.7832876. Analysis scripts are provided at Github from the link https://github.com/RippeLab/MLL-EDC4. Other data are available on request from the corresponding authors Karsten Rippe (karsten.rippe@dkfz.de) and Michael Lübbert (michael.luebbert@uniklinik-freiburg.de).

Counts (Research Letter): Abstract: **172** (max. 200), Text: **1167** (max. 1200), Figures/tables: **2** (max. 2), References: **25** (max. 25)

Translocations involving the mixed lineage leukemia (*MLL/KMT2A*) gene generally confer poor prognosis in acute myeloid leukemia (AML) and display a large inter- and intratumor heterogeneity[1]. By conducting a single cell RNA sequencing analysis (scRNA-seq), the different developmental stages along the hematopoietic stem cell (HSC) to myeloid trajectory can be resolved, which is relevant for self-renewal, interactions of leukemic cells with non-malignant cells in the microenvironment and therapy resistance[2-5]. However, information on *MLL*-rearranged (*MLL*-r) cases of AML is scarce as previous scRNA-seq studies of AML by van Galen et al.[2] and Slush et al.[3] include only one *MLL*-r patient each. In our previous work we have described a novel *MLL* fusion with the enhancer of mRNA decapping 4 (*EDC4*) gene (*MLL-EDC4*)[6] for which recently another case has been reported[7].

Here, we dissected cell types and developmental stages in five AML patients by scRNA-seq to compare the novel *MLL-EDC4* translocation to *MLL-MLLT3* and *MLL-ELL* fusions (**supplemental Table 1**). Mononuclear cells were collected from peripheral blood or bone marrow and subjected to scRNA-seq to yield 17,600 cells as described in further detail in the Supplemental Information. Transcriptome features of the merged scRNA-seq data obtained from the five patient samples were visualized by uniform manifold approximation and projection (UMAP) and clustering (**Figure 1A, B; supplemental Figure 1A, B**). We then annotated leukemic vs. non-malignant cells according to marker gene expression profiles and validated the results with the chromosome ploidy computed from the scRNA-seq data (**Figure 1C**). The scRNA-seq analysis revealed a significant intratumor heterogeneity of the *MLL-MLLT3* #2, *MLL-MLLT3* #3 and *MLL-ELL* patients with two distinctive clusters (c1, c2) of leukemic cells. In contrast, the *MLL-EDC4* and *MLL-MLLT3* #1 samples showed a more homogeneous phenotype. Non-malignant cells determined by marker gene expression clustered per cell type across all patients without further batch correction while leukemic cells from each patient sample clustered individually.

We characterized the differentiation state of leukemic cells with an automated cell type prediction approach using the Human Cell Atlas (HCA)[8] bone marrow dataset from eight healthy donors as a training data set. Genes signatures and scores for the different cell types were assigned based on the most expressed cell type markers from the HCA data (**Figure 1D-F**; **supplemental Figure 1C-E**, **supplemental Table 2**). Leukemic cells with *MLL-EDC4* translocation represented a distinct leukemic cell cluster and were almost exclusively classified as HSCs, multipotent progenitors (MPPs) or erythroblasts (ERPs), which is in line with their $CD34^{+}/CD14^{-}$ FACS signature (**supplemental Table 3**). In contrast, malignant cells from the common *MLL*-fusions presented a more differentiated phenotype that unveiled a trajectory from myeloid progenitors to monocyte-like cells from cluster 2 to 1 for *MLL-MLLT3* #2, *MLL-MLLT3* #3 and *MLL-ELL* (**Figure 1D; supplemental Figure 1F**). Interestingly, a fraction of cells from cluster 2 of *MLL-MLLT3* #3 stand out as it displayed signatures of MPP (21%) and HSC (14%) cells (**Figure 1E**, **F**). The *MLL-EDC4* patient showed elevated module scores for HSC- and MPP-genes, while no upregulation in monocytic *CD14+ related* genes was evident (**Figure 1E, F**). This phenotype was also partly present in cluster 2 of *MLL-MLLT3* #3 as apparent from the bimodal distribution of the violin plot and the low monocyte score of the whole cluster and in a minor fraction of cluster 2 from *MLL-ELL*. In contrast, leukemic cells from *MLL-MLLT3* #1 and #2 patients and cluster 1 of *MLL-MLLT3* #3 and *MLL-ELL* showed an almost opposite pattern. The analysis of the microenvironment revealed monocytes with an unusual gene expression signature in *MLL-EDC4* that was characterized by expression of CD36, cathepsins and CLEC receptors (**supplemental Figure 1G**)[9].

Next, we performed a differential gene expression analysis of gene sets and pathways for the different *MLL*-r cases. Gene set enrichment analysis (GSEA) showed a downregulation of myeloid leukocyte mediated immunity and activation and a dampened immune response in the *MLL-EDC4* positive leukemic cells. Pathways associated with *MYC* targets, interferon alpha response, eukaryotic translation initiation or elongation and reactive oxygen species (ROS) were upregulated (**Figure 2A**). The upregulation of various ribosomal proteins in *MLL-EDC4* positive AML may be linked to the malignant transformation of cells[10]. Furthermore, the upregulation of ROS pathways has been shown to interfere with hematopoiesis due to an increase in oxidative stress causing genomic instability[11]. Transcriptomes of leukemic cells from the *MLL-MLLT3* and *MLL-ELL* positive AML displayed an upregulation of classical monocyte markers in contrast to *MLL-EDC4* (**Figure 2B**). Interestingly, the most differentially expressed gene in patient *MLL-EDC4* positive cells was lactate dehydrogenase B (*LDHB),* mediating the switch on the anaerobic glycolysis and lactate production that could reflect a high proliferation rate of leukemic cells (Warburg effect) and/or adaption to hypoxia[12].

The *MLL-EDC4* fusion showed a distinctive upregulation of genes known to have an impact on cell-fate decision and cellular differentiation in hematopoiesis and endothelial-to-hematopoietic transition (*NPM1, CDK6, SOX4, GATA2, MYC, DACH1*) or leukemic stem cell activation (*FLT3, HOPX*, *HOXA9, RUNX1*)[13-19] (**Figure 2B**). It is noted that transcription factors (TFs) like *SOX4*, *GATA2*, *MYC and RUNX1* are well established master regulators of stem cell programs. These findings prompted us to systematically evaluate TF expression and their activity based on target gene expression. Compared to the other fusions (**Figure 2C**), *MLL-EDC4* displayed an increased activity of interferon-related TFs such as *STAT2* and *IRF9*, of oncogenes *MYC*, *MYB* as well as other TFs like *E2F4, ETS1, GATA1, NFYA, POU2F1, SPI1, TAL1* that have been linked to stemness in hematopoietic cells[20-22]. Based on this data, a network of interacting TFs was generated (**Figure 2D**). Unsupervised clustering highlighted MYC as a central node in the network that is linked to many TFs as first or second edge. *MYC* is known to play a crucial role in cell growth, proliferation and tumorigenesis[23]. In addition, TF activity showed an upregulation of *POU2F1* in *MLL-EDC4* positive AML, which can function in cell growth control, cellular stress response, stem cell identity and immune regulation[24]. Finally, activity of hematopoietic key regulator *RUNX1* was high as inferred from the aberrant expression of its downstream targets *UBB, PSNE1,* ARID1B and *KIAA0125* involved in differentiation of myeloid cells[25].

In summary, our scRNA-seq analysis of *MLL* fusions in AML reveals variable degrees of intratumor heterogeneity and differentiation stages. The *MLL-EDC4* positive AML is associated with a more primitive cell differentiation state compared to *MLL-MLLT3* or *MLL-ELL*. The unique hematological progenitor-like cell type in our *MLL-EDC4* case is evident from an extensive upregulation of a network of transcription factors that are known to be crucial for differentiation block and leukemic development. Furthermore, a fraction of leukemic cells with an HSC/progenitor cell type in one cluster of the *MLL-MLLT3* #3 sample was detected, which points to a complex interplay of MLL fusion partner and the cell type that develops the AML initiating translocation. It is well established that a more stem cell like phenotype is highly relevant for prognosis and therapy response[2-5]. Accordingly, it will be important to extend the approach described here to a larger patient cohort to reveal the relation between the developmental stage along the myeloid trajectory and clinical parameters for different *MLL* fusions.

Informed consent was obtained from all participants involved in the research reported in the manuscript at the Department of Medicine I of University Freiburg Medical Center.

Acknowledgements

The research was supported by the Deutsche Forschungsgemeinschaft (DFG) via grants RI 1283/15-2 (K.R.), LU 429/16-2 (M.L.), MA 7792/1-2 (J.-P.M.) and BE 6461/1-2 (H.B.) within Research Group FOR2674 and via subproject Z1 within SFB1074.

Contributions

Study design and coordination: ML, KR. Acquisition of patient samples: HB, TM, JDA, SO. Acquisition of data: LCS, JPM. Analysis of data and data curation: LCS, APS, SMT, SS, IS, JPM, KR. Drafting of manuscript: LCS, KR. Review and editing of manuscript: all authors. Supervision: KR, ML.

Conflict-of-interest disclosure

The authors declare no competing financial interests.

Figure legends

Figure 1. Intratumor heterogeneity and cell type assignment of *MLL-r samples.* (**A**) UMAP embedding of all AML samples colored by patient. (**B**) UMAP embedding colored by cell types determined from marker gene expression. AML cells form separate clusters for each patient whereas non-malignant cell types from different samples cluster together. (**C**) UMAP embedding colored by ploidy with AML cells annotated as aneuploid (red) and microenvironment cells as diploid (cyan). (**D**) UMAP embedding of AML cells colored by cell type prediction with SingleR according to the HCA as reference data set. (**E**) Pie charts of predicted cell type composition for AML cell clusters. (**F**) Violin plots of myeloid cell signature module scores according to **supplemental Table 2** for AML cell clusters. c1, cluster1; c2, cluster2; HSC, hematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte-monocyte progenitor; preDC, pre-dendritic cell; cDC2, type 2 conventional dendritic cell; Mono, CD14+ monocytes.

Figure 2. Gene expression and transcription factor activity in *MLL-EDC4* compared to other *MLL-r* cases. (**A**) Enriched gene sets in up- and down-regulated genes of *MLL-EDC4* AML cells compared to all other AML cells visualized as dot plots. Gene sets from Hallmark, Reactome and GO:BP were used. (**B**) Clustered single-cell transcriptomic heatmap of the most differentially expressed genes between AML cell clusters (**C**) Heatmap of transcription factor activities for AML cells based on scRNA-seq data. (**D**) Transcription factor network colored by transcription factor activity.

References

1. Meyer C, Burmeister T, Groger D, et al. The MLL recombinome of acute leukemias in 2017. *Leukemia*. 2018;32(2):273-284.

2. van Galen P, Hovestadt V, Wadsworth Ii MH, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*. 2019;176(6):1265-1281 e1224.

3. Shlush LI, Mitchell A, Heisler L, et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature*. 2017;547(7661):104-108.

4. Zeng AGX, Bansal S, Jin L, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat Med*. 2022;28(6):1212-1223.

5. Bottomly D, Long N, Schultz AR, et al. Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell*. 2022;40(8):850-864 e859.

6. Becker H, Greve G, Kataoka K, et al. Identification of enhancer of mRNA decapping 4 as a novel fusion partner of MLL in acute myeloid leukemia. *Blood Adv*. 2019;3(5):761-765.

7. Hu DY, Wang M, Shen K, et al. A new breakpoint fusion gene involving KMT2A::EDC4 rearrangement in de novo acute myeloid leukemia. *Int J Lab Hematol*. 2023;45(4):596-598.

8. Regev A, Teichmann SA, Lander ES, et al. The Human Cell Atlas. *Elife*. 2017;6.

9. Jakos T, Pislar A, Jewett A, Kos J. Cysteine Cathepsins in Tumor-Associated Immune Cells. *Front Immunol*. 2019;10:2037.

10. Warner JR, McIntosh KB. How common are extraribosomal functions of ribosomal proteins? *Mol Cell*. 2009;34(1):3-11.

11. Richardson C, Yan S, Vestal CG. Oxidative stress, bone marrow failure, and genome instability in hematopoietic stem cells. *Int J Mol Sci*. 2015;16(2):2366-2385.

12. Du Y, Zhang MJ, Li LL, et al. ATPR triggers acute myeloid leukaemia cells differentiation and cycle arrest via the RARalpha/LDHB/ERK-glycolysis signalling axis. *J Cell Mol Med*. 2020;24(12):6952-6965.

13. Menendez-Gonzalez JB, Sinnadurai S, Gibbs A, et al. Inhibition of GATA2 restrains cell proliferation and enhances apoptosis and chemotherapy mediated apoptosis in human GATA2 overexpressing AML cells. *Sci Rep*. 2019;9(1):12212.

14. Argiropoulos B, Humphries RK. Hox genes in hematopoiesis and leukemogenesis. *Oncogene*. 2007;26(47):6766-6776.

15. Tsapogas P, Mooney CJ, Brown G, Rolink A. The Cytokine Flt3-Ligand in Normal and Malignant Hematopoiesis. *Int J Mol Sci*. 2017;18(6).

16. Li BE, Ernst P. Two decades of leukemia oncoprotein epistasis: the MLL1 paradigm for epigenetic deregulation in leukemia. *Exp Hematol*. 2014;42(12):995-1012.

17. Lee JW, Kim HS, Hwang J, et al. Regulation of HOXA9 activity by predominant expression of DACH1 against C/EBPalpha and GATA-1 in myeloid leukemia with MLL-AF9. *Biochem Biophys Res Commun*. 2012;426(3):299-305.

18. Ahmadi SE, Rahimi S, Zarandi B, Chegeni R, Safa M. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. *J Hematol Oncol*. 2021;14(1):121.

19. Chen MJ, Yokomizo T, Zeigler BM, Dzierzak E, Speck NA. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*. 2009;457(7231):887-891.

20. Pellicano F, Park L, Hopcroft LEM, et al. hsa-mir183/EGR1-mediated regulation of E2F1 is required for CML stem/progenitor cell survival. *Blood*. 2018;131(14):1532-1544.

21. Saint-Andre V, Federation AJ, Lin CY, et al. Models of human core transcriptional regulatory circuitries. *Genome Res*. 2016;26(3):385-396.

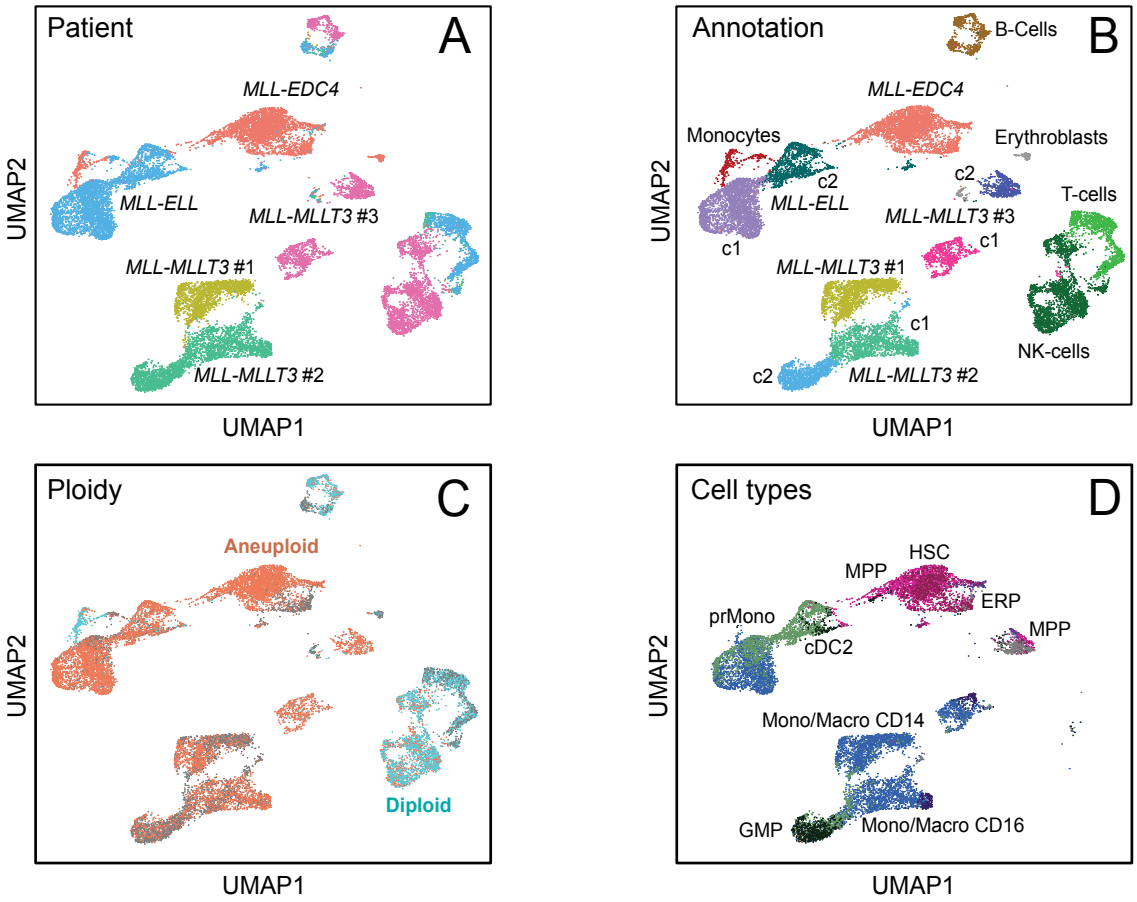
22. Lulli V, Romania P, Riccioni R, et al. Transcriptional silencing of the ETS1 oncogene contributes to human granulocytic differentiation. *Haematologica*. 2010;95(10):1633-1641.

23. Dang CV. MYC on the path to cancer. *Cell*. 2012;149(1):22-35.

24. Vazquez-Arreguin K, Tantin D. The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. *Biochim Biophys Acta*. 2016;1859(6):792-804.

25. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.

Figure 1



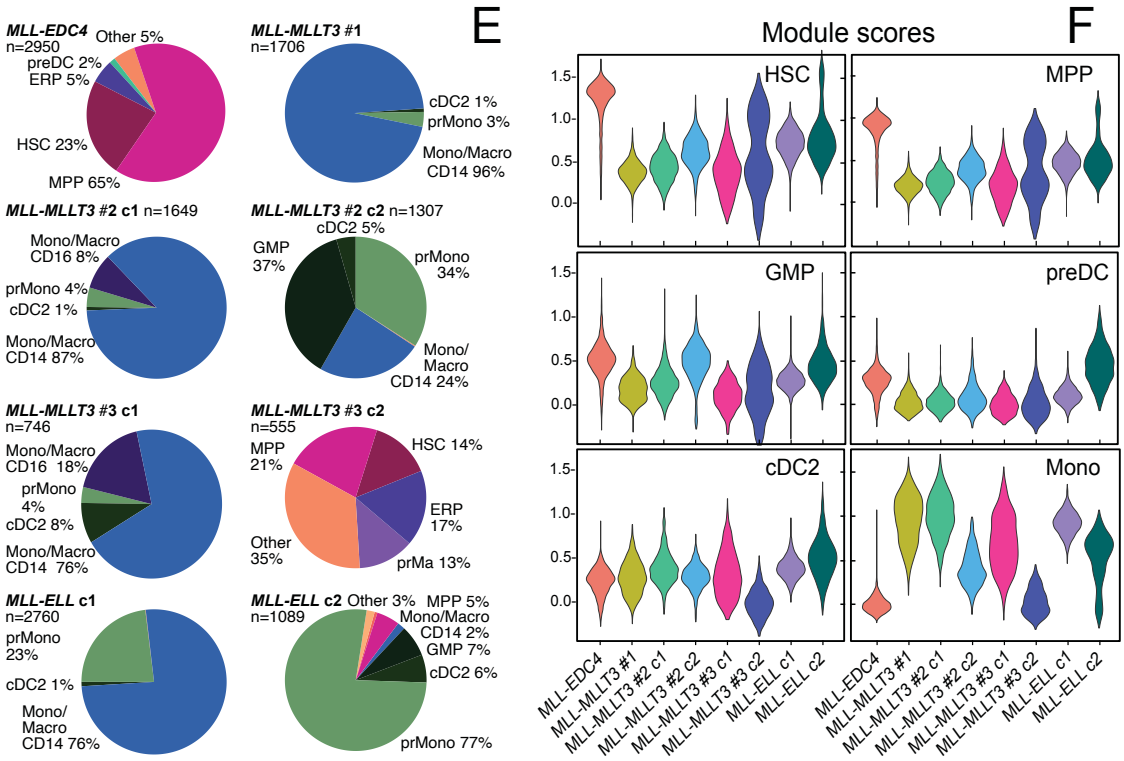
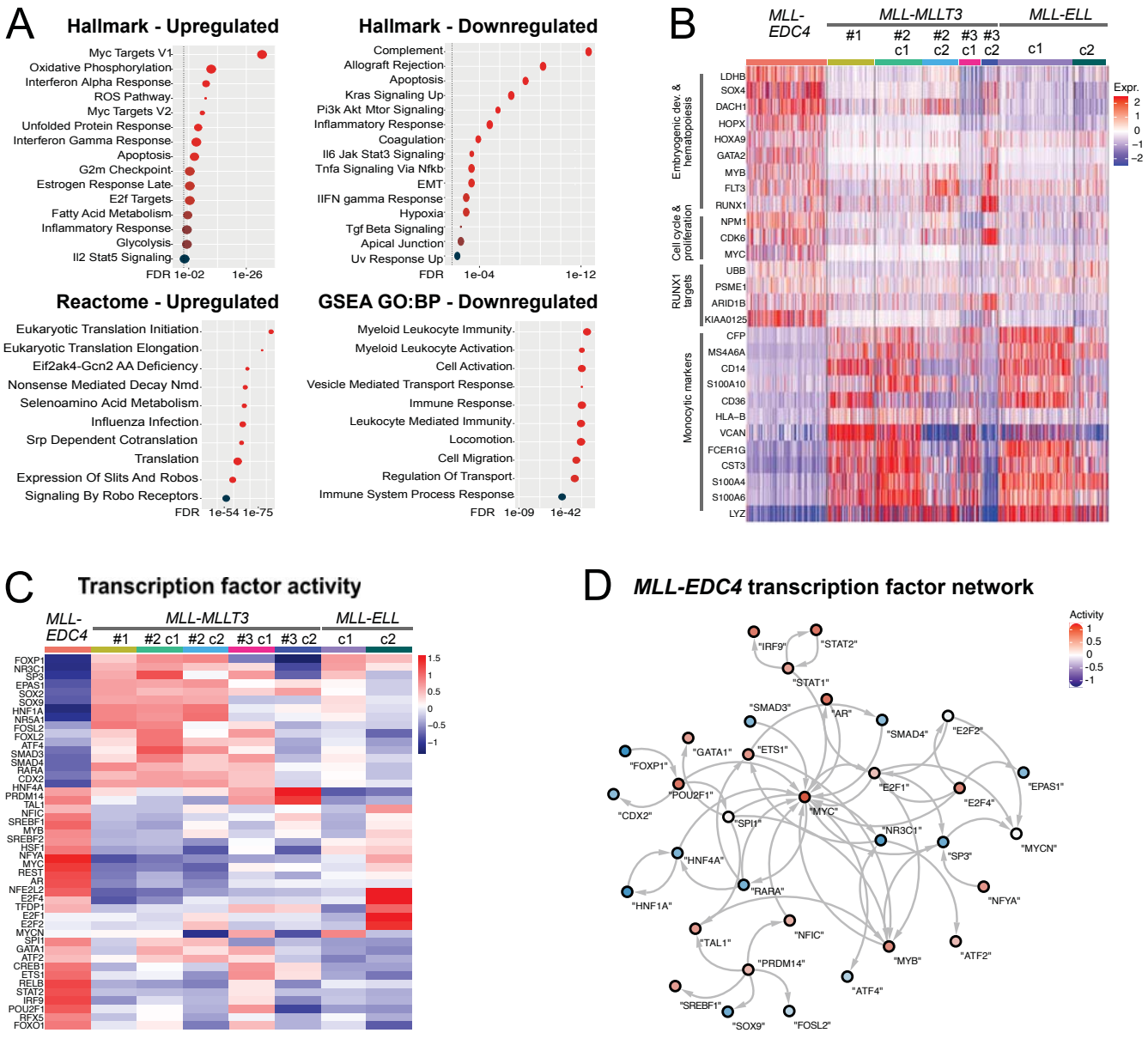


Figure 2



Supplemental files to

Progenitor like cell type of an *MLL-EDC4* fusion in acute myeloid leukemia

Linda C. Schuster, Afzal P. Syed, Stephan M. Tirier, Simon Steiger, Isabelle Seufert, Heiko Becker, Jesus Duque-Afonso, Tobias Ma, Seishi Ogawa, Jan-Philipp Mallm, Michael Lübbert and Karsten Rippe

Content

Supplemental Methods

Sample acquisition and clinical data

The study complied with all relevant ethical regulations for working with patients and patient samples in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants. The samples were enriched for mononuclear cells (MNCs) via ficoll-hypaque and depleted from $CD3^+$ cells via autoMACS (Miltenyi Biotec) as described previously[1].

Droplet-based scRNA-seq

Single-cell RNA sequencing was performed with PBMCs and BMNCs on the Chromium platform with the single cell 3' library and gel bead kit v2 (10x Genomics). Approximately 8,000 cells per sample were loaded and libraries were generated according to the manufacturer's protocol. The sequencing-ready library was cleaned up with SPRI-select beads (Beckman Coulter) and sequenced by the DKFZ Core Facility on the Illumina NovaSeq 6000 platform with S1 flow cell and paired-end sequencing with 26 and 96 bp read length.

Preprocessing and quality control of scRNA-seq data

Preprocessing of the scRNA-seq data was performed using Cell Ranger version 3.1.0 (10x Genomics). Each sample was aligned to the human reference genome assembly "refdata-cellranger- GRCh38-1.2.0_premrna" using the Cell Ranger command *count*. The scRNA-seq data are available as read count matrices at the Zenodo open repository at http://www.doi.org/10.5281/zenodo.7832876. Raw expression data were then loaded into R version 4.0.2 and analyzed using the Seurat package[2] version 4.0.0 with the parameters suggested by the developers. Specifically, single-cell profiles with less than 500 detected genes (indicating a dying cell or no cell in a droplet), more than 3000 detected genes (indicating cell doublets), or more than 15% of UMIs derived from mitochondrial genes were discarded. Additionally, cells with a doublet score >0.4 calculated via the Python package Scrublet[3] were removed using following parameters: *sim_doublet_ratio = 2; n_neighbors = 30; expected_doublet_rate = 0.1*.

For comparison of leukemic cells with healthy hematopoietic progenitors, bone marrow scRNA-seq raw-count data collected from eight healthy individuals (census of immune cells)

were downloaded from the HCA data portal[4]. This dataset was generated on the Chromium platform with the same single cell 3' reagent v2 chemistry as used for our experiments. For compatibility Gene symbols of the HCA dataset were converted from GENCODE v27 to v28.

Analysis of scRNA-seq data

Data of all samples was merged and ~17,600 single cell transcriptomes passed quality control criteria. Regularized negative binomial regression was used to normalize UMI count data using *sctransform*[5]. The number of UMIs per cell and the percent of mitochondrial reads per cell were regressed out using Seurat's standard analysis workflow. Data was integrated using canonical correlation analysis (CCA) available in Seurat package. Principal component analysis (PCA) was conducted using the top 3,000 variable genes. 30 and 15 principal components as determined by an elbow plot were used for downstream analysis of non-integrated and integrated data, respectively. K-nearest neighbor (KNN) graph of cells was calculated and used for Louvain-based clustering to assign cells to clusters. Low-dimensional embeddings of non-integrated as well as integrated data were computed using UMAP.

Cell type specific marker genes such as *CD3D*, *MS4A1*, *NKG7*, *HBB* and CD14 were clearly detectable in the scRNA-seq data, enabling a robust marker-based assignment of non-malignant cell identities. Cell clusters with transcriptomic signatures that could not be assigned as microenvironment were labelled as leukemic cells. There leukemic cell state was confirmed by their aneuploidy, which was computed with the R package copyKat[6] for each sample individually using the non-malignant cell types monocytes, T-cell and NK-cell as diploid reference. Differentially expressed genes between groups of cells were identified using Wilcoxon Rank Sum test ($p_{adj} < 0.05$, log2FoldChange > 0.1) using the *FindMarkers* function in Seurat. Gene set enrichment analysis was performed with the hypeR R package[7]. Cell type prediction of leukemic cells was determined via SingleR[8] using a down-sampled HCA data set (~1000 cells per cell type) and associated cell identity labels as training data set. Module scores for HSC- to CD14 monocyte-like transcriptional profiles were calculated via *AddModuleScore* from Seurat and using a signature gene list derived from the HCA data set (**supplemental Table 2**). Pseudotime inference was conducted via Slingshot[9].

Transcription factor activity was inferred from scRNA-seq data using the DoRothEA package[10] with the statistical method VIPER[11]. Only regulons with confidence levels A and B were used. A list of transcription factors that were also regulated by each other was imported into Cytoscape[12] and the transcription factors were colored based on DoRothEAs calculated activity levels.

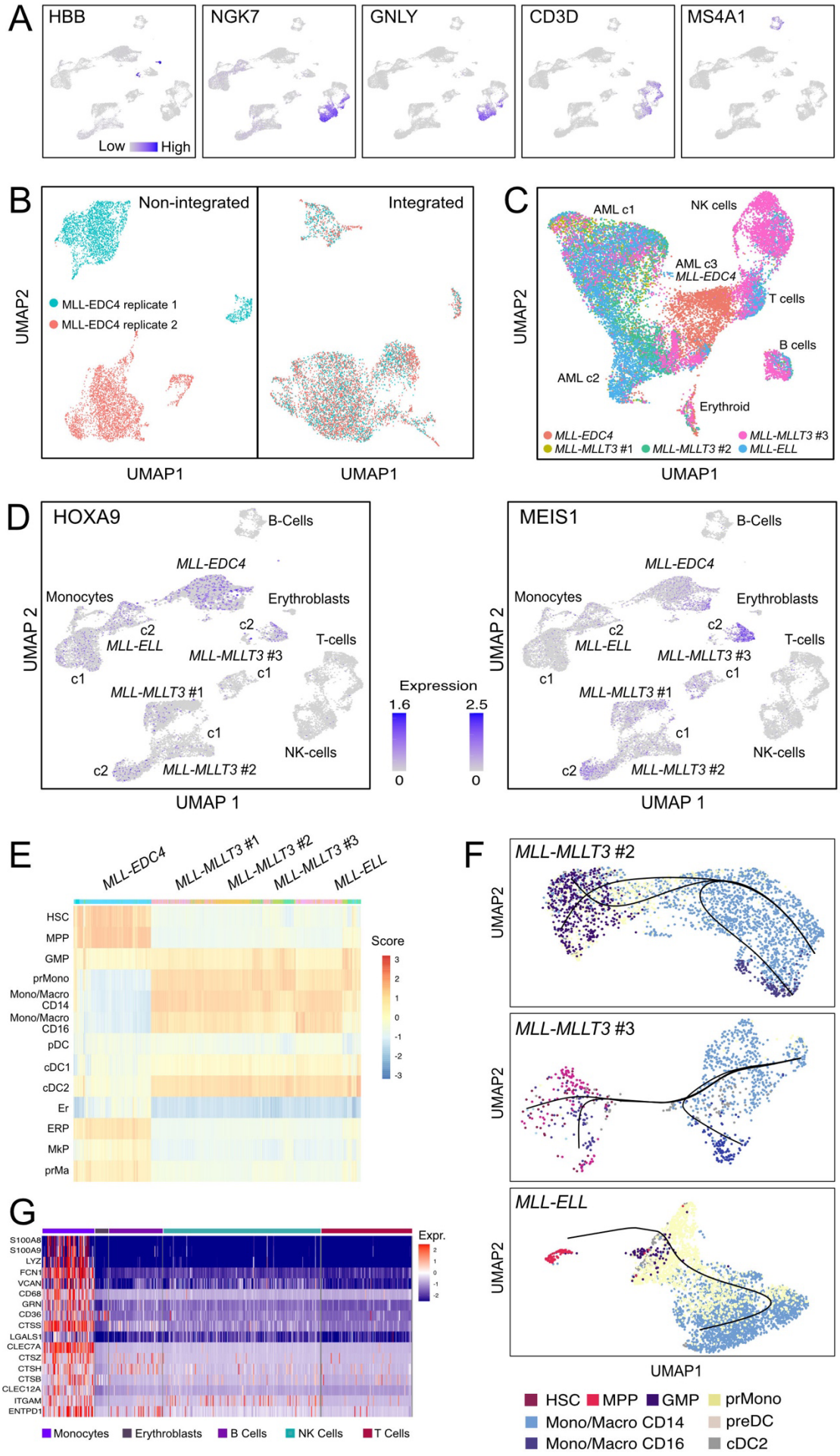

A
HBB
NGK7
GNLY
CD3D
MS4A1
Low
High
B
Non-integrated
Integrated
MLL-EDC4 replicate 1
MLL-EDC4 replicate 2
UMAP2
UMAP1
UMAP1
C
AML c1
NK cells
AML c3
MLL-EDC4
T cells
B cells
AML c2
Erythroid
MLL-EDC4
MLL-MLLT3 #3
MLL-MLLT3 #1
MLL-MLLT3 #2
MLL-ELL
UMAP2
UMAP1
D
HOXA9
MEIS1
B-Cells
MLL-EDC4
Monocytes
Erythroblasts
T-cells
NK-cells
c1
c2
MLL-ELL
MLL-MLLT3 #1
MLL-MLLT3 #2
MLL-MLLT3 #3
Expression
1.6
2.5
0
0
UMAP 2
UMAP 1
E
MLL-EDC4
MLL-MLLT3 #1
MLL-MLLT3 #2
MLL-MLLT3 #3
MLL-ELL
HSC
MPP
GMP
prMono
Mono/Macro CD14
Mono/Macro CD16
pDC
cDC1
cDC2
Er
ERP
MkP
prMa
Score
F
MLL-MLLT3 #2
MLL-MLLT3 #3
MLL-ELL
UMAP2
UMAP1
HSC
MPP
GMP
prMono
Mono/Macro CD14
preDC
Mono/Macro CD16
cDC2
G
S100A8
S100A9
LYZ
FCN1
VCAN
CD68
GRN
CD36
CTSS
LGALS1
CLEC7A
CTSZ
CTSH
CTSB
CLEC12A
ITGAM
ENTPD1
Expr.
Monocytes
Erythroblasts
B Cells
NK Cells
T Cells

Figure S1. Cell type annotation and transcriptional characterization of microenvironment. (**A**) UMAP embedding colored by expression of cell type specific markers for AML patients. (**B**) Comparison of *MLL-EDC4* replicate 1 from our previous scRNA-seq analysis[13] with the new data acquired here (replicate 2). Left, non-integrated data. The raw expression profiles of the replicates are separated due to confounding technical factors. Right, integrated data. Highly similar distributions of the gene expression profiles are obtained, which demonstrates the reliability of our scRNA-seq protocol. (**C**) Integrated scRNA-seq data set of all five AML patients. Blasts from the *MLL-EDC4* sample still form a distinct cluster (AML c3) that is separated from the other samples. (**D**) UMAP embedding colored by expression of HOXA9 and MEIS1 for the different samples. Both markers show their strongest simultaneous expression in *MLL-EDC4* and in *MLL-MLLT3* #3 c2, which supports the presence of HSC/MPP like cells in the corresponding clusters[14]. (**E**) Heatmap of label scores from cell type prediction for clustered AML cells. MPP, multipotent progenitor; GMP, granulocyte-monocyte progenitor; preDC, pre-dendritic cell; cDC2, type 2 conventional dendritic cell; prMono, promonocyte; Mono CD14, $CD14^+$ monocytes; Mono CD16, $CD16^+$ monocytes. (**F**) UMAP embeddings of *MLL-MLLT3* #2 and #3 and *MLL-ELL* leukemic cells colored according to predicted cell types. Black lines indicate pseudotime trajectories. (**G**) Clustered single-cell transcriptomic heatmap of differentially expressed genes in *MLL-EDC4* monocytes vs. other non-malignant cells in the microenvironment. The monocyte-like cells from *MLL-EDC4* were unusual with respect to expression of CD36, cathepsins and CLEC receptor.

Supplemental Table S1. Patient information and scRNA-seq data

Patient/MLL fusion	***MLL-EDC4***	***MLL-MLLT3 #1***	***MLL-MLLT3 #2***	***MLL-MLLT3 #3***	***MLL-ELL***
Sex / age	F / 56 years	F / 28 years	F / 58 years	F / 64 years	F / 57 years
AML type	Secondary AML evolving from MDS, no prior CTx	t-AML after RCTx for Hodgkin lymphoma, no MDS phase	De-novo AML, no MDS phase	tAML after RCTx for breast and CTx for ovarian cancer, no MDS phase	t-AML after RCTx for ovarian cancer, no MDS phase
Prior AML treatment	Decitabine	None	None	None	None
Timepoint of cell sampling	Before 4th cycle of decitabine	Diagnosis	Diagnosis	Diagnosis	Diagnosis
Karyotype	46,XX,t(11;16)(q23;q22)[12]; 46, XX[8]	46,XX,t(9;11)(p22;q23)[10]	46,XX,t(9;11)(p22;q23)	46,XX,t(9;11)(p22; q23)[6]/ 47,XX, idem, +21[14]	46,XX,t(11;19)(q23;p13)
Cell source	PBMCs, $CD3^+$ cell depleted	BMNCs	BMNCs	BMNCs	PBMC
% blasts	90	78	95	59	57
% nuclei *MLL* fusion (FISH)	85	90	95	90	82
Clinical course	Diagnosis 11/ 2012, start decitabine 05/ 2013, SD until progression and exitus letalis 02/2014	Induction chemotherapy 01/2016, CR 02/2016, HSCT 04/2016, alive in remission (04/2023)	Induction chemotherapy 01/2015, CR 03/2015, HSCT 03/ 2015, alive in remission (05/2023)	Induction chemotherapy 06/2014, CR 09/2014, relapse 02/ 2015, last seen alive 02/ 2015	Induction chemotherapy 11/2012, CR 11/2012, HSCT 02/ 2013, last seen alive 08/ 2022
scRNA-seq cell number	3255	1766	3464	3917	5154
scRNA-seq genes/cell (median)	951	624	856	1181	1225

Demographic, clinical, molecular, and diagnostic information for all samples in this study. CTx, chemotherapy; RCTx, radio-chemotherapy; HSCT, hematopoietic stem cell transplantation; MDS, myelodysplastic syndrome; t-AML, therapy-related AML. PBMCs, peripheral blood mononuclear cells; BMNCs, bone marrow mononuclear cells; CR, complete remission; SD, stable disease. Clinical presentation of *MLL-EDC4*: Initial disease course marked by typical MDS features without blast expansion, which is absent in *MLL-r*. This was followed by grade 2 bone marrow fibrosis, and erythrocyte transfusion dependence for 8 months, until progression to secondary AML occurred. Notably, the AML course was overall indolent (“smoldering”), thus necessitating only limited and intermittent low-dose treatment measures (decitabine, hydroxyurea), over a period of 10 months.

Supplemental Table S2. Single cell derived gene signatures

HSC	MPP	GMP	Mono CD14	Pre DC	cDC1
AVP	SPINK2	MPO	S100A12	STMN1	HLA-DPA1
RPS4X	IGLL1	PRTN3	S100A9	TUBA1B	HLA-DPB1
RPL5	NPM1	ELANE	S100A8	IGLL1	HLA-DRB1
EIF3E	TUBA1B	AZU1	CXCL8	TUBB	CD74
SPINK2	HSP90AB1	CTSG	TYROBP	SOX4	HLA-DQA1
HSP90AB1	LDHB	PRSS57	VCAN	PLAC8	HLA-DQB1
NPM1	EEF1B2	CLEC11A	DUSP1	IRF8	HLA-DRA
RPL31	STMN1	H2AFZ	FOS	C12orf75	HLA-DRB5
RPS6	RPLP0	CALR	FCN1	HMGB1	CPVL
EEF1B2	RPL5	PLAC8	S100A6	ACTG1	CST3
RPLP0	HNRNPA1	STMN1	NEAT1	HIST1H4C	SNX3
RPL3	PRSS57	IGLL1	CD14	ITM2C	C1orf54
RPS3	RPSA	AC020656.1	CTSS	HMGN2	IRF8
HINT1	TUBB	RNASE2	FTL	SRP14	ID2
FAM30A	HINT1	RPS18	NFKBIA	UBB	RGS10
EEF2	RPS4X	NUCB2	NAMPT	HMGN1	TMSB4X
RPL10A	NUCB2	HSP90B1	ZFP36L1	HMGB2	HLA-DMA
ZFAS1	HIST1H4C	RPS19	S100A4	HNRNPA1	DNASE1L3
RPS18	RPS5	DUT	SLC2A3	H2AFZ	S100A10
PRDX1	RPL7A	AREG	G0S2	SPINK2	HLA-DQA2
RPSA	RPL3	FABP5	RGS2	SCT	TUBA1B
RPS5	RPS3	RPLP1	CSTA	PCLAF	CLEC9A
RPL7A	ENO1	LDHB		CD74	CPNE3
NOP53	GYPC	NPM1		PLD4	ACTB
RACK1	RPS6	HSPB1		NUCB2	ACTG1
RPL15	C1QTNF4	RPL35		HSP90AA1	ARPC2
RPS2	PRDX1	RPS21		HNRNPA2B1	LMNA
ANKRD28	RPL4	RPL36		TCF4	LSP1
RPS8	EIF3E	RPS23		SLC25A5	C1orf162
C6orf48	SMIM24	HMGN1		LDHB	TXN
HOPX	HMGB1	CST7		RPSA	PPA1
LDHB	RPS18	MS4A3		SEC61B	GSTP1
SNHG8	RPL10A			CCDC50	TAGLN2
CD164	HSP90AA1			NPM1	PSMB9
RPS23	DUT			PLP2	HLA-C
RPL4	HMGA1			PPIB	PPT1
RPS12	UBB			NUCKS1	EEF1B2
HNRNPA1	EEF2			SEPT6	HLA-DMB
RPL30	BTF3			PCNA	NAP1L1

Gene signatures of healthy bone marrow donors from HCA were generated by differential gene expression analysis. These signatures were used to calculate module scores for feature expression programs in single cells.

Supplemental Table S3. Flow Cytometry

Patient/MLL fusion	***MLL-EDC4***	***MLL-MLLT3 #1***	***MLL-MLLT3 #2***	***MLL-MLLT3 #3***	***MLL-ELL***
Peripheral blood flow cytometry	4% CD117+ cells; partial co-expression of myeloperoxidase (20%); negative for CD14, CD3, CD19, CD34.	2% CD19+/ CD20+ cells; negative for CD5, CD10, CD200, CD34, CD117; CD14 not tested.	~85% CD33+ cells; negative for CD34, CD117; CD14 not tested.	37% CD34+ cells; partial co-expression of CD13, CD117 (60%) and CD33 (33%); negative for for CD3, CD7, CD19, CD20, CD79a, MPO; CD14 not tested.	10% CD34+/ CD117+ cells; 35% CD14+ with negativity for CD34/ CD117.
Bone marrow flow cytometry	25% CD117+ with coexpression of CD13 i.c., CD33 i.c., CD33s (65%), CD34 (65%), CD45 (50%), CD13s (55%), CD64 (50%), CXCR4 (45%) and MPO (20%). Negative for CD7, CD14, CD15, CD56, CD79a, CD3 and CD19.	65% CD14+ cells, positive for CD4, CD11b, CD11c, CD13, CD15, CD33, CD38, CD45, CD64, lysozyme and HLA- DR; partially positive for CD86 (75%); negative for CD2, CD3 i.c., CD5, CD7, CD19, CD34, CD56, CD65, CD79a, CD117 and MPO	85% CD33+ cells displayed in monocyte gate, positive for CD4, CD11c, CD15, CD38, CD45 lo, CD64 and HLA- DR; partially positive for lysozyme (85%), CD11b (85%), CD86 (80%), CD13 i.c. (70%), CD13 (35%), CD14 (30%), CD56 (25%), CD65 (25%) and MPO (10%). Negative for CD2, CD3 i.c., CD5, CD7, CD19, CD34, CD79a and CD117	40% CD117+ cells, positive for CD13, CD45, partially positive for CD38, CD34, CD86, CD33, HLA-DR and MPO. Negative for CD1a, CD2, CD3, CD5, CD7, CD10, CD11c, CD14, CD15, CD19, CD20, CD56, CD65, CD79a, CD200, Mo7.1, TdT and Lysozyme	95% CD33+ cells with projection in blast cell gate, negative for CD2, CD3, CD5, CD7, CD13, CD14, CD19, CD20, CD34, CD56, CD65, CD79a, CD117 and TdT
Blood cell differential	>80% blasts	5% blasts, 75% pro-monocytes/ monocytes	92% monoblasts, 8% erythro-blasts	54% blasts, 5% mono-cytes	57% mono-blasts/pro-monocytes, 27% mono-cytes

Supplemental references

1. Stosch JM, Heumuller A, Niemoller C, et al. Gene mutations and clonal architecture in myelodysplastic syndromes and changes upon progression to acute myeloid leukaemia and under treatment. Br J Haematol. 2018;182(6):830-842.

2. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33(5):495-502.

3. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 2019;8(4):281-291 e289.

4. Regev A, Teichmann SA, Lander ES, et al. The Human Cell Atlas. Elife. 2017;6.

5. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296.

6. Gao R, Bai S, Henderson YC, et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat Biotechnol. 2021;39(5):599-608.

7. Federico A, Monti S. hypeR: an R package for geneset enrichment workflows. Bioinformatics. 2020;36(4):1307-1308.

8. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163-172.

9. Street K, Risso D, Fletcher RB, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018;19(1):477.

10. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019;29(8):1363-1375.

11. Alvarez MJ, Shen Y, Giorgi FM, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet. 2016;48(8):838-847.

12. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-2504.

13. Becker H, Greve G, Kataoka K, et al. Identification of enhancer of mRNA decapping 4 as a novel fusion partner of MLL in acute myeloid leukemia. Blood Adv. 2019;3(5):761-765.

14. Collins CT, Hess JL. Deregulation of the HOXA9/MEIS1 axis in acute leukemia. Curr Opin Hematol. 2016;23(4):354-361.