Cell Stem Cell, Volume *30*

Supplemental Information

Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia

Sergi Beneyto-Calabuig, Anne Kathrin Merbach, Jonas-Alexander Kniffka, Magdalena Antes, Chelsea Szu-Tu, Christian Rohde, Alexander Waclawiczek, Patrick Stelmach, Sarah Gräßle, Philip Pervan, Maike Janssen, Jonathan J.M. Landry, Vladimir Benes, Anna Jauch, Michaela Brough, Marcus Bauer, Birgit Besenbeck, Julia Felden, Sebastian Bäumer, Michael Hundemer, Tim Sauer, Caroline Pabst, Claudia Wickenhauser, Linus Angenendt, Christoph Schliemann, Andreas Trumpp, Simon Haas, Michael Scherer, Simon Raffel, Carsten Müller-Tidow, and Lars Velten

Supplementary figures and legends

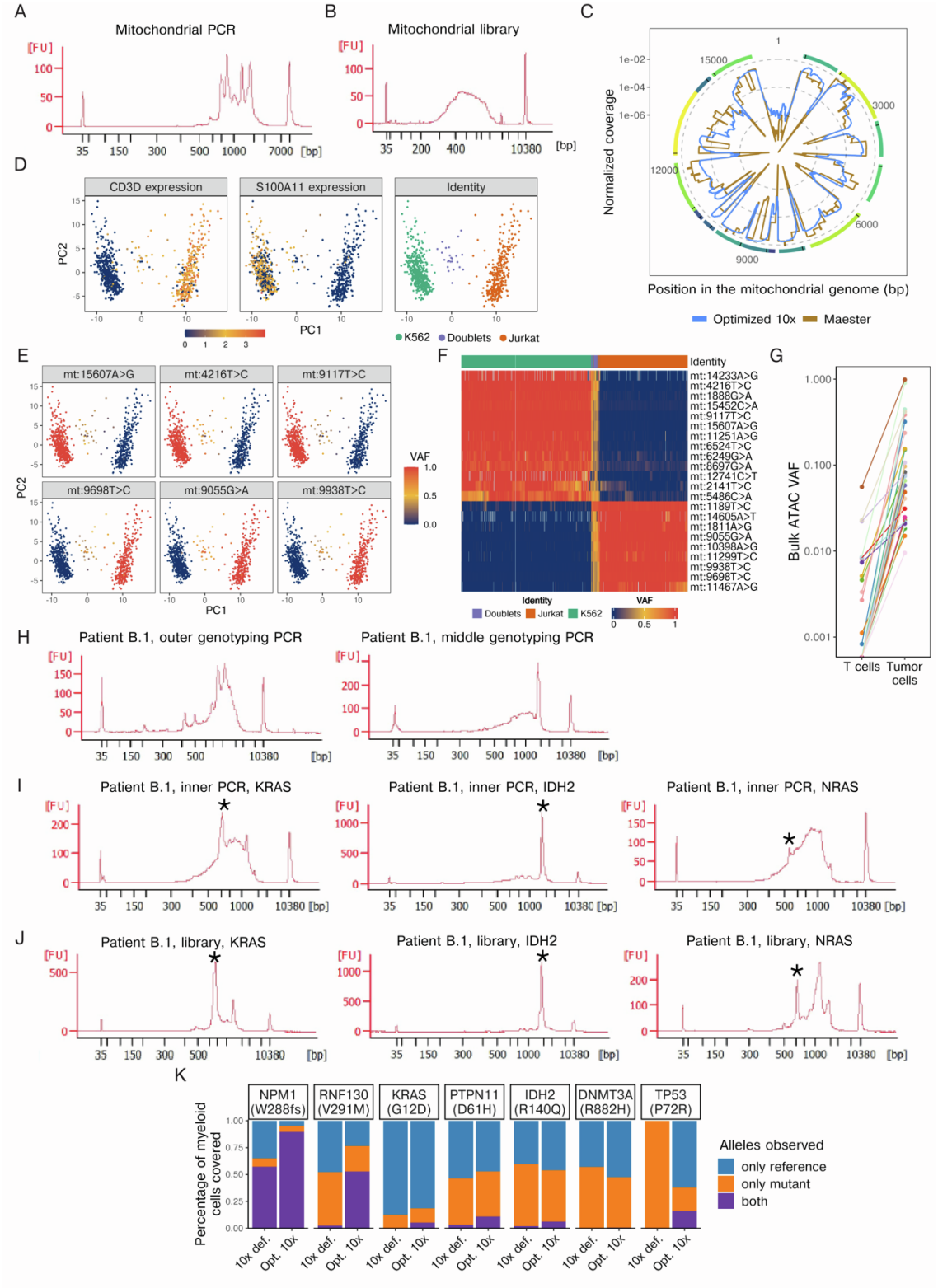

Figure S1. Technical evaluation of "Optimized 10x" libraries, related to Figure 1 and STAR Methods. **a.** Bioanalyzer traces of a mitochondrial library following PCR. One representative library is shown. **b.** Final mitochondrial library following tagmentation and library PCR. One representative library is shown. **c.** Normalized coverage across the mitochondrial genome obtained by Optimized 10x and data from MAESTER[14]. **d.** Cell line mixing experiment of Jurkat and K562 cells. Principal component analysis of gene expression. Score plots highlight the expression of a T cell gene (*CD3D*), a myeloid gene (*S100A11*) and assignment of cells as Jurkat, K562 or Doublets. Take note that doublets were removed from the main AML datasets using the scrublet algorithm prior to analysis. **e.** Score plots highlighting the variant allele frequencies of six mitochondrial variants that differ between Jurkat and K562. **f.** Heatmap depicting all mitochondrial mutations identified between these cell lines. **g.** mtSNVs are absent or present at negligible allele frequencies in T cells. mtSNV variant allele frequencies from bulk ATAC are shown for CD3+ and CD3- FACS sorted populations termed T cells and Tumor cells, respectively. **h.** Bioanalyzer traces of Nuclear mutation-targeting outer and middle PCRs, patient B.1. **i.** Inner PCRs for KRAS, IDH2 and NRAS mutations in patient B.1. Asterisks indicate the expected product size. **j.** Final libraries for KRAS, IDH2 and NRAS mutations in patient B.1. Asterisks indicate the expected product size. **k.** For all cells covered with default 10x or Optimized 10x, barcharts depict the fraction of cells where only the mutant allele, only the reference allele, or both alleles are captured. See also Figure 1c. Only immature and early myeloid cells are included.

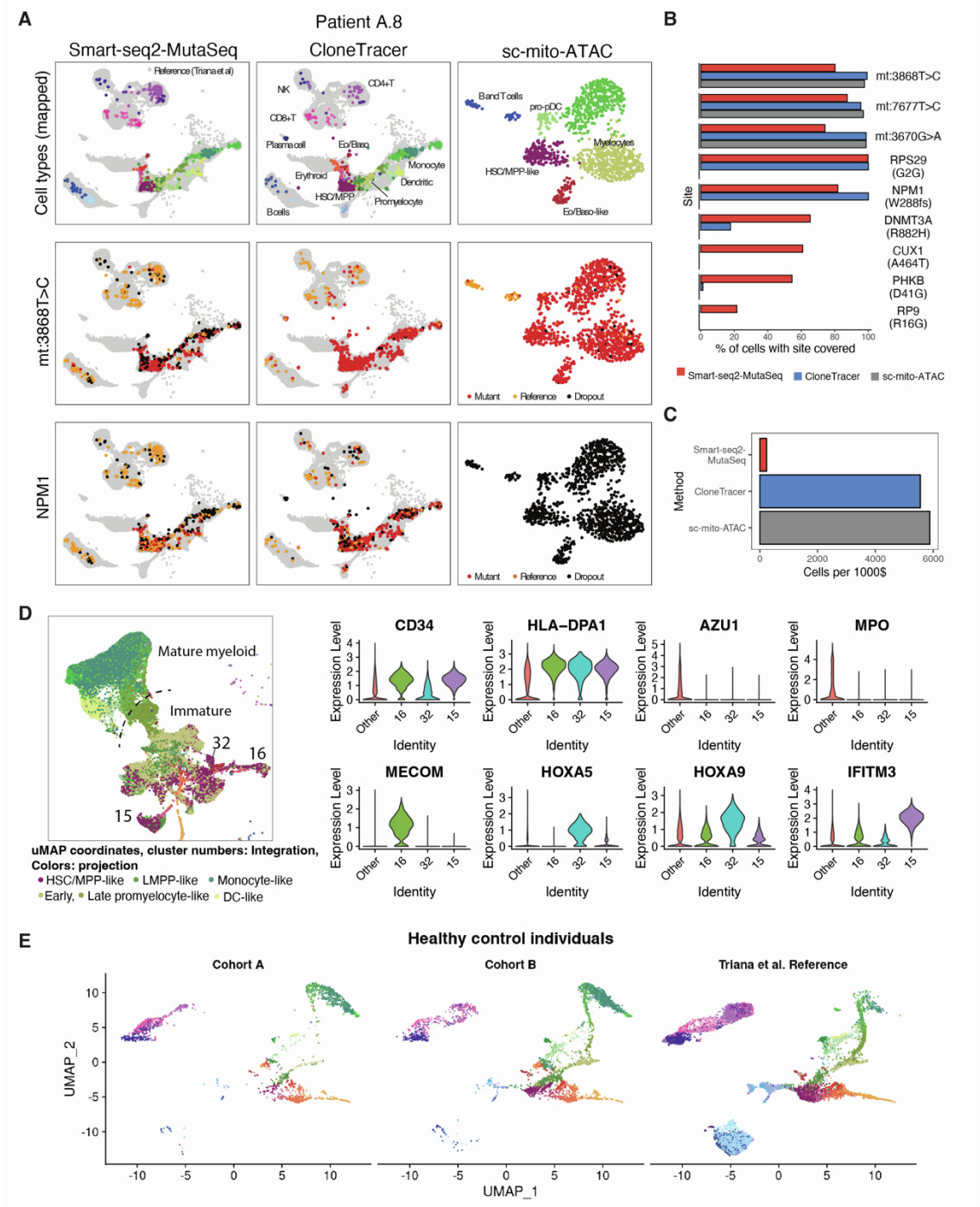

Figure S2. Comparison of protocols and evaluation of data integration strategies, related to Figure 1 and STAR Methods. **a.** uMAPs illustrating the application of three single-cell genomic methods for clonal tracking to biological material from the same AML patient (A.8). Columns: Cells were profiled using MutaSeq (well-based RNA-seq), "Optimized 10x" (droplet-based RNA-seq) or sc-mito-ATAC seq[22] (droplet-based ATAC-seq). Top row: Cells were mapped to a reference atlas using state of the art mapping algorithms[52], see also methods. Middle row: Mutational status for a mitochondrial mutation. Bottom row: Mutational status for a mutation in the gene NPM1. **b.** Coverage of nuclear and mitochondrial mutations identified in A.8. Only myeloid cells are included. **c.** Bar chart comparing the throughput of the three

methods. **d.** Scanorama integration maintains biological differences between cells. Left panel: uMAP reproduced from Figure 3g, highlighting cluster 15, 16 and 32 that all project to HSCs/MPPs. Right panels: Violin plots depicting the expression of several genes that are usually not expressed in HSCs/MPPs (AZU1, MPO, IFITM3), or co-expressed in HSCs/MPPs (all other genes). **e.** Scanorama integration removes technical differences between cells. Two healthy bone marrow donors processed as part of Cohort A and B map to the same cell state space as the reference individual. To create this plot, the same data integration algorithm used to generate the main figures was used, with identical settings, on these three individuals. For color code, see Figure S3a.

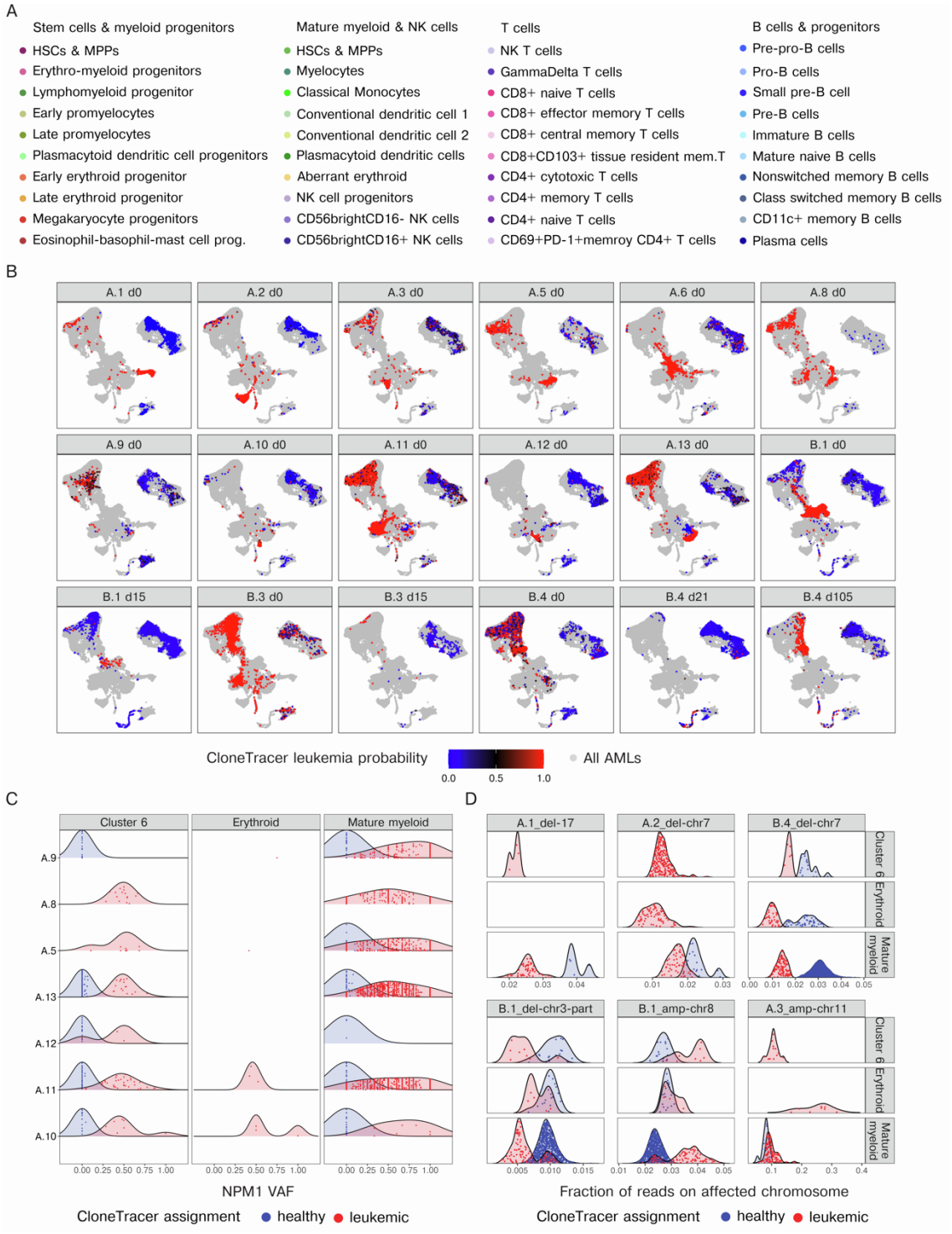

Figure S3. CloneTracer applied to two AML patient cohorts, related to Figure 3. **a.** Color legend for main Figure 3a. **b.** CloneTracer assignments stratified by patient and time point, color-coded by leukemia probability. **c.** Observed variant allele frequency of NPM1 mutations in patients carrying this mutation. Blue dots: Cells classified as healthy by CloneTracer. Red dots: Cells classified as leukemic. **d.** Coverage of the chromosome affected by copy number losses (del) and gains (amp).

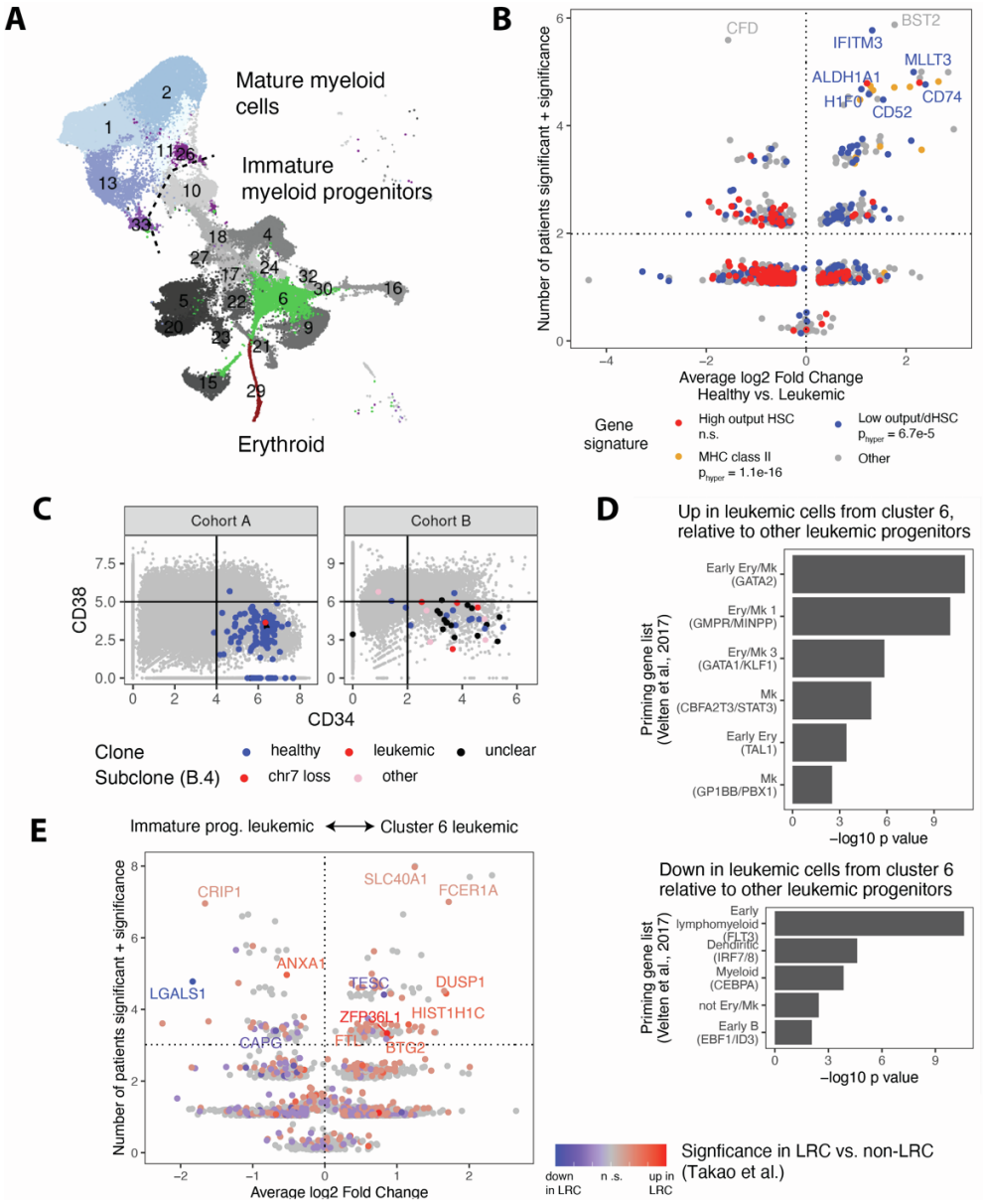

Figure S4. Analyses of healthy and leukemic stem cells, related to Figure 3 and 4. **a.** uMAP highlighting cluster identity. Clusters highlighted in shades of grey are referred to as "immature myeloid progenitors" in the main text, clusters highlighted in shades of blue as "mature myeloid cells". Cells of the stem cell cluster 6 are shown in green. **b.** Volcano plot as in main Figure 3f, comparing healthy and leukemic cells from C6. n=6 patients were analyzed. Genes are colored by gene signatures for high output and low output HSCs[36], identified by clonal tracking, and dormant HSCs identified by long-term label retention assays[33]. **c.** Scatter plots depicting the surface marker expression of CD34 and CD38, highlighting the cells that were identified as putative dormant stem cells in panel b. **d.** Bar charts depicting the enrichment of priming gene signatures among genes significantly up- or down-regulated in at least three patients from Figure 5a. **e.** Volcano plot as in main Figure 5a, comparing leukemic cells from C6 to other leukemic immature myeloid cells from the same patient. n=14 patients with confident CloneTracer leukemia assignments were analyzed. Genes are colored using information on their expression in label retaining vs. non label retaining AML cells[31].

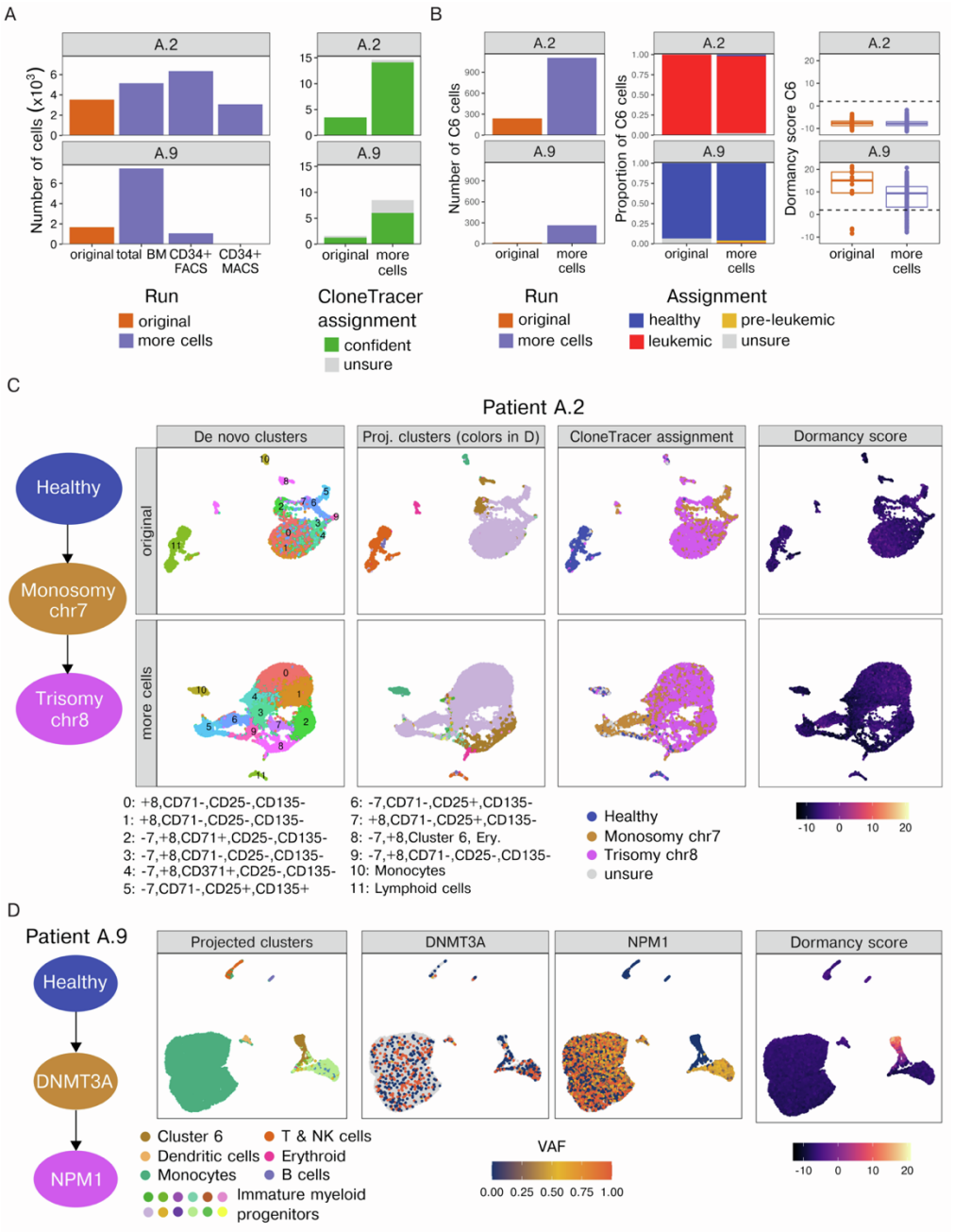

Figure S5. Inclusion of additional 23,110 cells from patients A.2 and A.9 confirmed original analyses and increased cell numbers of cluster 6, related to Figure 4. **a.** Number of cells with confident CloneTracer clonal assignment in the original and additional runs. **b.** Number of cells in C6, their clonal identity and dormancy status in the two patients. **c.** Unsupervised analysis of patient A.2, comparing original data (top row) and additional data (bottom row). uMAP coordinates were computed separately for the two data sets. From left to right: (i) Bottom: new clustering performed on the additional data. Top: new cluster labels transferred to the original data, highlighting high correspondence of clusters. (ii) Original cluster labels highlighted on the uMAPs (iii) Clonal assignments (iv) Dormancy score (see Methods). **d.** Unsupervised analysis of additional cells sequenced for patient A.9. Left panel highlights clusters obtained from original data and the same clusters mapped to the new data. Right panels highlight genotype and dormancy score.

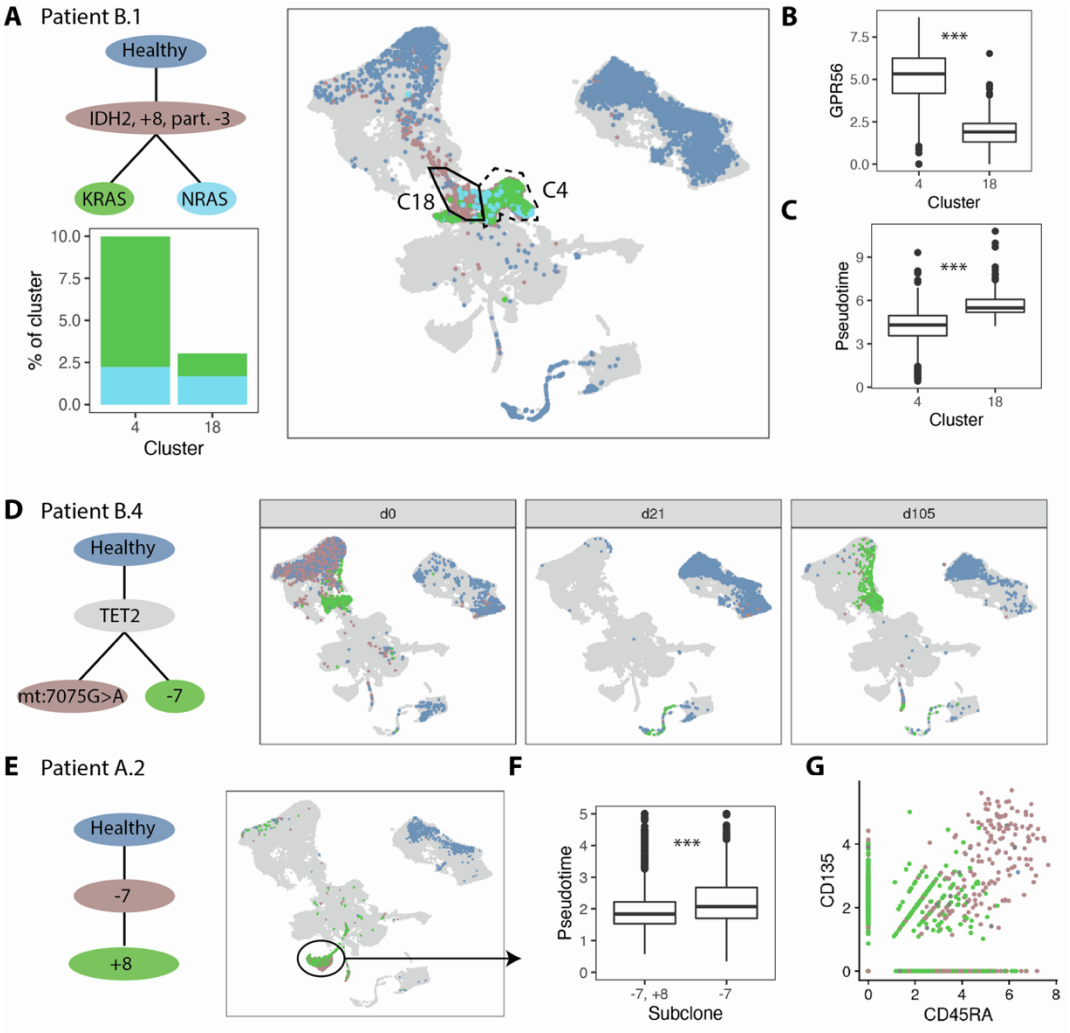

Figure S6. Analyses of sub-clones with CloneTracer, related to Figure 5. **a-c:** In patient B.1, we observed three clones: A trisomy 8/IDH2/partial chr3 deletion mutant clone, as well as a KRAS and an NRAS mutant sub-clones. While all clones existed both in a GPR56-high, more immature state (cluster C4), and a GPR56-low, late promyelocyte-like state (cluster C18), the KRAS clone was predominantly observed in the more immature state (C4). Top left: Clonal tree for patient B.1. Right: uMAP highlighting clonal identities at the time of diagnosis. Brown dots also include cells with dropout of KRAS/NRAS. Two clusters (C4 and C18) are highlighted. Bottom left: Representation of the KRAS and NRAS clones in clusters C4 and C18. **b.** GPR56 surface expression in clusters C4 and C18. **c.** Pseudotime on clusters 4 and 18. **d:** In patient B.4, we observed two sub-clones, one of which was marked by monosomy 7, and the other one marked by a mitochondrial mutation. According to bulk exome data, both clones carried a TET2 mutation. While the mitochondrial clone abundantly generated monocytes, the monosomy 7 clone existed in a promyelocyte-like state and relapsed after 105 days while maintaining its cell state. Clonal tree (left) and uMAP highlighting clonal identities (right) for patient B.4. at the time of diagnosis (d0), post induction therapy (d21) and relapse (d105). **e-g:** In patient A.2, a monosomy 7 clone co-existed with a -7/+8 sub-clone in a patient-specific aberrant cell state. Within this state, the sub-clone shifted to a more immature identity and altered its cell surface phenotype **e.** Clonal tree (left) and uMAP highlighting clonal identities (right) for patient A.2 at the time of diagnosis. **f.** Boxplot of pseudotime within cluster 15 (highlighted), stratified by clone. **g.** Surface expression of CD45RA and CD135 for cells from cluster 15, color coded by clone.

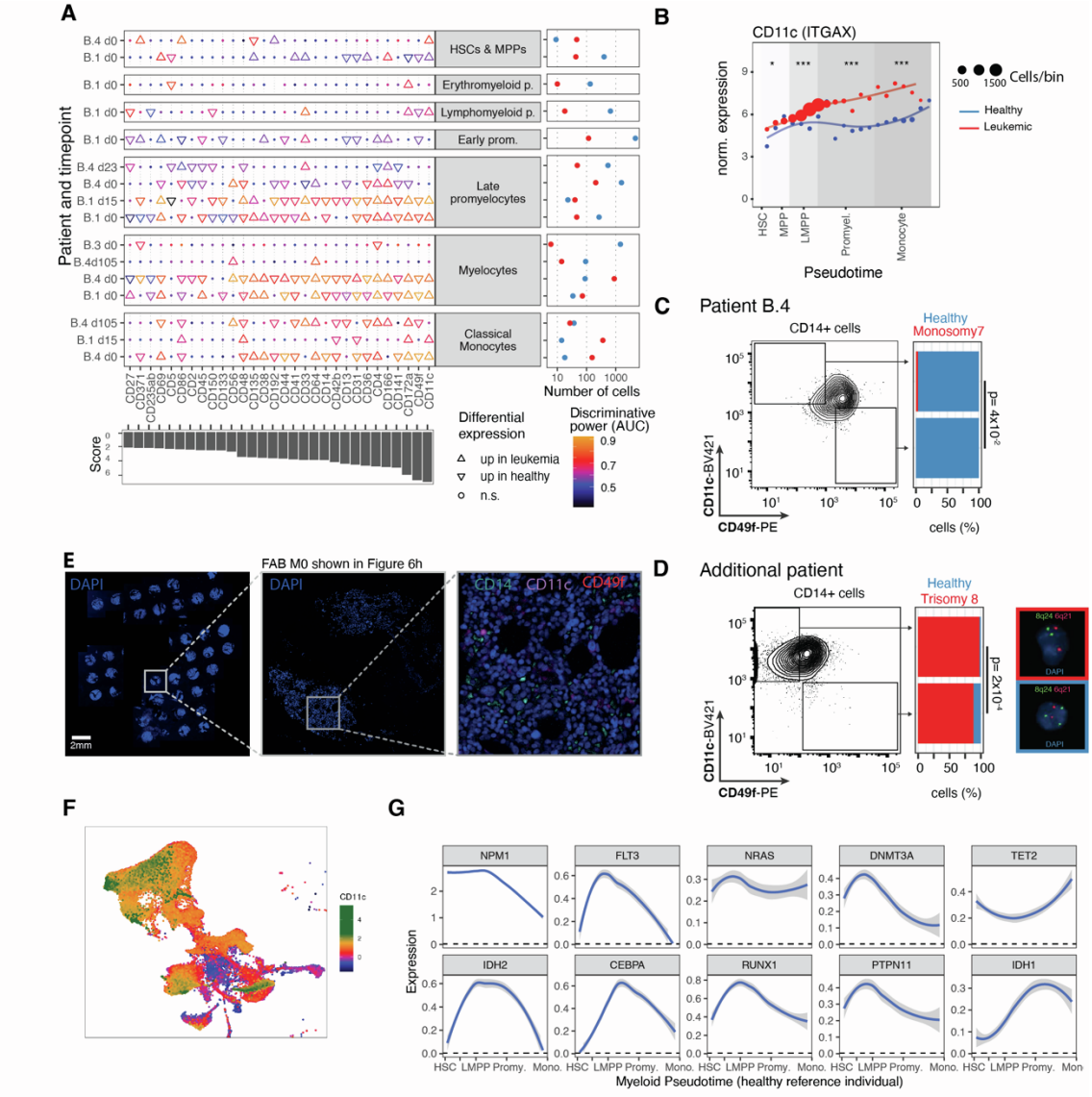

Figure S7. Discovery and validation of leukemia markers, related to Figure 6. **a.** Dot plot indicating intra-sample differential expression testing between healthy and leukemic cells. Upward triangles indicate surface markers with significant overexpression in leukemic cells, downward triangles indicate surface markers with significant overexpression in healthy cells. Color indicates discriminative power (area under the receiver operating characteristics curve). Right panel: Scatter plot indicating the number of healthy (blue) and leukemic (red) cells underlying these comparisons in each sample. Bottom panel: score evaluating the different markers, defined as the sum of AUC times the sign of the log fold change, if significant. **b.** Smoothened expression of CD11c over pseudotime of patient B.1 stratified by clone. Asterisks indicate significance of differential expression within the shaded area of pseudotime. ***: FDR < 0.001, **: FDR < 0.01, *: FDR < 0.1. *p* values are from a Wilcoxon test of library-size normalized ADT counts. Points indicate mean expression within 20 equally sized bins along pseudotime, point size indicates number of cells per bin. **c.** Quantification of healthy and leukemic cells in CD14+ cell fractions of patient B.4 by FACS and FISH. See main Figure 6c for a detailed legend. **d.** Like c, but for an additional patient carrying a trisomy 8. **e.** Representative immunofluorescence images of a tissue microarray (left)

and a single punch biopsy of a FAB M0 classified AML patient (middle, right) used for validating CD11c and CD49f as leukemia/healthy markers (see Figure 6f-h). Nuclei were counterstained with DAPI. Scale bar 2mm. **f.** Expression of CD11c on the uMAP of myeloid cells. **g.** Expression of commonly mutated leukemia driver genes[44] smoothened over myeloid differentiation pseudotime of the healthy reference individual[4].