Methods S1

Table of Contents

Statistical modelling of clonal hierarchies and clonal assignments

The goal of our model is to (1) compute the evidence for any biologically reasonable clonal hierarchy, (2) identify the best clonal hierarchy and (3) compute the posterior probability (confidence) of the clonal assignment, see Figures 1d.

As input data to the model, we consider $M \in \mathbb{N}_0^{n \times m}$, a matrix of UMI counts supporting the mutant allele across n single cells and m mutations of interest, and $N \in \mathbb{N}_0^{n \times m}$, a matrix of total UMI counts falling on that site (reference or mutant). In the case of CNVs, M contains the number of UMIs falling in the chromosomal region of interest and N the total number of UMIs observed in the cell. A vector $t \in \{SNV, mtSNV, CNV\}^m$ specifies for each mutation whether it is a nuclear SNV, mitochondrial SNV or CNV. A vector u specifies for each cell the cell type identified by default single cell transcriptomic analysis. A vector v specifies for each cell the sample (e.g. timepoint) it is from when present.

As further input data, we consider two integer-valued matrices $a, b \in \mathbb{N}_0^{m \times s}$ that contain information from bulk experiments across s samples (e.g. timepoints), such as read counts supporting the mutant allele and total read counts from bulk exome or panel sequencing. When no such information is available, a and b are set to 0. Finally, $r \in \mathbb{R}_+^m$ specifies the known ratio in chromosome counts between affected and non-affected cells for CNVs (e.g. 1.5 for trisomy, 0.5 for monosomy, etc).

We assume that in reality, cells belong to a number $l \leq m + 1$ of clones that are represented by a binary matrix A of dimensions $m \times l$. For example, a clonal hierarchy where a healthy cell gives rise to a founding clone that in turn gives rise to two independent sub-clones would be represented by

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad (1)$$

A vector $\pi \in \{1, \ldots, l\}^n$ maps every cell to a clone, i.e. a column in A. A list of simplices ψ_s specifies the marginal probabilities of each clone in each sample (e.g. timepoint), i.e. the fraction of each clone in the population of cells.

Assuming that A and π are known, data is generated through the following process, parametrized by a set of further parameters (FPR, c, h, g) defined below. For a graphical representation of the model, see Figure N1a.

If mutation j is a nuclear or mitochondrial SNV, we assume that mutant UMIs $M_{\cdot j}$ are either false positives created from a background poisson process, or true positives sampled from the total number of UMIs $N_{\cdot j}$ using a Beta-Binomial distribution. This modelling choice is justified in Figure N1b-c and full posterior predictive checks for all patients are provided in Figure N1d-f. The use of a Binomial sampling process is inappropriate, as both mitochondrial and, more so, nuclear counts are overdispersed (Figure N1b), possibly as a consequence of allele-specific gene expression. Thus,

$$p(M_{ij}|A_{j,\pi(i)}, h_{t(j)}, FPR_{t(j)}, c_j) = \begin{cases} pPoisson(M_{ij}|FPR_{t(j)}) \text{ if } A_{j,\pi(i)} = 0 \\ pBetaBinom(M_{ij}|N_{ij}, h_{t(j)}c_j, (1-h_{t(j)})c_j) \text{ else} \end{cases} \quad (2)$$

Here, FPR is a false positive rate defined independently for each class of mutation, h is the heteroplasmy of the mutation and c is a concentration parameter specifying the degree of over-dispersion. For nuclear SNVs, we assume that within one cell 50% of DNA molecules are mutated. Due to this assumption, we set strongly informative prior on h (e.g. a Beta distribution with shapes 1000,1000), thereby avoiding non-identifiabilities in the case of nuclear SNVs with low coverage. Different priors can be used e.g. for loss-of-heterozygosity SNVs, or mutations that strongly affect RNA stability. The concentration parameter is mutation-specific as we observed that mutations with lower coverage (e.g. *DNMT3A*) tend to be more overdispersed than SNVs falling in highly expressed genes such as *NPM1*. By contrast, for mitochondrial SNVs, the model allows to infer the heteroplasmy individually for each mutation (i.e. fraction of mutated mitochondrial transcripts in a cell carrying the mutation) from the data, and a flat prior is used while the concentration parameter is shared across mtSNVs. For FPR and c, weakly informative priors are used, see source code for a full specification of priors.

If j is a CNV, we assume that the UMIs on the affected chromosomal region are drawn from the total number of UMIs in the cell using beta-binomial sampling.

$$p(M_{ij}|A_{j,\pi(i)}, h_j, g_j, c_{t(j)}) = \begin{cases} pBetaBin(M_{ij}|N_{ij}, g_j c_{t(j)}, (1-g_j)c_{t(j)}) \text{ if } A_{j,\pi(i)} = 1 \\ pBetaBin(M_{ij}|N_{ij}, h_j c_{t(j)}, (1-h_j)c_{t(j)}) \text{if } A_{j,\pi(i)} = 0 \end{cases} \quad (3)$$

Importantly, the observed CNV ratio constrains the ratio between on-target reads in cells carrying the CNV vs. cells not carrying the CNV.

$$p(r_j|g_j, h_j) = pNorm(r_j|g_j/h_j, 0.05) \quad (4)$$

We observed that the fraction of reads on the affected chromosome depends on the cell type, e.g. the fraction of RNA molecules from chromosome 7 is much decreased in plasma cells compared to healthy cells, also in individuals not affected by the copy number variant (Figure N2a). We therefore additionally formulated a model accounting for this cell type covariate by replacing g and h by cell type specific parameters

$$p(M_{ij}|A_{j,\pi(i)}, h_{j,u(i)}, g_{j,u(i)}, c_{t(j)}) = \begin{cases} pBetaBin(M_{ij}|N_{ij}, g_{j,u(i)}c_{t(j)}, (1-g_{j,u(i)})c_{t(j)}) \text{ if } A_{j,\pi(i)} = 1 \\ pBetaBin(M_{ij}|N_{ij}, h_{j,u(i)}c_{t(j)}, (1-h_{j,u(i)})c_{t(j)}) \text{if } A_{j,\pi(i)} = 0 \end{cases} \quad (5)$$

For this more complex model, we inferred a cell type specific prior on the fraction of UMI on the affected chromosome from patients not affected by the CNV such that

$$logit(h_{j,u(i)}) \sim Norm(x_{u(i)}, \sigma) \quad (6)$$

Detail on the estimation of the prior (cell type specific average fraction of UMI on the affected chromosome $x_{u(i)}$ and its variation across patients σ) is provided in Figure N2b,c. Posterior predictive checks confirm that this model describes the data better, compared to the simple, cell-type invariant model (Figure N2d). Application of the cell type specific model was only necessary in samples B.2 and B.4.

Finally, to include data from bulk measurements, allele frequencies are computed from clonal frequencies as follows:

$$AF_j = 0.5 * \sum_{l \leq m} A_{jl}\psi_l \quad (7)$$

Then,

$$p(a_j|AF_j) = Binom(a_j|b_j, AF_j) \quad (8)$$

Together, equations 2 and 3 specify the likelihood of the single cell data, given the structure of the tree, clonal assignments, and all remaining parameters (Called θ in the following):

$$\mathcal{L}(M|A,\pi,\theta) = \prod_i \prod_j p(M_{ij}|N_{ij}, A_{j,\pi(i)}, \theta) \quad (9)$$

Equations 4 and 8 define the likelihood of additional observations available to the model, such that the final likelihood becomes:

$$\mathcal{L}(M,a,r|A,\pi,\psi,\theta) = \mathcal{L}(M|A,\pi,\theta) * \prod_j p(a_j|A,\psi) * \prod_{j\in CNV} p(r_j|\theta) \quad (10)$$

Where CNV is the set of all CNV mutations.

We implemented the model in pyro[57], a probabilistic programming language that efficiently fits variational posterior distributions using stochastic variational inference with a highly parallel, GPU-based backend based on pytorch. For a given tree A, we sample node attachment from the clonal fractions:

$$\pi_i \sim \text{Categorical}(\psi_{v(i)}) \quad (11)$$

and initially marginalize out the node attachment vector π. Priors on all other parameters of the likelihood functions (2-11) are specified above and in the source code and are weakly informative unless specified above. We noticed that the parameters of our model (in particular, the clonal fraction ψ) strongly depend on the choice of A, making a joint inference of A and the remaining parameters inefficient.

We therefore perform model comparison with fixed A, using the evidence lower bound (ELBO). CloneTracer offers the possibility of generating all possible trees that are compatible with the three-gamete rule and the infinite site hypothesis and compute the ELBO for each configuration, however, in the case of larger numbers of mutations, this brute-force approach leads to unreasonably long runtimes and we therefore additionally implemented a search heuristic, where all tree configurations for the two most highly covered variants are explored and compared, and mutations are subsequently added to this tree in the order of their coverage.

We ensure convergence of the ELBO by plotting its value over iterations; in general, 400 iterations, starting with 5 particles, are sufficient for the ELBO to converge in patients harboring SNVs. For patients with CNVs 1000 iterations were selected to ensure convergence.

Once A is identified based on the ELBO, and all other parameters with the exception of the node attachment vector π are inferred, we infer π given the variational posteriors for all other parameters.

For all analysis that require a discrete assignment of cells as leukemic or healthy, we assigned all cells as leukemic if their posterior probability for being from the leukemic

clone according to the CloneTracer model was >0.8, healthy if the posterior probability was <0.2, and excluded them from clonal analyses otherwise.

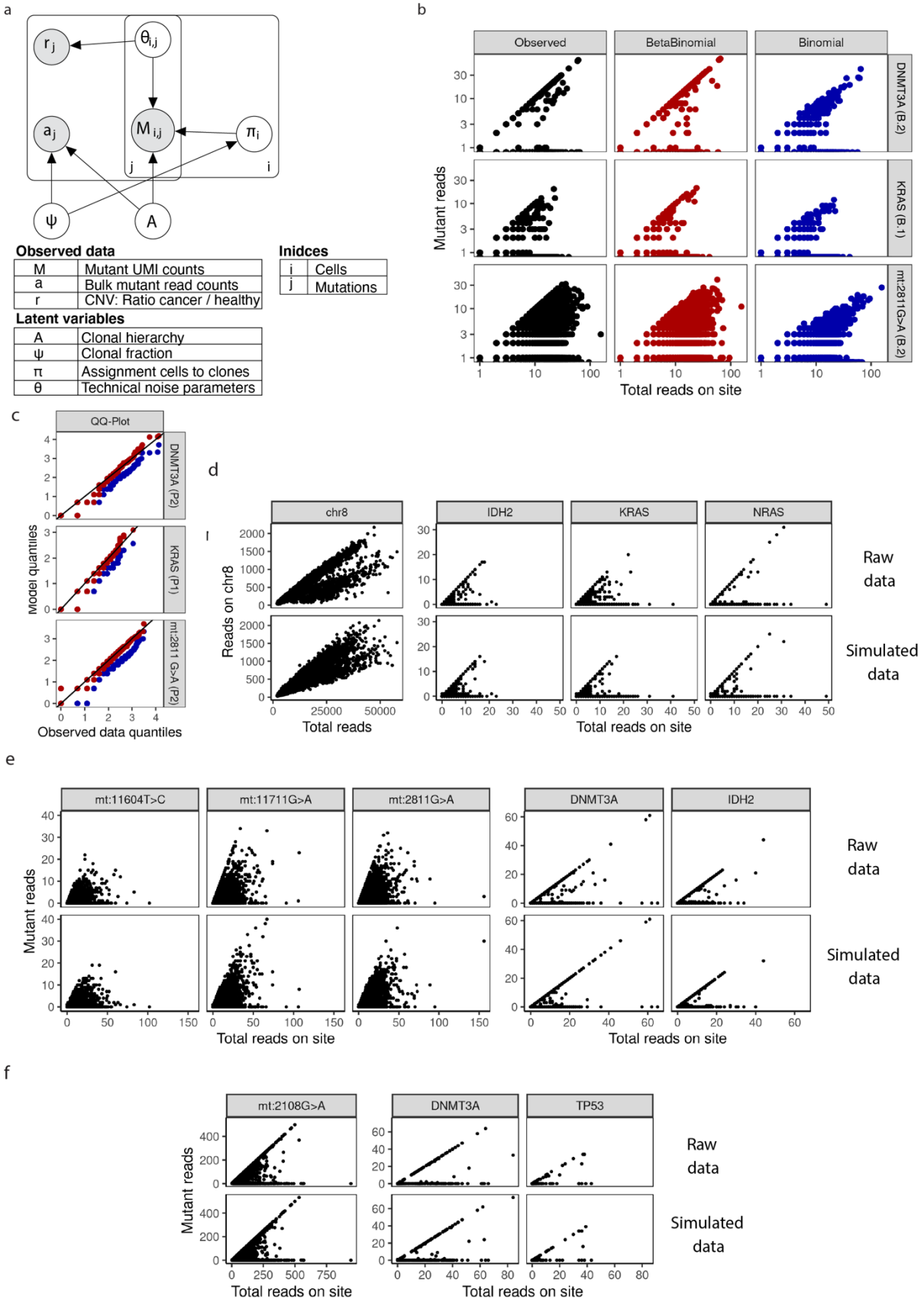

Figure N1. *Justification of basic modeling choices.* ***a.*** *Graphical display of the model, grey: observations, white: latent variables.* ***b.*** *Number of mutant reads plotted against number of total reads for three mutations (left column, black dots). The data was fitted using a mixture of a binomial sampling process to describe dropout of mutant reads in mutated cells and a poisson process to describe background false positive mutant reads in healthy cells (right column, blue dots). Alternatively, an overdispersed beta-binomial sampling process was used (central column, red dots).* ***c.*** *Quantile-quantile plots comparing the modeled beta-binomial and the binomial sampling process to the data. Red: Beta-Binomial model, blue: binomial model.* ***d.*** *Posterior predictive check for the full model. Top row: Raw data for patient B.1. Bottom row: The complete model described in the Supplementary Methods was fit to the data. Subsequently, mutant read counts were simulated from the model while keeping the total read counts and the clonal assignments fixed (posterior predictive check). Clonal assignments were kept fixed for this analysis since the model does not describe the relationship between total read count and cell type.* ***e.*** *Raw data and Posterior predictive check for patient B.2.* ***f*** *Raw data and Posterior predictive check for patient B.3.*

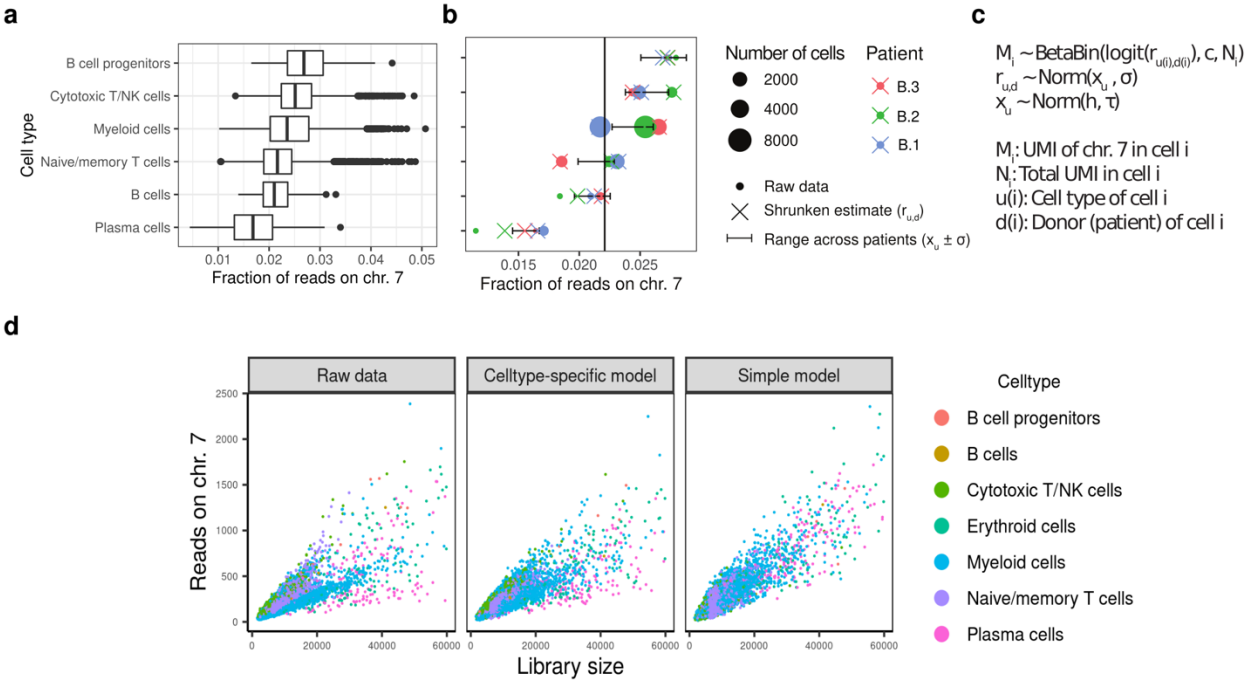

Figure N2. *Accounting for cell type as a covariate in CNV analyses.* ***a.*** *Boxplot depicting the UMIs on chromosome 7 as a fraction of the total number of UMIs per cell. Data from patients B.1-B.3 (not affected by a monosomy 7) are shown.* ***b.*** *Inference of a cell type specific prior. Fraction of UMIs on chromosome 7 in different cell types is plotted; circles denote raw observed UMI fraction per patient and cell type, crosses denote a shrunken estimate, and the horizontal bar indicates the posterior estimate on the cell type specific UMI fraction and its standard deviation across patients. Inference was done using stan from the hierarchical model specified in panel c, and all cells from patients B.1, B.2 and B.3.* ***c.*** *Model used for estimating the fraction of UMIs on chromosome 7 per cell type, and its variance across patients.* ***d.*** *Raw data and Posterior predictive check for patient B.4, using both the simple model with no cell type covariate, and the cell type specific model.*

Application of CloneTracer to cohorts A and B of AML patient samples

CloneTracer relies on well-covered clonal markers for the high-confidence identification of healthy and leukemic cells as well as sub-clones when present. Several types of genomic aberrations can be leveraged as clonal markers: nuclear single nucleotide variants (SNVs) affecting highly expressed genes (e.g. *NPM1*), mitochondrial SNVs (mtSNVs) and copy number variants (CNVs). For samples which only harbor nuclear SNVs occurring in lowly expressed genes, CloneTracer was unable to infer the clonal hierarchy due to the high dropout levels (see section *samples without well-covered clonal markers)*. As a result, healthy and leukemic cells could not be confidently identified.

We selected nuclear mutations to include in this analysis as follows: All mutations with a distance of <1.5kb to the 3' end an average expression of >0.2 UMIs per cell in the patient were selected for primer design. Mitochondrial mutations were selected based on bulk ATAC sequencing, where available, or as previously described[21] (see also STAR methods).

When it comes to selecting the clonal hierarchy for a particular sample, CloneTracer can output more than one possible tree configuration with equal statistical evidence. We often observed this behavior when poorly covered mutations are included (e.g. *DNMT3A*). In those cases, we always selected the mutation hierarchy with the lowest number of nodes.

In the following subsections we describe in detail the output of CloneTracer when applied to AML samples from cohorts A and B.

Samples with highly covered nuclear SNVs

NPM1 288fs mutation is observed in around 30% of AML patients[44]. Since the gene is highly expressed in immune cells, it can be leveraged as clonal marker (Figure N3).

In some instances, CloneTracer inferred that *DNMT3A* mutations occurred upstream of *NPM1* (patients A.8, A.9 and A.13). However, due to the low coverage of *DNMT3A* SNVs, confident identification of pre-leukemic cells was not possible for most cells. As a result, we computed the healthy probability as the sum of healthy and *DNMT3A* clonal probabilities. We used the resulting values for assigning cells as healthy and leukemic. In patients A.5, A.11 and A.12 pre-leukemic mutations in *DNMT3A* or *IDH2* co-occurred with *NPM1*. This may be due to the absence of pre-leukemic cells in these patient samples or that most pre-leukemic cells dropped out for *IDH2* or *DNMT3A* mutations and therefore there was no statistical evidence to place them upstream of *NPM1* mutation. In both scenarios we cannot rule out that among cells labelled as healthy there are pre-leukemic cells.

We showed that more sensitive methods such as MutaSeq[21] (Smartseq2-based method) or sequencing larger number of cells with optimized 10x enabled the confident distinction of healthy and pre-leukemic cells (Figure 4 and Figure S5 of the manuscript). Therefore, we recommend such approaches for the analysis of pre-leukemic cells.

It is worth noting that synonymous mutations affecting highly expressed genes can also be used as clonal markers for CloneTracer. Patient A.8 is an example of this with

a SNV in *RPS29*, a ribosomal gene which is covered in every single cell. *RPS29* mutation co-occurred with NPM1 mutation in this patient increasing the confidence in the distinction between healthy/pre-leukemic and leukemic cells.

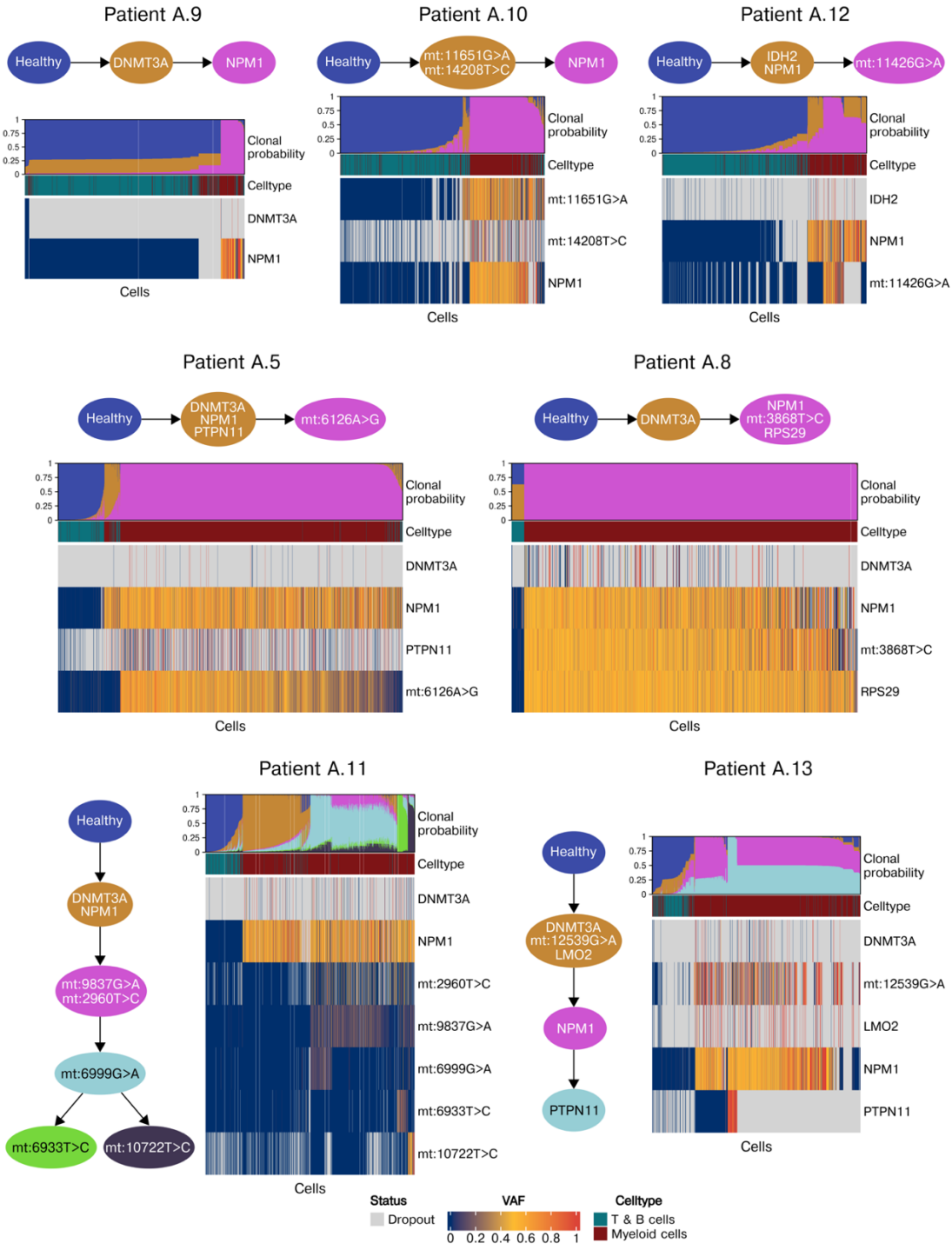

Figure N3. *Application of CloneTracer to AML samples with well-covered nuclear SNVs. For each patient the clonal hierarchy inferred by the model is shown. The heatmaps display for each single cell the fraction of mutant reads (VAF, variant allele frequency) in all mutations as well as the cell type identity. Gray indicates that neither reference nor mutant reads were observed, termed as dropout. The clonal probabilities inferred by CloneTracer are shown at*

the top of the heatmap. Cells are ordered by decreasing clonal probability from the top to the bottom of the inferred clonal hierarchy.

We encountered mitochondrial variants which labelled pre-leukemic clones such as in A.13. In other patient samples, mtSNVs were main leukemic clonal markers as they co-occurred with *NPM1* mutations (patient A.8). We also observed mitochondrial mutations happening downstream of the first leukemogenic hit (patients A.5, A.11 and A.12). Particularly in A.11 mtSNVs gave rise to further downstream subclones. A.10 was the only patient in which mtSNVs were inferred upstream of *NPM1* but no other (pre-) leukemic mutation was amplified. We considered the mtSNVs clone as pre-leukemic because we noticed that the patient had a somatic mutation in *TET2* with high allele frequency from exome sequencing results (see Supplementary Table 3). Similarly, to patients A.8, A.9 and A.13, the healthy probability of A.10 cells was computed adding the healthy and mitochondrial clone probabilities.

Samples with mitochondrial SNVs as clonal markers
In two patients for which no well-covered SNVs or CNVs were observed, mitochondrial variants were present in the founding clone. Therefore, we leveraged mtSNVs as clonal markers for the distinction of healthy and leukemic cells (Figure N4).

Patient A.6 (also shown in Figure 1h) contained mitochondrial variants which acted as main clonal markers and subclonal markers. All cells carrying the nuclear *MPO* mutation, which was observed in exome data with an allele frequency of ~0.5, were assigned to the mitochondrial clone *mt:3019G>C*, again indicating that all cells assigned to this mitochondrial clone are leukemic. DNAseq from single-cell colonies confirmed the validity of *mt:3019G>C* mutation as main leukemic marker and validated the inferred hierarchy by CloneTracer (Figure 1i).

For patient B.3, CloneTracer inferred a clonal hierarchy with a single clonal population. This clone also harbored mutations in *DNMT3A* and *TP53* which are often found in individuals with clonal hematopoiesis, indicating that all leukemic and possibly some residual pre-leukemic cells are contained in this clone.

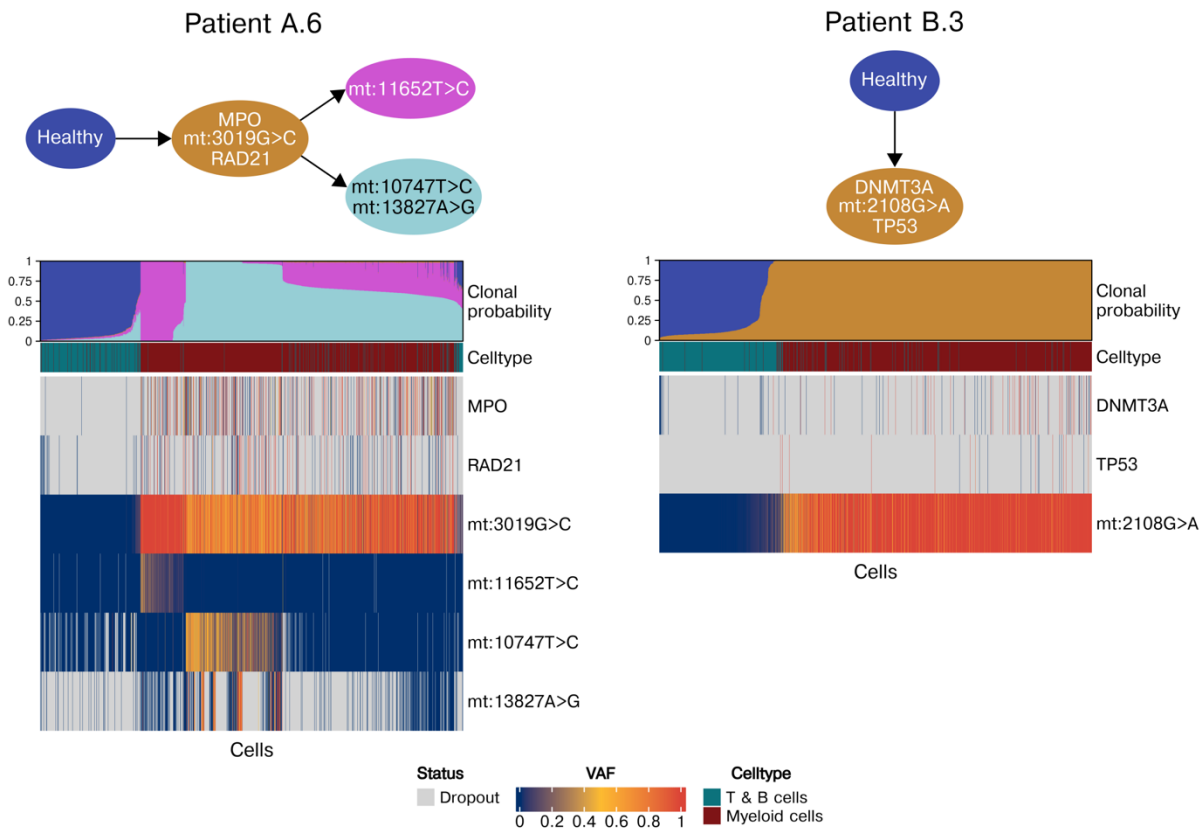

Figure N4. *Application of CloneTracer to AML samples with mitochondrial SNVs as clonal markers. For each patient the clonal hierarchy inferred by the model is shown. The heatmaps display for each single cell the fraction of mutant reads (VAF, variant allele frequency) in all mutations as well as the cell type identity. Gray indicates that neither reference nor mutant reads were observed, termed as dropout. The clonal probabilities inferred by CloneTracer are shown at the top of the heatmap. Cells are ordered by decreasing clonal probability from the top to the bottom of the inferred clonal hierarchy.*

Samples with CNVs

Similar to previous approaches[16,17], CloneTracer infers CNVs leveraging the fact that aneuploidies have an effect on the expression of genes in the affected regions compared to diploid cells. In the case of large chromosomal aberrations such as monosomies or trisomies the amount of data is much larger compared to SNVs, therefore facilitating clonal tracking.

CloneTracer requires knowledge of the location of chromosomal aberrations. We identified large aneuploidies from clinical karyotyping of the samples. In the case of partial deletions (labelled as chrX-part in Figure N5) we used Numbat[17] with default settings to identify the boundaries of the alterations.

The advantage that CloneTracer provides over current tools to infer CNVs from scRNAseq, is that the information on chromosomal alterations can be combined with mtSNVs and nuclear SNVs, enabling us to clarify if the CNV is really a clonal marker, and the discovery of potential subclones. This is particularly relevant in samples with a complex clonal hierarchy such as patient B.1 (Figure N5). In this sample, chromosomal aberrations in chromosomes 3 and 8 occurred early in the evolution of the leukemia and therefore are optimal markers for the distinction between healthy and leukemic cells. However, the inclusion of nuclear SNV data enabled the identification of 2 additional mutually exclusive subclones driven by mutations in *KRAS* and *NRAS* (Figure N5 and S6a).

We found other examples in which copy number alterations were part of the founding clone of the leukemia such as patients A.1 and A.3 (Figure N5). On the other hand, we also observed cases in which CNVs were subclonal as shown for patients B.4 and A.2 (Figure N5) and B.2 (Figure N6). The identified subclones were phenotypically distinct (see Figure S6d-g).

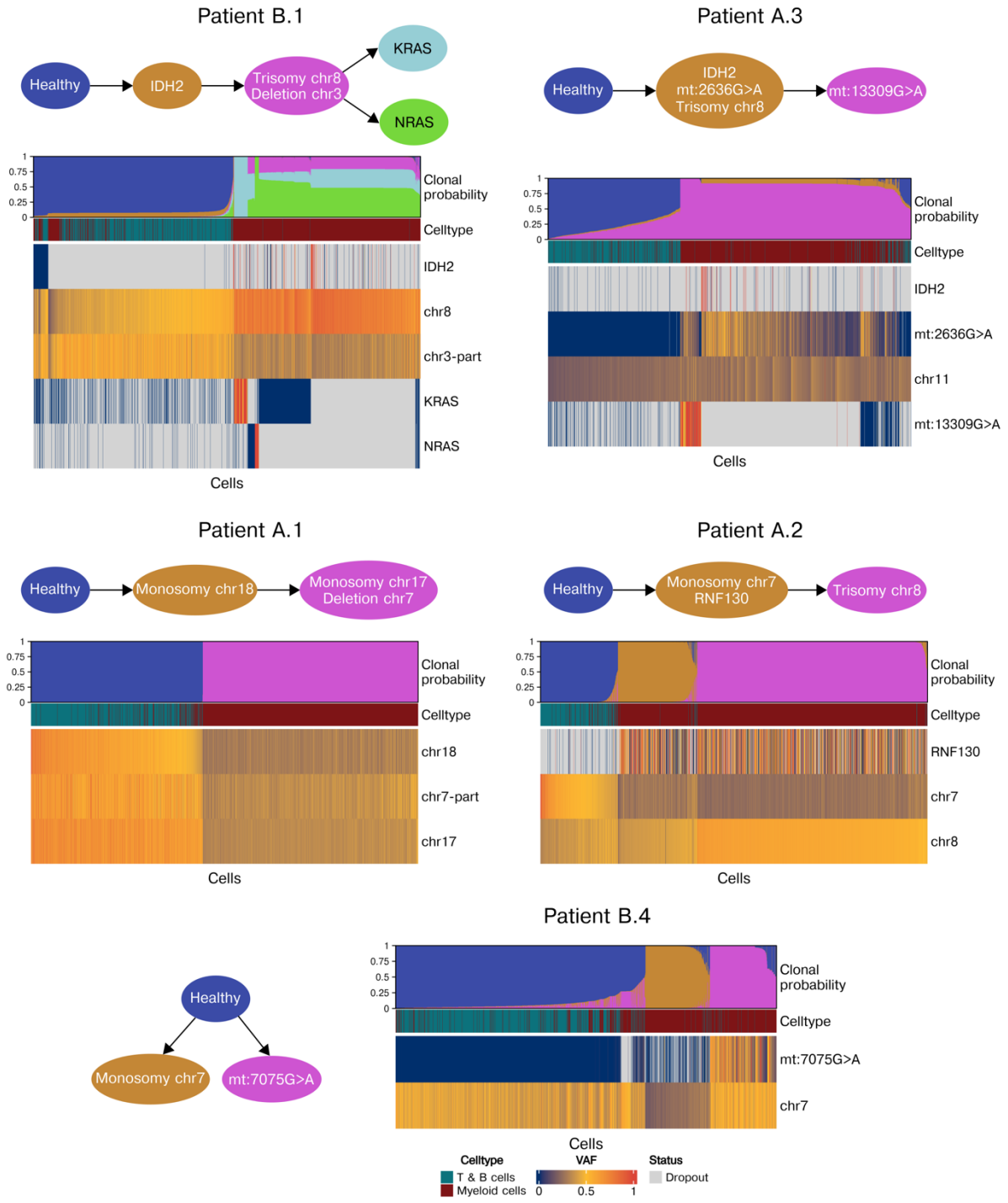

Figure N5. *Application of CloneTracer to AML samples with CNVs as clonal markers. For each patient the clonal hierarchy inferred by the model is shown. The heatmaps display for each single cell the fraction of mutant reads (VAF, variant allele frequency) in all mutations as well as the cell type identity. For chromosomal regions, the scaled fraction of reads falling into the region is shown. Gray indicates that neither reference nor mutant reads were observed, termed as dropout. The clonal probabilities inferred by CloneTracer are shown at the top of the heatmap. Cells are ordered by decreasing clonal probability from the top to the bottom of the inferred clonal hierarchy.*

Samples with subclonal markers

We have shown that CloneTracer can use mtSNVs and CNVs as clonal markers for the identification of healthy and leukemic cells. However, there are instances in which these alterations occur downstream of the initial leukemic founding mutations and thereby cannot be used to distinguish healthy and malignant cells. This phenomenon occurred in patient B.2 (Figure 1j and N6) in which the mitochondrial variants and the trisomy in chromosome 8 occurred downstream of mutations in *DNMT3A* and *IDH2* genes which presumably are the initial drivers of the leukemia[24]. In these situations, healthy cells could not be confidently identified (Figure N8a). Notwithstanding, a large fraction of cells could unambiguously be labelled as leukemic and assigned to different subclones if present.

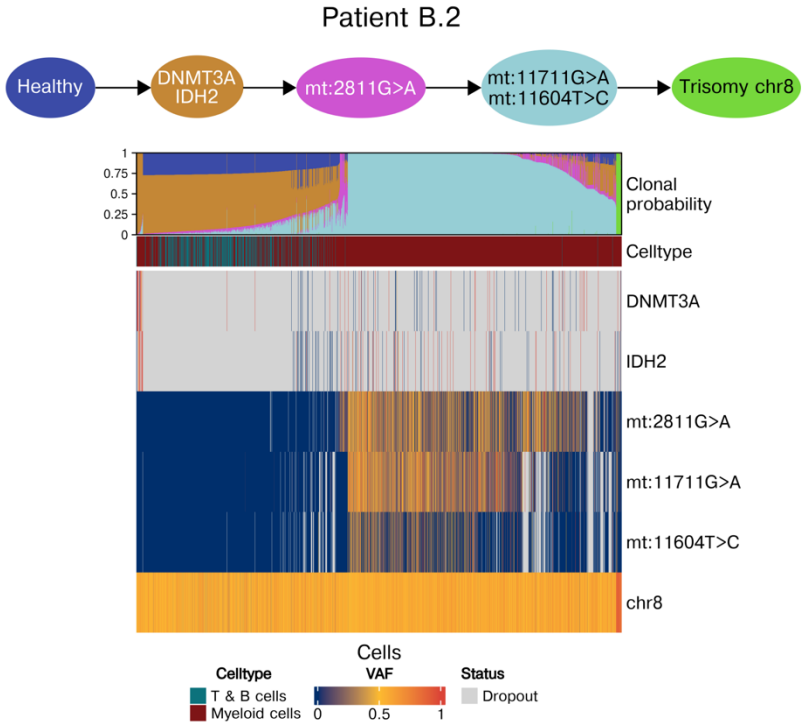

Figure N6. *Application of CloneTracer to AML samples in which the confident identification of healthy cells was not possible. For each patient the clonal hierarchy inferred by the model is shown. The heatmaps display for each single cell the fraction of mutant reads (VAF, variant allele frequency) in all mutations as well as the cell type identity. For chromosomal regions, the scaled fraction of reads falling into the region is shown. Gray indicates that neither reference nor mutant reads were observed, termed as dropout. The clonal probabilities inferred by CloneTracer are shown at the top of the heatmap. Cells are ordered by decreasing clonal probability from the top to the bottom of the inferred clonal hierarchy.*

Samples without well-covered clonal markers

For patient samples in which no good clonal markers were found, the inference of the hierarchy and subsequent identification of healthy and leukemic cells was not possible (Figure N7). We observed that samples in which less than 60% of cells were covered in at least one mutation CloneTracer failed (Figure N8b).

Among samples for which clonal tracking was not possible we observed cases in which neither mtSNVs nor CNVs were detected and only low-covered SNVs were

amplified (patients A.4 and A.14). In other cases, mtSNVs were present but their coverage was either insufficient (patient A.7) or they clearly labelled a small subclone (patient A.15, <2% of covered cells had a mutant read and tumor bulk ATAC VAF: 0.015 for *mt:14386T>C*).

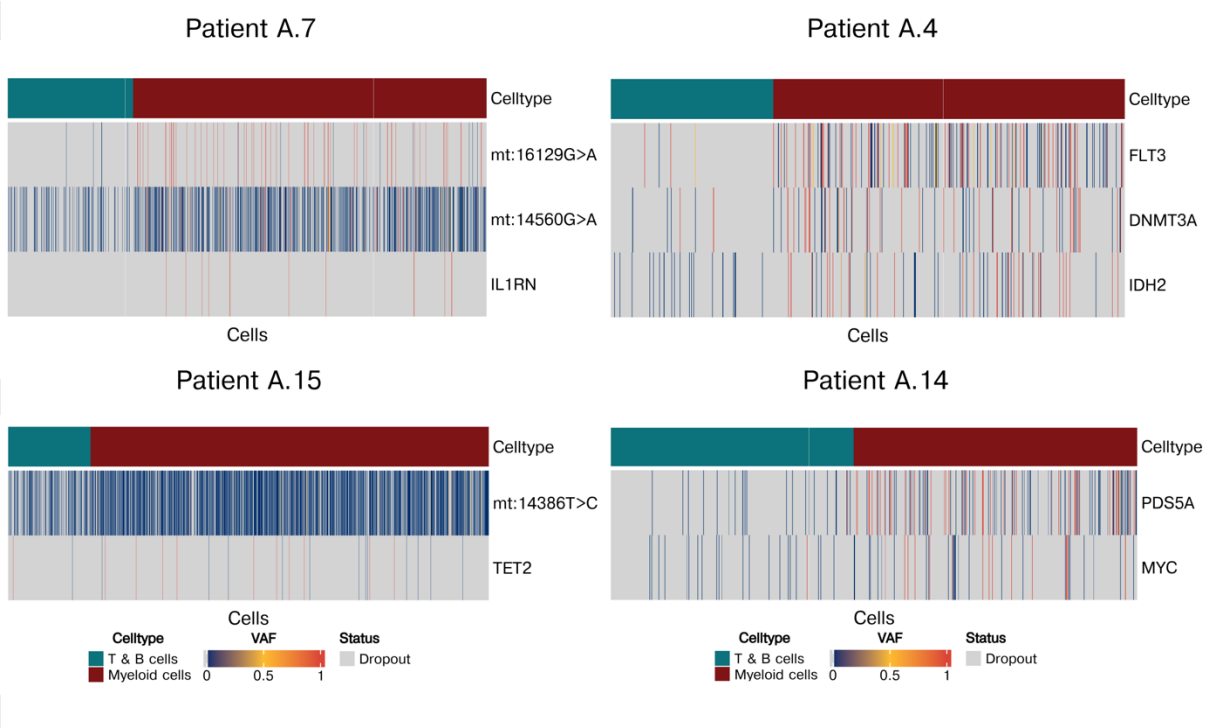

Figure N7. *Samples without well-covered clonal markers are not suited for clonal tracking with CloneTracer. For each patient a heatmap displays for each single cell the fraction of mutant reads (VAF, variant allele frequency) in all mutations as well as the cell type identity. Gray indicates that neither reference nor mutant reads were observed, termed as dropout. Cells are order by cell type identity.*

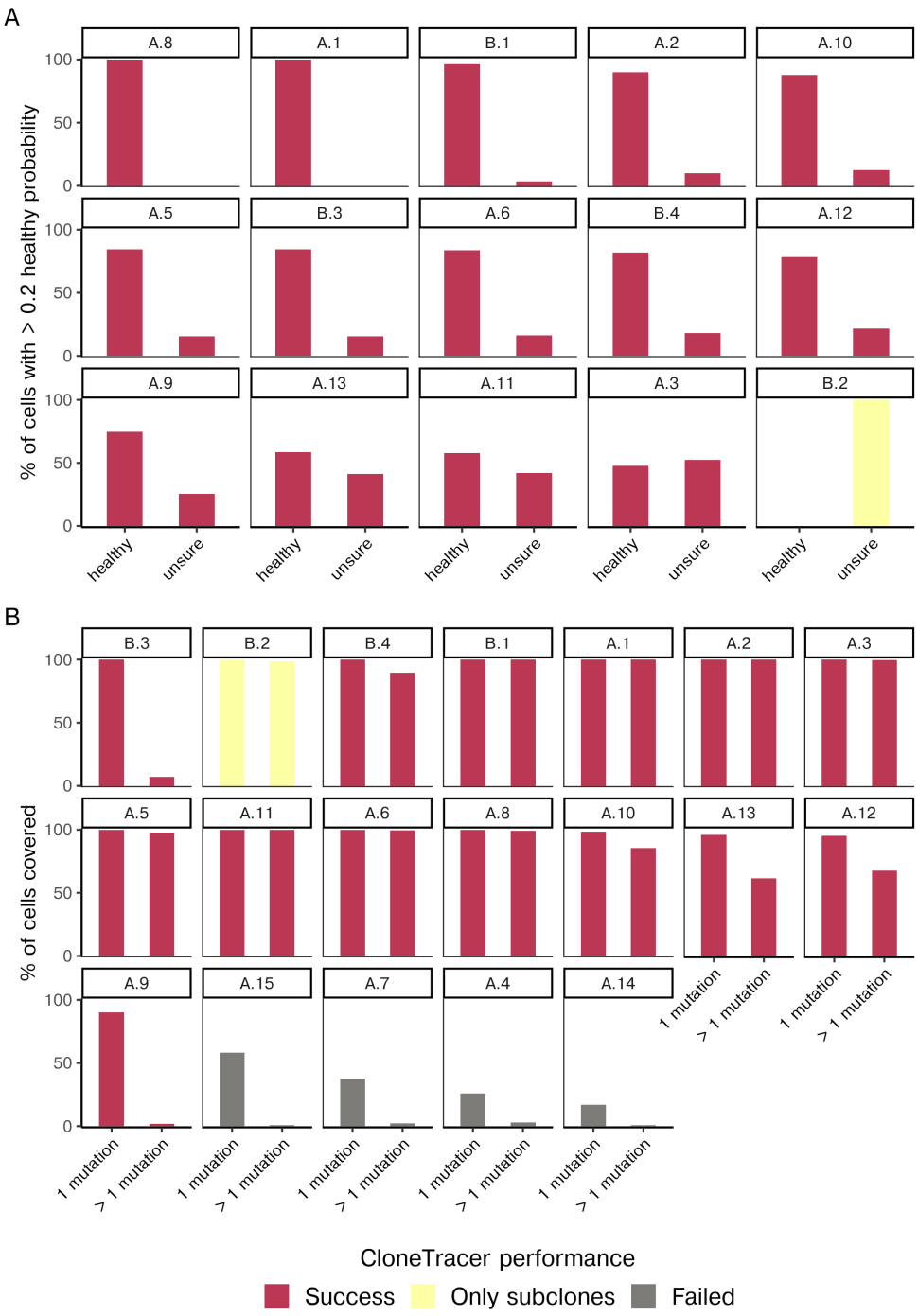

Figure N8. *Criteria to determine the performance of CloneTracer.* ***a.*** *The percentage of cells assigned as healthy and unsure among cells with >0.2 posterior probability of being healthy is shown for each patient in which CloneTracer was ran. Samples in which the fraction of unsure cells was >60 % were labelled as "Only subclones", while the rest were labelled as "Success" since healthy and leukemic cells could be confidently assigned.* ***b.*** *The percentage of cells covered in 1 mutation and at least 2 mutations is shown for each patient. Samples with <60% of cells covered with at least 1 mutation are labelled as "Failed".*

Different cell types have different confidence rates of clonal assignment
Lymphoid cells had a lower RNA content / library size compared to myeloid cells as shown in Figure N9. Furthermore, the expression of genes often mutated in AML (e.g., *NPM1*) and mitochondrial genes was lower compared to myeloid cell types. As the coverage of SNVs and mtSNVs is strongly dependent on the expression of the affected gene, lymphoid cells had lower coverage which resulted in higher uncertainty inferred by CloneTracer.

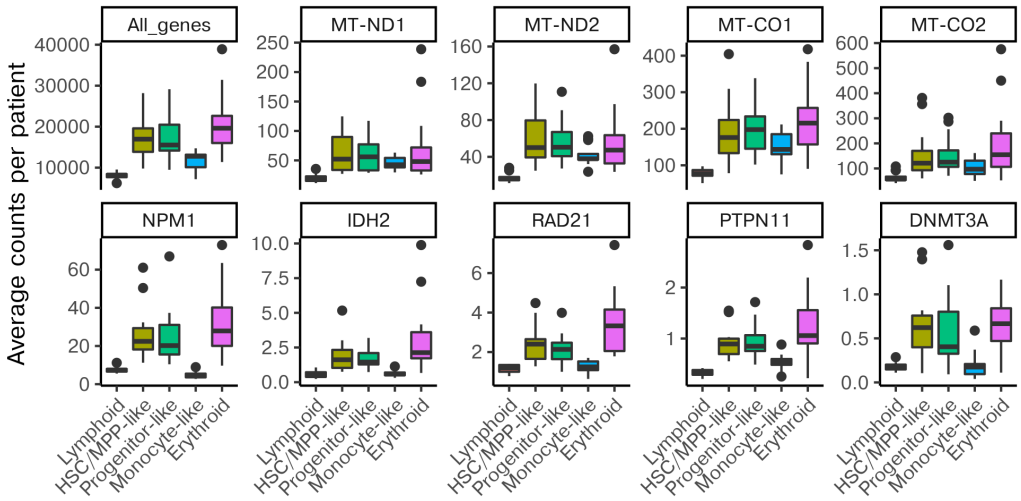

Figure N9*. Expression of selected genes across different populations. The average UMIs per patient is shown for different lineages.*

Sequencing depth requirements for Optimized 10x libraries
To determine how deep Optimized 10x libraries needed to be sequenced to achieve saturation, we computationally downsampled reads from different patient libraries and computed the fraction of cells covered. We concluded that for most nuclear SNVs an average of 500 reads/cell and site are sufficient to achieve sequencing saturation (Figure N10). We observed that for highly expressed genes such as NPM1 in some patients a depth of 1000-1500 reads/cell improved the coverage of individual cells.

We did the same analysis for the mitochondrial libraries and observed that sequencing between 10-15x10^3 reads per cell was sufficient to achieve sequencing saturation in most mtSNVs tested (Figure N11).

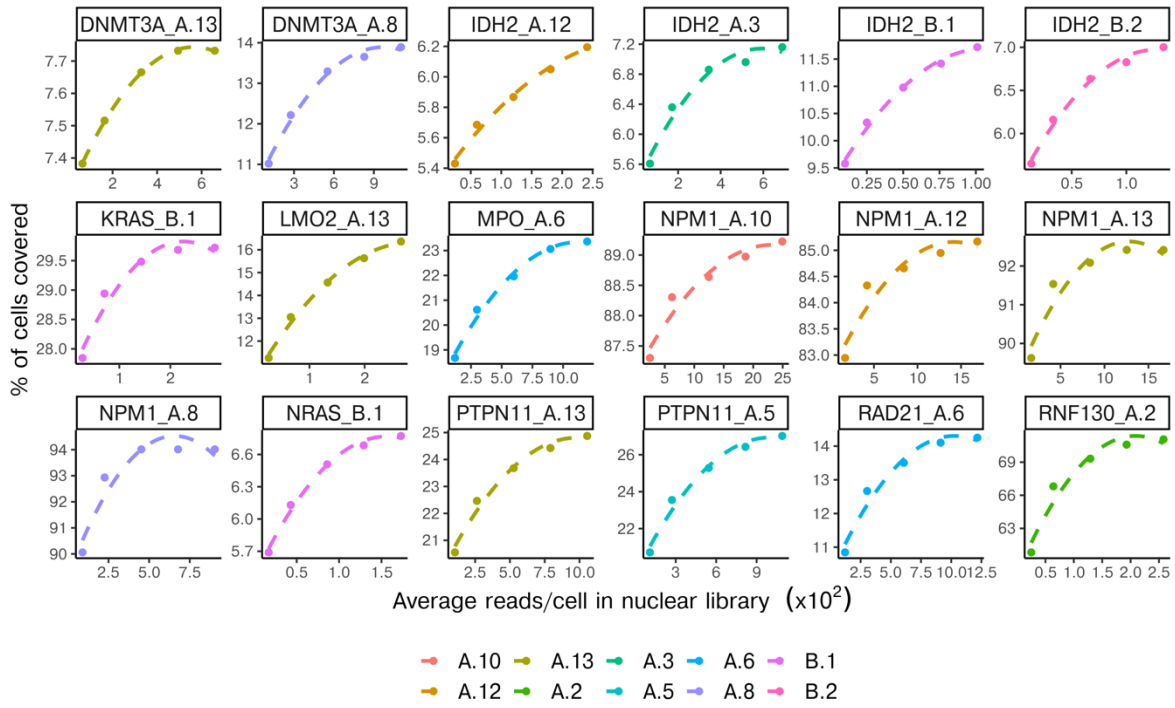

Figure N10. *Sequencing saturation of Optimized 10x SNV nuclear libraries was analyzed by downsampling. Sequence reads for nuclear mutation libraries were downsampled to 10, 25, 50, 75 and 100% of the original coverage (points). The average number of downsampled reads per cells is plotted against the percentage of cells covered for the nuclear SNV.*

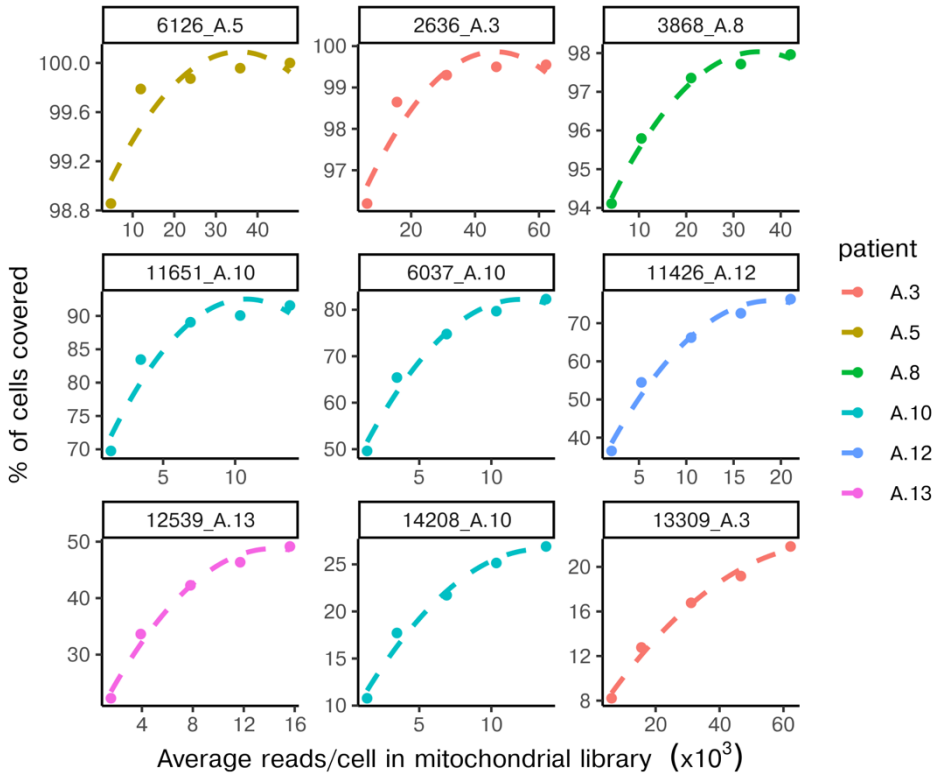

Figure N11. *Sequencing saturation of Optimized 10x mitochondrial libraries was analyzed by down-sampling. Sequence reads for mitochondrial mutation libraries were downsampled to 10, 25, 50, 75 and 100% of the original coverage (points). The average number of*

downsampled reads per cells is plotted against the percentage of cells covered for the mitochondrial SNV.

Evaluation of long-read sequencing for Optimized 10x libraries

The main coverage limitations of our SNVs Optimized 10x libraries are the distance from the mutation to the polyA and the expression of the affected gene. The former is due to the difficulty to amplify long fragments in nested PCRs and the inherent bias of Illumina sequencers towards short fragments[19]. In order to determine if long-read sequencing could improve the coverage of nuclear SNVs we took the PCR3 product of the Optimized 10x protocol and prepared it for Oxford Nanopore sequencing as previously described[20]. To permit for deep sequencing of 600/reads per cell, we selected 4 SNVs which led to short and long fragments as well as affecting highly and lowly expressed genes.

We processed the data using sicelore[20] (https://github.com/ucagenomix/sicelore-2.1) and observed that long-read sequencing does not lead to higher fraction of cells covered in SNVs which give rise to short fragments and fall in highly expressed genes such as *NPM1* (Figure N12a). Coverage of SNVs in lowly expressed genes which were located far away from the polyA only marginally improved (*DNMT3A* and *TET2*). For a mutation in *MPO*, a highly expressed gene, and located almost 2kb away from the polyA, we obtained a 2-fold increase in the fraction of cells covered hinting at the potential benefit of using long-read sequencing in such instances. Of note, in our cohort, the *MPO* mutation was the only instance of a mutation in a highly expressed gene with >1.5kb distance from the polyA site.

We naively assigned each cell as mutant or reference based on the presence of a mutant read for *NPM1* and *MPO* and used T cells and myeloid cells as ground truth of healthy and leukemic. We computed the FPR and FNR for Optimized 10x libraries sequenced with Illumina and Nanopore technologies. We observed that nanopore sequencing led to a higher false positive rate (Figure N12b**,** from 11% to 19% and from 0.3% to 5%, respectively) and to a larger false negative rate (from 8% to 46% in NPM1 and from 11 to 12% in MPO).

To see if the higher coverage obtained with long read sequencing leads to overall better clonal assignments, despite higher error rates, we computed the FPR and FNR of CloneTracer assignments in patient A.6 using nanopore and standard Optimized 10x data. We observed that nanopore data led to an increase of false positives (more T cells being assigned as leukemic) while no change in the FNR was observed.

Given the imbalances between FPR and FNR from long-read sequences we did not include this data for clonal tracking, although it may be beneficial in some instances.

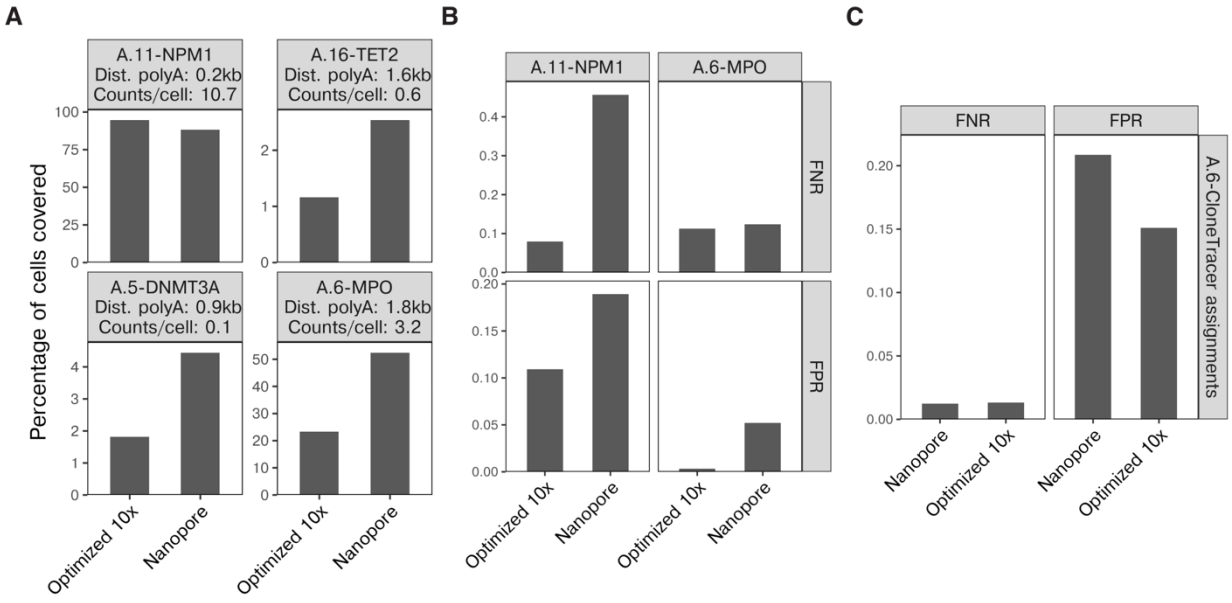

Figure N12. *Evaluation of Optimized 10x libraries sequenced with Nanopore.* ***a)*** *coverage comparison between standard Optimized 10x and nanopore sequenced libraries for 4 SNVs. Dist. PolyA refers to the approximate distance of a mutation to the polyA. It was estimated as described in TAPseq*[18] ***b)*** *FPR and FNR comparison between Optimized 10x and Nanopore sequenced libraries for NPM1 and MPO. Cells were naively assigned as reference or mutant based on the presence of a mutant read and T cells and myeloid cells were used as ground truth of healthy and leukemic, respectively.* ***c)*** *FPR and FNR computed as in* ***b)*** *but using the CloneTracer assignments for patient A.6.*

Evaluation of enrichment strategies for immature cells

In single cell studies of healthy hematopoietic differentiation[1,4,28], CD34 enrichment is commonly performed to increase stem- and progenitor cell numbers available for analyses. In AML, the consequences of such enrichment have never been systematically evaluated.

In this study, we compared FACS and magnetic bead (MACS) enrichment of CD34+ cells. For patient A.2, we did separate runs of total bone marrow, CD34+ cells isolated with FACS and CD34+ cells isolated with magnetic beads (MACS). We confirmed cell numbers by trypan blue staining and counting and loaded identical cell numbers on each 10x flow cell. Thereby we recovered half the number of cells in the MACS sample compared to FACS after quality filtering (Figure N13a). We did not observe significant differences in the quality of the cells after applying quality control filters (Figure N13b). Additionally, we observed a high correlation in the abundance of celltypes identified by projecting the two samples to either the Triana 2021 healthy reference map (to identify the most similar cell type, Figure N13c) and to our AML map (Figure N13d). Based on this analysis, we conclude that at least in our hands there are no substantial differences between isolating CD34+ cells via flow cytometry or magnetic beads except for a lower number of high-quality cells in the latter. This justifies the use of FACS enrichment for the main data set.

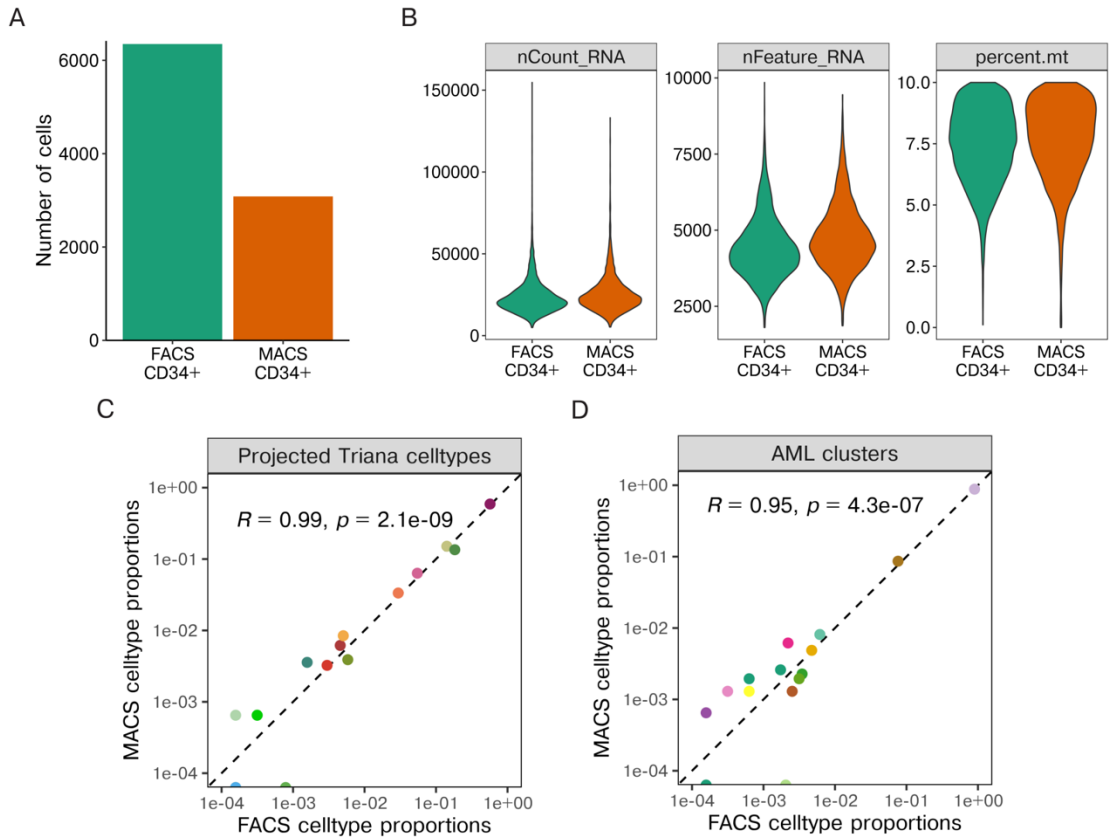

***Figure N13**. Quality control metrics comparison between FACS CD34+ and MACS CD34+ runs. **a.** Number of cells that passed quality control filters. **b.** Number of reads, number of genes and percentage of mitochondrial reads per cells are shown after filtering. **c.** Abundance of celltypes identified by mapping to the reference from Triana 2021. R indicates the Pearson correlation coefficient between runs and p the corresponding p_value. Color legend is equivalent to Figure S3a. **d.** Abundance of clusters identified by mapping to the AML map from this study. R indicates the Pearson correlation coefficient between runs and p the corresponding p_value. Color legend is equivalent to Figure S5d.*

To evaluate more generally the consequences of CD34+ enrichment in AML, we ran CD34+ and total BM on different 10x flow cells for the patients A.2 and A.9 in the context of the large cell number experiment outlined in Figure S5. This allowed us to comment on two relevant scenarios:

- In NPM1-mutated, CD34 negative AML cases, such as patient A.9, the total BM lacked the highly relevant CD34+ stem and progenitor cells. The stem cell population (‘cluster 6’) was only present with 7 cells in the total BM sample of patient A.9, vs. 260 in the CD34+ enriched sample (Figure N14a). Hence in these samples it is essential to sort CD34+ if statements about rare progenitor cells are intended.

- In an overall CD34 positive AML, such as patient A.2, the question was if enriching CD34+ cells would cause certain CD34- cell types to be missed or otherwise lead to false representations. This was not the case: All progenitor populations were well represented in both samples (Figure N14b). Only dendritic cells and lymphoid cells disappeared in the CD34+ enriched sample,

but were abundant and therefore sufficiently captured by the total bone marrow sample. Take note that for all patients, both CD34+ and CD34- cells were included in the sample subjected to 10x profiling.

In conclusion, CD34+ sorting is essential to capture stem and progenitor cells in leukemias with high degrees of differentiation or loss of CD34; in other samples it increases the number of relevant progenitor populations available for analysis at the expense of providing a lower representation of mature cell populations. This justifies the enrichment strategy used for the main data set.

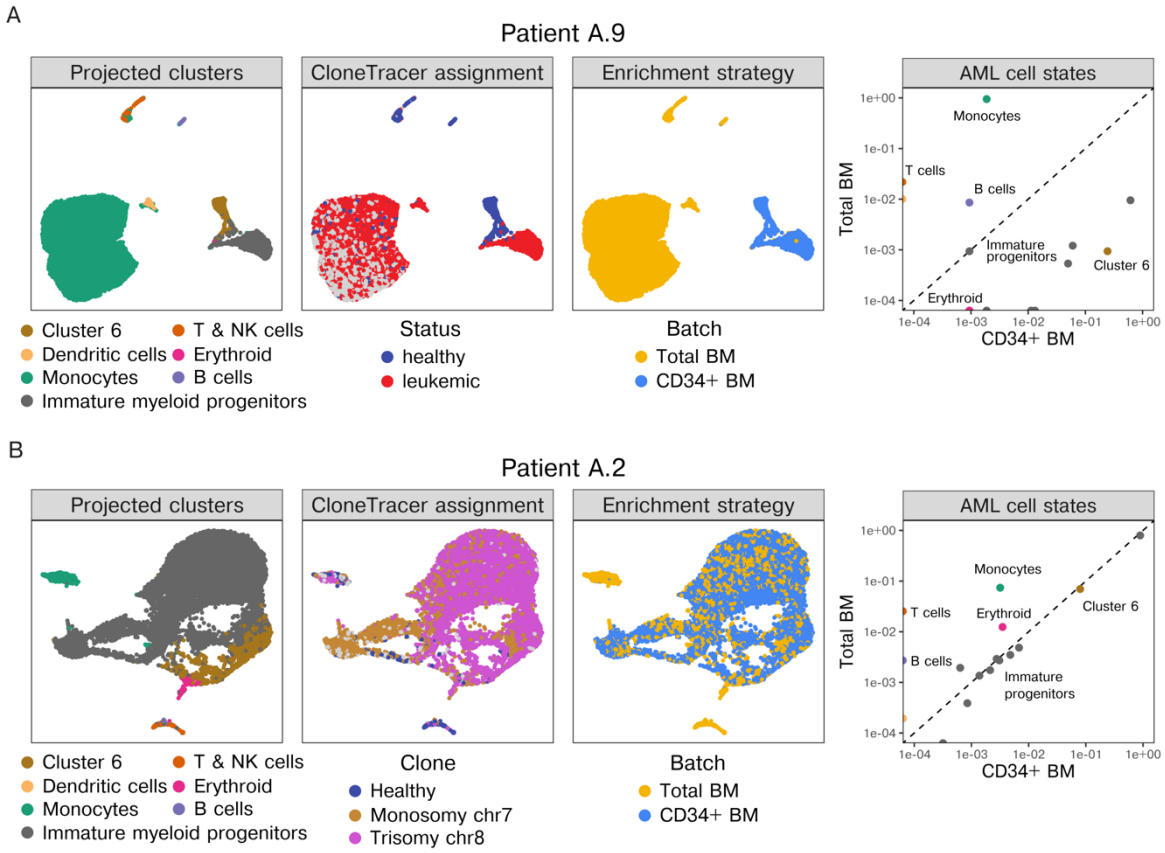

Figure N14. a. *Unsupervised analysis of cells from patient A.9, an overall CD34 negative AML patient. uMAPs, from left to right, highlight projected cluster identity, clone, and CD34 sorting gate. Right panel: Scatter plot comparing the abundance of different cell states/clusters in the total bone marrow and the CD34+ sample.* ***b.*** *Like panel A, but for patient A.2, a patient with abundant CD34 expression.*