Supplementary

1. Materials and methods

1.1 Acquisition and preprocessing

Sample acquisition

Bone marrow aspirates and blood samples (in the event of a dry tap) were obtained from AML patients at the time of diagnosis. EDTA tubes were used to collect the samples, which were then stained following the EuroFlow consortium's Standard Operating Procedures[1]. Following EuroFlow protocols[1,2], an acute leukemia orientation tube (ALOT) and six AML/MDS tubes were measured for each patient to establish and further refine the diagnosis of AML (table S2). At least 50 000 events were acquired on a FACSCanto II (BD Biosciences) which was calibrated and controlled as per the EuroFlow guidelines. After applying exclusion criteria (see above), 838 FCS files from 122 patients were retained. The leukemic blast cell population was extracted from these files using InfiniCyt software (Cytognos, BD), and both complete and blast FCS files underwent preprocessing using a computational pipeline.

Harmonization

Data needed to be harmonized due to varying marker and channel names over the files. The first step to ensure consistency in markers and channels nomenclature was to rename the markers Pacific Blue and Pacific Orange (PacO) to HV450 and HV500, respectively. The channels of interest (FSC-A, FSC-H, SSC-A, FITC-A, PE-A, PerCP-Cy5.5-A, PE-Cy7-A, APC-A, APC-H7-A, HV450-A, HV500-A) were then extracted. Additionally, the marker names were harmonized. This involved removing spaces, standardizing capitalization and hyphenation and ensuring consistency by adjusting alternative marker names. To validate correspondence between the filename and the actual measured panel in the tube, the filenames, which contained the tube number, were cross-referenced with the markers included in the corresponding FCS file. If any discrepancies were identified, the filenames were corrected accordingly or removed if no correct tube could be identified.

Blast extraction

The complete files were used to check the compensation (and compensation matrices were adjusted if necessary) and provide the appropriate transformation parameters for the logicle function. After preprocessing they were used to make a CytoNorm[3] model for batch effect removal. To get a more detailed insight into AML, the leukemic blasts were extracted manually from all original files by an expert using InfiniCyt software (Cytognos, BD).

Preprocessing leukemic cell files

Every FCS file containing only the leukemic cells was preprocessed to ensure data quality. The spillover matrix calculated at acquisition or a manually adapted matrix (n = 299) in case of compensation issues, was used to compensate the files. Additionally, a logicle transformation, with parameters determined on a complete representative file, was applied to the non-scatter channels. The scatters were linearly transformed to make the 0.01 and 0.99 quantiles in the complete representative file align with the median of these quantiles from the non-scatter channels. At last, all the 2D scatter plots were manually examined to guarantee data quality.

Preprocessing complete files

Every FCS file was preprocessed. The spillover matrix calculated at acquisition or a manually adapted matrix (in case of remaining compensation issues) was used to compensate the files. After removing all the margin events (the events that exceed the flow cytometer's detection range) of FSC-A, FSC-H and SSC-A using the PeacoQC package, a logicle transformation was applied on the non-scatter channels with the distributions determined on the correspondent markers of a representative file serving as the goal distribution. The scatters were linearly transformed where the medians of the 0.01 and 0.99 quantiles of the non-scatters of a representative file were used as goal distribution. To remove anomalies in the marker signal, PeacoQC[4] was used and debris was removed with the aid of FlowSOM using the algorithm below.

The lowest FSC-A and SSC-A density minimum was determined on an aggregate of all the files of a tube (FSC-A_{agg} and SSC-A_{agg}) using the flowDensity[5] deGate function. For tube 4 this cutoff was manually set to 0.2. The cells with an SSC-A higher than SSC-$A_{agg} + 1$ were separated in a flowframe (an object of the flowCore[33] package) and an FSC-A density minimum of this flowframe with high SSC-A was determined by deGate (FSC-A_{agg_high}). Each FCS file was clustered using a nine-by-nine SOM grid with the backbone channels of the non-ALOT tubes (FSC-A, SSC-A, FSC-H, HV450-A, HV500-A, PerCP-Cy5-5-A and PE-Cy7-A). First, the forward scatter density minimum closest to FSC-A_{agg} was established using deGate (FSC-A_{ff}). If FSC-A_{ff} was higher than 1, it was switched to FSC-A_{agg} and if it was lower than 0.05, it was replaced by FSC-A_{agg} / 2. The intersection with the y-axis was calculated of the line l drawn with FSC-A_{ff} as the intersection with the x-axis and a slope of 4, the slope of the debris cloud. The minimum SSC-A density minimum was determined by deGate (SSC-A_{ff}) and if it was lower than SSC-A_{agg} / 2, it was changed to SSC-A_{agg}. A flowframe was made with all the cells with an SSC-A higher than SSC-$A_{ff} + 1$ and of this flowframe all the FSC-A density minima (FSC-$A_{ff\_high\_all}$) were calculated. FSC-A_{ff_high} then equals the minimum of the absolute values of FSC-$A_{ff\_high\_all}$ - FSC-A_{agg_high}, which is the closest value to FCS_{agg_high}. If the rounded value of FSC-A_{ff_high} was larger or equal to 2 times FSC-A_{agg_high}, it was replaced with FSC-A_{agg_high}. All the cluster centers that lay left of line l and lower than FSC-A_{ff_high} were marked as debris. As additional criteria, clusters that lay between l and m, with m being a line with the same slope as l and moved with 1 to the right, were also marked as debris if 15% of the cells lay left of l. The clusters marked as debris were removed as were the cells left of FSC-A_{ff}. For 58 files, clusters were manually annotated as debris.

Afterward, doublets were removed using PeacoQC doublet detection tool with a manually adapted MAD score. At last, all the 2D scatter plots were manually checked to assure data quality (Figure S2).

Normalization

Since the data was acquired over five years, a lot of technical variation (also called batch effects), could arise due to machine defects, human error, acquisition problems etc. Three kinds of batch effects were observed. The first batch effect occurred in every tube and was caused by switching fluorochromes from PacO to HV500 (Figure S3A). This batch effect spread to APC in tubes 1 and 6. A second batch effect appears as a difference in the acquisition year 2019 versus the other years in tube 6 APC and APC-H7 (Figure S3B). The last batch effect (Figure S3C) arose between the files before and after August 1, 2016 and is noticed in tube 2 PE, PerCP-Cy5.5 and PE-Cy7 (Table S3). For these batch effects, no obvious causes could be identified, although we suspect the use of a brighter fluorochrome. Using the QuantileNorm function from the package CytoNorm[3], a normalization model was trained on the complete files, learning a transformation from each batch to the mean of all batches. The same transformation was then applied to normalize each batch of blast files.

1.2 Automated gating using FlowSOM and immune cell population extraction

FlowSOM was used for automated gating, using a 20 x 20 grid. FlowSOM works by first clustering all the cells together with a self-organizing map (SOM), in this case in 400 clusters. A separate FlowSOM model was created for each tube of the EuroFlow panel (Table S2). The model was built on an aggregate (containing all the markers included in the tube and a total of 3 000 000 cells equally distributed over all files) of all the manually gated leukemic cells. Labels were assigned to each cluster of the FlowSOM model based on marker expression (i.e. positive or negative for each marker in the tube), with cutoff values for positivity determined by density minimums (using the deGate function[5]) on the aggregates. Manual adjustments were made if necessary (6 out of 49 cases). Afterward, the clusters with the same label were grouped together. These are called metaclusters and resemble cell populations.

1.3 XGBoost leave-one-out cross-validation

Firstly, FlowSOM models were built for each fold (i.e. all the patients but one), and features were extracted as described before. Next, feature selection methods were performed to address overfitting (i.e. the ML model makes accurate predictions on training data but not on new data): statistically significant features with a threefold change or higher and uncorrelated features (maximum correlation = 0.2)[6]. These methods were applied to all tubes and individually per tube. Furthermore, metadata, including gender, ELN risk classification (excluded when predicting ELN risk), age, age squared, white blood cell count, peripheral blood blast percentage and bone marrow blast percentage (all determined at the time of diagnosis) were added to the selected features. Finally, an extreme gradient boosting (XGBoost) model was trained on the complete dataset minus one patient and validated on the remaining patient using the selected features or all the features.

XGBoost[7] hyperparameters were tuned beforehand. Hypertuning per fold (leave-one-out cross-validation) would be infeasible due to an exponential increase in computational runtime. A randomly collected 0.8% of the training data was used to grow trees to reduce overfitting. The learning rate was controlled by eta which we set to 0.05. This also prevents overfitting by making the boosting process more conservative.

2. Figures

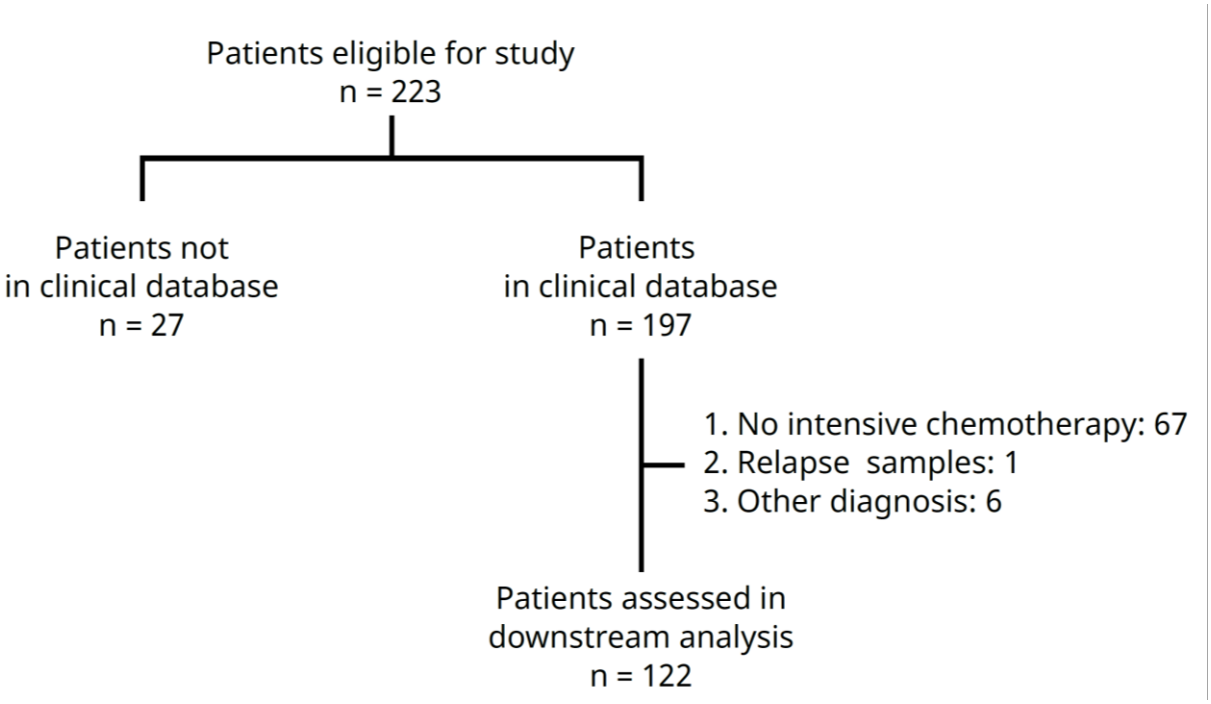


Figure S1: Patient flowchart. A cohort of 223 patients diagnosed with AML was collected at the Ghent University Hospital between 2015 and 2019. Patients were excluded for reasons indicated in the figure, leading to a final cohort of 122 patients assessed in downstream analysis.

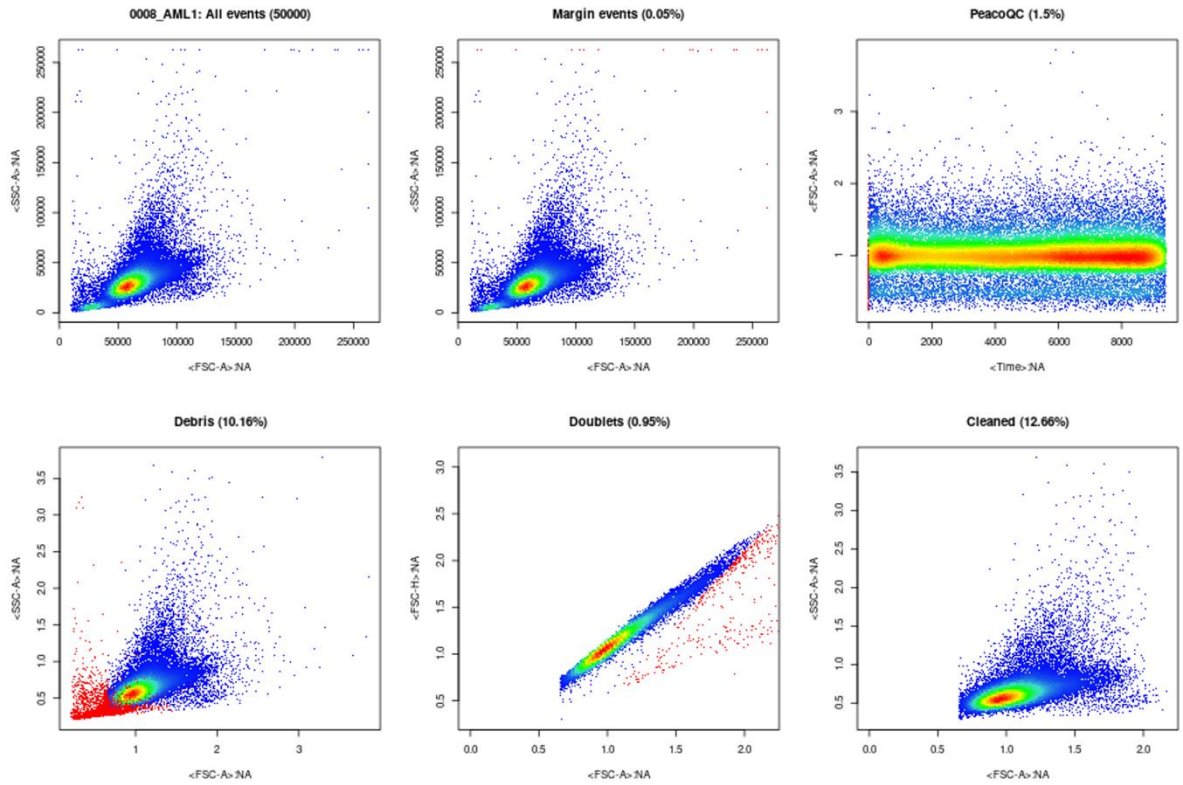


Figure S2: 2D scatter plots of preprocessing steps on complete file.

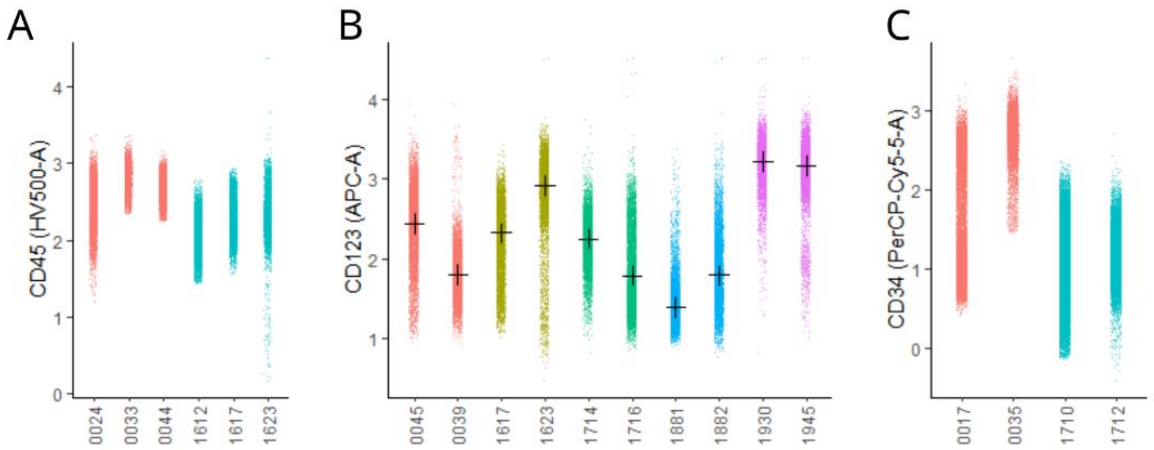


Figure S3: File scatters per batch effect. A: represents a batch effect between before (red) and after (blue) the fluorochrome change from PacO to HV500. **B:** represents a batch effect in tube 6 between 2019 (purple) and the other acquisition years. **C:** represents a batch effect before (red) and after (blue) August 1, 2016 in PerCP-Cy5.5

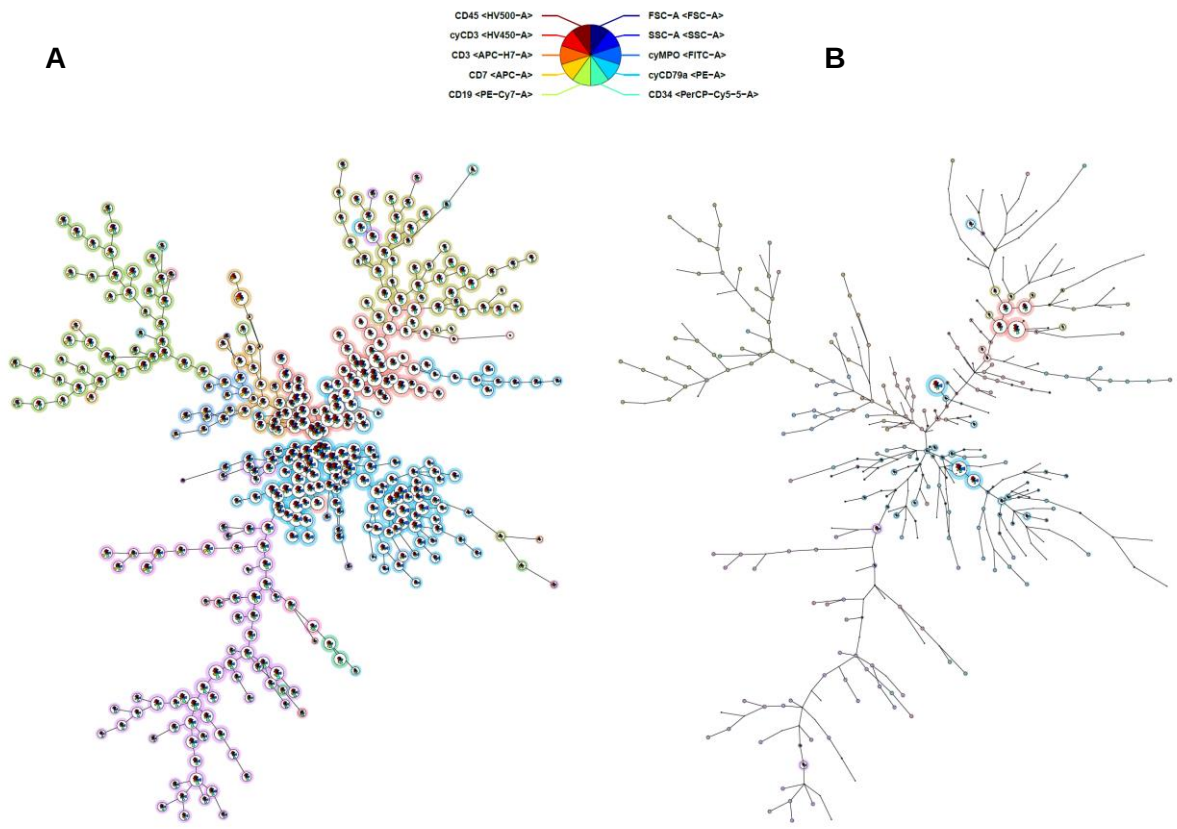


Figure S4: **FlowSOM tree of the ALOT tube.** Background colors indicate metaclusters (i.e. cell populations). **A:** Minimal spanning tree representation of the FlowSOM model built with an aggregate of ALOT tube files from all patients. **B:** Patient 1627 mapped on the aggregate FlowSOM from (A).

Significance
- p < 0.05
- p < 0.01
- p < 0.001
- not significant

Expression
- −
- +

Metacluster percentage
1
0.5
0

Genetic abnormalities
- −
- +

Outcome
- −
- +

relapse HSCT
- no relapse
- primary refractory
- relapse after chemotherapy
- relapse after transplant

ELN risk classification
- favorable
- intermediate
- poor

cytogenetics
- BCR−ABL t(9;22)
- complex karyotype
- del5, del5q
- del7, del7q
- inv(16), t(16;16)
- MLL = t(9;11) (p21−22;q23)
- monosomal karyotype
- normal karyotype
- other
- PML−RARA = t(15;17)
- RPN1−EVI1 = inv(3), t(3;3)
- t(8;21) = RUNX1−RUNXT (AML1−ETO
- t(v;11q23) MLL rearrang, excl t(9;11)

WHO classification
- ambiguous lineage
- NOS
- therapy−related
- with MDS−related changes
- with recurrent genetic abnormalities

genetic abnormalities
- BCR−ABL1
- inv(16)
- inv(3)
- mutated NPM1
- mutated RUNX1
- t(15;17)
- t(8;21)

Figure S5: Full heatmap of metacluster percentages displays distinct expression patterns in WHO 2016 subclasses. Patients are grouped based on genetic abnormalities according to the WHO 2016 guidelines, and based on expression patterns within "*NPM1*" and "Other" groups. Patient clinical data, including presence/absence of mutations and translocations, and immunophenotype of cell populations (presence "+" or absence "-" of markers present in the specific tube of the EuroFlow AML panel) are visualized left of and below the cell population percentages, respectively. Significance for differential expression of features between groups of interest (i.e. relapse after chemotherapy, relapse after transplant, survival status at two years post-diagnosis, ELN favorable vs poor, $NPM1^{mut}$ *vs.* $NPM1^{WT}$ and inv[16] vs no inv[16]) was calculated with a Wilcoxon signed-rank test and is represented at the bottom in pink. Grey color indicates absence of information for outcomes, genetic abnormalities and cell population percentages. In outcome columns, a plus stands for alive after two years or received a transplant while white stands for deceased after two years or did not receive a transplant. For relapse, it is indicated whether patients were primary refractory, relapsed after chemotherapy or relapsed after receiving a transplant.

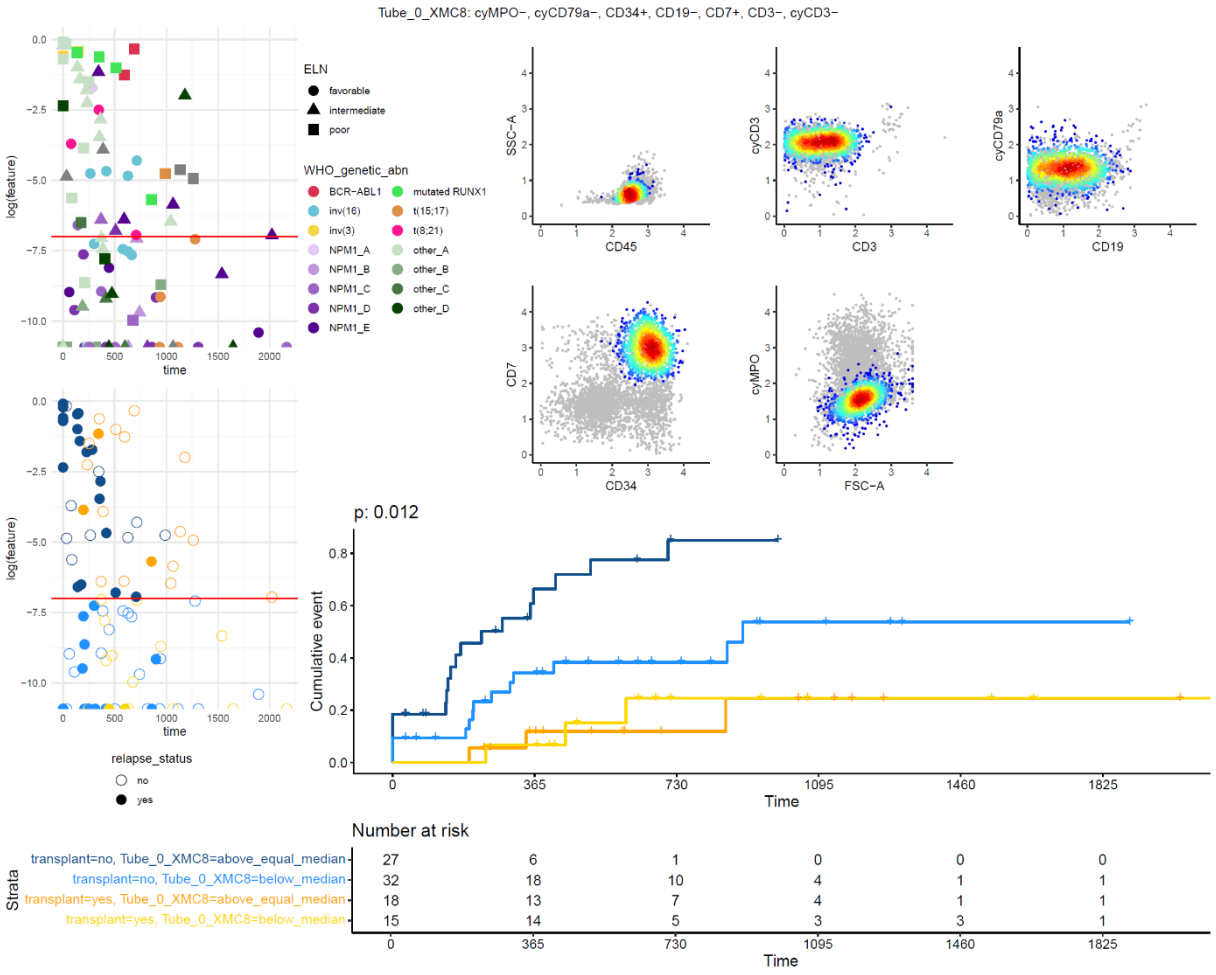


Figure S6: Representative feature (metacluster 8 of tube 0) for CD34+ cyMPO-. Patients with a percentage of this leukemic cell population above the median had a higher risk for relapse compared to patients with a below-median percentage. Receiving an HSCT lowers the risk of relapse.

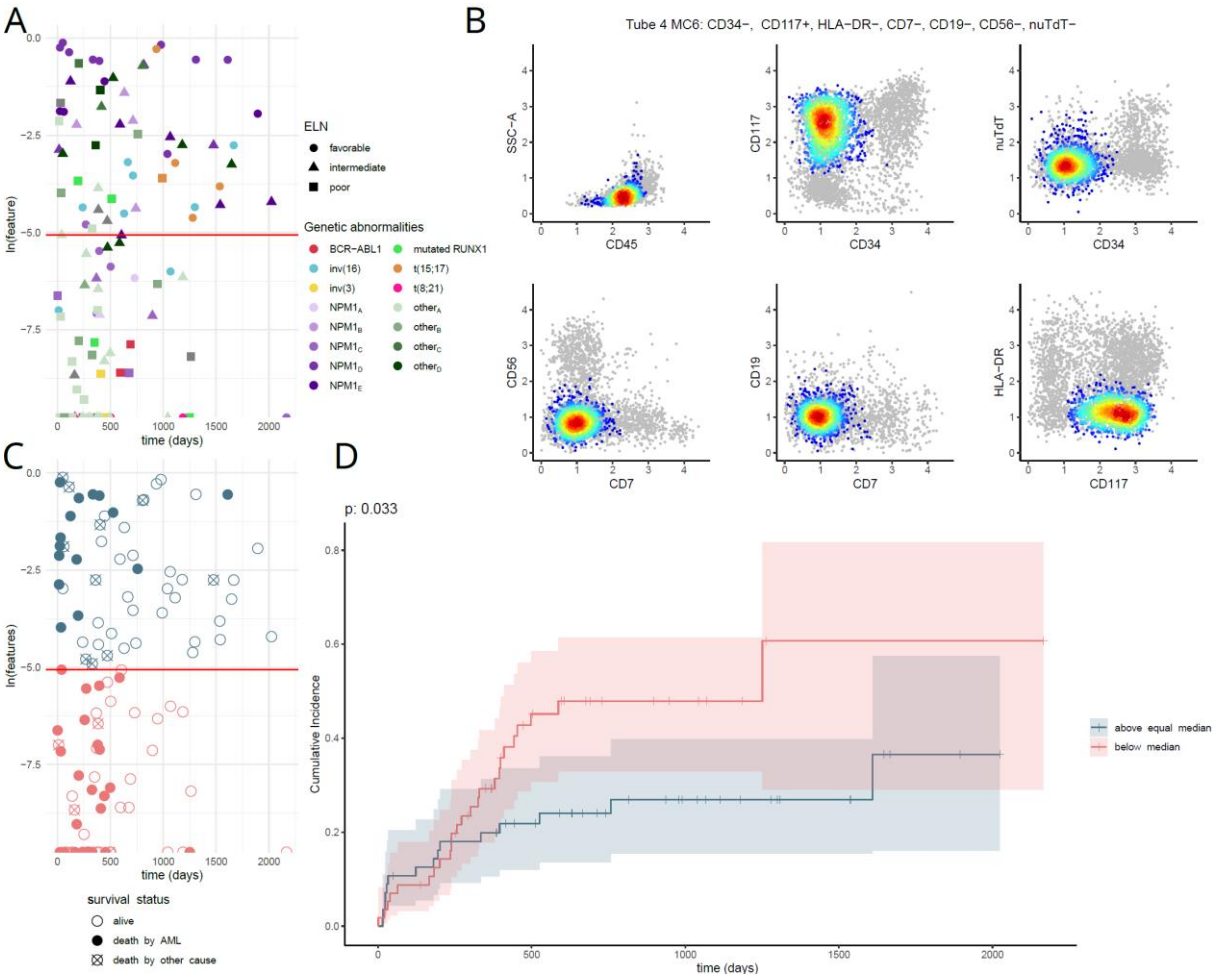


Figure S7: Patients with an above-median expression level of CD34- CD117+ HLA-DR- leukemic cells have a lower risk of AML-related mortality. Detailed visualizations of feature values, time to AML-related mortality and genetic abnormalities for a representative feature (metacluster 8 of tube 4)**. A:** Y-axis shows cell population percentages for the specific cell population investigated, represented as log(feature) values. Every point indicates the value for a specific patient, with icons representing ELN risk category and colors indicating the "genetic abnormalities" group as in Figure 2. -inf values are scaled to 0. X-axis shows the survival and/or follow-up time. Horizontal red line indicates the median value for this feature for all patients. **B:** 2D scatter plots for metacluster 8 of tube 4 shows a CD34- CD117+ HLA-DR- CD7- CD19- CD56- nuTdT- phenotype for this cell population. **C:** Plot displaying the survival status (symbols as indicated in legend) at last contact, survival time (x-axis) and feature value (y-axis) of each patient. **D:** Cumulative incidence plot showing results of the non-parametric estimate of the cumulative incidence test with non-AML related mortality as a competing risk. Patients with above/equal to median expression of the investigated cell population are indicated by the blue line, whereas patients with below median expression are shown by the red line. The 95% confidence interval is shown as ribbons.

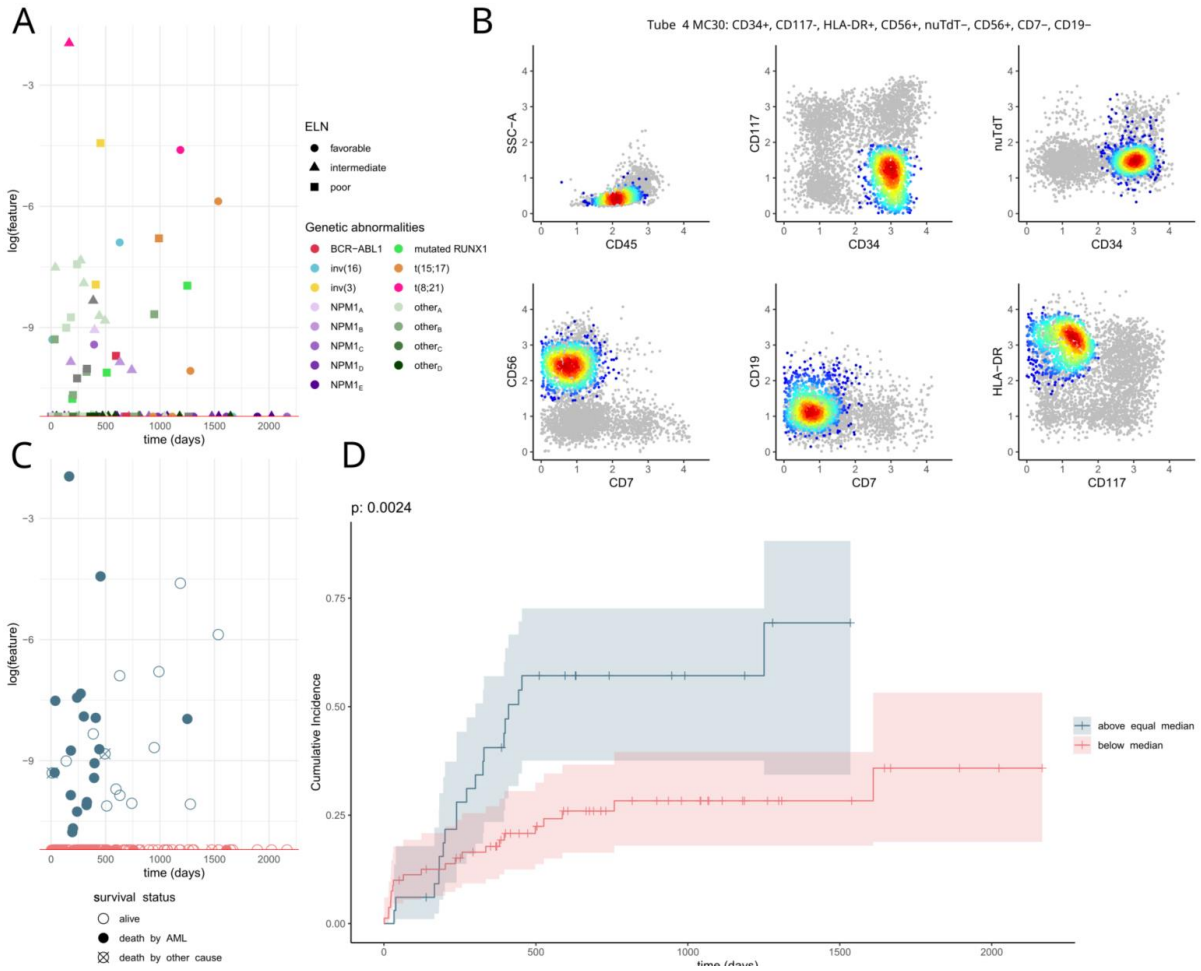
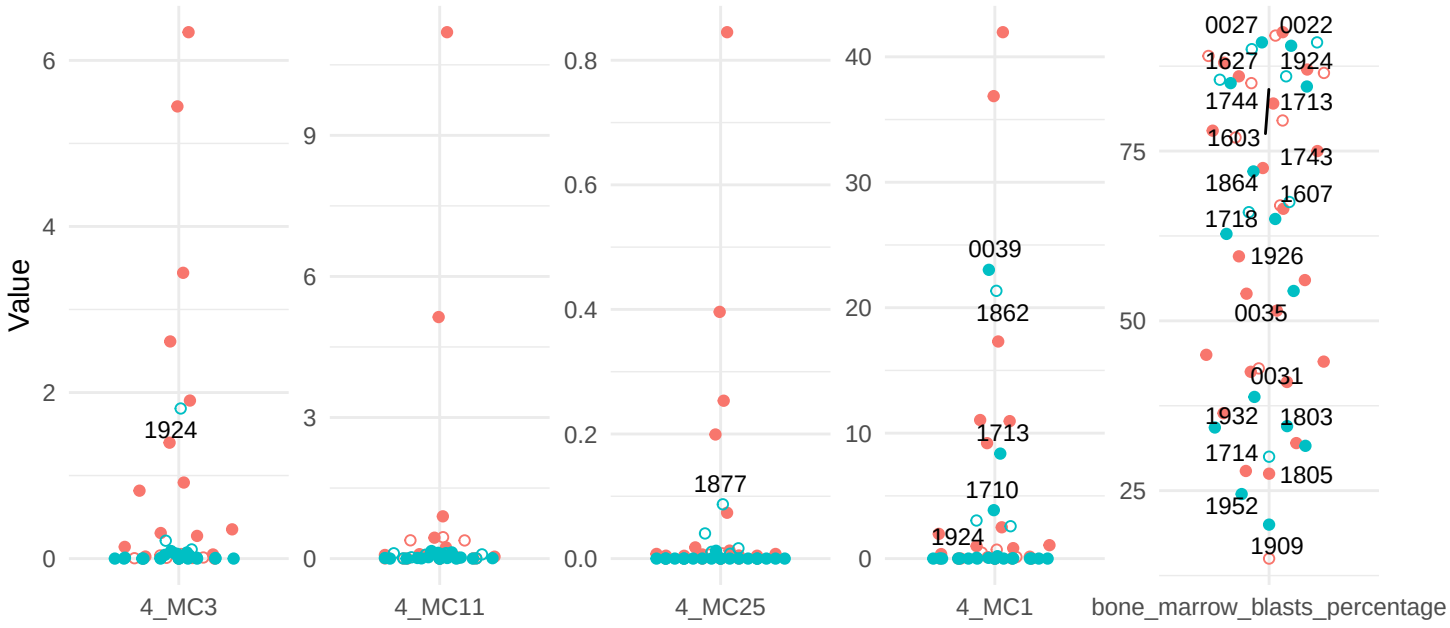


Figure S8: Patients with an above-median expression level of CD34+ CD117- leukemic cells succumb to AML-related mortality more quickly. Detailed visualizations of feature values, time to AML-related mortality and genetic abnormalities for a representative feature (metacluster 30 of tube 4). Detailed visualizations of feature values, time to AML-related mortality and genetic abnormalities for a representative feature**. A:** Y-axis shows cell population percentages for the specific cell population investigated, represented as log(feature) values. Every point indicates the value for a specific patient, with icons representing ELN risk category and colors indicating the "genetic abnormalities" group as in Figure 2. -inf values are scaled to 0. X-axis shows the survival and/or follow-up time. Horizontal red line indicates the median value for this feature for all patients. **B:** 2D scatter plots for metacluster 30 of tube 4 shows a CD34+ CD117- HLA-DR+ CD7- CD19- CD56+ nuTdT- phenotype for this cell population. **C:** Plot displaying the survival status (symbols as indicated in legend) at last contact, survival time (x-axis) and feature value (y-axis) of each patient. **D:** Cumulative incidence plot showing results of the non-parametric estimate of the cumulative incidence test with non-AML related mortality as a competing risk. Patients with above/equal to median expression of the investigated cell population are indicated by the blue line, whereas patients with below median expression are shown by the red line. The 95% confidence interval is shown as ribbons.

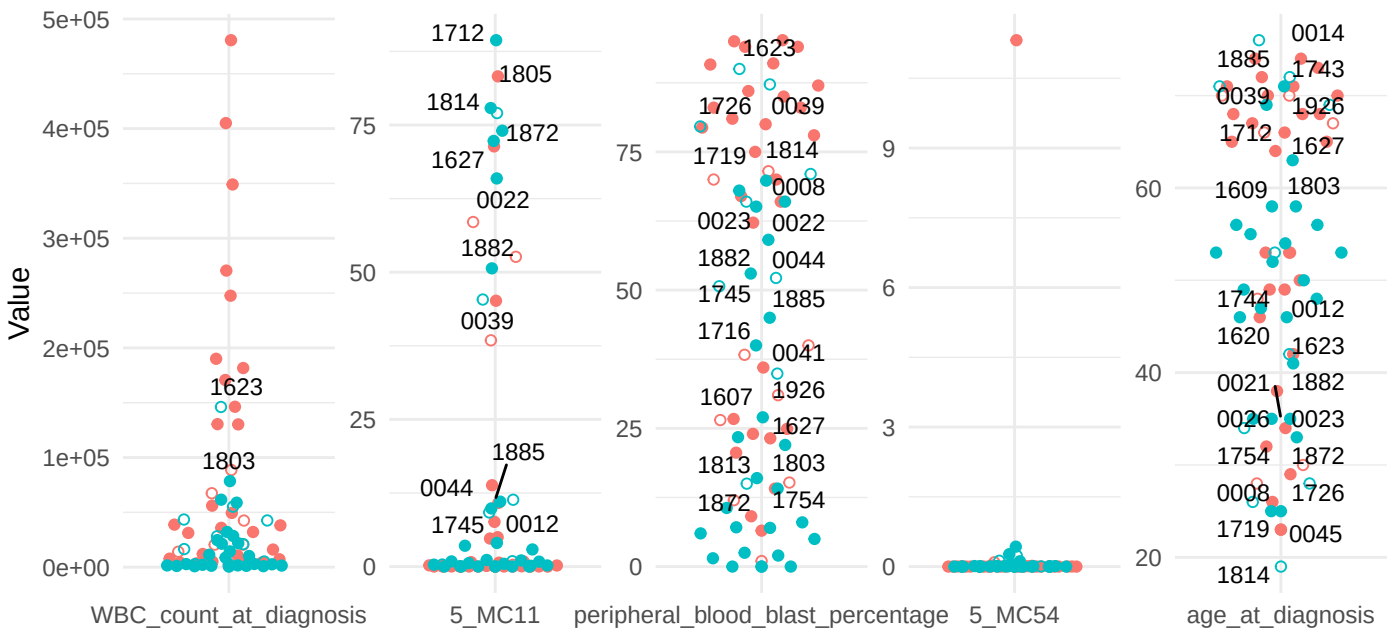
relapse vs noRelapse noTransplant

n = 25 vs n = 31



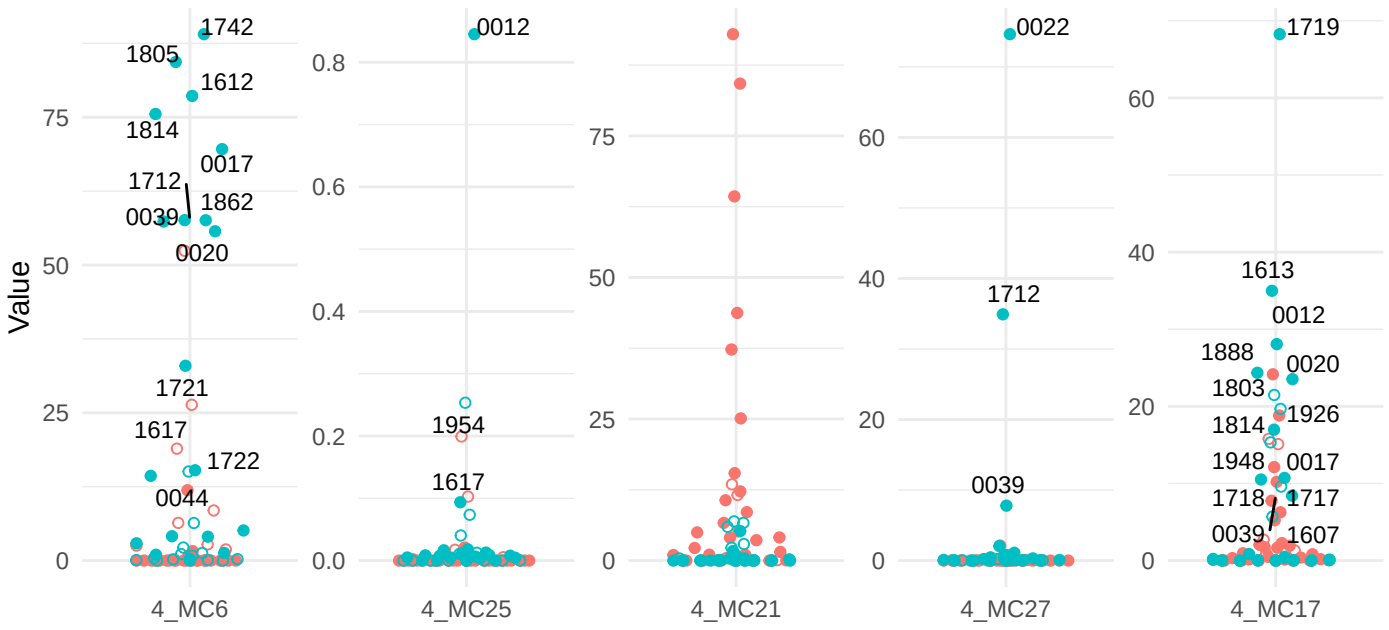
A2 vs D2

n = 35 vs n = 38



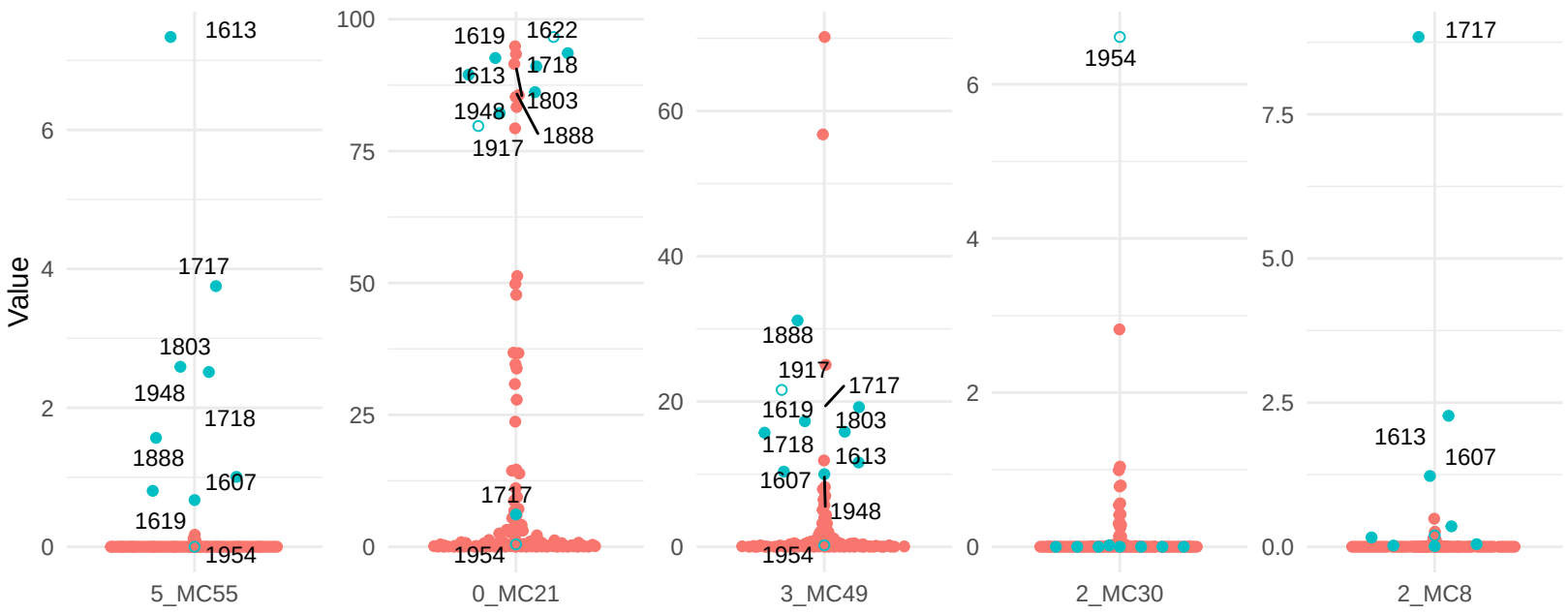
favorable vs poor

n = 39 vs n = 39



inv16 vs noInv16

n = 11 vs n = 111



NPM1 vs noNPM1

n = 42 vs n = 77

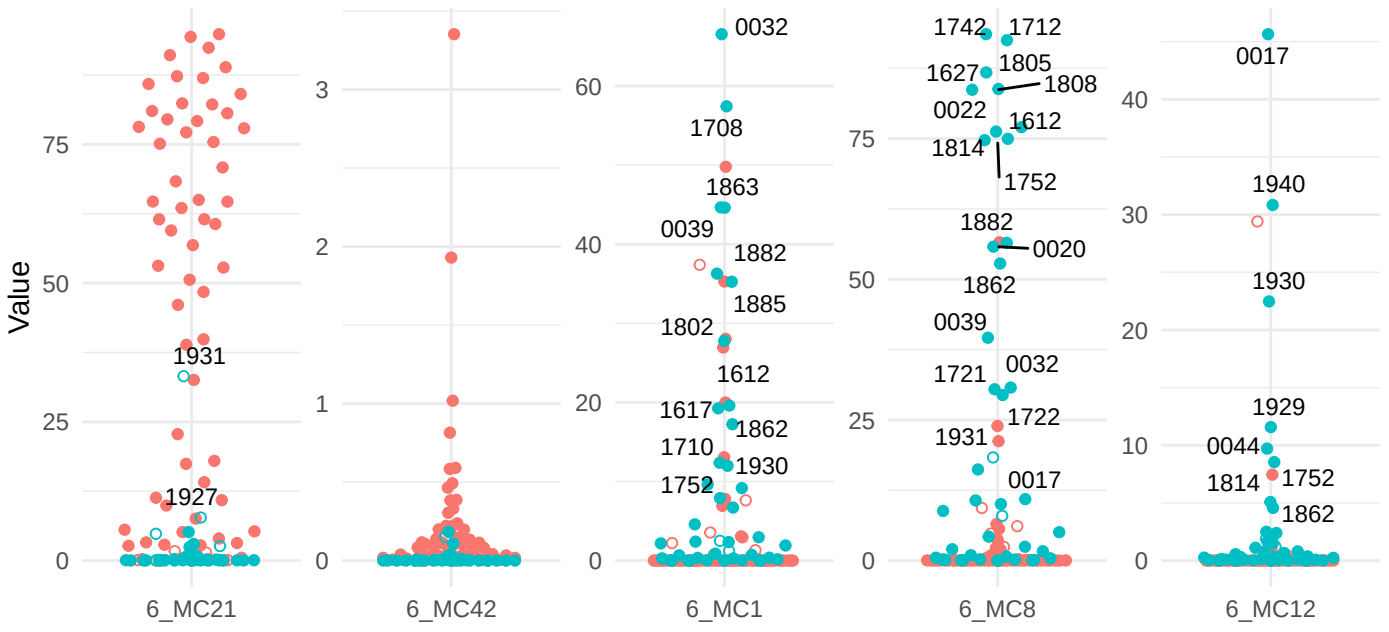


Figure S9: Feature values of top 5 most important cell populations determined with SHAP analysis of the ML models. Patients are colored according to their group (based on outcome, ELN or mutational status), as indicated in the legend. A filled circle means the patients is predicted correctly by the ML model, whereas a hollow circle means the prediction of the ML model is wrong. Patient IDs are indicated.

3. Tables

Table S1: Overview of patient characteristics and numbers

Patients included in study	
n	122
Age at diagnosis, years (median)	18 – 76 (54)
Sex, n (%)	
male	68 (55.7)
female	54 (44.3)
WBC count at diagnosis, range (median)	540 – 480650 (16290)
bone marrow blast percentage at diagnosis, % (median)	15.0 – 96.5 (61.5)
peripheral blood blast percentage at diagnosis, % (median)	0 – 95.2 (36.0)
ELN risk classification at diagnosis, n (%)	
favorable	39 (32.0)
intermediate	44 (36.0)
poor	39 (32.0)
WHO classification at diagnosis, n (%)	
ambiguous lineage	2 (1.6)
NOS	27 (22.1)
therapy-related	11 (9.0)
MDS-related changes	17 (13.9)
with recurrent genetic abnormalities	65 (53.3)
genetic abnormalities, n (%)	
BCR-ABL1	1 (0.8)
inv(16)	10 (8.2)
inv(3)	2 (1.6)
mutated NPM1	36 (29.5)
mutated RUNX1	5 (4.1)
t(15;17)	5 (4.1)
t(8;21)	5 (4.1)
NA	58 (47.5)

treatment received at diagnosis, n (%)	
intensive chemotherapy	122 (100)
patients undergoing HSCT, n (%)	57 (46.7)
follow-up, months (median)	0 – 71 (13)

Table S2:Overview of EuroFlow AML/MDS panel used for our analysis

Tube	HV450	HV500	FITC	PE	PerCP-Cy5.5	PE-Cy7	APC	APC-H7
ALOT	cyCD3	CD45	cyMPO	cyCD79a	CD34	CD19	CD7	CD3
1	HLA-DR	CD45	CD16	CD13	CD34	CD117	CD11b	CD10
2	HLA-DR	CD45	CD35	CD64	CD34	CD117	CD300e	CD14
3	HLA-DR	CD45	CD36	CD105	CD34	CD117	CD33	CD71
4	HLA-DR	CD45	nuTdT	CD56	CD34	CD117	CD7	CD19
5	HLA-DR	CD45	CD15	NG2	CD34	CD117	CD22	CD38
6	HLA-DR	CD45	CD42a + CD61	CD203c	CD34	CD117	CD123	CD4
7	HLA-DR	CD45	CD41	CD25	CD34	CD117	CD42a	CD9

Table S3: Overview of batch effects present in the data.

Tube	Marker (Channel)	Reason
ALOT	CD45 (HV500-A)	Switch from PacB/PacO to HV450/HV500
	CD7(APC-A)	Switch from PacB/PacO to HV450/HV500
1	CD45 (HV500-A)	Switch from PacB/PacO to HV450/HV500
2	CD45 (HV500-A)	Switch from PacB/PacO to HV450/HV500
	CD64 (PE-A)	Batch effect before 1 Aug 2016
	CD34 (PerCP-Cy5-5-A)	Batch effect before 1 Aug 2016
	CD117 (PE-Cy7-A)	Batch effect before 1 Aug 2016
3	CD45 (HV500-A)	Switch from PacB/PacO to HV450/HV500
4	CD45 (HV500-A)	Switch from PacB/PacO to HV450/HV500
5	CD45 (HV500-A)	Switch from PacB/PacO to HV450/HV500
6	CD45 (HV500-A)	Switch from PacB/PacO to HV450/HV500
	CD123 (APC-A)	Switch from PacB/PacO to HV450/HV500 and batch effect between 2019 and other years
	CD4 (APC-H7-A)	Batch effect between 2019 and other years

Table S4: Cell populations present in patients with shorter time to relapse (both patients with and without HSCT) if above/equal to median expression. p-values are calculated using a Cox proportional-hazards test.

Tube and metacluster	Immunophenotype	p-value
tube 3 MC45	CD34+ CD117+ HLA-DR+ CD36+ CD71+ CD105- CD33-	0.00027
tube 5 MC47	CD34+ CD117+ HLA-DR+ CD38+ CD22- CD15+ NG2-	0.0012
tube 4 MC18	CD34+ CD117+ HLA-DR+ CD7- CD56- CD19- nuTdT-	0.0024
tube 6 MC24	CD34+ CD117+ HLA-DR+ CD123+ CD4- CD42a/CD61- CD203c-	0.0031
tube 2 MC12	CD34+ CD117+ HLA-DR+ CD14- CD64- CD35- CD300e-	0.0036
tube 4 MC38	CD34+ CD117- HLA-DR+ CD7+ CD56- CD19- nuTdT+	0.005
tube 5 MC25	CD34+ CD117+ HLA-DR+ CD38+ CD22- CD15- NG2-	0.0053
tube 5 MC23	CD34+ CD117+ HLA-DR+ CD38- CD22- CD15- NG2-	0.0074
tube 6 MC25	CD34+ CD117+ HLA-DR+ CD123+ CD4+ CD42a/CD61- CD203c-	0.0077
tube 4 MC21	CD34+ CD117+ HLA-DR+ CD7+ CD56- CD19- nuTdT-	0.0087
tube 5 MC45	CD34+ CD117+ HLA-DR+ CD38- CD22- CD15+ NG2-	0.0092
tube 5 MC28	CD34+ CD117+ HLA-DR+ CD38+ CD22+ CD15- NG2-	0.01
tube 1 MC29	CD34+ CD117+ HLA-DR+ CD13+ CD11b- CD16- CD10-	0.013
tube 2 MC13	CD34+ CD117+ HLA-DR+ CD14+ CD64- CD35- CD300e-	0.013
tube 1 MC12	CD34+ CD117+ HLA-DR+ CD13- CD11b- CD16- CD10-	0.016
tube 4 MC42	CD34+ CD117+ HLA-DR+ CD7+ CD56- CD19- nuTdT+	0.019
tube 5 MC46	CD34+ CD117+ HLA-DR- CD38+ CD22- CD15+ NG2-	0.019
tube 3 MC25	CD34+ CD117+ HLA-DR+ CD36- CD71- CD105- CD33+	0.022
tube 4 MC19	CD34+ CD117+ HLA-DR+ CD7- CD56- CD19+ nuTdT-	0.023
tube 3 MC27	CD34+ CD117+ HLA-DR+ CD36- CD71+ CD105- CD33+	0.025
tube 4 MC13	CD34+ CD117- HLA-DR+ CD7- CD56- CD19- nuTdT-	0.034
tube 5 MC43	CD34+ CD117- HLA-DR+ CD38+ CD22- CD15- NG2-	0.037
tube 4 MC40	CD34+ CD117+ HLA-DR+ CD7- CD56- CD19- nuTdT+	0.041

Table S5: Cell populations present in patients who succumbed to AML-related mortality more quickly if above/equal to median expression. p-values are calculated using a non-parametric estimate of the cumulative incidence test.

Tube and metacluster	Immunophenotype	p-value
tube 4 MC30	CD34+ CD117- HLA-DR+ nuTdT- CD56+ CD7- CD19-	0.0024
tube 6 MC28	CD34+ CD117- HLA-DR+ CD123- CD4- CD203c+ CD42a+CD61-	0.0091
tube 4 MC12	CD34+ CD117- HLA-DR- nuTdT- CD56- CD7- CD19-	0.018
tube 1 MC9	CD34+ CD117- HLA-DR- CD10- CD11b- CD13- CD16-	0.026
tube 4 MC38	CD34+ CD117- HLA-DR+ nuTdT+ CD56- CD7+ CD19-	0.038

4. References

1 van Dongen JJM, Lhermitte L, Böttcher S, Almeida J, van der Velden VHJ, Flores-Montero J *et al.* EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia* 2012; **26**: 1908–1975.

2 EuroFlow Protocols. https://euroflow.org/protocols/ (accessed 31 Jul2023).

3 Van Gassen S, Gaudilliere B, Angst MS, Saeys Y, Aghaeepour N. CytoNorm: A Normalization Algorithm for Cytometry Data. *Cytometry A* 2019; : cyto.a.23904.

4 Emmaneel A, Quintelier K, Sichien D, Rybakowska P, Marañón C, Alarcón-Riquelme ME *et al.* PeacoQC: Peak-Based Selection of High Quality Cytometry Data. *Cytometry A* 2021; **n/a**. doi:10.1002/cyto.a.24501.

5 Malek M, Taghiyar MJ, Chong L, Finak G, Gottardo R, Brinkman RR. flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. *Bioinformatics* 2015; **31**: 606–607.

6 Van Gassen S, Vens C, Dhaene T, Lambrecht BN, Saeys Y. FloReMi: Flow density survival regression using minimal feature redundancy: FloReMi. *Cytometry A* 2016; **89**: 22–29.

7 Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H *et al. xgboost: Extreme Gradient Boosting*. 2023https://CRAN.R-project.org/package=xgboost.