LETTER TO THE EDITOR

A machine learning approach for the rapid identification of measurable residual disease in acute myeloid leukemia

Authors

Amirali Vahedi,[1,2] Mohammadreza Royaei,[3] Tahereh Madani,[3] François Mercier[1,2#] and Behzad Poopak[3,4#]

[1]Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, Canada; [2]Lady Davis Institute for Medical Research, Montreal, Canada; [3]Payvand Clinical, Specialty, Pathology, Medical Genetics and Molecular Laboratory, Tehran, Iran and [4]Hematology Department, School of Paramedics, Islamic Azad University, Tehran Medical Sciences, Tehran, Iran

[#]*FM and BP contributed equally as senior authors.*

Correspondence:
F. MERCIER - francois.mercier@mcgill.ca
B, POOPAK - bpoopak@gmail.com

https://doi.org/10.3324/haematol.2024.286019

Received: June 19, 2024.
Accepted: February 24, 2025.
Early view: March 6, 2025.

Table S1. Patient cohorts' response categories.

	Number (%)	%MRD, Median (Range)
Training/validation cohort	132	
Remission*	125 (94.7)	0** (0 - 2.6)
Relapse/Refractory	7 (5.3)	8.1 (5.2 - 26.1)
Retrospective test cohort	30	
Remission	22 (73.3)	0 (0 - 4.5)
Relapse/Refractory	8 (26.7)	12.8 (5.9 - 84.6)
Prospective test cohort	50	
Remission	42 (84.0)	0 (0 - 4.5)
Relapse/Refractory	8 (16.0)	21.4 (5.2 - 72.2)

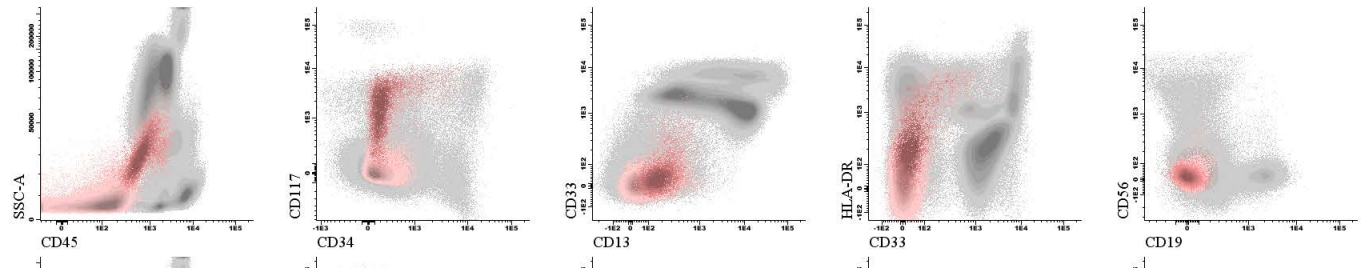
*Defined morphologically as blasts < 5%.
**MRD percentages of "undetectable" patients (patients with no MRD, considering the assay's lower limit of detection) were designated as 0.
MRD, measurable residual disease

Figure S1. Populations gated and annotated by manual analysis with Infinicyt™ software. All normal bone marrow populations including "Normal Myeloid Precursors" (last row, blue) and if present, "Residual Leukemic Cells" (last row, red) were gated and assigned based on their immunophenotypic profile. The designated populations were "Residual Leukemic Cells" (i.e., MRD), "Normal Myeloid Precursors", "Erythroid Cells", "Erythroid Precursors", "B-Cell Precursors", "Mature B-Cells", "Plasma Cells", "T- and NK-Cells", "Promyelocytes", "Myelocytes, Metamyelocytes and Band Cells", "Segmented Neutrophils", "Eosinophils", "Mature Monocytes", "Monoblasts and Promonocytes", "Plasmacytoid Dendritic Cells", "Basophils", and "Mast Cells". CD, cluster of differentiation; MRD, measurable residual disease; NK, natural killer; SSC, side scatter.

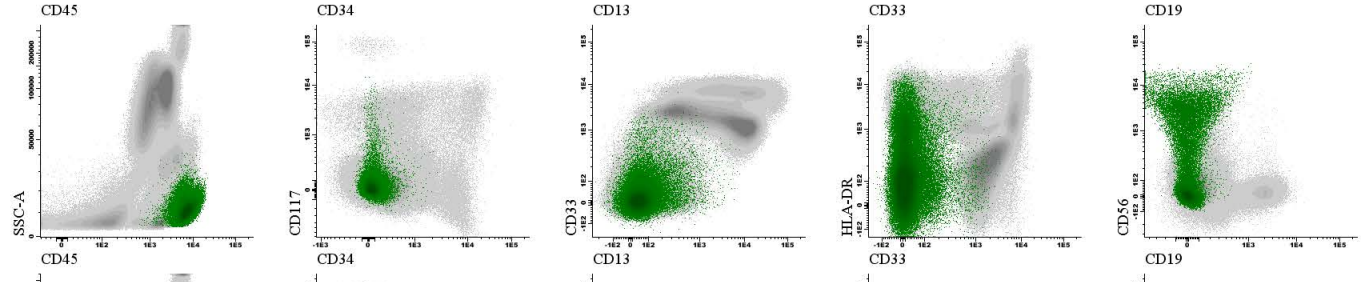
Figure S2. Detailed analyses of model performances. A) Classification report of test sets for 5-fold nested cross-validation comparing the performance of the three models tested; i.e., SVM, LGBM, and RFC. Each dot represents one fold. The boxplot demonstrates the median, quartiles, and spread of the precision, recall, and F1-score of the models. B) Batch effect analysis of the test cohorts (total n = 80). PCA used median signal intensities of all fluorescence and scatter parameters for 4 bone marrow populations including "Mature B Cells", "Mature Monocytes", "Normal Myeloid Precursors", and "T and NK Cells". Each dot represents one patient. The difference between populations is not driven by batch number (i.e., technical confounders). C) Classification report of 10 test cases comparing RFC prediction with manual analysis. Five cases from the retrospective test cohort (4 MRD-positive and 1 MRD-negative) and 5 from the prospective test cohort (3 MRD-positive and 2 MRD-negative) were randomly selected to report the classification performance. The boxplot demonstrates the median, quartiles, and spread of the recall, F1-score, and precision of the model for all predicted classes. D) The Spearman's correlation analysis of all populations between manual analysis and RFC in all cases in retrospective (n = 30) and prospective (n = 50) test cohorts. The light line shows perfect correlation. The model demonstrated good correlation with manual analysis in most of the classes in both retrospective and prospective test cohorts including "All Myeloid Precursors" (normal and abnormal), "T- and NK-Cells", and "Erythroid Cells". For the simplicity of analysis, the sum of "Myelocytes, Metamyelocytes and Band Cells", "Segmented Neutrophils", and "Eosinophils", and the sum of "Mature Monocytes"

and “Monoblasts and Promonocytes” were designated as “Neutrophil Series and Eosinophils” and “Monocyte Series”, respectively. Performance metrics including accuracy, precision, recall, and F1-score were calculated with scikit-learn package. LGBM, light gradient-boosting machine; MRD, measurable residual disease; NK, natural killer; PC, principal component; PCA, principal component analysis; RFC, random forest classifier; SVM, support vector machine.

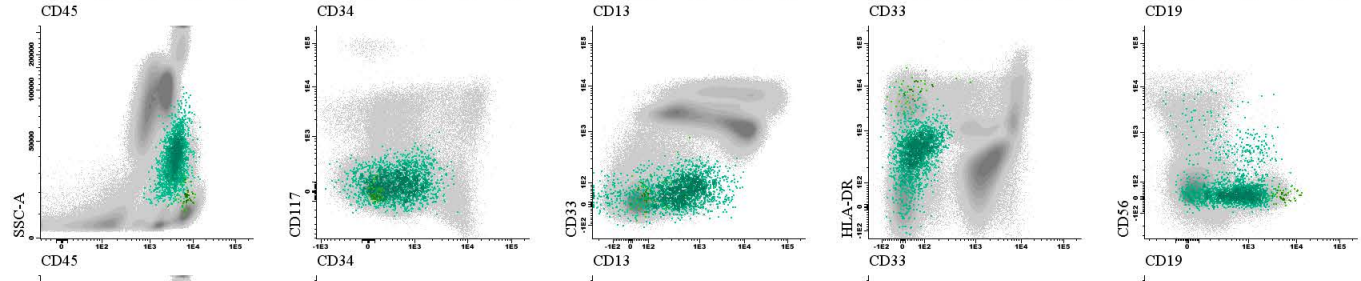
Erythroid cells and precursors



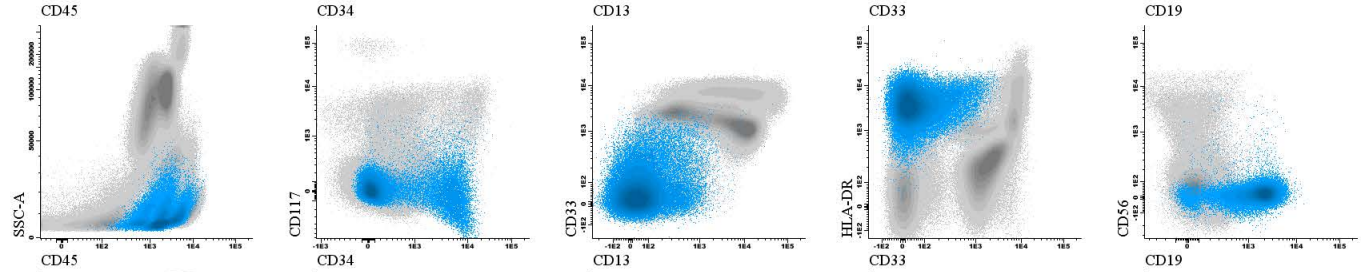
T- and NK-cells



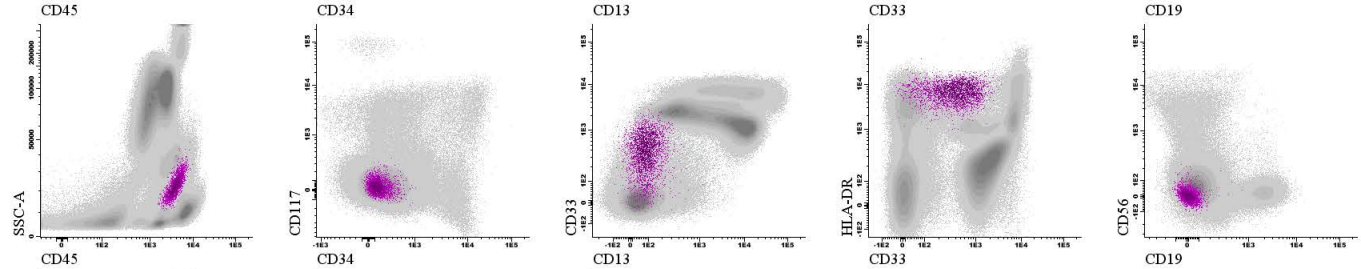
Mature B-cells and plasma cells



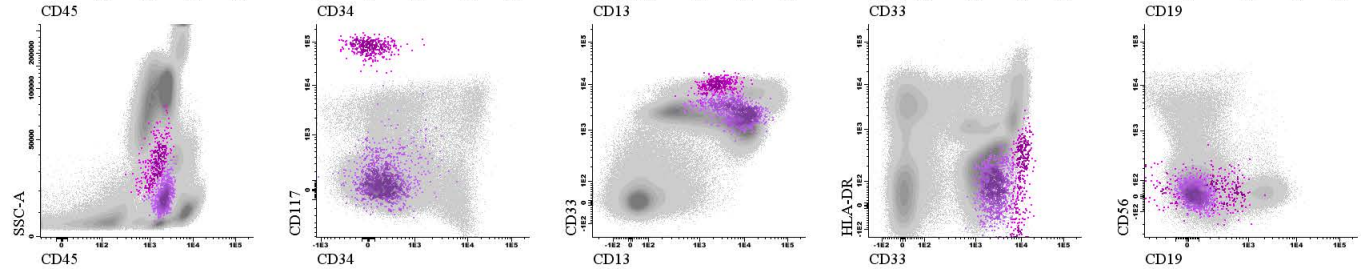
B-cell precursors



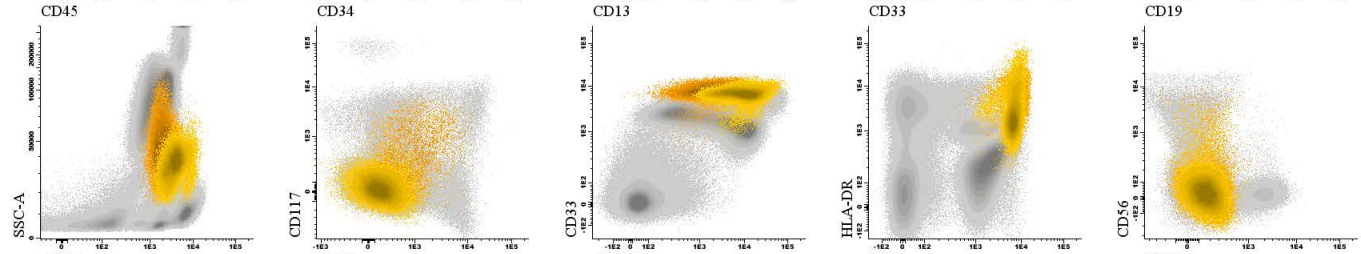
Plasmacytoid dendritic cells



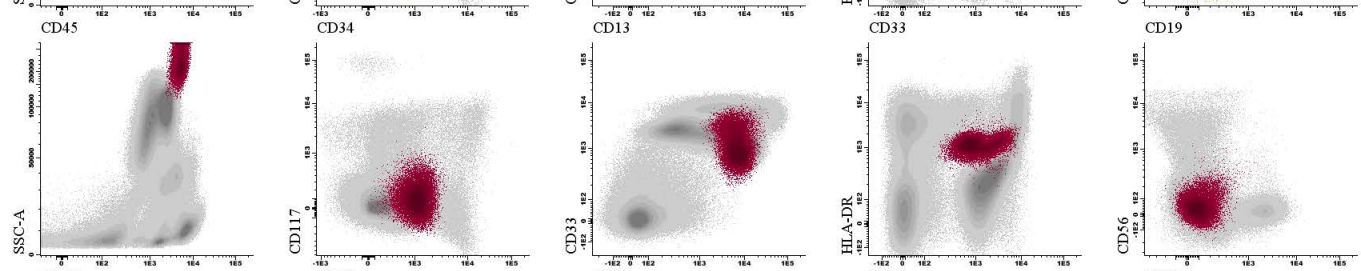
Basophils and mast cells



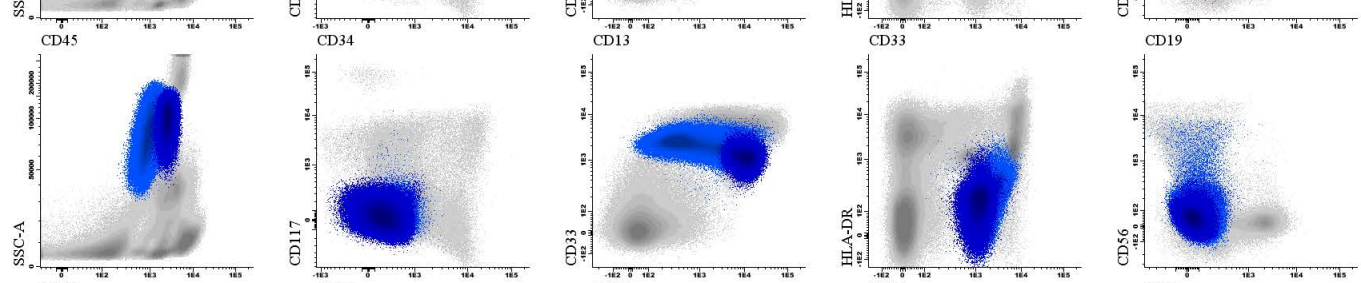
Monoblasts, promonocytes and mature monocytes



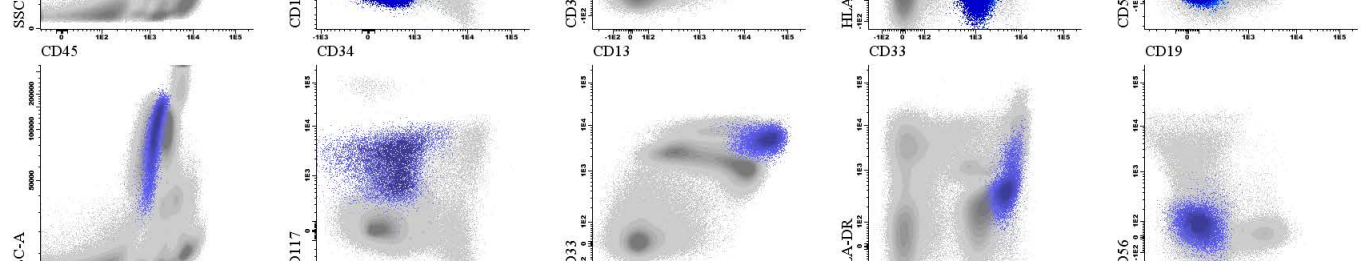
Eosinophils



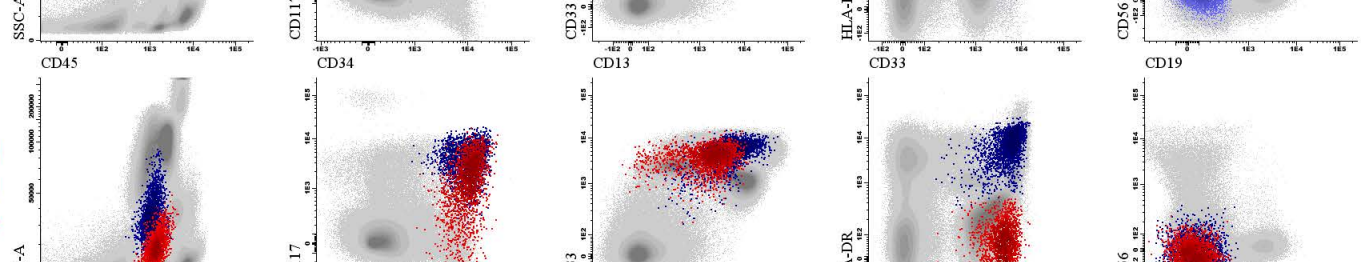
Myelocytes, metamyelocytes, band cells, and segmented neutrophils

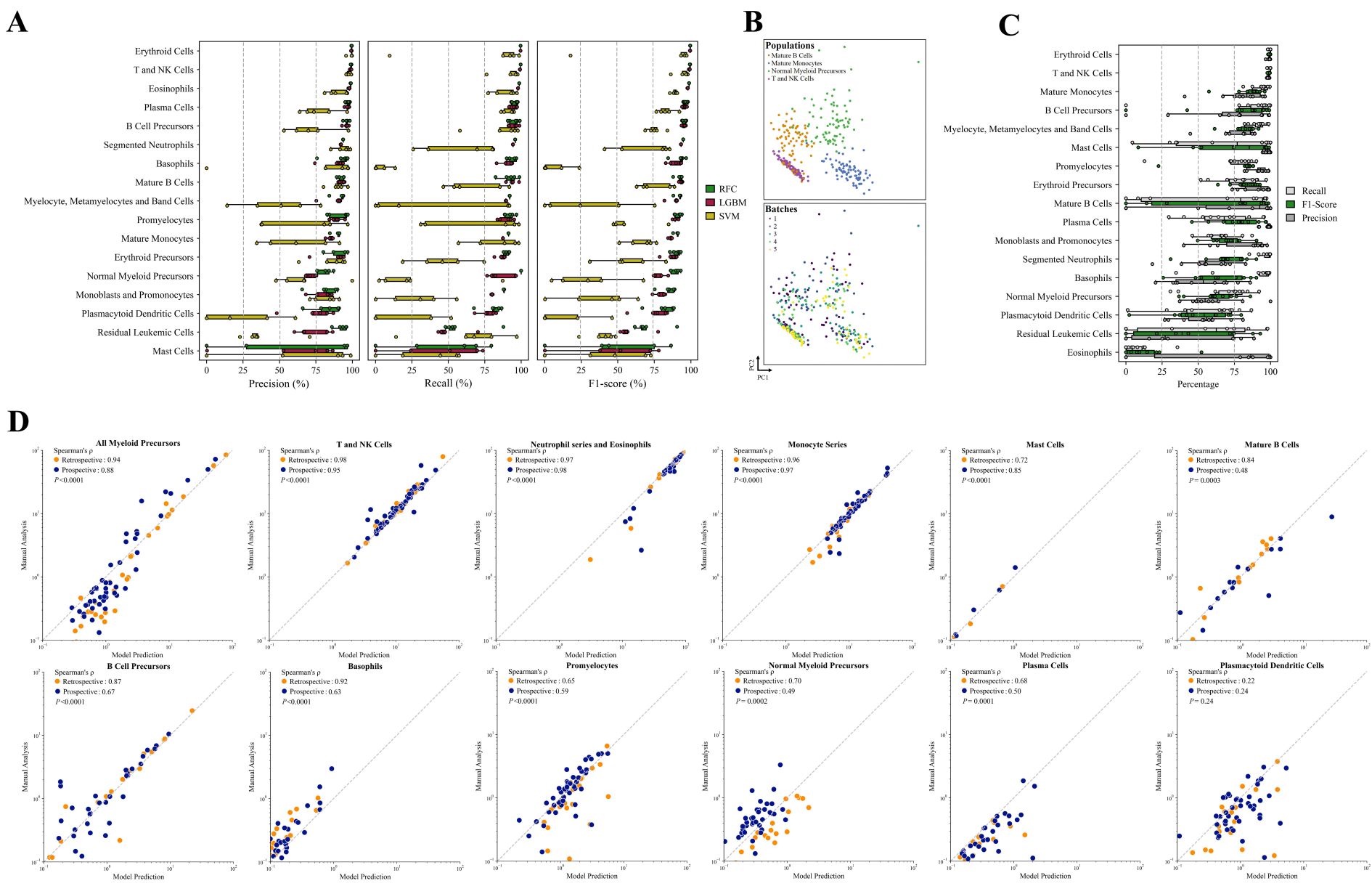


Promyelocytes



Normal myeloid precursors and residual leukemic cells




A
Erythroid Cells
T and NK Cells
Eosinophils
Plasma Cells
B Cell Precursors
Segmented Neutrophils
Basophils
Mature B Cells
Myelocyte, Metamyelocytes and Band Cells
Promyelocytes
Mature Monocytes
Erythroid Precursors
Normal Myeloid Precursors
Monoblasts and Promonocytes
Plasmacytoid Dendritic Cells
Residual Leukemic Cells
Mast Cells
0 25 50 75 100
Precision (%)
Recall (%)
F1-score (%)
RFC
LGBM
SVM
B
Populations
Mature B Cells
Mature Monocytes
Normal Myeloid Precursors
T and NK Cells
Batches
1
2
3
4
5
PC2
PC1
C
Erythroid Cells
T and NK Cells
Mature Monocytes
B Cell Precursors
Myelocyte, Metamyelocytes and Band Cells
Mast Cells
Promyelocytes
Erythroid Precursors
Mature B Cells
Plasma Cells
Monoblasts and Promonocytes
Segmented Neutrophils
Basophils
Normal Myeloid Precursors
Plasmacytoid Dendritic Cells
Residual Leukemic Cells
Eosinophils
0 25 50 75 100
Percentage
Recall
F1-Score
Precision
D
All Myeloid Precursors
Spearman's ρ
Retrospective : 0.94
Prospective : 0.88
P <0.0001
T and NK Cells
Retrospective : 0.98
Prospective : 0.95
P <0.0001
Neutrophil series and Eosinophils
Retrospective : 0.97
Prospective : 0.98
P <0.0001
Monocyte Series
Retrospective : 0.96
Prospective : 0.97
P <0.0001
Mast Cells
Retrospective : 0.72
Prospective : 0.85
P <0.0001
Mature B Cells
Retrospective : 0.84
Prospective : 0.48
P = 0.0003
B Cell Precursors
Retrospective : 0.87
Prospective : 0.67
P <0.0001
Basophils
Retrospective : 0.92
Prospective : 0.63
P <0.0001
Promyelocytes
Retrospective : 0.65
Prospective : 0.59
P <0.0001
Normal Myeloid Precursors
Retrospective : 0.70
Prospective : 0.49
P = 0.0002
Plasma Cells
Retrospective : 0.68
Prospective : 0.50
P = 0.0001
Plasmacytoid Dendritic Cells
Retrospective : 0.22
Prospective : 0.24
P = 0.24
Manual Analysis
Model Prediction