Received: 11 January 2021 | Revised: 6 April 2021 | Accepted: 26 April 2021

DOI: 10.1002/cyto.a.24360

ORIGINAL ARTICLE

CYTOMETRY Journal of Quantitative Cell Science PART A

Computational flow cytometry as a diagnostic tool in suspected-myelodysplastic syndromes

Carolien Duetz[1] | Sofie Van Gassen[2,3] | Theresia M. Westers[1] | Margot F. van Spronsen[1] | Costa Bachas[1] | Yvan Saeys[2,3] | Arjan A. van de Loosdrecht[1]

[1]Department of Hematology, Amsterdam UMC, VU University Medical Center, Cancer Center Amsterdam, Amsterdam, Netherlands

[2]VIB Inflammation Research Center, Ghent University, Ghent, Belgium

[3]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

Correspondence
Arjan A. van de Loosdrecht, Department of Hematology, Amsterdam UMC, VU University Medical Center, Cancer Center Amsterdam, Amsterdam, Netherlands.
Email: a.vandeloosdrecht@amsterdamumc.nl

Funding information
European Union's Horizon 2020 research and innovation programme, Grant/Award Number: 634789

Abstract

The diagnostic work-up of patients suspected for myelodysplastic syndromes is challenging and mainly relies on bone marrow morphology and cytogenetics. In this study, we developed and prospectively validated a fully computational tool for flow cytometry diagnostics in suspected-MDS. The computational diagnostic workflow consists of methods for pre-processing flow cytometry data, followed by a cell population detection method (FlowSOM) and a machine learning classifier (Random Forest). Based on a six tubes FC panel, the workflow obtained a 90% sensitivity and 93% specificity in an independent validation cohort. For practical advantages (e.g., reduced processing time and costs), a second computational diagnostic workflow was trained, solely based on the best performing single tube of the training cohort. This workflow obtained 97% sensitivity and 95% specificity in the prospective validation cohort. Both workflows outperformed the conventional, expert analyzed flow cytometry scores for diagnosis with respect to accuracy, objectivity and time investment (less than 2 min per patient).

KEYWORDS
diagnostic test, flow cytometry, hematological malignancies, machine learning, myelodysplastic syndromes

1 | INTRODUCTION

Myelodysplastic syndromes (MDS) are a group of malignant hematological disorders characterized by cytopenia, bone marrow dysplasia and a heterogeneous disease course [1]. In the 2016 World Health Organization (WHO) classification of myeloid malignancies, distinguishing MDS from reactive causes of cytopenia and dysplasia is described as "one of the biggest challenges" [2]. Currently, both bone marrow morphology and cytogenetics are mandatory in the diagnostic work-up, however, these measurements do often not suffice for a conclusive diagnosis. Cytopenic patients with normal cytogenetics and borderline dysplastic bone marrow features provide the main challenge, and additional measurements or watch full waiting are frequently required [3, 4].

Immunophenotypic assessment of bone marrow cells by flow cytometry (FC) has shown to be instrumental to discriminate MDS from non-neoplastic cytopenias [5]. MDS bone marrow frequently has an altered cell subset composition and hematopoietic cells express

Yvan Saeys and Arjan A. van de Loosdrecht contributed equally to this work.

aberrant levels of cells surface markers. Several diagnostic FC scores have been developed, that combine assessment of progenitor b-cells and myeloid progenitor percentages with evaluation of aberrant cell surface marker expression on myeloid and erythroid cell populations [6–10]. Despite their proven utility, FC scores may be improved in terms of accuracy, objectivity, and required time-investment.

In recent years, several machine learning based tools for cell population detection in (flow) cytometry data have been developed [11]. These tools use clustering methods to group cells with similar expression patterns and thereby, allow for computational identification of cell populations. One of the best performing methods is the algorithm FlowSOM, that has a high reproducibility and concordance with manual analysis combined with a limited running time [12, 13]. In this study, we developed a computational workflow for MDS diagnosis, combining FlowSOM and a machine learning classifier [13–15]. Thereby, we aimed to improve accuracy, objectivity and ease of use of flow cytometry based MDS diagnostics. In addition, using these data driven unbiased approach, we investigated which cellular properties matter most to diagnose patients suspected for myelodysplastic syndromes.

2 | MATERIALS AND METHODS

2.1 | Patients

We included 230 patients, grouped into the following categories: MDS patients, patients with a non-neoplastic cytopenia and age-matched controls (non-cytopenic, no hematological disease); age-matched controls and patients with a non-neoplastic cytopenia are further referred to as controls. All patients were subjected to a full diagnostic work-up for MDS at the Amsterdam UMC, location VUmc. Patients without clinical follow-up data and insufficient bone-marrow samples were excluded. Diagnosis was established based on clinical characteristics at diagnosis and follow-up, cytomorphology, cytogenetics, pathology and biochemical indicators. In case FC results were used to establish diagnosis, a hemato(patho)logist reevaluated all other diagnostic parameters to assess whether MDS diagnosis could be confirmed without including FC results.

The total cohort consists of cases used to develop the computational workflow, further referred to as the training cohort, and cases used to prospectively validate the computational workflow, further referred to as the validation cohort. To develop a diagnostic workflow focused on the most difficult to diagnose MDS patients, only MDS patients without excess of blasts (EB) were included in the training cohort ($n = 67$). The training cohort consisted of patients ($n = 148$) included between March 2013 and August 2017, the prospective validation cohort consisted of patients ($n = 57$) included between September 2017 and August 2018. We included an additional validation cohort with MDS EB patients ($n = 25$) to evaluate diagnostic performance in this MDS subgroup. Table 1 summarizes the characteristics of the patients in the training and validation cohorts. The study was approved by the Medical Ethics Committee of VU University Medical Center, Amsterdam, the Netherlands.

2.2 | Sample handling

All bone marrow samples were pre-processed and analyzed within 24 h according to European Leukemia Net guidelines [5]. Bone marrow samples were collected in heparin and erythrocyte lysis was performed with ammonium chloride-based lysing solution. The staining panel consisted of six 8-color FC tubes and is outlined in more detail in Table S1. For every tube at least 100.000 leukocyte events were acquired using a FACS CantoII™ (BD Biosciences). FC equipment settings were generated in a highly standardized fashion, following guidelines of Euroflow [16]. Standard operating procedures were followed to assess FC data, including standardized flow cytometer set up, performance checks and generation of compensation settings (for details on these general laboratory procedures we refer to Cloos et al. [17]). Expert analysis of MFC data were performed following ELN guidelines for two diagnostic MDS FC scores, the Ogata score and the integrated flow cytometry score (iFS) [7, 10]. Sample handling and analysis of flow cytometry standard (fcs) files was previously described in more detail by Cremers et al. and Alhan et al. [10, 18].

2.3 | Computational diagnostic workflow: Preprocessing, computational cell population detection and classification for disease status

First, we optimized fcs files for computational analysis in fully automated fashion. High quality events were selected using the R package FlowAI, the FlowAI package evaluates flow rate abnormalities, out of dynamic range expression signals and parameter stability [15]. In total, 0.1% to 26.3% (median of 2.0%) cells were removed by FlowAI per tube, in over 90% of all tubes less than 5% of cells were removed. Subsequently, cells with extreme values in scatter parameter and doublets were excluded using the R package FlowCore [19] (Text S1). After the FlowAI check, singlet selection and scatter outlier removal, all fcs files harbored at least 40.000 cells. In order to optimize data distribution and range for computational analysis, hyperbolic arcsine transformation with a cofactor of 150 and approximated min-max scaling per parameter were applied (Text S2).

In the next stage of the computational diagnostic workflow, we applied FlowSOM—an algorithm for cell population detection in cytometry data—to the fcs files of the training cohort. FlowSOM groups cells with similar expression patterns into nodes and subsequently groups nodes with similar expression patterns in metaclusters that resemble larger cell populations. The number of nodes was set to 225 clusters (grid size of 15×15) and the number of metaclusters to 30. Per file 40.000 cells were included in an aggregate for the FlowSOM analysis. Subsequently, all cells were mapped on the FlowSOM to compute metacluster percentages, mean fluorescent intensity (MFI) and coefficient of variation (CV) per file. To validate

TABLE 1 Patient characteristics of the training and validation cohorts

A. Patient characteristics training cohort			
	Control ($n = 81$)		**MDS ($n = 67$)**
Age (median)	64 (19–91)		71 (21–94)
Sex (m/f)	43/38		50/17
Diagnoses	Deficiencies/blood loss ($n = 18$)	**WHO**	MDS-SLD ($n = 2$)
	Immune mediated cytopenia ($n = 14$)		MDS-RS-SLD ($n = 6$)
	Spontaneous recovery ($n = 19$)		MDS-MLD ($n = 32$)
	Drug induced ($n = 4$)		MDS-RS-SLD ($n = 17$)
	Alcohol intoxication ($n = 4$)		MDS-U ($n = 7$)
	BM Infiltration solid tumor ($n = 1$)		Isolated del5q ($n = 3$)
	Hemolytic anemia ($n = 1$)	**IPSS-R**	Very low ($n = 13$)
	Cyclic neutropenia ($n = 4$)		Low ($n = 29$)
	Other ($n = 4$)		Intermediate ($n = 13$)
	Healthy donors ($n = 12$)[a]		High ($n = 5$)
			Very high ($n = 2$)
			Insufficient data ($n = 5$)
B. Patient characteristics validation cohort			
	Control ($n = 27$)		**MDS ($n = 30$)**
Age (median)	71 (43–88)		66 (34–78)
Sex (m/f)	15/12		18/12
Diagnoses	Deficiencies/blood loss ($n = 5$)	**WHO**	MDS-SLD ($n = 3$)
	Immune-mediated cytopenia ($n = 4$)		MDS-RS-SLD ($n = 1$)
	Spontaneous recovery ($n = 3$)		MDS-MLD ($n = 16$)
	Drug induced ($n = 2$)		MDS-RS-MLD ($n = 6$)
	Alcohol intoxication ($n = 1$)		MDS-U ($n = 1$)
	Splenomegaly ($n = 1$)		Isolated del5q ($n = 3$)
	Chronic neutropenia($n = 3$)	**IPSS-R**	Very low ($n = 3$)
	Healthy donors ($n = 8$)[a]		Low ($n = 17$)
			Intermediate ($n = 3$)
			High ($n = 3$)
			Very high ($n = 0$)
			Insufficient data ($n = 4$)
C. Patient characteristics validation cohort MDS with excess of blasts			
			MDS ($n = 25$)
Age (median)			72 (51–82)
Sex (m/f)			20-mei
Diagnoses[a]	**WHO**		MDS-EB 1 ($n = 11$)
			MDS-EB 2 ($n = 14$)
	IPSS-R		Intermediate ($n = 10$)
			High ($n = 6$)
			Very high ($n = 8$)
			Insufficient data ($n = 1$)

Abbreviations: del5q, with an isolated deletion of 5q; EB, excess of blasts; IPSS-R, Revised International Prognostic Scoring System; MDS, myelodysplastic syndrome; MLD, multi lineage dysplasia; RS, ring sideroblasts; SLD, single lineage dysplasia; U, undefined; WHO, World Health Organization Classification 2016.

[a]Healthy controls were age-matched non-cytopenic volunteers without a hematological disease.

performance of the FlowSOM method, we studied the expression patterns of the computationally identified cell populations using conventional expert analysis by sequential gating in bivariate plots. In addition, we evaluated concordance with expert analysis by comparing abundancy of several cell subsets (lymphocytes, myeloid progenitors, neutrophils, erythroid cells, plasmacytoid dendritic cells and basophils) identified by expert analysis and FlowSOM. Finally, we mapped fcs files annotated with expert identified cell populations on the FlowSOM models and the Jaccard index was calculated to evaluate concordance.

For the diagnostic workflow, the following cellular features were extracted from the clusters and metaclusters of the FlowSOM analysis per patients: abundance relative to the total number of cells, mean fluorescence intensity (MFI) per parameter and coefficient of variation (CV) per parameter.

Based on the features extracted from the computationally identified cell populations, we trained three machine-learning classifiers to distinguish MDS patients from controls: a Random Forest classifier [14], a Support Vector Machine (SVM) classifier [20], and Generalized Linear Mixed models [21]. To evaluate whether feature selection improved model performance, we performed feature selection using the top 200 and top 50 variables identified by minimum redundancy maximum relevance (MRMR) importance measures and the highest values of the Gini index derived from RF [22, 23]. We used standard settings for SVM, glmnet and RF, except for the number of trees in RF algorithm (10.000).

2.4 | Optimization and evaluation of the computational diagnostic workflow in the training cohort

To select the best features from the FlowSOM analysis and the best performing machine learning classifier, we performed five-fold cross validation in the training cohort. An overview of the explored options is summarized in Table 2. For the initial optimization and evaluation of the computational workflow, we used all six tubes of the FC panel. Subsequently, we aimed to identify the best performing single tube, as a single tube approach would be advantages with regard to antibody costs and processing time. Hereto, we trained the computational workflow based on all tubes separately and evaluated diagnostic accuracy.

We assessed performance using receiver operating characteristic (ROC) curve analysis, the area under the curve (AUC), sensitivity and specificity. Clinical diagnosis, established as described above, was used as reference standard. Accuracy and practical aspects of both computational diagnostic workflows were compared with two manually analyzed diagnostic FC scores, the iFS and the Ogata score. Practical aspects included the required amount of antibodies and bone marrow cells, and time required for analysis. For the iFS and the Ogata score, time required for analysis was measured 10 times for two experienced technicians.

TABLE 2 Basic workflow and optional adaptations

	Required approaches	Optional adaptations
Stage 1. Preprocessing and quality control	Quality control (FlowAI) [15]	
	Doublet removal	
	Removal of extreme values in light scatters	
	Hyperbolic arcsine transformation	
	Approximated min-max scaling	
Stage 2. Feature extraction	Unsupervised clustering on cell level (FlowSOM) [13]	Features derived from metaclusters versus features derived from clusters and metaclusters
	Feature extraction per case	
Stage 3. Classification	Classification for disease status by machine learning	Number of features, all versus 200 versus 50
		Classification by RF [14] versus SVM [33] versus GLM [21]

Note: For each stage, the basic workflow consisted of fixed approaches and optional adaptations. The fixed approaches were not evaluated for superior performance, the optional adaptations were evaluated for superior performance to develop the optimal computational workflow.
Abbreviations: GLM, generalized linear mixed models; MRMR, minimum redundancy maximum relevance; RF, random forest; SVM, support vector machine.

Currently used expert analyzed FC scores are less sensitive in MDS cases without specific morphologic or cytogenetic aberrancies and in low risk disease. Therefore, we evaluated performance in MDS with or without ring-sideroblasts, MDS with or without cytogenetic aberrancies in the combined cohorts and in revised international prognostic scoring system (IPSS-R) defined risk groups.

2.5 | Identification of the most relevant features in the training cohort

To gain insight in the diagnostic workflows, the top 50 and top 10 most important features for discriminating between MDS patients and controls were identified based on the highest values of the Gini index derived from RF [22]—the Gini index reflects the contribution of a feature to the accuracy of the RF analysis. Subsequently, we studied what type of features (i.e., scatter parameters, fluorescent parameters or abundancy) were present in the top 50 most relevant features. For the top 10 features, we performed an additional validation step: the top 10 features were reanalyzed manually in all files of the training cohort, correlated with the computationally identified features and compared between MDS patients and controls.

2.6 | Prospective validation of diagnostic accuracy and most relevant features

To prospectively validate diagnostic accuracy, we trained the single-tube and six-tube workflows using all cases of the training cohort. Subsequently, raw fcs files of the validation cohort were preprocessed, mapped on the FlowSOM and classified as control or MDS based on the RF trained on the training cohort cases.

Sensitivity and specificity were calculated as described for the training cohort. In an additional validation step, we compared the most relevant features—identified in the training cohort—with the same features in the validation cohort.

3 | RESULTS

3.1 | Computationally identified and expert analyzed cell populations show a high concordance

To evaluate performance of FlowSOM in the training cohort, we assessed cell populations identified by FlowSOM for marker expression and examined bivariate expression plots (Figures S1 and S2). Based on these assessments, we concluded that FlowSOM identified major cell subsets including neutrophils, monocytes, progenitor cells (myeloid, erythroid and lymphoid), basophils and plasmacytoid dendritic cells. In addition, several cell populations with unfamiliar expression patterns were encountered (e.g., CD36+/CD71−/CD105+ cells, and CD36+/CD71+/side scatter high). Subsequently, we compared abundancy of cell populations identified by FlowSOM and cell populations identified by expert analysis. Computational and expert analysis showed a high concordance (Table S2, Figure S3).

3.2 | The computational diagnostic workflows outperform currently used FC scores

In the training cohort, we selected the best performing computational workflow based on diagnostic accuracy (AUC, sensitivity and specificity) in the five-fold cross validation. The best performing workflow is visualized in Figure 1, and consists of FlowSOM metacluster features and the machine learning classifier Random Forest without prior feature selection (Table 3). The workflow based on all six-tubes of the FCM panel, reached a sensitivity and specificity of 90% and 96%, respectively, and an AUC of 0.969 (CI 0.945–0.993) (Table 3). Subsequently, we evaluated which single tube reached the highest diagnostic accuracy. Tube 3, developed to assess erythroid dysplasia, obtained the highest diagnostic accuracy, reaching a sensitivity and specificity of 85% and 95%, respectively, with an AUC of 0.964 (CI 0.937–0.991) (Table 2) (Table S3). We assessed accuracy of both the six-tube and the single-tube workflow in several MDS subgroups (i.e., MDS with or without ring sideroblasts and MDS with or without cytogenetic aberrancies, revised international prognostic scoring system (IPSS-R) subgroups), as diagnostic accuracy of traditional flow cytometry score may vary between these groups. We found negligible differences in accuracy between these groups for both the single-tube and six-tube workflows (Table S4). In the second validation cohort concerning MDS patients with EB, a sensitivity of 100% was obtained.

In a next step, we compared the accuracy of both workflows with two expert analyzed flow cytometry scores: the Ogata score and the iFS. Compared to the iFS, the six tubes workflow had a significantly improved specificity ($p = 0.013$) and the single tube showed a trend towards an increased specificity ($p = 0.077$). Both workflows showed a trend towards improved sensitivity (six tubes $p = 0.064$ and single tube $p = 0.078$). Compared with the Ogata score, both computational workflows had a significantly improved sensitivity (six tubes and single tube $p < 0.001$). In addition, the six tubes workflow showed a trend towards improved specificity (six-tube $p = 0.070$ and single tube $p = 0.27$).

With regard to the practical aspects of the computational workflows, total analysis time was reduced from a median of 60 min (range: 45–90 min) for the iFS approach and 10 min (range: 8–15 min) for the Ogata score, to less than 3 min for the six-tubes computational diagnostic workflow and less than 30 seconds for the single-tube computational diagnostic workflow. Introduction of the single-tube-computational diagnostic workflow in the diagnostic work-up of MDS would reduce antibodies and bone-marrow cells needed with 86% compared with the iFS.

3.3 | Evaluation of the cellular characteristics most relevant for diagnosis

In order to give insight into the cellular features contributing most to the accuracy of the computational diagnostic workflows, we identified the top fifty and top 10 features with the highest discriminative power based on the Random Forest-derived variable importance measure based on the Gini index for both the six-tube and the single-tube workflow (Table S5, Table 4) [22]. Fifty percent of the top 50 features of the single-tube and six-tube workflow concerned scatter parameters, 45% fluorescence parameters and 5% abundancy parameters. Subsequently, we evaluated whether the top 50 parameters are included in currently used diagnostic flow cytometry scores—iFS, FCSS, Ogata score, Red score. This was the case for 29% of all top 50 features (Table S5).

As a validation step, we identified the top 10 most discriminative cellular features by re-analyzing all fcs files manually to evaluate concordance with computational analysis. All manually analyzed features correlated significantly with the computationally identified features, and differed significantly between MDS patients and controls (Table S6).

Computationally identified erythroid and myeloid progenitors proved to be most relevant for diagnosis, as these cell populations harbored the most discriminative features (Table 4). Increased side scatter of erythroid cells in the poly- and orthochromatic stage (CD36+/CD71+/CD105-/CD117-/CD34-) was the most discriminative feature to separate patients with MDS from controls. Remarkably, in the six-tubes workflow, three features derived from tube four, five and six, reflected a relatively high side scatter in MDS derived from a cell population that could represent a

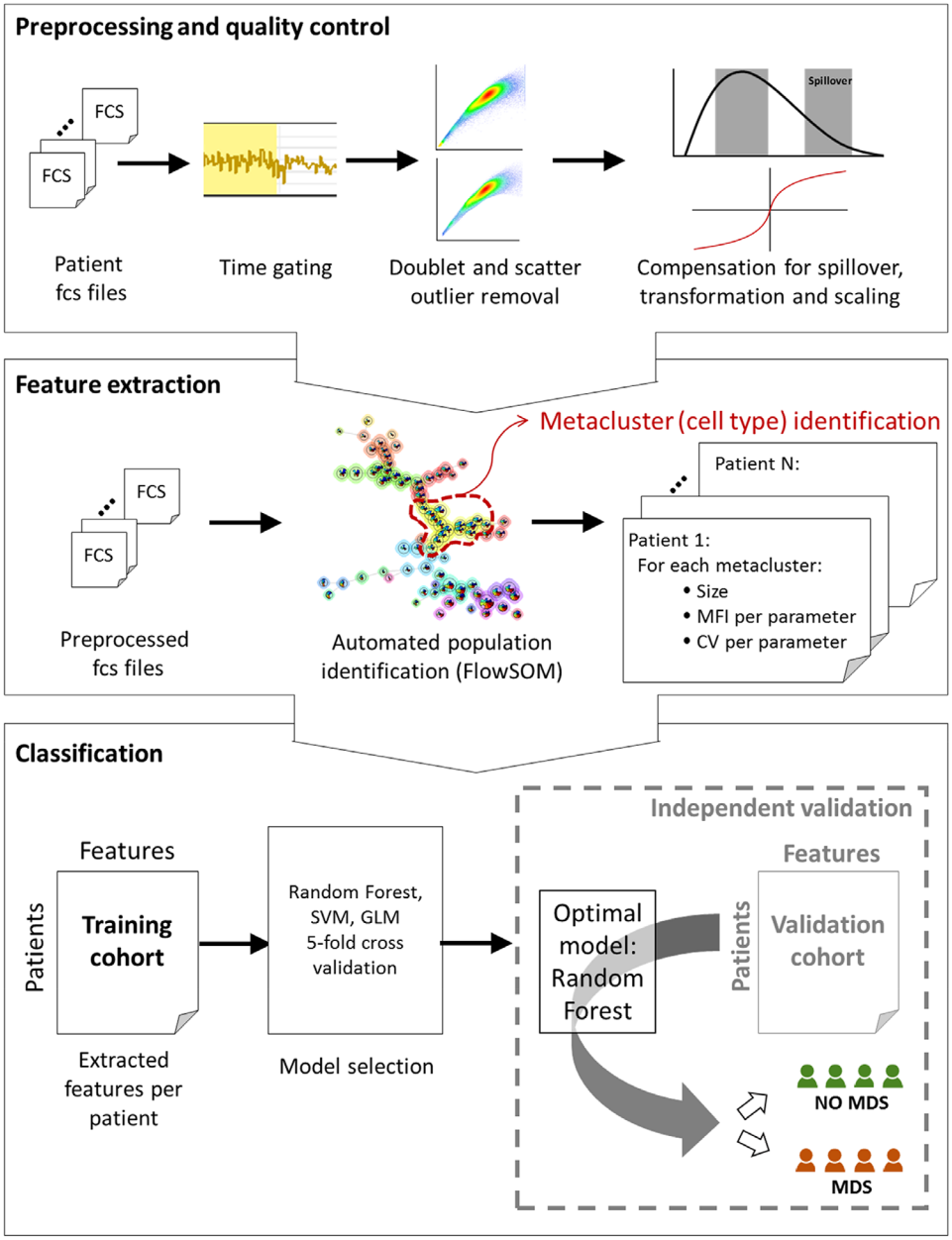

FIGURE 1 Flowchart of the computational diagnostic workflow. All three stages of the computational diagnostic workflow are visualized in the flowchart. In Stage 1, fcs files were preprocessed to ensure high quality data required for computational analysis. In Stage 2, unsupervised clustering was performed on cell level by the FlowSOM algorithm to automatically identify cell types and generate features per case. FlowSOM grouped cells in clusters, represented by the nodes in the depicted FlowSOM tree, and subsequently, grouped the clusters into metaclusters, indicated by the background coloring. Per case the following features were extracted from the FlowSOM analysis: metacluster abundancy, mean fluorescence intensity (MFI) and coefficient of variation (CV) for all parameters of the metaclusters. In Stage 3 of the best performing workflow, these features were supplied to the classification algorithm random forest (RF) to differentiate MDS from controls. Support vector machine (SVM) and Lasso and Elastic-Net Regularized Generalized Linear Models (GLM) were also evaluated in a five-fold cross validation, however, they were outperformed by RF. *fcs = flow cytometry standard, MFI = mean fluorescence intensity, CV = coefficient of variation, SVM = support vector machine, GLM = Lasso and Elastic-Net Regularized Generalized Linear Models, MDS = myelodysplastic syndromes [Color figure can be viewed at wileyonlinelibrary.com]

similar cell type. In order to evaluate whether these features reflected a similar cellular characteristic, correlations between these features were calculated using the Spearman's rank correlation coefficient (two-tailed); all features showed a strong and significant correlation (rho 0.69–0.96, $p < 0.0001$). Notably, this particular erythroid cell population also harbored the 4th and 10th most discriminative feature from the single tube workflow; the CV of CD71 and CD36, which were both increased in MDS patients (Figure 2). As an increased SSC of erythroid cells was not previously identified as relevant parameter in MDS-FC, we evaluated this feature in several MDS subgroups (i.e., low and high risk disease, with or without cytogenetic aberrancies and with or without ring sideroblasts). We found an increase in all MDS groups as compared with controls, yet most pronounced in patients with ring sideroblasts. In MDS patients, the side scatter of erythroid cells correlated significantly with the percentage

TABLE 3 Comparison of performance of expert flow cytometry scores and the computational diagnostic workflows

	Expert scores		Computational diagnostic workflow			
	Ogata score	iFS	Six tubes			Single tube
Training cohort			SVM[a]	GLM[a]	RF	RF
Sensitivity	69%	81%	94%	85%	90%	85%
Specificity	89%	86%	89%	93%	96%	95%
AUC	-	-	0.96 (CI 0.93–0.99)	0.90 (CI 0.85–0.95)	0.97 (CI 0.95–0.99)	0.96 (CI 0.94–0.99)
Validation cohort						
Sensitivity	67%	80%	[a]-	[a]-	90%	97%
Specificity	89%	85%	[a]-	[a]-	93%	95%

Abbreviations: GLM, lasso and elastic-net regularized generalized linear models; iFS, integrated flow score; RF, random forest; SVM, support vector machine.

[a]Accuracies of SVM and GLM with optimized feature selection (50 features for SVM selected by MRMR, all features for GLM).

of ring sideroblasts side scatter (rho 0.53, $p < 0.001$) and hemoglobin levels (rho −0.33, $p = 0.01$) (Figure 2C).

The third most discriminative feature of the six-tube workflow and the second in the single-tube workflow was the median CD117 expression on a cell population with immature erythroid blast-like marker expression (CD36+/CD71+/CD117+/CD105+/CD34+), that was relatively low in MDS. The seventh most discriminative feature of the six-tube workflow was a low CD117 expression in MDS on similar erythroid-like cells derived from tube six. This feature was significantly correlated with the feature derived from tube three (rho 0.815, $p < 0.001$), indicating once more that these features were likely derived from similar cell subsets. Notably, this cell population was characterized by a granulocyte-like side scatter (both in MDS and controls) which has not been described for erythroid cells in literature before. Further analysis and cytomorphology of the sorted population indicated that this subset with high side scatter was an artifact induced by our ammonium-chloride based erythrocyte lysing method (data not shown). In order to ensure that the workflow did not depend on this artificially-induced population, we recalculated sensitivity and specificity when excluding this cell population from classification. The changes in accuracy were negligible (Table S7).

Two discriminative features were not derived from myeloid or erythroid cell populations; the side scatter of HLA-DR negative lymphocytes and the forward scatter on HLA-DR positive lymphocytes. Both features were increased in patients with myelodysplastic syndromes compared with controls.

3.4 | Diagnostic accuracy and most relevant cellular features were confirmed in an independent prospective validation cohort

Finally, we validated performance of both workflows in the independent prospective validation cohort. In the validation cohort, performance for the six-tubes computational diagnostic workflows resulted in 90% sensitivity and 93% specificity, and for the single-tube computational diagnostic workflow 97% and 95%, respectively (Table 3).

As an additional validation step, we studied the most discriminative cellular features of the training cohort in the validation cohort. Cellular features of MDS patients of the training cohort were similar in 16 out of 20 features in the validation cohort, in controls these features were similar in all cases ($p > 0.05$) (Figure S4). Interestingly, the four features that were significantly different between MDS patients in the training cohort and the validation cohort were even more discriminative from controls in the validation cohort compared with the training cohort. Eighteen out of 20 cellular features that were identified as discriminative between MDS and controls in the training cohort, differed significantly between MDS and controls in the validation cohort ($p < 0.05$) as well.

4 | DISCUSSION

The diagnostic work-up in suspected MDS is challenging and heavily relies on morphology and cytogenetics. In this study, we developed a computational diagnostic workflow for flow cytometry to aid in distinguishing myelodysplastic syndromes from non-neoplastic cytopenias. By combining algorithms for pre-processing (FlowAI), feature generation (FlowSOM) and classification (Random Forest), we established a completely automated and accurate diagnostic tool.

We first optimized and evaluated the computational workflow based on six tubes of the FC panel, reaching a sensitivity and specificity of 90% and 93%, respectively, in the validation cohort concerning MDS patients without EB and controls. Subsequently, we identified the best performing single tube, as a single-tube approach has practical advantages with regard to processing time and materials required. In the prospective validation cohort, the single tube workflow obtained a 97% sensitivity and 95% specificity. Thereby, the computational workflows showed (a trend towards) significantly increased accuracies compared with expert analyzed flow cytometry scores (i.e., the Ogata score and the integrated flow cytometry score). In a second validation cohort concerning MDS patients with EB a sensitivity of a 100% was obtained for both workflows, illustrating high diagnostic accuracy across all MDS subtypes. In addition, analysis time for the computational workflows was less than 3 min and implementation

TABLE 4 Identification of the cellular features most relevant for diagnosis

Six-tube			
Populations	**Parameter (in MDS)**	**Relative marker expression based on marker enrichment modeling (MEM)**	**Tube**
1. Erythroid progenitors	SSC Mean (high)	**CD71+ CD36+,** CD105−, CD117−, CD33−, CD45−, FSC-A−, SSC-A−, HLA-DR−, CD34−	3
2. Erythroid progenitors	SSC mean (high)	**CD71+ CD235a+,** CD45− FSC-A− SSC-A− HLA-DR− CD7− CD13− CD34 −	6
3. Lysis artifact	CD117 MFI (low)	**CD105+, CD71+, CD117+, CD36+, SSC-A+, FSC-A+,** HLA-DR−, CD34−, CD33−, CD45−	3
4. Lymphocytes	SSC Mean (high)	**CD45+,** CD117−, CD13− FSC-A−, CD11b−, SSC-A−, CD34−, HLA-DR−, CD16−, CD10−,	1
5. Progenitor	FSC mean (high)	**HLA-DR+, CD34+,** CD117+/−, FSC-A+/−, SSC-A−, CD64−, CD45+/− CD2−, IREM2−	2
6. Erythroid**	SSC mean (high)	CD15−, FSC-A−, HLA-DR−, CD45−, SSC-A−, CD25−, CD123−, CD38−, CD34−, CD117−	5
7. Lysis artifact	CD117 MFI (low)	**CD71+, CD117+, SSC-A+,** CD7−, CD34−, CD13−, FSC-A+, HLA-DR−, CD235a−, CD45−	6
8. Lymphocytes	FSC mean (high)	**HLA-DR+** CD45+/−, CD64−, SSC-A−, FSC-A−, CD2−, CD34−, CD117−, IREM2−	2
9. Myeloid Progenitor	HLA-DR CV (low)	**CD34+, CD117+, HLA-DR+,** CD33+/−, CD71−, FSC-A+/−, CD36−, CD105−, SSC-A−, CD45−	3
10. Erythroid**	SSC mean (high)	SSC-A−, CD45−, FSC-A−, HLA-DR−, CD5−, CD56−, CD7−, CD34−, CD117−, CD19−	4
Single-tube			
Populations	**Parameter (in MDS)**	**Relative marker expression based on marker enrichment modeling (MEM)**	**Tube**
1. Erythroid progenitor	SSC Mean (high)	**CD71+, CD36+,** CD105−, CD117−, CD33−, CD45−, FSC-A, SSC-A, HLA-DR−, CD34−	3
2. Lysis artifact	CD117 MFI (low)	**CD105+, CD71+, CD117+, CD36+,** SSC-A, FSC-A+/−, HLA-DR−, CD34−, CD33−, CD45−	3
3. Early erythroid progenitor	FSC-A mean (high)	**CD105+, CD71+, CD117+, CD36+,** FSC-A−, HLA-DR+, CD33−, SSC-A−, CD34−, CD45−	3
4. Erythroid progenitor	CD71 CV (high)	**CD71+, CD36+,** CD105−, CD117−, CD33−, CD45−, FSC-A−, SSC-A−, HLA-DR−, CD34−	3
5. Early erythroid progenitor	FSC-A mean (high)	**CD117+, CD105+, CD36+, CD71+, FSC-A+,** HLA-DR+, CD34+, CD33−, SSC-A−, CD45−	3
6. Myeloid progenitor	HLA-DR CV (low)	**CD34+, CD117+, HLA-DR+,** CD33+/−, CD71-, FSC-A+/−, CD36−, CD105−, SSC-A−, CD45+/−	3
7. Myeloid progenitor	HLA-DR CV (low)	**HLA-DR+, CD34+, CD117+,** CD71−, CD36−, CD105−, CD33−, FSC-A−, SSC-A−, CD45−	3
8. Progenitor	SSC mean (high)	**CD34+,** HLA-DR+/−, CD117+/−, CD105− CD71−, SSC-A−, CD33−, CD45+/−, FSC-A−, CD36−	3
9. Myeloid progenitor	SSC Mean (high)	**HLA-DR+, CD34+, CD117+,** CD71−, CD36−, CD105−, CD33−, FSC-A−, SSC-A−, CD45−	3
10. Erythroid progenitor	CD36 CV (high)	**CD71+, CD36+,** CD105− CD117−, CD33−, CD45−, FSC-A−, SSC-A−, HLA-DR−, CD34−	3

Abbreviations: CV, coefficient of variation; FSC, forward light scatter; SSC, sideward light scatter.
**Features derived from the same erythroid cell population.

of the single-tube workflow will reduce monoclonal antibody usage with 86% compared with the iFS score. Based on the high accuracies and practical advantages, we suggest to focus further studies on the single tube approach.

As diagnostic FC scores, in particular the Ogata score, have shown to be less sensitive in cases without specific morphologic or cytogenetic aberrancies and in low risk disease, we compared accuracy of the diagnostic workflows in these subgroups [7, 10]. We did not find relevant differences in accuracy between these and other MDS cases without excess of blasts, which implies that this method is useful in the most difficult to diagnose MDS patients. Note, the number of patients in the subgroups were small which may limit the analysis.

The improved accuracy compared with traditional MDS FC scores may partially be explained by the application of novel computational

analysis methods for FC data. Computational tools, such as FlowSOM, are not hampered by upfront knowledge or assumptions and can therefore be applied to approach FC data in an unbiased fashion. This advantage has up till now mainly been used to detect novel cell (sub) populations in flow and mass cytometry data [24]. However, in this study we showed that it can also be a valuable approach to increase diagnostic accuracy by detecting novel features with discriminative power.

In order to give insight in the computational workflows, we studied the most diagnostically relevant features. The top 10 features showed a high accordance with expert analysis, thereby confirming that the features were not a result of a potential artifact caused by scaling or computational analysis. Interestingly, erythroid cells and myeloid progenitors contained the most diagnostically relevant features. The most discriminative computationally identified feature was a high side scatter of maturing erythroid cells in MDS patients that has not been described previously. The increased side scatter may be partially explained by the presence of ring sideroblasts, as the SSC correlated significantly with the percentage of ring sideroblasts. Yet, also patients with less than 15 percent ring, sideroblasts showed an increased erythoid side scatter, indicating that other dysplastic characteristics for example, multinucleation and cytoplasmic vacuolization contribute to an increased side scatter. The same cell population also harbored two other relevant features; the coefficients of variation of CD36 and CD71. Both were previously recognized as diagnostically relevant and are incorporated in multiple expert-analyzed FC scores including the iFS [9, 10].

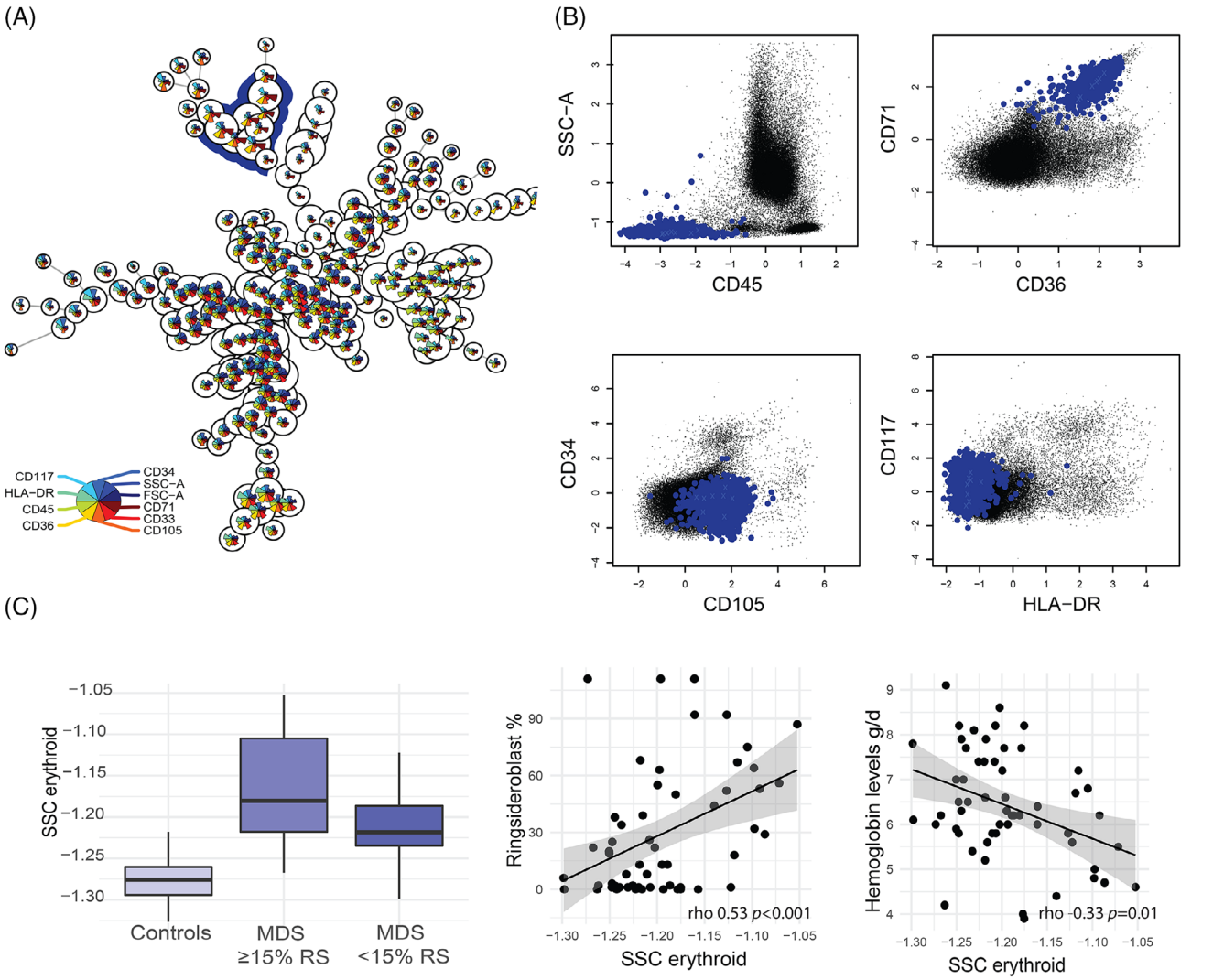

FIGURE 2 Characteristics of the most discriminative cell population. (A) The minimal spanning tree represents the computationally detected populations using the FlowSOM method for tube three (CD71, CD36, CD105, CD117, CD33, CD45, HLA-DR, CD34). The nodes harbor cell subpopulations with similar expression patterns (visualized in the plot pies). The group of clusters highlighted by the blue background coloring represents the erythroid population, CD36 and CD71 positive and CD105 negative, that proved to be important for discriminating between MDS and controls. (B) The computationally detected population depicted in blue represents the erythroid cells that harbor the most relevant features for discriminating between MDS patients and controls (plots from one representative patient). (C) In the left panel, the erythroid sidescatter of controls and MDS patients with and without >15% ring sideroblasts are depicted. In the middle and the right panel, correlation of the erythroid sidescatter with ring sideroblasts and hemoglobin levels are depicted for all MDS patients. Expression parameters underwent hyperbolic arcsine transformation and approximated min-max scaling (Text S2) [Color figure can be viewed at wileyonlinelibrary.com]

Myeloid progenitors harbored several relevant features as well, including an increased forward scatter, indicating larger cell size, and a reduced coefficient of variation of HLA-DR in MDS. The latter may be explained by the clonal nature of the MDS progenitors and thereby, reduced phenotypic heterogeneity resulting in a lower CV as has been previously described for CD117 [8].

Remarkably, two diagnostically relevant features were derived from lymphocytes, HLA-DR negative lymphocytes showed an increased side scatter, whereas HLA-DR positive lymphocytes showed an increased forward scatter in MDS. This may reflect a different functional state of lymphocytes or a different lymphocyte subset distribution [25]. The increased SSC may be explained by an increased number of NK-cells harboring lytic granules in MDS patients, since we observed significantly higher percentages of NK cells (CD56+, CD5−, CD45+) in tube four (Cluster 14, $p = 0.002$). Although interesting, it was beyond the scope of this study to further investigate this finding. Nevertheless, it shows that diagnostic information in suspected MDS is not limited to myeloid or erythroid cells.

Implementation of the computational diagnostic workflow in the diagnostic work-up of MDS will reduce subjectivity associated with expert analysis and interpretation of both FC and morphology information. More importantly, due to the increased accuracy, it may reduce the amount of follow-up bone-marrow aspirates needed, decrease uncertainty for patients and clinicians and improve timely access to suitable treatment options. Introduction of the computational workflow in the diagnostic work-up of patients with clonal hematopoiesis of indeterminate potential (CHIP), clonal cytopenia of undetermined significance (CCUS) and idiopathic cytopenia of undermined significance (ICUS) requires future research, but eventually may contribute to treatment decisions and risk stratification in these challenging conditions [26–28].

The follow-up of our study will have to focus on the validation of this method, preferably in multicenter setting. Studies into computational FC analysis of acute leukemia and multiple myeloma have shown that compliance to quality and harmonization standards allows for successful application in multicenter setting [29, 30]. Since erythroid progenitors proved to be highly relevant for diagnosis, erythrocyte-lysing method has to be considered carefully. Recent reports show that erythrocyte lysis also affects erythroid progenitors and a non-lysis method may be considered [31, 32]. One of the challenges for multicenter validation may be the importance of the side and forward light scatter, since these parameters are currently difficult to standardize. We propose several options for multicenter validation. One option is to diagnose samples from other center based on the workflow and samples described in this manuscript. As mentioned before, this will require far-reaching harmonization, and novel methods for scatter standardization need to be explored. To circumvent the challenging scatter parameter harmonization, the workflow described in this study may be further optimized without scatter parameters. When excluding scatter parameters, diagnostic accuracy of the current diagnostic workflow decreased by 10% (Table S8). In future studies, several options may be considered to increase accuracy without scatter parameters: for example, include (more comprehensive) panels for lymphocyte and progenitor subsets, and further tuning of FlowSOM parameters and machine learning classifiers. An alternative option for multicenter validation is to train the workflow on data acquired in every center separately, this would only require standardization within centers.

To summarize, our work presents the first successful machine learning-based application for FC to distinguish between MDS and non-neoplastic cytopenias, that improves upon currently used flow cytometry diagnostic tools. In addition to the improvements in accuracy, the computational workflow is fast, objective, needs fewer reagents, and offers novel insights into which cellular properties contribute to diagnoses of MDS patients.

ACKNOWLEDGMENTS

We would like to thank all technicians for collecting and analyzing the FC data, especially, A. Zevenbergen and C. Cali. We thank G.J. Ossenkoppele and J. Cloos for critical reading of the manuscript. This study was supported in part by research funding from MDS-RIGHT, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 634789 - "Providing the right care to the right patient with MyeloDysplastic Syndrome at the right time" to Arjan A van de Loosdrecht. Yvan Saeys and Sofie Van Gassen are ISAC Marylou Ingram Scholars.

AUTHOR CONTRIBUTIONS

Carolien Duetz: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; validation; visualization; writing-original draft. **Sofie Van Gassen:** Conceptualization; formal analysis; methodology; software; visualization; writing-review & editing. **Theresia Westers:** Conceptualization; investigation; methodology; resources; supervision; validation; writing-original draft; writing-review & editing. **Margot Van Spronsen:** Conceptualization; writing-review & editing. **Costa Bachas:** Conceptualization; writing-review & editing. **Yvan Saeys:** Methodology; software; supervision; writing-review & editing. **Arjan van de Loosdrecht:** Conceptualization; funding acquisition; methodology; resources; supervision; writing-review & editing.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORCID

Carolien Duetz https://orcid.org/0000-0002-2905-0699

REFERENCES

1. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Solé F, et al. Revised international prognostic scoring system for Myelodysplastic syndromes. Blood. 2012;120:2454–65.
2. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health

Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127:2391–405.
3. Haase D. Cytogenetic features in myelodysplastic syndromes. Ann Hematol. 2008;87:515–26.
4. Germing U, Strupp C, Giagounidis A, Haas R, Gattermann N, Starke C, et al. Evaluation of dysplasia through detailed cytomorphology in 3156 patients from the Düsseldorf registry on myelodysplastic syndromes. Leuk Res. 2012;36:727–34.
5. Porwit A, van de Loosdrecht AA, Bettelheim P, Brodersen LE, Burbury K, Cremers E, et al. Revisiting guidelines for integration of flow cytometry results in the WHO classification of myelodysplastic syndromes—proposal from the international/European LeukemiaNet working Group for Flow Cytometry in MDS. Leukemia. 2014;28:1793.
6. Duetz C, Westers TM, van de Loosdrecht AA. Clinical implication of multi-parameter flow Cytometry in Myelodysplastic syndromes. Pathobiology. 2019;86:14–23.
7. Ogata K, Della Porta MG, Malcovati L, Picone C, Yokose N, Matsuda A, et al. Diagnostic utility of flow cytometry in low-grade myelodysplastic syndromes: a prospective validation study. Haematologica. 2009;94:1066–74.
8. Wells DA, Benesch M, Loken MR, Vallejo C, Myerson D, Leisenring WM, et al. Myeloid and monocytic dyspoiesis as determined by flow cytometric scoring in myelodysplastic syndrome correlates with the IPSS and with outcome after hematopoietic stem cell transplantation. Blood. 2003;102:394–403.
9. Mathis S, Chapuis N, Debord C, Rouquette A, Radford-Weiss I, Park S, et al. Flow cytometric detection of dyserythropoiesis: a sensitive and powerful diagnostic tool for myelodysplastic syndromes. Leukemia. 2013;27:1981–7.
10. Cremers EMP, Westers TM, Alhan C, Cali C, Visser-Wisselaar HA, Chitu DA, et al. Implementation of erythroid lineage analysis by flow cytometry in diagnostic models for myelodysplastic syndromes. Haematologica. 2017;102:320–6.
11. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. Nat Rev Immunol. 2016;16:449–62.
12. Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. Cytometry A. 2016;89:1084–96.
13. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. Cytometry A. 2015;87:636–45.
14. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
15. Monaco G, Chen H, Poidinger M, Chen J, de Magalhães JP, Larbi A. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. Bioinformatics. 2016;32:2473–80.
16. Kalina T, Flores-Montero J, van der Velden VHJ, Martin-Ayuso M, Böttcher S, Ritgen M, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia. 2012;26:1986–2010.
17. Cloos J, Harris JR, Janssen JJWM, Kelder A, Huang F, Sijm G, et al. Comprehensive protocol to sample and process bone marrow for measuring measurable residual disease and leukemic stem cells in acute myeloid leukemia. J Vis Exp. 2018;133:56386.
18. Alhan C, Westers TM, Cremers EMP, Cali C, Ossenkoppele GJ, van de Loosdrecht AA. Application of flow cytometry for myelodysplastic syndromes: pitfalls and technical considerations. Cytometry B Clin Cytom. 2016;90:358–67.
19. Ellis B HP, Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J, Jiang M, Finak G (2018). flowCore: basic structures for flow cytometry data. R package version 1480. 2018.
20. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273–97.
21. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.
22. Louppe G, Wehenkel L, Sutera A, Geurts P. Understanding variable importances in forests of randomized trees. In: CJC B, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in neural information processing systems. Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'13); 2013. p. 431–439. Red Hook, NY: Curran Associates Inc.
23. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics. 2013;29:2365–8.
24. Kordasti S, Costantini B, Seidl T, Perez Abellan P, Martinez Llordella M, McLornan D, et al. Deep phenotyping of Tregs identifies an immune signature for idiopathic aplastic anemia and predicts response to treatment. Blood. 2016;128:1193–205.
25. Kittang AO, Kordasti S, Sand KE, Costantini B, Kramer AM, Perezabellan P, et al. Expansion of myeloid derived suppressor cells correlates with number of T regulatory cells and disease progression in myelodysplastic syndrome. Onco Targets Ther. 2016;5:e1062208.
26. Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015;126:9–16.
27. Valent P, Horny HP. Minimal diagnostic criteria for myelodysplastic syndromes and separation from ICUS and IDUS: update and open questions. Eur J Clin Investig. 2009;39:548–53.
28. Kwok B, Hall JM, Witte JS, Xu Y, Reddy P, Lin K, et al. MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. Blood. 2015;126:2355–61.
29. Flores-Montero J, Sanoja-Flores L, Paiva B, Puig N, García-Sánchez O, Böttcher S, et al. Next generation flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia. 2017;31:2094–103.
30. Lhermitte L, Mejstrikova E, van der Sluijs-Gelling AJ, Grigore GE, Sedek L, Bras AE, et al. Automated database-guided expert-supervised orientation for immunophenotypic diagnosis and classification of acute leukemia. Leukemia. 2018;32:874–81.
31. Wangen JR, Eidenschink Brodersen L, Stolk TT, Wells DA, Loken MR. Assessment of normal erythropoiesis by flow cytometry: important considerations for specimen preparation. Int J Lab Hematol. 2014;36:184–96.
32. Violidaki D, Axler O, Jafari K, Bild F, Nilsson L, Mazur J, et al. Analysis of erythroid maturation in the nonlysed bone marrow with help of radar plots facilitates detection of flow cytometric aberrations in myelodysplastic syndromes. Cytometry B Clin Cytom. 2020;98:399–411.
33. Hearst MA. Support vector machines. IEEE Intell Syst. 1998;13:18–28.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Duetz C, Van Gassen S, Westers TM, et al. Computational flow cytometry as a diagnostic tool in suspected-myelodysplastic syndromes. Cytometry. 2021;99:814–824. https://doi.org/10.1002/cyto.a.24360