Poster #1977

Development and Validation of Automated Flow Cytometry System for the Diagnosis and Prediction of Molecular Abnormalities in Myelodysplastic Syndrome

Maher Albitar, MD[1], Babak Shahbaba, PhD[2], Sally Agersborg, MD, PhD[1], Richard Chang, BS[1], Adam Albitar, BS[1], David Uyeji, BS[1], Gabriel Luchetta, BS[1], Hongjun Su, PhD[3], and Hong Zhang, PhD[3]

[1]NeoGenomics Laboratories, Aliso Viejo, CA; [2]University of California, School of Information and Computer Science, Irvine, CA; [3]Department of Computer Science and Information Technology, Armstrong State University, Savannah, GA

INTRODUCTION

Myelodysplastic syndrome (MDS) is a neoplastic disease of hematopoietic stem cells with abnormalities involving the immune system. Therefore abnormalities can be detected in myeloid lineage as well as reactive lymphoid cells. Diagnosis of MDS has improved significantly with the recent characterization of the molecular abnormalities of this disease. Detection of significant clonal molecular abnormalities typically detected in MDS is currently considered sufficient for the diagnosis of MDS. However, flow cytometry analysis remains the first line in the diagnosis of hematopoitic neoplasms. Absence of the expression of antigen on some cells, or increases or decreases in specific populations of cells can be diagnostic for MDS. Therefore, in principle, diagnosis of MDS relies on detecting these abnormalities in hematopoietic cells. However, other reactive processes can manifest with features overlapping with those of MDS, especially in early stages of the disease. Using pattern recognition-based approaches incorporating multiple variables from flow cytometry has been demonstrated to be the best approach for reliable diagnosis of MDS by flow cytometry. Multiple flow cytometry-based scoring systems have been developed for the diagnosis of MDS. However, most of these studies of various scores used conventional diagnostic confirmation of MDS diagnosis, which remains less objective.

The most commonly studied scoring system is the “Ogata score”, which uses the percentage of CD34+, percent of B-cell within the CD34+, intensity of CD45 on immature myeloid as compared with lymphoid cells, and the granularity in the mature granulocytes. The sensitivity of this score was reported between 65% and 89% and the specificity between 90% to 98%. However, this approach was reported to be very limited in hypocellular BM and pediatric patients. Other studies reported more significant limitations in low-risk MDS.

However, most of these scores involve subjective parameters that are difficult to standardize. We developed a flow cytometry software with a capability to automatically capture relevant parameters of each gated cell population and use the generated metadata in an algorithm for the diagnosis and prediction of molecular abnormalities in MDS.

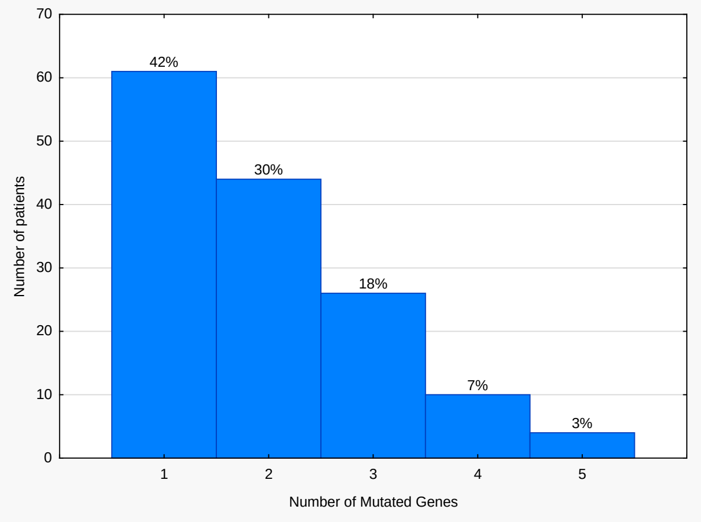
PATIENTS

- 294 bone marrow samples were used for training
- 115 bone marrow samples were used for validation
- 108 samples refereed with diagnosis of AML were also tested using the algorithm
- All samples were referred for suspected diagnosis of MDS due to cytopenia
- All samples had molecular evaluation by NGS using 54 gene panel
- Majority had cytogenetic data
- Patients classified as having MDS if molecular studies or cytogenetic data showed one or more abnormality associated with MDS
- Mutations at allele frequency ≥20% are considered adequate for the diagnosis of MDS

Results:

	Confirmed Negative	Confirmed Disease
MDSTraining (#294)	138	185
MDS Validation (#115)	92	39
AML (#108)	14	94

Classification of patients based on molecular and cytogenetic studies



Distribution of patients with one, two, three, four, or five mutations in the training set

METHOD

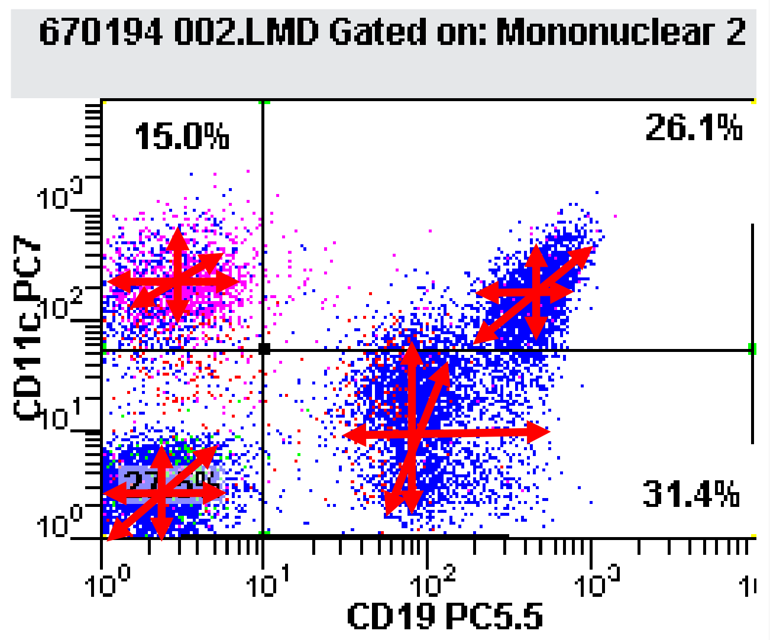
Flow panel and gating

- Standard 23-antibody panel for leukemia and lymphoma evaluation
- Conventional gating to capture on the average 2623 different data points
- Software which automatically captures and saves the following parameters from each quadrant from each gate:
  - Percentage of cells
  - Mean intensity
  - Dispersion in this quadrant (variance) for each antibody on the X and Y axes
  - Correlation coefficient between the X and Y dispersions

SOFTWARE

- Based on classical approach in flow cytometry data analysis.
- Feature added to provide more automated help in data analysis.
- Use of advanced machine learning technologies (SVM) and other mathematical algorithms with custom distribution kernel to detect abnormal flow distributions. Gaussian Mixture models (GMM) are applied to automatic clustering and gating. A special graph algorithm was developed for automatic gate recognition. This system retains the traditional features such as gating definition and adjustments, 2D plots, and statistical tables. However, it provides automation at all analysis steps. Furthermore, the SVM method facilitates analyses far beyond the 2D or 3D limitation in the traditional approach.
- The system provides automated automatic gate prediction, ability to be trained to provide automatic determination of normal versus abnormal plots and automatic determination of diagnosis. The training uses customized designation based on SVM, which is incorporated software.
- Software which automatically captures and saves all possible parameters from each gate. See figure below.

Automated capturing of all possible data

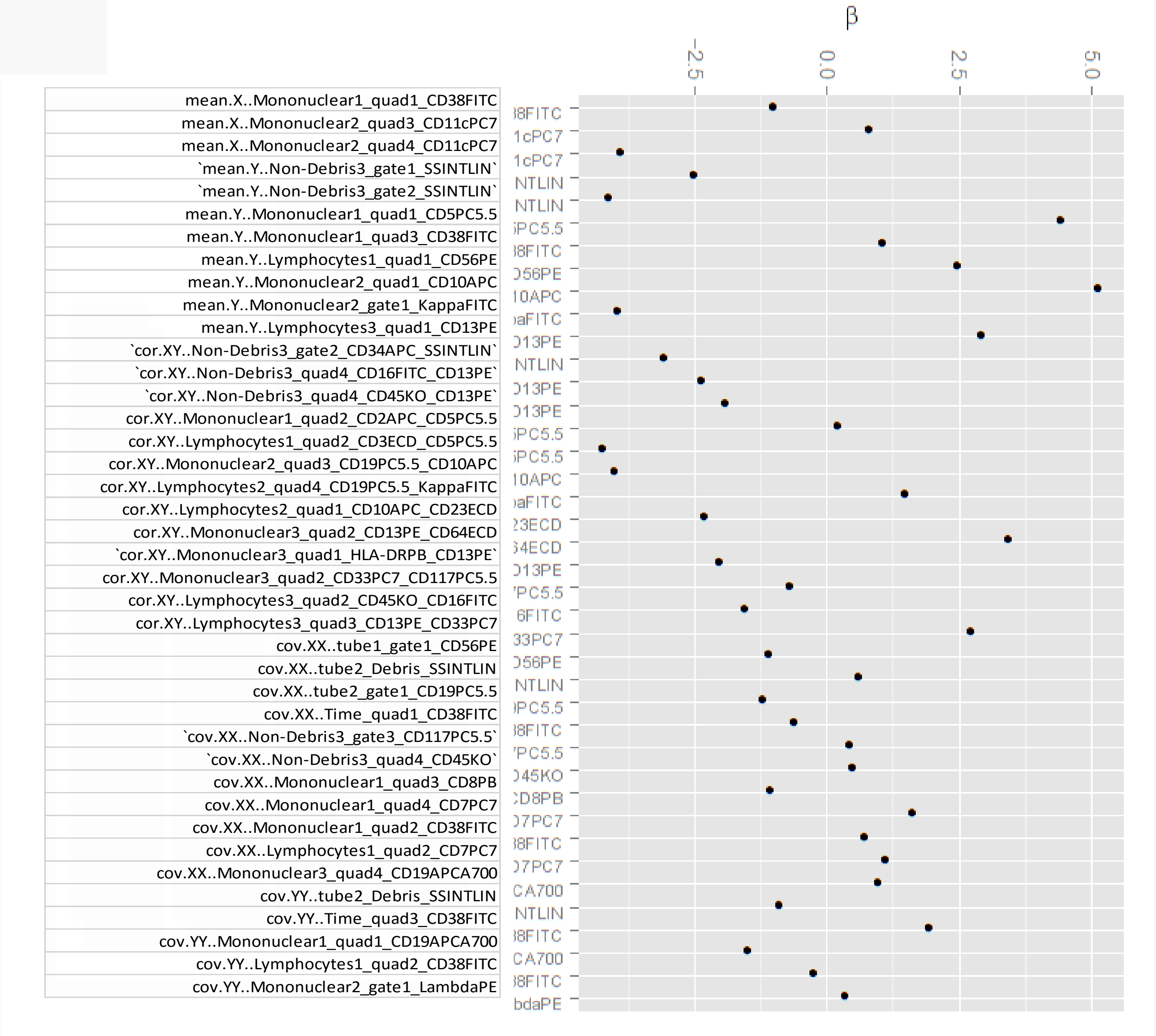


- Percent
- Mean intensity X and Y
- Dispersion (variance) X and Y
- Correlation coefficient between X and Y dispersions

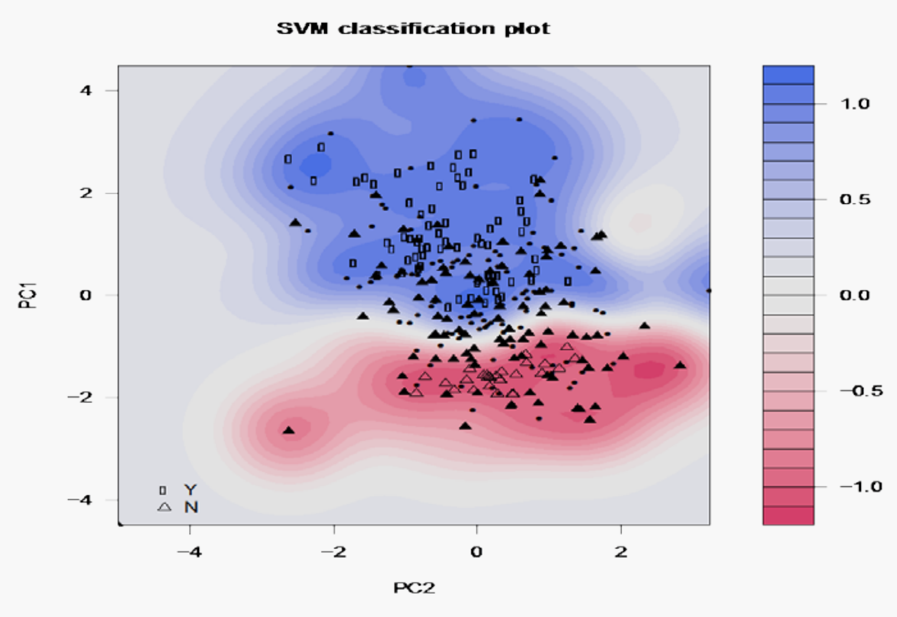
RESULTS

- Univariate analysis showed 103 variables to be statistically significant in distinguishing MDS with adjusted P-values less than 0.05 after controlling for false discovery rate (FDR).
- In multivariate analysis a lasso logistic regression model was used at first and selected 40 variables.
- Using these variables, a predictive model was developed using a support vector machine (SVM) to identify MDS.
- Upon testing this model using the leave-one-out procedure, the area under the ROC curve was 91.6%.

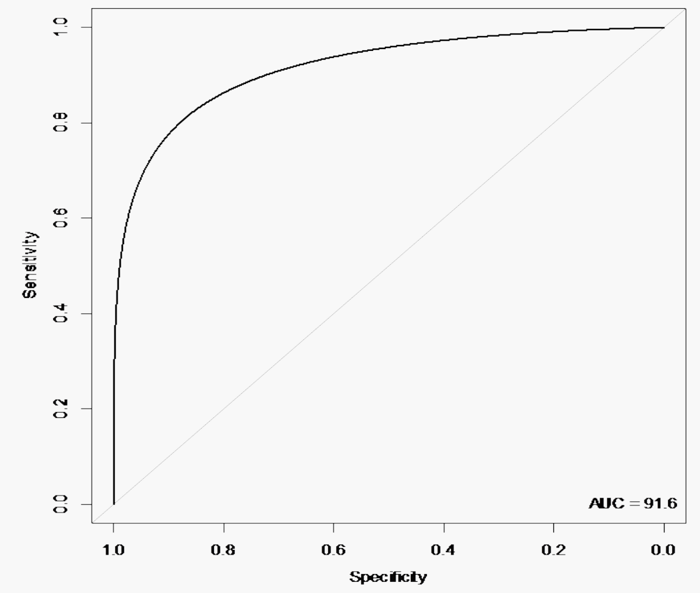
Coefficients from a logistic regression model using the 40 selected variables by LASSO



An illustration of the SVM model using two variables derived based on principal component analysis



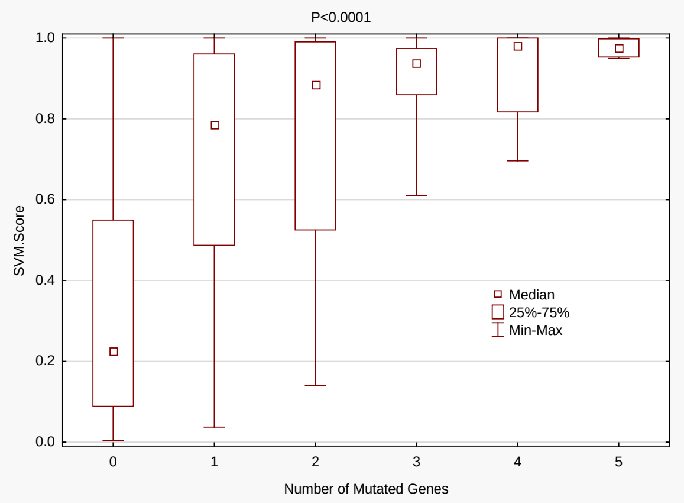
ROC curve for the SVM model based on all the test cases



- For further validation of this algorithm after integration into the software, we tested blindly an additional cohort of 115 patients that had bone marrow submitted for ruling out MDS. The algorithm correctly distinguished between MDS and non-MDS in 104 (90.4%) of these patients using a cut-off point at 0.5 and predicted the presence of cytogenetic abnormality or the presence of one or more genes mutated. However when corrected for cases misclassified, the sensitivity was at 97% and specificity at 93%.
- In addition, we tested cases with questionable diagnosis of AML. The same algorithm detected AML cases as abnormal with a sensitivity at 96% and specificity at 100% after correcting for misclassified cases.
- The algorithm classify AML cases with inv(16) or t(15;17) as normal.

	Score<0.5	Score≥0.5	False Positive	False Negative	Missclassified	Sensitivity after adjusting for missclassified	Specificity after adjusting for missclassified
MDSTraining (#294)	123	171	15	14	1 APL, 2 polymorphism, 2 complex cytogenetic abnormalities, 3 DNMT3A<30%, 3 SF3B1, 1 TP53 but lymphoma, 1 TET2<30%	93%	89%
MDS Validation (#115)	86	29	6	5	1 JAK2 mutation, 3 poor gating	97%	93%
AML (#108)	14	94	0	14	2 Inv16, 2 APL, 2 poor gating, 1 T-ALL, 7 No blasts	96%	100%

Sensitivity and specificity were calculated after correcting for the misclassified cases. Only cases shown in red are included in the correction.



Upon correlating the algorithm score with the number of mutated genes as a reflection of the severity of the disease, there was statistically significant (P< 0.0001) correlation between the score and the number of mutated genes