Phenotype in combination with genotype improves outcome prediction in acute myeloid leukemia: a report from Children’s Oncology Group protocol AAML0531

Andrew P. Voigt,1* Lisa Eidenschink Brodersen,1 Todd A. Alonzo,2,3 Robert B. Gerbing,2 Andrew J. Menssen,1 Elisabeth R. Wilson,1 Samir Kahwash,4 Susana C. Raimondi,5 Betsy A. Hirsch,6 Alan S. Gamis,7 Soheil Meshinchi,2,8 Denise A. Wells2 and Michael R. Loken1

1Hematologics, Inc, Seattle, WA; 2Children’s Oncology Group, Monrovia, CA; 3University of Southern California, Los Angeles, CA; 4 Nationwide Children’s Hospital, Columbus, OH; 5St. Jude’s Children’s Research Hospital, Memphis, TN; 6University of Minnesota Medical Center, Minneapolis, MN; 7Children’s Mercy Hospitals & Clinics, Kansas City, MO and 8Fred Hutchinson Cancer Research Center, Seattle, WA, USA

*APV and LEB contributed equally to this study

ABSTRACT

Diagnostic biomarkers can be used to determine relapse risk in acute myeloid leukemia, and certain genetic aberrancies have prognostic relevance. A diagnostic immunophenotypic expression profile, which quantifies the amounts of distinct gene products, not just their presence or absence, was established in order to improve outcome prediction for patients with acute myeloid leukemia. The immunophenotypic expression profile, which defines each patient’s leukemia as a location in 15-dimensional space, was generated for 769 patients enrolled in the Children’s Oncology Group AAML0531 protocol. Unsupervised hierarchical clustering grouped patients with similar immunophenotypic expression profiles into eleven patient cohorts, demonstrating high associations among phenotype, genotype, morphology, and outcome. Of 95 patients with inv(16), 79% segregated in Cluster A. Of 109 patients with t(8;21), 92% segregated in Clusters A and B. Of 152 patients with 11q23 alterations, 78% segregated in Clusters D, E, F, G, or H. For both inv(16) and 11q23 abnormalities, differential phenotypic expression identified patient groups with different survival characteristics (P < 0.05). Clinical outcome analysis revealed that Cluster B (predominantly t(8;21)) was associated with favorable outcome (P < 0.001) and Clusters E, G, H, and K were associated with adverse outcomes (P < 0.05). Multivariable regression analysis revealed that Clusters E, G, H, and K were independently associated with worse survival (P range < 0.001 to 0.008). The Children’s Oncology Group AAML0531 trial: clinicaltrials.gov Identifier: 00372593.

Introduction

Acute myeloid leukemia (AML) is a heterogeneous disease affecting multiple lineages of hematopoietic cells. The disease is classified by well-defined cytogenetic or molecular aberrations, and as one of eight broadly defined morphologic classes, each with a variety of immunophenotypic features.1 Such diverse assessment modalities are difficult to compare, preventing a more comprehensive understanding of the relationships between morphology, genotype, immunophenotype, and outcome in patients with AML.

Conventional characterization of leukemic immunophenotypes used for lineage assignment involves calculating the proportion of cells with antigen expression above a defined threshold, but does not quantify the amount of each gene product.2 We recently reported that antigen intensity relationships of normal hematopoietic cell populations are invariant throughout maturation from an uncommitted progenitor cell to a mature blood cell among both pediatric and adult individuals.3,4 The study helped confirm that with a high degree of quality control...
and system stability, precise quantification of surface gene product expression can provide a robust basis to assess phenotypic deviations from normal maturation patterns that occur as a result of neoplastic transformation. This concept is supported by our recent report of the recurrent multidimensional immunophenotype, RAM, which independently identifies high-risk pediatric AML at diagnosis.

In this study, we used the complete multidimensional, quantitative leukemic immunophenotype [immunophenotypic expression profile (IEP)] to improve the assessment of the heterogeneity seen in AML. In a study of 769 patients, those with similar global immunophenotypic patterns were grouped together by unsupervised hierarchical clustering. This approach provided a focal point to correlate continuous and categorical variables and determine the relationships among immunophenotype, genotype, morphology, and outcome in a sufficiently large cohort of similarly treated patients. The integration of testing modalities helps identify previously unrecognized patients with poor clinical outcomes, and further clarifies the relationship between a specific genetic event and its effect on the expression of surface gene products.

Methods

Patient samples

Of 1022 newly diagnosed pediatric patients with de novo AML enrolled on the Children’s Oncology Group (COG) protocol AAML0531, 769 satisfied three criteria for the study reported herein: (1) submitting a bone marrow aspirate (N=626, 81%) or peripheral blood specimen (N=143, 19%) (when bone marrow was unavailable) for multidimensional flow cytometry (MDF) at diagnosis, (2) providing consent for correlative biology studies, and (3) MDF analysis showing leukemia comprising >10% of non-erythroid cells. Patients with acute promyelocytic leukemia were not enrolled in the AAML0531 study and those with Down syndrome were excluded from analysis. Details of the AAML0531 protocol have been previously published. Centrally reviewed cytogenetic data and French–American–British (FAB) classification were available for 97.5% and 86.2% of patients, respectively. The study was approved by the institutional review board (IRB) at the National Cancer Institute and IRBs at each of the 184 enrolling centers. Patients and their families provided informed consent or assent as appropriate. The trial was conducted in accordance with the Declaration of Helsinki.

Risk stratification

AAML0531 defined diagnostic risk by cytogenetic or molecular markers. Patients with monosomy 7, deletion 5q, monosomy 5, or FLT3-ITD with a high allelic ratio (>0.4) were classified as high-risk. Patients that had inv(16) (including t(16;16) variants), t(8;21), a CEBPA mutation, or an NPM1 mutation were classified as low-risk. All other patients with known cytogenetics were allocated to the standard-risk group. Patients with persistence of disease, as identified by morphologic assessment at the end of initial induction therapy, were also stratified to the high-risk group.

Flow cytometric analysis

Bone marrow aspirates or peripheral blood samples were drawn in heparin or ethylenediaminetetraacetic acid (EDTA) and submitted for MDF assessment. For correlative biology studies, MDF was performed centrally at Hematologics with a standardized panel of monoclonal antibodies designed to detect measurable residual disease with a difference-from-normal approach. A comprehensive flow cytometric work up was performed at the contributing institution, but was not reviewed centrally. Specimens were processed as previously described.

Hierarchical clustering

Unsupervised hierarchical clustering of the 769 IEPs was performed with R Studio. A dendrogram was constructed using a Euclidian distance metric and a complete-linkage method without scaling of the IEPs. Morphologic and genetic data were not included in the clustering algorithm and did not influence the dendrogram. Selection of the number of phenotypic clusters was validated with the elbow method by comparing within- and between-cluster variation (Online Supplementary Figure S1).

Mutation screening

Genomic DNA was extracted from diagnostic bone marrow specimens by the Puregene® protocol (Gentra Systems, Inc.). CEBPA, FLT3-ITD, WT1, and NPM1 mutations were screened as previously described. Patients with inv(16) or t(8;21) were further analyzed for coinciding c-KIT mutations.

Morphologic assessment

The initial AML diagnosis was made at each contributing institution, and concurrence of the diagnostic morphologic assessment was centrally reviewed. In the central review, subtypes were assigned according to the FAB and World Health Organization (WHO) 2001 classifications (Online Supplementary Table S1), as the clinical trial began prior to the release of the 2008 WHO.

Results

Phenotypic clustering

Diagnostic specimens from 769 patients enrolled in AAML0531 were assessed for quantitative expression of several cell surface markers using a standardized panel of reagents (Figure 1A-D). The neoplastic cell population from each specimen was identified by using CD45 versus log right-angle light scatter (SSC) gating with WinList (Verity Software House, Topsham, ME, USA), and was subsequently verified with all combinations of reagents (Figure 1E). The log mean fluorescence intensities (MFI) of 12 cell surface antigens as well as the physical parameters forward scatter (FSC) and log SSC were then determined for the identified leukemic cell population. The coefficient of variation (CV) of CD34 expression was also calculated as an independent parameter for each patient’s leukemia, since CD34 has been shown to provide a measure of maturation for neoplastic cells. Together, these independently quantified characteristics defined the IEP for each patient as a location in a 15-dimensional data space (Figure 1F,G). Of note, the methodology of CD45 vs. SSC gating in defining the IEP precludes analysis of the influence of minor phenotypic (sub)clones on phenotype.

Unsupervised hierarchical clustering was performed using the calculated IEPs to segregate patients with similar multidimensional phenotypes into related regions of a dendrogram (Figure 2A). The relative intensities of each antigen assessed were depicted in a blue-to-yellow color gradient (extending over four log units) as a heatmap (Figure 2B). Although the dataset consisted of a heterogeneous collection of 769 unique quantitative diagnostic phenotypes, unsupervised clustering identified groups of patients with similar IEPs. Computational analysis sug-
gested that the dataset could be appropriately divided into eleven distinct clusters (Online Supplementary Figure S1) with similar IEPs (Clusters A–K, Figure 2A,B). Comparable phenotypic heterogeneity was observed across specimen types (peripheral blood and bone marrow).

**Association between phenotype and morphology**

Although the current WHO classification of AML is dependent on the molecular and genetic features of leukemia,1 morphologic classification of AML describes lineage and maturational features of the leukemic population.10 To determine the relationship between morphologic subtype and immunophenotype, phenotypic clusters were assessed for co-occurrence of FAB subtypes (Figure 2C, Online Supplementary Table S2). Patients classified as FAB-M0 or M1 (N=22 and N=90, respectively) were scattered throughout the dendrogram and had no identifiable groupings. Patients classified as FAB-M2 (N=161) (blue) segregated in two predominant regions of the dendrogram within Clusters A and B. The majority of patients classified as FAB-M4 (N=165) (green) segregated near the top of Cluster A. Patients classified as FAB-M5 (N=144) (yellow) were identified in a large region of the dendrogram corresponding to Clusters D, E, F, and G. Nine patients classified as FAB-M7 (N=30) did not segregate together. Patients classified as FAB-M7 did not segregate together. Patients classified as FAB-M7 (N=30) predominantly segregated to Clusters H and K. These findings suggest that some morphologic groups share similar patterns of expression of gene products. Furthermore, some FAB classes can be subdivided according to phenotypic differences.

**Association between phenotype and genotype**

The underlying cytogenetic and mutational status of each patient was appended to the dendrogram to analyze the association between genotype and phenotype (Figure 2C). Clear relationships between IEPs and underlying genotypes were identified, as many patients with the same genetic abnormality segregated in similar regions of the dendrogram. Each phenotypic cluster (A–K, Figure 2A,B) was analyzed for high-density regions of each genetic abnormality (consisting of at least 9 patients). A genotypic subcluster was assigned for each high-density region identified (Subclusters A-i to K-i, Figure 2D and Online Supplementary Table S3).

The major chromosomal abnormalities were highly correlated with IEPs. Of the 95 patients with inv(16), 79% were within Cluster A (Figure 2C). Subcluster analysis revealed that 53% of all inv(16) patients were tightly clustered within the A-ii region and 20% of all inv(16) patients segregated to the A-v region of Cluster A (Figure 2C,D). Patients with inv(16) made up 86% of Subcluster A-ii and 35% of Subcluster A-v. Subclusters A-ii and A-v had similar frequencies of patients with coinciding c-KIT mutations (30% and 26%, respectively). Both subclusters were associated with FAB M4 morphology (89% and 48%, respectively). Patients in Subclusters A-ii and A-v had distinct multidimensional phenotypes (Online Supplementary Figure S2).

Of the 109 patients with t(8;21), 92% segregated in Cluster A or B. Strikingly, 70% of the patients with t(8;21) were identified in Subclusters A-iii and B-i (Figure 2C,D).
Figure 2. Hierarchical clustering of IEPs. (A) A dendrogram was generated by unsupervised hierarchical clustering of the 769 IEPs. Eleven phenotypic clusters (A–K), selected by minimizing within-cluster variation and maximizing between-cluster variation, were identified for outcome analysis. (B) The IEP of each patient is presented in the form of a heatmap. (C) The morphologic, karyotypic, and mutational profiles of each patient were compared to the IEPs. (D) Genotypic (sub)clusters with associations among IEPs and morphologic, karyotypic, and/or mutational abnormalities were identified for further analysis. (E) Key denoting intensity of the surface gene product expression to color scale and mutational and morphologic classifications. Somatic mutations are denoted in red and those for wild-type patients are denoted in gray. FAB classifications are indicated by color.
These two phenotypic groups are largely distinguished by quantitative expression of CD56 (Online Supplementary Figure S3). Subclusters A-iii and B-i predominantly included patients with t(8;21) (85% and 83%, respectively). Further, these subclusters were strongly associated with FAB M2 morphology (79% and 80%, respectively). Interestingly Subcluster A-v, which was associated with inv(16), also included 17 patients with t(8;21) (all of which were inv(16) negative). Of all patients with t(8;21), 16% were identified within Subcluster A-v, which was associated with FAB M2 morphology (79% and 80%, respectively). Further, these subclusters were strongly associated with co-existing FLT3-ITD mutations. In Subcluster A-iv, 65% (11 of 17) of patients had FLT3-ITD mutations, whereas those in Cluster C-i had an WT1 mutation, therefore, 42% of all patients in the dataset had both mutations. In Subcluster C-i, only 16% (4/25) of patients with FLT3-ITD mutations also had a CECPA mutation; however this accounted for 44% of all patients that had co-existing FLT3-ITD and CECPA mutations. In Subcluster C-ii, 50% (9 of 18) of patients with FLT3-ITD mutations also had an NPM1 mutation, constituting 48% of all patients in the dataset with both FLT3-ITD and NPM1 mutations.

**Associations among phenotype, genotype, and outcome**

Kaplan–Meier analysis of outcomes was performed to define the 5-year event-free survival (EFS) of patients in different phenotypic clusters (Figure 5). The 5-year EFS of patients in each individual cluster was compared to the EFS of all other patients; statistically significant differences were observed for patients in Clusters B, E, G, H, and K (Table 1). Representations of phenotypes observed for these clusters are shown in Online Supplementary Figures S5-S9.

Univariable analysis revealed that 5-year EFS and overall survival (OS) varied among patients in different IEP clusters. Patients in Cluster B had more favorable 5-year EFS and patients within Clusters E, G, H, and K had more adverse OS and EFS than those in other clusters. Patients in Cluster B (who predominantly had t(8;21)) had significantly higher 5-year EFS (69%, CI: 57%–78%) than those in other clusters (46%, CI: 45%–50%; P<0.001). Interestingly, patients in Clusters E, G, H, and K had poor 5-year EFS (19%–39%; Table 1). After adjusting for age and molecular/cytogenetic risk groups, multivariable analysis revealed that patients in Clusters G, H, and K had significantly higher hazard ratios (HRs) for EFS and OS, whereas those in Cluster E had a significantly higher HR.
for OS, but not EFS (Table 2). Cluster B, with a high frequency of t(8;21), showed no additional favorable effect on EFS or OS.

A similar outcome analysis was performed on genotypic subclusters to determine whether the combination of phenotypic and genotypic features leads to a more accurate prediction of patient outcomes than genotypic features alone. Patients with inv(16) in Subclusters A-ii and A-v had significantly different outcomes (Figure 4A), which was not further explained by the frequency of corresponding c-KIT mutations (30% vs. 26%, respectively). The 5-year EFS for patients with inv(16) with a phenotype corresponding to Subcluster A-v was significantly higher (84%, CI: 57%–94%) than for those with a phenotype corresponding to Subcluster A-ii (54%, CI: 39%–67%; P = 0.039).

In further analysis of the role of c-KIT mutations in core binding factor (CBF) leukemias, CBF/c-KIT positive patients (N=50) demonstrated no statistically significant differences in EFS (P = 0.105) or OS (P = 0.192) than CBF/c-KIT negative patients (N=154). In addition, three clusters had sufficient (N>1) patients with CBF AML and c-KIT mutations: Clusters A, B, and H. For each of these clusters, the difference in EFS and OS was assessed between CBF/c-KIT positive and CBF/c-KIT negative patients. In Clusters B and H, there was no significant difference in OS or EFS between CBF/c-KIT positive and CBF/c-KIT negative patients. In Cluster A, CBF/c-KIT positive patients (N=29) had a significantly worse 5-year EFS than CBF/c-KIT negative patients (N=91) (50% +/- 19% vs. 71% +/- 10%, P = 0.046). However, a difference in outcome between CBF/c-KIT patients in Subcluster A-ii vs. A-v was not observed for either OS (A-ii: 71.1%, A-v: 77.8%, P = 0.915) or EFS (A-ii: 46.7%, A-v: 55.6%, P = 0.680).

The outcomes of patients with 11q23 abnormalities also differed by phenotype. Patients with 11q23 within Subcluster D-i or E-i, who were assigned to the standard-risk group at diagnosis, had a higher 5-year EFS (Subcluster D-i: 51%, CI: 36%–64%; Subcluster E-i: 42%, CI: 26%–58%) than those in Subclusters F-i, G-i, or H-i (Subcluster F-i: 25%, CI: 8%–47%; Subcluster G-i: 22%, CI: 5%–51%; Subcluster H-i: 20%, CI: 5%–47%), though this difference was not significant (P=0.063) likely due to low sample size. However, merging these clusters on the basis of their relationships within the dendrogram revealed two distinct 5-year EFS outcomes (Subclusters D-i+E-i: 47% vs. Subclusters F-i+G-i+H-i: 23%, P = 0.006). The subclusters in which patients with 11q23 had poorer outcomes did not have a higher frequency of MLL translocation partners associated with higher risk in other pediatric studies of MLL rearrangements.21,22 However, patients with t(9;11) were overrepresented in Subcluster D-i. Therefore, while phenotype did not further subset high-risk MLL rearrangements, it did further identify patients with t(9;11). Similar outcome comparison for patients with t(8;21) within Subclusters B-i, A-iii, and A-v showed no significant difference in outcome with 5-year EFS of 76% (CI: 64%–88%), 85% (CI: 69%–100%), and 58% (CI: 34%–82%), respectively (P=0.152). Likewise, comparison of patients with FLT3-ITD within Subclusters A-iv, A-vi, C-i, and C-ii revealed no significant difference in 5-year EFS.

A specific area of the dendrogram, which primarily comprised Clusters H, I, and J, was void of high-density genotypic subclusters. Although patients in these clusters had several genetic abnormalities, none of the patients with unifying abnormalities grouped together with the combined density and frequency observed in other regions of the dendrogram. The outcomes of patients in Clusters I and J were unremarkable, the absence of patients with inv(16) or t(8;21) is, however, notable.

Cluster H was marked by a large cohort size (N=81) and poor patient outcomes. Of note, 86% of patients within
Cluster H were classified in the low-risk or standard-risk group on the basis of cytogenetic or molecular markers. Strikingly, patients classified in the low-risk group by cytogenetic or molecular markers within Cluster H (N=25) had significantly poorer 5-year EFS (33%) and 5-year OS (66%) than all other favorable-risk patients (N=265) in the study (5-year EFS=72%, \( P<0.001 \); OS=84%, \( P=0.008 \); Online Supplementary Figure S10A,B). Furthermore, Group H predicts significantly worse EFS and OS for high-risk patients, but only predicts significantly worse OS for standard-risk patients (Online Supplementary Figure S10C-F).

**Supervised prediction of cluster and subcluster cohorts**

Unsupervised hierarchical clustering was employed to discover a previously unknown structure in the dataset, namely the relationship between immunophenotype, genotype, and outcome. To apply these identified relationships to new patients, a supervised boosted decision tree algorithm was constructed to replicate the original unsupervised cluster classifications using only the IEP. The 769 patients were divided into training (N=513, 2/3) and testing (N=256, 1/3) cohorts. This algorithm was applied

---

**Table 2. Univariable and multivariable Cox regression analysis of the phenotypic clusters cohorts by age and cytogenetic or molecular risk classification.**

<table>
<thead>
<tr>
<th></th>
<th>OS from study entry</th>
<th>EFS from study entry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>HR</td>
</tr>
<tr>
<td><strong>Univariable</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All other clusters</td>
<td>532</td>
<td>1</td>
</tr>
<tr>
<td>Cluster H</td>
<td>81</td>
<td>2.08</td>
</tr>
<tr>
<td>Cluster E</td>
<td>52</td>
<td>1.82</td>
</tr>
<tr>
<td>Cluster K</td>
<td>16</td>
<td>3.03</td>
</tr>
<tr>
<td>Cluster G</td>
<td>11</td>
<td>3.03</td>
</tr>
<tr>
<td>Cluster B</td>
<td>77</td>
<td>0.72</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3–10</td>
<td>231</td>
<td>1</td>
</tr>
<tr>
<td>0–2</td>
<td>174</td>
<td>1.15</td>
</tr>
<tr>
<td>≥11</td>
<td>364</td>
<td>1.27</td>
</tr>
<tr>
<td>Risk Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>360</td>
<td>1</td>
</tr>
<tr>
<td>Low</td>
<td>290</td>
<td>0.32</td>
</tr>
<tr>
<td>High</td>
<td>108</td>
<td>1.26</td>
</tr>
<tr>
<td>Karyotype by complexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–2</td>
<td>612</td>
<td>1</td>
</tr>
<tr>
<td>3+</td>
<td>138</td>
<td>1.40</td>
</tr>
<tr>
<td>Multivariable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All other clusters</td>
<td>519</td>
<td>1</td>
</tr>
<tr>
<td>Cluster H</td>
<td>77</td>
<td>2.09</td>
</tr>
<tr>
<td>Cluster E</td>
<td>51</td>
<td>1.85</td>
</tr>
<tr>
<td>Cluster K</td>
<td>16</td>
<td>4.18</td>
</tr>
<tr>
<td>Cluster G</td>
<td>11</td>
<td>3.64</td>
</tr>
<tr>
<td>Cluster B</td>
<td>76</td>
<td>1.20</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3–10</td>
<td>227</td>
<td>1</td>
</tr>
<tr>
<td>0–2</td>
<td>173</td>
<td>0.72</td>
</tr>
<tr>
<td>≥11</td>
<td>358</td>
<td>1.51</td>
</tr>
<tr>
<td>Risk Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>360</td>
<td>1</td>
</tr>
<tr>
<td>Low</td>
<td>287</td>
<td>0.31</td>
</tr>
<tr>
<td>High</td>
<td>103</td>
<td>1.26</td>
</tr>
<tr>
<td>Karyotype by complexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–2</td>
<td>612</td>
<td>1</td>
</tr>
<tr>
<td>3+</td>
<td>138</td>
<td>1.58</td>
</tr>
</tbody>
</table>

*Statistically significant hazard ratios with corresponding \( P \) values in bold type. OS: overall survival; EFS: event-free survival; HR: hazard ratio; CI: confidence interval.
to the test cohort, and accurately classified 84.0% of patients within an eleven-class prediction setting (average sensitivity = 0.824, average specificity = 0.982, average F1-score = 0.841). The sensitivity, specificity, and F1-score of predictions for each cluster in the test cohort are detailed in Online Supplementary Table S4. As patients with inv(16) and 11q23 showed divergent clinical outcomes based on subcluster designations, additional boosted tree-based models were trained to identify inv(16) patients within Subclusters A-ii and A-v and 11q23 patients within Subcluster H-i. Subclusters D-i, E-i, F-i, and G-i completely overlap with Clusters D, E, F, and G, hence no additional boosted decision tree models were trained to identify these subclusters. Patients with inv(16) were partitioned into A-ii and A-v subclusters with an overall accuracy of 92.3% (average sensitivity = 0.833, average specificity = 0.895, average F1-score = 0.800). Patients with 11q23 were partitioned into D-i, E-i, F-i, G-i, and H-i with an overall accuracy of 95.4% (average sensitivity = 0.743, average specificity = 0.979, average F1-score = 0.790). Additional details and performance metrics of subcluster models are provided in Online Supplementary Table S5.

Each of the eleven clusters demonstrated a unique pattern of dysregulated surface gene product expression. To characterize these immunophenotypic patterns, boosted decision tree models were trained to distinguish patients in each cluster from all other patients using the IEP. The relative influence of each IEP parameter in generating a correct prediction was quantified, where a high relative influence indicates that a given surface gene product is an important component of a cluster’s immunophenotypic expression pattern. As opposed to the evaluation of positive or negative expression of single antigens, the variable importance quantifications highlight the multidimensional nature of surface gene product dysregulation that defines each of the eleven clusters (Figure 5). This data is depicted in Figure 5, where the six most important IEP parameters for each cluster are displayed and each parameter is subsequently colored to illustrate the quantitative amount of each antigen (or non-antigen variable for SSC and FSC), as compared to the quantitative antigen expression of normal myeloid progenitor cells. For example, the six most important IEP parameters for Cluster A are, in order: CD34, CD56, CD13, HLA-DR, CD33, and CD117. CD34 is the most important parameter and the relative intensity of the antigen is essentially the same as that of normal myeloid progenitor cells. CD56 is the second most important parameter for Cluster A and has increased expression of CD56 compared to normal myeloid progenitor cells (which lack the CD56 antigen). In comparison CD34 is the most important parameter for Cluster J, but due to lack of expression, not present.

Discussion

In this study, we present a novel approach for the diagnostic classification of AML that uses quantitative MDF-based diagnostic classification of AML. This method generates a unique patient-specific profile, which, in combination with the diagnostic karyotype and/or somatic mutations, provides a more robust and precise prognostic tool than that of individual testing modalities. Historically, relationships among immunophenotype, genotype, morphology, and outcome have been loosely correlated, with phenotypic associations hinging largely on the expression of a single antigen. Although previous studies have performed clustering analysis of immunophenotypic data to identify small subgroups of patients with poor prognosis, such studies have not evaluated a sufficiently large cohort of uniformly treated patients. By defining the IEP as a continuous variable, patients with similar global immunophenotypic patterns can be grouped together with hierarchical clustering, thus providing a focal point to correlate continuous and categorical test results. As such, our findings clarify the heterogeneous relationships among phenotype, genotype, morphology at diagnosis, and clinical outcome in pediatric

---

Discussion

In this study, we present a novel approach for the diagnostic classification of AML that uses quantitative MDF-based diagnostic classification of AML. This method generates a unique patient-specific profile, which, in combination with the diagnostic karyotype and/or somatic mutations, provides a more robust and precise prognostic tool than that of individual testing modalities. Historically, relationships among immunophenotype, genotype, morphology, and outcome have been loosely correlated, with phenotypic associations hinging largely on the expression of a single antigen. Although previous studies have performed clustering analysis of immunophenotypic data to identify small subgroups of patients with poor prognosis, such studies have not evaluated a sufficiently large cohort of uniformly treated patients. By defining the IEP as a continuous variable, patients with similar global immunophenotypic patterns can be grouped together with hierarchical clustering, thus providing a focal point to correlate continuous and categorical test results. As such, our findings clarify the heterogeneous relationships among phenotype, genotype, morphology at diagnosis, and clinical outcome in pediatric
AML. Limiting the study to de novo AML in children and young adults avoids the increased complexity of multiple lineages resulting from the progression of myelodysplastic syndrome to AML in adults.

Phenotypic heterogeneity is observed in AML to such an extent that the detailed quantitative gene product expression of each leukemia is unique. The observed heterogeneity is presumably a result of the accumulation of multiple genetic abnormalities that can occur in myriad combinations. Leukemogenesis disrupts normal hematopoietic development by altering the precise amounts and timing of appearance of surface gene products required for proper maturation. The accumulation of multiple genetic mutations causes a loss of gene product

---

**Figure 5. Relative influence of IEP components in each cluster.** A boosted decision tree model was trained to identify patients in each cluster versus all other patients. Variable importance was computed by calculating the mean decrease in the Gini index relative to the maximum decrease in the Gini index. The relative influence of the six most important IEP components were plotted for each cluster. In addition, the relative influence of each IEP component is colored in comparison to the intensity of the gene product expression on normal, uncommitted progenitor cells for pediatric patients. For example, a blue-colored bar indicates that the average intensity of a surface gene product within a cluster is lower than the average intensity of that same surface gene product in normal pediatric patients. The combination of most influential IEP components provides insight regarding the multidimensional pattern of surface gene products that are expressed within each cluster. Of note, surface gene products need not be aberrantly expressed to have a high relative influence.
regulation resulting in a unique quantitative immunophenotype for each individual leukemia.

Previous efforts have applied computational algorithms to elucidate genomic (in one case fully genomic) classifications of adult AML, correlating overlapping genotypic profiles with clinical outcome. It is remarkable that by using immunophenotype as the discriminator of patient cohorts we observe several similarities between the current pediatric study and those (using genomic data as the discriminator) in adult AML. These similarities include: the number of computationally relevant AML subtypes, the high level of specificity with which the t(8;21) and inv(16) cohorts cluster together, and indications of further biologic and prognostic subdivisions within current cytogenetic classifications. Most notably, we observe a similar occurrence of multiple FLT3-ITD subgroups, with a subset exhibiting NPM1 co-mutations, in line with those reported by Papaemmanuil and colleagues. Additional commonalities include an observed subset of t(8;21) patients with co-occurring c-KIT mutation, and, to a lesser extent, a subset of patients with overlapping inv(16) and c-KIT mutations. Where a few previous studies have shown the negative impact of c-KIT on OS, relative risk (RR), complete response (CR), and/or EFS for CBF-AML patients, our results are in agreement with those studies which show no additional prognostic effect of c-KIT on the OS and EFS of CBF-AML patients.

Our novel approach of clustering diagnostic immunophenotypes facilitates the segregation of patients with potentially hundreds of different genotypes into clinically meaningful cohorts, thereby allowing a more accurate prognostic determination within apparently uniform genetic groupings. As patients with similar genotypes segregated in similar regions of the dendrogram, genetic subclusters with high phenotypic-genotypic associations could be identified. This begins to elucidate the relationship between a genetic hit and its phenotypic consequence and the subsequent impact on clinical outcome. We plan to further validate these findings in COG AAML1031.

**Funding**

This work was supported by grants U10CA098543 (Chair's grant), U10CA098413 (the Statistical Center Grant), U10CA180886 (NCTN Operations Center Grant), and U10CA180899 (NCTN Statistics & Data Center Grant).

**Acknowledgments**

The authors would like to thank the patients and families for participating in AAML0531. We also thank Vani Shanker for scientific editing.