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. Wells[1] and Michael R. Loken[1]

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

**APV and LEB contributed equally to this study*

Haematologica 2017
Volume 102(12):2058-2068

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.*

Correspondence:

lisa@hematologics.com

Received: March 21, 2017.

Accepted: September 6, 2017.

Pre-published: September 7, 2017.

doi:10.3324/haematol.2017.169029

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/102/12/2058

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 abnormalities, 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,[3] 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.[5] This concept is supported by our recent report of the recurrent multidimensional immunophenotype, RAM, which independently identifies high-risk pediatric AML at diagnosis.[6]

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.[7,8] 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.[7] A comprehensive flow cytometric work up was performed at the contributing institution, but was not reviewed centrally. Specimens were processed as previously described.[7]

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*).[9,10]

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.[11-14] 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[15] (*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).[7,8] 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.[16,17] 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.[18] 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-M6 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).

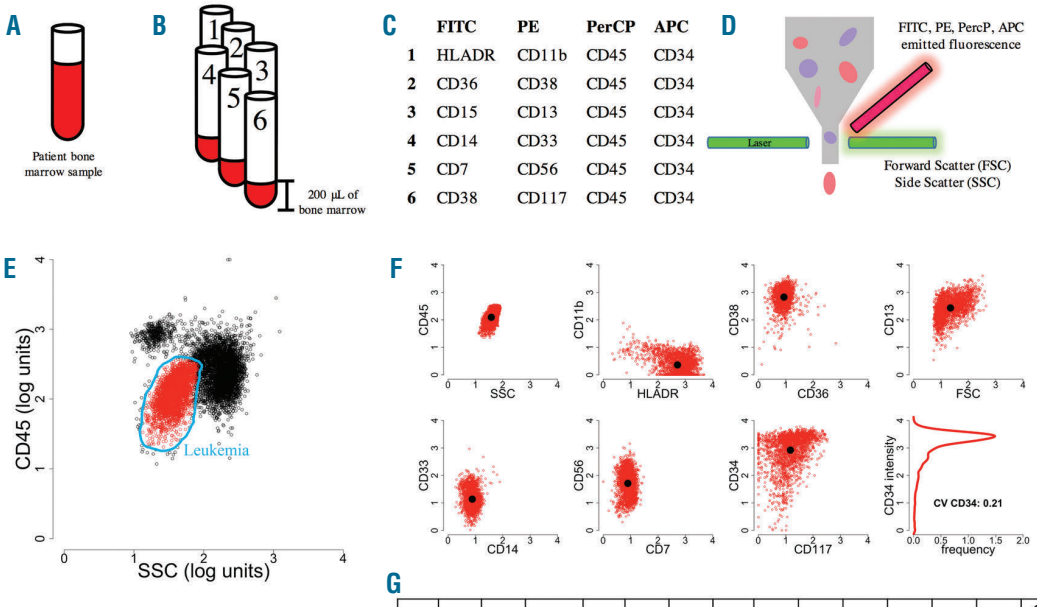

Figure 1. Overview of immunophenotypic expression profiling (IEP). (A) Diagnostic bone marrow specimens were acquired from each patient enrolled in the COG protocol AAML0531. (B) Then, 200 µL of bone marrow was added to 6 tubes containing (C) Fluorescein Isothiocyanate (FITC)-, Phycoerythrin (PE)-, Peridinin Chlorophyll Protein Complex (PerCP)-, and anti-Allophycocyanin (APC)-conjugated antibodies. (D) Flow cytometry was performed on samples in each tube, and fluorescence measurements, forward light scatter (FSC) and right-angle light scatter (SSC) characteristics were collected for 200,000 events. (E) Flow cytometry results were analyzed by an expert, and leukemic populations were identified by CD45 *vs.* SSC gating. (F) For cells identified in the leukemia gate, the mean intensity for each parameter (black dot) was computed. Mean fluorescence intensity was utilized as an unaltered quantification of signal. In addition, the coefficient of variation (CV) of CD34 was computed as a metric to assess cellular maturation. (G) Collectively, these 15 quantitative intensities constituted the IEP for each patient.

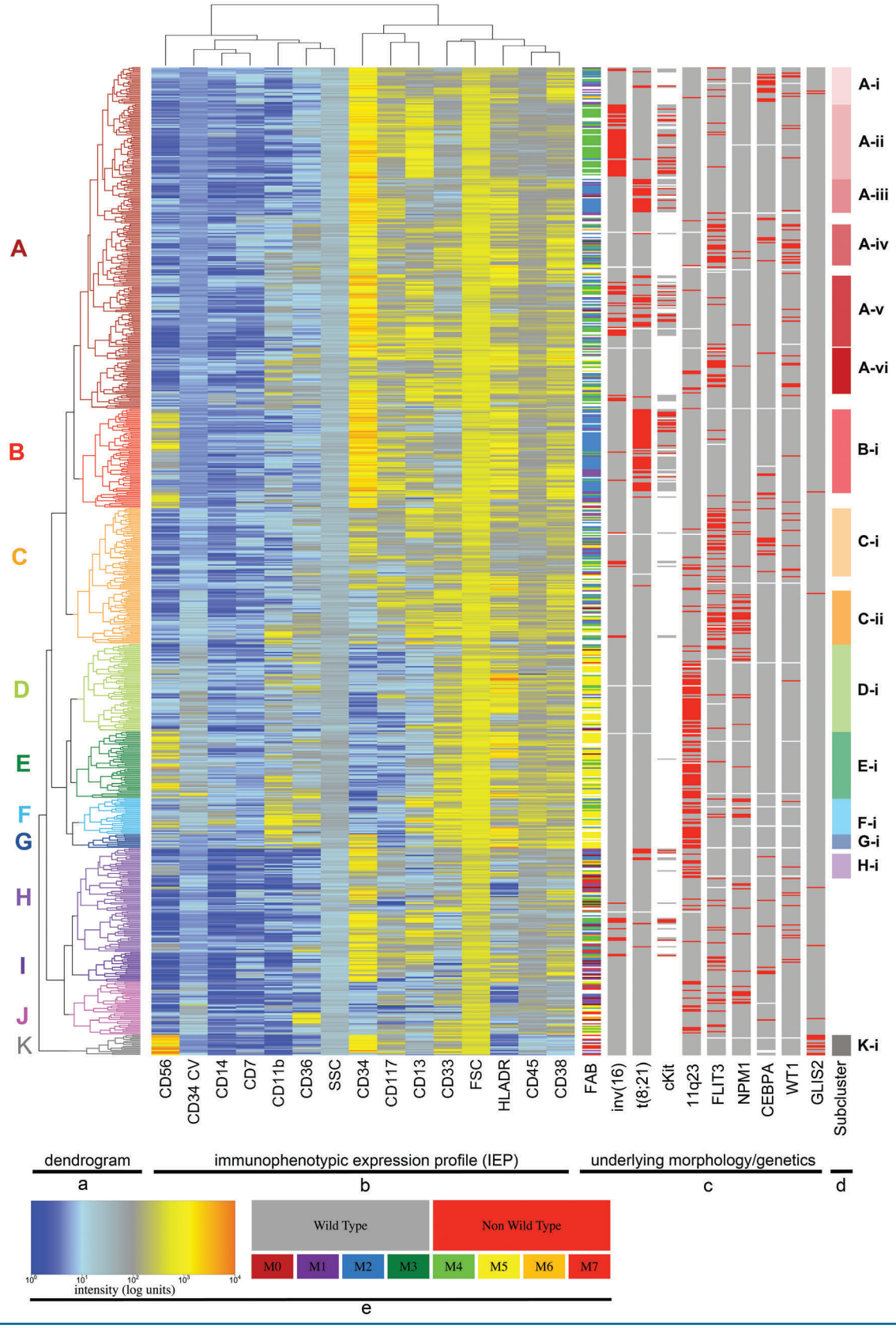

Figure 2. Hierarchical clustering of IEPs. (A) A dendrogram was generated by unsupervised hierarchical clustering of the 769 IEPs. Eleven phenotypic clusters (A–K), selected by minimizing within-cluster variation and maximizing between-cluster variation, were identified for outcome analysis. (B) The 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.

Table 1. Comparison of 5-year EFS (95%CI) of patients in individual phenotypic clusters with that of all other patients.

Cluster*	EFS of patients in cluster of interest	EFS of all other patients	P
Cluster A (*N*=266)	51% (44%–56%)	48% (43%–52%)	0.257
Cluster B (*N*=77)	69% (57%–78%)	46% (43%–50%)	<0.001
Cluster C (*N*=106)	50% (40%–69%)	48% (45%–52%)	0.803
Cluster D (*N*=68)	57% (44%–68%)	48% (44%–51%)	0.149
Cluster E (*N*=52)	39% (26%–52%)	49% (46%–53%)	0.041
Cluster F (*N*=28)	39% (22%–57%)	49% (45%–53%)	0.173
Cluster G (*N*=11)	27% (7%–54%)	49% (45%–53%)	0.027
Cluster H (*N*=81)	28% (18%–38%)	51% (47%–55%)	<0.001
Cluster I (*N*=23)	52% (30%–70%)	49% (45%–52%)	0.583
Cluster J (*N*=41)	56% (40%–70%)	48% (44%–52%)	0.479
Cluster K (*N*=16)	19% (5%–40%)	49% (46%–53%)	0.006

*Statistically significant phenotypic clusters (in comparison with all other patients) are highlighted in blue. EFS: event-free survival; CI: confidence interval.

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% segregated into Subcluster A-v.

The 152 patients with 11q23/*MLL (KMT2A)* alterations had distinct IEPs. Overall, 78% of all 11q23 patients segregated in Cluster D, E, F, G, or H. Within each cluster, a subcluster was defined to further investigate the clinical and biologic features of patients with *MLL* translocations. The majority of patients in each subcluster harbored *MLL* translocations (Subcluster D-i: 66%; Subcluster E-i: 67%; Subcluster F-i: 57%; Subcluster G-i: 82%; and Subcluster H-i: 53%), and each subcluster was strongly associated with FAB M5 morphology (Subcluster D-i: 69%; Subcluster E-i: 78%; Subcluster F-i: 63%; Subcluster G-i: 100%; and Subcluster H-i: 24%). The translocation partners for 11q23 did not appear to be associated with phenotypic heterogeneity (*Online Supplementary Figure S4*). *MLL* chromosomal rearrangements by abnormality (e.g., t(9;11) or t(11;19)) could not be subdivided further into more specific immunophenotypic associations.

Of the 17 patients with the *CBFA2T3–GLIS2* chimeric fusion gene transcript,[19,20] 59% were identified within Cluster K. Conversely, 63% (10 of 16) of patients within this cluster harbored *CBFA2T3–GLIS2* fusions. The IEPs of these patients revealed remarkably consistent bright expression of CD56, dim or negative expression of CD45 and CD38, and a lack of HLA-DR expression, which is consistent with the previously reported RAM phenotype.[6] Within this cluster, 54% of patients had FAB M7 morphology.

AML-associated somatic mutations also had a strong association with immunophenotype. Patients with *CEBPA* mutations segregated in several small groups throughout the dendrogram, most prominently in Subcluster A-i and Subcluster C-i (Figure 2C,D). Of 46 patients with *CEBPA* mutations, 30% were identified within A-i, and 24% were identified within Subcluster C-i. *CEBPA* mutations occurred in 48% of patients in Subcluster A-i and 21% of patients in Subcluster C-i.

FLT3-ITD mutations were associated with 4 genotypic subclusters, often in combination with other genetic mutations. Overall, 61% of all patients with *FLT3*-ITD mutations were identified in Subclusters A-iv, A-vi, C-i, and C-ii. In Subcluster A-iv, 65% (11 of 17) of patients with *FLT3*-ITD mutations also had a *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 *CEBPA* mutation; however this accounted for 44% of all patients that had co-existing *FLT3*-ITD and *CEBPA* mutations. In Subcluster C-ii, 50% (9 of 18) of patients with *FLT3*-ITD mutations also had an *NPM1* mutation, constituting 43% 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 3). 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: 43%–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

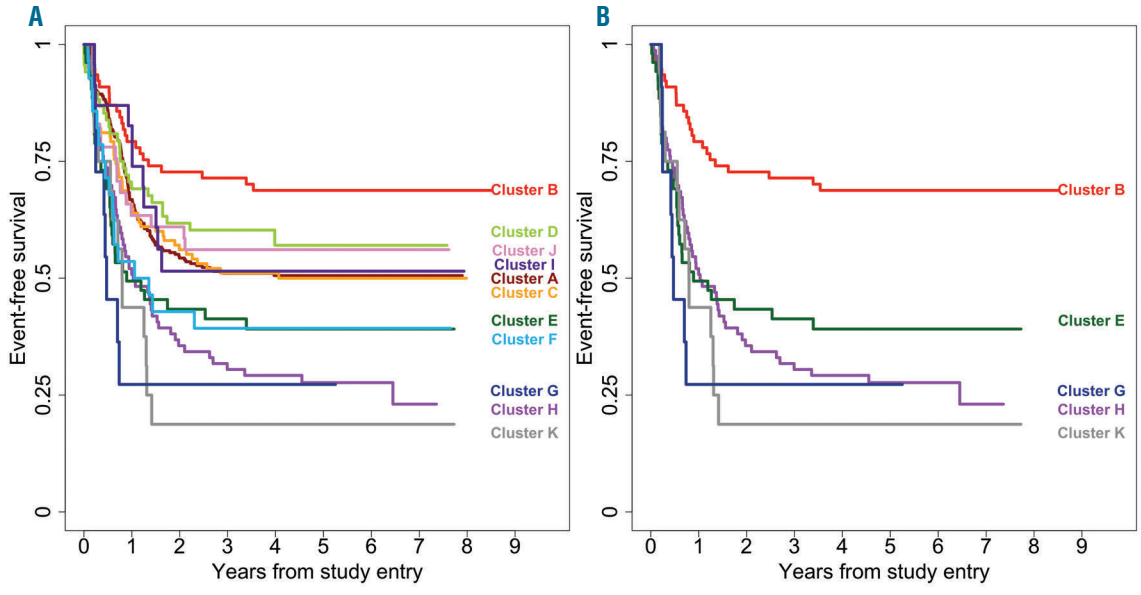

Figure 3. Kaplan–Meier analysis of 5-year EFS of patients by phenotypic cluster. (A) Curves showing differences in EFS for patients in the 11 IEP clusters. (B) Curves showing phenotypic clusters in which the 5-year EFS for patients was significantly different (P<0.05) from that of patients in other clusters. Although patients in Clusters E and F had identical EFS, Cluster F EFS was not statistically significant due to low sample size.

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 *vs.* 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. Within 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: 3%–51%; Subcluster H-i: 20%, CI: 3%–47%), though this difference was not significant (P=0.063) likely due to low sample size (Figure 4B). 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.

		OS from study entry			EFS from study entry		
Univariable	*N*	HR	95% CI	*P*	HR	95% CI	*P*
Cluster groups							
All other clusters	532	1			1		
Cluster H	81	**2.08**	1.50– 2.89	**<0.001**	**1.81**	1.36–2.41	**<0.001**
Cluster E	52	**1.82**	1.19–2.79	**0.006**	**1.55**	1.07–2.25	**0.022**
Cluster K	16	**3.03**	1.69–5.46	**<0.001**	**2.29**	1.31–3.99	**0.004**
Cluster G	11	**3.03**	1.42–6.47	**0.004**	**2.34**	1.16–4.73	**0.018**
Cluster B	77	0.72	0.45–1.16	0.178	**0.56**	0.37–0.85	**0.007**
Age (years)							
3–10	231	1			1		
0-2	174	1.15	0.82–1.61	0.640	1.23	0.93–1.63	0.142
≥11	364	1.27	0.95–1.68	0.102	1.12	0.89–1.42	0.336
Risk Group							
Standard	360	1			1		
Low	290	**0.32**	0.23–0.43	**<0.001**	**0.38**	0.29–0.48	**<0.001**
High	108	1.26	0.93–1.70	0.130	1.31	1.01–1.70	0.045
Karyotype by complexity							
0-2	612	1			1		
3+	138	1.40	1.05-1.87	0.024	1.20	0.93-1.54	0.162
Multivariable							
Cluster groups							
All other clusters	519	1			1		
Cluster H	77	**2.09**	1.48–2.94	**<0.001**	**1.79**	1.33–2.41	**0.001**
Cluster E	51	**1.85**	1.18–2.91	**0.008**	1.41	0.95–2.09	0.087
Cluster K	16	**4.18**	2.18–8.00	**<0.001**	**2.35**	1.29–4.27	**0.005**
Cluster G	11	**3.64**	1.65–8.05	**0.001**	**2.32**	1.12–4.83	**0.024**
Cluster B	76	1.20	0.73–1.98	0.480	0.87	0.56–1.35	0.539
Age (years)							
3–10	227	1			1		
0–2	173	0.72	0.49–1.07	0.103	0.88	0.64–1.21	0.428
≥11	358	**1.51**	1.13–2.02	**0.005**	1.27	1.00–1.62	0.051
Risk Group							
Standard	360	1			1		
Low	287	**0.31**	0.22–0.44	**<0.001**	**0.40**	0.30–0.52	**<0.001**
High	103	1.26	0.90–1.76	0.173	**1.34**	1.00–1.78	**0.047**
Karyotype by complexity							
0-2	612	1			1		
3+	138	**1.58**	1.17-2.14	**0.003**	1.24	0.96-1.61	0.106

*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 presence.

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,[23-27] with phenotypic associations hinging largely on the expression of a single antigen.[28,29] Although previous studies have performed clustering analysis of immunophenotypic data to identify small subgroups of patients with poor prognosis,[30-32] 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

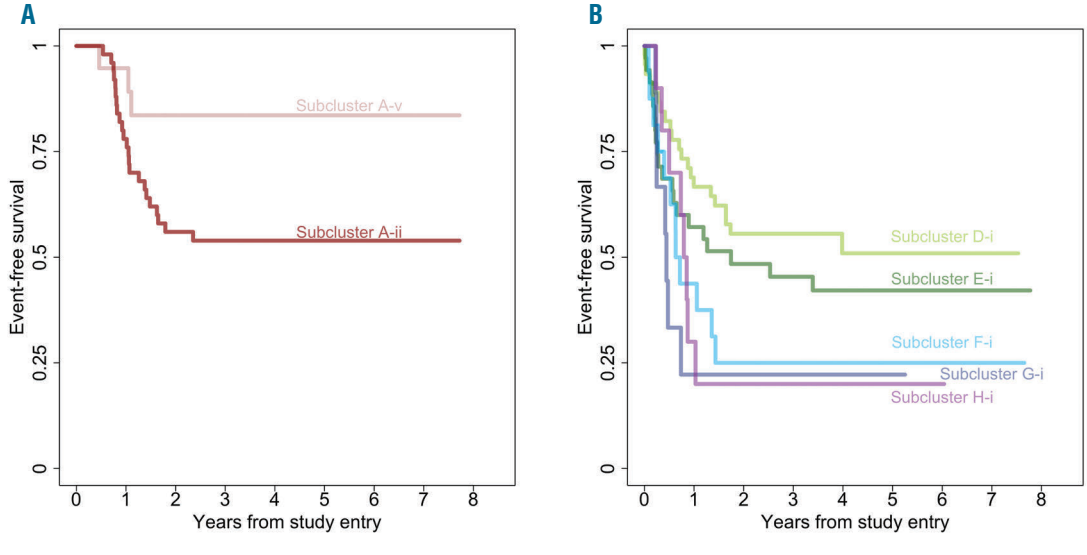

Figure 4. Kaplan–Meier analysis of the differences in 5-year EFS among patients with identical phenotypes in different genotypic subclusters. (A) Patients with inv(16) in Subcluster A-v (green) had a significantly better 5-year EFS than those with inv(16) in Subcluster A-ii (light purple) (*P*=0.039). (B) Patients with 11q23 in Subclusters D-i, E-i, F-i, G-I, and H-i had heterogeneous 5-year EFS.

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.[33] 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

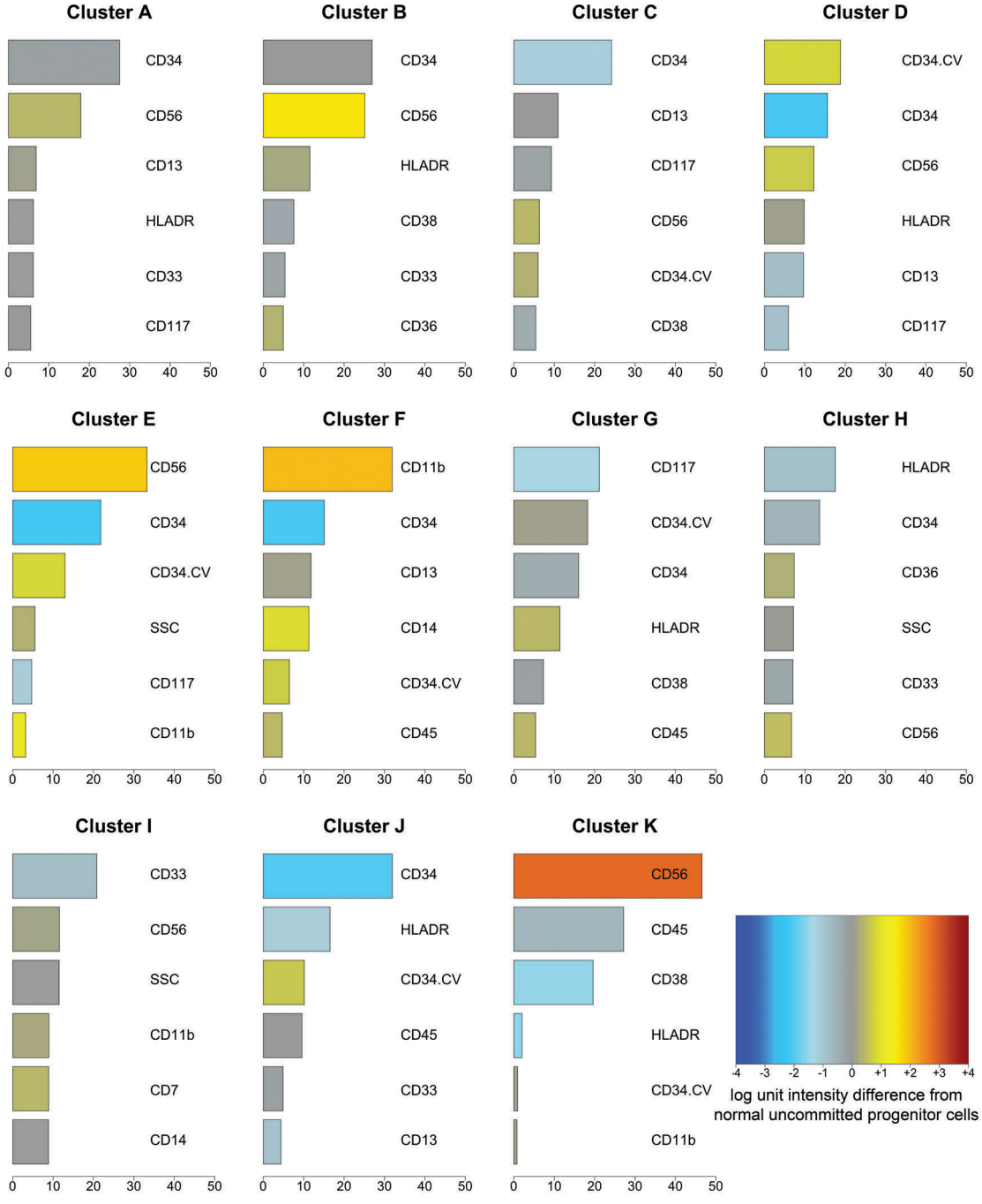

Figure 5. Relative influence of IEP components in each cluster. A boosted decision tree model was trained to identify patients in each cluster *versus* all other patients. Variable importance was computed by calculating the mean decrease in the Gini index relative to the maximum decrease in the Gini index.[10] 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.[3] 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.[34-36] 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.[36] 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,[37-39] 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.[40,41] Our study also revealed more diverse subgroups of the *MLL* fusion patients than previous studies, which is not surprising given the higher prevalence of *MLL* mutations observed in pediatric AML.

Interestingly, immunophenotype alone identifies patient subgroups with adverse clinical outcome. Patients in Clusters G, H, and K had poor 5-year EFS and OS, and both univariable and multivariable Cox regression analyses demonstrated that these phenotypes were independent predictors of poor outcome. Interestingly, Group H had markedly poor outcome and no unifying genetic features, yet a high frequency of patients in the cohort had genetic abnormalities. When comparing patients with favorable-risk cytogenetic/molecular markers in Group H to all other patients with the same favorable-risk markers, those in Group H have significantly worse survival, suggesting that additional uncharacterized mutations captured in the immunophenotype have an adverse effect on patient outcome.

We recently reported that the RAM immunophenotype independently identifies a cohort of very young pediatric AML patients with poor response to therapy and adverse outcome.[6] Herein we demonstrate that this cohort, which was originally identified by expert analysis, can be reproduced by hierarchical cluster analysis. In addition, 63% of RAM patients were discovered to have the *CBFA2T3–GLIS2* chimeric fusion gene, which also indicates a poor prognosis in AML.[19,20] Patients with RAM positive status but *CBFA2T3-GLIS* negative status have equally poor outcome (*data not shown*), highlighting one context in which a solely genomic approach would preclude identifying all poor-risk patients with these clinical features.

Multidimensional phenotypes can also help to further explain the heterogeneous response to therapy seen within conventional cytogenetic classifications. Although patients with inv(16) are considered to be low-risk,[1,42] patients with inv(16) in Subcluster A-ii had significantly worse 5-year EFS than those in Subcluster A-v. Patients with inv(16) in Subcluster A-ii had distinct immunophenotypic features from those within Subcluster A-v (*Online Supplementary Figure S2*), suggesting that additional genetic abnormalities may contribute to the differential expression of gene products and perhaps a more aggressive clinical course. However, both subclusters had a similar prevalence of corresponding *c-KIT* mutations, indicating that the specific addition of the *c-KIT* mutation does not explain the observed difference in outcome, as has been reported among pediatric patients with core binding factor previously defined, thus it should be deleted here and left as CBF-AML.[40] This finding further supports the fact that the combination of phenotype and genotype can provide a more accurate method to predict the risk of induction failure, relapse or death in these genetically defined low-risk 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.

References

1. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937-951.
2. Clinical and Laboratory Standards Institute. Clinical flow cytometric analysis of neoplastic hematolymphoid cells; approved guideline—Second edition. Wayne, PA: US Food and Drug Administration; 2008. FDA publication no. 7-150.
3. Loken MR, Voigt AP, Eidenschink Brodersen L, et al. Consistent gene product expression #2: antigen intensities on bone marrow cells are invariant between individuals. Cytometry A. 2016;89(11):987-996.
4. Loken MR, Voigt AP, Eidenschink Brodersen L, Fritschle W, Menssen AJ, Wells DA. Consistent quantitative gene product expression: #3. Invariance with age. Cytometry A. 2016;89(11):997-1000.
5. Loken MR. Residual disease in AML, a target that can move in more than one direction [editorial]. Cytometry B. 2014; 86(1):15-17.
6. Eidenschink Brodersen L, Alonzo TA, Menssen AJ, et al. A recurrent immunophe-

notype at diagnosis independently identifies high-risk pediatric acute myeloid leukemia: a report from Children's Oncology Group. Leukemia. 2016; 119(10):2077-2080.
7. Loken MR, Alonzo TA, Pardo L, et al. Residual disease detected by multidimensional flow cytometry signifies high relapse risk in patients with de novo acute myeloid leukemia: a report from Children's Oncology Group. Blood. 2012;120(8):1581-8.
8. Gamis AS, Alonzo TA, Meshinchi S, et al. Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: Results from the randomized phase III Children's Oncology Group trial AAML0531. J Clin Oncol. 2014; 32(27):3021-3032.
9. Ketchen DJ, Shook CL. The application of cluster analysis in strategic management research: an analysis and critique. Strat. Mgmt. J. 1996;17(6):441-458.
10. Tibshirani R, James G, Witten D, Hastie T. An introduction to statistical learning- with applications in R. New York, NY: Springer, 2013.
11. Ho PA, Alonzo TA, Gerbing RB, et al. Prevalence and prognostic implications of CEBPA mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. Blood. 2009;113(26):6558-6566.
12. Meshinchi S, Alonzo TA, Stirewalt DL, et al. Clinical implications of FLT3 mutations in pediatric AML. Blood. 2006; 108(12):3654-3661.
13. Ho PA, Zeng R, Alonzo TA, et al. Prevalence and prognostic implications of WT1 mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. Blood. 2010;116(5):702-710.
14. Brown P, McIntyre E, Rau R, et al. The incidence and clinical significance of nucleophosmin mutations in childhood AML. Blood. 2007;110(3):979-985.
15. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. Blood. 2002;100(7):2292-2302.
16. Civin CI, Loken MR. Cell surface antigens on human marrow cells: Dissection of hematopoietic development using monoclonal antibodies and multiparameter flow cytometry. Int J Cell Cloning. 1987;5(4):1-6.
17. Loken MR, Terstappen LWMM, Civin CI, Fackler MJ. Flow cytometric characterization of erythroid, lymphoid and monomyeloid lineages in normal human bone marrow. In: Laerum OD, Bjerksnes R., eds. Flow Cytometry in Hematology. New York, NY: Academic Press; 1992:31-42.
18. Bennett JM, Catovsky D, Daniel MT, et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) Co-operative Group. Br J Haematol. 1976;33(4):451-458.
19. Gruber TA, Larsen Gedman A, Zhang J, et al. An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell. 2012;22(5):687-697.
20. Masetti R, Pigazzi M, Togni M, et al. CBFA2T3-GLIS2 fusion transcript is a novel common feature in pediatric, cytogenetically normal AML, not restricted to FAB M7 subtype. Blood. 2013;121(17):3469-3472.
21. Balgobind BV, Raimondi SC, Harbott J, et al. Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study. Blood. 2009; 114(12):2489-2496.
22. Pigazzi M, Masetti R, Bresolin S, et al. MLL partner genes drive distinct gene expression profiles and genomic alterations in pediatric acute myeloid leukemia: an AIEOP study. Leukemia. 2011;25:560-563.
23. Lin LI, Chen CY, Lin DT, et al. Characterization of CEBPA mutations in AML: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res. 2005;11(4):1372-1379.
24. Hurwitz CA, Raimondi SC, Head D, et al. Distinctive Immunophenotypic features of t(8;21)(q22;22) acute myeloblastic leukemia in children. Blood. 1992;80(12):3182-3188.
25. Paietta E, Andersen J, Gallagher R, et al. The immunophenotype of acute promyelocytic leukemia (APL): an ECOG study. Leukemia. 1994;8(7):1108-1112.
26. Adriaansen HJ, te Boekhorst PAW, Hagemeijer AM, van der Schoot CE, Delwel HR, van Dongen JJ. Acute myeloid leukemia M4 with bone marrow eosinophilia (M4Eo) and inv(16)(p13q22) exhibits a specific immunophenotype with CD2 expression. Blood. 1993;81(11):3043-3051.
27. Baer MR, Stewart CC, Lawrence D, et al. Acute myeloid leukemia with 11q23 translocations: myelomonocytic immunophenotype by multiparameter flow cytometry. Leukemia. 1998; 12(3):317-325.
28. Ossenkoppele GJ, van de Loosdrecht AA, Schuurhuis GJ. Review of the relevance of aberrant antigen expression by flow cytometry in myeloid neoplasms. Br J Haematol. 2011;153(4):421-436.
29. Zeijlemaker W, Kelder A, Wouters R, et al. Absence of leukaemic CD34+ cells in acute myeloid leukaemia is of high prognostic value: a longstanding controversy deciphered. Br J Haematol. 2015;171(2):227-238.
30. Chen CY, Chou WC, Tsay W, et al. Hierarchical cluster analysis of immunophenotype classify AML patients with NPM1 gene mutation into two groups with distinct prognosis. BMC Cancer. 2013;13:107.
31. Hulkkonen J, Vilpo L, Hurme M, Vilpo J. Surface antigen expression in chronic lymphocytic leukemia: clustering analysis, interrelationships and effects of chromosomal abnormalities. Leukemia. 2002;16(2): 178-85.
32. Zangrando A, Luchini A, Buldini B, et al. Immunophenotype signature as a tool to define prognostic subgroups in childhood acute myeloid leukemia. Leukemia. 2006; 20(5):888-891.
33. Terstappen LW, Safford M , Konemann S, et al. Flow cytometric characterization of Acute Myeloid Leukemia. Part II. Phenotypic heterogeneity at diagnosis. Leukemia. 1991;5(9):757-767.
34. Valk P JM, Verhaak R GW, Beijen MA, Erpelinck C AJ, Barjesteh van Waaliwijk van Doorn-Khosrovani S, Boer JM. Prognostically useful gene-expression profiles in Acute Myeloid Leukemia. N Engl J Med. 2004;350(6):1617-1628.
35. Bullinger L, Döhner K, Blair E, Fröhling S, Schlenk RF, Tibshirani R. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med. 2004;350(16):1605-1616.
36. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND. Genomic classification and prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016;374:2209-2221.
37. Shimada A, Taki T, Tabuchi K, et al. KIT mutations, and not FLIT3 internal tandem duplication, are strong associated with a poor prognosis in pediatric acute myeloid leukemia with t(8;21): a study of the Japanese Childhood AML Cooperative study group. Blood. 2006;107(5):1806-1809.
38. Manara E, Bisio V, Masetti R, et al. Core-binding factor acute myeloid leukemia in pediatric patients enrolled in the AIEOP AML 2002/1 trial: screening and prognostic impact of c-KIT mutations. Leukemia. 2014;28:1132-1134.
39. Chen W, Xie H, Wang H, et al. Prognostic significance of KIT mutatins in core-binding factor acute myeloid leukemia: a systematic review and meta-analysis. PLoS ONE. 2016;11(1):e0146614.
40. Pollard JA, Alono TA, Gerbing RB, et al. Prevalence and prognostic significance of KIT mutations in pediatric core binding factor AML enrolled on serial pediatric cooperative trials for de novo AML. Blood. 2010;15(12):2372-2379.
41. Klein K, Kaspers G, Harrison CJ, et al. Clinical impact of additional cytogenetic aberrations, cKIT and RAS mutations, and treatment elements in pediatric r(8;21)-AML: results from an international retrospective study by the International Berlin-Frankfurt-Münster Study group. J Clin Oncol. 2015;33(36):4247- 4259.
42. Grimwade D, Walker H, Oliver F, et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukemia Working Parties. Blood. 1998;92(7):2322-2333.