Mass Cytometric Functional Profiling of Acute Myeloid Leukemia Defines Cell-Cycle and Immunophenotypic Properties That Correlate with Known Responses to Therapy

Gregory K. Behbehani\textsuperscript{1,2,3}, Nikolay Samusik\textsuperscript{1}, Zach B. Bjornson\textsuperscript{1}, Wendy J. Fantl\textsuperscript{1,4}, Bruno C. Medeiros\textsuperscript{2,3}, and Garry P. Nolan\textsuperscript{1}
ABSTRACT

Acute myeloid leukemia (AML) is characterized by a high relapse rate that has been attributed to the quiescence of leukemia stem cells (LSC), which renders them resistant to chemotherapy. However, this hypothesis is largely supported by indirect evidence and fails to explain the large differences in relapse rates across AML subtypes. To address this, bone marrow aspirates from 41 AML patients and five healthy donors were analyzed by high-dimensional mass cytometry. All patients displayed immunophenotypic and intracellular signaling abnormalities within CD34+CD38hi populations, and several karyotype- and genotype-specific surface marker patterns were identified. The immunophenotypic stem and early progenitor cell populations from patients with clinically favorable core-binding factor AML demonstrated a 5-fold higher fraction of cells in S-phase compared with other AML samples. Conversely, LSCs in less clinically favorable FLT3-ITD AML exhibited dramatic reductions in S-phase fraction. Mass cytometry also allowed direct observation of the in vivo effects of cytotoxic chemotherapy.

SIGNIFICANCE: The mechanisms underlying differences in relapse rates across AML subtypes are poorly understood. This study suggests that known chemotherapy sensitivities of common AML subtypes are mediated by cell-cycle differences among LSCs and provides a basis for using in vivo functional characterization of AML cells to inform therapy selection. Cancer Discov; 5(9): 988-1003. © 2015 AACR.

See related commentary by Do and Byrd, p. 912.

INTRODUCTION

Acute myeloid leukemia (AML) includes molecularly and biologically distinct subtypes of disease in which clonal populations of abnormal stem and progenitor cells give rise to a large population of proliferative myeloid blasts and other immature cell types (1). It has long been appreciated that recurrent cytogenetic and molecular abnormalities delineate distinct AML subtypes with differing biologic and clinical characteristics (2, 3). It is also now well established that early stem and progenitor cell subsets [leukemia stem cells (LSC)] within the abnormal myeloid cell compartment contain the leukemia-initiating activity of AML and likely mediate clinical relapse following therapy (4–6). Although LSCs have been studied extensively in recent years, the relationship between the specific properties of LSCs and the clinical features of different disease subtypes remains poorly understood.

LSCs are hypothesized to mediate relapse on the basis of their relative quiescence (7, 8) and protective interactions with the bone marrow niche (9). The evidence for the quiescence of AML stem cells is largely indirect, however. Most studies of human AML cells have been performed on samples treated in vitro or ex vivo with chemotherapy agents that kill bone marrow cells in S-phase, followed by the demonstration that surviving quiescent cells initiate disease in immunocompromised mice. Other studies have demonstrated that murine hematopoietic stem cells (HSC) are generally quiescent in vivo; however, even in normal HSC populations, 50% of the cells will enter S-phase within 6 days and >90% enter S-phase within 30 days (10). A more important weakness of this hypothesis is the marked differences in chemotherapy responsiveness of different AML disease subtypes. Chemotherapy alone cures 60% to 79% of patients with core-binding factor AML [patients with t(8;21), inv(16), or t(16;16) cell karyotypes; refs. 11, 12], whereas patients with FLT3-ITD–mutated AML have a 3- to 5-year overall survival (OS) rate of 20% to 30% (3, 13) in spite of high rates of initial remission. This suggests that the resistance of LSCs to chemotherapy depends on the disease subtype; however, the molecular mechanisms that underlie these differences are largely unknown.

Recent years have witnessed an explosion in the genetic characterization of human malignancies in general and in AML in particular (14, 15). The complete genomic sequencing of hundreds of AML samples has now been completed, and these analyses have demonstrated that the relatively small number of well-characterized mutations in AML (e.g., FLT3, NPM1, c-KIT, CEBPA) exist in a complex landscape of hundreds of other mutations (14, 15). These studies have yielded promising new molecular targets for AML treatment, but how these mutations relate to the functional behaviors of these cells is not well understood. Furthermore, analysis of the largest cohort of sequenced AML samples revealed more than 100 unique combinations of known or suspected oncogenic mutations among 200 patients with AML (14). This complexity suggests that personalized treatment approaches cannot simply target all of the mutations present in a given...
RESULTS

Immediate Sample Collection and Bar-Coded Staining Resulted in Consistent Immunophenotypic and Functional Measurements by Mass Cytometry

Bone marrow aspirates were collected from 35 patients with AML [18 newly diagnosed, 11 relapsed/refractory, 1 patient with relapsed myeloid sarcoma, and 5 patients with AML in complete remission (CR) at the time of sample collection], 4 patients with acute promyelocytic leukemia (APL), 2 patients with high-risk myelodysplastic syndromes (MDS; both transformed to AML within 60 days of biopsy), and 5 healthy donors. Most samples had been analyzed by clinical flow cytometry (15.7%) and intracellular functional markers (14.4%) (39/45) having CVs of less than 20% (Supplementary Table S2). The simultaneous measurement of up to 28 markers (staining panel A) allowed analysis of aberrant marker expression in each immunophenotypic population at a high resolution. Surface marker expression in AML samples was evaluated against normal samples by first gating cells from the normal samples into developmental immunophenotypic subsets based on standard surface markers. Each population from each of the normal and AML samples was then compared across the 28 surface markers as shown in Fig. 3A. Because the normal and AML samples were stained and analyzed in the same tubes simultaneously, levels of each surface marker in each of 35 immunophenotypic populations of each patient sample could be reliably assessed. The summed number of markers of the (28) with aberrant expression levels was calculated for each gated immunophenotypic population in each patient sample (Fig. 3B). Immunophenotypic aberrancies were detected at every stage of myeloid development spanning from HSCs to mature myeloid populations in all AML samples. In the majority of patients, aberrancies were also

Distribution of Cells across Developmental Stages Is AML Subtype Specific

To perform immunophenotypic analysis of the mass cytometry data, both traditional gating and high-dimensional SPADE clustering were performed using 19 of the surface markers common to both staining panels (Supplementary Table S2). The resulting SPADE analysis of the normal bone marrow was consistent across all of the healthy donors; an example from one healthy donor is shown in Fig. 1 and Supplementary Fig. S1. SPADE clustering yielded cell groupings that corresponded to commonly defined immunophenotypic subsets across normal hematopoietic development. Both SPADE clustering (Fig. 2A) and manual gating (Fig. 2B and C; Supplementary Fig. S2) demonstrated that patients with core-binding factor mutations [CBF–AML; n = 5; t(8;21), inv(16), and (16;16) karyotypes] and those with adverse-risk karyotypes (ARK–AML; n = 6) had the highest prevalence of immature immunophenotypes, particularly HSCs (linCD34+CD38+CD45RA CD90+CD33+) and multipotent progenitor cells (MPP, linCD34+CD38+CD45RA CD90−CD33+). The fractions of these two populations were increased more than 50-fold in CBF–AML and ARK–AML patient samples compared with healthy donors (P < 0.002, in all comparisons). In contrast, patients with normal karyotype AML (NK–AML; with or without FLT3 mutation; n = 17) and APL (n = 4) exhibited much smaller increases of approximately 3- to 10-fold in the HSC and MPP populations. The HSC and MPP populations of both CBF–AML and ARK–AML samples were significantly increased compared with these populations in NK–AML or APL (P values ranging from 0.0014 to 0.039; except for ARK–AML MPP vs. APL MPP, where P = 0.067). These findings demonstrate that high dimensionality of mass cytometry can detect unique patterns of cell development that can yield novel potentially diagnostic information.

High-Resolution Analysis Reveals Karyotype- and Genotype-Specific Patterns of Abnormal Surface Marker Expression

The simultaneous measurement of up to 28 markers (staining panel A) allowed analysis of aberrant marker expression in each immunophenotypic population at a high resolution. Surface marker expression in AML samples was evaluated against normal samples by first gating cells from the normal samples into developmental immunophenotypic subsets based on standard surface markers. Each population from each of the normal and AML samples was then compared across the 28 surface markers as shown in Fig. 3A. Because the normal and AML samples were stained and analyzed in the same tubes simultaneously, levels of each surface marker in each of 35 immunophenotypic populations of each patient sample could be reliably assessed. The summed number of markers of the (28) with aberrant expression levels was calculated for each gated immunophenotypic population in each patient sample (Fig. 3B). Immunophenotypic aberrancies were detected at every stage of myeloid development spanning from HSCs to mature myeloid populations in all AML samples. In the majority of patients, aberrancies were also
Figure 1. SPADE plots of normal bone marrow sample #6. SPADE clustering was performed on all samples (normal and AML) simultaneously to generate a single tree structure for all samples. All of the cell events from each sample were then mapped to the common tree structure. Each node of the SPADE tree is colored for the median expression of the indicated markers from low (blue) to high (red). The size of each node is correlated to the fraction of cells mapping to the node; however, a minimum size was enforced for most nodes to allow visualization of node color. Immunophenotypic grouping of nodes was performed manually on the basis of the median marker expression level of each node, and based on analysis of the relevant biaxial plots (e.g., CD38 vs. CD34). NK cell, natural killer cell; Plt, platelets.

detected in most nonmyeloid immunophenotypic populations (including B, T, and natural killer cells), suggesting that the AML clone has wide-ranging effects in the bone marrow.

In addition to analysis of the total number of immunophenotypic aberrancies, specific aberrancies were also detected in patients with different AML subtypes. These were particularly informative in the immunophenotypic hematopoietic stem and progenitor cell (HSPC) compartment (lin−CD34+CD38lo), where strong trends were observed across the different AML subsets (as shown in Fig. 4A; Supplementary Table S4). FLT3-ITD+ NK–AML samples (n = 11) were characterized by increased expression of CD7, CD33, CD123, CD45, CD321, and CD99, and decreased expression of CD34, CD117, and CD38 relative to expression in healthy patient samples. FLT3WT NK–AML samples (n = 6) were characterized by increased expression of CD99 and decreased expression of CD47. All P values were significant (ranging from 0.02 to 4.5 × 10−7). Expression of CD99 was elevated in most of the AML samples (30/36) but was normal in six samples, including two of the three samples with t(8;21) karyotypes. Each of the 36 leukemia samples displayed immunophenotypic abnormalities within the HSPC gated population (3.6 abnormalities on average).

In nine samples, unambiguous separation into immunophenotypically normal and abnormal cell populations was observed among the cells in this gate. Interestingly, among these samples, some markers were aberrantly high in abnormal cells from patients with one disease subtype and aberrantly low in the abnormal cells of another. For example, HLA-DR was extremely high in the abnormal HSPCs of sample #22 (MLL rearrangement) and aberrantly low in the abnormal cells of all four APL samples relative to immunophenotypically normal cells from the same patients (green circles; Fig. 4B). The high-resolution immunophenotypic analysis enabled by mass cytometry thus allowed both the detection of surface marker abnormalities at all stages of hematopoietic development in AML samples and the separation of immunophenotypically normal and abnormal stem cell populations in many patients.
**Figure 2.** Different AML samples demonstrate karyotype- and genotype-specific patterns of cell distribution across developmental populations. **A,** SPADE tree colored for the fraction of total cells in each node from lowest (blue) to highest (red). The size of each node is correlated to the fraction of cells mapping to the node; however, a minimum size was enforced for most nodes to allow visualization of node color. NK cell, natural killer cell; Plt, platelets. **B,** heatmap demonstrating the relative frequency of cells in each of the indicated manually gated immunophenotypic subsets averaged across the patients in the indicated karyotypic and genotypic groups. **C,** the frequency of cells in the indicated (manually gated) stem and progenitor cell compartments for patients of each AML subtype. Error bars, SEM. *, the 14 replicate normal samples came from 5 donors.

**Unique High-Dimensional Immunophenotypic Patterns within the HSPC Population Characterize Different AML Subtypes**

In order to simultaneously view and compare the entire aberrant expression pattern within the HSPC populations as a function of AML subtype, a visualization tool called viSNE was used. viSNE uses a nonlinear, iterative process of single-cell alignment to minimize the multidimensional distance between events and represents the separation of cell events in a two-dimensional map (19). The HSPC population (CD34+CD38−) of each sample was analyzed by viSNE using 19 surface markers. The viSNE patterns from the five healthy donors were consistent (Fig. 5). The AML and APL samples all displayed patterns divergent from normal with notably
Figure 3. Systematic detection of multiple small aberrancies defines large karyotype- and genotype-specific immunophenotypic changes across hematopoiesis in patients with AML. **A**, method for defining aberrant marker expression. All immunophenotypic gates were defined on the basis of the normal samples, and the same gates were applied to each of the AML samples. Once each population was gated, the median expression of each of the 28 surface markers was extracted and compared with the median expression in the 14 samples from five healthy donors for each gated population. AML samples that were 2-fold outside the total variance of the normal samples were considered to be aberrant for that marker (CD45 is shown as an example). **B**, the total number of aberrant markers (of 28 measured markers) was summed for each population and each patient. Each box is colored for the number of the 28 markers that was aberrant for each patient (rows) in each gated immunophenotypic population (columns). The color scale ranges from green, indicating no aberrant marker expression, to red, indicating the highest numbers of aberrancies. The exact number of aberrant markers expressed (of the 28 tested) is printed in each box. The high rates of aberrancy observed in the pre-B-cell population may be due in part to contamination of this gate with dimly CD19-positive malignant myeloid cells due to the limited number of markers defining this immunophenoctype subset (CD19 and CD10) and to the relatively dim staining of these antibodies in normal cells. NK cells, natural killer cells.
Figure 4. AML subtypes are characterized by specific marker aberrances in CD34+CD38lo cells. A, median expression level of the indicated markers in the total CD34+CD38lo population. Each data point represents the median expression level of one patient sample or one of the 14 sample aliquots from the five healthy donors. B, top, bi-axial plots of HLA-DR versus CD33 in the gated CD34+CD38lo subset from three healthy donors. Bottom, bi-axial plots of patient samples show that HLA-DR was either abnormally low (in APL samples) or abnormally high (AML sample #22) in immunophenotypically abnormal cells as compared with residual immunophenotypically normal cells (green circles).
**Figure 5.** viSNE analysis of CD34+CD38lo subset reveals distinct karyotype- and genotype-specific immunophenotypic patterns in high-dimensional space. Each sample was analyzed by viSNE (up to 5,000 sampled events per individual) using 19 dimensions (Supplementary Table S2). Two gates encompassing the vast majority of normal CD34+CD38lo events are shown for reference. The AML subtype is indicated for each sample. Each cell event is colored for its expression level of CD38 from blue (0 ion counts) to red (approximately 40 ion counts). Red cell events still fall within the CD34+CD38lo gate and demonstrated dim CD38 expression.
few cells falling within the viSNE space populated by the CD34+/CD38− cells from healthy donors. Separate, but distinct, patterns could be seen for the two samples with inv(16) karyotypes, two of the three samples with a t(8;21) karyotype, and 10 of the 11 NK–AML samples with FLT3-ITD mutations (of note, the one FLT3-ITD− patient lacking this common pattern, AML#20, did not harbor an FLT3-ITD mutation at disease relapse 6 months later). The HSPCs of the 4 patients with APL (all of which had FLT3-ITD mutations) and two of the three samples with FLT3 tyrosine kinase domain mutations (FLT3-TKD) also exhibited a pattern similar to that of the FLT3-ITD+ NK–AML patients. No clear patterns were observed across the patients with FLT3VT NK–AML or with ARK–AML; however, each sample was clearly different from the normal HSPCs of the healthy donor samples. These differences were statistically significant in at least one tSNE dimension (as described in Supplementary Methods).

To confirm that high-dimensional differences observed in the viSNE analysis could group samples based on AML subtype, the gated CD34+/CD38− subset from each sample was analyzed by binning samples into 100 bins using a K-means clustering of the expression levels of the 19 markers used in the viSNE analysis. This independent analysis produced a hierarchical grouping of the samples based on the pairwise correlation of the distribution of cells across 100 multidimensional bins. This grouped all of the normal bone marrow samples into a single branch of the dendrogram (Supplementary Fig. S3). Ten of the 11 FLT3-ITD+ NK–AML samples were grouped with two of the FLT3-TKD+ APL samples and two of the FLT3-TKD+ NK–AML samples into a separate branch. As in the viSNE analysis, AML#20 was dissimilar from the other FLT3-ITD+ NK–AML samples. Sample AML#18 (ARK–AML, FLT3-ITD+) was also not in this branch of the dendrogram. A z-transform of the correlation of each sample to the normal HSPCs of the healthy donor samples. These differences were statistically significant in at least one tSNE dimension (as described in Supplementary Methods).

To confirm that high-dimensional differences observed in the viSNE analysis could group samples based on AML subtype, the gated CD34+/CD38− subset from each sample was analyzed by binning samples into 100 bins using a K-means clustering of the expression levels of the 19 markers used in the viSNE analysis. This independent analysis produced a hierarchical grouping of the samples based on the pairwise correlation of the distribution of cells across 100 multidimensional bins. This grouped all of the normal bone marrow samples into a single branch of the dendrogram (Supplementary Fig. S3). Ten of the 11 FLT3-ITD+ NK–AML samples were grouped with two of the FLT3-TKD+ APL samples and two of the FLT3-TKD+ NK–AML samples into a separate branch. As in the viSNE analysis, AML#20 was dissimilar from the other FLT3-ITD+ NK–AML samples. Sample AML#18 (ARK–AML, FLT3-ITD+) was also not in this branch of the dendrogram. A z-transform of the correlation of each sample to the normal HSPCs of the healthy donor samples. These differences were statistically significant in at least one tSNE dimension (as described in Supplementary Methods).

To confirm that high-dimensional differences observed in the viSNE analysis could group samples based on AML subtype, the gated CD34+/CD38− subset from each sample was analyzed by binning samples into 100 bins using a K-means clustering of the expression levels of the 19 markers used in the viSNE analysis. This independent analysis produced a hierarchical grouping of the samples based on the pairwise correlation of the distribution of cells across 100 multidimensional bins. This grouped all of the normal bone marrow samples into a single branch of the dendrogram (Supplementary Fig. S3). Ten of the 11 FLT3-ITD+ NK–AML samples were grouped with two of the FLT3-TKD+ APL samples and two of the FLT3-TKD+ NK–AML samples into a separate branch. As in the viSNE analysis, AML#20 was dissimilar from the other FLT3-ITD+ NK–AML samples. Sample AML#18 (ARK–AML, FLT3-ITD+) was also not in this branch of the dendrogram. A z-transform of the correlation of each sample to the normal HSPCs of the healthy donor samples. These differences were statistically significant in at least one tSNE dimension (as described in Supplementary Methods).

To confirm that high-dimensional differences observed in the viSNE analysis could group samples based on AML subtype, the gated CD34+/CD38− subset from each sample was analyzed by binning samples into 100 bins using a K-means clustering of the expression levels of the 19 markers used in the viSNE analysis. This independent analysis produced a hierarchical grouping of the samples based on the pairwise correlation of the distribution of cells across 100 multidimensional bins. This grouped all of the normal bone marrow samples into a single branch of the dendrogram (Supplementary Fig. S3). Ten of the 11 FLT3-ITD+ NK–AML samples were grouped with two of the FLT3-TKD+ APL samples and two of the FLT3-TKD+ NK–AML samples into a separate branch. As in the viSNE analysis, AML#20 was dissimilar from the other FLT3-ITD+ NK–AML samples. Sample AML#18 (ARK–AML, FLT3-ITD+) was also not in this branch of the dendrogram. A z-transform of the correlation of each sample to the normal HSPCs of the healthy donor samples. These differences were statistically significant in at least one tSNE dimension (as described in Supplementary Methods).

To confirm that high-dimensional differences observed in the viSNE analysis could group samples based on AML subtype, the gated CD34+/CD38− subset from each sample was analyzed by binning samples into 100 bins using a K-means clustering of the expression levels of the 19 markers used in the viSNE analysis. This independent analysis produced a hierarchical grouping of the samples based on the pairwise correlation of the distribution of cells across 100 multidimensional bins. This grouped all of the normal bone marrow samples into a single branch of the dendrogram (Supplementary Fig. S3). Ten of the 11 FLT3-ITD+ NK–AML samples were grouped with two of the FLT3-TKD+ APL samples and two of the FLT3-TKD+ NK–AML samples into a separate branch. As in the viSNE analysis, AML#20 was dissimilar from the other FLT3-ITD+ NK–AML samples. Sample AML#18 (ARK–AML, FLT3-ITD+) was also not in this branch of the dendrogram. A z-transform of the correlation of each sample to the normal HSPCs of the healthy donor samples. These differences were statistically significant in at least one tSNE dimension (as described in Supplementary Methods).
Functional Mass Cytometry Analysis of AML Subtypes

RESEARCH ARTICLE

Figure 6. Karyotype- and genotype-specific patterns of S-phase fraction in AML. A, SPADE trees of representative patients of the indicated AML subtypes. Nodes are colored for the frequency of S-phase cells in each node from lowest (blue) to highest (red). NK cell, natural killer cell; Plt, platelets. B, AML cells exhibit a lower S-phase fraction than immunophenotypically similar normal cells. C, stem and progenitor cell populations from patients with CBF–AML exhibit a higher S-phase fraction than immunophenotypically similar cells from patients with other AML subtypes (APL samples excluded from analysis). D, stem and progenitor cell populations from patients with FLT3-ITD+ NK–AML exhibit a lower S-phase fraction than cell from patients with other AML subtypes (APL samples and patients with FLT3-TKD mutations excluded from analysis). Colored boxes group immunophenotypic populations: HSPC, hematopoietic stem and progenitor cells; B, blasts (immunophenotypic); Mono, monocyte lineage cells; Gran, granulocyte lineage; RBC, red blood cell lineage; B-cell, B-cell lineage. Error bars, SEs.
samples. Consistent with previous findings (22, 23), FLT3-ITD+ samples had significantly higher levels of phosphorylated (p) STAT5 (2.5–6-fold relative to normal) across all immunophenotypically defined myeloid populations (Supplementary Fig. S4A). This was not the case for FLT3-ITD+ AML vs. normal (P = 3 × 10⁻³ for FLT3-ITD+ AML vs. all other AML). In addition, levels of pSTAT5 were also significantly higher in the FLT3-ITD+ APL samples. A similar trend was observed for MAPKAPK2 phosphorylation, which was consistently higher in myeloid cell subsets of FLT3-ITD+ NK-AML samples (significant for all FLT3-ITD+ myeloid populations except for the CMP/GMP gate where P = 0.058; Supplementary Fig. S4B). Almost all AML subtypes exhibited higher levels of pERK relative to normal (an average increase of 4-fold; P < 0.05 for all populations; Supplementary Fig. S4C). Conversely, phosphorylation of 4EBP1 was lower than normal across all AML subtypes and in all gated myeloid populations with the exception of the metamyelocyte and mature granulocyte populations of patients with FLT3-ITD+ AML (Supplementary Fig. S4D).

As aberrant intracellular signaling could be detected as early as the HSC/MPP population, the multidimensional viSNE plots of individual CD34+CD38lo cells from each sample were analyzed for phosphoprotein expression level to verify that the abnormal signaling was found in the immunophenotypically abnormal cells. This analysis is shown in Supplementary Fig. S5A–S5B. AML cells in the distinct regions of the viSNE plot associated with the FLT3-ITD+ immunophenotype appear to exhibit higher activation of pSTAT5 and pMAPKAPK2. In samples from patients with other AML subtypes, cells with either abnormally high or low levels of pSTAT5 or pMAPKAPK2 were also more likely to appear outside regions of the viSNE plot populated by the CD34+CD38lo cells of healthy donors (Supplementary Fig. S5A–S5B). Analysis of other measured signaling molecules in the CD34+CD38lo population is shown in Supplementary Fig. S6A–S6I. Thus, mass cytometry identified immunophenotypically aberrant HSPCs in AML patients and allowed visualization of abnormal intracellular signaling states in these same cells.

High-Dimensional Analysis of AML Samples Enabled Direct Assessment of Chemotherapy Response

To assess the feasibility of using mass cytometry to monitor chemotherapy responses, samples from 10 of the patients who were receiving oral hydroxyurea (HU; to control elevated blast count) were compared with samples from the 23 patients not receiving HU. Consistent with the expected reduction in cellular deoxyribonucleic acid pools induced by HU, samples from treated patients exhibited 80% to 90% suppression of IdU incorporation relative to untreated patients (Fig. 7A and B). However, HU treatment had minimal effect on the fraction of cells in S-phase across the myeloid immunophenotypic populations. A significant decrease was observed only in the GMP (3.44% vs. 1.25% S-phase cells; P = 0.016) and promyelocyte populations (10.1% vs. 2.98% S-phase cells; P = 0.0013; Fig. 7C). The CD34+CD38lo cell populations were least affected by HU treatment; there was no significant difference in S-phase fraction between samples from patients who were treated with HU and those who were not (2.72% vs. 2.86%; P = 0.21), in spite of the >6-fold difference in IdU incorporation in S-phase cells from these populations (median counts of 129 vs. 19.8, P = 4 × 10⁻⁴; Fig. 7B and C). In addition, the fraction of actively cycling cells (indicated by phosphorylated Rb) was unchanged or increased in all cell populations from patients treated with HU (Fig. 7D).

Levels of apoptosis (as indicated by cleaved PARP) did not differ in any of the immunophenotypic cell populations of patients treated with HU (Supplementary Fig. S7). Although in contrast with the complete S-phase arrest and cell-cycle exit commonly observed when leukemic cell lines are treated with HU in vitro (24), the analyses of patient cells shown here are much more consistent with the modest clinical effects observed in AML patients receiving HU in the clinic (25). This discrepancy suggests that cultured leukemic cell lines may represent only a subpopulation of AML differentiation states (e.g., promyelocytes) and that in vivo functional assays using mass cytometry will enable a more comprehensive understanding how patients respond to leukemia treatment.

DISCUSSION

High-Dimensional Cytometry Enables Characterization of the In Vivo Functional Properties of Malignant Cells

The incredible genetic and developmental heterogeneity of human malignancies creates critical questions in the study of many cancers. To what extent do oncogenic mutations disrupt developmental signaling programs in cells? Do these mutations lead to reproducible aberrant developmental patterns? To what extent is cell-cycle state driven by these different trajectories? Answers to these questions could help target therapies for both debulking of tumors and elimination of cells with stem cell-like capacities.

In this study, high-dimensional cytometric characterization of minimally manipulated samples from patients with AML provided a wide range of information regarding the in vivo functional properties of malignant cells. Combined immunophenotypic analysis and assessment of cell function demonstrated several advantages. First, in all samples, we observed immunophenotypically abnormal cells at all developmental stages through analyses of combinations of common immunophenotypic markers. Second, the characteristics of individual AML cells could be understood within the context of their developmental state, allowing comparisons to be performed after controlling for changes mediated by differentiation. Thus, more sensitive comparisons between different leukemia patient samples could be performed by using immunophenotypically gated subpopulations than would be possible if the total cell populations or blast cell populations were compared (this was particularly true for the S-phase fraction, H3K9ac, pMAPKAPK2). This enabled the detection of AML subtype-specific changes in cell-cycle state and surface marker expression among rare immunophenotypic HSPC populations that would be obscured by traditional analysis of total cell populations. These findings are consistent with a recent study of gene expression analysis that demonstrated improved prognostic accuracy when analysis was performed based on the differential expression patterns between AML populations.
and developmentally similar normal cells (26). Third, the integration of measurements of cell cycle and intracellular signaling enabled the real-time assessment of HU response and demonstrated that the chemotherapy response of AML cells in vivo is different in many ways from the in vitro response of cultured AML cell lines (Fig. 7), a difference that may be due to the inability of cell culture systems to model the differentiation that occurs in malignant cell clones in vivo.

Distinct Subtype-Specific Immunophenotypic and Functional Properties of AML LSCs

The immunophenotypic patterns of LSCs differed in samples from patients with different AML disease subtypes; these cell surface markers might represent targets for antibody-mediated therapies. Previous studies of aberrant surface markers on AML LSCs yielded inconsistent results, which may stem from the incorrect assumption that all AML LSCs have a similar phenotype (5, 27). In the patient cohort analyzed here, no one marker was specific for the presumed LSCs of all patients. Strong subtype-specific trends were apparent, however, particularly in FLT3-ITD+ NK–AML patients, in whom immunophenotypic HSPCs exhibited characteristic increases in CD33 (8 of 11 patients) and CD123 (all 11 patients, a finding also observed in all FLT3-TKD+ samples and three of four FLT3-ITD+ APL samples; Fig. 4A; Supplementary Table S4). Consistent with previous reports (28), CD99 was the most consistently elevated marker in the LSCs of AML samples in general (30 of 36 samples). CD99 levels were normal, however, in 6 patients, including a majority of those with the t(8;21) karyotype.

These specific patterns were even more pronounced in the 19-dimensional viSNE and binning analyses of the CD34+CD38lo population, which accurately grouped the FLT3-ITD+ AML samples on the basis of immunophenotypic information alone (Fig. 5; Supplementary Fig. S3). The observation that the only FLT3-ITD+ NK–AML sample that did not have this HSPC phenotype had lost the FLT3-ITD mutation at relapse suggests that the HSPC immunophenotype might...
be more accurate than binary genetic testing in certain clinical situations. These findings greatly extend previous reports of other subtype-specific immunophenotypes, such as an aberrant increase in CD7 in patients with double CEBPA mutations (29) and increased CD56 in patients with t(8;21) karyotype (findings also observed in this study; Supplementary Table S4). High-dimensional mass or fluorescent cytometry approaches should therefore be able to rapidly identify at least some AML subtypes with high accuracy if internal staining controls are used and analysis is restricted to the most predictive subgated populations (i.e., CD34+CD38-).

Critically, the AML subtype–specific immunophenotypic changes observed were detected using four independent analyses to define the developmental subpopulations: manual gating by standard surface markers (gates defined solely on the basis of the normal samples), 19-dimensional SPADE clustering, viSNE analysis, and a K-means–based cell binning approach. The immunophenotypic populations within each AML sample with the greatest number of immunophenotypic aberrancies also had: (i) the greatest expansion relative to the normal population frequency (Figs. 2 and 3), (ii) the largest abnormalities in cell-cycle state (Fig. 6), and (iii) aberrant intracellular signaling (Supplementary Figs. S4–S6; Supplementary Table S4). Taken as a whole, these observations indicate that high-dimensional analysis provides qualitatively different information from that obtained using standard diagnostic approaches and should enhance classification of malignant hematologic diseases.

High-Dimensional Cytometry Enables Assessment of AML LSC Capabilities In Vivo

To date, the functional properties of human leukemia stem cells have been almost exclusively studied in vitro or in the setting of murine transgenic (30) or xenograft models (4, 31–33). Although these groundbreaking studies led to the current understanding of stem cell biology and the complex clonal structure of AML, several important issues remain unresolved. Studies in which large series of patient samples have been engrafted into immunocompromised mice consistently demonstrated that engraftment capacity differs significantly across AML disease subtypes and may not reliably capture the LSC activity of all patients (31–33). In addition, the engraftment process has been shown to be cell cycle–dependent (34), greatly complicating the use of xenograft models for the study of LSC proliferation. The protocol utilized in this study circumvented these limitations by collecting whole bone marrow samples from patients with AML immediately at the bedside, and stringently limiting sample manipulation prior to fixation. Although it remains possible that changes in cell immunophenotype or functional properties could have occurred in the 15 to 20 minutes between bone marrow aspiration and sample fixation (a delay required for incubation with IdU; ref. 16), this report represents the closest measurement of the in vivo functional properties of AML cells possible with current technology of which we are aware.

One limitation of the mass cytometry technology is that the cells studied are destroyed during measurement, thereby precluding subsequent functional analyses. As a result, stem and progenitor cells had to be identified on the basis of previously established immunophenotypic criteria. This is reasonable given that the identification of stem cells was performed primarily on the basis of very well-established immunophenotypic criteria (lin-CD34-CD38hi; ref. 4), and a variety of methods for immunophenotypic identification were used, yielding similar results [manual gating, SPADE clustering (35), and viSNE analysis (19)]. Although controversy persists regarding the leukemia-initiating potential of partially differentiated leukemia progenitor cell populations (30, 36), the majority of comparative studies of LSC activity have demonstrated that CD34+CD38- leukemia cells have a similar or greater capacity to xenograft AML than other more mature populations (9, 30, 36–39). Even in AML subtypes in which CD34+ LSCs are present (40), CD34+ LSCs engraft leukemia more efficiently than CD34- LSCs (41) and are enriched in patients at relapse (6). Most relevant to the hypothesis of LSC-mediated AML relapse is the consistent finding that increased residual CD34+CD38- cells are predictive of leukemia recurrence (5, 6).

Finally, the LSC populations gated based on immunophenotypic criteria in this study exhibited functional properties similar to normal HSPCs (particularly S-phase fraction, H3K9ac, and pATM; Fig. 6; Supplementary Fig. S8; and Supplementary Table S4). Regardless of assumptions about stem cell immunophenotype, the data presented here demonstrate that patients with CBF–AML do not exhibit any identifiable immature cell fraction with a lower than normal S-phase fraction, whereas samples from AML patients with FLT3-ITD mutations had immature cell populations with very low S-phase fractions that could potentially mediate disease relapse.

AML Subtype–Specific LSC Cell-Cycle Patterns Correlate with the Known Responses to Chemotherapy

The data reported here indicate that the cell-cycle properties of immunophenotypic LSCs depend on AML disease subtype. Though preliminary, these results have important implications for the understanding of consolidation chemotherapy, which primarily serves to eliminate residual stem and early progenitor cell populations. Currently, high-dose, single-agent cytarabine is one of the most commonly used consolidation chemotherapy regimens in the treatment of AML, largely based on its superior efficacy for the treatment of core-binding factor AML (42). The cytotoxic effect of cytarabine, however, is almost exclusively restricted to cells in S-phase of the cell cycle (43). Thus, the finding that the S-phase fraction of LSCs in CBF–AML patient samples is approximately 5-fold higher than that in samples from patients with other AML subtypes (Fig. 6C) is consistent with the known responsiveness of this patient subset.

Conversely, despite relatively high rates of initial response to chemotherapy, patients with FLT3-ITD+ AML are not commonly cured with consolidation chemotherapy alone (13). The data presented in this report also fit well with this observation: The LSCs in FLT3-ITD+ AML patient samples would be expected to be resistant to cytarabine-based treatment by virtue of their low S-phase fraction (allowing disease relapse). Interestingly, the much more common myelo-monoblast cells in patients with FLT3-ITD+ AML exhibited the same S-phase fraction as samples from other AML subtypes and would thus be expected to be comparably sensitive to cytotoxic therapy (allowing for the initial responses generally observed in these patients; Fig. 6D). Unique to this study is the ability to directly compare individual immunophenotypic populations in AML with the same
Functional Mass Cytometry Analysis of AML Subtypes

RESEARCH ARTICLE

populations in normal human bone marrow. This yielded the surprising finding that the S-phase fraction of AML cells in vivo is actually lower than normal cells of the same developmental stage. Although our results vis-à-vis cell cycle are consistent with a variety of prior investigations, they are not conventionally appreciated in clinical practice. Our ability to detect these differences appears to require rapid processing of samples, as we have anecdotal observed changes in cell-cycle markers and basal signaling after as little as an hour ex vivo. Halogenated uridine analogues have previously been utilized in the study of both normal murine bone marrow (10, 44) and human leukemia myeloid cell populations (45) in vivo, and our results yielded S-phase fractions that are consistent with these studies.

Although the sample size of this study allowed detection of highly significant differences between common AML subtypes, it remains small relative to the high degree of genomic complexity of AML and somewhat overrepresentative of patients with higher-risk subtypes. In the 24 samples from this cohort that were also tested for an expanded panel of somatic mutations, several additional mutations were detected. Of particular note was the common occurrence of DNMT3A mutations in FLT3-ITD+ NK-AML patients (14, 15, 46), observed in 4 of the 7 FLT3-ITD+ NK-AML patients from this cohort (Supplementary Table S1). Due to the sample size, it was not possible to determine the additive effect of DNMT3A mutation in FLT3-ITD+ NK-AML (though the 3 patients with wild-type DNMT3A all exhibited similar immunophenotypes and low S-phase fractions). Mass cytometric analysis of a much larger patient cohort should allow the functional consequences of these complex genetic interactions to be addressed.

In Vivo Functional Profiling of AML to Guide Design of Novel Therapeutic Strategies

This study provides an example of how characterization of the in vivo functional properties of LSCs can potentially inform the development of novel therapeutic strategies. The finding that FLT3-ITD+ AML LSCs have a significantly decreased S-phase fraction relative to other AML subtypes suggests that single-agent, high-dose cytarabine is unlikely to be a curative consolidation therapy for these patients. As such, these results also suggest that two immunophenotypic targets, CD33 and CD123, could be exploited to treat these cells in a cell cycle-independent manner with antibody–drug conjugates (ADC). Consistent with this hypothesis are the results of the phase III ALFA-0701 AML clinical trial (46, 47) that used fractionated treatment with the CD33-directed ADC gemtuzumab ozogamicin (GO) in both induction and consolidation therapies. This trial demonstrated a significant improvement in OS (estimated 2-year OS of 36% vs. 64%; P = 0.023), specifically among NK-AML patients with FLT3-ITD mutations who received GO. In addition to explaining this observed benefit of CD33 targeting, the results reported here also predict that the addition of a CD123-targeted drug might further enhance LSC clearance and improve the survival of patients with FLT3-ITD+ AML.

Finally, this study establishes that high-dimensional cytometry can potentially be used to monitor disease treatment in “real time” and proximal to the acquisition of clinical material. Detection of aberrant signaling in FLT3-ITD+ AML and measurement of cell-cycle effects of HU treatment also demonstrate that monitoring the responses of patients with hematologic malignancies to both targeted and cytotoxic therapies is clinically feasible. This would be particularly useful for therapies with incompletely understood mechanisms of action and those that target LSCs or other rare cell populations. Because mass cytometry allows the simultaneous assessment of many bone marrow cell types, the efficacy of immunotherapy approaches (e.g., chimeric antigen receptor T cells) could also be monitored. As understanding of the underlying genetic complexity of hematologic malignancies expands, real-time, functional measurement of therapy response will be increasingly important for understanding which of the many genetic lesions present in these diseases represent the most relevant targets for therapeutic intervention and how to best utilize the novel agents designed to target them.

METHODS

Antibodies

Antibodies, manufacturers, and concentrations are listed in Supplementary Table S2. Antibody staining was performed in two overlapping panels, as indicated. Primary antibody transition metal conjugates were either purchased or conjugated using 200-µg antibody lots combined with the MaxPAR antibody conjugation kit (Fluidigm Sciences) according to the manufacturer’s instructions. Following conjugation, antibodies were diluted to 100× working concentration in Candor PBS Antibody Stabilization solution (Candor Bioscience GmbH) and stored at 4°C. Antibody CVs were calculated by comparing the replicate analyses of aliquots of the healthy donor samples.

Human Samples

Fresh bone marrow aspirates were collected immediately (<1 minute) after aspiration into heparinized tubes containing IdU (at a final concentration of approximately 20 µmol/L) and incubated at 37°C (16). After 15 minutes, samples were fixed with a fixation/stabilization buffer (SmartTube), according to the manufacturer’s instructions, and then frozen at –80°C for up to 36 months prior to analysis by mass cytometry. Samples were collected from patients at Stanford University Hospital who were undergoing routine diagnostic bone marrow aspiration and provided informed consent to donate a portion of the sample for tissue banking as part of a protocol approved by the Stanford University Institutional Review Board in accordance with the Declaration of Helsinki. The clinical characteristics of each patient are shown in Supplementary Table S1. Healthy control samples were obtained from Allcells using the same protocol. The exception was healthy donor sample #1; this sample was incubated for approximately 1 hour with IdU at 37°C. All healthy cell samples were collected into 15 µmol/L IdU. Bone marrow cell samples were thawed just prior to analysis in a 4°C water bath, and red cells were lysed using a hypotonic lysis buffer (SmartTube). Cells were then washed twice in cell staining medium (CSM; 1× PBS with 0.5% BSA and 0.02% sodium azide) at room temperature.

Antibody Staining

Prior to antibody staining, mass tag cellular barcoding was performed as previously described, (17, 18), and full details are provided in Supplementary Methods. Bar-coded cells were incubated with surface marker antibodies in 2 mL of CSM for 50 minutes with continuous mixing. Cells were washed twice with CSM, and then surface antibodies were fixed in place by a 15-minute incubation with 1.5% paraformaldehyde (Electron Microscopy Sciences). Cells were pelleted by centrifugation and resuspended with vortexing in ice-cold methanol, and red cells were lysed using a hypotonic lysis buffer (SmartTube). Cells were then washed twice in cell staining medium (CSM; 1× PBS with 0.5% BSA and 0.02% sodium azide) at room temperature.
Mass Cytometry

After completion of antibody staining, cells were washed twice with CSM and then incubated overnight (staining panel A) or for 36 hours (staining panel B) in PBS with a 1:5,000 dilution of the indium intercalator pentamethylecyclopentadienyl-Ir(III)-diporphophenazine (Fluidigm Sciences) and 1.5% paraformaldehyde. Excess intercalator was then removed with one CSM wash and two washes in pure water. Cells were then resuspended in pure water at approximately 1 million cells per ml and mixed with mass standard beads (Fluidigm Sciences). Cell events were acquired on the CyTOF mass cytometer (Fluidigm Sciences) at an event rate of 100 to 500 events per second with instrument-calibrated dual-count detection (49). Noise reduction, a cell length of 10 to 90, and lower convolution threshold of 200 were used. After data acquisition, the mass bead signal was used to correct short-term signal fluctuation during the course of each experiment, and bead events were removed (50). Approximately 240,000 cell events were collected for each sample in each of the two staining panels (480,000 total events per sample; Supplementary Table S5).

Immunophenotypic Aberrancy Analysis

Aberrant immunophenotype analysis was performed by first gating the normal populations into developmental immunophenotypic subsets on the basis of standard surface markers (as in Supplementary Fig. S9). AML cells in each population were then compared with the normal samples across the 28 surface markers. Because the normal populations and AML samples were all stained and analyzed in the same tube simultaneously, the same gates were used to identify immunophenotypic populations from each of the AML samples. The median expression level of each marker in each gated population from each patient sample was then calculated. As the 14 replicate normal samples analyzed came from only five healthy donors, AML sample aberrancy was defined conservatively as an AML sample median expression level greater than or less than the median of the similarity bone marrow population plus or minus two times the absolute variance of the healthy control samples. For example, overexpression of CD33 in the MPP population of AML sample #23 was determined as follows: AML#23CD33 median CD33 expression > normal sampleCD33 median CD33 expression + 2 × absolute variance of normal sampleCD33 expression. This process was repeated for each of the other measured surface markers in each of 35 gated immunophenotypic populations in every AML sample. The summed number of markers with aberrant expression patterns was calculated for each gated immunophenotypic population from each patient sample.

Disclosure of Potential Conflicts of Interest

G.K. Behbehani has received funds from the speakers bureau of Fluidigm Corporation and is a consultant/advisory board member for the same. G.P. Nolan has ownership interest (including patents) in Fluidigm and is a consultant/advisory board member for the same. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: G.K. Behbehani, W.J. Fantl, B.C. Medeiros, G.P. Nolan

Development of methodology: G.K. Behbehani, W.J. Fantl

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): G.K. Behbehani, B.C. Medeiros

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): G.K. Behbehani, N. Samusik, Z.B. Bjornson, W.J. Fantl, B.C. Medeiros

Writing, review, and/or revision of the manuscript: G.K. Behbehani, Z.B. Bjornson, W.J. Fantl, B.C. Medeiros, G.P. Nolan

Study supervision: G.P. Nolan

Grant Support

G.K. Behbehani was supported by a developmental research grant from the Stanford Cancer Center. G.P. Nolan was supported by NIH/NCI grants U19 AI057229, 1U19AI100627, U54 CA141945, SU54CA149071, 1R01CA130826, 5R01AI073724, R01 GM109836, R01CA184968, R101MS088953, P01 CA163233, R33 CA183654, R33 CA183692, 41000412117, 201303028, N01-HV-00242, HHSN268201000034C, HHSN272201200028C, and HHSN272200700038C; CRIRM DR1-01477; Department of Defense OCI110674 and 11491122; FDA: HHSF223201210194C-FDA-RAA-12-00118; and the Bill and Melinda Gates Foundation (GF12141-137101; OPP1113682). G.P. Nolan was also supported by the Rachford and Carlota A. Harris Endowed Professorship.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 10, 2015; revised June 16, 2015; accepted June 17, 2015; published OnlineFirst June 19, 2015.

REFERENCES


Functional Mass Cytometry Analysis of AML Subtypes


# Mass Cytometric Functional Profiling of Acute Myeloid Leukemia Defines Cell-Cycle and Immunophenotypic Properties That Correlate with Known Responses to Therapy

Gregory K. Behbehani, Nikolay Samusik, Zach B. Bjornson, et al.


<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/2159-8290.CD-15-0298</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://cancerdiscovery.aacrjournals.org/content/suppl/2015/06/30/2159-8290.CD-15-0298.DC1">http://cancerdiscovery.aacrjournals.org/content/suppl/2015/06/30/2159-8290.CD-15-0298.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 50 articles, 19 of which you can access for free at: <a href="http://cancerdiscovery.aacrjournals.org/content/5/9/988.full#ref-list-1">http://cancerdiscovery.aacrjournals.org/content/5/9/988.full#ref-list-1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 15 HighWire-hosted articles. Access the articles at: <a href="http://cancerdiscovery.aacrjournals.org/content/5/9/988.full#related-urls">http://cancerdiscovery.aacrjournals.org/content/5/9/988.full#related-urls</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-mail alerts</th>
<th>Sign up to receive free email-alerts related to this article or journal.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprints and Subscriptions</td>
<td>To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a>.</td>
</tr>
<tr>
<td>Permissions</td>
<td>To request permission to re-use all or part of this article, use this link <a href="http://cancerdiscovery.aacrjournals.org/content/5/9/988">http://cancerdiscovery.aacrjournals.org/content/5/9/988</a>. Click on &quot;Request Permissions&quot; which will take you to the Copyright Clearance Center's (CCC) Rightslink site.</td>
</tr>
</tbody>
</table>