Functional Proteomic Profiling of AML Predicts Response and Survival

Steven M. Kornblau1,6, Raoul Tibes2,6, YiHua Qiu1, Wenjing Chen1, Hagop M. Kantarjian,3 Michael Andreeff,1 Kevin R. Coombes,4 Gordon B. Mills,5

1Departments of Stem Cell Transplantation and Cellular Therapy, 3Leukemia, 4Bioinformatics and Computational Biology and 5Systems Biology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA and 2The Translational Genomics Research Institute (TGen), Phoenix/Scottsdale, Arizona 6Both authors contributed equally to this work.

Address requests for reprints and all communications concerning this manuscript to Dr. Kornblau, Section of Molecular Hematology and Therapy, Unit 448, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030-4095.
E-mail to skornblau@mdanderson.org. Tel: (713) 794-1568, Fax: (713) 794-1938

Running Title: Proteomic Profiling using Reverse Phase Protein Arrays in AML

Key words: Proteomics, Reverse Phase Protein Arrays, RPPA, AML,
Abstract

Since protein function regulates the phenotypic characteristics of cancer, a functional proteomic classification system could provide important information for pathogenesis and prognosis. With the goal of ultimately developing a proteomic based classification of acute myeloid leukemia (AML), we assayed leukemia-enriched cells from 256 newly diagnosed AML patients, for 51 total and phospho-proteins from apoptosis, cell cycle, and signal transduction pathways, using reverse phase protein arrays (RPPA). Expression in matched blood and marrow samples were similar for 44 proteins; another 7 had small fold changes (8-55%), suggesting that functional proteomics of leukemia-enriched cells in the marrow and periphery are similar. Protein expression patterns were independent of clinical characteristics. However, 24 proteins were significantly different between FAB subtypes, defining distinct signatures for each. Expression signatures for AML with cytogenetic abnormalities involving –5 or -7, were similar suggesting mechanistic commonalities. Distinct expression patterns for FLT3-ITD were also identified. Principal component analysis defined 7 protein signature groups, with prognostic information distinct from cytogenetics that correlated with remission attainment, relapse, and overall survival. In conclusion, protein expression profiling patterns in AML correlate with known morphological features, cytogenetics and outcome. Confirmation in independent studies may also provide pathophysiological insights facilitating triage of patients to emerging targeted therapies.
Introduction

The current classification of Acute Myeloid Leukemia (AML) uses the French-American-British (FAB) system based on morphological features, along with flow cytometric analysis of surface marker expression, cytogenetics, and assessment of recurrent molecular abnormalities. These classification schemes have prognostic relevance but they generally do not alter therapeutic recommendations. Furthermore, current prognostic models, based on clinical and laboratory features, have low predictive power explaining less than half of the outcome. Abnormal activation of signal transduction pathways (STP), aberrant cell cycle regulation, and evasion of apoptosis are key events in malignant transformation. Expression of STP proteins is heterogeneous and prognostic in AML. Distinct molecular abnormalities and patterns of pathway activation in leukemic cells combine to suggest potential targets for therapeutic intervention. However, unlike chronic myelogenous leukemia, where all cases start with the same chromosomal abnormality, AML is markedly heterogeneous with numerous genetic aberrations. Consequently, knowledge facilitating individualization of targeted therapies in development in AML is sorely needed. Thus an improved understanding of leukemic cell biology might lead to improved classification schemes that more accurately explain the heterogeneity of response and which thereby guide the use of targeted therapies on an individualized basis.

Novel array technologies enable the analysis of numerous features at the level of DNA copy number, mutations, methylation, mRNA transcription, regulatory microRNA as well as emerging approaches to assess protein expression levels within cells and cytokine and chemokine arrays to assess external forces acting on leukemic cells.
Gene expression profiling arrays have demonstrated the ability to provide new classification schema and to define prognostic subgroups in AML\textsuperscript{11,12}. The contribution of these approaches to patient management and to understanding the pathophysiology of AML remains to be fully elucidated. Transcriptional profiles can provide superior tumor classification, response and prognostic information. However, RNA transcript levels often do not correlate with protein expression\textsuperscript{13-15} and differentially expressed genes may not necessarily be involved in the pathogenesis of a disease. Protein expression and post-translational modifications either alone or in concert with other profiling approaches could provide independent or complementary information not captured by transcriptional profiles. Thus the functional proteome provides an untapped information source, which may contribute prognostic information and provide important pathophysiological cues regarding response to current or emerging therapies as well as potential new targets for therapy. It is therefore critical to determine whether protein levels and post-translational modifications can provide novel determinants of cellular phenotype and biological behavior. Expression levels and activation of single, or a few proteins have been studied\textsuperscript{5-9,16} but comprehensive analysis of the complexity of intracellular STP is lacking due to lack of a feasible, quantitative, high throughput proteomic technology applicable to patient samples.

To generate a comprehensive proteomic profile of the level of expression and activation of apoptosis, STP, and cell cycle regulating proteins in AML, we utilized reverse phase protein arrays (RPPA), an approach where protein lysates from multiple samples are printed onto a slide and probed with a single antibody to generate a quantitative output.\textsuperscript{17-19} Proteins and their corresponding phosphoproteins can be
assessed, reflecting the activation state and functionality of a given protein, pathway, or network. The technique offers high sensitivity, throughput, inter- and intra-slide (array) reproducibility, thereby overcoming the limitations of conventional protein assay techniques (Western blotting or ELISA assays). The usefulness, high precision, and reliability of RPPA for analysis of leukemic specimens has been demonstrated.

In this report, we present a set of 320 samples from 256 newly diagnosed, AML patients profiled for 51 total and phosphoprotein epitopes involved in STP, apoptosis and cell cycle regulatory pathways. These protein expression signatures provide prognostic information, and could potentially guide the selection of targeted therapies.

Methods

Patient Population

This study utilized a total of 539 Bone marrow (BMA), peripheral blood (PB) or pheresis (PH) specimens from 442 patients with AML including 97 cases with a same day BMA and PB or PH sample. These samples were collected at diagnosis (320 samples from 256 patients), in a primary refractory state (59 samples from 47 patients), in relapse (138 samples from 116 patients), in remission (9 samples from 8 patients) and while receiving therapy (n=7). The same day BMA and PB or PH samples permit comparison of expression between blood and marrow. This report is restricted to the 256 newly diagnosed patients: the associated demographics are described in Table 1. Samples were collected for the Leukemia Sample Bank at the University of Texas M.D. Anderson Cancer Center between 1/15/1998 and 3/9/2006, on Institutional Review Board approval.
Board (IRB) approved protocol Lab01-473 and consent was obtained according to the Declaration of Helsinki. Samples were analyzed under an IRB approved laboratory protocol (Lab05-0654).

Among the 256 newly diagnosed, previously untreated (New) patients, 120 were treated with anthracycline and high dose ara-C (HDAC) based regimens, 49 on HDAC plus non anthracycline chemotherapy regimens (FLAG, FA, Clofarbine etc), 21 with other non HDAC based chemotherapy, 10 with targeted agents, 10 with Gemtuzumab Ozogamicin (mylotarg) based regimens, 6 with demethylating or histone deacetylating agents, one with low dose ara-c, one with stem cell transplant as initial therapy and 40 were not treated at MDACC.

Specimen Collection and Processing

Immediately after collection, samples were placed on ice and processed within 2 hours by Ficoll-Hypaque gradient separation to yield a mononuclear cell fraction. The isolated mononuclear cell fraction containing blasts was further enriched by depletion of CD3+/CD19+ B- and T-lymphocytes using magnetic antibody-conjugated sorting (Miltenyi Biotech, Auburn, CA) as previously described. The purity of the leukemic samples was generally greater than 80% blasts after enrichment. After isolation and washing, boiling hot protein lysis buffer (2x, 0.5 mol/L Tris-HCl (pH 6.8), 2% SDS, 10% glycerol and 4% β-mercaptoethanol) was added to make a whole-cell lysate which is then aliquoted into single use vials before cryopreservation. We have previously demonstrated that protein made from cryopreserved AML cells yield identical results on
RPPA with protein made from fresh cells on the day of collection, demonstrating that protein levels are not adversely affected by cryopreservation of cells or protein lysates.18

Controls, Reference Standard, and Cell Lines

To correct for staining, background, and loading variation across (array) slides, a positive control consisting of a mixture of 11 cell lines hereafter called “pooled control” lysate and the lysate buffer as negative control were utilized. For expression controls 18 cell lines, including baseline and growth factor or cytokine stimulated samples were used and shown on the schematic by the letter “C” (Fig. 1A). Eighteen peripheral blood samples from healthy volunteers (IRB approved collection) served as normal controls, shown by the letter “N” in figure 1. To permit absolute quantification of signal strength, we included 138 purified peptides from sequences used to generate antibodies in the study as well as additional controls, shown by the purple band encircling the patient samples.

Array Assembly and Printing Method

For quantification purposes, 5 serial dilutions (1:3 dilution steps) of each protein lysate were arrayed in 384 well plates (Genetix, Boston, MA). Samples were printed onto nitrocellulose coated glass slides (FAST Slides, Schleicher & Schuell BioScience, Inc. USA, Keene, NH) using an Aushon Biosytems 2470 Arrayer (Aushon BioSystems Inc., Burlington, MA) with 175 micron pins and a single touch. The samples were printed in replicate, one centrally located and the other split on either side and arranged in a reversed orientation (Fig. 1A). Based on the sample concentration of $1 \times 10^4$ cells per
microliter and a printing volume of 2 nanoliter/touch, we estimate that the spots ranged from 85 cell equivalents of protein in undiluted, with ~ 1 cell protein equivalent in the most diluted (1:81) spot (Fig. 1B). To permit topographical normalization a sample of pooled or negative control sample was printed at the end of each row of patient sample, creating a grid across the whole slide which appears as alternating spots in the sample array (Fig 1C). Each slide contained 7968 dots. A representative slide is shown in figure 1C.

**Antibody Detection and Array Staining**

A detailed description of the array methodology including antibody staining and detection has been published\(^1^8\). Briefly, after printing, slides were incubated for 15 min in biotin blocking solution to block endogenous peroxidase, avidin, and biotin prior to incubating slides in protein block at 4\(^0\) C over night. Primary antibodies (AB) in concentrations from 1:250-1:2000 were added for 1-2 hours with frequent rotation (see Supplementary Table 1 for dilutions and manufacturers of AB). A biotinylated secondary antibody (anti-mouse or –rabbit), diluted 1:10000-1:20000, used as starting point for signal amplification, was added for 1 hour. Subsequently, array slides were incubated using the DAKO (Copenhagen, Denmark) signal amplification system using catalyzed reporter deposition of substrate to amplify the signal detected by the primary antibody\(^2^2\). Slides were incubated in streptavidine-biotin-peroxidase and biotinyl-tyramide/hydrogen peroxide reagents each for 15 min with frequent washing in between. Finally, 3,3'-diaminobenzidine tetrachloride (DAB) was cleaved by tyramide-bound horseradish peroxidase, giving a stable brown precipitate with excellent signal-to-
noise ratio. This technique is sensitive and reproducible to the femtomolar range as reported.18;18;23

Fifty-two proteins were assayed including 30 total and 21 phosphoproteins (Supplemental Table 1) using antibodies validated as described previously.18 We have utilized a naming system where the protein name always comes first, followed by description of any modification (e.g. “p “for phosphorylation), followed by the amino acid site of the modification when more than one phosphorylation site was analyzed (e.g. AKT.p308 vs. AKT.p473). This differs from the convention of using a lower case “p” before the protein name. This system facilitates alphabetical sorting and provides additional detail regarding modification site. Many of these same samples (n=125) were used in our prior study looking at STP expression in AML by Western blot(WB).7 There was significant correlation between the RPPA and the WB results for ERK (p=0.03), ERK.p (p=.003), AKT.p473 (p=0.02) and PKCα (p=0.04) using the same antibodies.

Data Analysis, Normalization, and Statistics

Hybridized slides were scanned on a desktop scanner (HP, Sunnyvale CA) at an optical resolution of 1200 dpi in 16 bit grayscale and saved as TIFF files. Protein expression intensity of each spot was measured with an automated software program MicroVigene™ (VigeneTech, Inc., North Billerica, MA). The dilution series of the samples provide a dilution-concentration-expression curve providing relative expression intensities. These numbers were used for data processing and calculation after standardization and topographical normalization. All analyses were performed using the R Statistical Programming Environment, version 2.4.0.
Topographical Normalization

Positive and negative controls were positioned across the slide at the end of each sample as described above. This enables normalization to adjust for variations in background staining. Negative controls (lysis buffer) set the baseline noise for background correction and positive controls (cell lines, peptides) provide expression intensities to set the scale for protein quantification in the samples. Background variation was variable, as shown in the topographical maps ranging from negligible to the most extreme (Fig. 1D,1E) After subtracting local background estimates from the MicroVigene quantifications, we subtracted the median background-corrected values of the negative controls from all spots. This ensured that the negative controls were centered around zero and resulted in data where replicate spots yielded more reproducible values and where the slope of the linear part of the concentration curves more closely matched the known dilution step. Next, the median intensity, $M$, of the positive controls was computed, and each dilution series was normalized by dividing by $M$ times the intensity of the nearest positive control. Other processing methods were evaluated, but this procedure yielded superior results, resulting in the highest correlation ($R^2 > 0.8$) between the replicates on a slide. Correlation between sample 1 and sample 2 on the slide was very high with $R > .7$ for 59%, $> .5$ in 92% of the proteins studied after topographical background correction. A manuscript by Neeley et al with a more detailed description of the algorithm is in preparation. The average multipliers for correction for loading and regional staining usually ranged between 0.624 and 1.376.
Summarization of Dilution Series

After preprocessing, dilution series were summarized by fitting a joint four-parameter logistic model as described by Tabus et al.\textsuperscript{24} We used our own implementation of this algorithm in an R package, SuperCurve 0.931 (http://bioinformatics.mdanderson.org/Software/OOMPA).

Various clustering methods were employed. Validity of clusters was assessed using perturbation bootstraps\textsuperscript{25} and the gap statistic.\textsuperscript{26} Associations between protein expression levels and categorical clinical variables were assessed in R using standard t-tests, linear regression, or mixed-effects linear models. Conservative Bonferroni corrections were performed to account for multiple statistical parameters (numbers of samples and proteins/antibodies) when calculating statistical significance.

The dataset is available at (http://bioinformatics.mdanderson.org/supplements.html, under “RPPA Data in AML”) or as an excel file from skornblau@mdanderson.org.

Results

Blood and Marrow Specimens can be used Interchangeably

This array contained same day blood/pheresis and marrow samples from 97 patients permitting a direct comparison of protein levels in blast cells from blood and marrow. If expression was significantly different depending on the source then separate analysis would be required for blood and marrow. In contrast if they were similar then all
the samples could be combined into a single analysis. Similar to our previous observations in Western blot\textsuperscript{27} and RPPA\textsuperscript{18}, linear mixed effects models revealed no statistically significant differences in expression intensities between blood and marrow for most (44/51) proteins. However, 7 proteins had statistically significant differences (Table 2); 3 were higher in blood and 4 higher in marrow. While these differences were statistically significant after Bonferroni correction, four had absolute fold differences < 20\% making it unlikely that these differences have biological relevance. The others, pMTOR, pSRC, and pS6RP, had fold differences ranging from 43 to 55\%, which might reflect proliferation potential (see discussion). Consequently blood and marrow derived samples were combined for the subsequent analysis. Replicate measurements from the same patient were averaged, with the 7 proteins where blood and marrow levels differed being adjusted to remove this small effect.

**Correlation of Protein Profiles with Clinical Characteristics**

No correlation between individual protein expression levels and traditional clinical characteristics (e.g. age, gender, race, infection status, performance status, antecedent hematological disorder, history of prior malignancy, radiation or chemotherapy, WBC, platelet count, hemoglobin, creatinine, bilirubin, albumin, percent blood and marrow blasts, percent cells expressing CD 7, 10, 13, 19, 20, 33, 34, cytogenetics, and treatment regimen) was observed (data not presented).
Correlation of Protein Profiles with FAB Subtypes

Since the FAB classification distinguishes leukemias based on differentiation pathways and the degree of maturation, we speculated that there might be unique protein expression signatures for different FAB types. We used linear regression, with individual proteins, to perform a one-way ANOVA with Bonferroni correction and observed that 23 proteins showed significant heterogeneity in protein expression levels between FAB classes. Since the ANOVA does not describe which FAB classes have significantly different expression levels for a protein, we used Tukey’s test for honestly significant differences (HSD) to determine which mean differences between classes were statistically significant. Next, we clustered the significant proteins based on their patterns of protein expression across the newly diagnosed AML patients and produced a heat map of their (standardized) average expression by FAB category (Figure 2). We found three clusters of proteins that tended to track similarly within a FAB class. The first cluster was enriched for total and phosphorylated signal transduction proteins (PKCA, PKCA.p, ERK2, AKT.p308, P38.p P70S6K, P70S6K.p, and Src.p527); proteins in this cluster were primarily characterized by lower expression in M0, M1, M2, with higher levels in the other FAB types. The differences between early myeloid (M0, M1, M2) were statistically significant different in comparison to FAB subtypes with a monocytic component (M4, M5). The second cluster contains PTEN and PTEN.p; it is characterized by significantly lower expression in M6 and M7. The third cluster is enriched for apoptosis (BAD, BAK, BCL2, BCLXl, BAD.p136, SMAC), cell cycle or differentiation regulating proteins (Myc, CCND1, SSBP2) and activated STAT proteins. In general, these proteins have significantly higher expression in the myeloid subtypes.
(M0, M1, and M2) compared to the monocytic (M4, M5), erythroid (M6) or megakaryocytic (M7) subtypes. Within this group, β-catenin (BCAT) and BCLX\textsubscript{L} exhibit a slightly different pattern, with significantly higher expression in M6 or M7.

**Expression Patterns and Cytogenetics**

Recurrent cytogenetic abnormalities currently provide the strongest available predictors of therapy response and outcome in AML.\textsuperscript{28} To define the relationship between protein expression and cytogenetic groups, we computed the average protein profile for each cytogenetic profile, then clustered these profiles using complete linkage and Pearson correlation (Figure 3). Based on the gap statistic,\textsuperscript{26} there are five clusters in the data: (i) a cluster driven by changes involving chromosomes 5 or 7, (ii) Ph1, (iii) t(8;21), 11q23, and IM, (iv) the complex karyotype of simultaneous –5, -7, and 11q23, and (v) a miscellaneous cluster including diploid, t(6;9), inv(16), +8, and +21. AMLs with cytogenetic aberrations are associated with distinct protein patterns that differ from the mean. In contrast, the diploid and miscellaneous AMLs show an “average” rather than a specific expression pattern, likely reflecting heterogeneity in these cases.

As with the FAB subtypes, we performed gene-by-gene ANOVAs to determine which individual proteins were differentially expressed in patients with different cytogenetic abnormalities. After Bonferroni correction, only two proteins were found to be significantly different: TP53 (\(p = 1.32\times10^{-13}\)) and BAX (4.73x10\(^{-7}\)) across all groups. TP53 was significantly higher whenever there was an abnormality of chromosome 5 or chromosome 7. BAX, by contrast, had three distinct levels of expression: low when there were abnormalities of 5 or 7, intermediate in 11q23 (even in the presence of...
losses of 5 and 7), inv(16), diploid, or miscellaneous samples, and high in all other categories. Additionally, cases with a FLT3-ITD were observed to have significantly higher expression of pMTOR (P<0.0001), pStat5 (P<0.001) AKT, Badp136, BAX, MEK, MTOR, pPTEN, pStat1 (P<0.01) and MYC, PKCα, pPKCα, SRCp527 (P<0.05). Association between FLT3-ITD and Stat5 phosphorylation\textsuperscript{29,30} and PKCα mRNA expression\textsuperscript{31} have been previously described.

Principal Component Analysis and Protein Signatures Groups

We hypothesized that proteins would form “constellations” in which the components provided predictive information. The most commonly used similarity measures (Euclidean distance or Pearson correlation) treat negatively correlated variables as highly distinct. To deal with this issue, we clustered the proteins using a distance \( d = (1 - |\rho|) \) based on the absolute value of the Pearson correlation, \( \rho \). This clustering suggested the existence of 10 protein constellations (CNSTN) (Figure 4 and Supplementary Figure 2). In many constellations, the proteins have related functions supporting the idea that the approach has functional validity. For example, CNSTN-9 contain phosphorylated signaling pathway members TP38.p, MEK.p, ERK.p, and inactivating phosphorylation of two of their targets, GSK3.p and BAD.p\textsuperscript{112}; CNSTN-6 contains the cell survival signaling proteins AKT.p308 and AKT.p473 inversely correlated with apoptosis regulating proteins BLC2, Bad and SMAC, ; CNSTN-1 has apoptosis related proteins survivin, XIAP and Bak in a positive correlation with cyclinD1 and negative correlation with phosphorylation of S6 which would promote transcription and proliferation; and CNSTN-4 has phosphorylated forms of STAT1, 5 and 6. Other
constellations contain proteins active in separate pathways that might have overlapping functions; e.g., CNSTN-8 has both P53 and MCL1, which might be expected to affect apoptosis via separate mechanisms.

To integrate and summarize the information from negatively and positively correlated proteins within a constellation, we performed a principal components analysis on each constellation retaining the first principal component. For each constellation and patient, this method computes a score, which is a weighted sum of protein levels, which describes the extent to which the constellation is present in the protein expression pattern of that patient. For example, in the first principal component from CNSTN 7, the proteins SRC and SRC.p527 get negative weights (-0.83 and -0.47, respectively), while S6 and SSBP3 get positive weights (0.10 and 0.27, respectively). We then clustered the patients, using Pearson correlation and complete linkage, based on these scores. Based on the gap statistic, the scores separate the patients into 7 protein signature groups (SG), although there is some indication that the high level split into two groups (Group A = SG1, SG2, and SG6 and group B = SG3, SG4, SG5, and SG7) gives the most reliable separation (Figure 4). The heatmap showing the average normalized expression for each protein, within each SG, shows that in some SGs all proteins within a particular constellation are positively correlated (e.g., CNSTN-5, the GSK3, BCL-XL, TP27, BCAT, DJI group) while in others some proteins are positively correlated with some members and negatively correlated with others (e.g., CNSTN-6 with P70S6K.p, P70S6, AKT.p308, AKT.p473, which correlate with each other, but are negatively correlated with other constellation members BAD, SMAC, SSBP2, BCL2). This
arrangement allows for better visualization of relative over and under-expression proteins within each of the 10 protein constellations within each of the 7 SG.

The distribution of patients with favorable, intermediate, or unfavorable cytogenetics among the 7 SG was significantly uneven (Chi-square test, $\chi^2 = 40.63$ on 12 degrees of freedom, $p = 5.66 \times 10^{-5}$; Table 3.). Patients with favorable cytogenetics were significantly over-represented in SG2 and underrepresented in SG7. Patients with intermediate cytogenetics were over-represented in SG3 and SG6 and under-represented in SG5. Patient with unfavorable cytogenetics were over-represented in SG5 and SG7 and under-represented in SG3.

**Effect of Signature Group on Remission, Relapse and Survival.**

Although the complete response rates within the 7 SG ranged from 42% to 73%, these differences were not statistically significant by a Chi-square test ($\chi^2 = 8.04$ on 6 degrees of freedom, $p = 0.24$). Rather the combine groups arbitrarily we analyzed the data using a hierarchical Bayesian model. The response within signature group $SG_i$ was assumed to be binomial with rate $\theta_i$. The response rates were assumed to arise from a common beta distribution $\theta_i \sim \text{Beta}(\alpha, \beta)$. We used an uninformative prior on the natural transformed scale:

$$p(\log(\alpha/\beta), \log(\alpha + \beta)) \propto \alpha \beta (\alpha + \beta)^{-3/2}$$

For details on the model, see section 5.3 of Gelman et al. We found that SG7 (which is enriched for unfavorable cytogenetics and depleted of favorable cytogenetics, with an overall CR rate of 42%) has a posterior probability of at least 80% of having a lower CR rate than SG2, SG3, SG4 or SG6, and at least a 69% posterior probability of having a
lower CR rate than SG1 or SG5. We also found that SG2 (which is enriched for favorable cytogenetics, with an overall CR rate of 72%) has a posterior probability of at least 70% of having a higher CR rate than every other group except for SG6. The relapse rates, which ranged from 37% in SG3 to 75% in SG6 were not statistically significantly different by a Chi-squared test ($\chi^2 = 7.79$ on 6 degrees of freedom, $p = 0.25$).

Next, we looked at associations between the SG and overall survival. Using a Cox proportional hazards model, we found a significant difference in overall survival (log rank test; score = 16.6 on 6 degrees of freedom; $p = 0.0109$) (Figure 5A). We found that patients in signature groups SG2, SG3, and SG4 had the best survival (hazard ratios, $1 \leq HR \leq 1.28$) while patients in SG5 (HR = 2.29), SG7 (HR = 2.14) and SG6 (HR= 2.13) had the worst survival.

Having observed that SG were associated with cytogenetics, we also evaluated a Cox proportional hazards model that used both cytogenetics and signature groups as predictors of survival. In the joint model, cytogenetics remained highly significant ($p=4.18\times10^{-7}$) while the sample group showed weaker evidence of significance ($p=0.13$). When stratified by cytogenetics the response rate varied by SG for patients with intermediate (range 40-67%) and unfavorable cytogenetics (range 30-75%) (Table 4), but not for those with favorable cytogenetics where all patients achieved remission. For all three cytogenetic groups the relapse rates differed within the different SG (Table 4). As might be expected these SG based differences in remission and relapse rates combined to result in different survival outcomes when the effect of SG on survival within each cytogenetic group was determined. Within cytogenetics categories, SG was
not significant However, among FLT3-ITD positive diploid cases the protein expression signature was still prognostic (P<0.0007, ) with patients in SG2 having the best survival, SG3 having intermediate survival, and the other groups having extremely poor survival (Figure 5B).

**Discussion**

This study has demonstrated that proteomic profiling of AML provides significant details about AML that complement and expand on known classification and prognostic information. We utilized RPPA because of its high sensitivity, quantitation and high throughput capabilities. The sensitivity of RPPA is evident by the ability to frequently see signal from the fifth dilution, corresponding to the protein equivalent of a single cell. The high throughput capability is demonstrated by the large number of samples analyzed on a single slide (N=539) and the fact that the 51 slides were stained, scanned, and analyzed in a 10 day period. Several observations emerge from this dataset.

Protein levels in AML cells do not appear to exhibit biologically relevant differences between blood and marrow derived specimens for most proteins. The presence of 97 same day blood and marrow specimens permitted direct comparison of expression in these two compartments. The majority of proteins examined (44 of 51) had statistically indistinguishable levels of expression. The fold differences for BAD, BAK, MTOR.p and survivin, while statistically different, are sufficiently modest (ranging from 1.08 to 1.20) that they are unlikely to have biological impact. There may be biological relevance to the observations that levels of total SRC and SRC.p527, an
inactivating phosphorylation, are 25 to 28% higher in blood and levels of activated S6RP, phosphorylated on amino acids 235 or 240-244, was 40 to 55% higher in marrow. Together these suggest that circulating cells are in a lower proliferative and transcriptional state compared to marrow bound leukemic blasts. This finding supports our previous observations in 8 proteins using Western blotting and by others with GEP. The similarity between marrow and circulating cells indicates that either can be utilized for proteomic analysis of AML provided that similar methodology is utilized to generate a leukemia enriched fraction. For patients with circulating leukemic blasts this could obviate the need to undergo bone marrow aspiration with its associated procedure costs and discomfort.

Proteomic profiling identified differences between subtypes of AML in the FAB classification schema, which classifies AML, based on lineage differentiation and maturation. A subset of 24 proteins showed differences in expression across the FAB classifications and the patterns of expression easily distinguish purely myeloid subtypes M0, M1 and M2 from M4 and M5 subtypes with monocytic components, and all of these were distinct from erythroleukemia and megakaryocytic leukemia. This data might be used to suggest treatment targets and guide selection of targeted therapies in different FAB subtypes. For example, expression of phospho-Stat 1 and 5 and several apoptosis related proteins were higher in M0-M2 subtypes relative to M4 and M5. In contrast expression levels of phosphorylated signal transduction proteins AKT, TP38, PKCo and SRC.p527 were higher in the monocytic subtypes. These patterns suggest that chemoresistance within FAB M0-M2 leukemias may be more dependent on resistance
to apoptosis, while growth and resistance to therapy in M4/M5 subtypes might be more dependent on the proliferative and antiapoptotic signals derived from activation of STP.

Similarly, different patterns of protein expression distinguished the major recurrent cytogenetic abnormalities observed in AML. Historically patients with abnormalities on chromosome 5 and 7 have similar poor response rates and survival with current therapy. Hierarchical clustering of proteomic data demonstrated that -5, 5q-, -7 or the combinations of -5 and -7 or -5, -7 and +8, had similar protein expression patterns. A similar clustering of -5 and -7 were observed with GEP. Thus it appears that the different cytogenetic changes have similar effects on protein expression and activation within AML blasts. Dissecting the protein patterns and subjecting them to pathway analysis may yield direct insight into the transforming mechanism(s) underlying these cytogenetic abnormalities. Of note is that P53 expression was highest in these patients, a group shown to have the highest rates of P53 mutations in AML, suggesting that use of agents targeting P53 activity or expression, such as MDM2 agonists might be active in these patients. Other examples are a generally higher survivin and GSK3 expression in AML with -5, -7 abnormalities. Survivin is marker of poor prognosis in advanced stages of AML. Survivin inhibitors might be best tested in this group (chromosome 5 and 7 abnormalities) first. Despite having favorable cytogenetics many patients with inv16 or t(8;21) relapse and die. Differences in protein expression patterns between long term survivors and those that ultimately relapse may reveal the underlying differences that result in this dichotomous outcome.

The third major observation is that the overall pattern of protein expression can be used to classify patients into 7 signatures groups based on the summation of
expression for each of 10 separate constellations of related protein expression. These constellations consist of proteins that correlate most closely with each other and suggest certain phenotypes for the cells. For example, CNSTN 9 has phospho-ERK, MEK, TP38 and inactivating phosphorylation of BAD and GSK3, a combination of features that promote growth and survival. Signature groups 4, 5 and 7 contain higher levels of CNSTN-8 proteins P53, MCL1 and neuropilin-1. Mutations of P53 are rare in AML, occurring in ~5-10%, but correlate with worse outcome. Approximately 10% of the samples analyzed had very strong p53 and sequencing revealed a high rate of P53 mutation in high expressers and confirmed the worse prognosis associated with high P53 expression (separate manuscript in preparation). The increased expression of MCL1 is associated with therapy resistance in AML. High expression of Neuropilin-1, which functions in vascular endothelial growth factor and semaphorin receptor signaling as a mediator of angiogenesis, has also been demonstrated to confer an adverse prognosis in AML possibly establishing a VEGF dependent autocrine loop that promotes survival. The RPPA method thus identifies the combination of mutant P53 and high levels of MCL1 and NRP1 as a novel adverse prognostic combination associated with the lower remission rates, highest relapse, and worse survival. This provides a testable hypothesis that combined therapy directed at restoring P53 function (such as with an MDM2 antagonist), inhibiting the antiapoptotic effect of MCL1 and interfering with VEGF signaling through NRP1 would be beneficial in this subgroup.

In most of the SGs, the protein expression patterns of STP conform to canonical expectations based on the established pathways. For example, in all SG, the level of pAKT and pS6RP were well correlated with their respective known substrates GSK3.
or P70S6K. Consistent with our prior Western blot findings, we again observed that: there was evidence of crosstalk within STP as levels of PKCa.p, AKT.p308, MEK.p, ERK.p and TP38.p were highly correlated within the SGs and that patients were more likely to have no STP or multiple STPs activated than expected ($\chi^2 = 13.07$, 3 df, $p=0.0004$), and reconfirmed that STP activation was associated with a worse prognosis (supplemental Figure 3) Some described canonical pathway activities are not replicated in the data set. For example, AKT is known to inactivate BAD and prevent apoptosis by phosphorylating serine 136 or to activate MTOR by phosphorylation. However in this data set when AKT.p473 and AKT.p308 is high BAD.p136 and MTOR.p levels are low and vice versa, suggesting non-canonical interactions. Disconnects between canonical pathways and expression levels may provide clues to deregulated systems in leukemia cells. For example, in most SG, high levels of PTEN.p are associated with low levels of phosph-AKT consistent with known models but in SG6 levels of AKT.p308 and .p473 are high despite high levels of pPTEN suggesting a failure of PTEN to regulate phospho-AKT levels. This could represent a failed feedback loop where the cell upregulated phospho-PTEN in an unsuccessful attempt to regulate phospho-AKT. Ratios of phopho to total protein may provide additional insight. Detailed analysis of stored material from these cases, in combination with specific agents targeting these specific pathways is required to test this hypothesis.

This dataset shows the power of being able to simultaneously ascertain the activation status of multiple pathways, with different functional effects, within a leukemic cell: a feature not possible with any prior method of analysis including mRNA expression arrays. The combination of these patterns leads to the definition of protein
expression signature groups with prognostic significance for remission attainment, relapse and survival. These SG classifications have prognostic significance that differs depending on the associated cytogenetics suggesting that pathway activation must be considered in the context of the underlying genetics of the cell. Some cohorts, e.g. unfavorable cytogenetics and SG6 have an unusually high CR rate and an unusually low relapse rate suggesting that despite having “poor prognosis cytogenetics” patients with this signature do as well as favorable prognosis cytogenetics patients. A dismal outcome with current therapy can also be predicted in some groups, e.g. unfavorable cytogenetics and SG1, with a 30% CR rate and a 100% relapse rate, suggesting that these patients should receive non-traditional therapies at presentation. Likewise, despite having favorable prognosis, cytogenetics patients with SG 5, and 6 had high relapse rates. Currently stem cell transplant for patients with favorable cytogenetics is reserved for those that relapse. This data, if confirmed, would suggest that favorable prognosis patients with these SGs should receive SCT in first CR. Furthermore, in an era with a rapidly increasing number of targeted therapies, each with relatively narrow spectrum of activity, a means to rationally select which agent to use in what molecular background is both critical and unknown. The patterns of expression in these groups can be used to suggest rational combinations of targeted therapies that should be evaluated in combination with conventional therapies.

In conclusion we were able to show the potential applicability of proteomic profiling for AML classification, target identification, treatment response prediction and hypothesis generation for disease mechanisms. The complex pattern of intracellular protein expression and activation might harbor clues for disease pathobiology. It is
possible to classify AML based on protein expression signatures that may predict outcome and therapy responsiveness across cytogenetics and other disease characteristics. Thus proteomic profiling may have advantages over other approaches of classifying AML as proteomic profiling may provide therapeutic guidance at the same time. A major question is whether the protein expression patterns defined from this dataset will be validated and found to be consistent across multiple patient sample sets, from multiple institutions, or have sufficient power to alter therapy. We are generating a validation array with a larger number of AML samples, which will be probed with an expanded number of antibodies to confirm our observation. If our results hold true this would be a first step to the ultimate goal of personalized molecular therapy based on protein profiling.
FUNDING

Leukemia and Lymphoma Society Grant # 6089 and NIH PO1 grant # CA-55164

Author Contributions:

Steven M. Kornblau designed research, performed research, analyzed data, and wrote the paper.

Raoul Tibe designed research, performed research, analyzed data, and wrote the paper.

YiHua Qiu performed research and analyzed data.

Wenjing Chen generated the database and analyzed data.

Hagop Kantarjian performed clinical care and edited manuscript.

Michael Andreeff performed clinical care and edited manuscript.

Kevin R. Coombes performed statistical analysis and wrote the paper.

Gordon B. Mills designed research, assisted with data analysis, and wrote the paper.

Conflict of interest disclosure: SMK, RT, YHQ and KC have filed for a patent related to these discoveries. There are no other conflicts of interest or financial relationships to disclose.
Reference List


33. Schoch C, Dicker F, Herholz H et al. Mutations of the TP53 Gene Occur in 13.4% of Acute Myeloid Leukemia and Are Strongly Associated with a Complex Aberrant Karyotype [abstract]. Blood 2006;108:


Table 1 Demographics and clinical characteristics of the 256 newly diagnosed AML patients in the study set.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Number or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td># of Cases</td>
<td>256</td>
</tr>
<tr>
<td>Gender</td>
<td>% Male</td>
<td>53.8%</td>
</tr>
<tr>
<td>Race</td>
<td>Asian</td>
<td>1.9%</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>8.3%</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>12.8%</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>76.7%</td>
</tr>
<tr>
<td>Age</td>
<td>Mean</td>
<td>61.1</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>63.8</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>16.1</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>87.2</td>
</tr>
<tr>
<td>FAB</td>
<td>M0</td>
<td>6.5%</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>13.8%</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>26.8%</td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>23.4%</td>
</tr>
<tr>
<td></td>
<td>M4EOS</td>
<td>3.4%</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>2.3%</td>
</tr>
<tr>
<td></td>
<td>M5A</td>
<td>5.4%</td>
</tr>
<tr>
<td></td>
<td>M5b</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>M6</td>
<td>2.7%</td>
</tr>
<tr>
<td></td>
<td>M7</td>
<td>1.9%</td>
</tr>
<tr>
<td></td>
<td>RAEBT</td>
<td>1.9%</td>
</tr>
<tr>
<td></td>
<td>UNK</td>
<td>8.0%</td>
</tr>
<tr>
<td>WHO Classification</td>
<td>AML with recurrent cytogenetic changes</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>AML with multilineage dysplasia</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>AML, Therapy related</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>AML not otherwise categorized</td>
<td>134</td>
</tr>
<tr>
<td>Cytogenetics</td>
<td>Favorable</td>
<td>8.3%</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>45.7%</td>
</tr>
<tr>
<td></td>
<td>Unfavorable</td>
<td>44.5%</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>1.5%</td>
</tr>
<tr>
<td>Zubrod PS</td>
<td>0-1-2</td>
<td>94.1%</td>
</tr>
<tr>
<td></td>
<td>3 or 4</td>
<td>5.9%</td>
</tr>
<tr>
<td>AHD &gt;=2 Mo</td>
<td></td>
<td>32.1%</td>
</tr>
<tr>
<td>Response</td>
<td>CR</td>
<td>56.8%</td>
</tr>
<tr>
<td></td>
<td>CRP</td>
<td>3.2%</td>
</tr>
<tr>
<td></td>
<td>Early Death</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>Fail</td>
<td>8.1%</td>
</tr>
<tr>
<td></td>
<td>Resistant</td>
<td>31.1%</td>
</tr>
<tr>
<td></td>
<td>Inevaluable</td>
<td>0.5%</td>
</tr>
<tr>
<td>Relapse</td>
<td>Yes</td>
<td>54.1%</td>
</tr>
<tr>
<td>Alive</td>
<td>Yes</td>
<td>29.4%</td>
</tr>
</tbody>
</table>

AHD>=2 = antecedent hematological disorder of greater than or equal to 2 months. CR= complete remission, CRp= Complete remission criteria met except that platelets fail to reach 100,000μL. Early death = death within the first two weeks after the start of induction therapy. Fail= death occurring more than 2 weeks after the start of therapy.
## Table 2. Proteins with different expression between blood and marrow

<table>
<thead>
<tr>
<th>Protein</th>
<th>P-Value</th>
<th>Fold</th>
<th>Protein</th>
<th>P-Value</th>
<th>Fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAK</td>
<td>2.5x10^{-4}</td>
<td>1.082</td>
<td>Survivin</td>
<td>3.8x10^{-7}</td>
<td>1.113</td>
</tr>
<tr>
<td>BAD</td>
<td>1.3x10^{-4}</td>
<td>1.113</td>
<td>MTOR.p</td>
<td>8.5 x10^{-4}</td>
<td>1.123</td>
</tr>
<tr>
<td>SRCp517</td>
<td>5.5x10^{-8}</td>
<td>1.278</td>
<td>S6RP.p235</td>
<td>2.8x10^{-7}</td>
<td>1.400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S6RP.p240-244</td>
<td>1.0x10^{-16}</td>
<td>1.553</td>
</tr>
</tbody>
</table>
Table 3: Distribution of Cytogenetic Risk Groups

<table>
<thead>
<tr>
<th>SG Group</th>
<th>Favorable</th>
<th>Intermediate</th>
<th>Unfavorable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed (Expected) Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 (2.83)</td>
<td>19 (15.84)</td>
<td>14 (15.33)</td>
</tr>
<tr>
<td>2</td>
<td>12 (4.17)</td>
<td>19 (23.30)</td>
<td>19 (22.54)</td>
</tr>
<tr>
<td>3</td>
<td>4 (3.42)</td>
<td>25 (19.10)</td>
<td>12 (18.48)</td>
</tr>
<tr>
<td>4</td>
<td>2 (3.25)</td>
<td>19 (18.17)</td>
<td>18 (17.58)</td>
</tr>
<tr>
<td>5</td>
<td>1 (2.67)</td>
<td>8 (14.91)</td>
<td>23 (14.42)</td>
</tr>
<tr>
<td>6</td>
<td>2 (2.92)</td>
<td>21 (16.31)</td>
<td>12 (15.78)</td>
</tr>
<tr>
<td>7</td>
<td>0 (2.75)</td>
<td>12 (15.38)</td>
<td>21 (14.88)</td>
</tr>
</tbody>
</table>

The three major cytogenetic risk groups are unevenly distributed across the seven protein signature groups. Each cell contains the observed (and expected under the null hypothesis) number of patients in each group. Observations that differed most from the expected values (and thus contributed significantly to the chi-squared test) are indicated by boldface.
Table 4: Complete remission and relapse rate by signature group stratified by cytogenetic risk groups

The three major cytogenetic risk groups are unevenly distributed across the seven signature groups. The left part of each column* under the cytogenetic risk categories shows what percentage of patients within each SG have achieved complete remission and the second # shows the percentage that relapsed.

<table>
<thead>
<tr>
<th>SG Group</th>
<th>Favorable</th>
<th>Intermediate</th>
<th>Unfavorable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>*CR/Relapse</td>
<td>CR / Relapse</td>
<td>CR / Relapse</td>
</tr>
<tr>
<td>1</td>
<td>*100 / 0</td>
<td>67 / 30</td>
<td>30 / 100</td>
</tr>
<tr>
<td>2</td>
<td>100 / 33</td>
<td>65 / 55</td>
<td>54 / 43</td>
</tr>
<tr>
<td>3</td>
<td>100 / 50</td>
<td>67 / 31</td>
<td>33 / 67</td>
</tr>
<tr>
<td>4</td>
<td>100 / 0</td>
<td>67 / 70</td>
<td>47 / 43</td>
</tr>
<tr>
<td>5</td>
<td>100 / 100</td>
<td>40 / 50</td>
<td>53 / 67</td>
</tr>
<tr>
<td>6</td>
<td>100 / 100</td>
<td>60 / 75</td>
<td>75 / 67</td>
</tr>
<tr>
<td>7</td>
<td>--/--</td>
<td>58 / 29</td>
<td>32 / 83</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1: Array Assembly, Printing Method, Topographical Normalization and Example Slide. (A) Each letter represents a group of specimens. Replicates of the same patient sample were printed in area “A” and in a reversed orientation in area “B”. As controls, 18 cell lines (MDA-MD-231, MDA-MD-231 stimulated with IGF, MDA-MD-468, MDA-MD-468 stimulated with EGF, Hela HL-60, Jurkat (x2 from separate sources), Jurkat incubated with anti-FAS antibody, K562, Kasumi-1, MV4-11, NB4, OCIAML3, Raji, THP-1, U937, Y79) were printed in area “C” and 18 normal peripheral blood samples in area “N”. Purified peptides (n=138) were printed encircling the patient samples, as shown by the purple band. (B) Schematic showing the dilution series of patient samples. At the end of each row of patient sample a positive control made of a mixture of 11 cell lines (Hela, HL60, Jurkat, Kasumi1, K562, KG1, MV4-11, OCI-AML3 NB4, U937, Y79) (pooled control, shown as a red “+”) or negative control (protein lysis buffer, shown as a “-“) is printed, creating a grid across the whole slide for negative and positive controls to permit topographical normalization. (C) Representative slide of phospho-AKT (Thr308). The overlays in purple, red and green covers the areas where the purified peptides, cell lines and normal peripheral blood samples and positive and negative control are located as depicted in (Fig 1a). The insert demonstrates phosphorylated and unphosphorylated AKT peptide specifically detected by the anti-phospho-AKT (Thr308) antibody. (D, E) The 3 dimensional (D) topographical map of the negative control generates a 3-D topographical map that can be used to correct for background. While the 3D grid of the positive controls sets the scale for quantification. A
topographical map from a slide (probed for BAK) with low background that is even across the slide (D) and another (E) from the slide (probed for MCL1) with the most background variation across the slide are shown.

**Figure 2: Hierarchical Clustering of Antibodies (AB) Differentially Expressed Among FAB Groups.** The heatmap graphically represents the 24 AB whose expression levels were significantly different between the FAB subgroups. The number of patients in each group is shown along the bottom of the figure. Protein levels were normalized such that the expression level of M1 was set to 0 and the expression of all other proteins are shown as fold change from this baseline. Each protein was standardizes so that each protein has mean 0 and standard deviation one across the categories.

**Figure 3: Clustering by cytogenetics** Average protein expression intensity of patients within cytogenetic risk groups based on unsupervised hierarchical clustering.

**Figure 4: Protein Expression against Signature Groups.** Proteins were clustered into 10 constellations shown by the 10 colored boxes along the left axis and the yellow lines surrounding each constellation. Based on the score for each constellation for each patient an overall score was generated and this divided patients into 7 protein signature groups (bottom of heatmap). The average score for each protein constellation within each signature group is shown. An additional heatmap showing the expression of each protein in each individual patient is available online as supplementary figure 2. The fold
change between -2 standard deviations and +2 standard deviations for each individual protein, corresponding to the -2 (green) to +2 (red) color gradient bar shown in the upper left corner, is listed to the right of the protein name.

**Figure 5: Overall Survival in Signature Groups in All Patients and in Diploid FLT3-ITD Cases** Kaplan Meier estimates of proportion of surviving patients over time in weeks for A) all cases and B) for cases with diploid cytogenetics and a FLT3-ITD are shown.
Figure 1
Figure 5A

Proportion Surviving

Time (weeks)

log rank p = 0.011
Figure 5 B

log rank p = 0.001

Proportion Surviving

Time (weeks)
Functional proteomic profiling of AML predicts response and survival

Steven M Kornblau, Raoul Tibes, YiHua Qiu, Wenjing Chen, Hagop M Kantarjian, Michael Andreeff, Kevin R Coombes and Gordon B Mills

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include digital object identifier (DOIs) and date of initial publication.