Published in final edited form as:
J Proteome Res. 2012 October 5; 11(10): 5005–5010. doi:10.1021/pr300567r.

Proteomic classification of acute leukemias by alignment-based quantitation of LC-MS/MS data sets

Eric J. Foss[1], **Dragan Radulovic**[2,**], **Derek L. Stirewalt**[1], **Jerald Radich**[1], **Olga Sala-Torra**[1], **Era L. Pogosova-Agadjanyan**[1], **Shawna M. Hengel**[3,4], **Keith R. Loeb**[1], **H. Joachim Deeg**[1], **Soheil Meshinchi**[1], **David R. Goodlett**[4], and **Antonio Bedalov**[1,*]

[1]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

[2]Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, Florida 33431, USA

[3]Pacific Northwest National Laboratory, Richland, WA 99352, U.S.A.

[4]Department of Medicinal Chemistry, University of Washington, Seattle, Washington 98195, USA

Abstract

Despite immense interest in the proteome as a source of biomarkers in cancer, mass spectrometry has yet to yield a clinically useful protein biomarker for tumor classification. To explore the potential of a particular class of mass spectrometry-based quantitation approaches, label-free alignment of liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) data sets, for the identification of biomarkers for acute leukemias, we asked whether a label-free alignment algorithm could distinguish known classes of leukemias on the basis of their proteomes. This approach to quantitation involves (1) computational alignment of MS1 peptide peaks across large numbers of samples; (2) measurement of the relative abundance of peptides across samples by integrating the area under the curve of the MS1 peaks; and (3) assignment of peptide IDs to those quantified peptide peaks on the basis of the corresponding MS2 spectra. We extracted proteins from blasts derived from four patients with acute myeloid leukemia (AML, acute leukemia of myeloid lineage) and five patients with acute lymphoid leukemia (ALL, acute leukemia of lymphoid lineage). Mobilized CD34+ cells purified from peripheral blood of six healthy donors and mononuclear cells (MNC) from the peripheral blood of two healthy donors were used as healthy controls. Proteins were analyzed by LC-MS/MS and quantified with a label-free alignment-based algorithm developed in our laboratory. Unsupervised hierarchical clustering of blinded samples separated the samples according to their known biological characteristics, with each sample group forming a discrete cluster. The four proteins best able to distinguish CD34+, AML and ALL were all either known biomarkers or proteins whose biological functions are consistent with their ability to distinguish these classes. We conclude that alignment-based label-free quantitation of LC-MS/MS data sets can, at least in some cases, robustly distinguish known classes of leukemias, thus opening the possibility that large scale studies using such algorithms can lead to the identification of clinically useful biomarkers.

*to whom correspondence should be addressed. **to whom correspondence related specifically to protein quantitation algorithm should be addressed.

Supporting Information

Supplemental Notes S1. A detailed description of how we prepared proteins and carried out mass spectrometry.

Author Contributions: E.J.F., D.R.G. and A.B. designed experiments, E.J.F. made figures, E.J.F. and A.B. wrote manuscript, E.J.F., O.S. and E.L.P. prepared samples, D.L.S., J.R., O.S. and E.L.P. provided critical help in obtaining samples, E.J.F. and S.M.H. performed mass spectrometry, D.R. quantified proteins, K.R.L., H.J.D., S.M., D.L.S., J.R., O.S., E.L.P. and H.J.D. participated in helpful discussions.

Introduction

Modern genomics technologies such as high throughput DNA sequencing, [1, 2] SNP chips, [3] digital transcriptome sequencing [4] and DNA methylation analysis [5–7] are being applied in numerous areas of cancer biomarker discovery. Indeed, we are on the brink of being able to routinely sequence the genome of each cancer patient.[8] But despite the promise of these techniques, they are intrinsically limited to detecting clues to cellular behavior that can be inferred from nucleic acids. Being essential in the control and execution of virtually every biological function and process, proteins are expected to provide a more immediate readout of cellular physiology than nucleic acids. Unfortunately, techniques for the direct analysis of large numbers of proteins have lagged behind those for analysis of nucleic acids. Some of these techniques rely on spectral counting, but these are generally considered to be only semiquantitative.[9] Other techniques, while highly quantitative, rely on isotopic labeling and are most applicable to pairwise experiment/control comparisons rather than analysis of large numbers of individuals. This is because larger numbers of samples have to be compared to each other through a common control, which multiplies errors, and because isotopic labeling techniques generally require analysis of multiple fractions so that sample complexity is reduced to the point that paired isotopic peaks are easily identified.[10, 11] Still other techniques, like multiple-reaction monitoring,[12] are limited to biomarker validation rather than discovery.

In recent years, several academic and commercial techniques have emerged that allow computational alignment of MS1 peptide peaks across large numbers of samples, relative quantitation of the peptides by integration of area under the curve, and then assignment of peptide IDs according to the corresponding MS2 spectra. [13–24] Such techniques hold the promise of accuracy sufficient for biomarker discovery without the limitations of isotopic labeling. Although alignment-based techniques have been used in leukemia biomarker studies using surface-enhanced laser desorption/ionization-time of flight (SELDI-TOF),[25] analogous techniques using LC-MS/MS have not been attempted. Because LC-MS/MS is not prone to the same artifacts that have troubled biomarker discovery using SELDI-TOF,[26–28] alignment-based quantitation of LC-MS/MS data sets constitutes a potentially fruitful technique for biomarker discovery. We have developed a label-free quantitation algorithm which we previously described and used for measuring protein variation in a yeast population consisting of 95 genetically distinct yeast strains.[14] The analysis of protein and RNA abundance showed that most of protein level variation in this genetically diverse yeast population is due to variation in translation and/or protein stability rather than variation in transcript levels. (To the best of our knowledge, this is the only published analysis of gene-specific RNA-protein correlations, i.e. correlations for a single gene across many individuals; in contrast, several reports have analyzed RNA-protein correlations between genes within a single individual.[29, 30]) This discovery is relevant to the application of label-free techniques to protein biomarker discovery in cancer because (1) it emphasizes the importance of looking directly at the proteome rather than inferring protein levels from transcript levels and (2) it shows that our particular label-free technique is sufficiently accurate to extract biologically relevant proteomic signatures from large numbers of individuals in a genetically diverse population.

In order to explore the applicability of label-free alignment based protein quantitation in cancer biomarker discovery, we tested our algorithm in the context of known classes of acute leukemias. For two reasons, acute leukemias provide an ideal situation in which to test the approach. First, a variety of well established biomarkers make distinguishing AML from ALL completely unambiguous, and therefore we can ask whether our algorithm classifies samples correctly without uncertainty about what is correct. Second, among the well

established biomarkers for distinguishing these diseases are the levels of several proteins; therefore if our algorithm does correctly classify the two diseases, we can ask whether it did so, at least in part, on the basis of known protein biomarkers. We chose 3 classes of cells with varying degrees of similarity, thus providing a range of difficulty for distinguishing the classes: ALL, AML and healthy CD34+ controls, which serve to control for the stage at which differentiation of the leukemic cells is blocked. Here we demonstrate that a label-free alignment-based method for protein quantitation can robustly distinguish all three classes of patient samples and identify previously validated biomarkers among the proteins it deems most useful for distinguishing these classes. This is the first instance in which such a method has been used to identify protein biomarkers of leukemia and opens the possibility of larger studies applying this techniques.

Methods

Samples were obtained from FHCRC leukemia repository (AML, ALL) or from volunteer donors at FHCRC (MNC, CD34+) under Institutional Review Board (IRB) approved protocols, and consent was provided according to the Declaration of Helsinki. Sterile, viable cryopreserved leukemia cell suspensions, obtained through Ficoll separation of diagnostic bone marrow aspirates and collection of mononuclear fractions are stored in dedicated freezers at –135°C. These cryopreserved leukemia cells, approximately 5×10^6 cells per aliquot, are backed by a database describing disease, cell counts, cytogenetics, flow cytometric analysis and treatment response linked to SWOG Biostatistical Center clinical databases. The quality of the samples in the repository is generally very high. Cell viability of >80%, as measured by trypan blue exclusion, is found in approximately 75% of samples. Our experimental procedure involves: 1) cell lysis; 2) digestion of total proteins to peptides; 3) quadruplicate microcapilary HPLC tandem mass spectrometry analysis of total peptides using a Thermo-Electron Orbitrap instrument; 4) alignment of MS1 peptide peaks from different MS runs; 5) extraction of changes in protein expression based on ion intensities; 6) protein identification through data base searches using SEQUEST ;[31] and 7) assignment of protein IDs to MS1 peaks based on the alignment algorithm. Details of this procedure can be found in Supplemental Notes S1. All mass spectrometry data collection preceded data analysis. Profiling a single sample took 105 minutes, and because each sample was profiled in quadruplicate, a single sample to 420 minutes. Details of our label-free alignment-based algorithm have been described previously.[14, 15] To access the code, contact D.R.

Results and Discussion

Fifteen frozen mononuclear fractions from peripheral blood were analyzed, 5 from patients with ALL, 4 from patients with AML and 6 samples of CD34+ cells from healthy donors. All ALL samples were B cell leukemias. Disease samples were chosen without biases, e.g. they were not chosen specifically from a drug-resistant groupt of patients. We also prepared two samples from ficoll-purified MNC fractions from peripheral blood of healthy donors in order to control for the effect of contaminating healthy monocytes. All leukemic samples contained a high fraction of blasts (60–92%) and had high viability, as judged by trypan blue staining (80–95%). Blasts were not purified by flow cytometry because we wanted to test our method under conditions that are both more challenging and more clinically relevant (i.e. with a heterogeneous collection of cells rather than a flow-sorted homogeneous collection). Each sample was analyzed in quadruplicate using a ThermoElectron LTQ-Orbitrap mass spectrometer for a total of 68 mass spectrometric runs. In order to prevent batch effects in which samples group according to the order in which they were analyzed, samples were analyzed in four groups, each group containing 1 of each sample type, and within each group the order in which the samples were analyzed was randomized.

We quantified peptides belonging to 639 different proteins, generally with multiple peptides corresponding to the same protein. In order to determine whether the expression levels of peptides could reveal the biological structure that underlies the different pathologies of AML and ALL, we analyzed protein expression levels by unsupervised hierarchical clustering of blinded sample quantitation results. The clustering algorithm accurately separated all major groups, with ALL and MNC fractions appearing most different from the other samples, while AML and CD34+ formed separate branches in a shared tree (Figure 1). The CD34+ and AML are expected to be similar because the CD34+ cells were harvested from the peripheral blood of healthy individuals that had been stimulated with GCSF, which induces production of myeloid precursors. Thus, even though we have only a small number of samples to identify biologically relevant groups, an unbiased group discovery algorithm (unsupervised hierarchical clustering) performed perfectly.

In order to estimate the probability that 17 samples could have been classified correctly by chance, we performed simulations. Specifically, we created arrays of lengths 6, 5, 4 and 2 for the CD34+, ALL, AML and MNC samples, respectively, and filled the first bin in each array with the name of the appropriate sample type ("CD34+", "ALL", "AML" and "MNC", respectfully). We then randomly chose names from the 13 remaining names (5 CD34+'s, 4 ALL's, etc.) to fill the remaining bins in each array, always removing each chosen name from the pool. We then asked whether the arrays had been filled correctly (the 6 member array containing 6 "CD34+" names, etc.) and repeated this one million times. The arrays were not filled correctly even a single time, giving a p value for the chance that our unsupervised hierarchical clustering classified samples corrrectly of $p < 10^{-6}$. For three reasons, this exercise is tremendously biased toward success and therefore the true p value will be far smaller than this. First, this is supervised rather than unsupervised clustering; i.e. we pre-specified the sizes of the different groups to be correct, whereas in our unsupervised clustering above, groups of any sizes could have been identified and there are thousands of incorrect group sizes possible. Second, we seeded each array correctly, putting "ALL" into the first bin of the 5 member array, etc., so not only was the seeding correct but there were also just 13 remaining samples to classify. Third, this classification is not hierarchical, i.e. it does not demonstrate that, for example, CD34+ and AML samples are more similar to each other than either is to ALL samples. Thus the true probability that our unsupervised hierarchical test succeeded by chance is $p \ll 10^{-6}$.

We next wanted to determine the best protein biomarkers for distinguishing CD34+, AML and ALL. Ideally, one would split the samples into a hypothesis-generating "training set" to identify promising protein biomarkers and then test these protein biomarkers in a "testing set". However, because our small sample size precluded such an analysis, we turned instead to a "leave one out" approach.[32] For example, in looking for proteins capable of distinguishing CD34+ and ALL, we omitted one CD34+ sample, generated candidate biomarkers by comparing the remaining 5 CD34+ samples with the AML samples, and then asked whether the omitted CD34+ sample would be classified correctly on the basis of the protein candidates identified. This was repeated until each sample had been omitted and only proteins that performed perfectly by this criterion were retained. As expected, the more different the samples were, the more distinguishing proteins we identified: 91, 71 and 17 proteins distinguished ALL from CD34+, ALL from AML, and AML from CD34+, respectively (Table S3).

CD34+ versus ALL

With 91 proteins capable of distinguishing CD34+ from ALL (Figure 2A), it was important to prioritize potential biomarkers. We reasoned that the best biomarkers would be those that are present in every individual in one group and absent in every individual in the other group. Ten out of 91 proteins displayed such an "all or none" pattern, and among these 10

was CD10, which was present in all ALL samples but not CD34+ samples. CD10 is a neutral endopeptidase that is already used as a marker for B-cell ALL;[33] thus, we are not only classifying patient samples correctly based on disease, we are also identifying validated protein biomarkers for such classification. For brevity, below we will use the term "all or none biomarker" as it has been defined in this paragraph.

AML versus ALL

Among the 71 proteins capable of distinguishing AML from ALL (Figure 2B) were 14 "all or none" biomarkers. Among these, nicastrin was notable for its ability to distinguish AML from either ALL or CD34+. Because nicastrin is a component of the gamma secretase complex,[34] an intramembrane cysteine protease that controls Notch proteolysis, identification of high levels of nicastrin in AML raises the possibility of disregulated Notch signaling in AML pathogenesis. Given the availability of gamma secretase inhibitors, this finding may have important therapeutic implications.[35]

CD34+ versus AML

Among the 17 proteins capable of distinguishing CD34+ cells from AML were three "all or none" biomarkers: retinal dehydrogenase, nicastrin (NCSTN) and 3-ketoacyl CoA thiolase (Figure 2C). Retinal dehydrogenase catalyzes the final step in retinoic acid synthesis, and retinoic acid drives differentiation in the myeloid lineage and is used in treatment of acute promyelocytic leukemia, a subtype of AML. Thus, it is intriguing to see undetectable levels of retinaldehyde dehydrogenase in the AML samples in contrast to high levels of this enzyme observed in the CD34+ samples. Two proteins (mitochondrial phosphate carrier protein and glucosidase 2 subunit beta) were higher in both AML and ALL than in CD34+, thus acting as lineage-independent markers of disease.

In order to determine the degree to which information obtained by direct measurement of the proteome could have been inferred from the transcriptome, we subjected transcriptional data from CD34+, mononuclear, AML and ALL samples (10, 9, 26 and 32 samples, respectively) to the same analysis presented above [36]. These data were collected by hybridization of RNA to Affymetrix HG-U133A arrays. We found that although transcripts corresponding to the 639 proteins we measured generally classified samples according to the known biology (Figure S1), the genes that were most valuable in this regard for the proteome and the transcriptome were significantly different. In fact, the transcripts corresponding to the proteins most useful for classification were completely unable to distinguish any pair of sample types (Figure S2 A–C), thus these two techniques show promising complementarity.

In summary, a label-free protein profiling technology proved capable not only of robustly distinguishing known classes of leukemias as well as healthy controls despite a very limited sample size, but also of doing so on the basis of known biomarkers, thus validating an alignment-based protein quantitation approach as tool for leukemia classification. It is worth noting that AML and ALL are each very heterogeneous diseases, yet despite the challenge posed by this intra-disease heterogeneity, our algorithm performed robustly. The success of this proof-of-principle experiment in the context of known classes of leukemia opens the door for large studies aimed at further refinements in tumor classification of leukemias as well as other cancers using not only our own algorithm[14] but also the other label-free algorithms academically and commercially available,[16–20] each of which is likely to have its own strengths and weaknesses. Because acute leukemias present a group of malignancies that differ in their molecular pathogenesis and the required treatment approach, a combination of proteome and transcriptome profiling in larger sample sets is likely to yield new prognostic and treatment response classes and their corresponding biomarkers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Priska Von Haller for help with mass spectrometry, Martin Morgan for computational advice, Phil Gafken for helpful discussions and Vid Leko for comments on the manuscript. This work was funded by N.I.H. grants UL1RR025014, CA015704, CA129132 and CA164545 to A.B and 5R33CA099139-04 to D.R.G. This work was also supported in part by the University of Washington's Proteomics Resource (UWPR95794).

References

1. Roukos DH. Trastuzumab and beyond: sequencing cancer genomes and predicting molecular networks. Pharmacogenomics J. 2011 Apr; 11(2):81–92. [PubMed: 20975737]
2. Bacher U, Schnittger S, Haferlach C, Haferlach T. Molecular diagnostics in acute leukemias. Clin Chem Lab Med. 2009; 47(11):1333–1341. [PubMed: 19817644]
3. Li L, Fridley BL, Kalari K, Jenkins G, Batzler A, Weinshilboum RM, et al. Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. PLoS One. 2009; 4(11):e7765. [PubMed: 19898621]
4. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol. 2011 Jul 31.
5. Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, et al. DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. Genome Res. 2011 Jul; 21(7):1017–1027. [PubMed: 21521786]
6. Cheung HH, Lee TL, Davis AJ, Taft DH, Rennert OM, Chan WY. Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. Br J Cancer. 2010 Jan 19; 102(2):419–427. [PubMed: 20051947]
7. Cheung HH, Lee TL, Rennert OM, Chan WY. DNA methylation of cancer genome. Birth Defects Res C Embryo Today. 2009 Dec; 87(4):335–350. [PubMed: 19960550]
8. Diamandis M, White NM, Yousef GM. Personalized medicine: marking a new epoch in cancer patient management. Mol Cancer Res. 2010 Sep; 8(9):1175–1187. [PubMed: 20693306]
9. Lundgren DH, Hwang SI, Wu L, Han DK. Role of spectral counting in quantitative proteomics. Expert Rev Proteomics. 2010 Feb; 7(1):39–53. [PubMed: 20121475]
10. Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics. 2007 Feb; 7(3):340–350. [PubMed: 17177251]
11. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002 May; 1(5):376–386. [PubMed: 12118079]
12. Rodriguez H, Rivers R, Kinsinger C, Mesri M, Hiltke T, Rahbar A, et al. Reconstructing the pipeline by introducing multiplexed multiple reaction monitoring mass spectrometry for cancer biomarker verification: an NCI-CPTC initiative perspective. Proteomics Clin Appl. 2010 Dec; 4(12):904–914. [PubMed: 21137031]
13. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem. 2003 Sep 15; 75(18):4818–4826. [PubMed: 14674459]
14. Foss EJ, Radulovic D, Shaffer SA, Goodlett DR, Kruglyak L, Bedalov A. Genetic Variation Shapes Protein Networks Mainly through Non-transcriptional Mechanisms. PLoS Biol. 2011 Sep. 9(9):e1001144. [PubMed: 21909241]
15. Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Bedalov A, Goodlett DR, et al. Genetic basis of proteome variation in yeast. Nat Genet. 2007 Nov; 39(11):1369–1375. [PubMed: 17952072]

16. Khan Z, Bloom JS, Garcia BA, Singh M, Kruglyak L. Protein quantification across hundreds of experimental conditions. Proc Natl Acad Sci U S A. 2009 Sep 15; 106(37):15544–15548. [PubMed: 19717460]

17. Monroe ME, Shaw JL, Daly DS, Adkins JN, Smith RD. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. Comput Biol Chem. 2008 Jun; 32(3):215–217. [PubMed: 18440872]

18. Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, et al. SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. Proteomics. 2007 Oct; 7(19):3470–3480. [PubMed: 17726677]

19. Hoekman B, Breitling R, Suits F, Bischoff R, Horvatovich P. msCompare: a framework for quantitative analysis of label-free LC-MS data for comparative biomarker studies. Mol Cell Proteomics. 2012 Feb 7.

20. Noy K, Towfic F, Wittenberg GM, Fasulo D. Shape-based feature matching improves protein identification via LC-MS and tandem MS. J Comput Biol. 2011 Apr; 18(4):547–557. [PubMed: 21417940]

21. Mueller LN, Brusniak MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res. 2008 Jan; 7(1): 51–61. [PubMed: 18173218]

22. Radulovic D, Jelveh S, Ryu S, Hamilton TG, Foss E, Mao Y, et al. Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics. 2004 Oct; 3(10):984–997. [PubMed: 15269249]

23. Palagi PM, Walther D, Quadroni M, Catherinet S, Burgess J, Zimmermann-Ivol CG, et al. MSight: an image analysis software for liquid chromatography-mass spectrometry. Proteomics. 2005 Jun; 5(9):2381–2384. [PubMed: 15880814]

24. America AH, Cordewener JH. Comparative LC-MS: a landscape of peaks and valleys. Proteomics. 2008 Feb; 8(4):731–749. [PubMed: 18297651]

25. Forshed J, Pernemalm M, Tan CS, Lindberg M, Kanter L, Pawitan Y, et al. Proteomic data analysis workflow for discovery of candidate biomarker peaks predictive of clinical outcome for patients with acute myeloid leukemia. J Proteome Res. 2008 Jun; 7(6):2332–2341. [PubMed: 18452325]

26. Hu J, Coombes KR, Morris JS, Baggerly KA. The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. Brief Funct Genomic Proteomic. 2005 Feb; 3(4):322–331. [PubMed: 15814023]

27. Diao L, Clarke CH, Coombes KR, Hamilton SR, Roth J, Mao L, et al. Reproducibility of SELDI Spectra Across Time and Laboratories. Cancer Inform. 2011; 10:45–64. [PubMed: 21552492]

28. Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J Natl Cancer Inst. 2005 Feb 16; 97(4):307–309. [PubMed: 15713966]

29. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol. 2007 Jan; 25(1):117–124. [PubMed: 17187058]

30. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999 Mar; 19(3):1720–1730. [PubMed: 10022859]

31. Yates, Jr, 3rd; Eng, JK.; McCormack, AL.; Schieltz, D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem. 1995 Apr 15; 67(8):1426–1436. [PubMed: 7741214]

32. Chiaretti S, Li X, Gentleman R, Vitale A, Vignetti M, Mandelli F, et al. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood. 2004 Apr 1; 103(7):2771–2778. [PubMed: 14684422]

33. Kamen BA. CALLA and protein kinase CK2 revisited: targetable molecules for treating patients with leukemia? J Pediatr Hematol Oncol. 2010 Jan.32(1):1. [PubMed: 20051778]

34. Shah S, Lee SF, Tabuchi K, Hao YH, Yu C, LaPlant Q, et al. Nicastrin functions as a gamma-secretase-substrate receptor. Cell. 2005 Aug 12; 122(3):435–447. [PubMed: 16096062]

35. Rizzo P, Osipo C, Foreman K, Golde T, Osborne B, Miele L. Rational targeting of Notch signaling in cancer. Oncogene. 2008 Sep 1; 27(38):5124–5131. [PubMed: 18758481]

36. Stirewalt DL, Meshinchi S, Kopecky KJ, Fan W, Pogosova-Agadjanyan EL, Engel JH, et al. Identification of genes with abnormal expression changes in acute myeloid leukemia. Genes Chromosomes Cancer. 2008 Jan; 47(1):8–20. [PubMed: 17910043]

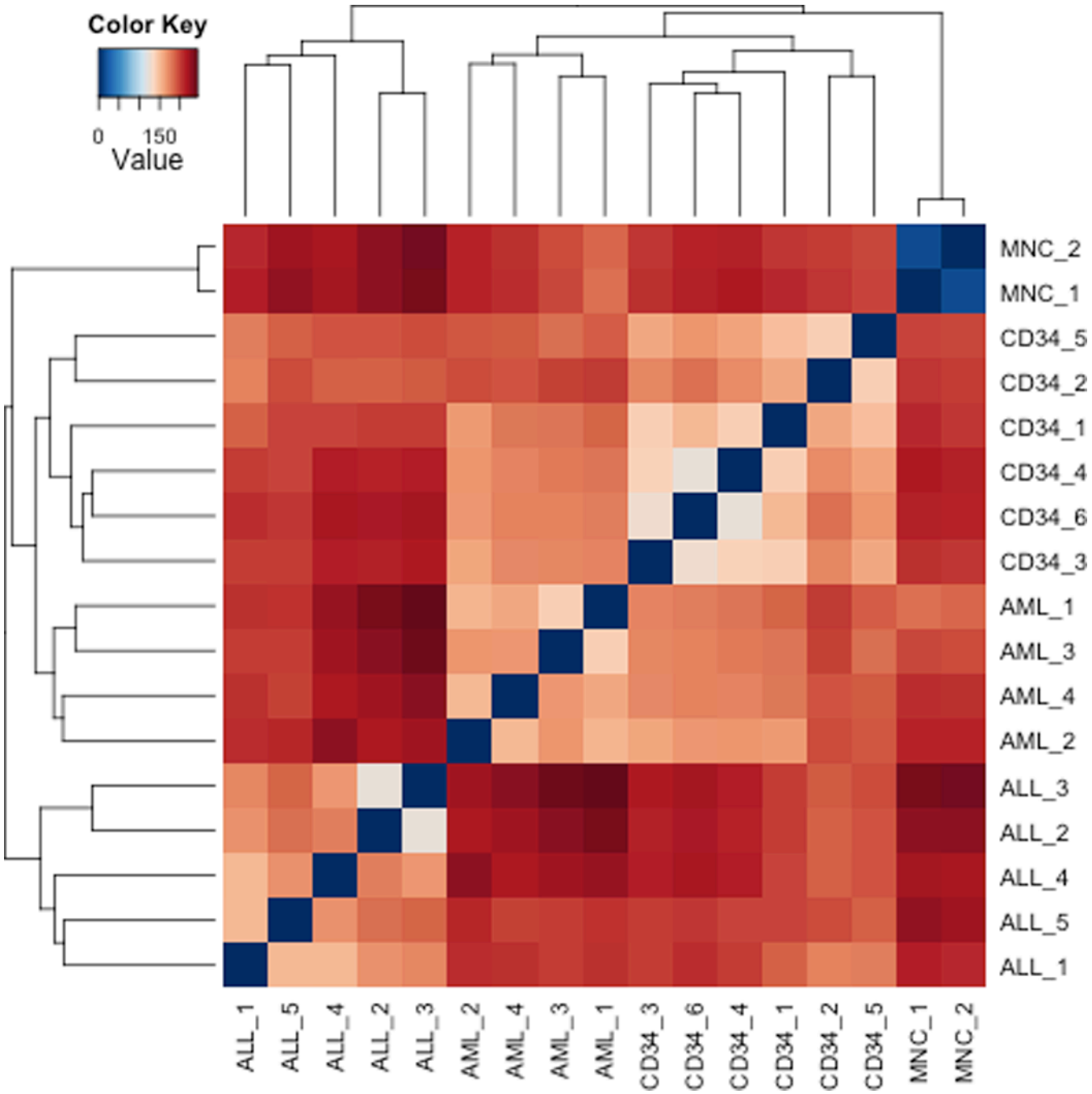
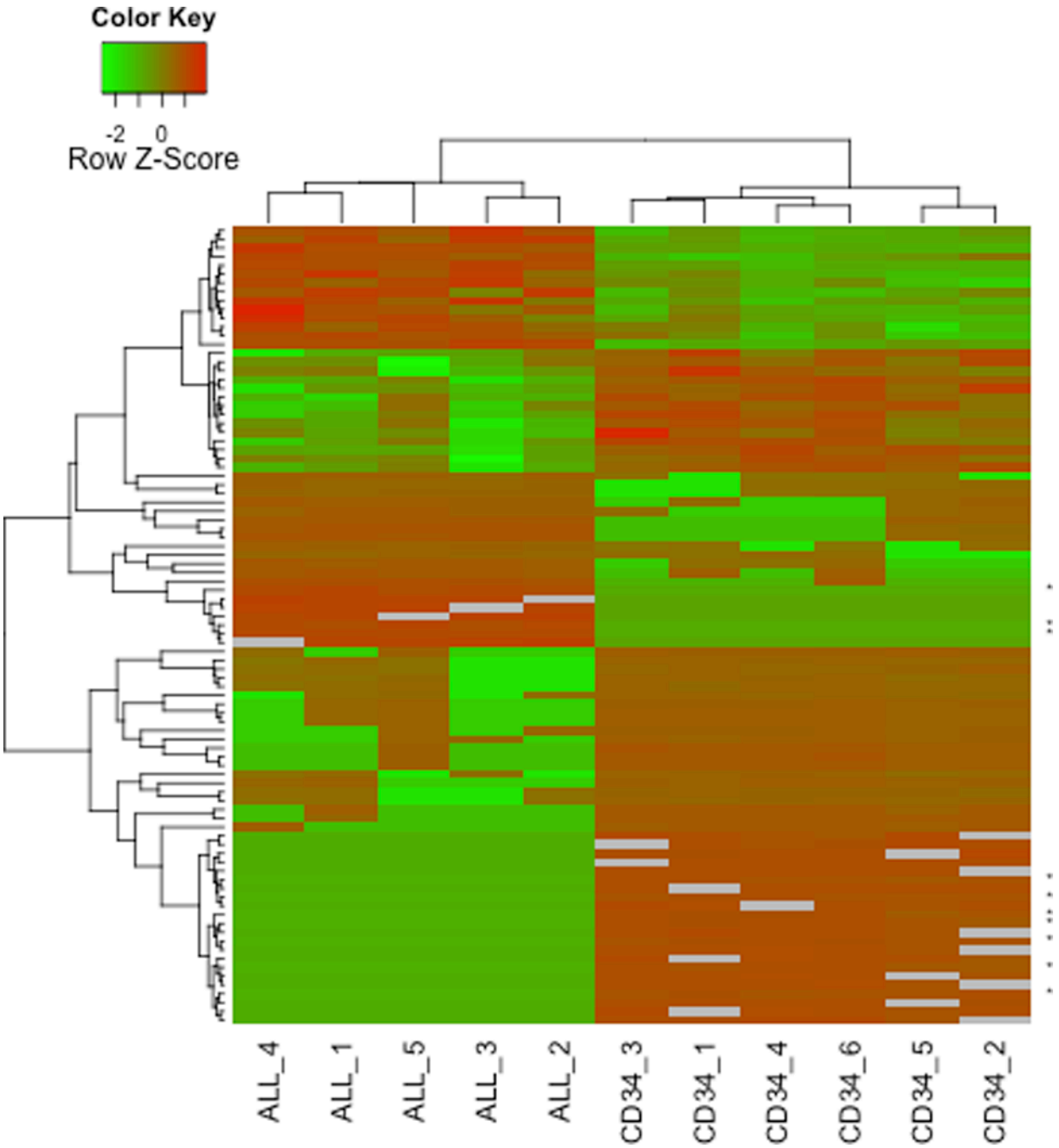
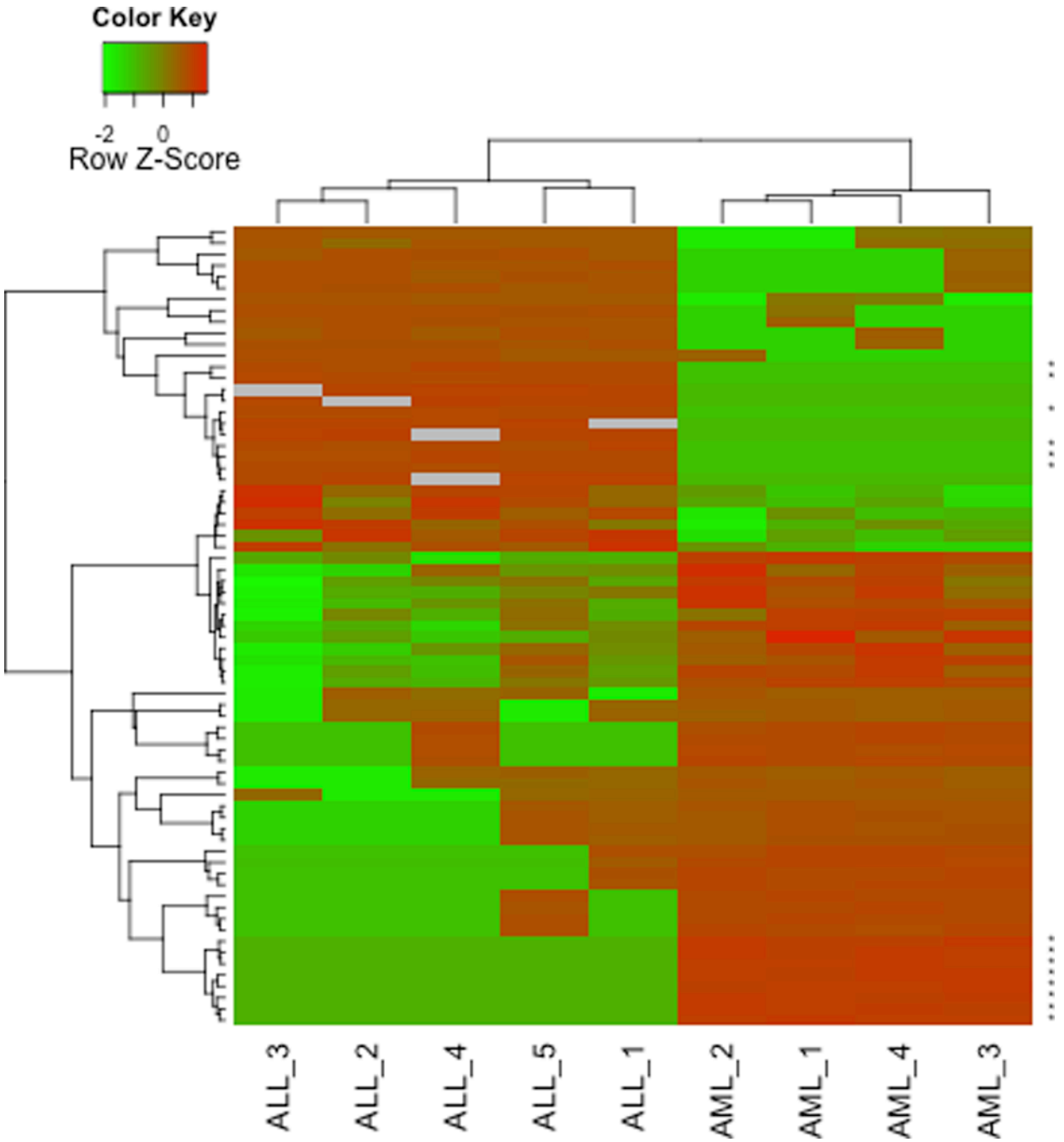


Figure 1.

Dendrogram and heat map showing similarities between 9 patient (ALL and AML) and 8 healthy (CD34+ and MNC) samples. Levels of 639 proteins were used to cluster samples according to Euclidean distance. Visual inspection of both the dendrogram and heat map demonstrate clear clustering of the four sample types, with AML and CD34+ showing the greatest degrees of similarity, as is expected. Colors indicate Euclidean distances between pairs of samples.





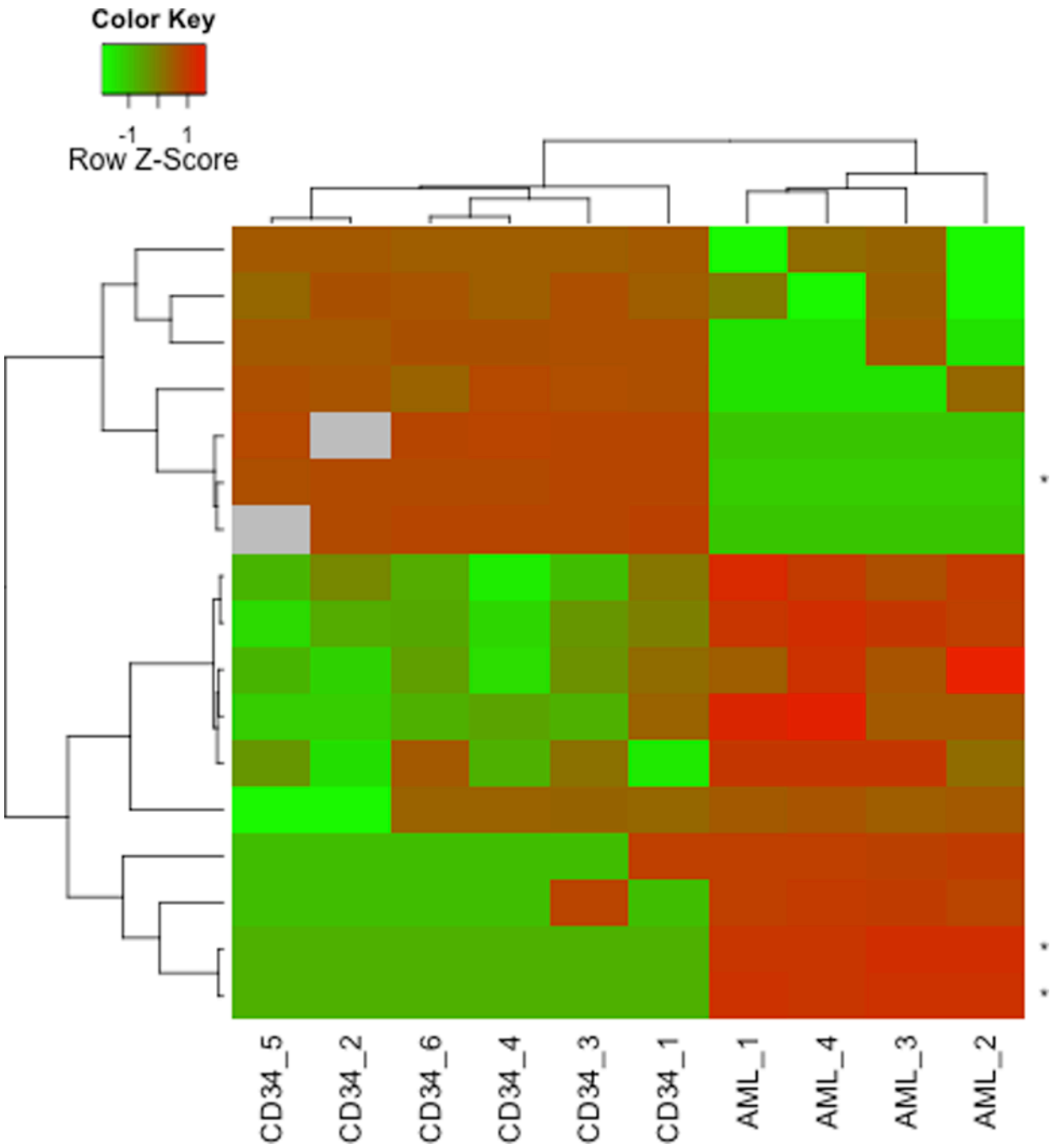


Figure 2.

A. Heat map depicting relative protein levels and dendrogram constructed on the basis of Euclidean distances between these protein levels showing the 91 proteins we identified as capable of distinguishing the 5 ALL samples from the 6 CD34+ samples. Each column represents a patient and each row represents a protein. Higher and lower protein levels are indicated in red and green, respectively. Gray boxes indicate missing values. The dendrograms on the top and left show unsupervised hierarchical clustering of patient samples and proteins, respectively. The "all or none" proteins are indicated by asterixes on the right.

B. Heat map and dendrogram as in figure 2A showing the 71 proteins we identified as capable of separating 5 ALL samples from 4 AML samples.
C. Heat map and dendrogram as in figure 2A showing the 17 proteins we identified as capable of separating 6 CD34+ samples from 4 AML samples.