Supporting Information

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

Figure S1. Dendrogram and heat map showing similarities between 58 patient (32 ALL and 26 AML) and 19 healthy (10 CD34+ and 9 MNC) samples. Levels of transcripts that correspond to the 639 proteins we measured were used to cluster samples according to Euclidean distance. Colors indicate Euclidean distances between pairs of samples.

Figure S2. Testing the transcripts corresponding to the proteins in figures 2A-C for their ability to distinguish classes. A. Heat map and dendrogram (constructed on the basis of Euclidean distance) showing the transcripts corresponding to 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 transcript. Higher and lower transcript 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 transcripts, respectively. There are more than 91 lines representing transcripts for 91 genes because multiple genes were represented by more than one probe. B. Heat map and dendrogram as in figure S2A showing transcripts corresponding to the 71 proteins we identified as capable of separating 5 ALL samples from 4 AML samples. C. Heat map and dendrogram as in figure S2A showing transcripts corresponding to the 17 proteins we identified as capable of separating 6 CD34+ samples from 4 AML samples.

Table S1

Quantitation of 639 proteins. Samples are listed in the first row according to sample type, an underscore and then sample number, e.g. the five ALL samples are described as ALL_1 through ALL_5. Quantitation was log base 2 of the median value for the sample divided by the average of the medians for all samples. If a particular class of samples had only a single missing value, we assumed that this was due to a misalignment rather than to the value being zero, and thus assigned it "NA"; otherwise, missing values were assigned zero, as they were assumed to be to low to be detected.

Table S2

Quantitation of transcripts corresponding to 639 proteins. The format is the same as for Table S1 except that the quantitation shows log base 2 ratios rather than absolute numbers. "NA" indicates missing data. Details of transcriptome measurements have been previously described.[36] Complete transcript measurements are available here:

http://www.interscience.wiley.com/jpages/1045-2257/suppmat

Table S3

List of proteins useful in distinguishing two sample types. The three possible comparisons are listed in columns C-E with the "winner" being the sample with the higher level of the protein in question. "All or none" proteins are listed in the last 3 columns.