cDNA microarray analysis of non-selected cases of acute myeloid leukemia demonstrates distinct clustering independent of cytogenetic aberrations and consistent with morphological signs of differentiation

ANNE MARGRETE ØYAN[1,2], TROND HELLEM BØ[2-4], INGE JONASSEN[3,4], BJØRN TORE GJERTSEN[5,6], ØYSTEIN BRUSERUD[5,6] and KARL-HENNING KALLAND[1,2]

[1]Department of Microbiology and Immunology, Haukeland University Hospital; [2]The Gade Institute, [3]Department of Informatics, University of Bergen; [4]Computational Biology Unit, Bergen Center for Computational Science, University of Bergen; [5]Section of Hematology, Haukeland University Hospital; [6]Institute of Medicine, University of Bergen, Bergen, Norway

Received November 8, 2005; Accepted December 27, 2005

Abstract. Acute myeloid leukemia (AML) is a heterogeneous disease with respect to biology and clinical course. Until now the basis for prognostic evaluation and therapeutic decision has been the karyotype, genetic FLT3 abnormalities and the initial chemotherapy response. A question that has emerged is if extensive gene expression analysis may supplement or partly replace current diagnostics. In an attempt to address this question, we performed cDNA microarray analysis on peripheral blood samples of 25 patients with newly diagnosed AML with high blast counts. The patients were randomly selected from a large group of consecutive patients. Leave-one-out crossvalidation (LOOCV) showed with high accuracy that gene expression classifiers could predict if leukaemia samples belonged to the FAB AML-M1 or to the FAB AML-M2 groups. An unsupervised two-dimensional hierarchical cluster analysis generated 3 patient subgroups. Except for an accumulation of samples classified as FAB M1 and M2 in cluster 3, there was no evident relationship between the clusters and the FAB classification. Each subgroup displayed clearly distinguished gene expression patterns validated using real-time quantitative PCR analysis. The identification of specific gene expressions that together constitute regulatory modules must complement cluster analyses in order to achieve an accurate basis for prognosis and prediction.

Correspondence to: Dr Anne Margrete Øyan, Department of Microbiology and Immunology, The Gade Institute, Haukeland University Hospital, Centre for Research in Virology, BBB, Jonas Lies v. 91, N-5009 Bergen, Norway
E-mail: anne.oyan@gades.uib.no

Key words: acute myeloid leukemia, cDNA microarray, gene expression, differentiation

Introduction

Acute myeloid leukaemia (AML) is an aggressive malignant disorder characterized by accumulation of immature myeloid progenitors in the bone marrow (1). AML can be divided into a heterogeneous majority and the less frequent acute promyelocytic leukaemia (APL) variant characterized by promyelocyte accumulation, chromosomal translocations involving 17q21, susceptibility to differentiation induction and good prognosis (2). The major prognostic parameters for non-APL variants are response to the initial chemotherapy cycle and the genetic abnormalities of the malignant cells (3,4). However, several prognostically important genetic abnormalities are associated with specific signs of differentiation in the malignant cells [e.g. inv(16); t(8;21)], suggesting that molecular mechanisms involved in regulation of differentiation may also contribute to the prognostic impact of these abnormalities (5). Recent DNA microarray studies suggest that AML classification can be further improved in terms of distinguishing AML from acute lymphoblastic leukaemia (ALL) (6-8), defining acute leukemias with MLL-abnormalities as a possible new entity separate from AML and ALL (9) and identifying new subgroups of AML with specific patterns of gene expression (10,11). Specific gene expression patterns are associated with prognostically important genetic abnormalities [i.e. t(8;21), inv16, t(8;21), t(15;17), 11q23-aberrations] (12-14). Prognostic molecular markers have also been found in AML blasts with normal cytogenetics (15,16).

Several new therapeutic approaches are now considered in AML, including differentiation induction through inhibitors of intracellular signaling and drugs targeting gene transcription (17). In a recent article, we emphasised the importance of collecting biological material as a part of future population-based clinical studies of AML, and at least two approaches are possible for integration of DNA microarray analysis in future clinical studies (18). Firstly, microarrays may be used to monitor effects of treatment, e.g. the effects of differentiation-induced therapy or therapeutic approaches targeting gene transcripts. This may imply patterns associated with the

Table I. Clinical and biological characteristics of acute myeloid leukemia patients.

Pts.	Sex	Age	Previous malignant disease	FAB classification	Non-supervised AML cluster[d]	Membrane molecule expression[a]					Cytogenetic abnormality	FLT3 abnormality[b]	WBC counts[c]
						CD13	CD14	CD15	CD33	CD34			
0043	F	64		AML-M1	C1	+	-	nt	-	-	Normal	-	84
0041	F	48		AML-M1	C3	+	-	-	+	+	del(7)(q22)	ITD	30
0050	M	64	MDS	AML-M1	C3	+	-	+	+	+	37-46XY	D835	12
0047	M	30		AML-M1	C1	-	-	nt	+	-	Normal	-	78
0048	F	54	Breast cancer	AML-M1	C3	+	-	nt	+	+	nt	ITD	35
0033	F	63		AML-M1	C3	+	-	+	-	+	nt	-	76
0031	M	83		AML-M2	Outgroup	+	-	nt	+	+	nt	-	49
0032	F	45		AML-M2	C2	+	-	-	+	-	Normal	ITD, D835	123
0034	F	52		AML-M2	C3	+	-	nt	+	+	Normal	-	58
0036	F	51		AML-M2	C1	+	-	+	+	-	Normal	ITD, wt^-	154
0040	M	77	Hodgkin's disease	AML-M2	Outgroup	+	-	-	+	+	del(7),-20	ITD	62
0049	F	58		AML-M2	C3	+	-	+	+	-	Normal	ITD, wt^-	41
0008	F	64		AML-M4	C2	-	-	+	+	+	Normal	-	123
0024	M	29		AML-M4	Outgroup	+	-	+	+	+	Normal	ITD, D835	19
0038	F	38		AML-M4	C1	-	+	+	+	-	Normal	-	228
0037	F	45	Ovary carcinoma	AML-M4	C2	+	-	+	+	-	Normal	-	70
0051	M	81	MDS	AML-M4	Outgroup	-	+	nt	-	-	nt	-	61
0023	F	78		AML-M4-5	C1	+	+	-	+	-	-4-5,+der(8)T(8;?)	-	93
0025	M	64		AML-M4-5	C2	+	-	-	+	+	Normal	ITD, wt^-	135
0007	M	82		AML-M5	C1	+	+	+	+	-	45X	-	198
0027	F	75		AML-M5	C1	+	-	+	+	+	Normal	ITD	104
0035	F	34		AML-M5	C1	-	-	+	+	-	t(9;11)(p22;q23)	D835	286
0045	F	58		AML-M5	C1	-	+	+	+	-	t(2;3)(q37;q21), t(2;4;10)(q13; q21)q21), der(11q),19q+	-	58
0044	M	33		AML-M5	C1	+	-	+	+	-	Normal	ITD	131
0046	F	71		AML-M5	C2	+	+	+	+	-	Normal	D835	64

F, female; M, male; AML, acute myeloid leukemia; MDS, myelodysplastic syndrome. The patient age is given in years. [a]Patients were regarded as positive when >20% of the blast cells stained positive judged by flow cytometric analysis. [b]FLT3 abnormalities were internal tandem duplicatons (ITD), Asp(D) 835 mutations (D835), and loss of wild-type (wt^-); nt, not tested. [c]White blood cell (WBC) counts in peripheral blood are expressed as $x10^9$/l (normal range 3.5-10.5$x10^9$/l). The WBC included at least 80% leukemia blasts.

various morphologically defined subsets within both the French-American-British (FAB) and the more recent World Health Organization classification (5,19). Secondly, the experience from highdose cytarabine therapy has demonstrated how the genetic characteristics of malignant cells can be used for prognostic classification of AML patients (20). DNA microarray analysis may then represent an additional approach for identification of new subsets with different responses to therapy. In the present study, we have therefore investigated native AML blasts derived from 25 patients by DNA microarray analysis. Patients were randomly selected from a larger consecutive group (21), and sample collection as well as RNA preparation were highly standardized (18). Despite considerable patient heterogeneity, the DNA microarray analysis allowed us to: i) identify differentiation-associated gene expression patterns; and ii) classify the patients into three major subgroups with abnormalities in the expression of genes involved in intercellular signalling, intracellular signaling, regulation of transcription, and tyrosine kinase-associated signaling using unsupervised hierarchical cluster analysis.

Materials and methods

Patient materials. The study was approved by the local Ethics Committee and samples collected after informed consent. During the time period of 1991-2001 we collected peripheral blood AML blasts from 64 consecutive patients with high blood blast counts (21). Cytogenetic analysis was performed for the last 48 of these 64 patients; 28 patients had a normal karyotype, 3 patients had a favourable [all inv(16)] and 5 had an unfavourable karyotype according to the definitions used (3,4). A total of 98 patients with AML were admitted to our institution during the same time period. A similar karyotype distribution was also observed for the whole patient group (the last 73 of the 98 patients examined, 6 patients having a

favorable karyotype). We investigated AML blasts derived from 25 patients (Table I; median age 58 years with variation range 29-83 years) that were randomly selected from the 64 consecutive patients with high peripheral blood blast counts (21). According to the FAB classification patients were classified as AML-M0/M1 (undifferentiated, 6 patients), AML-M2 (neutrophil differentiation, 6 patients) and AML-M4/M5 (monocyte differentiation, 13 patients). One ALL sample was included as a test sample.

Preparation of AML blast. Peripheral blood samples were collected on glass tubes (sample tubes with acidum cotrose-dextrose solution A; Becton-Dickinson) and leukemic peripheral blood mononuclear cells (PBMC) were isolated by density gradient separation (Ficoll-Hypaque specific density 1.077; NycoMed, Oslo, Norway;) immediately after sampling. We selected patients with a high percentage of AML blasts among the blood leukocytes (Table I), and highly enriched AML cell populations could therefore be prepared by density gradient separation alone (>95% blast cells judged by light microscopy). Cells were frozen without delay and stored frozen in liquid nitrogen (22). Cells were thawed and RNA prepared according to strictly standardized procedures (18).

Microarray analysis. Total RNA was extracted according to standard protocols (23). Synthesis of T7 RNA polymerase promoter-containing double-stranded cDNA and the generation of T7 RNA polymerase amplified RNA (cRNA) were performed as previously described (18) and according to the manufacturer's instructions (Ambion). Aminoallyl-dUTP incorporation followed by cross-coupling of Amersham Cy5- and Cy3-NHS esters was used for fluorochrome labeling of nucleic acids (18). AML sample cRNAs and the referance cRNAs made out of a pool of 18 different cell lines (Supplementary Table I) were labelled with Cy5 and Cy3, respectively.

The Human 1 cDNA Microarray system from Agilent containing 12.814 cDNA clones selected from the Incyte collection was used for the detection of gene expression in the AML samples. Hybridization and washing of the arrays were performed as recommended by Agilent (cDNA Microarray Kit Protocol). An Axon scanner recorded signal intensities at 532- and 633-nm laser lines, and the GenePix Pro software was used for feature extraction and creation of the GenePix intensity report file.

Normalization and analysis of DNA microarray data. After subtraction of background intensities for each spot, missing signals were floored using the value of 20. This procedure may include potentially interesting genes, i.e. genes that are detectably expressed only for a subset of patients, and we therefore performed data preparation using this method. The flooring of missing signals allowed us to produce log ratios for all spots, including those spots where the signal was only present in one channel. The resulting intensities were again subjected to the intra-array normalization procedure lowess and a gene expression data matrix was produced (24). The lowess method (25) was used to correct for dye-specific effects and log-ratios calculated producing a data matrix with one row per spot (gene) and one column per sample (array). We calculated each gene's two-sample t-score in one FAB class versus the other FAB classes, thereby identifying genes most consistently high or low in one FAB class relative to the other samples. P-values corresponding to the t-scores were found using the t-distribution with n-2 degrees of freedom, where n is the number of samples included in the calculation.

Crossvalidation, validation using permutation test and two-way cluster analysis. We used a leave-one-out crossvalidation (LOOCV) with the diagonal linear discriminant (DLD) (26) approach to test whether it is possible to predict FAB class M1, M2, M4, and M5 from the gene expression data. In this analysis, the samples labelled M4-5 were excluded. Each class was treated separately and we evaluated whether a classifier could be learnt that discriminates this one class from the others. In LOOCV, one sample (patient) is held back, a classifier is learnt using data on the remaining training samples and tested on the held back sample. This procedure is repeated with each patient as test sample and the prediction accuracy is the percentage of (held back) samples that are correctly classified. In this way, we obtained prediction accuracies for each FAB class. For each class, a permutation test was performed (1000 permutations of FAB labels followed by LOOCV) to assess whether an equally good accuracy could be obtained by chance.

The gene expression differences in AML samples between FAB classes M1, M2, M4 and M5 were visualized by taking the genes with the highest t-scores for each class as input to a two-way hierarchical cluster analysis using the J-Express Pro software (www.molmine.com) (27). We also included the two samples belonging to the AML FAB class M4-5 in the two-way cluster analysis to visualize their distribution compared to the other FAB classes.

Unsupervised cluster analysis. In order to explore possible new subtypes by analysis of the expression data, we performed an unsupervised hierarchical clustering of the 25 AML samples. For this, we selected genes with the highest variance among the AML blasts. Before clustering using J-Express Pro, the expression profile of each gene was centered by subtracting the mean over all samples for that gene. The genes and tissue were then hierarchically clustered using average linkage (WPGMA) (28) and Pearson correlation (29) as similarity metrics.

Testing of association between clusters and FAB classes. We performed a contingency table Chi-squared test (29) for association between clusters and FAB classes. Due to the relatively small number of samples, the samples were divided into two classes: one class containing samples with FAB class M1 and M2, and one class containing samples with FAB class M4 and M5. For each cluster, we performed a 2x2 contingency table test concerning significant high or low FAB representation in each cluster.

qPCR in the low density array (LDA) format. TaqMan low density arrays (LDA) are customisable, 384-well microfluidic cards for real-time PCR (Applied Biosystems). Each LDA card was configured for 95 different genes in duplicates

including Celera gene IDs. Hexamer-primed single-stranded cDNA corresponding to 10 ng of total RNA was diluted in TaqMan Universal buffer (Applied Biosystems) and added to each loading well. Using the above configuration each sample occupied 4-wells or one half of each card. The samples were distributed to the microwells by centrifugation for 1 min at 343 x g. The cards were sealed and placed in the ABI 7900HT Sequence Detection System using the following cycling parameters: 2 min at 50˚C, 10 min at 95˚C, and 40 alternate cycles of 15 sec at 95˚C and 60 sec at 60˚C. SDS2.2 software was applied for relative quantitation ($RQ/2^{-\Delta\Delta Ct}$) analysis using GAPDH as normalizer and one sample as calibrator. The data were exported to Excel for further exploration and visualization.

Pathway assist analysis of molecular interactions and cellular processes. Pathway Assist is a software supplied with a molecular interaction and pathway database (ResNet), which contains 500000 links to >50000 proteins extracted from 5000000 Medline full-length articles. Selected genes from each cluster (C1- C3) were used as input genes in an analysis using Pathway Assist software v3.2 (www.ariadnegenomics.com) to explore and visualise biological pathways, molecular interaction and cellular processes characteristic for each cluster. The pathways were built by using three commands: i) direct interaction between input genes; ii) common targets for input genes; and iii) common regulators for input genes.

Results

Differential gene expression for each FAB subclass versus the other AML samples. The gene expression profile was examined in native human AML blasts derived from 25 patients using the Human 1 cDNA Microarray from Agilent. In order to analyse the expression data for all genes, we first performed a flooring of background intensity values and calculated $\log_2$-ratio values based on all spots (30). We had previously found both variation and considerable overlap between gene lists based upon either floored or filtered datasets (30). A disadvantage with filtered data is the possible removal of genes that are expressed in only a minor subset of the tumors (30,31). In total, 6502 out of 12814 genes were over- or under-expressed compared to the reference probe, using a one sample t-test (29) and setting the p-value threshold to 0.05.

All arrays were selected for building a class prediction model: M1 AML (n=6), M2 AML (n=6), M4 AML (n=5), M5 AML (n=6). We calculated each gene's two-sample t-score in one FAB class versus the other FAB classes, thereby identifying the genes with the most consistent high or low expression in each FAB group compared with the remaining samples. For the 100 top scoring genes the t-scores were as follows for the different FAB classes: M1 from 10.18 to 4.02 (p-values of 1.4143E-9 to 6.1389E-4), M2 from 5.31 to 3.16 (p-values of 2.8911E-5 to 4.7359E-3), M4 from 6.02 to 3.02 (p-values of 5.6065E-6 to 6.3893E-3), M5 from 4.67 to 3.11 (p-values of 1.3043E-4 to 5.3606E-3). Considering that 12814 genes were tested, we expected some genes to be high scoring by chance. We expected 7.7 genes to score higher than 4.0, given the number of tests performed. Thus, we suspect only 8 of the 100 genes in our list for M1 to be false positives. We also expected 55 genes to score higher than 3.2, and 69.2 genes to score higher than 3.1 by chance. As a consequence of this, we expected the false discovery rate to be rather high for the gene lists for M2, M4 and M5. Still, we observe more genes than expected with high scores for these three classes.

Class prediction by use of leave-one-out crossvalidation (LOOCV). In order to evaluate the ability of subsets of expressed genes to predict FAB subclass specificity, LOOCV was employed. The prediction accuracy did not increase when >10 genes were included (we tried 10, 20, 30, ...100). The prediction success rates were 88% for M1, 84% for M2 and 80% for M4 and 60% for M5 versus all other samples when 10 genes were included in the classifier. Prediction accuracy did not increase when >10 genes were included. It should be emphasized that, for each of the n test samples, new gene subsets were extracted based only upon the (n-1) training samples in each round of the LOOCV. In this way, it was avoided that information based upon the test sample was included in the classifiers. Classification of M1 versus non-M1, M2 versus non-M2, etc. gives an impression of the degree of separability of each FAB class from the other classes. This is a preliminary step to multiclass prediction, e.g. direct classification into categories M1, M2, M4 and M5.

Validation using random permutations of AML sample labels. To further test the significance of the prediction accuracies achieved using LOOCV, a permutation test was performed. The FAB labels were permuted randomly followed by LOOCV as described above. Again the design was such that information did not leak from the test samples into the classifiers, which were derived from only the training set of samples. This permutation process was repeated 1000 times for each FAB subclass in order to obtain sufficient data for estimating the distribution of prediction success on random labels. The prediction success rates achieved with randomly permuted labels were then compared with that of the original correct labelling. For M1, 981 of 1000 rounds had lower prediction accuracy than the correct labelling. Using the corresponding tests for M2, M4 and M5, 960, 922 and 812 of 1000 rounds of permuted labels, respectively, gave a lower prediction accuracy than with correctly labelled samples. The permutation results gave approximate p-values of the prediction success rates obtained for the four FAB classes: 0.019 for M1, 0.040 for M2, 0.078 for M4 and 0.188 for M5. Thus, all class prediction success rates are higher than expected on average, although only significantly higher for M1 and M2.

Study of the 100 most differentially expressed genes for each FAB class. To investigate the correlations between FAB class labels and the expression profiles of the genes, we applied two-way hierarchical cluster analysis. The 100 genes were identified using a supervised analysis step where we calculated the t-score of each gene, quantifying how well it separates AML samples according to the FAB classes, and selected the top 100 from the resulting list for each FAB class. The resulting dendrogram visualises the relationships between the expression profiles of these genes and how they enable separation between the FAB classes (Supplementary Fig. 1). The 100 top t-scores ranked from 10.2 to 3.1 (p-values 1.4143E-9 to 6.3893E-3,

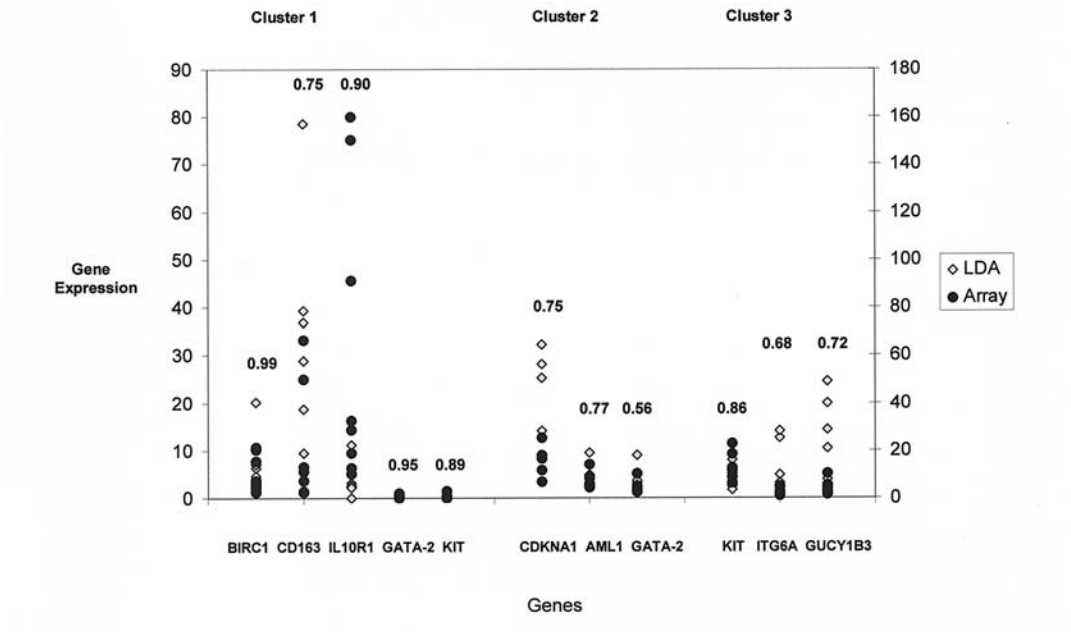


Figure 1. Correspondence between cDNA microarray and LDA gene expression. Total RNA was extracted from AML blasts and the relative mRNA expression levels were determined by qPCR TaqMan low density arrays (●, LDA) and Agilent cDNA microarrays (◇, Array). LDA analysis used GAPDH as normalizer and one sample as calibrator. GAPDH was used for normalisation of cDNA microarray values. The correlation coefficients between Array and LDA values were calculated and visualized for each of the following genes: BIRC1, CD163, IL10R1, GATA-2, KIT (cluster 1, 8 patients); CDKNA1, AML1, GATA-2 (cluster 2, 5 patients); and KIT, ITG6A, GUCY1B3 (cluster 3, 6 patients).

respectively). In FAB M1, only 27% of the 100 genes with the highest t-score showed overexpressed mRNAs relative to the other FAB classes. For comparison, 92% were upregulated in M2, 77% in M4 and 43% in FAB M5. More detailed information on the most consistently differentially expressed genes between FAB classes ranked by p-value is shown in Supplementary Table II. GenBank accession numbers, descriptions of genes and fold changes (D) are given.

Unsupervised hierarchical cluster analysis of AML samples. Based upon the 500 genes with highest variance in the floored dataset, an unsupervised hierarchical clustering of the 25 samples generated three main clusters of AML samples and an outgroup close to clusters 2 and 3 (Supplementary Fig. 2, black). The robustness of the clusters was tested by the inclusion of various numbers of genes in the cluster analysis. The main clusters persisted although the outgroup samples redistributed in different ways when 800, 500, 250, 100 or 50 genes were used. LOOCV gave the samples of cluster 1 versus the other samples (Supplementary Fig. 2, blue) a prediction accuracy of 96%; samples of cluster 2 versus the other samples (purple, middle of Supplementary Fig. 2), a 92% accuracy; and samples of cluster 3 versus the other samples (red, right of Supplementary Fig. 2), a 92% accuracy. A permutation test revealed the same success or better for randomly permuted labels for 0/1000 permutations for cluster 1 and 4/1000 permutations for clusters 2 and 3. The analysis was based upon the 50 genes with the highest t-score in the model.

Testing for association between clusters and FAB classes. A 2x2 contingency table Chi-square test did not demonstrate significant associations between FAB class and cluster 1 or cluster 2 [Chi-square values of 2.16 (p=0.23) and 1.96 (p=0.32), respectively]. The Chi-square test revealed, however, a significant association between FAB M1 and M2 samples and patient cluster 3 (p=0.0052).

Independent validation of gene expression data using real-time quantitative PCR and flow cytometry. Using the real-time quantitative PCR LDA format, 95 different genes were independently validated. In total, 17 of the genes in clusters C1-C3 were included in this analysis. The correspondence between gene expression according to the cDNA microarray (Agilent) and the quantitative PCR data (LDA) is displayed in Fig. 1. Validation of CD34 expression using flow cytometry analyses of the same samples was published previously (31).

Single genes that differed between the AML sample clusters of the unsupervised cluster analysis. In total, 76 genes exhibited a fold change between 50 and 3 for samples of one cluster compared to all other samples. In Supplementary Table III, the genes are ranked according to fold change accompanied by overall very strong p-values. Table II contains functional sorting in addition to fold change ranking. Cluster 1 samples shared a 25- to 4-fold higher expression of many immuno-regulatory cytokine receptors compared to the remaining samples (Table II). Monocyte/macrophage markers, MAC-1α (CD11b) and CYBB/gp91PHOX typically expressed in myeloid cells differentiated beyond the promyelocyte stage, were consistently upregulated in cluster 1. c-KIT mRNA was relatively strongly repressed in cluster 1 samples, while colony stimulating factor 1 receptor (CSF1R) that may substitute for KIT was overexpressed. TNFRSF1B forms a heterocomplex with TNF-receptor 1 and mediates the recruitment of two anti-apoptotic proteins, c-IAP1 and c-IAP2. TNFRSF1B, BIRC1/NAIP as well as MDR that mediate anti-apoptotic signals were

Table II. Genes with altered expression in AML clusters 1-3.

Altered expression of genes with important common characteristics				Additional genes with altered expression			
GenBank	Symbol	Gene name	D	GenBank	Symbol	Gene name	D
AML cluster 1							
Receptors and membrane molucules involved in cell signalling				Regulation of transcription and cell cycle			
↑ NM_006847	LILRB4	Leukocyte immunoglobulin-like receptor sfB4	25				
↑ U95626	CCRL2	Chemokine (C-C motif) receptor-like 2	20	↑ AA573434	KLF4	Kruppel-like factor 4 (gut)tf	8
↑ D10925	CCR1	Chemokine (C-C motif) receptor 1	17	↑ AA844153	AHR	Aryl hydrocarbon receptor	5
↑ U20350	CX3CR1	Chemokine (C-X3-C motif) receptor 1	16	↑ NM_004536	GAS7	Growth arrest-specific 7	4
↑ AF245703	TLR8	Toll-like receptor 8	15	↑ NM_004536	NAIP	Baculoviral IAP repeat-containing 1	4
↑ AA402981	KCTD12	Voltage-sensitive potassium channel complex	13	↑ M81934	CDC25B	Cell division cycle 25B	3
↑ NM_004244	CD163	CD163 antigen, Scavenger receptor	12	↑ U66306	RXRA	Retinoid X receptor, α	3
↑ U00672	IL10RA	Interleukin 10 receptor, α	11	↓ NM_006022	TSC22	TGF ß-stimulated protein	6
↑ X04011	CYBB	Cytochrome b-245, ß, gp91PHOX	10				
↑ Z82244	HO-1	Heme oxygenase (decycling) 1	9				
↑ M_003264	TLR2	Toll-like receptor 2	8				
↑ M18044	ITGAM	MAC-1α, CD11b (p170)	8				
↑ U37518	TNFSF10	Tumor necrosis factor (ligand) superfamily 10	8				
↑ AL034562	PTPNS1	Protein tyrosine phosphatase, non-receptor S1	6				
↑ M32315	TNFRSF1B	TNFR superfamily, member 1B	4				
↑ X03663	CSF1R	Colony stimulating factor 1 receptor (v-fms)	4				
↑ Y00081	IL-6	Interleukin 6	3				
↓ NM_000222	KIT	Stem cell factor receptor (SCFR)	14				
AML cluster 2							
Regulation of transcription and cell cycle				Membrane molecules, soluble mediators			
↑ M68891	GATA-2	GATA binding protein 2 transcription factor	14	↑ J03745	MMP2	Matrix metalloproteinase 2	20
AL031846	CBX7	Chromobox protein homolog 7	13	↓ V00522	HLA-DR ß3	MHC, class II, DR ß3	15
↑ U66838	CCNA1	Cyclin A1	11	↓ X13334	CD14	CD14 antigen	11
↑ U79260	MGC5149	Similar to FTO protein and MLL5	8	↓ NM_003474	ADAM12	A disintegrin and metalloproteinase 12	6
↑ U51869	CPBP	Core promoter element binding protein	5				
↑ AW380330	PBXIP1	Hematopoietic PBX-interacting protein	4	Others			
↑ AI825989	ZFP36L2	EGF-response factor 2	4	↑ M73720	CPA3	Carboxypeptidase A3 (mast cell)	81
↑ D43968	RUNX1	Acute myeloid leukemia 1 transcription factor	3				
↓ M32011	NCF2	Neutrophil cytosolic factor 2 (gp67PHOX)	10				
↓ NM_004166	CCL15	Chemokine (C-C motif) ligand 15	5				
↓ AA203476	PTTG1	Pituitary tumor-transforming 1	4				
↓ D14134	RAD51	DNA repair protein RAD51 homolog 1	3				
↓ Z93016	PRG5	p53-responsive gene 5	3				
AML cluster 3							
Soluble mediators, cell membrane receptors				Regulation of transcription, development andcell cycle			
↑ AL035091	CD34	CD34 antigen	15	↑ AB037762	MEF-2	Myelin gene expression factor 2	5
↑ X06182	KIT	Stem cell factor receptor (SCFR)	4	↓ AB007854	GAS7	Growth arrest-specific 7	3
↑ M62424	F2R	Coagulation factor II (thrombin) receptor	4	↓ U49278	UBE2V1	Ubiquitin-conjugating enzyme E2 v 1	3
↑ AF186111	EGFL7	EGF-like-domain, multiple 7/NEU1	4				
↑ NM_002845	PTPRM	Protein tyrosine phosphatase, receptor type, M	4				
↑ AB032994	CYFIP2	p53 inducible protein	4				
↑ X66533	GUCY1B3	Guanylate cyclase 1, soluble, ß 3	4				
↑ M34667	PLCG1	Phospholipase C, γ 1	3				
↑ X53586	ITGA6	Integrin, α 6/CD49f	2				
↓ NM_080792	PTPNS1	Protein tyrosine phosphatase, non-receptor S1	5				
↓ NM_004119	FLT3	fms-related tyrosine kinase 3	5				
↓ M62880	ITGB7	Integrin, ß 7	5				
↓ AI828515	TNFSF13	Tumor necrosis factor (ligand) SF13	5				

D is relative gene expression.

7- to 4-fold overexpressed in this cluster. For the possible functional relationship of upregulated CD163, IL10RA, HO-1 and MDR1, see Supplementary Fig. 3 and Discussion.

Cluster 2 was characterized by upregulation (D=2.5-14) of several hematopoietic transcription factors. GATA-2 (D=14) is ubiquitously expressed in hematopoietic cells, with particularly high expression in early hematopoietic progenitors as well as mast cell lineages, and declines with blood cell maturation (32). Other myeloid-specific factors that may enforce leukemisation are RUNX1, PBXIP. A striking feature of cluster 2 was an average 50-fold overexpression of carboxypeptidase A3, an enzyme associated with mast cell granules. The body's sole histamine-producing enzyme, histidine decarboxylase, was also highly upregulated in these samples (D=35.0). EGF-2 response factor, also known as butyrate response factor 2, is among the most consistently upregulated genes in cluster 2 (D=3.9).

Cluster 3 samples exhibited an average of 15-fold upregulation of CD34 mRNA, and 2.5- to 4-fold overexpression of integrin α 6 (ITGA6), coagulation factor II (thrombin) receptor (F2R), and c-KIT, all membrane molecules associated with myeloid progenitor cells. As shown in the Pathway Assist-derived Supplementary Fig. 4, c-KIT and thrombin receptor F2R and CD34 may all activate the PI3K pathway and AKT1. The receptor tyrosine kinase c-KIT was relatively overexpressed and FLT3 was underexpressed, while tyrosine phosphatases PTPNS1 (D=5) and PTPRM (D=4) were over- and underexpressed, respectively, in cluster 3 (Table II).

Discussion

The most important prognostic parameters in AML are currently the karyotype, genetic FLT3 abnormalities and response to the initial chemotherapy course (3,4). These parameters probably mirror important parts of the complex epigenetic and genetic alterations in AML, and demonstrate that the peripheral AML cells comprise information that determines highly effective anti-leukemic treatment in a subset of the cases (20). Based on the fact that most AML patients are >60 years of age, and that the overall long-term survival in AML is <50%, there is a striking need for new therapeutics with more acceptable systemic toxicity (33). Our present study suggests that additional use of DNA microarrays can identify differentiation-associated gene expression patterns and define new patient subsets. The results thereby indicate that this approach may become useful for understanding the pathogenesis involved, for additional prognostic classification of patients and monitoring of differentiation induction therapy. We only included patients with high blast counts in the peripheral blood, and highly enriched AML blast populations could be prepared by density gradient separation of blood samples. This simple technique has a minimal risk of inducing functional alterations in the blasts [for a detailed discussion and additional references, see Bruserud *et al* (34)]. Furthermore, cells were collected, separated, frozen and thawed according to strictly standardized procedures (22).

Our patients were randomly selected from a larger group of consecutive patients with high peripheral blood blast counts (21). Our previous study demonstrated that a normal karyotype was detected for nearly 60% of our patients, whereas favourable and unfavourable karyotypes were detected only for small patient subsets (6% and 11%, respectively) (21). This distribution is similar to other studies (3,4,35) except for the low frequency of favourable karyotypes that is probably due to the relatively large number of elderly patients in our study (median age 59 years) (21,36). Furthermore, the higher frequencies of FLT3/ITD [40% versus 27%, Bruserud *et al* (21)] and D-835 mutations [20% versus 7%, Glenjen *et al* (37)] among our patients may reflect our selection of patients with a high degree of leukemization.

The present study found gene expression classifiers with a significant, although not very high, ability to predict FAB subclasses of AML. Both LOOCV and permutations were used to validate the prediction accuracies. Only M1 and M2 samples had expression signatures that were strong enough for use in a classification model. One biological reason may be that samples of FAB classes M1 and M2 are more homogeneous, while FAB classes M4 and M5 contain samples of more diverse gene expression. It is not unexpected that the most striking characteristic of undifferentiated AML cells (FAB-M0/M1) was downregulation of several genes involved in cell differentiation and organ development. On the other hand, granulocytic differentiation (FAB-M2) was associated with a heterogeneous gene expression pattern with regard to molecular function. In contrast, AML cells with monocytic differentiation (FAB-M4/M5) showed altered expression of genes mainly involved in intracellular signal transduction and regulation of DNA-repair, cell cycle events or DNA transcription (Supplementary Table II). This last observation is consistent with reports regarding essential regulation of gene expression during monocyte differentiation (38,39). Previous studies have reported that unsupervised cluster analysis sorted AML samples according to cytogenetic abnormalities rather than FAB phenotypes (12) and that gene expression correlates with the common AML translocations [t(8;21), t(15;17), inv(16)], and translocations involving 11q23 have also been described (12,40-42). However, these abnormalities are uncommon, especially among the elderly majority of the consecutive population-based patient group. Even though we investigated highly heterogeneous patients randomly selected from a consecutive group with another distribution of cytogenetic abnormalities due to high age, our results are in agreement with these previous studies of relatively young selected patients: gene expression classifiers did not predict FAB subclasses with a very high accuracy but, according to t-scores, we were able to distinguish patients of a particular FAB subclass from the others.

Unsupervised hierarchical cluster analysis has a potential to reveal gene clusters shared by subsets of patient samples. This analysis is independent of previous sample classification such as FAB classification. Instead, the samples may segregate into new clusters or groups that share significant gene expression patterns. When the 500 genes with highest variance across the samples were selected for the unsupervised hierarchical cluster analysis, as visualized in Supplementary Fig. 2, three main clusters of 25 AML samples resulted. In addition, four patient samples segregated as an outgroup. LOOCV and permutation tests showed that classifiers were able to predict the three main clusters of heterogeneous AML samples with a very high accuracy. Except for a significant accumulation

of FAB M1 and M2 samples in cluster 3, there was no evidence that the non-supervised clustering was affected by FAB subclasses. The segregation of many M1 and M2 samples into cluster 3 therefore supports the above proposal that the M1 and M2 samples of this study may share more gene expression patterns than M4 and M5 samples. It has been pointed out that the differentiation stage of the lineage, as reflected in the FAB classification, may direct unsupervised clustering to an extent that could obscure gene expression of more critical significance for prognosis and prediction (16). This idea is partly supported by the results described above. Supervised and unsupervised cluster analysis may therefore have to be supplemented by directed search and identification of expression of a number of genes that together constitute a critical regulatory module. In the future, the identification of precisely characterized modules of genes may provide the best targets for tailored treatment. We present several putatative gene expression modules.

The expression of a number of genes in clusters 1, 2 and 3 (Table II and Supplementary Table III) appears relevant for the pathogenesis of AML. When several consistently regulated genes fit together into one regulatory module, the probability of functional relevance is high compared to observations on single genes. This is exemplified in cluster 1 by the relative overexpression of CD163 (D=12) and IL10RA (D=10). Although not previously described in AML, there are several reports that CD163 and IL-10 may be involved in a regulatory loop (43-45) that leads to heme oxygenase 1 (HO-1) induction (46,47). Examination of the 100 top scoring genes revealed that HO-1 is indeed consistently upregulated (D=8.6, p=2.8274E-4) in cluster 1 samples. Recently, resistance to apoptosis has been reported in gastric cancer cells with elevated HO-1 and c-IAP2 activity (48). We noticed that NAIP/BIRC1 (IAP repeat-containing 1) was also overexpressed (D=4, p=7.5774E-4) in cluster 1 (Supplementary Fig. 3). It is therefore of interest to further explore whether the IL-10, CD163, HO-1 regulatory module has been diverted and selected to counteract terminal differentiation and apoptosis (49) and thus enhance the proliferation and survival of these AML blasts. Another pathway that may be specifically activated in this group of AML patients is the NF-κB pathway indicated by upregulation of NF-κB-targeted genes such as ICAM-1, IL-6, TNFRSF1B and NAIP (50).

Altered expression of genes encoding transcription and chromatin condensation and cell cycle regulation factors was common for cluster 2 samples. Increased expression of the GATA-2, RUNX1, CBX7, cyclin A1 and defender against cell death 1 (DAD1) and repression of NCF2 (D=11) that is transcribed exclusively in myeloid cells that have differentiated beyond the promyelocyte stage, may contribute to leukemisation of hematopoietic cells with normal cytogenetics.

In cluster 3, hematopoietic stem cell marker CD34 mRNA was highly expressed. Attempts have been made to define a core gene expression pattern associated with stem cells ('stemness') (51,52). This concept has met with difficulty since there is very little overlap between gene lists extracted by three different groups in order to define 'stemness' (53,54). A comparison of three independently derived lists of 'stemness' genes showed only one common gene, integrin α 6 (ITGA6), in the three studies (54). Intriguingly, this particular gene (ITGA6, p=5.8212E-5) is the second most consistently upregulated gene in cluster 3. CD34, KIT and thrombin receptor have all been associated with stem cell expression (55) and were all consistently and highly overexpressed in cluster 3 (Table II) suggesting that samples of cluster 3 share important features associated with gene expression of early myeloid precursors and possibly bone marrow stem cells (Supplementary Fig. 4).

Cluster 1 to cluster 3 subsets defined by unsupervised cluster analysis indicate that patient subpopulations may differ with regard to pathogenesis. Cluster 1 was characterized by altered expression of several soluble mediators and membrane molecules, including receptors for immunoregulatory cytokines, which may in turn activate anti-apoptotic genes and genes that mediate drug resistance. Cluster 2 (patients with normal cytogenetics) was characterized by altered expression of genes encoding the transcription/cell cycle machinery and intracellular signaling molecules. The characteristic pattern of receptor tyrosine kinase and phosphatase expression of cluster 3 samples was striking, including a prominent KIT/FLT3 (56) expression ratio in combination with downregulation of the tyrosine phosphatase PTPNS1. Such patterns may be helpful in the pursuit of new therapeutic approaches.

To conclude, our present results demonstrate that gene expression profiles can be used to: i) characterize the differentiation status of native human AML cells; and ii) define new subsets among highly heterogeneous AML patients. We suggest that supervised and unsupervised cluster analyses will have to be complemented with exact identification of regulatory gene modules that may or may not segregate into specific clusters. The new tools for global analysis of gene expression supplemented with appropriate validation and experiments have the potential to achieve this goal.

Acknowledgements

This work was funded by the Haukeland University Hospital Research Fund, the Norwegian Cancer Society, and the Olaf Ruunshaugens legat. Hua My Hoang is acknowledged for excellent technical assistance with nucleic acid labeling and hybridization. Beth Johannessen is acknowledged for excellent work with cell lines and RNA purification.

References

1. Smith M, Barnett M, Bassan R, Gatta G, Tondini C and Kern WK: Adult acute myeloid leukaemia. Crit Rev Oncol Hematol 50: 197-222, 2004.
2. Avvisati G, Lo Coco F and Mandelli F: Acute promyelocytic leukemia: clinical and morphologic features and prognostic factors. Semin Hematol 38: 4-12, 2001.
3. Wheatley K, Burnett AK, Goldstone AH, Gray RG, Hann IM, Harrison CJ, Rees JK, Stevens RF and Walker H: A simple, robust, validated and highly predictive index for the determination of risk-directed therapy in acute myeloid leukaemia derived from the MRC AML 10 trial. United Kingdom Medical Research Council's Adult and Childhood Leukaemia Working Parties. Br J Haematol 107: 69-79, 1999.
4. Kottaridis PD, Gale RE, Frew ME, Harrison G, Langabeer SE, Belton AA, Walker H, Wheatley K, Bowen DT, Burnett AK, Goldstone AH and Linch DC: The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. Blood 98: 1752-1759, 2001.

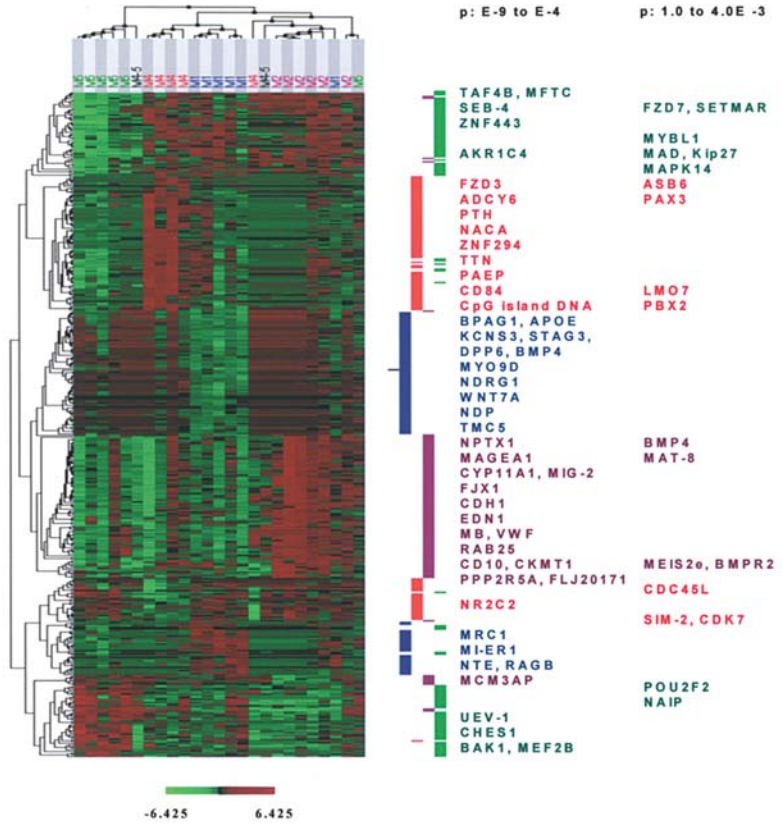
5. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J, Lister TA and Bloomfield CD: World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee Meeting-Airlie House, VA, 1997. J Clin Oncol 17: 3835-3849, 1999.
6. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
7. Moos, PJ, Raetz EA, Carlson MA, Szabo A, Smith FE, Willman C, Wei Q, Hunger SP and Carroll WL: Identification of gene expression profiles that segregate patients with childhood leukemia. Clin Cancer Res 8: 3118-3130, 2002.
8. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, Cheng C, Campana D, Wilkins D, Zhou X, Li J, Liu H, Pui CH, Evans WE, Naeve C, Wong L and Downing JR: Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 1: 133-143, 2002.
9. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR and Korsmeyer SJ: MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 30: 41-47, 2002.
10. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, Beverloo HB, Moorhouse MJ, van der Spek PJ, Lèowenberg B and Delwel R: Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 350: 1617-1628, 2004.
11. Bullinger L, Dèohner K, Bair E, Frèohling S, Schlenk RF, Tibshirani R, Dèohner H and Pollack JR: Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 350: 1605-1616, 2004.
12. Debernardi S, Lillington DM, Chaplin T, Tomlinson S, Amess J, Rohatiner A, Lister TA and Young BD: Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. Genes Chromosomes Cancer 37: 149-158, 2003.
13. Yagi T, Morimoto A, Eguchi M, Hibi S, Sako M, Ishii E, Mizutani S, Imashuku S, Ohki M and Ichikawa H: Identification of a gene expression signature associated with prognosis of pediatric AML. Blood 102: 1849-1856, 2003.
14. Kohlmann A, Schoch C, Schnittger S, Dugas M, Hiddemann W, Kern W and Haferlach T: Molecular characterization of acute leukemias by use of microarray technology. Genes Chromosomes Cancer 37: 396-405, 2003.
15. Marcucci G, Mrâozek K and Bloomfield CD: Molecular heterogeneity and prognostic biomarkers in adults with acute myeloid leukemia and normal cytogenetics. Curr Opin Hematol 12: 68-75, 2005.
16. Bullinger L and Valk PJ: Gene expression profiling in acute myeloid leukemia. J Clin Oncol 23: 6296-6305, 2005.
17. Hatfield KJ, Olsnes AM, Gjertsen BT and Bruserud Ø: Anti-angiogenic therapy in acute myelogenous leukemia: targeting of vascular endothelial growth factor and interleukin 8 as possible antileukemic strategies. Curr Cancer Drug Targets 5: 229-248, 2005.
18. Gjertsen BT, Øyan AM, Marzolf B, Hovland R, Gausdal G, Døskeland SO, Dimitrov K, Golden A, Kalland KH, Hood L and Bruserud Ø: Analysis of acute myelogenous leukemia: preparation of samples for genomic and proteomic analyses. J Hematother Stem Cell Res 11: 469-481, 2002.
19. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR and Sultan C: Proposals for the classification of the acute leukaemias. French-American-British (FAB) Co-operative Group. Br J Haematol 33: 451-458, 1976.
20. Löwenberg B, Downing JR and Burnett A: Acute myeloid leukemia. N Engl J Med 341: 1051-1062, 1999.
21. Bruserud Ø, Hovland R, Wergeland L, Huang T and Gjertsen BT: Flt3-mediated signaling in human acute myelogenous leukemia (AML) blasts: a functional characterization of Flt3-ligand effects in AML cell populations with and without genetic Flt3 abnormalities. Haematologica 88: 416-428, 2003.
22. Bruserud Ø: Effect of dipyridamole, theophyllamine and verapamil on spontaneous *in vitro* proliferation of myelogenous leukaemia cells. Acta Oncol 31: 53-58, 1992.
23. Chomczynski P and Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156-159, 1987.
24. Causton HC, Quackenbush J and Brazma A: Microarray Gene Expression Data Analysis - A Beginner's Guide. Blackwell Publishing, Oxford, 2003.
25. Yang YH, Buckley MJ and Speed TP: Analysis of cDNA microarray images. Briefings Bioinformatics 2: 341-349, 2001.
26. Mardia KV, Kent JT and Bibby JM: Multivariate Analysis. Academic Press, London, p300, 1979.
27. Dysvik B and Jonassen I: J-Express: exploring gene expression data using Java. Bioinformatics 17: 369-370, 2001.
28. Sokal RR and Michener CD: A statistical method for evaluating systematic relationships. Univ Kans Sci Bull 38: 1409-1438, 1958.
29. Bhattacharyya GR and Johnson RA: Statistical Concepts and Methods. Wiley, New York, p400, 1977.
30. Halvorsen OJ, Øyan AM, Bø TH, Olsen S, Rostad K, Haukaas SA, Bakke AM, Marzolf B, Dimitrov K, Stordrange L, Lin B, Jonassen I, Hood L, Akslen LA and Kalland KH: Gene expression profiles in prostate cancer: association with patient subgroups and tumour differentiation. Int J Oncol 26: 329-336, 2005.
31. Øyan AM, Bø TH, Jonassen I, Ulvestad E, Gjertsen BT, Kalland KH and Bruserud Ø: CD34 expression in native human acute myelogenous leukemia blasts: differences in CD34 membrane molecule expression are associated with different gene expression profiles. Cytometry 64: 18-27, 2005.
32. Persons DA, Allay JA, Allay ER, Ashmun RA, Orlic D, Jane SM, Cunningham JM and Nienhuis AW: Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. Blood 93: 488-499, 1999.
33. Bruserud Ø, Gjertsen BT and Huang TS: Induction of differentiation and apoptosis - a possible strategy in the treatment of adult acute myelogenous leukemia. Oncologist 5: 454-462, 2000.
34. Bruserud Ø, Gjertsen BT, Foss B and Huang TS: New strategies in the treatment of acute myelogenous leukemia (AML): *in vitro* culture of aml cells - the present use in experimental studies and the possible importance for future therapeutic approaches. Stem Cells 19: 1-11, 2001.
35. Thiede C, Steudel C, Mohr B, Schaich M, Schèakel U, Platzbecker U, Wermke M, Bornhèauser M, Ritter M, Neubauer A, Ehninger G and Illmer T: Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. Blood 99: 4326-4335, 2002.
36. Mrâozek K, Heinonen K and Bloomfield CD: Clinical importance of cytogenetics in acute myeloid leukaemia. Best practice and research. Clin Haematol 14: 19-47, 2001.
37. Glenjen N, Hovland R, Wergeland L, Wendelbo, Ernst P and Bruserud Ø: The angioregulatory phenotype of native human acute myelogenous leukemia cells: influence of karyotype, Flt3 abnormalities and differentiation status. Eur J Haematol 71: 163-173, 2003.
38. Kowenz-Leutz E, Herr P, Niss K and Leutz A: The homeobox gene GBX2, a target of the myb oncogene, mediates autocrine growth and monocyte differentiation. Cell 91: 185-195, 1997.
39. Glynne RJ and Watson SR: The immune system and gene expression microarrays - new answers to old questions. J Pathol 195: 20-30, 2001.
40. Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W and Schoch C: Global approach to the diagnosis of leukemia using gene expression profiling. Blood 106: 1189-1198, 2005.
41. Schoch C, Kern W, Kohlmann A, Hiddemann W, Schnittger S and Haferlach T: Acute myeloid leukemia with a complex aberrant karyotype is a distinct biological entity characterized by genomic imbalances and a specific gene expression profile. Genes Chromosomes Cancer 43: 227-238, 2005.
42. Schoch C, Kohlmann A, Schnittger S, Brors B, Dugas M, Mergenthaler S, Kern W, Hiddemann W, Eils R and Haferlach T: Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. Proc Natl Acad Sci USA 99: 10008-10013, 2002.
43. Buechler C, Ritter M, Orsâo E, Langmann T, Klucken J and Schmitz G: Regulation of scavenger receptor CD163 expression in human monocytes and macrophages by pro- and anti-inflammatory stimuli. J Leukoc Biol 67: 97-103, 2000.

44. Sulahian TH, Hèogger P, Wahner AE, Wardwell K, Goulding NJ, Sorg C, Droste A, Stehling M, Wallace PK, Morganelli PM and Guyre PM: Human monocytes express CD163, which is upregulated by IL-10 and identical to p155. Cytokine 12: 1312-1321, 2000.
45. Walter RB, Bachli EB, Schaer DJ, Ruegg R and Schoedon G: Expression of the hemoglobin scavenger receptor (CD163/HbSR) as immunophenotypic marker of monocytic lineage in acute myeloid leukemia. Blood 101: 3755-3756, 2003.
46. Otterbein LE and Choi AM: Division of Pulmonary A: heme oxygenase: colors of defense against cellular stress. Am J Physiol 279: L1029-L1037, 2000.
47. Philippidis P, Mason JC, Evans BJ, Nadra KM, Taylor DO, Haskard RC and Landis RC: Hemoglobin scavenger receptor CD163 mediates interleukin-10 release and heme oxygenase-1 synthesis: anti-inflammatory monocyte-macrophage responses *in vitro*, in resolving skin blisters *in vivo*, and after cardiopulmonary bypass surgery. Circ Res 94: 119-126, 2004.
48. Liu ZM, Chen GG, Ng EK, Leung WK, Sung JJ and Chung SS: Upregulation of heme oxygenase-1 and p21 confers resistance to apoptosis in human gastric cancer cells. Oncogene 23: 1-11, 2003.
49. Brouard S, Otterbein LE, Anrather J, Tobiasch E, Bach FH, Choi AM and Soares MP: Carbon monoxide generated by heme oxygenase 1 suppresses endothelial cell apoptosis. J Exp Med 192: 1015-1026, 2000.
50. Takada Y, Murakami A and Aggarwal BB: Zerumbone abolishes NF-kappaB and IkappaBalpha kinase activation leading to suppression of antiapoptotic and metastatic gene expression, upregulation of apoptosis, and downregulation of invasion. Oncogene 24: 6957-6969, 2005.
51. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA and Lemischka IR: A stem cell molecular signature. Science 298: 601-604, 2002.
52. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC and Melton DA: 'Stemness': transcriptional profiling of embryonic and adult stem cells. Science 298: 597-600, 2002.
53. Evsikov AV and Solter D: Comment on ''Stemness': 'transcriptional profiling of embryonic and adult stem cells' and 'a stem cell molecular signature'. Science 302: 393, 2003.
54. Fortunel NO, Otu HH, Ng HH, Chen J, Mu X, Chevassut T, Li X, Joseph M, Bailey C, Hatzfeld JA, Hatzfeld A, Usta F, Vega VB, Long PM, Libermann TA and Lim B: Comment on ' 'Stemness': transcriptional profiling of embryonic and adult stem cells' and 'a stem cell molecular signature'. Science 302: 393, 2003.
55. Ramalho-Santos J, Schatten G and Moreno RD: Control of membrane fusion during spermiogenesis and the acrosome reaction. Biol Reprod 67: 1043-1051, 2002.
56. Care RS, Valk PJ, Goodeve AC, Abu-Duhier FM, Geertsma-Kleinekoort WM, Wilson GA, Gari MA, Peake IR, Lèowenberg B and Reilly JT: Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. Br J Haematol 121: 775-777, 2003.

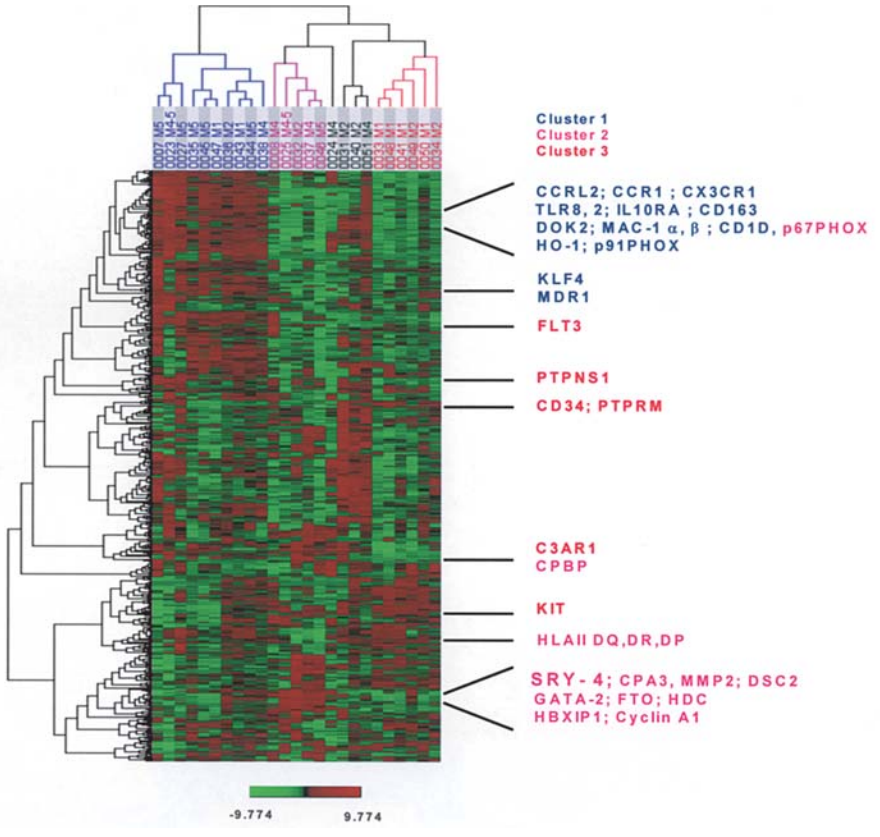
Supplementary Table I. Cell lines used as reference probe.

Cell lines	Tissue description	ATCC No./References
GA-10	B lymphocyte; Burkitt's lymphoma	CRL-2392
MOLT-4	T lymphoblast; acute lymphoblastic leukaemia	CRL-1582
U-937	Histiocytic lymphoma	CRL-1593.2
ST486	Ascites; B lymphocyte; Burkitt's lymphoma	CRL-1647
K-562	Bone marrow; chronic myelogenous leukemia (CML)	CCL-243
KG-1a	Bone marrow; acute myelogenous leukaemia (AML)	CCL-246.1
Jurkat γ1.wt	T lymphocyte; acute T cell leukaemia	CRL-2679
HL-60	Peripheral blood; promyeloblast; promyeloblast; acute promyelocytic leukaemia	CCL-240
MCF7	Mammary gland; breast; epithelial; metastatic site: pleural effusion adenocarcinoma	HTB-22
U-2 OS	Bone; osteosarcoma	HTB-96
HEL 299	Lung; fibroblast; normal	CCL-137
HeLa	Cervix; epithelial; adenocarcinoma	CCL-2
Placenta	Human placenta; normal	[a]
HUV-EC-C	Umbilical vein; vascular endothelium; endothelial; normal	[b]
SaOS-2	Bone; osteosarcoma	HTB-85
RF-48	Stomach; metastatic site: ascites gastric adenocarcinoma	CRL-1863
NCI-N87	Stomach; metastatic site: liver gastric carcinoma	CRL-5822
HLF-a	Lung; epidermoid carcinoma	CCL-199

[a]Obtained from Professor Helga Salvesen, Haukeland University Hospital. [b]HUVEC cell line was made by Therese Visted, University of Bergen (e-mail: therese.visted@pki.uib.no).



Supplementary Figure 1. Two-way cluster analysis of AML samples. Supervised two-way cluster analysis based upon 100 genes with top t-scores discriminating patients of FAB classes M1 (blue), M2 (violet), M4 (red) and M5 (green). The two samples classified as FAB M4-5 are included in the two-way cluster analysis (black). The corresponding p-values of the genes characteristically expressed in M1, M2. M4 and M5 ranged from 1.4143E-9 to 6.3893E-3. Gene clusters are indicated by vertical bars and a few selected genes of each cluster were extracted and highlighted in the same colour.



Supplementary Figure 2. Unsupervised two dimensional hierarchical cluster analysis of gene expression in AML. Unsupervised two dimensional hierarchical cluster analysis of different FAB classes based upon 500 of 12814 genes with highest variance. The clustering defines three sample clusters: cluster 1 (blue), cluster 2 (purple), cluster 3 (red). The gene profiles were centered by subtracting the mean before clustering. Characteristic genes of each cluster are indicated to the right.

Supplementary Table II. Genes that separate the AML FAB classes.

A GenBank	Gene symbol	Gene name	Relative gene expression D	p-value
Genes overexpressed in M1				
AK024443	NTE	Similar to neuropathy target esterase, anti apoptotic	3.0	3.7247E-4
NM_006064	RAGB	GTP-binding protein ragB	2.6	1.9828E-4
X93093	ICAM4	Intercellular adhesion molecule 4	2.1	1.6744E-4
X87344	MDR/TAP	Transporter 2, ATP-binding cassette, sub-family B	2.1	3.9473E-4
AB046830	MI-ER1	Mesoderm induction early response 1	2.0	2.0080E-4
X91826	ZNF75a	Zinc finger protein 75a	1.9	1.3696E-4
M84337	PAP	Pancreatitis-associated protein	1.7	4.9653E-4
M93215	MRC1	Mannose receptor, C type 1	1.6	1.9629E-6
Genes underexpressed in M1				
J03202	LAMC1	Laminin, γ 1	2.7	3.9831E-4
AW843848	PLA2G2A	Phospholipase A2, group IIA	2.1	3.6911E-4
AF196481	MID2	Midline 2	2.0	4.8791E-4
AW971101	ARHGAP8	Rho GTPase activating protein 8	2.0	5.8626E-6
D87953	NDRG1	N-myc downstream regulated gene 1	1.9	5.4704E-6
NM_002252	KCNS3	Potassium voltage-gated channel	1.9	1.2604E-5
U53476	WNT7A	Proto-oncogene Wnt7a	1.9	3.7536E-5
U42391	MYO9B	Myosin IXB	1.8	1.0559E-5
AB011103	KIF5C	Kinesin family member 5C	1.8	3.4434E-5
NM_002387	MCC	Mutated in colorectal cancers	1.8	2.4439E-4
AW160589	APOE	Apolipoprotein E	1.6	2.7438E-7
NM_006489	NOVA1	Neuro-oncological antigen 1	1.6	1.9385E-5
AL035702	RASAL1	RAS protein activator like 1	1.6	1.7686E-4
L33404	KLK7	Kallikrein 7 (stratum corneum)	1.6	9.3908E-6
AB009849	KLK8	Kallikrein 8	1.5	1.8622E-7
AI572906	STAG3	Stromal antigen 3	1.5	1.0456E-5
M69225	BPAG1	Bullous pemphigoid antigen1	1.5	1.4143E-9
AK023655	TMC5	Transmembrane channel-like 5	1.4	1.3093E-6
U43842	BMP4	Bone morphogenetic protein 4	1.4	2.3826E-4
M96860	DPP6	Dipeptidylpeptidase 6	1.3	4.2772E-6
NM_000266	NDP	Norrie disease (pseudoglioma)	1.2	8.0898E-9
Genes overexpressed in M2				
AW269972	TSPAN-2	Tetraspan 2	6.2	0.0020256
AK000996	DNAPTP4	DNA pol transactivated protein 4	5.0	3.1239E-4
Z24725	MIG-2	Pleckstrin homology domain containing C1	4.3	5.5451E-4
J04469	CKMT1	Creatine kinase, mitochondrial 1 (ubiquitous)	4.0	5.7904E-4
AF080071	CTNNAL1	Catenin (cadherin-associated protein), α-like 1	4.0	0.0024472
S56805	EDN1	Endothelin 1	3.6	2.6668E-4
L34789	CDH1	E-cadherin	3.4	3.0812E-4
AA143153	CYP11A1	Cytochrome P450, 11 sf A1	3.4	1.2900E-4
X52003	TFF1	Trefoil factor 1	3.3	0.0015604
AK000178	FLJ20171	Hypothetical nucleic acid binding protein	3.3	1.0714E-4
U82671	MAGEA1	Melanoma antigen, family A, 1	3.1	6.3487E-4
NM_002522	NPTX1	Neuronal pentraxin I	3.1	4.1349E-4
AI219825	CDH1	E-cadherin (epithelial)	3.0	1.4154E-4
AJ245599	FJX1	Four jointed box 1 (*Drosophila*)	3.0	8.7494E-5
NM_005368	MB	Myoglobin	3.0	9.8454E-5
AF083124	RAB25	RAB25, member of RAS oncogene family	3.0	1.0446E-4
X04385	VWF	Von Willebrand factor	3.0	2.4684E-4
M26628	MME	Membrane metallo-endopeptidase/CD10	2.9	2.7408E-4
NM_001204	BMPR2	Bone morphogenetic protein receptor, type II	2.8	0.0012506
U28249	MAT-8	Mammary tumor protein 8	2.7	0.0022710
M22490	BMP4	Bone morphogenetic protein 4	2.6	0.0025805
AF179896	MEIS2e	Meis1, myeloid ecotropic viral int 1 homolog 2	2.6	0.0037418
AJ222700	TSC22	Transforming growth factor ß-stimulated	2.5	0.0023206
AA234460	PPP2R5A	Protein phosphatase 2, regulatory subunit B (B56)	1.7	1.1564E-4

Supplementary Table II. Continued.

A GenBank	Gene symbol	Gene name	Relative gene expression D	p-value
Genes underexpressed in M2				
AA993270	RNASE6	RNase A k6	8.0	0.0042152
AI828515	TNFSF13	Tumor necrosis factor (ligand) superfamily, 13	6.0	0.0041853
NM_003906	MCM3AP	Minichromosome maintenance deficient 3	1.8	0.0010823

B GenBank	Gene symbol	Gene name	Relative gene expression D	p-value
Genes overexpressed in M4				
Z62661		CpG island DNA genomic Mse1 fragment	3.3	7.0084E-4
U02368	PAX3	Paired box gene 3	3.1	0.0043684
AL137735	ABS2	Ankyrin repeat and SOCS box-containing 2	3.0	0.0027760
AK000555	ASB6	Ankyrin repeat and SOCS box-containing 6	3.0	0.0011355
AA777749	LMO7	LIM domain only 7	2.8	0.0022797
F17759	GPD1	Glycerol-3-phosphate dehydrogenase 1	2.6	8.5218E-6
X59842	PBX2	Pre-B-cell leukemia transcription factor 2; HOX12	2.5	0.0044071
V00597	PTH	Parathyroid hormone	2.4	4.2872E-4
AF141882	APMCF1	Signal recognition particle receptor	2.4	4.0106E-4
AA779336	C4BPB	Complement component 4 binding protein, ß	2.3	5.3546E-5
AL163248	ZNF294	Zinc finger protein 294	2.3	1.7194E-4
X90568	TTN	Titin	2.3	2.0182E-4
NM_003874	CD84	CD84 antigen	2.3	2.8752E-4
M34046	PAEP	Progestagen-associated endometrial protein	2.1	7.5400E-5
AB002361	KIAA0363	Similar to α NAC protein, NACA	2.1	2.3867E-4
AF250226	ADCY6	Adenylate cyclase 6	2.0	1.4458E-4
AP000500		Chromosome 3p21.3, anti-oncogene region	1.8	2.4579E-4
AB039723	FZD3	Frizzled homolog 3	1.4	9.2867E-4
AF155104	MED6	Mediator of RNA polymerase II transcription, S6 homolog	1.4	3.2020E-4
Genes underexpressed in M4				
U80456	SIM2	Single-minded homolog 2	3.5	0.0022416
U10990	NR2C2	Nuclear receptor subfamily 2,C	3.3	5.6064E-6
AF081535	CDC45L	Cell division cycle 45-like	3.2	0.0011646
X77743	CDK7	Cyclin-dependent kinase 7	2.7	0.0039741
Genes overexpressed in M5				
AK027180	MEF2D	MADS box transcription enhancer factor 2	5.5	5.2806E-4
NM_004536	NAIP	Baculoviral IAP repeat-containing 1	4.5	0.0037580
BE271713	PTPN18	Protein tyrosine phosphatase	4.3	0.0016970
NM_002698	POU2F2	POU domain, class 2, transcription factor	3.3	0.0015483
AI741331	BAK1	BCL2-antagonist/killer 1	3.2	1.3042E-4
U68723	CHES1	Checkpoint suppressor 1	2.9	1.9165E-4
U49278	UEV-1	Ubiquitin-conjugating enzyme E2 variant 1	2.5	9.4665E-4
Genes underexpressed in M5				
AK001241	FLJ10379	Hypothetical protein FLJ10379	5.9	0.0015801
S68287	AKR1C4	Aldo-keto reductase family 1, member C4	4.3	9.2731E-4
AB046663	MFTC	Mitochondrial folate transporter/carrier	3.5	4.8732E-4
U59912	MAD	MAX dimerization protein 1	3.5	0.0038577
L35263	MAPK14	Mitogen-activated protein kinase 14	3.3	0.0013878
AA761901	MYBL1	v-myb myeloblastosis viral oncogene like 1	3.2	0.0013150
AF054989	SETMAR	SET domain and mariner transposase fusion gene	3.0	0.0013803
Y09321	TAF4B	TATA-binding protein associated factors	2.9	4.5509E-4
AAF81070	SEB-4	Similar to ssDNA binding protein SEB4	2.6	5.3294E-4
AB011414	ZNF443	Kruppel-type zinc finger (C2H2)	2.5	7.3397E-4
AB017365	FZD7	Frizzled homolog 7, Frizzled receptor	2.5	0.0037855
AA455410	CDKN4	Cyclin-dependent kinase inhibitor 1B (Kip1, p27)	2.2	0.0026842

D is relative gene expression.

Supplementary Table III.

A, Differential expressed genes in AML cluster 1 (C1)

GenBank	Gene symbol	Gene name	Relative gene expression D	p-value
Genes overexpressed in C1				
NM_006847	LILRB4	Leukocyte immunoglobulin-like receptor sfB 4	25.0	6.0010E-5
U95626	CCRL2	Chemokine (C-C motif) receptor-like 2	20.0	2.1689E-5
D10925	CCR1	Chemokine (C-C motif) receptor 1	16.9	2.7644E-6
U20350	CX3CR1	Chemokine (C-X3-C motif) receptor 1	16.4	5.7489E-6
AF245703	TLR8	Toll-like receptor 8	14.8	5.8377E-7
AA402981	KCTD12	Potassium channel tetramerisation domain containing 12	13.9	3.1412E-5
NM_004244	CD163	CD163 antigen, Scavenger receptor	12.0	3.7383E-5
AX017610	WO9947655	Similar to CD20-like precusor mRNA	11.8	4.0021E-4
AF012281	PDZD1	PDZ domain containing 1	11.1	1.6133E-5
U00672	IL10RA	Interleukin 10 receptor, α	10.6	1.4853E-4
J04142	CD1D	CD1D antigen, d polypeptide	10.5	9.9480E-6
X04011	CYBB	Cytochrome b-245, ß polypeptide, gp91phox	10.4	1.1829E-4
M87842	LGALS2	Lectin, galactoside-binding, soluble, 2	9.9	2.5241E-6
Z82244	HO-1	Heme oxygenase (decycling) 1	8.6	3.3551E-4
J03745	ANX5	Annexin A5	8.3	3.1703E-5
NM_003264	TLR2	Toll-like receptor 2	8.0	1.4047E-4
U37518	TNFSF10	Tumor necrosis factor (ligand) superfamily 10	8.1	3.6638E-5
AA573434	KLF4	Kruppel-like factor 4 (gut) transcription factor	8.1	3.8469E-4
M18044	ITGAM	MAC-1α, CD11b (p170)	7.9	4.3479E-4
AW015376	NCF1	Neutrophil cytosol factor 1	7.6	1.9885E-4
X83289	MDR1/ABCB1	Multidrug resistance 1	7.5	1.4778E-4
AF282618	RISC	Retinoid-inducible serine carboxypeptidase	7.0	8.2971E-5
X78947	CTGF	Connective tissue growth factor	6.9	1.6649E-4
AL034562	PTPNS1	Protein tyrosine phosphatase, non-receptor S1	6.0	6.2263E-6
AF034970	DOK2	Docking protein 2	6.0	4.9222E-5
BE293414	ANXA2	Annexin A2	5.7	3.4704E-7
AA844153	AHR	Aryl hydrocarbon receptor	5.1	3.5884E-5
X03663	CSF1R	Colony stimulating factor 1 receptor (v-fms)	4.5	2.3262E-4
M32315	TNFRSF1B	TNFR superfamily, member 1B	4.3	6.8904E-5
NM_004536	NAIP	Baculoviral IAP repeat-containing 1	4.3	8.6550E-4
M15395	ITGB2	MAC-1ß, CD18	3.7	2.0915E-4
AB007854	GAS7	Growth arrest-specific 7	3.6	5.8188E-4
U66306	RXRA	Retinoid X receptor, α	3.2	3.0909E-4
M81934	CDC25B	Cell division cycle 25B	3.0	2.0141E-4
Y00081	IL-6	Interleukin 6	2.5	1.6578E-4
Genes underexpressed in C1				
NM_000222	KIT	Stem cell factor receptor (SCFR)	14.3	5.2996E-5
AI819896	ITM2A	Integral membrane protein 2A	11.9	3.6744E-4
Z35227	ARHH	Ras homolog gene family, member H	6.7	8.8347E-5
X70683	SOX4	SRY (sex determining region Y)-box 4	5.9	1.9885E-4
NM_006022	TSC22	TGF ß-stimulated protein	5.5	5.5508E-5

B, Differential expressed genes in AML cluster 2 (C2)

GenBank	Gene symbol	Gene name	Relative gene expression D	p-value
Genes overexpressed in C2				
M73720	CPA3	Carboxypeptidase A3 (mast cell)	81.8	2.0007E-5
M27717	CPA3	Carboxypeptidase A3 (mast cell)	50.0	1.4622E-5
M60445	HDC	Histidine decarboxylase	35.0	1.4716E-4
J03210	MMP2	Matrix metalloproteinase 2	20.1	2.9270E-9
AI133467	DSC2	Desmocollin 2	19.0	3.9789E-5
AW873072	DSC2	Desmocollin 2	14.8	1.7857E-4
M68891	GATA-2	GATA binding protein 2 transcription factor	14.5	5.0162E-5
AL031846	CBX7	Chromobox protein homolog 7	13.0	3.6592E-5
U66838	CCNA1	Cyclin A1	11.3	3.1566E-4
U79260	FTO	Fatso	7.8	8.1133E-4
D25217	MLC1	Megalencephalic leukoencephalopathy	7.5	6.3287E-4
U51869	CPBP	Similar to core promoter element binding protein	5.5	5.2311E-4

Supplementary Table III. Continued.

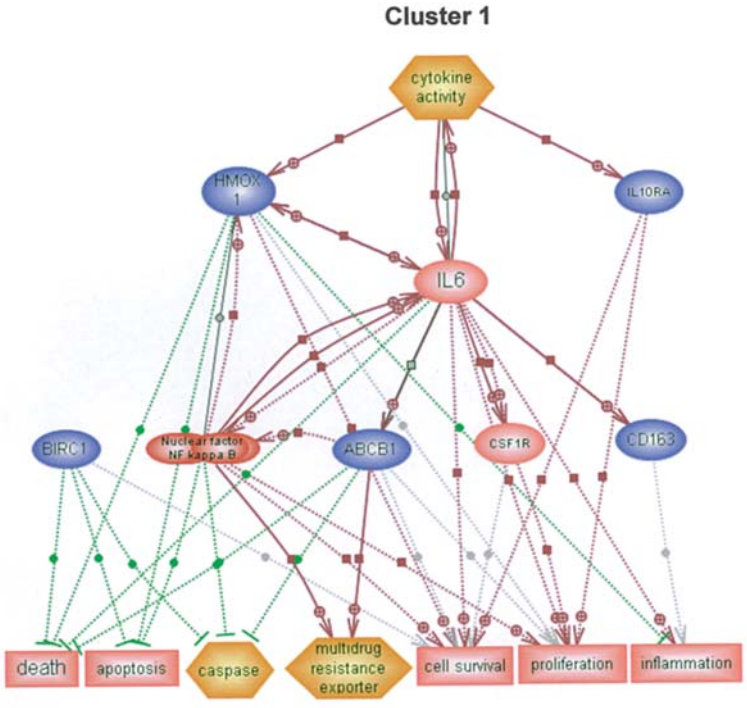
B, Differential expressed genes in AML cluster 2 (C2)

GenBank	Gene symbol	Gene name	Relative gene expression D	p-value
Genes overexpressed in C2				
AW380330	PBXIP1	Hematopoietic PBX-interacting protein	4.4	9.0979E-4
AI825989	ZFP36L2	EGF-response factor 2	3.9	8.2812E-4
M61906	PIK3R1	Phosphoinositide-3-kinase, regulatory subunit 1	3.6	1.8857E-4
D43968	RUNX1	Acute myeloid leukemia 1 transcription factor	2.5	2.4381E-4
U84214	DAD1	Defender against cell death 1	1.8	5.2159E-5
Genes underexpressed in C2				
V00522	HLA-DRB3	Major histocompatibility complex, class II, DR ß3	14.6	1.6546E-4
M14662	HLA-DRB1	Major histocompatibility complex, class II, DR ß1	12.0	2.3413E-4
X00457	HLA-DRA1	Major histocompatibility complex, class II, DP α1	11.6	3.5939E-4
NM_000433	NCF2	Neutrophil cytosolic factor 2 (p67PHOX)	11.3	0.0010103
X13334	CD14	CD14 antigen	10.9	8.4406E-4
J03745	ANXA5	Annexin V	10.4	3.1472E-4
M32011	NCF2	Neutrophil cytosolic factor 2 (p67PHOX)	10.4	5.6149E-4
J03745	ANX5	Annexin A5	10.3	3.1472E-4
M24364	HLA-DQB1	Major histocompatibility complex, class II, DQ ß1	10.1	1.6546E-4
AI765830	BLVRA	Biliverdin reductase A	8.1	3.3305E-4
D14665	ADAM9	A disintegrin and metalloproteinase domain 9	7.8	0.0010558
M23254	CANP2	Calpain 2	7.3	8.0661E-4
NM_003474	ADAM12	A disintegrin and metalloproteinase domain 12 (meltrin α)	6.2	4.6973E-5
M26038	HLA-DRB3	Major histocompatibility complex, class II, DR ß5	6.2	1.4960E-4
BE244440	HLA-DRA	Major histocompatibility complex, class II, DR α	5.2	1.6802E-4
NM_004166	CCL15	Chemokine (C-C motif) ligand 15	4.6	7.0302E-5
AA203476	PTTG1	Pituitary tumor-transforming 1	4.4	4.6634E-4
AK001323	ECT2	Epithelial cell transforming sequence 2 oncogene	3.4	0.0010097
D14134	RAD51	DNA repair protein RAD51 homolog 1	3.3	0.0010610
Z93016	WAP1	p53-responsive gene 5	3.0	8.6981E-4

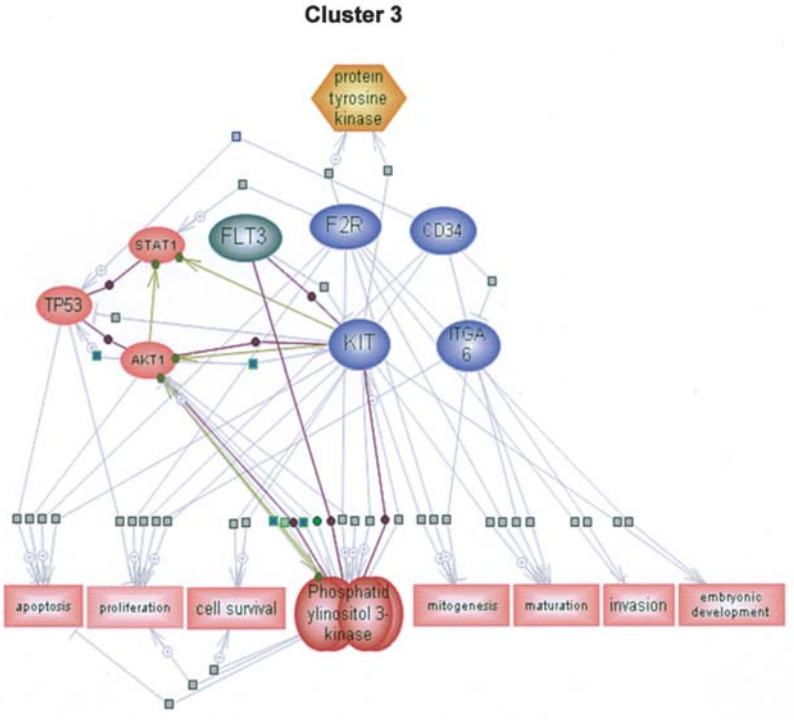
C, Differential expressed genes in AML cluster 3 (C3)

GenBank	Gene symbol	Gene name	Relative gene expression D	p-value
Genes overexpressed in C3				
AL035091	CD34	CD34 antigen	15.0	3.3273E-4
M86609	AKR1C1	Aldo-keto reductase C	6.3	6.4644E-6
AB037762	MEF-2	Myelin gene expression factor 2	4.5	8.4979E-5
S68287	AKR1C1	Aldo-keto reductase C	4.3	5.9716E-5
M62424	F2R	Coagulation factor II (thrombin) receptor	4.2	2.3010E-4
X66533	GUCY1B3	Guanylate cyclase 1, soluble, ß 3	4.0	3.0817E-4
AI678022	PTPRM	Protein tyrosine phosphatase, receptor type, M	4.0	3.4608E-4
X06182	KIT	Stem cell factor receptor (SCFR)	3.9	0.0019252
AF186111	EGFL7	EGF-like-domain, multiple 7/NEU1	3.6	0.0031686
AA100426	MOX2	OX-2 membrane glycoprotein precursor	3.3	0.0021809
U96922	NPP4B	Inositol polyphosphate-4-phosphatase, type II	3.1	1.8965E-4
M34667	PLCG1	Phospholipase C, γ 1	2.7	0.0014567
X53586	ITGA6	Integrin, α 6/CD49f	2.3	7.4027E-5
Genes underexpressed in C3				
M62880	ITGB7	Integrin, ß 7	5.2	0.0010669
AI828515	TNFSF13	Tumor necrosis factor (ligand) superfamily, 13	5.1	0.0024154
NM_004119	FLT3	Fms-related tyrosine kinase 3	4.8	0.0024840
AL034562	PTPNS1	Protein tyrosine phosphatase, non-receptor S1	4.6	3.4608E-4
U02687	FLT3	Fms-related tyrosine kinase 3	4.4	0.0024840
D00017	ANXA2	Annexin A2	3.5	0.0024005
AB007854	GAS7	Growth arrest-specific 7	3.2	0.0018785
U13697	CASP1	Caspase 1	3.0	0.0025037
AX011749	WO9955858	Similar to AHNAK	3.0	1.3911E-6
U49278	UBE2V1	Ubiquitin-conjugating enzyme E2 variant 1	2.8	2.5605E-4

D is relative gene expression.



Supplementary Figure 3. Analysis of molecular interactions and cellular processes using Pathway Assist. Highly expressed genes in cluster 1 (C1), supposed to be involved in a regulatory loop that lead to heme oxygenase 1 induction, were used as input genes (blue) for analysis. Three genes, CD163, BIRC1 and IL10RA, were validated using QPCR LDA. The common regulator for these genes selected by Pathway Assist was IL-6 (pink) or the functional class cytokines (yellow). Both IL-6 and other cytokines and CSF1R (v-fms) were represented among the most differentially overexpressed genes in C1. Red lines, positive effect; green lines, negative effect; gray dotted lines, effect of unknown function; grey solid lines, a direct binding between HMOX1 and NF-κB complex.



Supplementary Figure 4. Networks of molecular interactions and cellular processes activated in cluster 3 using CD34, c-KIT, F2R and ITGA6 as input genes. CD34, c-KIT, F2R and ITGA6 (blue) are membrane molecules characteristic of myeloid progenitor cells and were highly expressed in this cluster, while FLT3 (green) was underexpressed. The expression profile of the genes was mainly controlled with QPCR LDA. CD34 has previously been validated using flow cytometry (31). Downstream targets for the selected genes were p53, AKT1, STAT1 (pink) and PI3-kinase (PI3K) complex. Lines with arrows, positive effect; violet lines, protein-protein binding; green lines, protein modification; blue square on the line between CD34 and TP53, positive effect of CD34 on molecular synthesis of p53; green square, positive effect on gene expression; line with green circle, possibility for PI3K to bind to the AKT1 promoter.