Genomics & Informatics Vol. 4(3) 97-102, September 2006

Identification of Gene Expression Signatures in Korean Acute Leukemia Patients

Kyung-Hun Lee, Se Won Park, Inho Kim, Sung Soo Yoon, Seonyang Park* and Byoung Kook Kim

Department of Internal Medicine, Seoul National University College of Medicine, Seoul 110-744, Korea

Abstract

Background: In acute leukemia patients, several successful methods of expression profiling have been used for various purposes, i.e., to identify new disease class, to select a therapeutic target, or to predict chemo-sensitivity and clinical outcome. In the present study, we tested the peripheral blood of 47 acute leukemia patients in an attempt to identify differentially expressed genes in AML and ALL using a Korean-made 10K oligo-nucleotide microarray.

Methods: Total RNA was prepared from peripheral blood and amplified for microarray experimentation. SAM (significant analysis of microarray) and PAM (prediction analysis of microarray) were used to select significant genes. The selected genes were tested for in a test group, independently of the training group.

Results: We identified 345 differentially expressed genes that differentiated AML and ALL patients (FWER<0.05). Genes were selected using the training group (n=35) and tested for in the test group (n=12). Both training group and test group discriminated AML and ALL patients accurately. Genes that showed relatively high expression in AML patients were deoxynucleotidyl transferase, pre-B lymphocyte gene 3, B-cell linker, CD9 antigen, lymphoid enhancer-binding factor 1, CD79B antigen, and early B-cell factor. Genes highly expressed in ALL patients were annexin A1, amyloid beta (A4) precursor protein, amyloid beta (A4) precursor-like protein 2, cathepsin C, lysozyme (renal amyloidosis), myeloperoxidase, and hematopoietic prostaglandin D2 synthase.

Conclusion: This study provided genome wide molecular signatures of Korean acute leukemia patients, which clearly identify AML and ALL. Given with other reported signatures, these molecular signatures provide a means of achieving a molecular diagnosis in Korean acute leukemia patents.

*Corresponding author: E-mail Seonpark@plaza.snu.ac.kr, Tel +82-2-2072-3347, Fax +82-2-744-6075
Accepted 9 June 2006

Keywords: acute myeloid leukemia, acute lymphoblastic leukemia, microarray, expression profiling

Introduction

The exact diagnosis and classification of acute leukemias are of clinical importance due to the different treatment approaches required. Acute myelogenous leukemia (AML) is a malignant disease of the bone marrow in which hematopoietic precursors are transformed and arrested in an earlier stage of development. AML is diagnosed on the presence of greater than 20% blasts in the marrow (Jaffe *et al.*, 2001). Acute lymphoblastic leukemia (ALL) is also a malignant clonal disease of the bone marrow in which early lymphoid precursors proliferate and replace the normal hematopoietic cells of the marrow. Although AML and ALL are distinct clinical and morphological disease entities, no single test is currently capable of differentiating them. The classification proposed by the French-American-British (FAB) is based on morphology and cytochemistry, and involves eight major subtypes, i.e., M0 to M7 (Bennet *et al.*, 1976; Bennet *et al.*, 1985).

The new World Health Organization (WHO) classification incorporates molecular, cytogenetic, and clinical features with morphology (Jaffe *et al.*, 2001), and provides new prognostic markers and suggests disease-specific therapeutic approaches. For example, all-trans-retinoic acid and arsenic trioxide have revolutionized the treatment of acute promyelocytic leukemia characterized by t (15:17) (q22:q21) (Kelly *et al.*, 2001). However, the standard diagnostic methods required for leukemia are complex, time-consuming, and require experienced specialists. Moreover, the diagnoses of some patients remain obscure based on current classifications. In addition, therapeutic strategies determined on the basis of classification are unsatisfactory. For example, the most appropriate treatment for standard-risk AML, which accounts for over half of all cases of this disease, has not been firmly established. Moreover, this subgroup is difficult to classify, because it includes cases with various numerical and structural cytogenetic abnormalities that occur infrequently, which makes it difficult to determine their prognostic significances (Grimwade *et al.*, 2001). Moreover, cytogenetic analysis provides no clues in cases of AML with a normal karyotype, which account for the majority of cases in the standard-risk group.

Gene expression profiles have recently been featured in research programs concerning hematologic malignancies. The diagnosis and classification of leukemia have been mainly issued for several years. Golub *et al.* successfully classified AML and ALL using molecular signatures (Golub *et al.*, 1999), and the studies by Bullinger *et al.* (Bullinger *et al.*, 2004) and Valk *et al.* (Valk *et al.*, 2004) suggested that molecular classifications based on gene expression profiles could provide more accurate diagnoses and classifications and predict clinical outcomes better. Bullinger *et al.* identified a new molecular subtype of AML that includes two prognostically relevant subgroups of AML with a normal karyotype. Valk *et al.* reported several novel clusters, some consisting of specimens with a normal karyotype and a unique cluster in a patient with a poor clinical outcome. Several clusters identified by Schoch *et al.* (Schoch *et al.*, 2002) and Debernardi *et al.* (Debernardi *et al.*, 2003) were found to correspond to cytogenetic and molecular definitions, such as, t (15;17), t (8;21), and inv (16). These, collectively, reflect that genome-wide expression profiles and can be integrated into clinical practice to broaden our understanding of these diseases.

Here, we report on an application of gene-expression profiling to the differentiation of AML and ALL. As far as we are aware, this is the first report of gene expression profiling by oligo-nucleotide Microarray analysis in hematological malignancies in Korea.

Materials and Methods

Patients and Clinical data

Patients diagnosed with AML or ALL with available peripheral blood samples available for analyses were enrolled in this study. A total of 47 patients were analyzedat the Seoul National University Hospital, Inha University Hospital, Gachon University Gil Medical Center, Hallym University Medical Center, and at the Korean Cancer Center Hospital (Table 1).

Table 1. Patients characteristics (n=47)

Characteristic	Total
Sex - no. (%)	
Male	30 (63.8)
Female	17 (36.2)
Age group - no.(%)	
∠35 yr	9 (19.2)
35~60 yr	27 (57.4)
≥60 yr	11(23.4)
Disease - no.(%)	
AML	37 (78.7)
ALL	10 (21.3)

Preparation of total RNA

Mononuclear cells were isolated from peripheral blood bycentrifugation on a Ficoll Hypaque (density, 1.077; APB, Upsalla, Sweden) density gradient. Total RNA from these cells was isolated using TRIZOL (Gibco BRL, NY) according to the manufacturer's instructions. Reference RNA was obtained from the Jerket cell line.

Oligo-neucleotidenucleotide Microarray analysis

Total RNA (5 μg) was converted into double stranded cDNA using the cDNA synthesis system (Roche) using T7-(dT)24 primer. The synthesized cDNA was purified using RNeasy kits (Qiagen, Valencia, *http://www.qiagen.com*), and prepared cDNAs were labeled with Cy5-UTP (for reference RNA) or Cy3-UTP(for test samples) using Megascript T7 kits (Ambionm Austin). cRNAs were cleaned using RNeasy (Qiagen). Labeled 15 μg of each cRNA was mixed and fragmented by heating to 94°C for 15 min, and hybridized with a Human 10 K microarray (Macrogen, Seoul, Korea) for 16h at 42°C. Arrays were then washed and scanned with an Array scanner (Molecular Dynamics). Acquired images were processed and analyzed statistically with respected to spot intensities using an Imagene v4.1 software (Biodiscovery, CA, USA).

Statistical analysis

Normalization and filtering

To normalize microarray data, we used the within pin group, intensity dependent Loess method. An 'MA-plot' was used to represent the R and G data, where $M=\log_2 R/G$ and $A=\log_2 (RG)1/2$; R represents the F635 signal from Cy-5 and G the F532 signal from Cy-3 labeling. To correct for pin variation, within pin group normalization'was performed. Raw data was normalized relative to a (pin tip+A), i.e. $\log_2 R/G$ $\log_2 R/G\text{-}ci(A) = \log_2 R/[ki\ (A)\ G]$ where ci (A) is the Lowess fit to the MA-plot for the i th pin group only, i=1,2,.../ and I denotes the number of pin groups. The normalized data was filtered using the criteria of missing value >20% in studied samples. Remained data with missing values were adjusted using the K nearest neighborhood joining method.

Selection of differentially expressed genes

We used 35 samples (28 AML and 7 ALL) as a training group and 12 samples (9 AML and 3 ALL) as a test group. To identify differentially expressed genes, which discriminate AML from ALL, namely classifier genes, we used SAM (Significant Analysis of Micorarray). Classifier genes were determined using FDR (<0.05). Selected classifier genes were validated by class prediction using

Gaussian linear discriminant analysis in the independent test group. CLUSTER and TREEVIEW were used for data clustering and visualization (Eisen *et al.*, 1998).

Results

Identification of classifier genes that differentiate AML and ALL

We selected 35 samples as a training group from the 47 samples to reduce the effect of individual variability. After normalization and filtering as described above, we obtained 10,000 gene expression data, which were available for further analysis. Based on the unsupervised clustering of whole samples, AML and ALL were not differentiable due to high noise levels in the expression data (Fig. 1).

To remove noise signals and identify genes differentially expressed in AML and ALL, supervised analysis was performed using Significant Analysis of Microarray (SAM). Genes were selected using a False Discovery Rate (FDR) of <0.05 as threshold, and 345 genes were found to be at significantly different levels in AML and ALL (Fig. 2). We selected 265 over-expressed genes in ALL and 80over-expressed genes in AML compared to the other classes (Fig. 2).

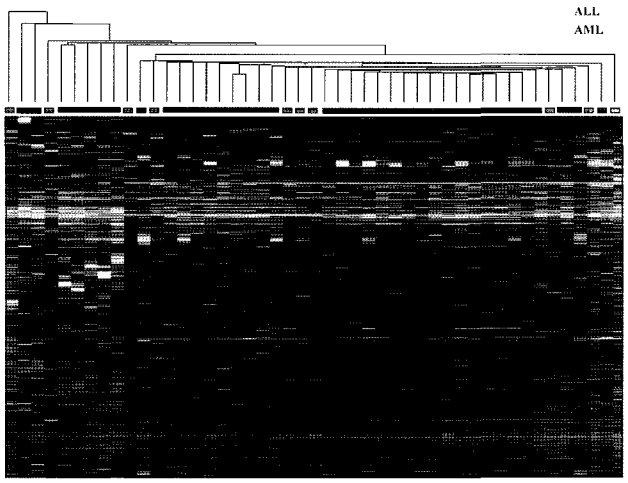

Fig. 1. Unsupervised hierarchical Cluster Analysis of all 47 samples with 10000 genes

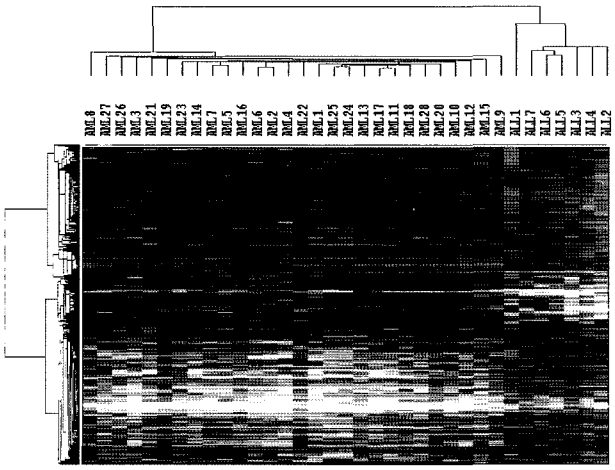

Fig. 2. Hierarchical Cluster Analysis of AML and ALL samples with 345 genes in the training group (n=35)

Validation of classifier genes probability

To verify the selected classifier genes, we performed class predictions with classifier genes in the test group, which consisted of 9 AML and 3 ALL patients. A cross-validation and a test error rate calculation were performed with PAM using 345 genes. Our data showed a remarkable ability to discriminate between AML and ALL in the test group. Finally, we successfully discriminated all AML and ALL patients using these 345 classifier genes (Fig. 3).

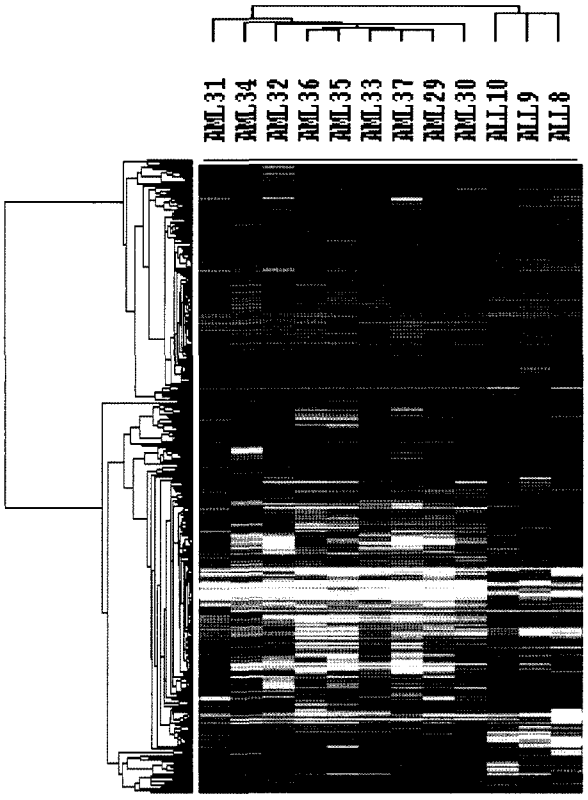

Fig. 3. Hierarchical Cluster Analysis of AML and ALL samples with 345 genes in the test group (n=12)

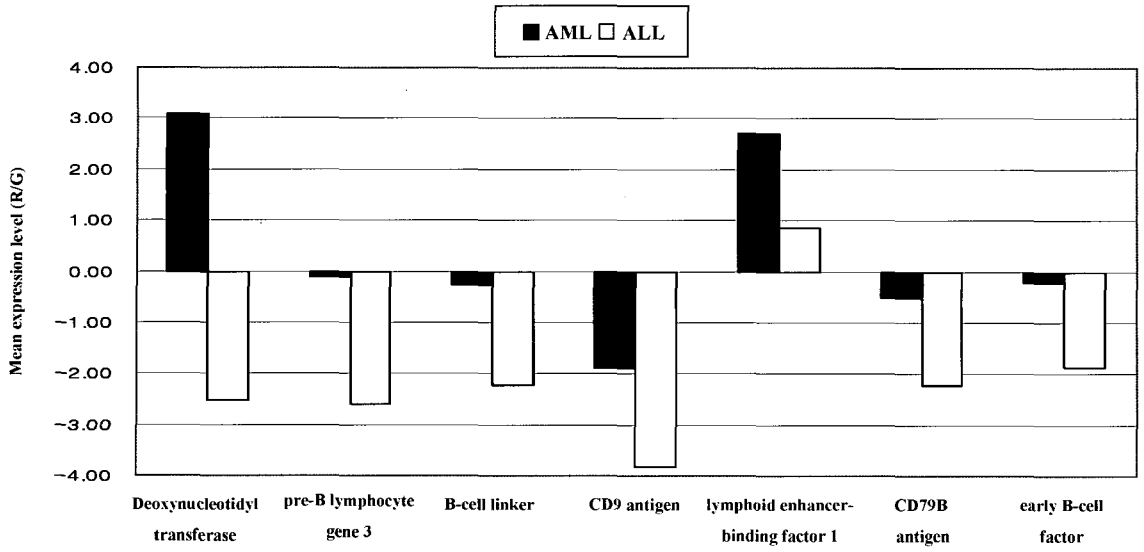

Fig. 4(a). Mean expression level of 7 genes which showed relatively high expression in AML patients

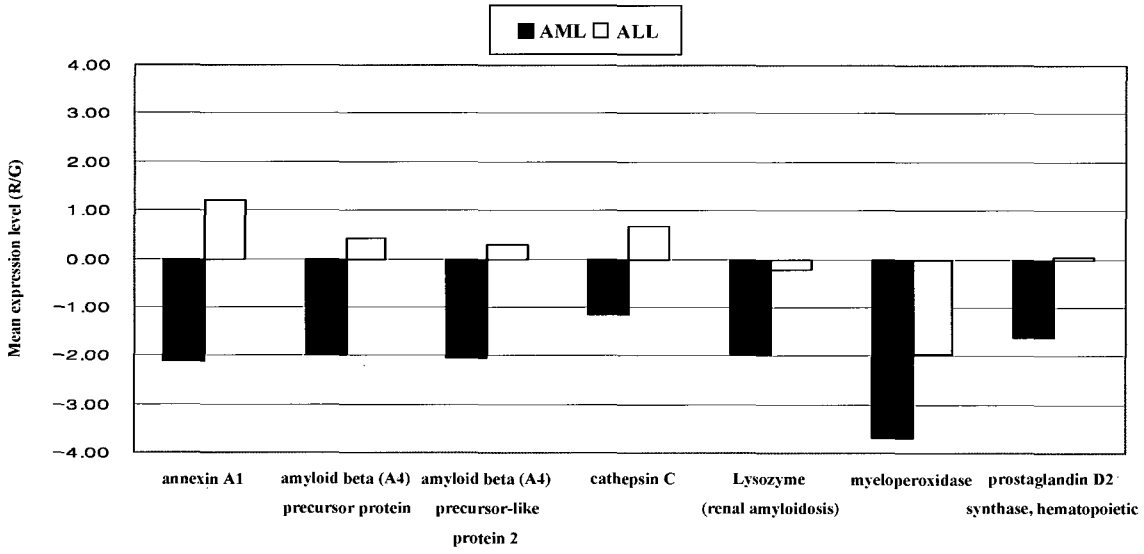

Fig. 4(b). Mean expression level of 7 genes showed relatively high expression in ALL patients

Classifier genes differentially expressed in AML and in ALL

Genes that showed relatively high expression in AML patients were deoxynucleotidyl transferase, pre-B lymphocyte gene 3, B-cell linker, CD9 antigen, lymphoid enhancer-binding factor 1, CD79B antigen, and early B-cell factor (Fig. 4a). Genes highly expressed in ALL patients were annexin A1, amyloid beta (A4) precursor protein, amyloid beta (A4) precursor-like protein 2, cathepsin C, lysozyme (renal amyloidosis), myeloperoxidase, and hematopoietic prostaglandin D2 synthase (Fig. 4b).

Discussion

Microarray technology has been applied to haematological malignancy studies on disease classification, outcome prediction, pathway delineation, and target identification. Some of this work has been very successful, and these successes encouraged some to declare the dawn of a new era of quantitative and predictive biology, in which networks of gene interactions will be unraveled and mathematical algorithms will be used to diagnose diseases, model disease progression, and predict outcome and response to treatment.

By the middle of 2003 over a hundred publications concerned direct or indirect applications of DNA microarrays to haematological malignancies (Ebert *et al.*, 2004; Margalit *et al.*, 2005). These studies fell into several overlapping areas. The first type involved basic profiling studies, which concerned the origins and aetiologies of tumors. These studies often involved comparing tumor gene signatures with signatures of cells at a particular haematological stage of development. Alternatively, the nature of aberrant gene expression compared with the nearest normal cell type may be examined to reveal details of the mechanism of tumorigenesis or tumor maintenance. The second type of study could be described as classification studies, in which the aim was to identify the minimum set of genes required to define a particular tumor or tumor subclass. Typically this study type follows on profiling studies and employs supervised clustering coupled with a machine learning algorithm or another type of statistical classifier. The best of these studies have attempted to independently cross-validate the classifier set using further arrays or conventional approaches in a new tumor sets. The third type of study involves the identification of diagnostic and prognostic markers. In this case, supervised learning approaches are employed that are similar to associate a particular profile with a knownoutcome, the aim being to identify a small subclass of genes with predictive capability. This might, for example, be coupled to a Kaplan-Meier survival curve to predict outcome for different tumor subsets, or to predict relapse or response to treatment. It is important to note that such classifiers can only be applied to defined tumor subsets.

Genes highly expressed in ALL in the present study have been associated with the pathogenesis of lymphoid leukemias. For example, annexin A1 was highly expressed in ALL patients. It was also reported to be a simple diagnostic marker for hairy cell leukemia with an accuracy of 100% according to a previous report (Falini *et al.*, 2004). Amyloid precursor protein has been hypothesized to be related with unknown molecular mechanisms involved in leukemogenesis with ERG and ETS2 (Baldus *et al.*, 2004). Moreover, cathepsin C was proposed to be a significant gene for large granular lymphocyte (LGL) leukemia identification by expression profiling (Kothapalli *et al.*, 2003).

Genes highly expressed in AML include B-cell related genes, such as pre-B lymphocyte gene 3, B-cell linker, lymphoid enhancer-binding factor 1, and early B-cell factor. The mechanism linking these genes to AML is not clear, which needs further investigations.

The classifiers of this study do not show high concordance compared to previous studies. Golub *et al.* used 50 predictor genes to discriminate AML from ALL (Golub *et al.*, 1999). Twenty-five genes were chosen that were more highly expressed in AML and 25 genes were more highly expressed in ALL. None of the genes, however, were chosen in the present study in Korean population. Although in the past years gene expression profiling has indeed shown great promise with regard to the classification of leukemia, one should also take its limitations into account. The limitations of gene expression profiling are evident if the results of different gene expression profiling studies are compared. The quality of materials used, sample processing, the application of different expression platforms and the wide variety of methods used for pattern discovery, make the proper validation of the results of different studies difficult.

In Korea, several reports have been issued on microarray technology. However, to date no report has been issued on the use of this technology in acute leukemia. In the present study, we identified gene expression profiles that differentiae AML and ALL. We hope that this approach may provide broad access to high-quality diagnoses and treatment guidance in leukemia.

Acknowledgements

This study was supported by grants of the International Mobile Telecommunications 2000 R&D Project, Ministry of Information & Communication, Republic of Korea (01-PJ11-PG9-01BT00A-0048), and the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (0405-BC02-0604-0004).

References

Baldus, C.D., Liyanarachchi, S., Mrozek, K., Auer, H., Tanner, S.M., Guimond, M., Ruppert, A.S., Mohamed, N., Davuluri, R.V., Caligiuri, M.A., Bloomfield, C.D., and de la Chapelle, A. (2004). Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses overexpression of APP, ETS2, and ERG genes. *Proc. Natl. Acad. Sci. USA* 101, 3915-3920.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R., and Sultan, C. (1985). Proposed revised criteria for the classification of acute myeloid leukemia: a report of the French-American- British Cooperative Group. *Ann. Intern. Med.*103, 620-625.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R., and Sultan, C. (1976).Proposals for the classification of the acute leukemias: French-American-British (FAB) Co-operative Group. *Br J.Haematol.* 33,

451-458.

Bullinger, L., Döhner, K., Bair, E., Frohling, S., Schlenk, R.F., Tibshirani, R., Dohner, H., and Pollack, J.R. (2004). Use of gene-expression profi.ing to identify prognostic subclasses in adult acute myeloid leukemia. *N. Engl. J. Med.* 350, 1605-1616

Debernardi, S., Lillington, D.M., Chaplin, T., Tomlinson, S., Amess, J., Rohatiner, A., Lister, T.A., and Young, B.D. (2003). 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.

Ebert, B.L. and Golub, T.R. (2004). Genomic approaches to hematologic malignancies. *Blood* 104, 923-932.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95, 14863-14868.

Falini, B., Tiacci, E., Liso, A., Basso, K., Sabattini, E., Pacini, R., Foa, R., Pulsoni, A., Dalla Favera, R., and Pileri, S. (2004). Simple diagnostic assay for hairy cell leukaemia by immunocytochemical detection of annexin A1 (ANXA1). *Lancet.* 363, 1869-1870.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-537.

Grimwade, D. (2001).The clinical significance of cytogenetic abnormalities in acute myeloid leukaemia. *Best Parct. Res. Clin. Haematol.* 14, 497-529.

Jaffe, E.S., Harris, N.L., Stein, H., and Vardiman, J.W. (2001). *World health organiztion classification of tumours: Pathology and genetics of tumours of haematopoietic and lymphoid tissues* (Lyon, France: IARC Press).

Kelly, L.M. and Gilliland, D.G. (2002). Genetics of myeloid leukemias. *Annu. Rev. Genomics Hum. Genet.* 3, 179-198.

Kothapalli, R., Bailey, R.D., Kusmartseva, I., Mane, S., Epling-Burnette, P.K., and Loughran, T.P. Jr. (2003). Constitutive expression of cytotoxic proteases and down-regulation of protease inhibitors in LGL leukemia. *Int. J. Oncol.* 22, 33-39.

Margalit, O., Somech, R., Amariglio, N., and Rechavi, G. (2005). Microarray-based gene expression profiling of hematologic malignancies: basic concepts and clinical applications. *Blood Rev.* 19, 223-234.

Schoch, C., Kohlmann, A., Schnittger, S., Brors, B., Dugas, M., Mergenthaler, S., Kern, W., Hiddemann, W., Eils, R., and Haferlach, T. (2002). Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc. Natl. Acad. Sci. USA* 99, 10008-10013.

Valk P.J., Verhaak, R.G., Beijen, M.A., Erpelinck, C.A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J.M., Beverloo, H.B., Moorhouse, M.J., van der Spek, P.J., Lowenberg, B., and Delwel, R. (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. *N. Engl. J. Med.* 350, 1617-1628.