

Identification of gene expression signatures for molecular classification in human leukemia cells

JU HAN SONG^{1,2}, HYEOUNG-JOON KIM², CHANG HYUN LEE⁴, SEUNG JUN KIM⁴,
SEUNG YONG HWANG⁴ and TAE SUNG KIM^{2,3}

¹College of Pharmacy and ²Genome Research Center for Hematopoietic Diseases, Chonnam National University, Gwangju; ³School of Life Sciences and Biotechnology, Korea University, Seoul; ⁴Department of Biochemistry and Molecular Biology, Hanyang University, Ansan, Republic of Korea

Received January 30, 2006; Accepted March 29, 2006

Abstract. Although the methods by which leukemia is classified have been improved for effective therapies, leukemia patients occasionally exhibit diverse, sometimes unpredictable, responses to treatment. Consequently, these patients also evidence individually different clinical courses when administered with anti-leukemia drugs. In order to find new, more precise molecular markers for leukemia classification, we have analyzed the gene expression profiles from 65 diagnostic bone marrow specimens of adult patients with AML, ALL, CML or CLL by using high-throughput DNA microarrays harboring approximately 8300 unique human genes or expression sequence tags. In the present study, we identified a group of leukemia-specific genes, which manifest gene expression profiles distinctly representative of normal bone marrow samples, as determined by a significance analysis of microarray (SAM) and GeneSpring 6.1 programs. We also determined the minimal number of genes showing a difference between acute and chronic leukemia patient groups. Furthermore, the unsupervised cluster analysis revealed a gene subset which can be used to distinguish between AML, ALL, CML and CLL patient groups, based on expression signatures. The expression levels of differentially regulated genes were verified via the principle component analysis (PCA). Our results may provide a novel set of molecular criteria for the classification of leukemia patients, and may also facilitate efforts to discover new targets, allowing for more effective treatment of leukemia patients.

Introduction

Leukemia is an aggregate of hematopoietic malignancies, and is generally characterized by the accumulation of lymphocytic or myelocytic lineage precursor cells arrested at early differentiation stages. These precursor cells rapidly proliferate, immortalize, and then persist as immature cells (1). This malignancy has been classified into major four categories namely, acute myelocytic leukemia (AML), acute lymphocytic leukemia (ALL), chronic myelocytic leukemia (CML) and chronic lymphocytic leukemia (CLL), and these classifications are assigned on the basis of rate of progression and origins of the leukemic blast. The classification of these subsets has, thus far, been fundamentally predicated on morphological, cytochemical, immunophenotypic, and karyotypic features (1-3). In accordance with these criteria, leukemia patients are usually treated via a standardized therapy. However, although the methods used to classify leukemia have been under steady development for the past several decades, resulting in marked improvements in the treatment of leukemia, a few inherent limitations remain. Leukemia patients occasionally evidence diverse or unexpected responses to therapy, and often exhibit idiosyncratic and individual clinical courses during treatment (4,5).

Due to the incidence of the unexpected responses, many researchers have embarked on investigations designed to discover novel molecular targets for class prediction and therapeutic strategies, not only for leukemia, but also for a host of other tumor types. DNA microarray technology has constituted a new driving force for the surmounting of these challenges (6,7). This new trend was ushered in by a preliminary study conducted by Golub *et al* intended to discover genes that differed between ALL and AML (8). Using a similar genome-wide approach, this group subsequently reported 15 highly specific genes distinct to both AML and ALL, all of which were associated with MLL translocation (9). Also, using mass patient specimens and the results of gene expression profiling trials, several research groups have attempted to locate molecular markers useful in subtype classification and the prediction of outcomes (10-14).

Correspondence to: Dr Tae Sung Kim, School of Life Sciences and Biotechnology, Korea University, 5-ga, Anam-dong, Seongbuk-gu, Seoul 136-701, Republic of Korea
E-mail: tskim@korea.ac.kr

Key words: human, leukemia, DNA microarray, classification, marker

In the present study, we conducted DNA microarray-based assays in order to locate leukemia-specific genetic markers that are distinct from those contained in normal hematopoietic cells, and that might prove useful in the prediction or identification of leukemia types. Furthermore, we also identified a pathogenesis-related gene subset, which can be used to distinguish between acute and chronic leukemia.

Materials and methods

Patient samples. All specimens were acquired from the bone marrow samples of 65 adults with newly diagnosed leukemia, all of whom had been admitted to the Chonnam National University Hospital (Gwangju, Korea). Informed consent for the use of all leukemic samples was obtained from all patients. Prior to biopsy, no patient received any form of anti-cancer therapy. Normal bone marrow samples, used as references for the microarray analyses, were obtained from healthy donors during the bone marrow transplantations. Mononuclear cells were isolated via Ficoll-Hypaque density gradient centrifugation, and cryo-preserved using RNAlater (Ambion, Austin, TX) at -20°C until later use. The leukemic specimens utilized in this study were classified as follows: 35 samples of AML (AM number), 8 of ALL (AL number), 13 of CML (CM number) and 9 of CLL (CL number). The clinical characteristics of the patients are listed in Table I. In order to maintain anonymity and ensure the impartiality of the study, we linked the gene expression profiles of all patient specimens in the microarray analysis only with age, sex, disease type, and chromosomal abnormalities.

DNA microarray. The Platinum Biochip™ Human 8.3 K microarray (GenoCheck, Ansan, Korea) used in the present study harbored approximately 5800 known genes, 2500 expressed sequence tags (ESTs), and housekeeping genes for internal normalization. All clones were spotted robotically onto slide glasses, using a Pixsys 5500 arrayer (Cartesian Technologies, CA). The slide glasses were subsequently post-processed, as was described in a previous report (15).

Probe synthesis. In order to purify the total RNA from the bone marrow mononuclear cells, the leukemic samples, stored in RNAlater, were washed twice with phosphate-buffered saline (PBS). These cells were then immediately applied to TRI reagent (MRC, Cincinnati, OH), and processed in accordance with the manufacturer's instructions. Total RNA was then stored at -80°C until use. The preserved total RNA specimens were amplified using an RNA amplification kit (Ambion) in accordance with the manufacturer's instructions. Using 2 μg of total RNA as starting material, we obtained 10 μg of amplified RNA (aRNA) from each specimen, after one round of amplification. Using the same method as was used for the leukemia samples, we also obtained aRNA from the total RNA of each normal sample. After the normal samples had been subjected to amplification, we prepared an aRNA pool, which was mixed with at least three aRNA samples in order to minimize the unique specificity between each individual as reference. The aRNA samples were then reverse-transcribed to labeled cDNA, using each Cy3 for the normal samples, or Cy5 for the leukemic samples.

Hybridization and data normalization. Equal amounts of the two probes were then combined and hybridized in DNA microarrays, in the manner described in a previous report (15). After hybridization, the raw images generated by the Axon Instruments GenePix 4000B scanner were processed and normalized using GenePix Pro 5.1 software (Axon, CA). In order to filter out any unreliable data, the cDNA spots with signal-to-noise values below 100 were excluded from the data. After normalization for further study, we included only the results from the 65 patients with reliable signal genes in a minimum of 50% of all genes spotted on the microarray.

Data analyses. Using the gene expression profiles obtained from selected patients, we used the GeneSpring 6.1 program (Silicon Genetics, CA) to determine which leukemia type-specific genes or gene subsets allowed us to distinguish between acute and chronic leukemia samples. We also calculated the values for significant gene expression, via the significance analysis of microarray (SAM) method. The input values for the SAM plot calculator for (Δ) and fold change parameters were determined to be 0.967 and 1.3, respectively. The SAM score (d) is predicated on the ratio of change in gene expression to the standard deviation in the data for that gene. Also, the estimated median false discovery rate for these input values was determined to be 2.36%. The gene subsets from two independent analyses were ultimately confirmed via 2-dimensional hierarchical clustering, using the Tree View microarray program (<http://www.microarrays.org/software>) and/or the 3-dimensional clustering of training patients ($n=65$) via Principle Component Analysis (PCA). The detailed methods by which each of the groups was classified are described in Results and Figure legends.

Results

Gene expression profiling of diagnostic leukemia cells. Using a DNA microarray harboring 8352 human probe sets, we obtained gene expression data from the diagnostic bone marrow samples of 65 adult patients. As leukemia is a generally heterogeneous malignancy, the gene expression profiles of these subjects were expected to differ between each of the types or subgroups. Surprisingly, after unsupervised hierarchical clustering analysis with the expression profiles of each patient, using the Tree View software, we observed a patient tree with four branches, which is shown in Fig. 1. The expression profiles of the ALL and AML patients overlapped to some degree, thereby indicating that leukemia is not only a heterogeneous malignancy, but also exhibits unique gene profiles within each of its subtypes.

Nevertheless, the existence of genes with common differences in these four types as compared to normal samples may indicate only the existence of general genes, which may be associated with common characteristics of leukemia, or common oncogenic features. In order to test this notion, the gene expression profiles of each of the leukemia patients were collected, and then evaluated with a focus on the determination of mean signal intensities, and to determine any significant differences with the profiles of the samples from normal

Table I. Clinical characteristics of 65 diagnostic patients with leukemia.

Disease type	Patients no.	Sex (M/F)	Age (year)	Blast (%)	Karyotype (no.)
AML	35 (AM#)	17/18	16-75 (mean: 43.9)	3-100 (mean: 74.1)	Normal: 13 Abnormal: 19 No data: 3
ALL	8 (AL#)	6/2	14-59 (mean: 39.5)	90-98 (mean: 93.3)	Normal: 1 Abnormal: 2 No data: 5
CML	13 (CM#)	9/4	27-75 (mean: 55.1)	3-95 (mean: 13.4)	Normal: 1 Abnormal: 5 No data: 7
CLL	9 (CL#)	6/3	52-73 (mean: 63.4)	50-95 (mean: 80.8)	No data: 9

donors. As shown in Table II, based on the mean intensities and SAM scores of the sample profiles ($n=65$), we were able to identify the top 50 probe sets, 25 of which were over-expressed and 25 of which were down-regulated, as compared to the normal samples. These differentially-expressed genes were found to include phosphoinositide-3-kinase p110 α catalytic subunit, CD69 antigen, hematopoietically expressed homeobox (all over-expressed) and synuclein α , transcription factor Dp family member 3, p300/CBP-associated factor (all down-regulated). Contrary to our expectations, however, the profiles of many of the genes in all of the leukemia samples were down-regulated, as compared to what was measured in normal bone marrow samples.

Genes for pathological differences between acute and chronic leukemias. Leukemia can be divided into acute and chronic types. In the absence of anti-leukemic therapies, patients with acute or chronic leukemia tend to follow distinct clinical courses, normally resulting in more rapid lapses in acute leukemia patients than are seen in chronic leukemia patients (17). Although chronic leukemia evidences a slow rate of progression with regard to its clinical course, the treatment of chronic leukemia has proven rather more difficult than that of acute leukemia (18). Therefore, we attempted to delineate genes that were differentially expressed in cases of acute and chronic leukemia. The gene expression signatures detected exhibited remarkable agreement with the leukemia types assigned at diagnosis, thereby indicating the existence of pathological differences between the two leukemia groups. As shown in Fig. 1, we had already confirmed that the tree for patient groupings, predicated on basal profiles, exhibited two distinct branches. In order to further verify the existence of the genes closely related to the pathological differences between the two groups, we divided the entire patient group ($n=65$) into two groups, with 43 acute patients (35 of AML and 8 of ALL) and 22 chronic patients (13 of CML and 9 of CLL). The genes evidencing significant expression distance between these two groups were identified using the SAM

program. These genes, sorted by SAM, were then compared once more with the expression profiles, using GeneSpring 6.1 software. Finally, we generated 32 discriminatory gene probes (20 in the acute cases and 12 in the chronic cases) associated with differences between acute and chronic leukemias, on the basis of gene expression profiles and genetic distances (Fig. 2A). Using the PCA program, we tested these 32 genes with regard to clinical course. When the 32 genes were reversely applied to training sets consisting of 65 patients, as shown in Fig. 2B, two major compartments were formed in the multi-dimensional analysis (purple-colored spots for acute patients and green-colored spots for chronic patients). Among the 32 genes selected in this analysis, 16 (50%) were previously associated with tumors, and only 10 (31.3%; PROM1, HSPD1-overlapping, LDHB, HLA-DOB, FLT3, CD99, MYB, LATS2 and CAPN3) were reported to have been associated with leukemia.

Genes for class prediction in leukemia. Using gene expression profiling coupled with high-throughput microarray techniques, several of the groups have already been studied extensively with regard to tumor classification, in an effort to identify genes associated with diagnosis (8-14). Leukemia, in particular, is an excellent disease model for use in studies regarding class prediction. However, no previous report exists on the detection of molecular markers specific to each of the four types of leukemia. Therefore, we attempted to simultaneously analyze gene expression profiles obtained from a group of patients suffering from the four types of leukemia. Prior to comparison, the DNA microarray data of the patient specimens were divided into four groups, namely AML ($n=35$), ALL ($n=8$), CML ($n=13$), and CLL ($n=9$) patients. Similar to that determined in the analysis shown in Fig. 2, we determined significant distances using the SAM software. In brief, each of the four groups was compared with the others. After candidate genes were selected via these analyses, those genes were once more filtered, via a comparison of mean signal intensities. We ultimately settled

Table II. Common genes over- or down-expressed in leukemia cells.

Gene Symbol	Accession no.	Annotation	Mean Intensity	p-value	SAM score (d)
Over-expression					
PIK3CA	AA993785	Phosphoinositide-3-kinase p110alpha catalytic subunit	1.4	0.00049	8.462
CD69	BF477645	CD69 antigen (p60, early T-cell activation antigen)	1.5	0.00042	9.849
PMS2L1	AA459266	Postmeiotic segregation increased 2-like 1	1.3	0.00055	8.235
EXT1	T91083	Exostoses (multiple) 1	1.4	0.00049	8.964
WBSCR16	AA430677	Williams-Beuren syndrome chromosome region 16	1.3	0.00054	8.580
EST	A1003832	EST	1.3	0.00054	8.353
EST	H25510	EST	1.3	0.00046	9.618
KIAA0828	H99883	KIAA0828 protein	1.4	0.00047	9.530
FLJ34306	H85475	Homo sapiens cDNA FU38031 fis, clone CT	1.6	0.00039	10.730
PFAAP5	N59893	Phosphonoformate immuno-associated protein 5	1.5	0.00040	11.094
ATPIA2	R73570	ATPase, Na ⁺ /K ⁺ transporting, alpha 2 (+) polypeptide	1.4	0.00051	8.649
EST	R00965	EST	1.7	0.00040	9.623
ZNF2b2	N51853	Zinc finger protein 262	1.5	0.00049	8.427
RPL37A	BQ432698	Ribosomal protein L37a	1.6	0.00045	8.986
ETSV6	AA983191	ets variant gene 6 (TEL, oncogene)	1.6	0.00046	8.623
ARID3B	AA907181	AT rich interactive domain 3B (BRIGHT-like)	1.4	0.00049	8.629
HHEX	A1285592	Hematopoietically expressed homeobox	2.2	0.00030	11.612
MGEA5	BQ431821	Meningioma expressed antigen 5 (hyaluronidase)	1.4	0.00050	8.321
EST	AA994738	ESTs	1.4	0.00049	9.179
CRYZL1	AA459008	Crystallin, zeta (quinone reductase)-like I	1.4	0.00048	9.191
EST	AA282183	ESTs	1.4	0.00052	8.403
KPNB1	AA495898	karyopherin (importin) beta 1	1.6	0.00040	9.595
KPNB1	AA424912	karyopherin (importin) beta 1	1.5	0.00049	8.214
TP53BP1	U09477	Tumor protein p53 binding protein, 1	1.5	0.00047	8.792
GALNT7	A1681849	UDP-N-acetyl-alpha-U-galactosamine polypeptide N-acetylgalactosaminyltransferase 7	1.8	0.00040	9.039
Down-expression					
SNCA	AA455067	Synuclein, alpha (non A4 component of amyloid precursor)	-4.3	0.00019	-15.717
TFDP3	Al126424	Transcription factor Dp family, member 3	-2.7	0.00023	-15.807
EST	H08749	EST	-2.2	0.00025	-15.244
PCAF	N74637	P300/CBP-associated factor	-2.7	0.00022	-16.400
EST	R83876	EST	-2.8	0.00021	-17.616
OAT	M12267	Ornithine aminotransferase (gyrate atrophy)	-3.1	0.00021	-15.961
OAT	AA446820	Ornithine aminotransferase (gyrate atrophy)	-3.0	0.00022	-15.325
CPOX	AA700808	Coproporphyrinogen oxidase	-3.5	0.00019	-18.788
FLT1	AA058828	Fms-related tyrosine kinase 1 (vascular endothelial growth factor receptor)	-5.3	0.00016	-19.297
RHCE	N53959	Rhesus blood group, CcEe antigens	-4.1	0.00020	-15.244
PRDX2	H68845	Peroxiredoxin 2	-3.4	0.00020	-16.645
EST	BG571204	EST	-4.2	0.00019	-16.645
EST	AA252395	EST	-3.0	0.00021	-17.117
MYBL2	AA456878	V-myb myeloblastosis viral oncogene homolog (avian)-like 2	-4.2	0.00020	-15.735
EST	AA988630	EST	-3.1	0.00021	-16.459
TNFAIP3	AA476272	Tumor necrosis factor, alpha-induced protein 3	-3.0	0.00022	-15.402
MCF2	H05800	MCF2 cell line derived transforming sequence	-3.3	0.00019	-18.413
EST	HA2864	EST	-4.6	0.00018	-18.007
MAFB	T50121	V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	-4.3	0.00019	-15.997
ALDH1A1	AA664101	Aldehyde dehydrogenase 1 family, member A1	-5.0	0.00019	-15.449
EST	Al147237	EST	-7.9	0.00016	-17.845
IGLC2	T67053	Immunoglobulin lambda variable 3-21	-4.8	0.00017	-19.407
EST	AA663981	EST	-4.4	0.00019	-16.621
UBE2L3	AA669526	Ubiquitin-conjugating enzyme E2L 3	-2.9	0.00022	-15.485
IL1B	AA150507	Interleukin 1, beta	-2.5	0.00021	-16.056

SAM score (d) represents a significant distance of leukemia samples against normal bone marrow samples, obtained as described in Materials and methods. The positive value indicates over-expression compare to normal samples and negative (-) value indicates down-expression. The p-value was calculated by SAM's own t-test.

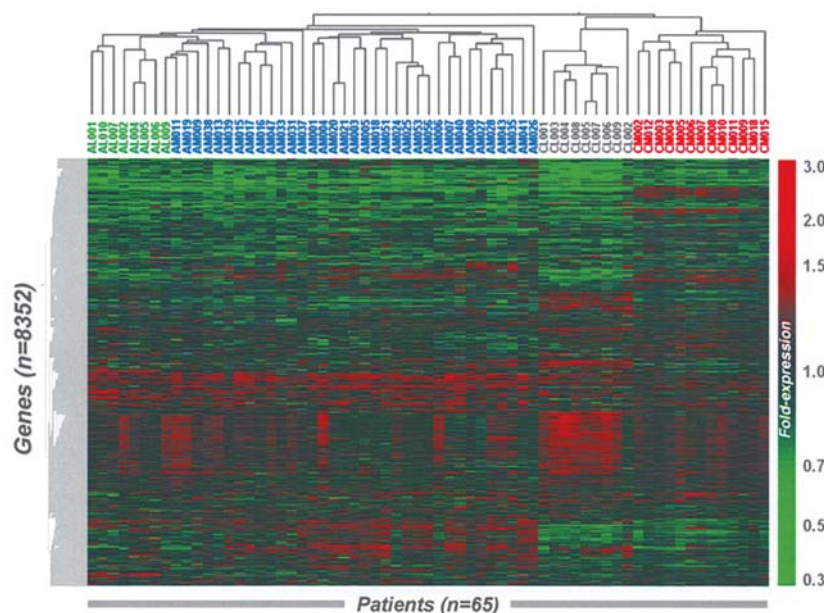


Figure 1. Hierarchical clustering analysis of diagnostic leukemia samples. Expression profiles of diagnostic adult leukemia samples (n=65) were obtained using a cDNA microarray, harboring 8352 DNA probes (rows). The leukemia specimens (columns) contained 8 samples of ALL (green), 35 of AML (blue), 9 of CLL (dark gray), and 13 of CML (red). The fold-change gene expression values are represented by colors, with red representing over-expression as compared to normal donor samples, and green representing down-regulation as compared to normal donor samples. Using both the GeneSpring 6.1 and Gene Tree View software packages, 2-dimensional hierarchical clusters were prepared on the basis of the gene expression profiles of each patient.

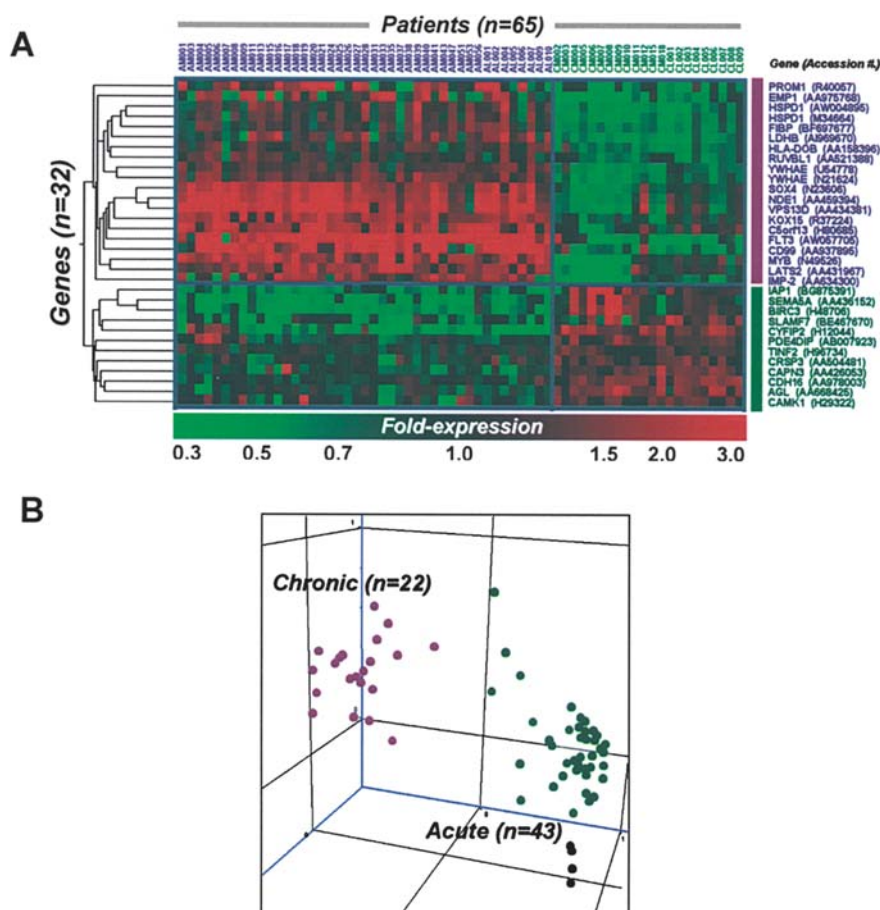


Figure 2. Representative genes for prediction of clinical course (acute or chronic). Prior to earnest analysis, the expression profiles (n=65) of leukemia patients were divided into 2 groups, containing the acute (n=43, purple) and chronic (n=22, green) samples. (A) We selected the top 100 genes on the basis of significant difference as determined by the SAM algorithm. These genes were then filtered once again using the GeneSpring 6.1 program, and 32 genes were ultimately generated. The fold-change gene expression values are represented by colors, with red representing overexpression as compared to that of normal donor samples, and green representing down-regulation. These genes are designated by symbols and GenBank numbers, on the left side of the figure. (B) The selected genes are applied to the PCA algorithm in order to confirm their ability to distinguish leukemia classes. Green- and purple-colored spots represent acute and chronic specimens, respectively.

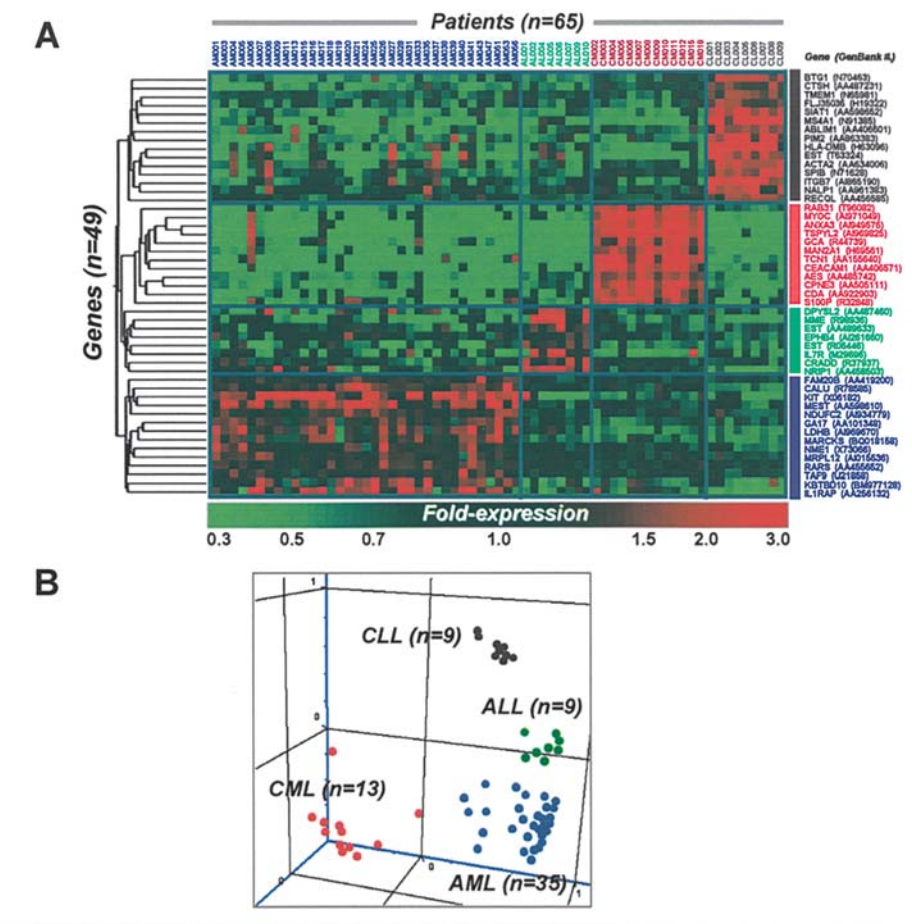


Figure 3. Representative genes for prediction of four-leukemia types. Prior to earnest analysis, the expression profiles (n=65) of leukemia patients were divided into 4 groups, corresponding to AML (n=35, blue), ALL (n=8, green), CML (n=13, red), and CLL (n=9, dark gray) samples. (A) We selected the top 100 genes with regard to significant difference, using the SAM algorithm. These genes were then filtered once again by the GeneSpring 6.1 program, and 49 genes (14 for AML, 8 for ALL, 12 for CML and 15 for CLL) were finally generated. The fold-change gene expression values are represented by colors, with red representing overexpression compared to normal donor samples, and green representing down-regulation. These genes are listed by their symbols and GenBank numbers on the right side of the figure. (B) The selected genes are subjected to the PCA algorithm in order to verify their efficacy with regard to the differentiation of leukemia classes. The AML, ALL, CML and CLL specimens are represented by blue-, green-, red- and dark gray-colored spots, respectively.

on 49 genes (14 for AML, 8 for ALL, 12 for CML and 15 for CLL), each representative of only one type (Fig. 3A). Previous studies have associated 15 (30.6%) and 11 (22.5%) of these genes with tumors and leukemia, respectively. In order to confirm whether the minimum genes selected would allow us to distinguish between each type of leukemia, although these genes originated from the same training sample set, we applied the genes to 65 patient samples. The PCA results suggested that, indeed, the gene subsets were sufficient for the discrimination of the four leukemia types (Fig. 3B).

Discussion

As leukemia is an aggregate disease of heterogeneous hematopoietic lineage cells, the optimal treatment choice for patients have depended primarily on its diagnostics (1). Due to the fact that no sole method, including RT-PCR and fluorescence *in situ* hybridization (FISH)-based detection assays, provides perfect diagnostic accuracy, over the last decades a variety of studies have focused on the detection of new predictive

genetic markers, in order to render therapy for leukemia, as well as for other tumors, more effective (2,3). Also, the coupling of conventional assays with novel diagnostic approaches may result in more accurate prediction, thus also increasing the efficacy of treatments. In this regard, leukemia is considered in this field to be an excellent disease model. From the preliminary studies of Golub *et al* through the genome-wide high throughput approach, to most recent studies, leukemia has been dissected in detail into novel subclasses, as have other tumors, albeit perhaps to a lesser degree (6-14). Furthermore, Bullinger and Valk recently reported that only few molecules are sufficient for a prediction of AML subclasses, strictly on the basis of gene expression profiles (16).

In the present report, we employed DNA microarray analyses to evaluate the transcriptional profiles of samples of clinical leukemia. These profiles were then utilized in order to differentiate between normal and leukemia clones, and to clarify which genes were expressed differentially in each of the leukemia types, or between acute and chronic leukemia samples. The ability to differentiate between such groups harbors clear diagnostic and therapeutic implications. The

expression profiles obtained from the 65 leukemia patients clearly demonstrated that leukemia cells and normal cells exhibit different expression patterns. Using both SAM and GeneSpring 6.1 analysis programs, we determined that several genes might prove useful as classification markers and/or therapeutic targets for our goals.

When the gene expression data from all specimens (n=65) was evaluated via unsupervised clustering analysis, and the result then visualized by 2-dimensional imaging and a hierarchical tree, we were surprised to see the results spontaneously divided into the respective disease types (AML, ALL, CML and CLL) generated at diagnosis (Fig. 1). In addition to this finding, further filtration of the results through the two programs provided the minimum number of genes (n=49) distinct to each human leukemia type (Fig. 3). This result was generally consistent with our expectations, supporting the notion that there are unique gene expression profiles for each disease. We also found a mixed tree zone, in which we found genes from both ALL and AML (AM011, 019, 009, 038, 013, 039, 015, 017, 016, 047, 033, 031 and 037) patients, clustered in the ALL group region. As there had been a previous report of a distinct gene expression pattern for the mixed-lineage leukemia gene (MLL) in cases of acute leukemia (9), we conducted a comparison of the ALL and AML patients in the mixed tree zones to their genetic backgrounds. However, we detected no meeting point which might clarify this abnormal result.

The patient tree shown also demonstrates the intimate relationship existing between each type of leukemia, as well as a certain amount of discriminatory data. As shown in Fig. 2, the patient samples were separated into two major branches on the basis of their genome profiles, even in cases in which the patient's groupings were not artificially manipulated. One of these is a sample of acute leukemia, and the other is chronic. With regard to this result, we would like to carefully suggest that each leukemic genetic background is closely genetically connected with its clinical course (acute and chronic), rather than with a specific lineage (lymphocytic or myelocytic).

The majority of the genes selected in this study have not, thus far, been the focus of interest in studies of oncogenesis and/or leukemogenesis. This is, perhaps, reasonable, as the functions and roles of only a small proportion of these genes has been discovered. Only about a quarter of the selected genes have been correlated with leukemia and/or tumors. Among our several biological observations, we have focused on the common characteristics of leukemia, without reference to type class. To this aim, we selected the 50 genes ranked highest with regard to significant expression changes, as compared the profiles of normal samples (Table II). Among these genes, a few, including PIK3CA (p110 α catalytic subunit of phosphoinositide 3-kinase) (19), CD69 (p60, early T-cell activation antigen) (20), SNCA (synuclein, α) (21) and OAT (ornithine aminotransferase) (22) have been previously associated with tumors. However, the mRNA expression level of FLT1, which is known as a member of the vascular endothelial growth factor (VEGF) receptor tyrosine kinase family (23), evidenced an opposite expression pattern. This is involved in the transduction of extracellular signals, resulting in cell growth and differentiation. Although a previous study conducted by Ratajczak *et al* demonstrated

that FLT1 mRNA was expressed in some AML and CML patients (24), and FLT1 RNA levels were down-regulated in all of the patients in our study. This indicates that defective FLT1 expression probably affects the blockage of normal signaling for differentiation in the early stages of hematopoietic cells, leading to the arrest of immature cells and leukemogenesis.

The other biologically relevant observation in which we became interested was a marker for class prediction and the gene as a pathological cause of each of the leukemia types. Based on this question, we identified 32 and 49 genes as predictive markers for clinical course and each leukemia type, using both of the algorithms referenced in this study, respectively (Figs. 2 and 3). Furthermore, we evaluated the possibility of using these genes as clinical type predictors, utilizing the PCA program. Oncogenic factors, including FLT3 (Fms-like tyrosine kinase 3) (25) and MYB (v-myb myeloblastosis viral oncogene homologue-like 2) (26) were over-expressed in the acute leukemic patients, but CAPN3 (calcium-activated neutral protease 3) (27) was over-expressed in the chronic patients (Fig. 2). However, no evidence was collected supporting the notion that these genes play pathological roles in each of the groups. Although a key relationship appears to exist between the selected genes and pathogenic functions inherent to leukemogenesis, the newly identified type prediction genes may provide new understanding into the unique biological characteristics of each group or type of leukemia. Consequently, detailed studies into cellular functions, as well as prediction markers, will be required, and are clearly warranted.

In conclusion, our results suggest that expression profiling, via a genome-wide approach, may offer a new method for the more accurate diagnosing of adult leukemia. The significant genes identified in our study may provide some insight into the abnormal biology of leukemia, and may also indicate a potential new target for the development of novel therapeutic drugs.

Acknowledgements

This work was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (01-PJ10-PG6-01GN16-0005).

References

1. Harris NL, Jaffe ES, Diebold J, *et al*: World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting 1997. *J Clin Oncol* 17: 3835-3849, 1999.
2. 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.
3. Foon KA and Todd RF III: Immunologic classification of leukemia and lymphoma. *Blood* 68: 1-31, 1986.
4. Kalwinsky DK, Raimondi SC, Schell MJ, *et al*: Prognostic importance of cytogenetic subgroups in *de novo* pediatric acute non-lymphocytic leukemia. *J Clin Oncol* 8: 75-83, 1990.
5. Ma SK, Wan TS and Chan LC: Cytogenetics and molecular genetics of childhood leukemia. *Hematol Oncol* 17: 91-105, 1999.
6. Su AI, Welsh JB, Sapinoso LM, *et al*: Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 61: 7388-7393, 2001.

7. Ramaswamy S, Tamayo P, Rifkin R, *et al*: Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 98: 15149-15154, 2001.
8. Golub TR, Slonim DK, Tamayo P, *et al*: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537, 1999.
9. Armstrong SA, Staunton JE, Silverman LB, *et al*: 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, *et al*: Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 350: 1617-1628, 2004.
11. Bullinger L, Dohner K, Bair E, *et al*: Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 350: 1605-1616, 2004.
12. Yeoh EJ, Ross ME, Shurtleff SA, *et al*: Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1: 133-143, 2002.
13. Ross ME, Mahfouz R, Onciu M, *et al*: Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* 104: 3679-3687, 2004.
14. Boer JM, Huber WK, Sultmann H, *et al*: Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. *Genome Res* 11: 1861-1870, 2001.
15. Song JH, Kim JM, Kim SH, *et al*: Comparison of the gene expression profiles of monocytic versus granulocytic lineages of HL-60 leukemia cell differentiation by DNA microarray analysis. *Life Sci* 73: 1705-1719, 2003.
16. Bullinger L and Valk PJ: Gene expression profiling in acute myeloid leukemia. *J Clin Oncol* 23: 6296-6305, 2005.
17. Preisler HD: The leukemias. *Dis Mon* 40: 525-579, 1994.
18. Szer J: Current thinking in the management of leukaemia. *Aust Fam Physician* 23: 1500-1505, 1994.
19. Estilo CL, O-Charoenrat P, Ngai I, *et al*: The role of novel oncogenes squamous cell carcinoma-related oncogene and phosphatidylinositol 3-kinase p110alpha in squamous cell carcinoma of the oral tongue. *Clin Cancer Res* 9: 2300-2306, 2003.
20. Esplugues E, Sancho D, Vega-Ramos J, *et al*: Enhanced anti-tumor immunity in mice deficient in CD69. *J Exp Med* 197: 1093-1106, 2003.
21. Prasad JE, Kumar B, Andreatta C, Nahreini P, Hanson AJ, Yan XD and Prasad KN: Overexpression of alpha-synuclein decreased viability and enhanced sensitivity to prostaglandin E₂, hydrogen peroxide, and a nitric oxide donor in differentiated neuroblastoma cells. *J Neurosci Res* 76: 415-422, 2004.
22. Matthaai KI and Williams JF: Ornithine aminotransferase turnover in host tissues of tumor-bearing rats. *J Natl Cancer Inst* 79: 805-809, 1987.
23. Bellamy WT, Richter L, Frutiger Y and Grogan TM: Expression of vascular endothelial growth factor and its receptors in hematopoietic malignancies. *Cancer Res* 59: 728-733, 1999.
24. Ratajczak MZ, Ratajczak J, Machalinski B, *et al*: Role of vascular endothelial growth factor (VEGF) and placenta-derived growth factor (PlGF) in regulating human haemopoietic cell growth. *Br J Haematol* 103: 969-979, 1998.
25. Lopes de Menezes DE, Peng J, Garrett EN, *et al*: CHIR-258: a potent inhibitor of FLT3 kinase in experimental tumor xenograft models of human acute myelogenous leukemia. *Clin Cancer Res* 11: 5281-5291, 2005.
26. Santin AD, Zhan F, Bignotti E, *et al*: Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology* 331: 269-291, 2005.
27. Witkowski JM, Zmuda-Trzebiatowska E, Swiercz JM, *et al*: Modulation of the activity of calcium-activated neutral proteases (calpains) in chronic lymphocytic leukemia (B-CLL) cells. *Blood* 100: 1802-1809, 2002.