Two Distinct Gene Expression Signatures in Pediatric Acute Lymphoblastic Leukemia with MLL Rearrangements

Shuichi Tsutsumi, Takeshi Taketani, Kunihiro Nishimura, Xijin Ge, Tomohiko Taki, Kanji Sugita, Eiichi Ishii, Ryoji Hanada, Misao Ohki, Hiroyuki Aburatani, and Yasuhide Hayashi

ABSTRACT

Acute lymphoblastic leukemia (ALL) with 11q23 translocations is usually associated with MLL gene rearrangement, but little is known about its leukemogenesis. We analyzed the gene expression profiles of pediatric ALL samples according to their translocations. Using oligonucleotide microarray analysis, we identified distinct expression profiles for 23 ALL samples with 11q23 translocations, including t(4;11) (n = 15), t(11;19) (n = 6), and t(5;11) (n = 2), compared with 9 ALL samples with other translocations, including t(12;21) (n = 6) and t(11;19) (n = 3). Gene expression scores of FLT3, MEI1, and CD44 for samples with MLL rearrangements were particularly high compared with those for other ALL samples. Statistical analysis of the gene expression profiles for the 21 ALL samples with MLL rearrangements at diagnosis revealed two subgroups that exclusively correlated with prognosis but not with any other clinico-pathological factor. The transcription factors CBF2 and CDP were highly expressed in the poor and good prognosis subgroups, respectively. In addition, their downstream target genes were differentially expressed. These findings provide new insights into the biological mechanisms of leukemogenesis and prognosis for pediatric ALL with MLL rearrangements.

INTRODUCTION

The prognosis of children with ALL has improved remarkably over the last 2 decades (1–3). This success has been achieved by using risk-directed therapy, which was developed after the realization that pediatric ALL is a heterogeneous disease (4). However, 20–25% of ALL patients still experience a relapse. Attempts to classify pediatric ALL into therapeutically relevant risk categories have relied mainly on clinical parameters, including age and WBC count at diagnosis, as well as early response to treatment (4). Recent advances in molecular biology have identified several genes involved in chromosomal translocations (10–12) and is highly conserved across species. The t(11;19), is strongly associated with poor prognosis (6). Thus, cytogenetic or direct molecular genetic methods have become an essential part of the routine diagnosis and follow-up of acute leukemia patients, as well as increasing our understanding of leukemogenesis.

The MLL gene (also known as ALL-1 or HRX), located at 11q23, encodes a protein of 3969 amino acids containing zinc fingers and AT-hook motifs and has homology with Drosophila trithorax protein (7–9). The MLL gene fuses with >30 genes on various partner chromosomes (10–12) and is highly conserved across species. Through its regulation of the HOX genes, MLL is essential for normal mammalian development and hematopoiesis. Although the function of the various MLL fusion genes and proteins is poorly understood, it appears that their fusion proteins disrupt the ability of wild-type MLL to regulate HOX gene expression, leading to leukemogenesis (13).

Recently, a genomic approach to cancer classification, including leukemia classification (14–17), based on gene expression monitoring using DNA microarrays, has been reported, with a distinct gene expression in pediatric T-ALL shown to be associated with a poor/good prognosis (17). MLL-RE-ALL has been reported to have characteristic, distinct gene expression profiles that are consistent with an early hematopoietic progenitor cell expressing selected multilineage markers and individual HOX genes. Clustering algorithms reveal that, based on their gene expression patterns, acute leukemia with MLL rearrangements can clearly be separated from conventional ALL and AML (18), suggesting that they constitute a distinct disease. Among MLL-RE-ALLs, infant patients have a poor prognosis. However, children >1 years old have a relatively good prognosis (4). We used an oligonucleotide microarray to analyze the expression of >12,600 genes in leukemic cells from 31 pediatric ALL patients, including 15 with t(4;11), 6 with t(11;19), and 2 with t(5;11). We found that MLL-RE-ALL could be identified from the distinct expression pattern of several genes, including FLT3, CD44, HOXA9, and MEIS1. Furthermore, using the gene expression profiles, each of the t(4;11), t(11;19), or t(5;11) found in MLL-RE-ALL could be classified into two distinct groups, with differential prognosis, irrespective of their translocation partner chromosomes.

MATERIALS AND METHODS

Leukemia Samples. Leukemia cells from the bone marrow or peripheral blood of ALL patients were obtained with informed consent at diagnosis or relapse. In each case, the percentage of blasts was >90%. CD19 was expressed in all samples, but CD2, CD5, and CD7 were not expressed in any samples. We analyzed 32 ALL samples with chromosomal translocations, comprising 3 samples with t(1;19), 6 with t(12;21), and 23 with MLL rearrangements, including 15 (t(4;11), 6 (t(11;19), and 2 (t5;11). Samples were obtained both at diagnosis and relapse from one patient with t(4;11) and only at relapse from one MLL-RE-ALL sample. Therefore, the remaining 21 samples were obtained only at diagnosis. All of the translocations were subjected to karyotype analysis, fluorescence in situ hybridization, and/or Southern blot analyses, and MLL partner genes were confirmed by RT-PCR as described elsewhere (11, 19–21). The t(1;19), t(12;21), t(4;11), t(11;19), and t(5;11) samples were found...
to have E2A-PBX1 (22), TEL-AML1, MLL-AF4 (19), MLL-ENL (19), and MLL-AF5q31 (11) fusion genes, respectively. Infant MLL-Rec-ALL patients were mainly treated according to the MLL-96 protocol (23).

**RNA Extraction and High-Density Oligonucleotide Array Analysis.** Total RNA and genomic DNA were isolated from frozen cells using the ISOGEN reagent (Nippon Gene, Tokyo, Japan) according to the manufacturer’s protocol. The quality of total RNA was examined by gel electrophoresis to confirm that the ribosomal 28S and 18S RNA bands were intact. The experimental procedures for GeneChip (Affymetrix, Santa Clara, CA) were performed according to the Affymetrix GeneChip expression analysis technical manual as described previously (24, 25). Briefly, 3–5 μg of total RNA were used to synthesize biotin-labeled cRNA, which was then hybridized to a GeneChip Human U95 V2 oligonucleotide array (Affymetrix). After washing, the arrays were stained with streptavidin-phycoerythrin and analyzed on a Hewlett-Packard Scanner to collect the image data. GeneChip Analysis Suite software 4.0 was used to calculate the AD for each gene probe set on the array, which was shown as an intensity value of the gene expression. The AD values were normalized for each array so that the average of all AD values was 100. Raw data are available on the Internet.*

**Statistical Analysis.** For each expression data set, where the AD values lay outside the range (10–8000), the value was reset to a minimum of 10 and a maximum of 8000. Subsequently, all values were log transformed for further analysis. Hierarchical clustering analysis was performed using GeneSpring (Silicon Genetics, Inc., Redwood, CA), CLUSTER, and TREEVIEW software (Eisen Lab.; Ref. 26).

Genes that correlated with particular class distinctions were identified as described by Golub et al. (14). We used the signal-to-noise statistic \((\sigma_s + \sigma_r)/\sqrt{\sigma_s \sigma_r}\) (\(\sigma_s\) and \(\sigma_r\) represent the mean and SD of expression, respectively, for each class). We also carried out 100,000 permutations of the samples by Mann-Whitney U and Kruskal-Wallis H tests to determine whether the correlations were more significant than would be expected by chance alone. Applying PCA, the coordinates of the first three principal components for each sample were selected. An SVM algorithm (27) was also applied to classify the samples using a modified version of the SVM light.

**RT-PCR and Sequence Analysis.** cDNA was reverse transcribed from 5 μg of total RNA using a cDNA synthesis kit (Invitrogen, Carlsbad, CA). PCR amplification was performed with the Advantage 2 PCR kit (Clontech, Palo Alto, CA) by incubating at 94°C for 2 min, followed by 35 cycles of 94°C for 30 s and 68°C for 2 min using FLT3-specific primers (F1; 5′-CCCAACTG-CACAGAAGAGATCACAG-3′ and F2; 5′-TACAGCTGTATTGGAATG-GTGGAAGG-3′). The PCR products were purified using a Qiaquick PCR purification kit (Qiagen, Hilden, Germany) and subjected to direct sequencing with the primers (5′-CCAGACGTCCGCTTCAAGAG-3′, 5′-GCCCTGAGATTGTATCCGGTCAGTC-3′, 5′-GGGGAATCCTTCTACATTG-3′, 5′-ATCTCTAGTCATTTCCAAAC-3′, F1 and F2) using DYEnamic ET Terminator Kits (Amersham Biosciences, Piscataway, NJ) on an Applied Biosystems DNA sequencer. Internal tandem duplications of the FLT3 gene were investigated by RT-PCR using the primers (5′-TGTCGAGCATACCTTAAAAC-3′ and 5′-ATCTCTAGTACCTTCCAAAC-3′) and electrophoresis as described previously (20).

**Genomic PCR and Restriction Fragment-length Polymorphism Analysis.** We amplified the exon 20 of the FLT3 gene by genomic PCR using the primers (5′-GTTGTGTGCACATCTCATGCCC-3′ and 5′-CCACAGTGAGTGCTTTGTACCAGT-3′) inoculating at 94°C for 2 min, followed by 35 cycles of 94°C for 30 s and 68°C for 30 s. Amplified products were digested with EcoRV and subjected to electrophoresis on an agarose gel (Fig. 4).

**RESULTS**

**Gene Expression Profiling Can Identify the Translocation Type in ALL Samples.** We analyzed 32 pro-B or early pre-B ALL samples, including those with MLL rearrangements (n = 23), TEL-AML1 (n = 6), and E2A-PBX1 (n = 3), with Affymetrix oligonucleotide microarrays containing 12,600 probe sets. All samples showed high CD19 expression signals. Relatively higher expression of CD44 and lower of CD10 (MME), CD22, CD24, and CD79B were found in patients with MLL rearrangements rather than in those with TEL-AML1 and E2A-PBX1 (supplementary information is available on the Internet).*

The results of the PCA were plotted with three-dimensional scaling to determine whether we could identify the ALL translocation types from their gene expression profiles. Samples carrying MLL rearrangements, those with TEL-AML1 and E2A-PBX1, were resolved with this method. In contrast, no distinct subgroups were observed for defined MLL rearrangements, such as MLL-ENL, MLL-AF4, or MLL-AF5q31 fusion genes (Fig. 1A). To classify the samples according to the similarity of their gene expression patterns and classify the genes according to the expression similarities over the samples, we applied a two-dimensional hierarchical clustering algorithm. In this analysis, samples with TEL-AML1 and E2A-PBX1 fusion were also subclassified into their respective clusters (Fig. 1B).

*Internet address: http://www2.genome.rcast.u-tokyo.ac.jp/MLL.
Gene expression profiling in pediatric ALL

MLL rearrangements, E2A-PBX1 and TEL-AML1) using signal-to-noise analysis (supplementary Fig. 7; Ref. 14). Among 50 genes that were uniquely expressed in a subset of MLL rearrangements, 13 genes were also found in 50 unique genes for MLL-Re-ALL in a previous report (18). These included LGALS1, CD44, CD45, and PMX1. Furthermore, we found that FLT3 and MEIS1 were expressed highly, and CD10 (MME) was less expressed in the samples with MLL rearrangements.

Two Distinct Patterns of Gene Expression among MLL-Re-ALL Samples. Gene expression patterns for leukemias with MLL rearrangements have been reported to be unique compared with those in other ALL without MLL rearrangements or AML (18). To investigate the variations in gene expression patterns among MLL-Re-ALL samples, we used a two-dimensional hierarchical clustering analysis on gene expression profiles for our 23 MLL-Re-ALL samples. This analysis produced two major sample clusters (Fig. 2). Two expression profiles generated from samples at onset and relapse from the same patient were classified in the same cluster. Excluding the two relapse samples, a random permutation test showed that the expression profiles of these two groups were statistically different (P < 0.05 by Mann-Whitney U test), gender, or treatment (by Mann-Whitney U test). Among 50 genes that were uniquely expressed in a subset of MLL rearrangements, 13 genes were also found in 50 unique genes for MLL-Re-ALL in a previous report (18). These included LGALS1, CD44, CD45, and PMX1. Furthermore, we found that FLT3 and MEIS1 were expressed highly, and CD10 (MME) was less expressed in the samples with MLL rearrangements.

Gene Expression Signature Has Prognostic Relevance for MLL. To elucidate the possible clinical significance associated with the two main expression profile groups, Kaplan-Meier analysis was performed for relapse and survival (Fig. 3). Excluding the two samples obtained at relapse, Kaplan-Meier analysis showed that Cluster B in Fig. 2 was associated with a distinctly favorable prognosis. As shown in Fig. 3B, the overall probability of survival at 3 years was 92 ± 8% SE for Cluster B and 0% for Cluster A (P = 0.0005 by Log-rank analysis). The probability of event-free survival at 3 years was 73 ± 14% for Cluster B and 0% for Cluster A (P = 0.01). MLL-Re-ALL patients < 1 year old are reportedly associated with a poor prognosis, whereas those with t(4;11) and >1 year old have relatively good prognosis (2, 4, 6). In our study, only 1 patient (No. 22 on Fig. 2) was older than 1 year and subclassified in the favorable cluster B.

As reported previously (18), we found that MLL-Re-ALL is characterized by elevated levels of FLT3 expression (shown in supplementary Fig. 8A). The AD values of FLT3 showed no significant difference between the two clusters. Internal tandem duplication of the FLT3 gene has been reported in 20–30% of adult AML (28, 29) and 15% of childhood AML (30). In addition, mutations of the FLT3 gene have been reported in 5% of adult AML (31). We investigated Asp835 mutations of the FLT3 gene by direct sequence and restriction fragment-length polymorphism analysis of genomic PCR products. Three (14%) of 21 MLL-Re-ALL patients, 1 from Cluster A and 2 from Cluster B, were found to have Asp835 mutations (Fig. 4). This was confirmed by sequence analysis of FLT3 cDNA. No other mutations were found in the intracellular region of FLT3 by sequence analysis of

**Fig. 2.** Two-dimensional hierarchical clustering analysis on MLL-Re-ALL samples. Each column represents a gene and each row a sample. Relative expression levels are shown in red (relatively high) and cyan (relatively low). The colored bar on either side of the cluster identifies the fusion gene of each sample (blue, MLL-AF4; green, MLL-ENL; purple, MLL-AF5q31). 0. samples from the same patient. Two samples were taken at relapse (A). Asp835 mutations were found in three samples (B).

**Fig. 3.** Survival analysis of MLL-Re-ALL. Kaplan-Meier curves for Cluster A (n = 9) versus Cluster B (n = 12). A, disease-free survival; B, overall survival.

**Fig. 4.** Detection of Asp835 mutation of the FLT3 gene. The arrow (356 bp) indicates Asp835 mutation of the FLT3 gene. The lane numbers correspond to the sample numbers in Fig. 2.
Biological Aspects between the Two Differential Prognostic Clusters. To investigate the biological features of the two clusters, we used gene expression profiling to identify differentially expressed genes. Of the 30 genes selected for further analysis, 7 were found to be specifically expressed in Cluster A, and 17 in Cluster B. The remaining 6 genes were expressed in both clusters.

The precision of subclassification of unknown samples into the two clusters by gene expression profiling is especially important in MLL-Remission-ALL because there are few conventional methods for predicting the prognosis of those types of ALL. A supervised SVM was used against these higher and lower signal-to-noise genes to classify the samples. The test sample was classified using a leave-one-out model against these higher and lower signal-to-noise genes to classify the samples. This result suggests that the prognosis of MLL-Remission-ALL can be predicted reliably by using the expression profiles of selected genes.

DISCUSSION

To make an accurate diagnosis of pediatric ALL, many clinical diagnostic examinations are required. It is necessary to consider the interrelationship of various prognostic factors, including chromosomal translocations (1, 2, 4, 12). ALL patients with t (12;21; TEL-AML1) are associated with a good clinical outcome (2, 4), whereas MLL-Remission-ALL patients are associated with a poor outcome (1, 2, 4, 10, 12). Our data suggest that MLL-Remission-ALL can be diagnosed from gene expression profiles.

One of the most interesting findings in this study was the remarkable variance in the gene expression signatures of MLL-Remission-ALL. This difference was more significant than that between 13 samples with t (4;11) and 6 with t (11;19). This result strongly suggests that at least two subgroups exist in MLL-Remission-ALL independent of the MLL partner genes, with patients in one subgroup (Cluster A) having a remarkably poor prognosis.

It was reported that an internal tandem duplication of FLT3 in AML predicted poor prognosis, and recently, mutations ofFLT3 have been reported to be rare in ALL (31). Our result showed Asp835 mutations ofFLT3 in 3 (14%) of 21 MLL-Remission-ALL patients, but we found no tandem duplications ofFLT3 similar to our previous report (20). Elevated expression of the FLT3 gene was not associated with either cluster, and Asp835 mutations were not associated with prognosis.

Except for SPN, expression of these leukocyte markers, CD10, CD19, CD22, CD24, CD44, CD79B, and TdT, were not significantly correlated with the two clusters (supplementary Fig. 6). It was reported that, with intensive treatment, including hematopoietic stem cell transplantation, 30–40% of MLL-Remission-ALL infants remained free of relapse.

**Fig. 5. Genes specifically expressed in Cluster A or Cluster B.** The top 30 genes (Gene list U) and bottom 30 genes (Gene list F) by signal-to-noise values are shown. Each column represents a leukemia sample and each row a gene. The signal-to-noise value (SN value) for the probe set of U95A array, gene symbol, and gene ontology is shown on the right. Each gene list, the genes are arranged according to their ontologies. Relative expression levels are shown in red (high) and cyan (low).
(4, 23, 33, 34), suggesting the existence of two patient groups with differential prognosis. Our results demonstrate that gene expression profiling is able to predict the prognosis of these distinct groups of MLL-Re-ALL more accurately than the conventional methods, such as karyotype analysis.

The top 30 and bottom 30 genes that were differently expressed between the two clusters (Clusters A and B) provided us with an insight into the biological behavior of ALL. In Gene list U, we found the transcriptional co-activators TRIP2 and CBF2. CBF2 was reported as a co-activator of NF-Y, which also bound the CCAAT motif (35–38). On the other hand, CDP in (Gene list F) also recognizes the CCAAT motif (32). CDP plays an essential role in the differentiation of hematopoietic cells (39). Loss of heterozygosity and reduced CDP expression has been observed in human uterine leiomyoma and breast cancer, providing the first evidence that CDP can act as a potential tumor suppressor (40, 41). Our analysis of the promoter sites of the listed 60 genes suggested that CDP might suppress the expression of four genes in Cluster B samples. The transcription factors, CBF2 and CDP, may regulate the expression of, and be correlated quantitatively with, many genes that were differently expressed between the two clusters. For example, CBF2 was reported to induce CDC2, which has tandem CCAAT motifs in its promoter site (37, 42). Actually, CDC2 showed a higher expression in Cluster A samples (P < 0.05 by unpaired t-test).

In the 30 genes expressed highly in Cluster B, we found 8 signal transduction genes, including four Rho family genes (RAC2, ARHGAP1/GRAF, and CFL1) and two GTP-associated genes (DNM2 and Sipa1). This result suggests that different pathways are activated in Cluster B. ARHGAP1/GRAF is one of the partner genes of MLL (46). Biallelic mutations of the ARHGAP1/GRAF gene have been identified in samples of myelodysplastic syndrome and AML. Recent studies have confirmed the oncogenic potential of Rho proteins (47, 48), and several studies suggest that Rho GTPases might be overfunctional in human cancers (49). It seems probable that the Rho pathway plays some roles in the leukemogenesis of patients with Cluster B.

In conclusion, these different gene expressions between the subgroups provided us with valuable information for clarifying the mechanism of leukemogenesis in MLL-Re-ALL. Further analysis of MLL-Re-ALL should lead to more accurate characterization of the key molecules of leukemogenesis and help in the search for new drug targets and diagnostic markers.

ACKNOWLEDGMENTS
We thank Shoko Shohama, Hisae Soga, Hiroko Meguro, Daisuke Komura and Shogo Yamamoto for their excellent comments. We also thank Drs. Akira Morimoto, Masahiro Sako and all of the participants in the Japan Infant Leukemia Study Group for providing samples and clinical data.

REFERENCES


Two Distinct Gene Expression Signatures in Pediatric Acute Lymphoblastic Leukemia with MLL Rearrangements

Shuichi Tsutsumi, Takeshi Taketani, Kunihiro Nishimura, et al.