MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia

Scott A. Armstrong[1–4], Jane E. Staunton[5], Lewis B. Silverman[1,3,4], Rob Pieters[6], Monique L. den Boer[6], Mark D. Minden[7], Stephen E. Sallan[1,3,4], Eric S. Lander[5], Todd R. Golub[1,3,4,5*] & Stanley J. Korsmeyer[2,4,8*]

*These authors contributed equally to this work.

Published online: 3 December 2001, DOI: 10.1038/ng765

Acute lymphoblastic leukemias carrying a chromosomal translocation involving the mixed-lineage leukemia gene (*MLL*, *ALL1*, *HRX*) have a particularly poor prognosis. Here we show that they have a characteristic, highly distinct gene expression profile that is consistent with an early hematopoietic progenitor expressing select multilineage markers and individual *HOX* genes. Clustering algorithms reveal that lymphoblastic leukemias with *MLL* translocations can clearly be separated from conventional acute lymphoblastic and acute myelogenous leukemias. We propose that they constitute a distinct disease, denoted here as MLL, and show that the differences in gene expression are robust enough to classify leukemias correctly as MLL, acute lymphoblastic leukemia or acute myelogenous leukemia. Establishing that MLL is a unique entity is critical, as it mandates the examination of selectively expressed genes for urgently needed molecular targets.

A subset of human acute leukemias with a decidedly unfavorable prognosis possess a chromosomal translocation involving the mixed-lineage leukemia gene (*MLL*, *HRX*, *ALL1*) on chromosome segment 11q23 (refs 1–4). The leukemic cells, which typically have a lymphoblastic morphology, have been classified as acute lymphoblastic leukemia (ALL). Unlike other types of childhood ALL, however, the presence of the *MLL* translocation in ALL often results in an early relapse after chemotherapy. As *MLL* translocations are typically found in infant leukemias and in chemotherapy-induced leukemia, it has remained uncertain whether host-related factors or tumor-intrinsic biological differences are responsible for poor survival.

Lymphoblastic leukemias with a rearranged *MLL* or germline *MLL* are similar in most morphological and histochemical characteristics. Immunophenotypic differences associated with lymphoblasts bearing an *MLL* translocation include a lack of the early lymphocyte antigen CD10 (ref. 5), expression of the proteoglycan NG2 (ref. 6) and a propensity to co-express the myeloid antigens CD15 and CD65 (ref. 5). This prompted the corresponding gene to be called mixed-lineage leukemia[1] and gave rise to models that remain largely unresolved, in which the leukemia reflects disordered cell-fate decisions or the transformation of a more multipotent progenitor.

Translocations in *MLL* result in the production of a chimeric protein in which the amino–terminal portion of MLL is fused to the carboxy–terminal portion of 1 of more than 20 fusion partners[7]. This has led to models of leukemogenesis in which the MLL fusion protein either may confer gain of function or neomorphic properties or may interfere with normal MLL function (with the *MLL* translocation representing a dominant-negative gene). Moreover, mice heterozygous for *Mll* ($Mll^{+/-}$) show developmental aberrations[8,9], suggesting that the disruption of one allele by chromosomal translocation may also manifest itself as haplo-insufficiency in leukemic cells.

The MLL protein is a homeotic regulator that shares homology with *Drosophila trithorax (trx)* and positively regulates the maintenance of homeotic (*Hox*) gene expression during development[8]. Studies of *Mll*-deficient mice indicate that *Mll* is required for proper segment identity in the axioskeletal system and also regulates hematopoiesis[9]. As Mll normally regulates the expression of *Hox* genes, its role in leukemogenesis may include altered patterns of *HOX* gene expression. Much evidence shows that *HOX* genes are important for appropriate hematopoietic development[10]. In addition, the t(7;11) (p15;p15) found in human acute myelogenous leukemia (AML) results in a fusion of *HOXA9* to the nucleoporin *NUP98* (refs 11,12). Thus, *HOX* genes represent one set of transcriptional targets that warrant assessment in leukemias with *MLL* translocation.

We considered that *MLL* translocations might maintain a gene expression program that results in a distinct form of leukemia

[1]*Departments of Pediatric Oncology,* [2]*Cancer Immunology and AIDS and* [8]*Howard Hughes Medical Institute, Dana-Farber Cancer Institute, Boston, Massachusetts, USA.* [3]*Division of Pediatric Hematology/Oncology, Children's Hospital, Boston, Massachusetts, USA.* [4]*Harvard Medical School, Boston Massachusetts, USA.* [5]*Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research, Cambridge Massachusetts, USA.* [6]*Division of Pediatric Hematology/Oncology, Sophia Children's Hospital, University of Rotterdam, The Netherlands.* [7]*Princess Margaret Hospital, The University of Toronto, Ontario, Canada. Correspondence and requests for materials should be addressed to S.K. (e-mail: stanley_korsmeyer@dfci.harvard.edu) or T.G. (e-mail: golub@genome.wi.mit.edu).*

and reasoned that RNA profiles might resolve whether leukemias bearing an *MLL* translocation represent a truly biphenotypic leukemia of mixed identity, a conventional B-cell precursor ALL with expression of limited myeloid genes, or a less committed hematopoietic progenitor cell. In addition, comparing gene expression profiles of lymphoblastic leukemias with and without rearranged *MLL* is important because of their markedly different response to standard ALL therapy and because such analysis may identify molecular targets for therapeutic approaches. The expression profiles reported here show that ALLs possessing a rearranged *MLL* have a highly uniform and distinct pattern that clearly distinguishes them from conventional ALL or AML and warrants designation as the distinct leukemia MLL.

Results

MLL is distinct from conventional ALL

To further define the biological characteristics specified by *MLL* translocations, we compared the gene expression profiles of leukemic cells from individuals diagnosed with B-precursor ALL bearing an *MLL* translocation against those from individuals diagnosed with conventional B-precursor ALL that lack this translocation. Initially, we collected samples from 20 individuals with conventional childhood ALL (denoted ALL), 10 of which had a *TEL/AML1* translocation. In addition, we collected samples from 17 individuals affected with the *MLL* translocation (denoted MLL). Details of the affected individuals and expression data are available online (Methods).

First, we determined whether there were genes among the 12,600 tested whose expression pattern correlated with the presence of an *MLL* translocation. We sorted the genes by their degree of correlation with the MLL/ALL distinction (Fig. 1) and used permutation testing to assess the statistical significance of the observed differences in gene expression[13]. For the 37 samples tested, roughly 1,000 genes are underexpressed in MLL as compared with conventional ALL, and about 200 genes are relatively highly expressed (data not shown). Thus, MLL shows a gene expression profile markedly different from that of conventional ALL.

MLL shows multilineage gene expression

Inspection of the genes differentially expressed between MLL and ALL is instructive (Fig. 1). Many genes underexpressed in MLL have a function in early B-cell development. These include genes expressed in early B cells[14,15], *MME*, *CD24*, *CD22* and *DNTT* (mouse *TdT*); genes required for appropriate B-cell development[16–19], *TCF3*, *TCF4*, *POU2AF1* and *LIG4*; and *SMARCA4* (mouse *Snf2b*), which is correlated with B-precursor ALL in an AML/ALL comparison[13] (Fig. 1 and Web Note A). Genes encoding certain adhesion molecules are relatively overexpressed in MLL, including *LGALS1*, *ANXA1*, *ANXA2*, *CD44* and *SPN*.

Several genes that are expressed in hematopoietic lineages other than lymphocytes are also highly expressed in MLL. These include genes that are expressed in progenitors[20–22], *PROML1*, *FLT3* and *LMO2*; myeloid-specific genes[23–25], *CCNA1*, *SERPINB1*, *CAPG* and *RNASE3*; and at least one natural killer cell–associated gene[26], the gene encoding NKG2D (Fig. 1 and Web Note A). Overexpression of *HOXA9* and *PRG1* in MLL is of particular interest, as these genes have been reported to be highly expressed in AML[13] and overexpression of *HOXA9* has been associated with a poor prognosis[13].

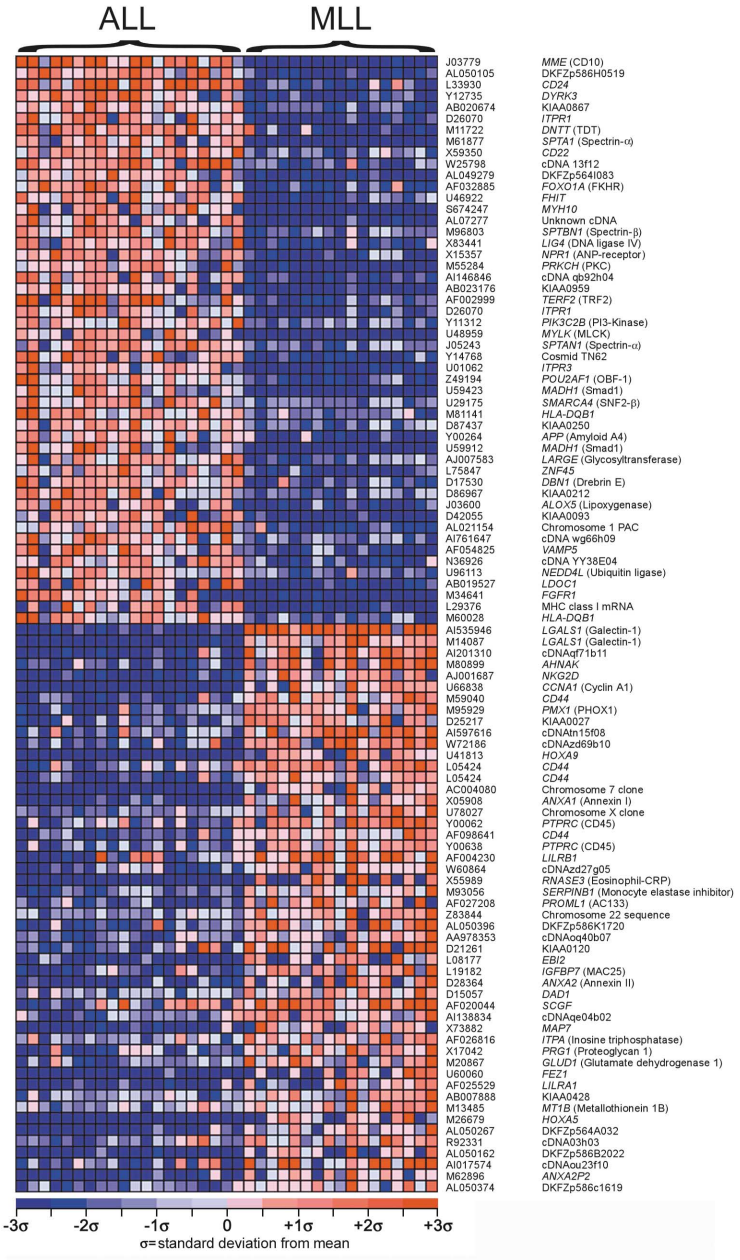

Fig. 1 Genes that distinguish ALL from MLL. The 100 genes that are most highly correlated with the class distinction are shown. Each column represents a leukemia sample, and each row represents an individual gene. Expression levels are normalized for each gene, where the mean is 0, expression levels greater than the mean are shown in red and levels less than the mean are in blue. Increasing distance from the mean is represented by increasing color intensity. The top 50 genes are relatively underexpressed and the bottom 50 genes relatively overexpressed in MLL. Gene accession numbers and the gene symbols or DNA sequence names are labeled on the right. Individual samples are arranged such that column 1 corresponds to ALL patient 1, column 2 corresponds to ALL patient 2, and so on. Information about the samples along with the top 200 genes that make the ALL/MLL distinction and their accession numbers can be found on our web site (http://research.dfci.harvard.edu/korsmeyer/MLL.htm).

MLL is arrested at an early stage of hematopoiesis

Because lymphoblasts with *MLL* rearrangement express many myeloid-specific genes, we carried out a detailed assessment of the expression of lymphoid genes. Genes known to mark early B-lymphoid commitment, such as *CD79B* (mouse *Igb*) and *CD19*, are expressed in MLL, albeit at lower levels than in ALL (Fig. 2). *MME* is not expressed in MLL (Fig. 2), whereas the interleukin-7 receptor gene (*IL7R*) is expressed at similar levels in ALL and MLL (Web Note A).

Several genes have been shown to vary their expression level as murine hematopoietic cells differentiate from stem cell to common lymphoid progenitor, to pro-B and then to pre-B cells. *CD79B*, *CD24*, *CD44* and *SPN* expression represent early steps of lymphoid development[14,27]. Expression of *CD79B* and *CD24* increases with maturation, whereas *CD44* and *SPN* expression decrease[27]. The MLL samples express relatively low levels of *CD24* and *CD79B* but high levels of both *CD44* and *SPN* (Fig. 2). Together, these data indicate that MLL represents a maturational arrest at an early lymphoid progenitor stage of development.

Some *HOX* genes are expressed at higher levels in MLL than in conventional ALL

The fact that several members of the class I *Hox* genes are regulated by *Mll*[8] prompted us to make a detailed comparison of the patterns of *HOX* gene expression in ALL and MLL. Several of the 20 class I *HOX* genes present on the microarrays show significant and consistent differences in expression. The genes *HOXA9* and *HOXA5* are not expressed in conventional ALL but are expressed, often at high levels, in most MLL samples (Fig. 3).

Similarly, *HOXA4* is typically expressed in MLL but rarely in conventional ALL (Fig. 3). Expression of *HOXC6* is mildly elevated in MLL (Web Note A). The *HOX* patterns show selectivity, however, as other genes such as *HOXA7* show no obvious difference in their expression pattern (Fig. 3). The gene *MEIS1*, encoding a cofactor for HOX proteins that can accelerate *Hoxa9*-dependent leukemia[28], is also overexpressed significantly in MLL (Web Note A), as has been reported for the subset of leukemias containing t(4;11)[29].

MLL is distinct from both AML and conventional ALL

MLL is characterized by the expression of myeloid-specific genes, which raises the possibility that MLL may be more closely related to AML. To examine this possibility and also whether MLL can be separated as a distinct type of leukemia, we carried out a principal component analysis (PCA) using the gene expression profiles of MLL, ALL and AML specimens. The clustering algorithm of PCA reduces complex multidimensional data to a few specified dimensions so that it can be visualized effectively[30].

Initially, we carried out the analysis in an unsupervised manner, using the 8,700 genes that show some variability in expression level. As expected, the ALL and AML samples have substantial separation (Fig. 4*a*). Notably, the MLL samples are largely separate from the AML or ALL samples (Fig. 4*a*). To determine whether this separation could be attributed to a difference in hematopoietic identity, we carried out a similar analysis using the 500 genes whose expression best distinguishes the separation of AML and ALL. When projected into this 500-gene space using PCA, the MLL samples are again separate from the AML and ALL samples (Fig. 4*b*).

As the above clustering analyses supported three distinct entities of ALL, AML and MLL, we examined whether selected genes could be identified that distinguish each type of leukemia from the other two (Fig. 5). Compared with the other leukemias, conventional ALL samples express high levels of lymphoid-specific genes (*MME*, *CD24*, *DNTT* and *LIG4*). AML samples express high levels of myeloid-specific genes *DF*, *CTSD* and *ANPEP*, whereas MLL samples express high levels of genes associated with hematopoietic progenitors (*PROML1*, *LMO2*, *FLT3*; Fig. 5). Permutation analysis indicates that roughly 200 genes are significantly overexpressed in MLL as compared with the other two leukemia categories (data not shown). Together, the PCA and gene expression comparisons (Figs 4 and 5) indicate that MLL is a separable, distinct disease on the basis of its gene expression profile.

Gene expression profiles correctly classify ALL, MLL and AML

A more stringent assessment of the power of the difference in gene expression profiles is their capacity to assign individual samples as MLL, conventional ALL or AML. Currently, the detection of *MLL* translocations in leukemia samples is most often

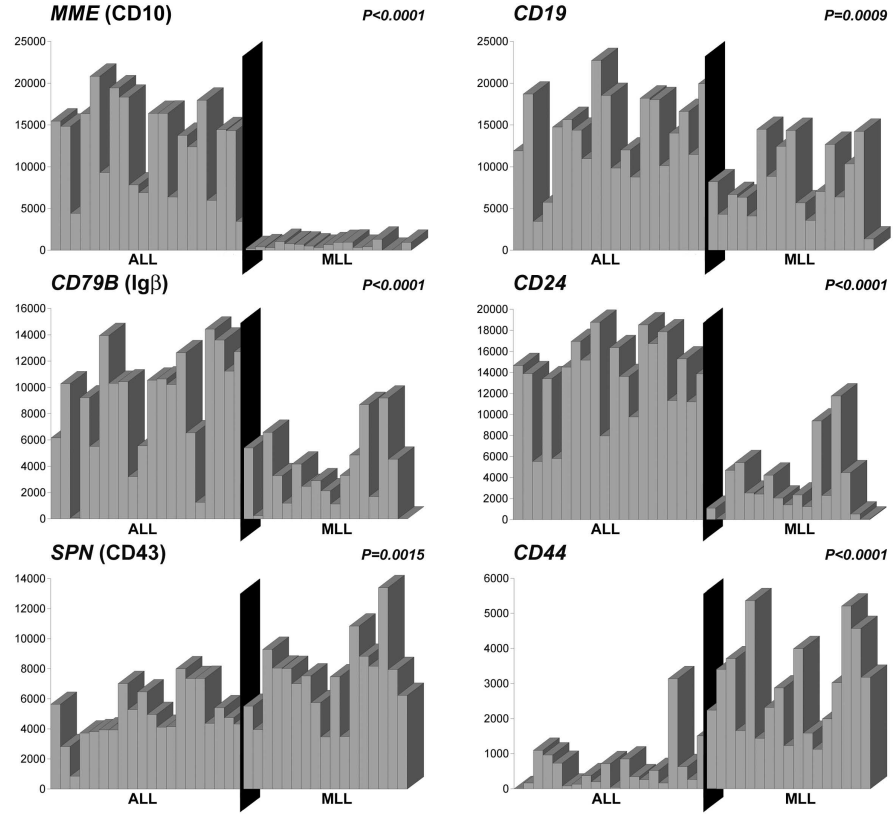

Fig. 2 Selected early lymphocyte gene expression in ALL and MLL. Relative expression levels of *MME* (*CD10*), *CD19*, *CD79B*, *CD24*, *SPN* (*CD43*) and *CD44* in ALL and MLL samples are shown. Each bar represents an individual leukemia sample. The expression values are raw data obtained from Affymetrix GENECHIP analysis after the arrays are scaled as described in Methods.

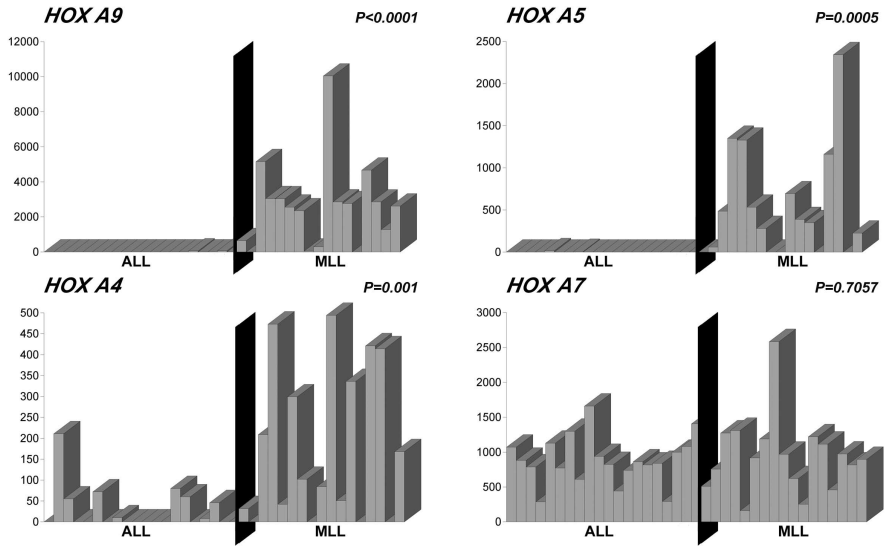

Fig. 3 Selected *HOX* gene expression in ALL and MLL. Relative levels of expression of *HOXA9*, *HOXA5*, *HOXA4* and *HOXA7* in ALL and MLL samples are shown. The expression values are obtained using Affymetrix GENECHIP analysis after the arrays are scaled (Methods).

carried out by cytogenetic analysis or by fluorescence *in situ* hybridization (FISH), both of which can sometimes fail technically and are not always available. Thus, other approaches to assign individual cases correctly to meaningful subsets of leukemia would be useful. To test this possibility, we developed a three-class predictor based on a *k*-nearest-neighbors algorithm[31]. This algorithm assigns a test sample to a class by identifying the *k* nearest samples in the training set and choosing the most common class among these *k* nearest neighbors. For this purpose, distances were defined by a euclidean metric on the basis of the expression levels of a specified number of genes.

We assessed the accuracy of this method using a cross-validation approach with a training set of 20 ALL, 17 MLL and 20 AML samples. When 1 of the 57 samples was removed, the genes that most closely correlated with the ALL/MLL/AML class distinction were identified and the expression of these genes was used to determine the class of the withheld sample. The model assigned the withheld sample to the appropriate class with 95% accuracy. Moreover, this accuracy was maintained as the number of genes used to build the predictor was increased from 40 to 250 (Fig. 6), which is further testimony to the strong distinction among these categories of leukemia.

To assess whether the unique signature of gene profiles in MLL samples could be attributed to their occurrence in infants, we tested the above model using ten independent leukemia samples. The test set comprised three childhood (older than 1 y) conventional ALLs, two lymphoblastic leukemias of childhood carrying cytogenetically verified *MLL* translocations, two infant (1 y) leukemias in which cytogenetic analysis did not detect an *MLL* translocation and three AML samples. Using the 100 genes that best correlated with the three-class distinction, nine of ten samples were correctly classified as MLL, ALL or AML. The one apparent error was an infant that was reported to be negative for an *MLL* rearrangement by cytogenetics but was consistently predicted to have a rearrangement on the basis of gene expression profile. This prompted further analysis by FISH, which confirmed that this infant leukemia did indeed possess an *MLL* translocation and that the prospective assignment by expression profiling was correct. Together, these data show that the unique gene expression profile characteristic of MLL is largely independent of the age of the affected individual.

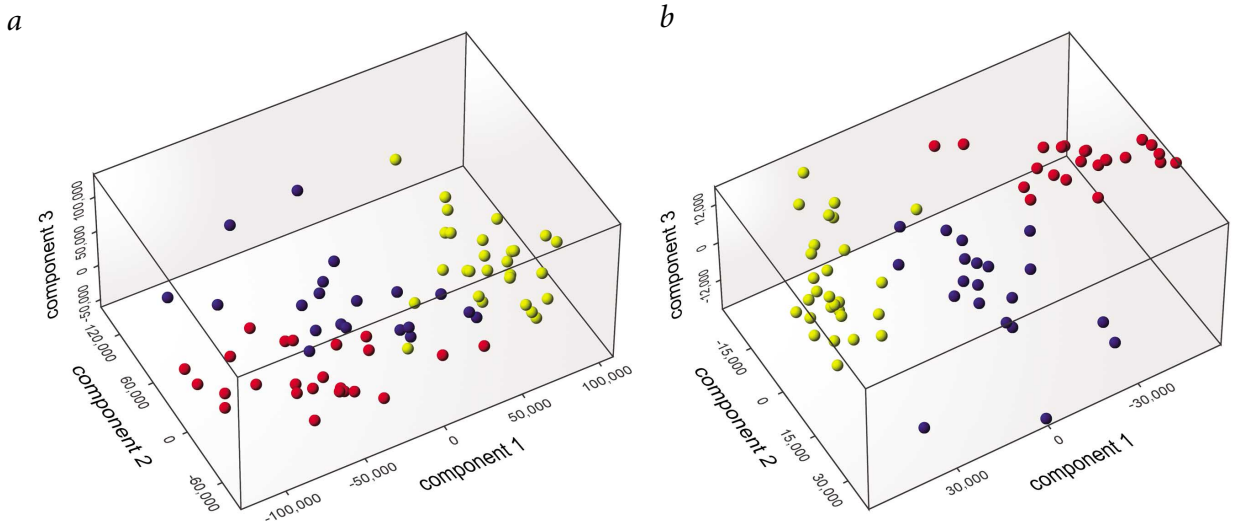

Fig. 4 Comparison of gene expression between ALL, MLL and AML. ***a***, Principal component analysis (PCA) plot of ALL (red), MLL (blue) and AML (yellow) carried out using 8,700 genes that passed filtering. ***b***, PCA plot comparing ALL (red), MLL (blue) and AML (yellow) using the 500 genes that best distinguished ALL from AML. Three-dimensional virtual reality modeling language (VRML) plots can be viewed at our web site (http://research.dfci.harvard.edu/korsmeyer/MLL.htm).

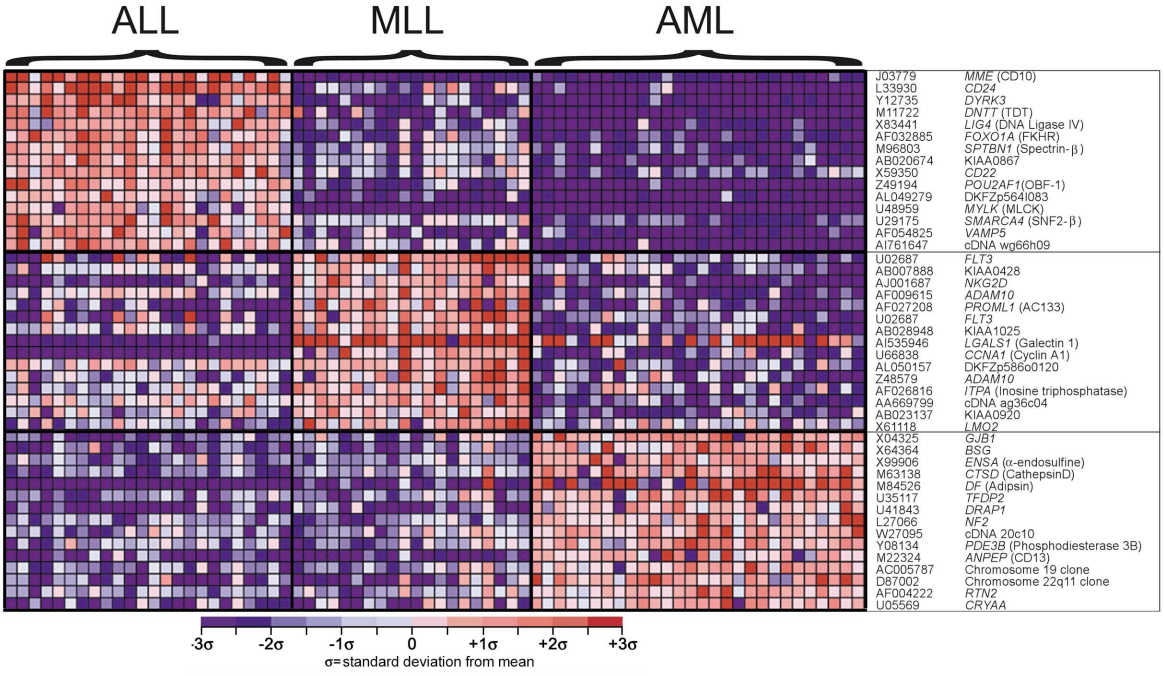

Fig. 5 Genes that are specifically expressed in MLL, ALL or AML. The top 15 genes that are most highly correlated with one type of leukemia as compared with the other two are shown. Each column represents a leukemia sample and each row a gene. The gene accession numbers and the gene symbols or DNA sequence names are shown on the right. Relative expression levels are shown in red (relatively high) and blue (relatively low) and ordered as described in Fig. 1.

Discussion

Gene expression profiles of lymphoblastic leukemias that possess an *MLL* translocation are remarkably consistent and differ significantly from those of other leukemias. We therefore propose that they should be considered a distinct disease termed 'mixed-lineage leukemia'. This concept is supported by our comparison of MLL samples with conventional B-cell precursor ALL samples that lack *MLL* rearrangement, in which roughly 1,000 genes were underexpressed and 200 overexpressed in the group possessing an *MLL* rearrangement. Moreover, evaluation of the expression profiles using PCA indicates that MLL is clearly separable from both conventional ALL and AML.

The expression differences are so robust that we could correctly classify roughly 95% of our leukemic samples as MLL, ALL or AML. As testimony to the extent of divergence of MLL, it remained separable from ALL and AML when 250 genes were used to build the class predictor. Future studies should examine whether further subdivision of ALL is warranted. Of note, *MLL* rearrangements can also occur in AML, and gene expression analysis of cytogenetically well characterized AML will determine whether their gene profiles are distinct. Our data strongly suggest a model in which a specific chromosomal translocation results in a distinct type of lymphoblastic leukemia, rather than a model in which all translocations merely provide transformation events that subsequently converge on a common pathway of leukemogenesis.

The gene expression patterns of MLL provide insight into the models proposed for its cellular origin. A summary of expression profiles shows that MLL expresses some lymphocyte-specific and myeloid-specific genes, but at lower levels than either conventional ALL or AML, respectively. On the basis of murine studies that have defined gene expression patterns during lymphocyte commitment[14,27], the low expression of *CD24* and *CD79B*, along with the high expression of *SPN* and *CD44*, suggest that MLL is arrested at an earlier stage of development than conventional ALL. Moreover, the expression of genes that are typically found in progenitor cells suggests that MLL represents an early hematopoietic progenitor. This is consistent with studies showing that multilineage gene expression occurs in hematopoietic progenitors before full lineage commitment[32].

Another possibility is that MLL might represent the expansion of a bipotential B-macrophage progenitor[33,34]. Early B cells can be induced to differentiate into myelomonocytic cells under certain conditions[35], and the derivation of macrophages from leukemia cell lines has been well documented[36]. An attractive model is that the MLL fusion protein drives the 'transdifferentiation' of an early lymphocyte progenitor. The expression of many myeloid and monocyte/macrophage-specific genes is consistent with MLL reflecting a very early B-cell progenitor that has initiated transdifferentiation.

The *HOX* genes that are selectively expressed in MLL are candidates for direct targets of the MLL fusion proteins. Studies of *Mll*-ablated mice have indicated that select members of the clustered *Hox* genes require MLL for their expression. As overexpression of *Hoxa9* has also been shown to induce AML in mouse models[28], and its expression is controlled by levels of *Mll*[37], misexpression of *HOXA9* might be an important component of leukemogenesis driven by *MLL* translocations. Future studies should determine whether MLL fusion proteins directly activate *HOX* genes and thus lead to defects in hematopoietic differentiation.

To our knowledge, this is the first whole-genome profiling study to show that a chromosomal translocation can specify a unique gene expression program. In addition, the finding that MLL is a distinct disease is of pathogenic and therapeutic importance. Lymphoblastic leukemias with *MLL* translocations are recognized as having a poor prognosis because standard ALL therapies have been relatively ineffective. The unique identity of MLL provides insight into this poor response and suggests that new therapeutic approaches are needed.

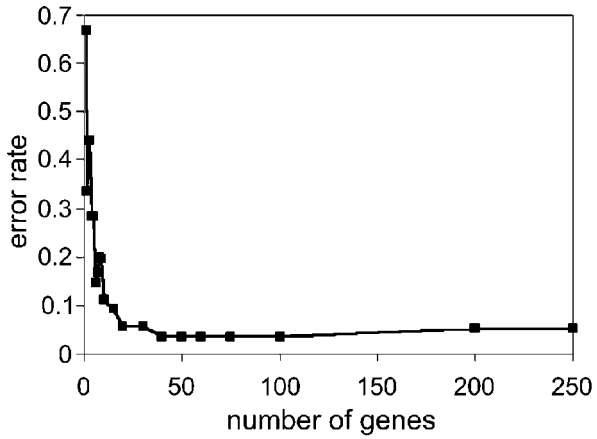

Fig. 6 Classification of ALL, MLL and AML on the basis of their gene expression profile. The error rate in class prediction (*y* axis) is plotted against the number of genes used to build the model (*x* axis). This plot was generated using a cross-validation approach. One of 57 samples (20 ALL, 17 MLL, 20 AML) was withheld, and the specified number of genes that best correlated with the ALL/MLL/AML three-class distinction were identified. The class of the withheld sample was then determined on the basis of the expression pattern of the genes identified in this manner. This process was repeated for all 57 samples, and the error rate in class prediction was calculated. The models were generated using between 1 and 250 genes.

Notably, pilot studies have shown that cytarabine, an important agent in myeloid leukemia treatment, may improve the outcome for MLL-affected individuals[38–40]. But it is the translocation-specific therapies that are proving attractive for their efficacy and lack of toxicity. Other leukemias in which a translocation specifies a distinct disease include chronic myelogenous leukemia (CML), with its BCR–ABL fusion, and acute promyelocytic leukemia, with its PML–RARα fusion. The tailored development of the tyrosine kinase inhibitor STI571 for treating CML and the use of all-*trans* retinoic acid in APL have substantially improved the outcome in people with those diseases[41,42].

Although pharmacological approaches to the complex regulatory capacity of MLL may prove challenging, the distinct gene expression signature defined here for MLL may provide unanticipated molecular targets. Of note, *FLT3* is the most differentially expressed gene that distinguishes MLL from ALL and AML (Fig. 5). Aberrations of *FLT3*, especially duplication of its juxtamembrane domain, have been noted in some cases of AML and may be leukemogenic[43–45]. As a tyrosine kinase receptor, FLT3 represents an attractive target for rational drug development. The power of whole-genome profiling to define unique disease entities such as MLL will stimulate the development of therapies based on molecular targets.

Methods

Leukemia samples. After obtaining informed consent, we obtained leukemia samples from the peripheral blood or bone marrow of affected individuals at diagnosis or relapse. All patients were treated on a protocol approved by the Institutional Review Board. All MLL and ALL samples were diagnosed as $CD19^+$ B-precursor ALL cells by pathologists at the institution where the samples were collected. When the samples were obtained from peripheral blood, the percentage of blasts was greater than 60% of the total white blood cells present. Fifteen of the samples with an *MLL* translocation and all of the conventional childhood ALL samples were obtained from affected individuals treated by Dana Farber Cancer Institute protocols between 1980 and 2001. Three of the infant leukemia samples with *MLL* rearrangements were obtained from individuals treated by the Interfant99 protocol, and the two adult samples with *MLL* rearrangements were obtained from individuals treated at Princess Margaret Hospital in Toronto. Except for three of the conventional ALL samples and four of the MLL samples that were obtained at relapse, all samples were diagnostic specimens. The AML samples have been described[13] and were from both adults and children. Eight of the *MLL* rearranged samples contain t(4;11), one t(9;11), three t(11;19), one t(3;11) and one t(1;11). Six of the *MLL* rearrangements were detected by either FISH or Southern blot, and thus the translocation partner is unknown.

We purified mononuclear cells from red blood cells and neutrophils by Ficoll–Hypaque density centrifugation, and they were either frozen in liquid nitrogen with 10% dimethyl sulfoxide in fetal calf serum or put directly into Trizol (Gibco-BRL) for RNA purification. Further details regarding leukemia samples can be found at our web site (http://research.dfci.harvard.edu/Korsmeyer/MLL.htm).

Assessment for the presence of *MLL* translocations. We assessed all leukemia samples by standard cytogenetics. We screened all childhood ALL samples for the presence of a *TEL/AML1* translocation by RT–PCR as described[46]. Any sample for which cytogenetics failed and that had no *TEL/AML1* translocation was further assessed either by FISH using a probe that spans the 11q23 breakpoint or by Southern blotting[39,47]. We did not assess AML samples for chromosomal translocations.

RNA purification, labeling and hybridization. We used a total of 10×10^6–20×10^6 cells to prepare total RNA using the Trizol (Gibco-BRL) purification method. This generally yielded between 5 μg and 20 μg of total RNA, the quality of which was examined by gel electrophoresis. If the ribosomal RNA bands were intact, the RNA was determined to be of good quality and 5–15 μg was used for subsequent production of biotinylated antisense RNA as described[13]. A detailed protocol can be found at Whitehead/MIT Genome Center Molecular Pattern Recognition web site (http://www.genome.wi.mit.edu/MPR). We excluded samples if less than 15 μg of labeled RNA was produced.

Labeled RNA was then hybridized to Affymetrix U95A or U95A V2 oligonucleotide arrays at 45 °C for 16 h. Arrays were washed and stained with streptavidin–phycoerytherin (SAPE, Molecular Probes). The signal was amplified using a biotinylated anti-streptavidin antibody (Vector Laboratories) at 3 μg ml^{-1}. This was followed by a second staining with SAPE. We used normal goat IgG as a blocking agent. We carried out the scans on Affymetrix scanners and calculated the expression values using Affymetrix GENECHIP software. The chip image was then scanned visually for obvious differences between arrays. If there were obvious abnormalities present in the image, the sample was rehybridized. The scans were then normalized on the basis of a linear scaling method as described in Web Note A. Samples were disregarded if the scaling factor was greater than threefold.

Data analysis. We identified the genes that correlated with particular class distinctions as described[13]. We used the signal-to-noise statistic $(\mu_0-\mu_1)/(\sigma_0+\sigma_1)$, where μ and σ represent the median and standard deviation of expression, respectively, for each class. We carried out 100 permutations of the samples to determine whether the correlations were greater than would be expected by chance with a 99% confidence.

The class predictor was carried out using a cross-validation approach and the *k*-nearest-neighbors algorithm as described in Web Note A. We removed a sample, identified the genes that correlated with the MLL, AML, ALL class distinction and assigned the removed sample to a class on the basis of the expression of these genes.

We carried out principal component analysis using S-plus statistical software and the default settings. The coordinates of the three principal components for each sample were then used to project the samples in three dimensions. Further details of the analysis can be found at our web site (http://research.dfci.harvard.edu/korsmeyer/MLL.htm).

Note: Supplementary information is available on the Nature Genetics web site (http://genetics.nature.com/supplementary_info/).

Acknowledgments

We thank V. Dalton, S. Waters, A. Cardoso and N. Haining for help with accumulating samples and obtaining clinical data; J. Korsmeyer and D. Neuberg for help with statistical analysis; J. Fletcher for help with interpretation of

cytogenetic data; E. Smith for assistance with figures and editing; C. Ladd, M. Angelo and other members of the Whitehead/MIT Center for Genome Research Program in Cancer Genomics for technical help and developing data analysis tools and P. Ernst and J. Hsieh for discussions. This work was supported in part by an NIH grant, an American Society of Hematology Fellow Scholar Award (S.A.A.), an American Society of Hematology Junior Faculty Scholar Award (T.R.G.), the Belfer Cancer Genomics Center and Bristol-Myers Squibb, Millennium Pharmaceuticals, and Affymetrix (T.R.G.).

Received 25 April; accepted 9 October 2001.

1. Rowley, J.D. The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.* **32**, 495–519 (1998).
2. Gu, Y. *et al.* The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila trithorax*, to the AF-4 gene. *Cell* **71**, 701–708 (1992).
3. Tkachuk, D.C., Kohler, S. & Cleary, M.L. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**, 691–700 (1992).
4. Domer, P.H. *et al.* Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL–AF4 fusion product. *Proc. Natl Acad. Sci. USA* **90**, 7884–7888 (1993).
5. Pui, C.H. *et al.* Clinical characteristics and treatment outcome of childhood acute lymphoblastic leukemia with the t(4;11)(q21;q23): a collaborative study of 40 cases. *Blood* **77**, 440–447 (1991).
6. Behm, F.G., Smith, F.O., Raimondi, S.C., Pui, C.H. & Bernstein, I.D. Human homologue of the rat chondroitin sulfate proteoglycan, NG2, detected by monoclonal antibody 7.1, identifies childhood acute lymphoblastic leukemias with t(4;11)(q21;q23) or t(11;19)(q23;p13) and MLL gene rearrangements. *Blood* **87**, 1134–1139 (1996).
7. Dimartino, J.F. & Cleary, M.L. Mll rearrangements in haematological malignancies: lessons from clinical and biological studies. *Br. J. Haematol.* **106**, 614–626 (1999).
8. Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A. & Korsmeyer, S.J. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505–508 (1995).
9. Hess, J.L., Yu, B.D., Li, B., Hanson, R. & Korsmeyer, S.J. Defects in yolk sac hematopoiesis in Mll-null embryos. *Blood* **90**, 1799–1806 (1997).
10. Buske, C. & Humphries, R.K. Homeobox genes in leukemogenesis. *Int. J. Hematol.* **71**, 301–308 (2000).
11. Nakamura, T. *et al.* Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nature Genet.* **12**, 154–158 (1996).
12. Borrow, J. *et al.* The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nature Genet.* **12**, 159–167 (1996).
13. Golub, T.R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
14. Hardy, R.R. & Hayakawa, K. B cell development pathways. *Annu. Rev. Immunol.* **19**, 595–621 (2001).
15. LeBien, T.W. Fates of human B-cell precursors. *Blood* **96**, 9–23 (2000).
16. Murre, C. Role of helix–loop–helix proteins in lymphocyte development. *Cold Spring Harb. Symp. Quant. Biol.* **64**, 39–44 (1999).
17. Fruman, D.A. *et al.* Impaired B cell development and proliferation in absence of phosphoinositide-3-kinase p85α. *Science* **283**, 393–397 (1999).
18. Schubart, K. *et al.* B cell development and immunoglobulin gene transcription in the absence of Oct-2 and OBF-1. *Nature Immunol.* **2**, 69–74 (2001).
19. Frank, K.M. *et al.* Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173–177 (1998).
20. Yin, A.H. *et al.* AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood* **90**, 5002–5012 (1997).
21. Rosnet, O. *et al.* Human FLT3/FLK2 gene: cDNA cloning and expression in hematopoietic cells. *Blood* **82**, 1110–1119 (1993).
22. Dong, W.F., Billia, F., Atkins, H.L., Iscove, N.N. & Minden, M.D. Expression of rhombotin 2 in normal and leukaemic haemopoietic cells. *Br. J. Haematol.* **93**, 280–286 (1996).
23. Yang, R. *et al.* Cyclin A1 expression in leukemia and normal hematopoietic cells. *Blood* **93**, 2067–2074 (1999).
24. Remold-O'Donnell, E., Chin, J. & Alberts, M. Sequence and molecular characterization of human monocyte/neutrophil elastase inhibitor. *Proc. Natl Acad. Sci. USA* **89**, 5635–5639 (1992).
25. Rosenberg, H.F., Ackerman, S.J. & Tenen, D.G. Human eosinophil cationic protein. Molecular cloning of a cytotoxin and helminthotoxin with ribonuclease activity. *J. Exp. Med.* **170**, 163–176 (1989).
26. Ho, E.L. *et al.* Murine Nkg2d and Cd94 are clustered within the natural killer complex and are expressed independently in natural killer cells. *Proc. Natl Acad. Sci. USA* **95**, 6320–6325 (1998).
27. Kondo, M., Weissman, I.L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
28. Nakamura, T., Largaespada, D.A., Shaughnessy, J.D. Jr, Jenkins, N.A. & Copeland, N.G. Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nature Genet.* **12**, 149–153 (1996).
29. Rozovskaia, T. *et al.* Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4:11) abnormality. *Oncogene* **20**, 874–878 (2001).
30. Venables, W.N. & Ripley, B.D. *Modern Applied Statistics with S-Plus* (Springer, New York, 1994).
31. Dasarathy, V.B. (ed.) *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques* (IEEE Computer Society Press, Los Alamitos, 1991).
32. Hu, M. *et al.* Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785 (1997).
33. Montecino-Rodriguez, E., Leathers, H. & Dorshkind, K. Bipotential B-macrophage progenitors are present in adult bone marrow. *Nature Immunol.* **2**, 83–88 (2001).
34. Cumano, A., Paige, C.J., Iscove, N.N. & Brady, G. Bipotential precursors of B cells and macrophages in murine fetal liver. *Nature* **356**, 612–615 (1992).
35. Nutt, S.L., Heavey, B., Rolink, A.G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556–562 (1999).
36. Borrello, M.A. & Phipps, R.P. The B/macrophage cell: an elusive link between $CD5^+$ B lymphocytes and macrophages. *Immunol. Today* **17**, 471–475 (1996).
37. Hanson, R.D. *et al.* Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc. Natl Acad. Sci. USA* **96**, 14372–14377 (1999).
38. Ludwig, W.D. *et al.* Immunophenotypic and genotypic features, clinical characteristics, and treatment outcome of adult pro-B acute lymphoblastic leukemia: results of the German multicenter trials GMALL 03/87 and 04/89. *Blood* **92**, 1898–1909 (1998).
39. Silverman, L.B. *et al.* Intensified therapy for infants with acute lymphoblastic leukemia: results from the Dana-Farber Cancer Institute Consortium. *Cancer* **80**, 2285–2295 (1997).
40. Pieters, R. *et al.* Relation between age, immunophenotype and *in vitro* drug resistance in 395 children with acute lymphoblastic leukemia—implications for treatment of infants. *Leukemia* **12**, 1344–1348 (1998).
41. Tallman, M.S. *et al.* All-*trans*-retinoic acid in acute promyelocytic leukemia. *N. Engl. J. Med.* **337**, 1021–1028 (1997).
42. Druker, B.J. *et al.* Efficacy and safety of a specific inhibitor of the BCR–ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344**, 1031–1037 (2001).
43. Nakao, M. *et al.* Internal tandem duplication of the *flt3* gene found in acute myeloid leukemia. *Leukemia* **10**, 1911–1918 (1996).
44. Zhao, M. *et al. In vivo* treatment of mutant FLT3-transformed murine leukemia with a tyrosine kinase inhibitor. *Leukemia* **14**, 374–378 (2000).
45. Tse, K.F., Mukherjee, G. & Small, D. Constitutive activation of FLT3 stimulates multiple intracellular signal transducers and results in transformation. *Leukemia* **14**, 1766–1776 (2000).
46. Loh, M.L. *et al.* Incidence of TEL/AML1 fusion in children with relapsed acute lymphoblastic leukemia. *Blood* **92**, 4792–4797 (1998).
47. Cuthbert, G., Thompson, K., Breese, G., McCullough, S. & Bown, N. Sensitivity of FISH in detection of MLL translocations. *Genes Chromosom. Cancer* **29**, 180–185 (2000).