npg
Leukemia (2005) 19, 1042–1050
© 2005 Nature Publishing Group All rights reserved 0887-6924/05 $30.00
www.nature.com/leu

Gene expression profiling of leukemic cell lines reveals conserved molecular signatures among subtypes with specific genetic aberrations

A Andersson[1], P Edén[2], D Lindgren[1], J Nilsson[3], C Lassen[1], J Heldrup[4], M Fontes[3], Å Borg[5], F Mitelman[1], B Johansson[1], M Höglund[1] and T Fioretos[1]

[1]*Department of Clinical Genetics, Lund University Hospital, Sweden;* [2]*Department of Complex Systems, Theoretical Physics, Lund University, Sweden;* [3]*Center for Mathematical Sciences, Lund University, Sweden;* [4]*Department of Pediatrics, Lund University Hospital, Sweden; and* [5]*Department of Oncology, Lund University Hospital, Sweden*

Hematologic malignancies are characterized by fusion genes of biological/clinical importance. Immortalized cell lines with such aberrations are today widely used to model different aspects of leukemogenesis. Using cDNA microarrays, we determined the gene expression profiles of 40 cell lines as well as of primary leukemias harboring 11q23/*MLL* rearrangements, t(1;19)[*TCF3/PBX1*], t(12;21)[*ETV6/RUNX1*], t(8;21)[*RUNX1/CBFA2T1*], t(8;14)[*IGH@/MYC*], t(8;14)[*TRA@/MYC*], t(9;22)[*BCR/ABL1*], t(10;11)[*PICALM/MLLT10*], t(15;17)[*PML/RARA*], or inv(16)[*CBFB/MYH11*]. Unsupervised classification revealed that hematopoietic cell lines of diverse origin, but with the same primary genetic changes, segregated together, suggesting that pathogenetically important regulatory networks remain conserved despite numerous passages. Moreover, primary leukemias cosegregated with cell lines carrying identical genetic rearrangements, further supporting that critical regulatory pathways remain intact in hematopoietic cell lines. Transcriptional signatures correlating with clinical subtypes/primary genetic changes were identified and annotated based on their biological/molecular properties and chromosomal localization. Furthermore, the expression profile of tyrosine kinase-encoding genes was investigated, identifying several differentially expressed members, segregating with primary genetic changes, which may be targeted with tyrosine kinase inhibitors. The identified conserved signatures are likely to reflect regulatory networks of importance for the transforming abilities of the primary genetic changes and offer important pathogenetic insights as well as a number of targets for future rational drug design.
Leukemia (2005) **19,** 1042–1050. doi:10.1038/sj.leu.2403749
Published online 21 April 2005
Keywords: cDNA microarray; pediatric leukemia; AML; ALL; CML-BC

Introduction

Hematologic malignancies are characterized by balanced chromosomal abnormalities – translocations and inversions – that lead to deregulated expression of genes located in the proximity of the breakpoints or result in tumor-specific fusion genes.[1,2] In acute leukemias, the genes participating in such rearrangements often encode evolutionarily conserved transcription factors that are important also in normal hematopoiesis.[3,4] In chronic myeloproliferate disorders, the chimeric fusion proteins more often involve genes encoding proteins with tyrosine kinase activity, the prototypic example being the ABL1 tyrosine kinase that becomes constitutively activated by its fusion to BCR in chronic myeloid leukemia (CML).[5,6]

Although hematologic malignancies have been extensively studied and more than 350 recurrent balanced chromosomal abnormalities have been identified,[2] many of them providing essential biological and prognostic information, our understanding of how individual fusion genes elicit their leukemogenic properties is still quite limited. However, during recent years, genome-wide gene expression analyses, performed by microarray-based technologies, have demonstrated that hematologic malignancies can be classified based on their gene expression profiles, which have been shown to be strongly associated with specific morphologic, immunophenotypic, genetic, and prognostic parameters.[7–10] Moreover, such studies have identified new molecular subgroups and increased our knowledge about leukemia biology.

Today, the perhaps most widely used model system for studying the biological consequences of leukemia-associated chromosomal rearrangements is based on immortalized hematopoietic cell lines. Although cell lines are known to differ from both neoplastic and normal tissues,[11] they provide powerful tools for investigating basic and applied aspects of leukemia cell biology. An illustrative example is provided by Philadelphia (Ph) chromosome-positive cell lines, carrying the t(9;22)(q34;q11) and the *BCR/ABL1* chimera, established from CML patients in blast transformation (BC). These cell lines remain sensitive to the ABL1 tyrosine kinase inhibitor imatinib mesylate (Gleevec)[12] despite their numerous passages *in vitro*, indicating that the primary genetic change and its impact on downstream transcriptional networks still are indispensable for their growth.

Surprisingly, however, only a few studies have addressed systematically whether hematopoietic cell lines with characteristic cytogenetic aberrations display conserved gene expression signatures despite their diverse origin and numerous passages *in vitro*. Recently, Fine *et al*[13] studied the gene expression pattern of a set of hematopoietic cell lines and B-lineage acute lymphoblastic leukemias (ALL), and showed that cell lines and leukemias segregate based on their genetic aberration upon cluster analyses and that they have a shared gene expression signature. In the present study, we confirm and further extend these observations by investigating a larger number of cell lines, derived from all hematopoietic lineages, and by analyzing a larger number of specific genetic aberrations.

Patients, materials and methods

Cell lines and patient material

A total of 40 hematologic cell lines were investigated in this study and their clinical subtype, genetic features, as well as their providers are summarized in Table 1. The cell lines were cultured according to the instructions provided by the cell repositories.

Bone marrow samples from 11 children with acute myeloid leukemia (AML) or ALL were obtained at the time of diagnosis.

Correspondence: Dr A Andersson, Department of Clinical Genetics, University Hospital, SE-221 85 Lund, Sweden; Fax: +46 46 131061; E-mail: anna.andersson@klingen.lu.se
Received 28 October 2004; accepted 24 February 2005; Published online 21 April 2005

Table 1 The 40 hematologic cell lines included in the study, their subtype and genetic rearrangements

Cell line	*Subtype*	*Genetic rearrangement*	*Fusion gene*
MV4:11[a]	AML	t(4;11)(q21;q23)	*MLL/MLLT2(AF4)*
ML-2[a]	AML	t(6;11)(q27;q23)	*MLL/MLLT4(AF6)*
Mono-Mac 6[a]	AML	t(9;11)(p22;q23)	*MLL/MLLT3(AF9)*
THP-1[a]	AML	t(9;11)(p22;q23)	*MLL/MLLT3(AF9)*
Kasumi-1[a]	AML	t(8;21)(q22;q22)	*RUNX1(AML1)/CBFA2T1(ETO)*
Skno-1[b]	AML	t(8;21)(q22;q22)	*RUNX1(AML1)/CBFA2T1(ETO)*
ME-1[a]	AML	Inv(16)(p13q22)	*CBFB/MYH11*
OCI-AML2[a]	AML	Hyperdiploid	*MLL*[c]
SIG-M5[a]	AML	Hypertriploid/hypotetraploid	
KG-1[a]	AML	Hypodiploid	
Eol-1[a]	AML	dup(11)(q23q23), del(4)(q12q12)	*MLL* PTD, *FIP1L1/PDGFRA*
CMK[a]	AML	Hypotetraploid	
CTV-1[a]	AML	t(12;16)(q24;q11)	
HL-60[a]	AML	Hypotetraploid	
U937[a]	AML	t(10;11)(p12;q14)	*PICALM/MLLT10(AF10)*
NB4[a]	APL	t(15;17)(q22;q12)	*PML/RARA*
RS4:11[d]	B-lineage ALL	t(4;11)(q21;q23)	*MLL/MLLT2(AF4)*
Kopn-1[b]	B-lineage ALL	t(11;19)(q23;p13)	*MLL/MLLT1(ENL)*
Kocl-33[b]	B-lineage ALL	t(11;19)(q23;p13)	*MLL/MLLT1(ENL)*
Mhh-Call 2[a]	B-lineage ALL	Hyperdiploid	
Sup-B15[a]	B-lineage ALL	t(9;22)(q34;q11)	P190 *BCR/ABL1*
SD-1[a]	B-lineage ALL	t(9;22)(q34;q11)	P190 *BCR/ABL1*
Mhh-Call 3[a]	B-lineage ALL	t(1;19)(q23;p13)	*TCF3(E2A)/PBX1*
697 (EU-3)[a]	B-lineage ALL	t(1;19)(q23;p13)	*TCF3(E2A)/PBX1*
REH[a]	B-lineage ALL	t(12;21)(q13;q22)	*ETV6(TEL)/RUNX1(AML1)*
Nalm-6[a]	B-lineage ALL	t(5;12)(q33;p13)	
MN-60[a]	B-lineage ALL	t(8;14)(q24;q32)	*IGH@/MYC*
Tanoue[a]	B-lineage ALL	t(8;14)(q24;q32)	*IGH@/MYC*
Mutz-5[a]	B-lineage ALL	t(12;13)(p12;q13–14)	
Mhh-Call 4[a]	B-lineage ALL	Hypodiploid	
Nalm-1[a]	CML BC	t(9;22)(q34;q11)	P210 *BCR/ABL1*
K562[a]	CML BC	t(9;22)(q34;q11)	P210 *BCR/ABL1*
EM-3[a]	CML BC	t(9;22)(q34;q11)	P210 *BCR/ABL1*
Lama-84[a]	CML BC	t(9;22)(q34;q11)	P210 *BCR/ABL1*
Meg-01[a]	CML BC	t(9;22)(q34;q11)	P210 *BCR/ABL1*
JK-1[a]	CML BC	t(9;22)(q34;q11)	P210 *BCR/ABL1*
KU-812[a]	CML BC	t(9;22)(q34;q11)	P210 *BCR/ABL1*
Molt-16[a]	T-cell ALL	t(8;14)(q24;q11)	*TRA@/MYC*
Karpas-45[a]	T-cell ALL	t(X;11)(q13;q23)	*MLL/MLLT7(AFX)*
PF-382[a]	T-cell ALL	Hypodiploid	

AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; APL, acute promyelocytic leukemia; CML, chronic myeloid leukemia; BC, blast crisis; PTD, partial tandem duplication.
[a]Obtained from the German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.
[b]The Skno-1 cell line and cell pellets from Kocl-33 and Kopn-1 were kindly provided by Dr K Chihara, Third Department of Internal Medicine, Kobe University School of Medicine, Japan, and Dr K Sugita, Department of Pediatrics, Yamanashi Medical University, Japan, respectively.
[c]*MLL* rearrangement as determined by Southern blot analysis in the present study.
[d]Obtained from the American Type Culture Collection, Manassas, VA.

The collection of bone marrow specimens was approved by the Research Ethics Committee of Lund University, Sweden. Cells were frozen in TRIzol (Invitrogen, Carlsbad, CA, USA) and all cases were analyzed cytogenetically at the Department of Clinical Genetics, Lund, Sweden. In addition, the ALL cases were screened molecularly for *MLL* rearrangements, *BCR/ABL1*, *ETV6/RUNX1*, and *TCF3/PBX1* fusions. The clinical and genetic features are listed in Table 2.

Microarray design

cDNA microarray slides were obtained from the Swegene DNA Microarray Resource Center at Lund University, Sweden (http://swegene.onk.lu.se). cDNA from cell lines and primary leukemias were hybridized to 32 and 27K microarray slides, respectively, of different design, but both containing 25 648 clones representing 17 492 Unigene clusters and 10 949 Locus link entries. Gene names are according to Unigene build 165.

RNA extraction, probe labeling, and hybridization conditions

Cells from the cell lines were collected 24 h after medium exchange, and total RNA was extracted using TRIzol reagent (Invitrogen) and RNeasy (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions. As a reference, the Universal Human Reference RNA (Stratagene, La Jolla, CA, USA) was used. cDNA synthesis, labeling of the poly(A) RNA, and hybridizations were performed using standard protocols (Supplementary data). Slides were scanned in the G2565AA Agilent DNA Microarray Scanner (Agilent Technologies, Palo Alto, CA, USA) and analyzed with the GenePix4.0 software (GenePix, Foster City, CA, USA).

Table 2 Patients characteristics of the 11 primary leukemias

Case no.	Age (y)	Sex	Subtype	Partial karyotype[a]	Genetic rearrangement
1	6 *mo*	F	B-lineage ALL	t(11;19)(q23;p13)[b]	*MLL/MLLT1(ENL)*
2	12 *mo*	M	B-lineage ALL	t(4;11)(q21;q23)[c]	*MLL/MLLT2(AF4)*
3	3.8	M	B-lineage ALL	der(19)t(1;19)(q23;p13)[d]	*TCF3(E2A)/PBX1*
4	8.9	M	B-lineage ALL	t(9;22)(q34;q11)[e]	P190 *BCR/ABL1*
5	3.5	F	B-lineage ALL	t(12;21)(p13;q22)[f]	*ETV6(TEL)/RUNX1(AML1)*
6	3.6	F	B-lineage ALL	t(12;21)(p13;q22)[f]	*ETV6(TEL)/RUNX1(AML1)*
7	9.6	M	T-cell ALL	46,XY	
8	9 *mo*	F	AML M5	t(9;11)(p22;q23)[b]	*MLL/MLLT3(AF9)*
9	1.7	F	AML M5	del(11)(q23q23)[c]	*MLL*
10	10	F	AML M2	t(8;21)(q22;q22)	*RUNX1(AML1)/CBFA2T1(ETO)*
11	9	M	AML M2	t(8;21)(q22;q22)	*RUNX1(AML1)/CBFA2T1(ETO)*

Abbreviations are explained in Table 1; y, years; mo, months; F, female; and M, male.
[a]The partial karyoptype is based on the cytogenetic analysis.
[b]Also verified by Southern blot, RT-PCR, and FISH.
[c]Also verified by Southern blot.
[d]Also verified by RT-PCR.
[e]The karyotype is based on RT-PCR and FISH.
[f]Also verified by FISH.

Patient RNA was extracted as above, and one round of linear amplification was performed using the RiboAmp-RNA Amplification kit (Arcturus, Mountain View, CA, USA) as described by the manufacturer. The quality of total and amplified RNA (aRNA) was assessed using the Agilent 2100 BioAnalyzer (Agilent). As a reference, aRNA from the Universal Human Reference (Stratagene) was used. Labeling was performed as above. For prehybridization, hybridization, and posthybridization washes, the Pronto Universal Microarray Reagent System (Corning, Acton, MS, USA) was used, as described by the manufacturer.

Microarray data acquisition

The data matrix was uploaded to the BioArray Software Environment (BASE)[14] and normalization was performed by a BASE implementation of the intensity-dependent normalization based on a Lowess fit.[15] To correct for bad quality spots, an error model was used (Supplementary data) and a variation and presence filter was applied. To analyze the cell line and primary leukemia data sets together, which were hybridized on slides of two different designs (32 and 27 K, respectively), the cell lines and acute leukemias were mean-centered individually, and for the few reporters that occurred in duplicate, measurements were merged and the data were filtered as described (Supplementary data).

Hierarchical clustering analyses (HCA) were performed in TMEV[16] and to view the similarities between samples in three dimensions, multidimensional scaling (MDS) using geodesic distances was performed in STATISTICA 6 software (Statsoft Inc., Tulsa, OK, USA).[17] This analysis[18] has previously been shown by us to visualize efficiently the biological relationship between samples subjected to microarray analysis.[17]

Four different measurements (Pearson, ROC area, Fisher statistic, and Golub score) were used for discriminatory analysis, yielding highly consistent results. For the gene lists presented herein, the Golub score[7] was used to rank genes. A random permutation test was performed using 1000 sample label permutations with $P \leqslant 0.001$ considered significant ($P \leqslant 0.002$ if <70 reporters) (Supplemental data). The gene lists were subjected to the gene ontology (GO) software EASE,[19] and an EASE score <0.05 was considered significant. We also used the EASE tool for basic pathway analysis using the same score as above as significant. Primary data are available through www.klingen.lu.se/E/projects.htm.

Verification of gene expression data

To confirm the results from the cDNA microarray analyses, Northern blot analyses were performed on 10 selected genes, using standard protocols.[20] The verified genes were selected based on their rank in the discriminatory lists as well as of their biological interest (Supplementary Table S4; Figure S4). As part of the gene verification process, Western- and Southern blot, and RT-PCR analyses were performed as previously described.[21,22]

Results

Gene expression analysis segregates cell lines according to clinical and molecular subtype

To investigate if gene expression profiling of the 40 hematologic cell lines could separate B-lineage ALL, T-cell ALL, AML, and CML BC, we first applied unsupervised HCA and MDS using geodesic distances. As outlined in Figure 1a and b, a nearly perfect segregation of the cell lines according to their lineage was observed, indicating that cell lines of diverse origin maintain a conserved gene expression pattern correlating with specific clinical subtypes. Only three cell lines (EM-3, CMK, and CTV-1) failed to segregate according to their subtype. EM-3, a Ph-positive CML BC cell line, segregated with the AMLs; CTV-1, a monocytic cell line, segregated with the T-cell ALLs; and CMK, a megakaryocytic cell line, segregated with the CML BCs. Using Western blot analysis, we showed that, in contrast to the other investigated *BCR/ABL1*-positive CML BCs, EM-3 did not display phosphorylation of STAT5, which may indicate that the original molecular signature correlating with *BCR/ABL1* expression had been lost or significantly altered in EM-3 (data not shown). CMK, segregating with the CML BC group, was further examined by RT-PCR for the presence of a *BCR/ABL1*

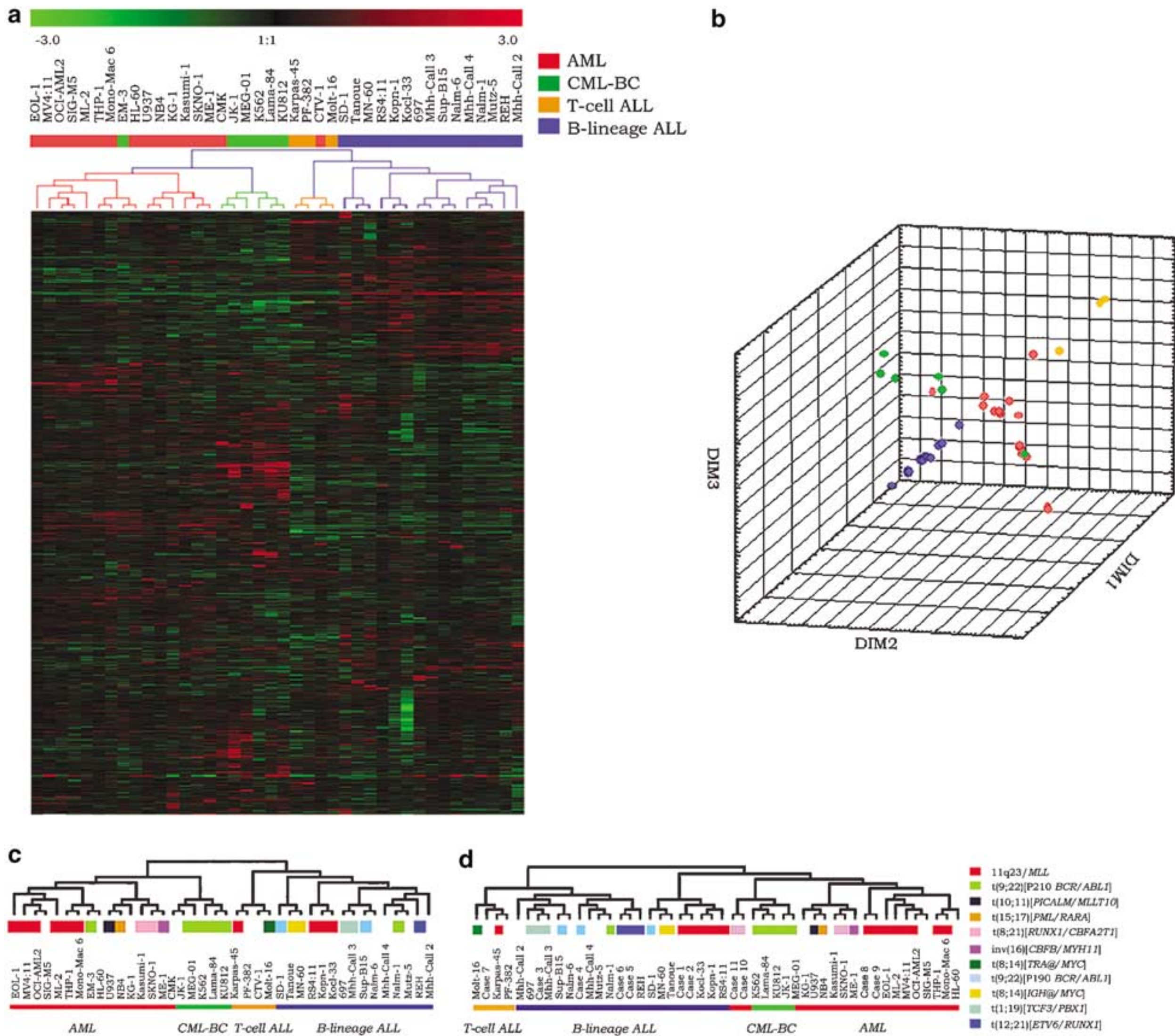

Figure 1 Hierarchical dendogram and MDS using geodesic distances of the cell lines and primary leukemias according to clinical subtype and genetic rearrangement. (a) Unsupervised HCA, using Pearson correlation and complete clustering, of the 40 cell lines, showing a nearly perfect segregation of the cell lines according to clinical subtype. The analysis is based on 3491 reporters. The color scale above each cluster indicates the log2 ratios; green represents downregulated genes and red upregulated genes. The cell lines are color-coded based on their subtype; AML (red), CML BC (green), B-lineage ALL (blue), and T-cell ALL (orange). (b) Projection of the 40 cell lines in 3D using MDS and geodesic distance, colored according to their subtype as above. The projection clearly separates the four clinical subtypes of leukemia. Moreover, the three misclassified cell lines, as determined by unsupervised HCA, mis-segregate also in this analysis. (c) Unsupervised HCA, using Pearson correlation and complete clustering of the 40 cell lines, color-coded according to their genetic abnormality; *MLL* (red), P210 *BCR/ABL1* (green), *PICALM/MLLT10(AF10)* (black), *PML/RARA* (orange), *RUNX1/CBFA2T1* (light pink), *CBFB/MYH11* (dark pink), *TRA@/MYC* (dark green), P190 *BCR/ABL1* (light blue), *IGH@/MYC* (yellow), *TCF3/PBX1* (blue/gray), and *ETV6/RUNX1* (dark blue). (d) Visualization of the 37 cell lines and 11 acute leukemias according to genetic rearrangements. Cell lines and fresh leukemic samples were mean-centered individually and unsupervised HCA was applied on the 3841 reporters. The cell lines and patient samples are color-coded based on their genetic aberration, as above. This figure clearly shows that cell lines and all primary leukemias, apart from the two *RUNX1/CBFA2T1*-positive cases and a Ph-positive ALL, segregate based on their genetic rearrangement.

rearrangement, but no such fusion transcript was detected (data not shown). To rule out contamination by other cell lines, the three misclassified cell lines were analyzed cytogenetically, confirming the karyotypes provided by the cell repository (data not shown). For the above reason, these cell lines were removed from subsequent discriminatory analysis.

Using unsupervised HCA, cell lines with identical or closely related primary genetic aberrations were also found to segregate together (Figure 1c). The cosegregation of cell lines with the same primary genetic changes indicates a shared molecular signature among cell lines with identical genetic aberrations. The two cell lines harboring P190 *BCR/ABL1* did not cluster together among the B-lineage ALLs. Moreover, three cell lines (OCI-AML2, SIG-M5, and HL-60) segregated with the five AMLs characterized by *MLL* abnormalities, but none of these have previously been reported to carry such rearrangements. Upon Southern blot analysis, we were, however, able to show that OCI-AML2 harbored a rearranged *MLL* gene (Supplementary Figure S1); the two other cell lines displayed germline configuration of *MLL*.

Primary acute leukemias and cell lines with the same primary genetic change cosegregate

Having shown that cell lines of diverse origin, but with the same primary genetic change, segregate together using unsupervised algorithms, we next investigated whether the maintained gene expression signatures identified in cell lines with identical primary genetic abnormalities also could be found in primary leukemias harboring the same aberrations. Accordingly, 11 acute leukemias with selected genetic aberrations were added to the data set (Table 2). Since only 27K arrays were available when analyzing the acute leukemias, the obtained data sets were mean-centered individually (Supplementary data). Following mean-centering, unsupervised HCA was applied and eight of the 11 acute leukemias cosegregated with the respective cell lines according to their genetic aberration (Figure 1d). This strongly suggests that the original gene expression pattern in primary leukemia samples is preserved in immortalized cell lines. Two acute leukemias with *RUNX1/CBFA2T1* failed to segregate with the two *RUNX1/CBFA2T1*-positive cell lines. Also, the Ph-positive ALL failed to cluster with the Ph-positive ALL cell lines. We next applied the discriminatory gene lists obtained from the cell lines correlating with leukemia subtype and genetic aberrations (see below) on the combined data set, resulting in a successful clustering according to clinical and genetic subtype, indicating that the majority of the top genes discriminating the clinical subtypes in cell lines were preserved in the primary acute leukemias investigated (data not shown).

Identification of genes discriminating the clinical subtypes

To identify genes that significantly correlate with B-lineage ALL, T-cell ALL, AML, and CML BC, gene discriminatory analysis was performed. Discriminating genes were subsequently categorized based on their GO annotations to facilitate biological and functional interpretations. We also performed pathway analysis using the EASE tool to facilitate functional interpretation of the gene expression data.

Discriminatory analysis of AML and ALL ($P\leqslant 0.001$), AML, and CML BC ($P\leqslant 0.001$), and B-lineage ALL and T-cell ALL ($P\leqslant 0.002$) revealed 367, 214, and 88 genes, respectively (Supplementary Figure S2; Supplementary Tables S5–7). Each gene list was subjected to GO and pathway analyses (Supplementary Tables S1 and S2). A selected set of the GO categories as well as the top 10 ranked genes are shown in Supplementary Table S3.

Comparing AML and ALL, we identified *CEBPA*, a myeloid transcription factor of importance for maturation along the neutrophilic lineage, and *MEIS2*, a homeobox-containing gene of the TALE-family, as upregulated in AML. In ALL, *CBFB* and *LEF1*, both involved in the *CBF*-transcription factor complex, were highly expressed. Using pathway analysis, five pathways were found, including proteasome degradation. Genes discriminating AML and CML BC included the transcription factor *JUND* and the serin/threonine kinase *PIM1*, upregulated in CML BC. Three pathways were enriched, including heme biosynthesis. In the B-lineage and T-cell ALL comparison, the transcription factor and lineage regulator *GATA3* and the signal transducer *CD3G* were expressed in T-cell ALL, the latter gene involved in the T-cell receptor complex. The B-cell regulator *PAX5* was upregulated in B-lineage ALL. Pathway analysis revealed five pathways, including activation of the T-cell receptor.

Genes discriminating the molecular subgroups

Having identified genes discriminating the various clinical subtypes, we next identified genes discriminating cell lines with identical genetic aberrations ($P\leqslant 0.002$): *RUNX1/CBFA2T1,* 35 genes; CBF-leukemias, 62; *TCF3/PBX1*, 10; and *IGH@/MYC,* 36 (Supplementary Figure. S3a–d; Supplementary Tables S8–11). Three methods were used to analyze cell lines with 11q23/*MLL* rearrangements ($P\leqslant 0.002$), the first compared B-lineage *MLL* and B-lineage ALL, 19 genes; the second AML-positive *MLL* and AML, 18 genes; and the third cell lines with *MLL* abnormalities and all other cell lines, 40 genes (Figure 2a–c; Supplementary Tables S12–14). All gene lists were subjected to GO analyses (Supplementary Table S1).

Two cell lines harbored *RUNX1/CBFA2T1* and among the upregulated genes *CBFA2T1,* were found, but no enriched GO categories were identified. Given the fact that the two cell lines with *RUNX1/CBFA2T1* and the cell line with *CBFB/MYH11* cosegregated upon unsupervised HCA (Figure 1c) and that both fusion genes converge on the *CBF*-transcription factor complex,[3] we performed discriminatory analysis of AMLs with (herein designated CBF-leukemias) or without these two genetic rearrangements. The *PIK3R1*, a regulatory subunit of the *PI3*-kinase pathway, were upregulated and three GO categories were enriched. Comparing *TCF3/PBX1* with other B-lineage ALLs revealed two enriched categories: signal transducer activity and transcription factor activity. The first category included genes such as *MAP3K1* and the latter *PBX1*. In cell lines with *IGH@/MYC*, the signal transducer *BMP7* was upregulated and the transcription factor *DP1* was downregulated. Two categories were enriched. In the B-lineage ALL and *MLL* comparison, two GO categories were identified, and the transcription factor and homeobox-containing gene *PBX3* was upregulated. We also compared AMLs with or without *MLL* rearrangements, identifying *PBX3*; no enriched category was found. Finally, when cell lines with *MLL* abnormalities were compared to all other cell lines, we again identified *PBX3*, suggesting an important role of this gene in *MLL*-positive leukemias. GO analysis revealed two enriched categories.

Correlation of deregulated gene expression with their chromosomal localization and expression pattern of tyrosine kinase-encoding genes

In order to investigate if the deregulated genes identified by discriminatory analysis were associated with a certain chromosomal region, the EASE tool was used. The calculations were performed for each of the discriminating gene lists, two of which resulted in a significant correlation. Comparing AML and ALL, 21 of the 367 deregulated genes were located on chromosome arm 17p (score 9.21×10^{-12}), all but one downregulated in AMLs. Moreover, when in AMLs with rearrangements involving *CBF*, 15 of the 39 downregulated genes in cell lines with *CBF*-rearrangements were located on chromosome arm 2q (score 3.98×10^{-11}).

Given the recent development of tyrosine-kinase inhibitory drugs,[23] we investigated selectively the expression signature of protein tyrosine kinase-encoding genes across all cell lines. Hence, all genes on our array annotated with protein-tyrosine kinase activity (GO: 0004713, $n=170$) were extracted. In total, 41 genes of this family remained after stringent quality filtering. To investigate the relative expression of these genes, the data set was median-centered. In order to view the tyrosine kinase-encoding genes with respect to the segregation pattern observed

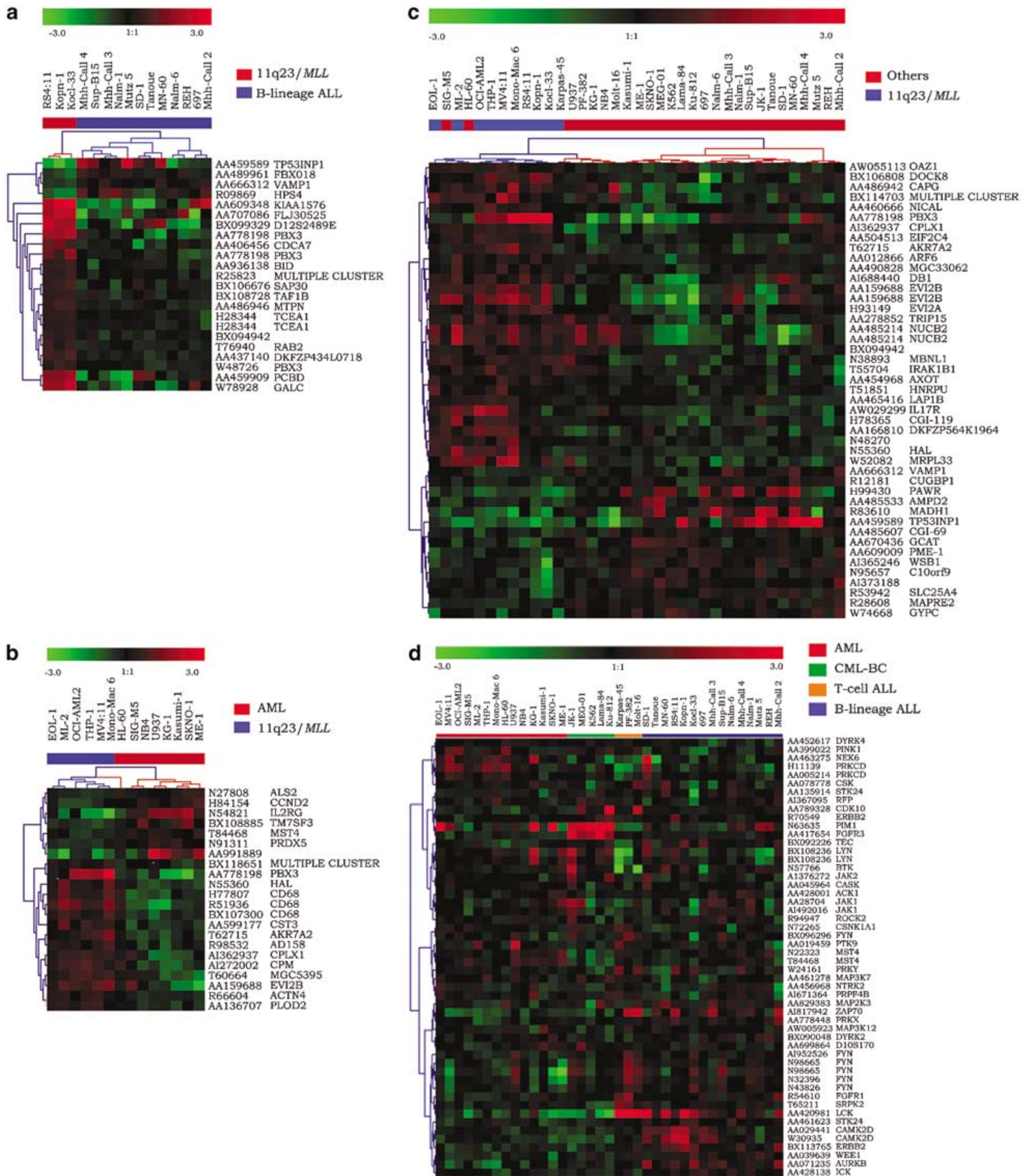

Figure 2 Genes discriminating cell lines with *MLL* rearrangements and visualization of the expression signature of tyrosine kinase-encoding genes. Discriminatory analysis ($P \leqslant 0.002$) of B-lineage ALLs with or without *MLL*, AMLs with or without *MLL*, and cell lines harboring *MLL* compared to the entire data set, was performed, followed by HCA. The color scale above each cluster indicates the log2 ratios; green represents downregulated genes and red upregulated genes. (a) Cluster analysis using Euclidean distance and complete linkage of the 21 reporters discriminating B-lineage ALLs (blue) with or without (red) *MLL*. (b) HCA using Pearson correlation and complete linkage of the 24 reporters discriminating AMLs with (red) or without *MLL* (blue). (c) Clustering using Pearson correlation and complete linkage of the 48 reporters discriminating cell lines harboring *MLL* rearrangements (blue) from all other lineages (red). (d) Visualization of the expression of the 41 tyrosine kinase-encoding genes. Clustering was only applied on the genes; the order of the cell lines is according to the unsupervised HCA in Figure 1a. The cell lines are colore coded as follows: AML (red), CML BC (green), B-lineage ALL (blue), and T-cell ALL (orange).

using unsupervised methods, clustering was only applied on the genes with the order of the cell lines obtained by unsupervised HCA being maintained (Figure 2d). As seen in Figure 2d, *PIM1* and *FGFR3* were expressed in CML BC. *LCK* was upregulated in the majority of the B-lineage ALLs and particularly in T-cell ALL. In a large subset of the B-lineage ALLs, *ZAP70* and *ERBB2* were

expressed. Another cluster of genes, upregulated in T-cell ALL, included *FYN* and *ZAP70*. In AML, a cluster of *NEK6*, *PRKCD*, and *PINK1* was seen. In particular, AMLs with *MLL* abnormalities showed a high expression of these genes; the other AMLs displayed expression of *MST4*.

Discussion

Immortalized hematopoietic cell lines have in recent years been instrumental in the cloning, characterization, and functional analyses of a large number of leukemia-associated fusion genes.[11] These cell lines have also become increasingly important tools in evaluating biological effects of newly developed inhibitory drugs, capable of interfering with critical signaling pathways or with altered proteins or transcripts resulting from the primary genetic changes.[6,24] While it is generally assumed that cell lines acquire additional aberrations in culture, little is known about to what extent pathogenetically important transcriptional programs remain conserved upon establishment and passaging of individual cell lines.

By applying cDNA microarray analysis and unsupervised algorithms on 40 hematopoietic cell lines with selected primary genetic rearrangements, we here show that the cell lines segregate based on their clinical subtype and, more importantly, with few exceptions, according to their primary genetic change. Thus, cell lines harboring 11q23/*MLL* rearrangements, *TCF3/PBX1*, *IGH@/MYC*, *RUNX1/CBFA2T1*, P210 *BCR/ABL1*, *ETV6/RUNX1*, or *CBFB/MYH11*, were shown to segregate based on their primary genetic abnormality. This cosegregation suggests that critical transcriptional programs remain conserved in hematopoietic cell lines despite their diverse origin and numerous passages *in vitro*. We obtained further support for this by analyzing primary ALL and AML samples with 11q23/*MLL* rearrangements, *TCF3/PBX1*, *RUNX1/CBFA2T1*, P190 *BCR/ABL1*, or *ETV6/RUNX1*, showing cosegregation of most of these samples with cell lines harboring identical genetic rearrangements. In a similar approach, Fine *et al*[13] recently determined the gene expression pattern in 19 cell lines, mainly of B-lineage origin, with various 11q23/*MLL* rearrangements, *TCF3/PBX1*, P190 and P210 *BCR/ABL1*, *ETV6/RUNX1*, or *E2A/HLF*, and in primary acute leukemia samples with *MLL/MLLT2(AF4)*, P190 *BCR/ABL1*, or *ETV6/RUNX1*. They also found that cell lines and primary leukemias segregated primarily according to their genetic changes.[13] In contrast to Fine *et al*, we found that cell lines with *MLL* rearrangements segregated primarily according to their morphologic subtype, although closely together within each lineage. Two AML cases with *RUNX1/CBFA2T1* failed to segregate with the two cell lines harboring this abnormality, indicating that such cell lines may have acquired additional genetic changes, altering their transcriptional profiles, when adapting to *in vitro* conditions.

Surprisingly, the myeloid cell lines HL-60, SIG-M5, and OCI-AML2 segregated with cell lines harboring *MLL* rearrangements. Although this, in principle, could be a reflection of their lineage specificity, we performed Southern blot analysis to investigate if the *MLL* gene was rearranged in these cell lines. Interestingly, one cell line, OCI-AML2, established from a patient with AML M4, had a rearranged *MLL* gene, not previously reported. The SIG-M5 cell line was established from a patient with AML M5a, and given the high frequency of *MLL* abnormalities in this FAB type,[25] it cannot be excluded that this cell line may carry an atypical *MLL* rearrangement. HL-60 on the other hand, established from a patient with AML M2, has been reported to carry a sideline with del(11)(q23), possibly resulting in loss of heterozygosity of *MLL*. The similarity in gene expression pattern of the latter two cell lines with *MLL*-positive cell lines indicates that they have activated downstream genes similar to *MLL* target genes, warranting further investigations.

Supervised analysis of cell lines of different lineages generated a list of genes differentially expressed among each subtype. The majority of these genes reflected lineage characteristics, but genes of possible pathogenetic importance were also identified. For example, discriminatory analysis of P210 *BCR/ABL1*-positive CML BC and AML identified *JUND* and *PIM1,* both previously shown to become activated upon BCR/ABL1 expression.[21,26] Interestingly, GO and pathway analyses also revealed upregulation of genes involved in heme biosynthesis. Although the enrichment of erythroid genes could be a reflection of the stem cell involvement in CML BC, it could also reflect that BCR/ABL1 indeed activates erythroid pathways as was recently demonstrated *in vitro* by retroviral transduction of human CD34 + cord blood cells in which BCR/ABL1 causes erythroid lineage switching.[27]

By comparing cell lines harboring specific genetic aberrations, we identified a number of genes correlating with genetic subtypes, providing biological insights into the downstream genes deregulated by the primary genetic change. Three different discriminatory analyses were performed to identify genes correlating with *MLL* and all three independently identified the homeobox transcription factor *PBX3* as upregulated, indicating that expression of *PBX3* may be important for *MLL*-mediated transformation. Interestingly, *PBX3* was recently shown to become upregulated after experimentally induced expression of *MLL/MLLT1(ENL)* in mouse hematopoietic cells.[28] In our data set, two cell lines harbored *RUNX1/CBFA2T1* and *CBFA2T1* was upregulated, as has previously also been identified in microarray analyses of primary AML with this fusion gene.[29] Analysis of *TCF3/PBX1*-positive cell lines revealed upregulation of *PBX1*, as has also been seen in primary ALLs carrying this fusion gene.[8] Two cell lines harbored *IGH@/MYC*, and both showed downregulation of *DP1*, possibly mediating an antiapoptotic effect by loss of E2F1-mediated apoptosis.[30]

We also determined the chromosomal location of the deregulated genes in cell lines of different clinical subtypes or with different genetic abnormalities. A significant correlation was obtained comparing AML and ALL; 20 of the 21 genes located on chromosome arm 17p were downregulated in the AML cell lines. Loss of chromosome 17 material, in particular of 17p, is one of the most frequent cytogenetic alterations seen in AML[31] and our results suggest that large chromosomal deletions may result in downregulation of several genes located in the lost regions. Interestingly, in AMLs with rearrangements involving the *CBF*-transcription factor complex, 15 genes, located on the distal part of 2q33–37, were downregulated. Loss of 2q is not a common cytogenetic abnormality in acute leukemias, but the present finding may suggest that cryptic deletions at this locus are present in CBF-leukemias or, alternatively, that deregulation of the *CBF*-transcription factor complex may lead to repression of genes on selected chromosomal loci, certainly warranting future investigations.

It is well known that genes encoding protein tyrosine kinases frequently are implicated in human cancer,[32] and given the development of selective tyrosine kinase inhibitors,[23] such genes have been targeted and extensively studied *in vitro* as well as *in vivo*. In order to investigate the expression of tyrosine kinase-encoding genes, we selectively analyzed these genes in our data set. Interestingly, *FGFR3* was upregulated in CML BC cell lines. Recently, the expression of *FGFR3* was found to

correlate with disease relapse in a series of transplanted CML patients[33] and to be expressed at increased levels in CD34$^+$ CML cells.[34] Hence, the high expression of *FGFR3* in CML BC cell lines suggests that targeting of FGFR3 using available inhibitors may provide an important additional treatment modality of *BCR/ABL1*-positive leukemias.

Apart from demonstrating that acute leukemias and immortalized hematologic cell lines of similar clinical subtypes share a molecular signature, we also show that cell lines and primary leukemic samples harboring identical or closely related primary genetic aberrations cosegregate. Moreover, we have identified a large number of differentially expressed genes that correlate with the most frequently observed fusion genes in leukemia and provide examples of genes, for example, *FGFR3*, which may be evaluated as novel targets for treatment. Furthermore, the present study also provides an extensive gene expression data set of hematopoietic cell lines, offering an important source for further explorations to obtain additional insights into the molecular consequences of recurrent genetic changes in leukemia.

Acknowledgements

We thank the Swegene DNA microarray resource center in Lund, supported by the Knut and Alice Wallenberg Foundation through the Swegene consortium, Eva Pålsson for cytogenetic analysis, and Srinivas Veerla for useful computational program construction. PE was supported by the Swedish Foundation for Strategic Research through the Lund Center for Stem Cell Biology and Cell Therapy. JN was supported by the Knowledge Foundation and AstraZeneca R&D Lund. This study was supported by the Swedish Cancer Society, The Swedish Children's Cancer Foundation, the Medical Faculty of Lund University, and the IngaBritt and Arne Lundberg foundation.

Supplementary Information

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu).

References

1 Rowley JD. The critical role of chromosome translocations in human leukemias. *Annu Rev Genet* 1998; **32**: 495–519.
2 Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat Genet* 2004; **36**: 331–334.
3 Downing JR. AML1/CBFbeta transcription complex: its role in normal hematopoiesis and leukemia. *Leukemia* 2001; **15**: 664–665.
4 Scandura JM, Boccuni P, Cammenga J, Nimer SD. Transcription factor fusions in acute leukemia: variations on a theme. *Oncogene* 2002; **21**: 3422–3444.
5 Cross NC, Reiter A. Tyrosine kinase fusion genes in chronic myeloproliferative diseases. *Leukemia* 2002; **16**: 1207–1212.
6 Deininger MW, Goldman JM, Melo JV. The molecular biology of chronic myeloid leukemia. *Blood* 2000; **96**: 3343–3356.
7 Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286**: 531–537.
8 Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R *et al.* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002; **1**: 133–143.
9 Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004; **350**: 1617–1628.
10 Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004; **350**: 1605–1616.
11 Drexler HG, Matsuo AY, MacLeod RA. Continuous hematopoietic cell lines as model systems for leukemia-lymphoma research. *Leuk Res* 2000; **24**: 881–911.
12 Druker BJ, Lydon NB. Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest* 2000; **105**: 3–7.
13 Fine BM, Stanulla M, Schrappe M, Ho M, Viehmann S, Harbott J *et al.* Gene expression patterns associated with recurrent chromosomal translocations in acute lymphoblastic leukemia. *Blood* 2004; **103**: 1043–1049.
14 Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg A, Peterson C. BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. *Genome Biol* 2002; **3**:SOFTWARE0003.
15 Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002; **30**: e15.
16 Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003; **34**: 374–378.
17 Nilsson J, Fioretos T, Höglund M, Fontes M. Approximate geodesic distances reveal biologically relevant structures in microarray data. *Bioinformatics* 2004; **20**: 874–880.
18 Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000; **290**: 2319–2323.
19 Hosack DA, Dennis Jr G, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003; **4**: R70.
20 Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: a Laboratory Manual*, 2nd edn. NY: Cold Spring Harbor, 1989.
21 Håkansson P, Segal D, Lassen C, Gullberg U, Morse H, Fioretos T *et al.* Identification of genes differentially regulated by the P210 BCR/ABL1 fusion oncogene using cDNA microarrays. *Exp Hematol* 2004; **32**: 476–482.
22 Fioretos T, Nilsson PG, Aman P, Heim S, Kristoffersson U, Malm C *et al.* Clinical impact of breakpoint position within M-bcr in chronic myeloid leukemia. *Leukemia* 1993; **7**: 1225–1231.
23 Gschwind A, Fischer OM, Ullrich A. Timeline: the discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat Rev Cancer* 2004; **4**: 361–370.
24 Stirewalt DL, Meshinchi S, Radich JP. Molecular targets in acute myelogenous leukemia. *Blood Rev* 2003; **17**: 15–23.
25 Schoch C, Schnittger S, Klaus M, Kern W, Hiddemann W, Haferlach T. AML with 11q23/MLL abnormalities as defined by the WHO classification: incidence, partner chromosomes, FAB subtype, age distribution, and prognostic impact in an unselected series of 1897 cytogenetically analyzed AML cases. *Blood* 2003; **102**: 2395–2402.
26 Nosaka T, Kitamura T. Pim-1 expression is sufficient to induce cytokine independence in murine hematopoietic cells, but is dispensable for BCR-ABL-mediated transformation. *Exp Hematol* 2002; **30**: 697–702.
27 Chalandon Y, Jiang X, Hazlewood G, Loutet S, Conneally E, Eaves A *et al.* Modulation of p210(BCR-ABL) activity in transduced primary human hematopoietic cells controls lineage programming. *Blood* 2002; **99**: 3197–3204.
28 Zeisig BB, Milne T, Garcia-Cuellar MP, Schreiner S, Martin ME, Fuchs U *et al.* Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol Cell Biol* 2004; **24**: 617–628.
29 Schoch C, Kohlmann A, Schnittger S, Brors B, Dugas M, Mergenthaler S *et al.* Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc Natl Acad Sci USA* 2002; **99**: 10008–10013.

30 Hitchens MR, Robbins PD. The role of the transcription factor DP in apoptosis. *Apoptosis* 2003; **8**: 461–468.
31 Mitelman F, Johansson B, Mertens F. Mitelman database of chromosome aberrations in cancer. http://cgap.nci.nih.gov/Chromosomes/Mitelman.
32 Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R *et al.* A census of human cancer genes. *Nat Rev Cancer* 2004; **4**: 177–183.
33 Dvorakova D, Krejci P, Mayer J, Fajkus J, Hampl A, Dvorak P. Changes in the expression of FGFR3 in patients with chronic myeloid leukaemia receiving transplants of allogeneic peripheral blood stem cells. *Br J Haematol* 2001; **113**: 832–835.
34 Dvorak P, Dvorakova D, Doubek M, Faitova J, Pacholikova J, Hampl A *et al.* Increased expression of fibroblast growth factor receptor 3 in CD34+ BCR-ABL+ cells from patients with chronic myeloid leukemia. *Leukemia* 2003; **17**: 2418–2425.