Cloning of genes involved in chromosomal translocations by high-resolution single nucleotide polymorphism genomic microarray

Norihiko Kawamata[a,b,c], Seishi Ogawa[b,d], Martin Zimmermann[b,e], Birte Niebuhr[f], Carol Stocking[f], Masashi Sanada[d], Kari Hemminki[g], Go Yamatomo[d], Yasuhito Nannya[d], Rolf Koehler[h], Thomas Flohr[h], Carl W. Miller[a], Jochen Harbott[i], Wolf-Dieter Ludwig[j], Martin Stanulla[e], Martin Schrappe[k], Claus R. Bartram[h,l], and H. Phillip Koeffler[a,l]

[a]Hematology/Oncology, Cedars-Sinai Medical Center/UCLA School of Medicine, Los Angeles, CA 90048; [d]Regeneration Medicine of Hematopoiesis, School of Medicine, University of Tokyo, Tokyo 113-8655, Japan; [e]Department of Pediatric Hematology and Oncology, Children's Hospital, Hannover Medical School, 30625 Hannover, Germany; [f]Molecular Pathology Heinrich-Pette-Institute, 20251 Hamburg, Germany; [g]Division of Molecular Genetic Epidemiology, German Cancer Research Center, 69120 Heidelberg, Germany; [h]Institute of Human Genetics, University of Heidelberg, 69117 Heidelberg, Germany; [i]Department of Hematology and Oncology, Center for Pediatrics, 35390 Giessen, Germany; [j]Department of Hematology, Oncology and Tumor Immunology, Robert-Rössle-Clinic at the HELIOS-Clinic Berlin-Buch, Charité, 13125 Berlin, Germany; and [k]Department of Pediatrics, University of Kiel, 69117 Kiel, Germany

Edited by Joe W. Gray, Lawrence Berkeley National Laboratory, Berkeley, CA and accepted by the Editorial Board June 13, 2008 (received for review November 21, 2007)

High-resolution single nucleotide polymorphism genomic microarray (SNP-chip) is a useful tool to define gene dosage levels over the whole genome, allowing precise detection of deletions and duplications/amplifications of chromosomes in cancer cells. We found that this new technology can also identify breakpoints of chromosomes involved in unbalanced translocations, leading to identification of fusion genes. Using this technique, we found that the *PAX5* gene was rearranged to a variety of partner genes including *ETV6*, *FOXP1*, *AUTS2*, and *C20orf112* in pediatric acute lymphoblastic leukemia (ALL). The 3′ end of the *PAX5* gene was replaced by the partner gene. The PAX5 fusion products bound to PAX5 recognition sequences as strongly as wild-type PAX5 and suppressed its transcriptional activity in a dominant-negative fashion. In human B cell leukemia cells, binding of wild-type PAX5 to a regulatory region of *BLK*, one of the direct downstream target genes of PAX5, was diminished by expression of the PAX5-fusion protein, leading to repression of *BLK*. Expression of PAX5-fusion genes in murine bone marrow cells blocked development of mature B cells. PAX5-fusion proteins may contribute to leukemogenesis by blocking differentiation of hematopoietic cells into mature B cells. SNP-chip is a powerful tool to identify fusion genes in human cancers.

chromatin immunoprecipitation | dominant negative | fusion gene | PAX5 | SNP-chip

Pediatric acute lymphoblastic leukemia (ALL) is the most common malignant disease in children (1–3). It is a genetic abnormality resulting from accumulation of mutations in tumor suppressor genes and oncogenes (1–3). Fusion genes including *ETV6*/*RUNX1* and *E2A*/*PBX1* are frequently detected in pediatric ALL (1). Deletion of the *INK4A*/*ARF* gene (9p21) is also a common abnormality in ALL (1). However, other genetic changes remain to be elucidated in this disease.

Identification of mutated genes in ALL has evolved with improvements in technology. A very recent approach is single nucleotide polymorphism (SNP) analysis using an array based technology (4–6) that allows identification of amplifications, deletions, and allelic imbalances, such as uniparental disomy (represents doubling of the abnormal allele due to somatic recombination or duplication, and loss of the other normal allele) (7, 8). However, SNP-chip analysis is only able to detect changes of gene dosage and is unable to identify balanced translocations, which commonly occur in ALL.

Previously, we analyzed 399 pediatric ALL cases by SNP-chip analysis and found a number of genomic abnormalities, in addition to well known common alterations (9). This technique is sensitive enough to identify genes involved in start sites of deletions/duplications. Indeed, this method allowed us to identify that the *PBX1* gene was involved in start sites of duplication of 1q23 generated by der(19)t(1;19)(q23;p13) (9). Furthermore, correlation analysis of the individual genomic abnormalities suggested the presence of der(12)t(12;21)(p13;q22) and der(21)t(12;21)(p13;q22), as well as dic(9;20)(p13;q11) (9).

In this study, we found that this new technology permitted us to identify genes involved in well known unbalanced translocations including *ETV6*/*RUNX1*. Further, we found previously undetected fusion genes between *PAX5* and a number of other partner genes by using this technique.

Results

Genes Involved in Unbalanced Translocations Were Identified by SNP-Chip Analysis. Because SNP-chip analysis can only detect changes of gene dosage including deletions, duplications, and amplifications (Fig. 1*A*), this technique is unable to identify balanced translocations (Figs. 1*Aii*). However, when one of a pair of reciprocally translocated chromosomes is lost, SNP-chip analysis can detect this abnormality as partial deletions of involved chromosomes (Fig. 1*Aiii*). Similarly, when one of a pair of reciprocally translocated chromosomes becomes duplicated, SNP-chip can also detect this abnormality as partial duplication of the involved chromosomes (Fig. 1*Aiv*). Furthermore, high resolution SNP-chip analysis allows us to identify the genes involved in these unbalanced translocations.

To prove that SNP-chip analysis can detect unbalanced translocations and the genes involved in these translocations, we

Author contributions: N.K., S.O., B.N., C.S., C.R.B., and H.P.K. designed research; N.K., S.O., B.N., C.S., M. Sanada, G.Y., and Y.N. performed research; N.K., S.O., and M. Sanada contributed new reagents/analytic tools; N.K., S.O., M.Z., M. Sanada, K.H., G.Y., Y.N., R.K., T.F., C.W.M., J.H., W.-D.L., M. Stanulla, M. Schrappe, C.R.B., and H.P.K. analyzed data; and N.K., S.O., M.Z., C.S., M. Stanulla, M. Schrappe, C.R.B., and H.P.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.W.G. is a guest editor invited by the Editorial Board.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EU784145, PAX5-FOXP1; EU784146, PAX5-AUTS2; EU784147, PAX5-C20orf112 short isoform; and EU784148, PAX5-C20orf112 long isoform).

[b]N.K., S.O., and M.Z. contributed equally to this work.

[c]To whom correspondence should be addressed at: Hematology/Oncology, Cedars-Sinai Medical Institute/UCLA Geffen School of Medicine, 8700 Beverly Boulevard, Los Angeles, CA 90048. E-mail: kawamatan@cshs.org.

[l]C.R.B. and H.P.K. contributed equally to this work.

This article contains supporting information online at www.pnas.org/cgi/content/full/0711039105/DCSupplemental.

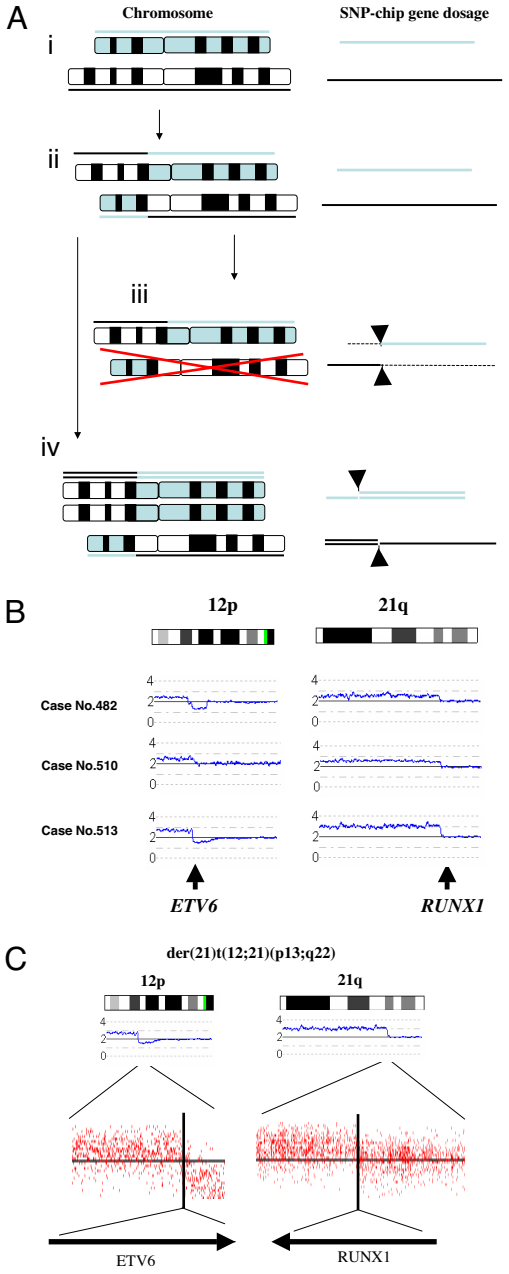

Fig. 1. SNP-chip analysis detected genes involved in unbalanced translocations. (*A*) SNP-chip analysis can identify breakpoints of translocations when one of the paired translocated chromosomes is either lost or duplicated/amplified. (*Left*) Chromosomal status. Gene dosages are indicated either above or beneath the chromosomes. (*Right*) Results of SNP-chip analysis. (*Ai*) Normal chromosomes; gene dosage is normal. (*Aii*) Reciprocal translocation; gene dosage is normal. (*Aiii*) One of the paired translocated chromosomes is lost; gene dosage is lower than normal on the left side of the upper chromosome and the right side of the lower chromosome. Arrow heads indicate the breakpoint of the translocation in each chromosome. (*Aiv*) One of the paired translocated chromosomes is duplicated; gene dosage is higher than normal on the right side of the upper chromosome and the left side of the lower chromosome. Arrow heads indicate the breakpoint of this translocation in each chromosome. (*B*) Representative cases with unbalanced translocation of der(21)t(12;21)(p13;q22). (*Left*) Start sites of duplication at 12p13 involving the *ETV6* gene. (*Right*) Start sites of duplication at 21q22 involving the *RUNX1* gene. SNP-chip data of representative cases with dup(12)(p13) and dup(21)(q22) are shown. These abnormalities were validated by FISH and/or RT-PCR (data not shown). Results of SNP-chip data were visualized by CNAG software. Lines above each chromosome show total gene dosage; level 2 indicates diploid (2N) amount of DNA, which is normal. (*C*) Magnified view of SNP-chip data. (*Upper*) Start sites of duplications at 12p13 and 21q22 are magnified. Signals of individual probe signals are shown. Vertical lines indicated the positions of start sites of duplications. (*Lower*) Genes involved in the start sites of duplications.

analyzed cases having extra copies of *ETV6*/*RUNX1* fusion genes generated by der(21)t(12;21)(p13;q22) (Fig. 1*B*), which were initially identified by FISH and/or RT-PCR (data not shown). SNP-chip was clearly able to identify this abnormality as duplications involving chromosome 12 and 21 (Fig. 1*B*). Further, the result of high-resolution (250k) SNP-chip clearly identified *ETV6* (12p13) and *RUNX1* (21q22) as the target genes involved in this unbalanced translocation (Fig. 1*C*).

PAX5 Gene Is Frequently Fused to Partner Genes. Our previous data showed the presence of dic(9;20)(p13;q11) in 11 cases of ALL (9), 5 of which had deletion 9p13.2-pter. These 5 cases had start sites of this deletion at 9p13.2 mapping to the *PAX5* gene (Fig. 2*A* and data not shown). This prompted us to reexamine all cases of B-ALL that had deletion of 9p [supporting information (SI) Table S1]. We found a total of 9 cases with similar start sites (9p13.2), mapping to the *PAX5* gene (Fig. 2*A* and data not shown). In 2 of these cases, simple abnormalities were detected by SNP-chip: case 514 had only del9p13.2-pter and del7q11.2-pter; case 458 had only del9p13.2-pter and dup3p13-pter (Table S1 and Fig. 2*A*). Three cases (536, 543, 572) had complex abnormalities including del9p13.2-pter and del20q11.21-qter, all with the *C20ORF112* gene within the start site of del20q (Table S1 and Fig. 2*A*). The other 2 cases (659, 767) had complex abnormalities that included *ETV6* on 12p13 (Table S1 and Fig. 2*A*).

Thus, we found four candidate partner genes fused to *PAX5* in seven cases by SNP-chip analysis; *ETV6* on 12p13 (two cases) (12), *C20orf112* on 20q11.1 (three cases), *AUTS2* on 7q11.1 (one case) and *FOXP1* on 3p13 (one case) (Fig. 2*A*). Because these translocations could lead to fusion transcripts between *PAX5* and different partner genes, the presence of the predicted fusion transcript was examined by RT-PCR using the mapping information from the SNP-chip data. RT-PCR and nucleotide sequencing data of the PCR products confirmed that the *PAX5* gene was fused to either the *ETV6* (two cases), *C20orf112* (three cases), *AUTS2* (one case), or *FOXP1* (one case) gene and transcribed into aberrant fusion messages (Fig. 2 *B* and *C*). Each fusion gene was mutually and exclusively detected in the samples studied. In one case with dic(9;20), exon 5 of *PAX5* was fused to exon 8 of *C20orf112*, and in two cases with dic(9;20), exon 8 of *PAX5* was fused to exon 3 of *C20orf112*. *PAX5*/*ETV6* involved exon 4 of *PAX5* and exon 3 of *ETV6*.

Cellular Localization and DNA Binding Affinity of PAX5 Fusion Products. In the *PAX5*/*FOXP1* fusion transcript, the amino acid coding frame of the *FOXP1* gene was not identical to that of *PAX5*, leading to a frame-shift and an early termination codon after the fusion point of these two genes (Fig. 2*D*). However, all other fusion genes were in frame and were predicted to encode chimeric proteins. Two proteins (a short and long form) with different breakpoints were predicted from the *PAX5*/*C20orf112* fusion genes (Fig. 2*D*).

To confirm cellular localization of PAX5-fusion proteins, we transfected vectors encoding wild-type *PAX5* and *PAX5* fusion genes (*PAX5-ETV6*, *PAX5-FOXP1*, *PAX5-C20ORF112S*, and *PAX5-C20ORF112L*) into 293T cells, fractionated the cytoplasmic and nuclear proteins, and examined the wild-type PAX5 and PAX5-fusion proteins by Western blot analysis (Fig. 2*E*). PAX5-ETV6 protein was detected in both the cytoplasm and nucleus; PAX5-FOXP1 and PAX5-C20ORF112L proteins were predom-

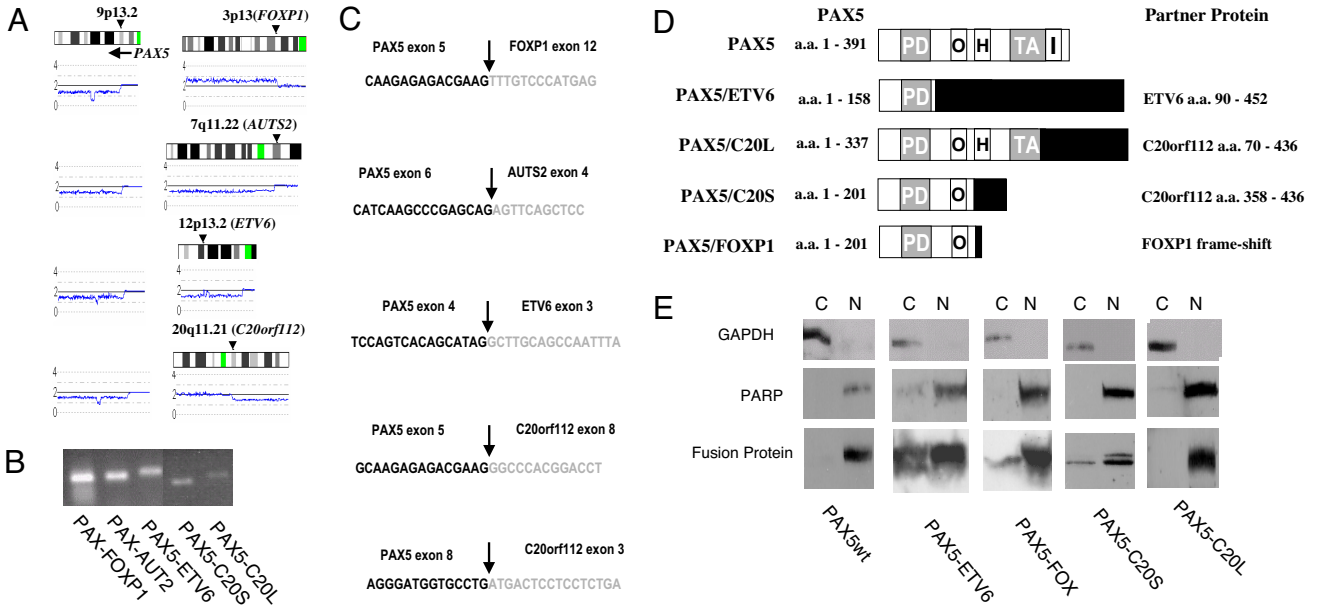

Fig. 2. *PAX5* gene is fused to partner genes. (*A*) Start sites of deletion at 9p13.2 involving the PAX5 gene. (*Left*) SNP-chip data of representative cases with 9p13.2 deletions. A vertical arrow indicates the start sites of 9p deletion that involves the *PAX5* gene. A horizontal arrow shows the direction of transcription of the *PAX5* gene. (*Right*) Chromosomal abnormalities of partner chromosomes. Arrow heads indicate the start sites of duplication or deletions. Genes involved in the start sites are shown. (*B*) Result of RT-PCR. The ALL samples suggesting the presence of PAX5 fusion genes by SNP-chip analysis were examined by RT-PCR using the primers of PAX5 and the respective partner genes. (*C*) Fusion sequences of the *PAX5* and partner genes. Joining sequences of fused transcripts are shown from the indicated exon of the fused gene. (*D*) Schematic structure of wild-type and mutant PAX5. Amino acid positions (aa) of each protein are indicated. PAX5/FOXP1 fusion construct has an early termination codon caused by a frame-shift. PD, paired domain; TA, transcription activation domain; O, octapeptide H, homeodomain-like; I, inhibitory domain. (*E*) Subcellular fractionation of PAX5-fusion proteins. pcDNA vector encoding wild-type PAX5, PAX5-ETV6, PAX5-FOXP1, PAX5-C20ORF112S, or PAX5-C20ORF112L was transfected into 293T cells. Nuclear and cytoplasmic proteins were separated and electrophoresed in the gel. Localization of PAX5-fusion proteins was examined by PAX5 N-terminal specific antibody. Purity of cytoplasmic protein was examined with anti-GAPDH antibody and purity of nuclear proteins with the anti-PARP antibody. C, cytoplasmic fraction; N, nuclear fraction.

inantly localized in the nucleus; and 20% and 80% of PAX5-C20ORF112S proteins were localized in the cytoplasm and the nucleus, respectively (Fig. 2*E*). Localization of the fusion proteins was also confirmed by immunohistochemical staining (data not shown).

Because PAX5-fusion proteins were localized in the nucleus, we analyzed DNA binding affinity of these PAX5-fusion proteins *in vitro*. DNA binding affinity of the PAX5 wild-type and fusion proteins expressed in 293T cells was analyzed by electrophoretic mobility shift assay (EMSA), and signals of probes bound to the proteins were plotted graphically (Fig. 3*A*). Binding activity of each protein in the absence of cold competitor oligonucleotide probe was regarded as 1.0, and the binding activity in the presence of cold competitor oligonucleotide probes was measured. All PAX5-fusion proteins showed similar binding activity to the PAX5 recognition sequences as the wild-type PAX5 (Fig. 3*A*).

***PAX5* Fusion Products Suppressed Transcriptional Activity of Wild-Type PAX5 in a Dominant Negative Fashion, Leading to Inhibition of B-Cell Development.** To examine the effect of PAX5-fusion proteins on transcriptional activity of wild-type PAX5, we performed a reporter gene assay using 293T cells. Cotransfection reporter gene assays using wild-type and fusion *PAX5* expression vectors along with a reporter gene driven by the murine CD19 promoter (which contains three repeats of PAX5 binding sequences) showed that the PAX5 fusion products suppressed transcriptional activity of PAX5 in a dominant-negative fashion (Fig. 3*B*). Expression of wild-type PAX5 proteins was minimally affected by coexpression of PAX5-fusion proteins (Fig. 3*C*), suggesting that PAX5-fusion proteins competed with wild-type PAX5 for the PAX5 binding sequences on the reporter gene.

Further, we transfected vectors encoding either PAX5-C20orf112S or PAX5-C20orf112L, each coexpressing the GFP marker, into Nalm 6 cells (a human B cell ALL cell line, which expresses endogenous PAX5) (data not shown). After transfection, GFP-positive cells were sorted by FACS and expression of PAX5-downstream genes was examined by semiquantitative RT-PCR (Fig. 3*D* and data not shown). We examined 10 downstream target genes (seven positively regulated direct target genes and three negatively regulated genes) of PAX5 (10–12) and found that four, including *ATP1B1*, *BLK*, *NEDD5* and *TCF7L2*, were down-regulated by induction of either PAX5-C20orf112S or PAX5-C20orf112L protein. However, expression of other reported PAX5 downstream target genes, including three positively regulated direct target genes (*IRF8*, *BST1*, *CD19*) and three negatively regulated genes (*CCR2*, *CCR5*, *NOTCH1*) were not affected by the induction of expression of the fusion proteins in these cells.

To examine the effect of PAX5 fusion protein on binding of wild-type PAX5 to the direct target gene *BLK* in the leukemic cells, we performed chromatin-immunoprecipitation (ChIP) assay using Nalm 6 cells transfected with either an empty vector or a construct encoding PAX5-C20orf112S. We used a PAX5 antibody detecting the C-terminal region of the protein, which could detect wild-type PAX5, but not PAX5-C20orf112S, as the C-terminal end of PAX5 was replaced by C20orf112S in this fusion protein. Although wild-type specific PAX5 antibody precipitated the promoter region of *BLK* after transfection of the empty vector, the amount of DNA of the *BLK* promoter region bound to wild-type PAX5 was reduced after transfection of the PAX5-C20orf112S gene (Fig. 3 *E* and *F*).

To examine the effect of PAX5-fusion proteins on B cell development in murine hematopoietic cells, we infected murine

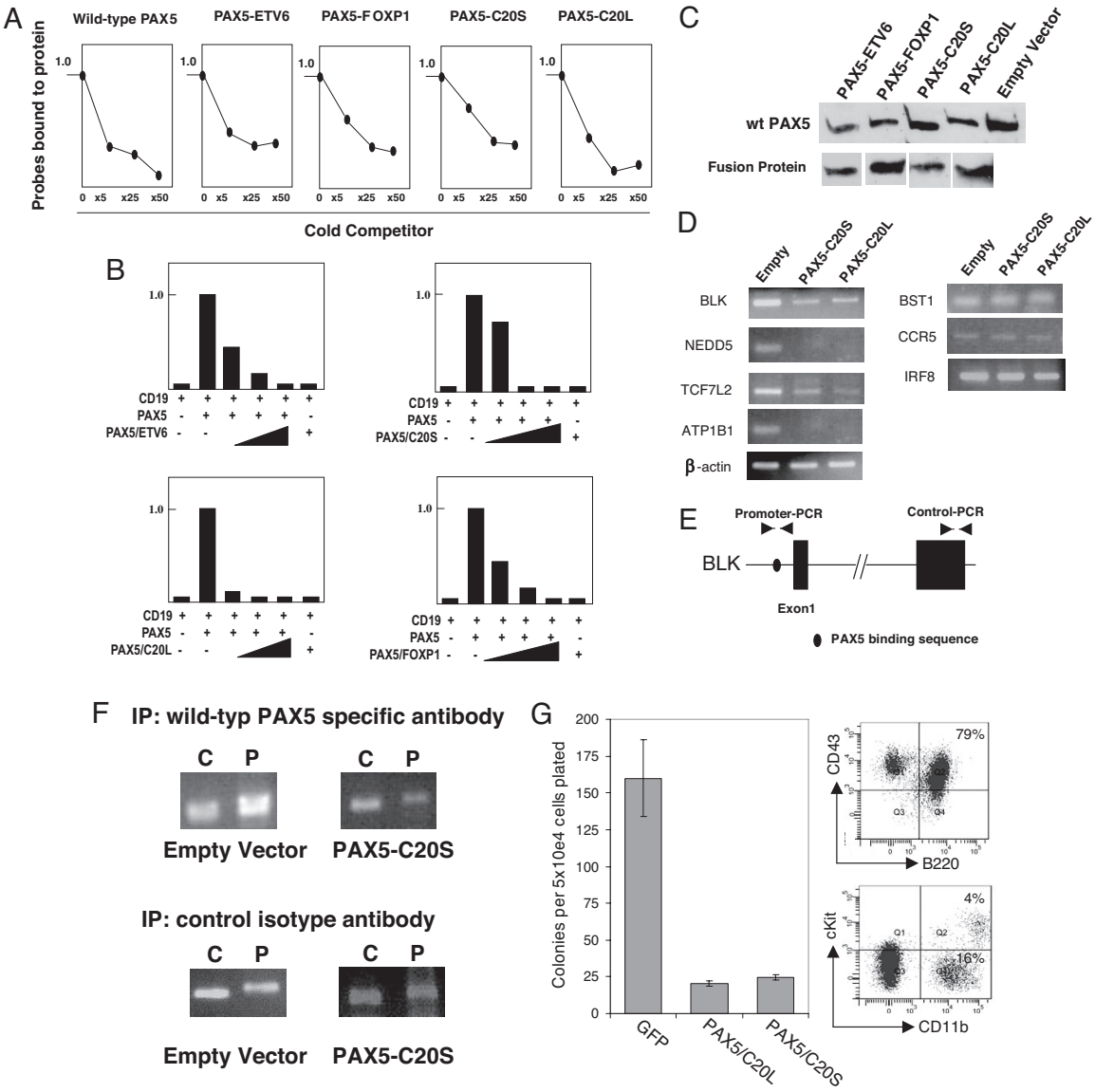

Fig. 3. *PAX5*-fusion proteins suppress transcriptional activity of PAX5 in a dominant-negative fashion and block the growth of B cells. (*A*) Result of EMSA: Wild-type PAX5 and PAX5 fusion were expressed in 293T cells, and nuclear proteins were purified. The purified nuclear proteins were mixed with radioisotope labeled double-strand oligonucleotide DNA, in either the presence or absence of cold competitor oligonucleotides (5-, 25-, and 50-fold cold competitor probes). Intensity of each shifted band indicating DNA–protein complex was measured and plotted graphically. Intensity of shifted bands in the absence of cold competitor probes was regarded as 1.0. (*B*) Reporter gene assay. Wild-type and mutant PAX5 were mixed at a various ratios (1:0, 1:0.3, 1:1, 1:3, respectively, 1 = 500 ng of construct) and transfected. Forty-eight hours later, relative activity of firefly luciferase was measured and plotted. Results represent the mean values of the three experiments. CD19, PAX5 luciferase reporter construct having PAX5 binding region of CD19 promoter; PAX5, wild-type PAX5; PAX5/ETV6, PAX5/ETV6 fusion; PAX5/C20L, long form of PAX5/C20orf112 fusion in which PAX5 exon 8 is fused to C20orf112 exon 3; PAX5/C20S, short form of PAX5/C20orf112 fusion in which PAX5 exon 5 is fused to C20orf112 exon 8; PAX5/FOXP1, PAX5/FOXP1 fusion with an early termination codon caused by a frame-shift after the site of fusion. (*C*) Results of expression of wild-type PAX5 and PAX5-fusion proteins. After cotransfection of equal amounts of vector encoding either wild-type or fusion PAX5 genes into 293T cells, the expression of respective proteins was examined by Western blot. Levels of expression of wild-type PAX5 protein were minimally affected by coexpression of the PAX5-fusion proteins. (*D*) Semiquantitative RT-PCR of downstream target genes of PAX5. Expression of PAX5 downstream target genes was examined by semiquantitative RT-PCR. Nalm 6, a human B cell ALL cell line expressing endogenous PAX5, was transfected with pMSCV-GFP (Empty), pMSCV-GFP-PAX5-C20orf112S (PAX5-C20S), or pMSCV-GFP-PAX5-C20orf112L (PAX5-C20L). GFP-positive cells were sorted and subject to semiquantitative RT-PCR. Optimal cycle numbers to semiquantify the expression of respective genes are as follows; BLK: 25 cycles; Nedd5; 25 cycles; TCF7L2: 25 cycles; ATP1B1: 25 cycles; β-actin: 22 cycles; CCR2: 25 cycles; CCR8: 30 cycles; IRF8: 30 cycles. (*E*) Structure of human *BLK* gene. Structure of *BLK* and primers used for ChIP assay within the 5′ regulatory region (Promoter-PCR) and 3′ end (Control-PCR) of the *BLK* gene is schematically shown. PAX5 binding site in the promoter region is indicated. (*F*) ChIP analysis of the PAX5 binding site in the *BLK* gene promoter. pMSCV-GFP (empty vector) or pMSCV-GFP-PAX5-C20S (PAX5-C20S) was transfected into human Nalm 6 B cell leukemia cells expressing endogenous PAX5. GFP-positive cells were subject to ChIP assay. The cells were fixed in formaldehyde solution and sonicated by ultrasound. DNA-protein complex was incubated with wild-type PAX5 specific antibody, which detected the C-terminal region of PAX5 but not the PAX5-C20orf112S protein (*Upper*). As a control, the DNA-protein complex was reacted with isotype nonspecific antibody (*Lower*). Immunoprecipitated DNA was subjected to PCR to amplify either the *BLK* promoter region containing PAX5 binding sequence (P) or, as an internal control, the 3′ end of the *BLK* gene (C). (*G*) Retrovirus infection experiments. Murine bone marrow cells were collected at 5 days after injection of 5FU. The hematopoietic cells were infected by retrovirus containing pMSCV-GFP empty vector (GFP), pMSCV-GFP-C20orf112L (PAX5/C20L), or pMSCV-GFP-C20orf112S (PAX5/C20S). GFP-positive murine hematopoietic cells were sorted and plated at 5×10^4 cells per plate in methylcellulose containing mSCF, mIL7, and hFL. At 8 days after the plating, the colony numbers were counted (*Left*; results represent means and SD of three experiments). Cell surface antigens on the GFP-positive cells infected with pMSCV-GFP (GFP) at Day 11 were examined by FACS using antibodies against CD43 and B220 (*Upper Right*), c-kit and CD11b (*Lower Right*) antibodies, to confirm the development of B cells.

hematopoietic cells from the bone marrow with retroviral vectors encoding either PAX5-C20orf112S or PAX5-C20orf112L. GFP-positive infected bone marrow cells were sorted by FACS and plated in media containing cytokines that are know to stimulate B cell differentiation (Fig. 3*F*). Murine hematopoietic cells infected with the empty vector showed abundant colonies (Fig. 3*G Left*), and 79% of the cells were B220 positive B cells (Fig. 3*G Upper Right*). In contrast, murine hematopoietic cells infected with either PAX5-C20orf112S or PAX5-C20orf112L formed very few colonies (Fig. 3*G Left*). Most of these colonies were GFP-negative (data not shown), suggesting that these PAX5-fusion proteins impaired B cell development from murine hematopoietic cells.

Discussion

In this study, we describe a paradigm for discovering fusion genes in malignancy by taking advantage of samples with unbalanced translocations and using high density SNP-chip analysis. This technique allows us to identify genes involved in translocations even if chromosomal analysis is not available, especially in solid tumors.

Steps to identify novel fusion genes using SNP chip analysis include (*i*) identify either a deletion or duplication that occurs within two genes; (*ii*) determine whether transcription of both genes is in the same direction; (*iii*) take advantage of ancillary tests such as standard chromosomal analysis or spectral karyotyping (14), which can grossly show that two chromosomes are fused; and (*iv*) design primers of candidate genes and perform RT-PCR to clone fusion genes. Rapid amplication of cDNA ends (RACE) (15) or long-distance PCR (12) also help the cloning of genes involved in translocations. In our SNP-chip data, a number of regions of segmental deletions or duplications were detected (9). Although some of them are simple deletions or duplications at the original sites of the chromosomes, the others are deletions that occurred during chromosomal translocations or when duplicated fragments were inserted into chromosomal sites other than the original region (data not shown). Therefore, data of chromosomal analysis help to define translocations, leading to identification of candidate genes in novel fusion genes.

Recently, Tomlins *et al.* found the fusion genes *TMPRSS2*/*ERG* and *TMPRSS2*/*ETV1* in prostate cancers by using expression microarray data (16). They focused on the genes *ERG* and *ETV1*, which are highly expressed in this cancer and examined levels of individual exons of these two genes (16). They found differences in expression of 5′ and 3′ regions of the genes, suggesting that these genes are fused to each other (16). In these fusion genes, the 5′ regions were replaced by the *TMPRSS2* gene, resulting in the differences in the expression of the 5′ and 3′ region of the *ERG*/*ETV1* genes (16). They also used SNP-chip analysis to identify these fusion genes and found a deletion of a genomic region between TMPRSS2 (21q22.3) and ERG (21q22.2), leading to fusion of these two genes (17). These new technologies, based on oligonucleotide microarrays and bioinformatics, will help to identify fusion genes in cancers.

Our study found that the *PAX5* gene was frequently fused to one of a variety of partner genes. PAX5 is a key transcription factor in the development of B cells (18, 19). We found that these PAX5 fusion proteins suppressed the function of wild-type PAX5 in a dominant-negative fashion and suppressed expression of downstream target genes of wild-type PAX5 in leukemic cells.

We found that when *PAX5* was joined to one of its fusion partner genes, its C-terminal end was replaced by one of the partner genes. Elimination of the C-terminal end of *PAX5* may play an important role in generation of a dominant negative form of mutated *PAX5*. In *in vitro* assays, PAX5-fusion proteins showed a similar affinity as wild-type PAX5 for the PAX5 recognition sequences. Although expression of several downstream targets of wild-type PAX5 was repressed by expression of PAX5-fusion proteins, others were not affected. Binding of transcription factors to DNA can be modulated by cofactors and/or neighboring transcription factors (20). Compared to PAX5, PAX5-fusion proteins may bind more strongly to some target genes and more weakly to others, depending on the contextual environment of the target genes.

Further, our data showed that PAX5-fusion protein inhibited B cell development of hematopoietic cells in a colony formation assay. This result may suggest that PAX5 fusion protein blocked differentiation of hematopoietic cells into mature B cells. PAX5-deficient mice have impairment of B cell differentiation (18). These data suggest that PAX5-fusion proteins may contribute to leukomegenesis by blocking B cell differentiation. It has been suggested that two distinct genetic abnormalities contribute to leukemogenesis in acute myelogenous leukemia (AML); one is mutations promoting cellular proliferation, for example *FLT3* or *RAS* mutations, and the other is mutations blocking differentiation, for example *PML-RARA* or *RUNX1-ETO* (21, 22). PAX5-fusion proteins may cooperate with unidentified mutations promoting cellular proliferation in the ALL cells.

Recently, Mullighan *et al.* have analyzed pediatric ALL samples by high density SNP-chips and found frequent abnormalities of *PAX5* gene (23). Their data also showed that PAX5 fusion products suppressed transcriptional activity of PAX5 in a dominant-negative fashion (23). In addition, other researcher have reported *PAX5* fusion genes, including *PAX5* fused to *ETV6* (12p13) (23, 24), *FOXP1* (3p14) (23), *ZNF521* (18q11) (23), *ELN* (7q11.23) (25), and *PML* (15q24) (26). We have found *PAX5* fused to either *ETV6*, *FOXP1*, *C20orf112* (20q11), or *AUTS2* (7q11.22).

In our study, the function of PAX5 was attenuated by the dominant-negative forms of the fusion products in B cell lineage ALL, suggesting that PAX5 behaves as a tumor suppressor in early B cells, and that impairment of its function can be associated with the development of ALL. In contrast, translocation of the *PAX5* gene to the enhancer region of the Ig heavy chain gene [t(9;14)(p13.2;q32)] or point mutations of the 5′ regulatory region of the *PAX5* gene leads to its overexpression, which is associated with B cell lineage lymphomas (27–29). Also, experimental overexpression of wild-type PAX5 can transform lymphocytes (30, 31). Therefore, an aberrant PAX5 may behave in a dominant-negative fashion at the pre-B stage of B cell development, resulting in ALL; its forced expression in a more mature B cell can lead to lymphoma. Our study showed that PAX5-fusion proteins blocked differentiation of B cells but did not transform them. B cells at different stages of differentiation may need alteration of distinct sets of pathways to transform. Why PAX5 can act as a tumor suppressor in ALL and as an oncoprotein in lymphoma is unclear. Further studies are needed to clarify the mechanism of this paradoxical phenomenon in carcinogenesis.

In summary, we identified multiple fusion genes in ALL by SNP-chip analysis, leading to the exploration of a B cell differentiation block as a contributing factor to the development of ALL. This methodology should help researchers to identify oncogenic fusion genes and explore the mechanism of tumorigenesis in other types of cancers as well.

Materials and Methods

Samples and DNA/RNA Preparation. SNP-chip was performed on 399 pediatric ALL patients consecutively enrolled in the ALL-BFM 2000 trial of the Berlin-Frankfurt-Münster (BFM) study at diagnosis and during remission (350 cases were B cell lineage ALL and 49 cases were T cell lineage ALL) (9). Detailed results of the SNP-chip analysis are published separately (9). The ALL-BFM 2000 study was approved by the local ethics committee. DNA and RNA were extracted from the ALL samples and cell lines by using standard techniques (32). Nalm 6, a human pre-B ALL cell line, was generously provided by Dr. G. Crook (Los Angeles Children's Hospital, Los Angeles, CA) and maintained in RPMI medium 1640 with 10% FBS.

SNP-Chip Analysis. SNP-chip of GeneChip Human mapping 50k array XbaI 240 and/or 250k Nsp were used for this study (Affymetrix Japan). Preparation of samples was reported previously (4, 5). The data were analyzed by CNAG program as previously described (4, 5). All 399 ALL samples and their matched control samples were analyzed by using 50K-SNP chip; selected cases with genomic abnormalities were also analyzed by using 250K SNP-chip.

RT-PCR. RT-PCR was performed by using ThermoScript RT-PCR Systems (Invitrogen) according to the manufacturer's protocol. The primers used for detection of *PAX5* fusion transcripts are listed in Table S2. Expression of PAX5 downstream target genes in Nalm 6 cells after transfection was examined by semiquantitative RT-PCR. The gene names and their primer sequences are listed in Table S3.

Reporter Gene Constructs and Expression Vectors. The PAX5 reporter gene construct with the luciferase gene and PAX5 binding region of the CD19 promoter, as well as the human PAX5 cDNA constructs, were kindly provided by Dr. M. Busslinger (Research Institute of Molecular Pathology, Vienna, Austria). PAX5-fusion constructs were generated by using PCR. All coding regions were ligated into the pcDNA3.1 vectors (Stratagene). Wild-type PAX5 cDNA was ligated into pMSCV vector (Clontech), and EGFP cDNA was ligated under the control of pGK promoter as a marker (pMSCV-GFP-wtPAX5). PAX5-C20orf112S and PAX5-C20orf112L cDNA sequences were also ligated into pMSCV-GFP vectors.

Transfection and Reporter Gene Assay. For reporter gene assays, pMSCV-GFP-wtPAX5 and pcDNA vectors encoding PAX5-fusion genes were cotransfected with the PAX5 reporter construct and pRL (*Renilla* luciferase) vector into 293T cells by using the Effecten transfection kit (Qiagen). Firefly and *Renilla* luciferase activities were measured with the Dual-Luciferase Reporter Assay System (Promega). Transfection into Nalm6 human pre-B cell ALL cell line was performed with Amaxa nucleofector. GFP-positive cells were sorted by using the MoFlo cell sorter (Dako). Detailed information about the procedure is described in *SI Text*.

Retrovirus Transduction into Murine Hematopietic Cells. Retrovirus containing pMSCV-GFP (empty), pMSCV-GFP-PAX5-C20orf112S, and pMSCV-GFP-PAX5-C20orf112L was generated. The retrovirus was transfected into murine bone marrow cells as previously reported (33). After the transfection, GFP-positive cells were sorted and plated into methylcellulose cultures (M3231; Stem Cell Technologies) as previously described (33). Surface antigens (CD43, B220, c-kit, and CD11b) of these GFP-positive cells were examined by using FACScan (Becton–Dickinson). Detailed information of the procedure is described in the *SI Text*.

Subcellular Fractionation of Proteins and EMSA. Forty-eight hours after transfection of vectors into 293T cells, the cells were subjected to subcellular fractionation with the CelLytic NuCLEAR Extraction Kit (Sigma-Aldrich). Detailed information of the procedure is described in the *SI Text*.

Purified nuclear proteins from the cells were also subjected to EMSA as previously reported (34). Detailed information of the procedure is described in the *SI Text*.

Chromatin Immunoprecipitation (ChIP) Assay. ChIP assay was performed with the Magna ChIP A kit from Millipore according to the manufacturer's protocol. pMSCV-GFP or pMSCV-GFP-C20orf112S were transfected into Nalm 6 cells as described above, and GFP-positive cells were sorted by MoFlo (Dako). Precipitated DNA was recovered and subjected to PCR to amplify the *BLK* promoter region and the 3′ end of the *BLK* gene (internal control). The primer sequences used for ChIP assay are listed in Table S4. Detailed information of the procedure is described in the *SI Text*.

ACKNOWLEDGMENTS. We thank Dr. M. Busslinger (Research Institute of Molecular Pathology, Vienna, Austria) for generously providing the CD19 promoter reporter construct and the human PAX5 cDNA. This study was supported by the Parker Hughes Fund and by grants from the National Institutes of Health. N.K. is supported by a fellowship from The Tower Cancer Research Foundation. H.P.K. holds the Mark Goodson Chair in Oncology Research at Cedars-Sinai and is a member of the Jonsson Cancer Center and the Molecular Biology Institute of the University of California, Los Angeles. This work was also supported by a grant-in-aid from the Department of Health, Welfare and Labor; by the Ministry of Education, Culture, Sports, Science and Technology (Japan); by European Union Grant FOOD-CT-2005-016320; by grants from the Deutsche Krebshilfe to Cambridge Research Biochemicals; and by the Fritz-Thyssen Foundation (C.S.). The ALL-BFM 2000 trial is supported by Grant 50-2698-Schr 1 of the Deutsche Krebshilfe.

1. Armstrong SA, Look AT (2005) Molecular genetics of acute lymphoblastic leukemia. *J Clin Oncol* 23:6306–6315.
2. Pui CH, Evans WE (2006) Treatment of acute lymphoblastic leukemia. *N Engl J Med* 354:166–178.
3. Pui CH, Relling MV, Downing JR (2004) Acute lymphoblastic leukemia. *N Engl J Med* 350:1535–1548.
4. Nannya Y, *et al.* (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 65:6071–6079.
5. Yamamoto G, *et al.* (2007) Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am J Hum Genet* 81:114–126.
6. Lindblad-Toh K, *et al.* (2000) Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat Biotechnol* 18:1001–1005.
7. Raghavan M, *et al.* (2005) Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res* 65:375–378.
8. Lehmann S, *et al.* (2008) Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. *Cancer* 112:1296–1305.
9. Kawamata N, *et al.* (2008) Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* 111:776–784.
10. Lin YH, Shin EJ, Campbell MJ, Niederhuber JE (1995) Transcription of the blk gene in human B lymphocytes is controlled by two promoters. *J Biol Chem* 270:25968–25975.
11. Schebesta A, *et al.* (2007) Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function. *Immunity* 27:49–63.
12. Delogu A, *et al.* (2006) Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity* 24:269–281.
13. Schröck E, Padilla-Nash H (2000) Spectral karyotyping and multicolor fluorescence in situ hybridization reveal new tumor-specific chromosomal aberrations. *Semin Hematol* 37:334–347.
14. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 85:8998–9002.
15. Cheng S, Fockler C, Barnes WM, Higuchi R (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc Natl Acad Sci USA* 91:5695–5699.
16. Tomlins SA, *et al.* (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310:644–648.
17. Perner S, *et al.* (2006) TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res* 66:8337–8341.
18. Urbanek P, Wang ZQ, Fetka I, Wagner EF, Busslinger M (1994) Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* 79:901–912.
19. Xie H, Ye M, Feng R, Graf T (2004) Stepwise reprogramming of B cells into macrophages. *Cell* 117:663–676.
20. Kumaran RI, Thakar R, Spector DL (2008) Chromatin dynamics and gene positioning. *Cell* 132:929–934.
21. Gilliland DG, Griffin JD (2002) The roles of FLT3 in hematopoiesis and leukemia. *Blood* 100:1532–1542.
22. Tenen DG (2003) Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* 3:89–101.
23. Mullighan CG, *et al.* (2007) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446:758–764.
24. Cazzaniga G, *et al.* (2001) The paired box domain gene PAX5 is fused to ETV6/TEL in an acute lymphoblastic leukemia case. *Cancer Res* 61:4666–4670.
25. Bousquet M, *et al.* (2007) A novel PAX5-ELN fusion protein identified in B cell acute lymphoblastic leukemia acts as a dominant negative on wild-type PAX5. *Blood* 109:3417–3423.
26. Nebral K, *et al.* (2007) Identification of PML as novel PAX5 fusion partner in childhood acute lymphoblastic leukaemia. *Br J Haematol* 139:269–274.
27. Busslinger M, Klix N, Pfeffer P, Graninger PG, Kozmik Z (1996) Deregulation of PAX-5 by translocation of the Emu enhancer of the IgH locus adjacent to two alternative PAX-5 promoters in a diffuse large-cell lymphoma. *Proc Natl Acad Sci USA* 93:6129–6134.
28. Iida S, *et al.* (1996) The t(9;14)(p13;q32) chromosomal translocation associated with lymphoplasmacytoid lymphoma involves the PAX-5 gene. *Blood* 88:4110–4117.
29. Pasqualucci L, *et al.* (2007) Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412:341–346.
30. Anderson K, *et al.* (2007) Ectopic expression of PAX5 promotes self renewal of biphenotypic myeloid progenitors co-expressing myeloid and B-cell lineage associated genes. *Blood* 109:3697–3705.
31. Souabni A, Jochum W, Busslinger M (2007) Oncogenic role of Pax5 in the T-lymphoid lineage upon ectopic expression from the immunoglobulin heavy-chain locus. *Blood* 109:281–289.
32. Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Press, Cold Spring Harbor, NY). 3rd Ed.
33. Schwieger M, *et al.* (2002) AML1-ETO inhibits maturation of multiple lymphohematopoietic lineages and induces myeloblast trnsformation in syngergy with ICSBP deficiency. *J Exp Med* 196:1227–1240.
34. Sato H, Wang D, Kudo A (2001) Dissociation of Pax-5 from KI and KII sites during kappa-chain gene rearrangement correlates with its association with the underphosphorylated form of retinoblastoma. *J Immunol* 166:6704–6710.