

# *HOXA* genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL)

Jean Soulier, Emmanuelle Clappier, Jean-Michel Cayuela, Armelle Regnault, Marina García-Peydró, Hervé Dombret, André Baruchel, Maria-Luisa Toribio, and François Sigaux

**Using a combination of molecular cytogenetic and large-scale expression analysis in human T-cell acute lymphoblastic leukemias (T-ALLs), we identified and characterized a new recurrent chromosomal translocation, targeting the major homeobox gene cluster *HOXA* and the *TCRB* locus. Real-time quantitative polymerase chain reaction (RQ-PCR) analysis showed that the expression of the whole *HOXA* gene cluster was dramatically dysregulated in the *HOXA*-rearranged cases, and also in *MLL* and *CALM-AF10*-related T-ALL cases, strongly suggesting that**

***HOXA* genes are oncogenic in these leukemias. Inclusion of *HOXA*-translocated cases in a general molecular portrait of 92 T-ALLs based on large-scale expression analysis shows that this rearrangement defines a new homogeneous subgroup, which shares common biologic networks with the *TLX1*- and *TLX3*-related cases. Because T-ALLs derive from T-cell progenitors, expression profiles of the distinct T-ALL subgroups were analyzed with respect to those of normal human thymic subpopulations. Inappropriate use or perturbation of specific molecular networks**

**involved in thymic differentiation was detected. Moreover, we found a significant association between T-ALL oncogenic subgroups and ectopic expression of a limited set of genes, including several developmental genes, namely *HOXA*, *TLX1*, *TLX3*, *NKX3-1*, *SIX6*, and *TFAP2C*. These data strongly support the view that the abnormal expression of developmental genes, including the prototypical homeobox genes *HOXA*, is critical in T-ALL oncogenesis. (Blood. 2005;106:274-286)**

© 2005 by The American Society of Hematology

## Introduction

T-cell acute lymphoblastic leukemias (T-ALLs) are highly malignant tumors representing 10% to 15% of pediatric and 25% of adult ALLs in humans.<sup>1,2</sup> T-ALL cells derive from partially differentiated thymocytes. These cells originate from pluripotent progenitors that are progressively committed to the  $\alpha\beta$  or  $\gamma\delta$  T-cell lineages in the thymus.<sup>3</sup> Somatic V(D)J recombination rearranges the T-cell-receptor (TCR) gene segments, generating protein products that allow cells to pass through several cellular processes, so-called  $\beta$ -selection and selection.<sup>4,5</sup> Hence, V(D)J recombination both directly compromises genome integrity and indirectly controls cellular functions considered to be critical for oncogenesis, such as cell cycle, proliferation, and apoptosis.<sup>6</sup> However, T-ALLs remain rare, consistent with efficient control mechanisms that can only be overcome by accumulation of oncogenic events. Several classes of proto-oncogenes can be activated by chromosomal rearrangements or epigenetic mechanisms in T-ALL.<sup>7-9</sup> The resulting oncoproteins include basic helix-loop-helix proteins (TAL1/SCL, TAL2, LYL1), homeodomain-containing proteins (TLX1/HOX11, TLX3/HOX11L2), and LIM only proteins (LMO1, LMO2), which are likely to be involved in transcription factor complexes.<sup>8,10</sup> In addition, chromosomal alterations or DNA mutations can activate genes involved in signal transduction and thymus differentiation, such as *NOTCH1* and *LCK*,<sup>11-13</sup> or can lead to gene fusion (*MLL-ENL*, *CALM-AF10*, *NUP214-ABL1*).<sup>14-16</sup> Moreover, the *CDKN2A/INK4A/ARF* tumor suppressor gene is inacti-

vated by deletion or mutation in most T-ALL cases.<sup>17,18</sup> Recently, T-ALL classifications based on selected oncogene expression and corresponding microarray gene expression patterns have been reported.<sup>19</sup>

In the present study, a combination of molecular cytogenetic and large-scale expression analyses allowed us to identify and characterize a new recurrent chromosomal translocation directly targeting and dysregulating the *HOXA* locus. This new rearrangement extends the catalog of T-cell oncogenes to these prototypical homeobox genes and strongly suggests that *HOX* dysregulation can be not only a consequence of cancer in humans but also a cause.<sup>20</sup> Moreover, we show that *HOXA* genes can be included in biologic networks that associate oncogenesis and thymocyte differentiation processes. Our data suggest that inappropriate use or perturbation of specific biologic pathways involved in thymic maturation, and ectopic expression of genes, often developmental genes, are critical for T-ALL oncogenesis.

## Patients, materials, and methods

### Patients, leukemic cells, and annotations

T-ALL was diagnosed in 92 patients who were treated at Saint Louis Hospital (Paris, France). Seven patients were studied at diagnosis and relapse (total 99 T-ALL samples). There were 56 children (median age,

From Institut National de la Santé et de la Recherche Médicale (INSERM) U462 'Lymphocyte et Cancer,' and Molecular Hematology Laboratory, Institut Universitaire d'Hématologie, Hôpital Saint Louis, Paris, France; Centro de Biología Molecular Severo Ochoa, CSIC, Universidad Autónoma de Madrid, Madrid, Spain; and Adult Hematology Department and Pediatric Hematology Department, Hôpital Saint Louis, Paris, France.

Submitted October 7, 2004; accepted February 27, 2005. Prepublished online as *Blood* First Edition Paper, March 17, 2005; DOI 10.1182/blood-2004-10-3900.

Supported by grants from the Ligue Nationale contre le Cancer ('Programme Cartes d'Identité des Tumeurs'), the Ministère de la Recherche ('Programme

Génomique'), the Ministère de la Santé ('Programme Tumorothèques'), and INSERM.

The online version of the article contains a data supplement.

**Reprints:** François Sigaux, INSERM U462, Institut Universitaire d'Hématologie, Centre Hayem, Hôpital Saint Louis, 1 Ave Claude Vellefaux, 75010 Paris, France; e-mail: fs@chu-stlouis.fr.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2005 by The American Society of Hematology

9 years; range, 1-16 years) and 36 adults (median age, 27 years; range, 17-66 years). Informed consent was obtained from the patients or relatives. The study was approved by the Review Board of the Fédération d'Hématologie, Hôpital Saint-Louis, Paris, France. T-ALL diagnosis was based on morphologic and immunophenotypic criteria using flow cytometry and an extended monoclonal antibody panel. Bone marrow samples from healthy donors ( $n = 4$ ), unpurified thymus cells, and MOLT4, HSB2, CEM, and Jurkat T-cell lines were also included in the study. DNA and RNA were extracted from the cryopreserved leukemic cells. Expression of the main T-ALL oncogene transcripts, *TALI*, *TLX1*, *TLX3*, *LMO1*, *LMO2*, and *LYL1*, and the *SIL-TALI*, *CALM-AF10*, and *NUP214-ABL1* transcripts were analyzed by real-time quantitative polymerase chain reaction (RQ-PCR; shown in Supplemental Figure S1, available on the *Blood* website; see the Supplemental Materials link at the top of the online article), as described.<sup>21</sup> Because *TALI* is highly expressed in erythroblasts,<sup>22</sup> 10 cases with high expression of erythroid-associated genes (Table S1) were excluded from the final analysis of *TALI* expression to avoid spurious results. *MLL* and *CDKN2A/INK4A/ARF* genomic abnormalities were searched for in the *HOXA*-expressing cases by fluorescent in situ hybridization (FISH; LSI *MLL* DualColor, VYSIS/ABOTT, Rungis, France, and RP11-149I2 BAC probe, respectively). *MLL* partners were identified in the 3 *MLL*-translocated cases using the *MLL* fusionChip kit (IPSOGEN, Marseille, France). *BMI1*, *NKX3-1*, *SIX6*, *TFAP2C*, *RAG1*, and *CD8A* gene expression was tested by RQ-PCR (see "PCR systems for gene expression analysis" in the Supplemental Materials and methods).

### Isolation of normal human thymocyte subsets

Postnatal thymocytes were obtained from thymus samples removed during corrective cardiac surgery in 3 patients aged 1 month to 3 years old.

Thymocyte subpopulations (for a review, see Carrasco et al<sup>4</sup>) were isolated and analyzed by fluorescence-activated cell sorting (FACS) for purity as previously described.<sup>23-26</sup> Briefly, after rosetting to remove CD2<sup>bright</sup> thymocytes, CD34<sup>+</sup> thymocytes were magnetically sorted and then treated with CD1a microbeads to isolate the CD34<sup>+</sup>CD1a<sup>+</sup> cells (Pre-T1, 96.5% purity). The most immature CD34<sup>bright</sup>CD33<sup>low</sup>CD1a<sup>-</sup> cells (early thymic progenitors [ETPs]), were sorted from the CD1a-depleted fraction with anti-CD33-phycoerythrin (PE) and anti-PE microbeads (94.6% purity).<sup>26</sup> The remaining CD34<sup>+</sup>CD33-depleted thymocytes corresponded to CD34<sup>+</sup>CD1a<sup>-</sup> Pro-T thymocytes (99.4%). Other thymic subpopulations were purified after centrifugation of total thymus cells on stepwise Percoll density gradients as previously described.<sup>4,24</sup> Characteristics of these subsets were: Pre-T2 and Pre-T3(E) (not separated here and further referred as Pre-T2/Pre-T3[E]), 99% CD4<sup>+</sup>CD8 $\alpha\beta$ <sup>-</sup>CD3<sup>-</sup>TCR $\alpha\beta$ <sup>-</sup>; Pre-T3(L), 98% CD4<sup>+</sup>CD8 $\alpha\beta$ <sup>+</sup>CD3<sup>-</sup>TCR $\alpha\beta$ <sup>ic</sup>; immature DPs (immature double-positive cells), 98% CD4<sup>+</sup>CD8 $\alpha\beta$ <sup>+</sup>CD3<sup>-</sup>TCR $\alpha\beta$ <sup>-</sup>; DP-TCR $\alpha\beta$ <sup>low</sup>, 63.2% CD4<sup>+</sup>CD8 $\alpha\beta$ <sup>+</sup>CD3<sup>low</sup>TCR $\alpha\beta$ <sup>low</sup>; DP-TCR $\alpha\beta$ <sup>int</sup>, CD8 $\alpha\beta$ <sup>+</sup>CD4<sup>+</sup>CD3<sup>int</sup>TCR $\alpha\beta$ <sup>+</sup>, 99% purity with 74% CD3<sup>int</sup>; DP-TCR $\alpha\beta$ <sup>br</sup>, 99.9% CD8 $\alpha\beta$ <sup>+</sup>CD4<sup>+</sup>CD3<sup>br</sup>TCR $\alpha\beta$ <sup>br</sup>; CD4-SP (CD4 single-positive cells), 97.2% CD4<sup>+</sup>CD3<sup>br</sup>TCR $\alpha\beta$ <sup>br</sup>; and CD8-SP, 99.8% CD8 $\alpha\beta$ <sup>+</sup>CD3<sup>br</sup>TCR $\alpha\beta$ <sup>br</sup>; see "Reagents used for human thymocyte subsets isolation" in the Supplemental Materials and methods.

### Genomic analysis of the *TCRB* and *HOXA* loci

*TCRB*-flanking FISH probes were prepared from the CTD-3092H9 P1 artificial chromosome (PAC; 5' probe, centromeric) and RP11-368I15 bacterial artificial chromosome (BAC; 3' probe, telomeric). *HOXA*-flanking FISH probes were prepared from the RP11-167F23 and RP5-1200I23 BACs (telomeric and centromeric of the *HOXA* locus, respectively). FISH mapping within the *HOXA* locus was performed using the RP11-163M21, RP11-1132K14, RP11-1025G19, and RP11-1148E13 BACs. Images were captured and analyzed using a DMRXA epifluorescence microscope (Leica), a Plan Apo 63  $\times$ /1.32 oil immersion objective lens (0.132 mm/pixel; Leica), a CCD camera (Cohu), and Q-FISH software (Leica). Southern blot analysis of the *HOXA* locus was performed using a panel of probes shown in Supplemental Materials and methods. Gene information was collected in the UCSC Genome Bioinformatics site (<http://genome.ucsc.edu>).

For inverse PCR amplification of the *TCRB*-*HOXA* breakpoint regions, leukemic gDNAs were digested by the *MaeI* and *MseI* restriction enzymes (cases TL43 and TL46, respectively) and ligated at low concentration. The circularized breakpoint-containing fragments were amplified by nested PCR. Primers were 5'-ATACCCACCACCAACTTG-3' and 5'-ATGCTT-TGTAAGGCTGTTTG-3' (first round PCR), and 5'-CGCGGATCCATGT-CACCCACCCTCCA-3' and 5'-CGCGGATCCAGTACCCACCCCTCTCC-3' (nested PCR). Reciprocal breakpoint sequences were long-range PCR amplified using the forward *HOXA* primers 5'-GAGCCTCGT-GTCTTTCTCC-3' (case TL43) and 5'-GTTCTAGCTTATTTACT-TCTA-3' (case TL46), and a set of reverse J $\beta$  primers, including 5'-CTGCTCAGCTCTTTTGGCC-3' (case TL43), and 5'-CCTCCCCACC-GCTGAGAG-3' (case TL46). Translocation breakpoints and flanking sequences were sequenced both directly and after cloning in the PCR-Script vector (Stratagene, La Jolla, CA).

### RQ-PCR analysis of genes from the *HOXA* locus

Leukemic and control cDNAs were analyzed for expression of the 11 *HOXA* and *EVX1* genes using SYBRgreen reagents (Applied Biosystems, Courtabouef, France). Primers are indicated in Supplemental Materials and methods. Results were expressed in Ct of the amplification curve and normalized on the basis of the expression of 2 housekeeping genes, *ABL1* and *TBP*. Because the *HOX* paralogs share high sequence similarity, primers were carefully designed to avoid cross-priming. The specificity of the amplified fragments was further checked by analyzing the dissociation curves, migration on polyacrylamide gels, and by direct sequencing of the PCR products for at least one product of each type. In most cases, the primers overlap exon boundaries to prevent contamination by gDNA.

### Microarray experiments and data analysis

Large-scale expression and analyses were performed on quality-controlled RNAs using the Affymetrix U133A technology, according to standardized procedures at the Strasbourg Genopole national facility. See "Microarray experiments and data analysis" in the Supplemental Materials and methods for details on RNA control and data normalization.

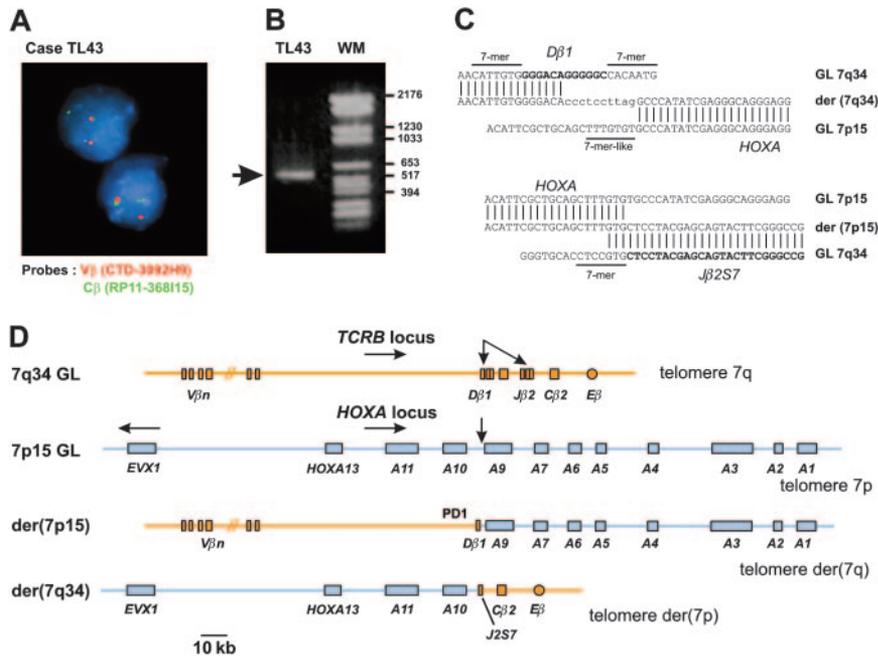
Microarray analyses were performed on the quantile-normalized data using dChip1.3 (<http://biosun1.harvard.edu/complab/dchip/>) and BRB (<http://linus.nci.nih.gov/BRB-ArrayTools.html>) software, as described in the Supplemental Materials and methods. LocusLink symbols and ID (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Gene>) were used when the information was available to label genes corresponding to the probe sets.

## Results

### A new recurrent chromosomal translocation in T-ALL targets the *HOXA* locus

In an attempt to identify new oncogenic abnormalities in T-ALL, we carried out an interphasic FISH screening in 15 T-ALL cases that did not show expression of the 3 oncogenes most commonly involved in this type of leukemia (*SIL-TALI*, *TLX1*, and *TLX3*). We used FISH probes of the T-cell-receptor  $\beta$  genes locus (*TCRB*) because this locus is often involved in oncogenic rearrangements of T-ALL and considering that the near-telomeric location of *TCRB* could lead to cryptic translocations. In this way, we identified 2 T-ALL cases (TL43, TL44) with a *TCRB* translocation (Figure 1A). Cloning of the TL43 translocation identified the *HOXA* locus as the *TCRB* partner sequence (Figure 1B-D). The *HOXA* locus was also involved in the second case (TL44), as demonstrated by FISH and Southern blot (data not shown).

We addressed the hypothesis that this new recurrent translocation could define a homogeneous T-ALL subgroup by analyzing global gene expression on a series of unselected T-ALL samples.



**Figure 1. A new chromosomal translocation involves the *TCRB* and *HOXA* loci.** (A) A *TCRB* chromosomal translocation was identified in leukemic cells of patient TL43 by dissociation of *TCRB*-flanking FISH probes. The centromeric  $V\beta$ -flanking probe (CTD-3092H9) and the telomeric  $C\beta$ -flanking probe (RP11-168115) were labeled with Texas red and fluorescein isothiocyanate (FITC), respectively. Original magnification,  $\times 63$ . (B) A 541-bp fragment containing the  $D\beta 1$  breakpoint was produced by circled nested PCR (left lane, the rearranged band is shown by an arrow; right lane, molecular weight marker). (C) Breakpoint and flanking sequences of both derivative chromosomes of the *TCRB*-*HOXA* translocation TL43. Recognition sequence signal (RSS) heptamer and putative heptamer-like sequence are indicated according to consensus.<sup>27</sup> Untemplated nucleotides (N-diversity) are typed in lowercase. GL indicates germline; der, derivative chromosomes. (D) Schematic representation of the *TCRB*-*HOXA* translocation TL43, drawn to scale according to data from the IMGT Information System (<http://imgt.cines.fr>) and the UCSC Genome Bioinformatics site (<http://genome.ucsc.edu>), and to the DNA sequences of the breakpoint regions of the 2 derivative chromosomes. Chromosomes are depicted according to the direction of the *TCRB* and *HOXA* gene transcription (except for the *EVX1* gene, which is transcribed in the opposite direction as indicated by arrow). Breakpoints are indicated with arrows.

### Large-scale expression analysis identifies a *HOXA*-expressing subgroup within HOX-related T-ALL samples

Leukemic cells from a series of 92 patients, including the 2 patients with the newly identified *TCRB*-*HOXA* rearrangement, were studied by transcriptome analysis of tumor cells using Affymetrix U133A microarrays. Sample and gene classifications were performed with unsupervised hierarchical methods using a list of 255 probe sets selected on the basis of expression variability and reliability among samples (the general microarray strategy is shown in Figure 2A; see also "Microarray experiments and data analysis" in the Supplemental Materials and methods for probe-set selection and methods). Major clusters of T-ALL were obtained, and significant correlations with the immunologic and oncogene expression annotations allowed definition of major T-ALL oncogenic groups (Figure 2B).

The first branch of the hierarchy (labeled as *TAL1*-related, *TAL\_R*) includes all samples with *SIL-TAL1* rearrangement ( $P < .001$ , 16 cases). Cases of this branch are associated with strong *TAL1* expression ( $P < .001$ , 33 of 40 T-ALL cases with

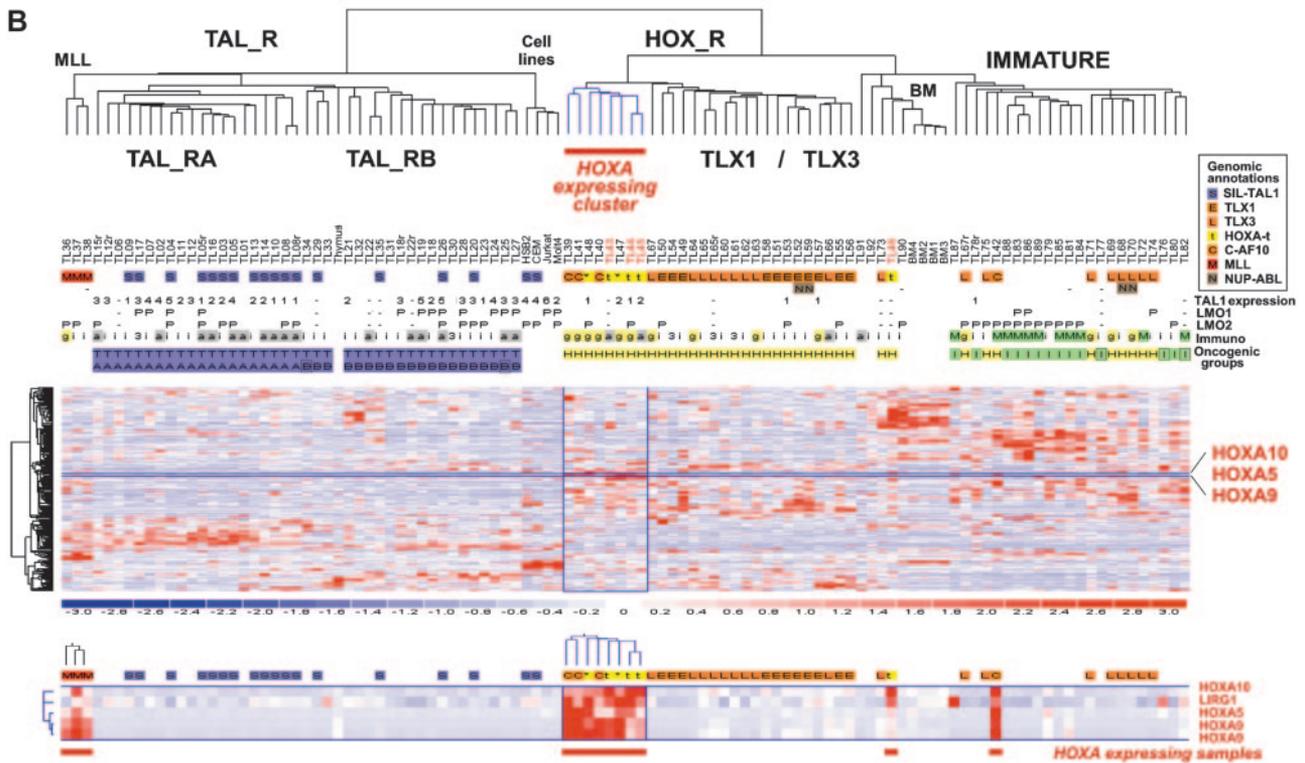
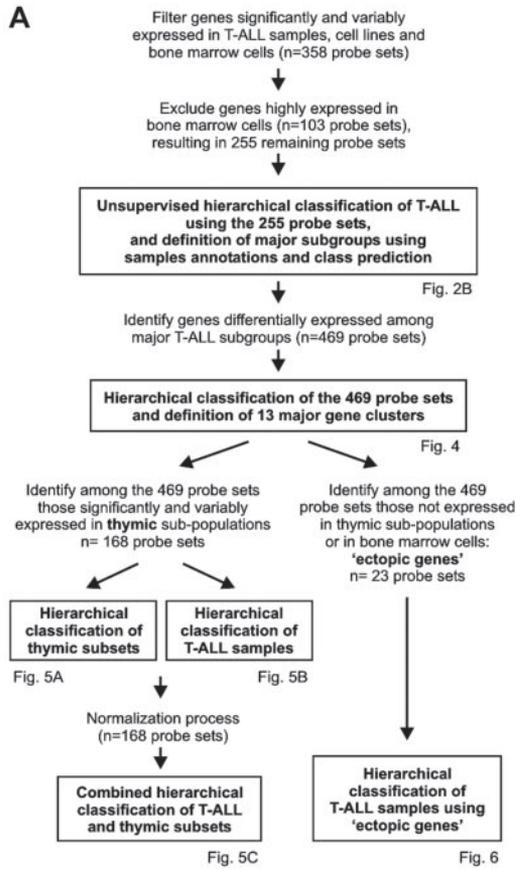
significant *TAL1* expression are found in this branch; see also Figure S1). The cell-surface TCR $\alpha\beta$  receptor is frequently expressed ( $P < .001$ , 16 of 20  $\alpha\beta$  T-ALL cases are included in this branch but no  $\gamma\delta$  cases), in line with previous work connecting *SIL-TAL1* rearrangement with  $\alpha\beta$  T-cell differentiation.<sup>19,28,29</sup> Furthermore, this analysis delineates 2 stable subgroups (*TAL\_RA* and *TAL\_RB*) not identified by previous works.

The second branch of the hierarchy includes all cases with *TLX1* and *TLX3* expression ( $P < .001$  in both cases). All the  $\gamma\delta$  T-ALL cases are included in this branch ( $P < .001$ , 14 cases). In addition to the *TLX1*- and *TLX3*-expressing cases, 2 clusters of samples are found in this branch. A first cluster contains 8 cases, including the 2 *HOXA*-translocated cases. Stability of this cluster was demonstrated by sampling analysis (Figure S2). The genes that mainly define this cluster are *HOXA5*, *HOXA9*, *HOXA10*, and *LRIG1*, and accordingly, this subgroup of samples was provisionally named the "*HOXA*-expressing cluster" (Figure 2B). These cases, and the *TLX1*- and the *TLX3*-expressing samples, were collectively named the *HOX\_R* group. Another cluster of samples is characterized by

**Figure 2. Unsupervised hierarchical classification of samples and genes based on expression microarray data.** (A) General flowchart describing the sequence of microarray analysis procedures throughout the study. (B) This bidimensional classification was performed using expression data of probe sets selected on expression variability and reliability through the samples ( $n = 358$  probe sets), and exclusion of probe sets highly expressed in bone marrow cells; the resulting list of 255 probe sets is shown in Table S1. The top panel shows a hierarchical tree of samples (columns), sample subgroups, and sample annotations. Major subgroups corresponding to main hierarchy branches are indicated on the tree. *MLL* indicates cases with *MLL* rearrangements; *TAL\_R*, *TAL1*-related cases, with 2 *TAL\_R* subgroups (*TAL\_RA* and *TAL\_RB*); *HOX\_R*, *HOX*-related cases; the *HOXA*-expressing cluster of cases is highlighted in red, and the main *TLX1*- and *TLX3*-expressing subgroup is indicated; IMMATURE, subgroup of cases characterized by strong expression of genes expressed in immature cells and frequent coexpression of myeloid genes; BM, normal human bone marrow cells. TL is the unique identification number for samples. Sample group annotations: for genomic annotations, S indicates cases with *SIL-TAL1* transcripts; E, *TLX1*-expressing cases, L, *TLX3*-expressing cases (both also quoted as genomic annotations for simplicity); t, *HOXA*-rearranged cases; C, cases with *CALM-AF10* transcripts; M, cases with *MLL* rearrangement as detected in these cases by FISH and Southern blot; and N, cases with *NUP214-ABL* transcripts. \**HOXA\_R* cases without identified rearrangement. For *TAL1*, *LMO1*, and *LMO2* gene expression, RQ-PCR evaluations of expression are indicated; - indicates cases not annotated to avoid bias due to erythroid contamination ( $n = 9$  cases, including the *HOXA*-rearranged cases TL46), or not available ( $n = 2$  cases). *TAL1* expression was scored for significant increased levels compared to normal thymus level, evaluated from 1 to 5 (moderate to highest levels), and significant *LMO1* and *LMO2* expressions compared to normal thymus levels are indicated as P (positive); see Figure S1. Immuno indicates immunologic markers: i, immature; 3, cCD3 expression; g,  $\gamma\delta$  expression; a,  $\alpha\beta$  expression; and M, myeloid markers (CD13 or CD33 or both). For oncogenic groups, T indicates *TAL\_R* cases; H, *HOX\_R* cases; I, immature cases without *TLX3* expression. For *TAL\_R* subgroup annotations, A indicates *TAL\_RA*; and B, *TAL\_RB*. These labels were assigned based on sample annotations and microarray analyses (Figure S2). Group assignment by prediction models was questionable in 2 *TAL\_RB* cases and 3 immature cases and these cases are indicated by boxes (TL25, TL34, TL76, TL77, and TL82; Figure S2D). The "immature group" label is provisional because oncogenic events are unidentified in this group. In the middle panel, left, hierarchical tree of genes (rows); center, for all samples (columns), relative expression levels are indicated according to the color scale shown on the bottom of the figure (from deep blue, lower expression, to deep red, higher expression); right, rows corresponding to the *HOXA* genes are indicated. The bottom panel shows magnification of the cluster of genes defining the *HOXA*-expressing cluster. These cases, and the 5 additional samples with *HOXA* expression, are indicated by red boxes (*HOXA*-expressing samples).

strong expression of genes expressed in immature cells (eg, *CD34* or *LMO2*) and frequent lymphoid and myeloid CD13 or CD33 surface antigen coexpression ( $P < .001$ , 9 of 14 cases, excluding the *TLX3* samples). Considering that oncogenic events are uncharacterized in these cases but probably occurred at an early stage of

differentiation (with lymphoid and myeloid potential), this group was provisionally named “Immature group” and constitutes the third major group of T-ALL. Half the *TLX3*-expressing samples cocluster with cases of this group, due to expression of immature genes.



Therefore, this analysis identifies a “*HOXA*-expressing cluster” of T-ALL that includes the 2 *HOXA*-rearranged samples. These cases cluster in the same branch as *TLX1* and half the *TLX3*-related cases, suggesting a biologic link between these leukemias.

### Genomic and expression analysis links *HOXA* expression to *TCRB-HOXA*, *MLL*, and *CALM-AF10* chromosomal rearrangements

We investigated the 8 cases of the *HOXA*-expressing cluster defined by the unsupervised hierarchy, and 5 additional samples with detectable *HOXA* expression (*HOXA*-expressing samples,  $n = 13$ ; Figure 2B). FISH, Southern blot, and reverse transcription-PCR (RT-PCR) were performed. Clinical, immunologic, and genetic features of the corresponding patients are indicated in Table 1.

In line with previous work,<sup>30,31</sup> *MLL* rearrangements were found in 3 cases, which clustered as an independent homogeneous group (Figure 2B). In addition and unexpectedly, 4 *HOXA*-expressing cases demonstrated a *CALM-AF10* translocation, as detected by RT-PCR.

Considering that the *HOXA*-expressing cases included 2 cases (TL43 and TL44) with a *TCRB-HOXA* translocation, we searched for additional *HOXA* rearrangements in this subgroup (Table 1). Two more *TCRB-HOXA* translocations were identified using FISH (cases TL45 and TL46). The *HOXA* breakpoint location was mapped by Southern blot analysis in the 4 cases, and the TL46 translocation was fully characterized by sequencing of the 2 derivative breakpoint regions, in addition to case TL43 (Figure S3). Strikingly, the 4 breakpoints clustered in a genomic region of 2.6-kb length located between the *HOXA10* and *HOXA9* genes (*BCR<sub>HOXA</sub>*; Figure 3A). In the final 2 *HOXA*-expressing cases (TL47 and TL48), no *HOXA* rearrangement was detected, suggesting additional mechanisms of *HOXA* activation, as has also been reported for *TAL1* and *TLX1* activation.<sup>35-37</sup> The 4 *TCRB-HOXA* cases were collectively named the *HOXA*-related (*HOXA\_R*) subgroup, along with cases TL47 and TL48, because these 2 cases most frequently clustered with *HOXA*-rearranged cases in clustering experiments (Figure S2).

To analyze the consequence of the translocations on the expression of the *HOXA* locus, all 11 *HOXA* genes, as well as the

non-*HOX* homeobox gene *EVX1* located in the vicinity of the *HOXA* locus, were analyzed for expression using specific RQ-PCR. All cases with *TCRB-HOXA*, *CALM-AF10*, and *MLL* rearrangements consistently demonstrated global overexpression of most *HOXA* genes (but no *EVX1* deregulation), compared to levels in normal thymus and control T-ALLs (Figure 3B). Strikingly, *HOXA* genes located on both sides of the *TCRB-HOXA* breakpoints were overexpressed. In addition, whereas we failed to amplify any potential *TCRB-HOXA* fusion transcript (data not shown), we found that a *HOXA10* gene transcript, referred to as *HOXA10B*, was present at high levels in all the 4 *HOXA*-rearranged cases and in the 2 additional *HOXA\_R* cases (Figure 3B). By contrast, no consistent expression of this *HOXA10B* transcript was detected in the thymus and other samples.

Therefore, the combination of large-scale expression, molecular cytogenetic, and gene-specific expression analyses allowed detailed description of the new recurrent *HOXA* chromosomal rearrangement. This rearrangement strongly deregulates the *HOXA* multigene cluster. In addition, a new association was detected between the *CALM-AF10* rearrangement and expression of *HOXA* genes.

### Inclusion of *HOXA*-expressing cases in a general molecular portrait of T-ALL

To analyze gene pathways differentially involved in T-ALL (including those of the *HOXA\_R* subgroup), we studied expression profiles of the oncogenic T-ALL groups. First, a compendium of 469 probe sets differentially expressed between the previously defined T-ALL groups was obtained (see “Microarray experiments and data analysis” in the Supplemental Materials and methods). Then, a supervised hierarchical classification of the genes was performed, allowing definition of 13 gene clusters based on correlated expression in the T-ALL groups (Figure 4; Table S2 provides a list of genes and clusters). Tentative examination of the known biologic function of the genes shows that a high number of them are linked to thymocyte differentiation or general cellular functions like signaling, cell cycle, DNA repair, and antiapoptotic pathways.

As expected, the *HOXA\_R*, *CALM-AF10*, and *MLL* cases express a cluster of genes containing *HOXA5*, *HOXA9*, *HOXA10*, and *HOXA11* (cluster C9; Figure 4B). *HOXA\_R* cases are also characterized by high average expression of genes, which have

**Table 1. Clinical and biologic features of the *HOXA*-expressing T-ALL cases**

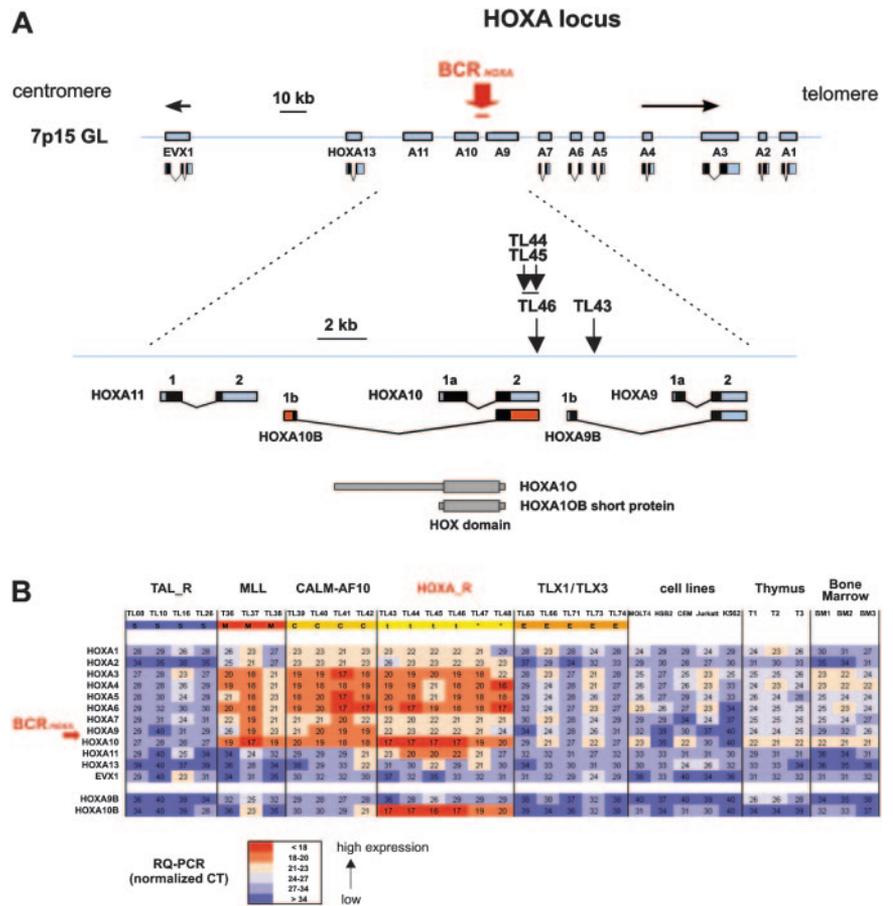
Case ID	Sex	Age, y	WBC count, $\times 10^9/L$	Immunophenotype	Chromosomal rearrangements			<i>CDKN2A</i> deletion
					<i>MLL</i>	<i>CALM-AF10</i>	<i>HOXA</i>	
TL36	M	4	290	$\gamma\delta$	<i>MLL-AF9</i>	—	—	ND
TL37	F	16	199	i	<i>MLL-ENL</i>	—	—	—
TL38	M	5	41	i	<i>MLL-ENL</i>	—	—	—
TL39	M	3	147	$\gamma\delta$	—	<i>CALM-AF10</i>	—	—
TL40	M	17	296	$\gamma\delta$	—	<i>CALM-AF10</i>	—	+
TL41	M	4	215	$\gamma\delta$	—	<i>CALM-AF10</i>	—	+
TL42	F	48	112	i	—	<i>CALM-AF10</i>	—	—
TL43	F	25	85	$\alpha\beta$	—	—	<i>TCRB-HOXA</i>	+
TL44	M	9	412	$\gamma\delta$	—	—	<i>TCRB-HOXA</i>	+
TL45	F	43	159	$\alpha\beta$	—	—	<i>TCRB-HOXA</i>	+
TL46	M	13	17	$\alpha\beta$	—	—	<i>TCRB-HOXA</i>	—
TL47	F	29	109	$\gamma\delta$	—	—	—*	+
TL48	M	39	245	$\gamma\delta$	—	—	—*	+

Chromosomal rearrangements were investigated by *MLL* and *HOXA* FISH and Southern blot, and by *CALM-AF10* RT-PCR. *MLL* partners were further identified in *MLL*-rearranged cases as described in “Materials and methods.” In addition, deletion of the *CDKN2A* (*CDKN2A/INK4A/ARF*) gene was analyzed by FISH.

WBC indicates peripheral white blood cells; M, male;  $\gamma\delta$ , TCR $\gamma\delta$  expression; —, no rearrangement detected; ND, not done; F, female; i, immature immunophenotype; +, *CDKN2A* deletion;  $\alpha\beta$ , TCR $\alpha\beta$  expression.

\*These cases were classified as *HOXA*-related considering their *HOXA* dysregulated pattern and frequent coclustering with the *TCRB-HOXA* rearranged cases.

**Figure 3. Clustering of the breakpoints and dysregulated expression of the HOXA genes in leukemias with TCRB-HOXA rearrangement.** (A) The 4 breakpoints cluster in a 2.6-kb breakpoint cluster region (BCR<sub>HOXA</sub>) as indicated. The direction of transcription is shown by arrows (centromere to telomere for the HOXA genes, and opposite direction for EVX1). Transcription products of the HOXA genes are drawn according to the data of the UCSC site. The HOXA10B exons are shown as red boxes. The rearrangements disrupt the HOXA locus at the 3' end of HOXA10 gene (cases TL44, TL45, and TL46), and within the intron 1b region of HOXA9 (TL43), as indicated. Open reading frames are shown as black boxes. The 2 isoforms of the HOXA10 protein (HOXA10A, quoted HOXA10, and HOXA10B) are drawn according to previous reports, and the homeobox DNA-binding domain is indicated.<sup>32-34</sup> (B) Analysis of expression of the HOXA genes by specific real-time RT-PCR in T-ALL samples. The HOXA and EVX1 expression values are shown in the 4 HOXA-rearranged cases (marked t), and the 2 HOXA-related cases without detected chromosomal rearrangement (TL47 and TL48, marked \*), which constitute the HOXA\_R group. Values in the 3 MLL-rearranged cases, the 4 CALM-AF10 cases, a panel of TAL\_R and TLX1/TLX3 samples, T-cell lines, normal human thymus, and normal bone marrow samples are shown. HOXA9B and HOXA10B transcripts were also analyzed (bottom rows). Median normalized Ct values are indicated and a color scale was used as shown.



been involved in T-cell differentiation or oncogenesis (or both), including *NOTCH2*, *NOTCH3*, *PTCRA* (Pre-T $\alpha$  gene), *TCRG*, *TCRD*, and *PIMI1* (clusters C7, C8, and C11), also shared with the TLX1 and TLX3 subgroups. Together with the fact that HOXA-rearranged, TLX1-, and most of the TLX3-expressing cases cluster in the unsupervised hierarchy (Figure 2B), these data reinforce the view that these subgroups share biologic similarities by using common pathways involved in thymocyte differentiation and, most likely, in T-cell oncogenesis.

In addition to the expression of HOXA genes, the CALM-AF10- and MLL-related cases are characterized by a high expression of the *BMI1* gene (Figure 4B; *Bmi1* expression was confirmed by specific RQ-PCR, not shown). *Bmi1* has been implicated in experimental murine B-cell leukemias<sup>38</sup> and in self-renewal of normal and leukemic stem cells.<sup>39</sup> This gene is considered a repressor of the HOX genes,<sup>40</sup> which contrasts intriguingly with the high expression of HOXA genes in the MLL- and CALM-AF10-related cases (Figure 4B).

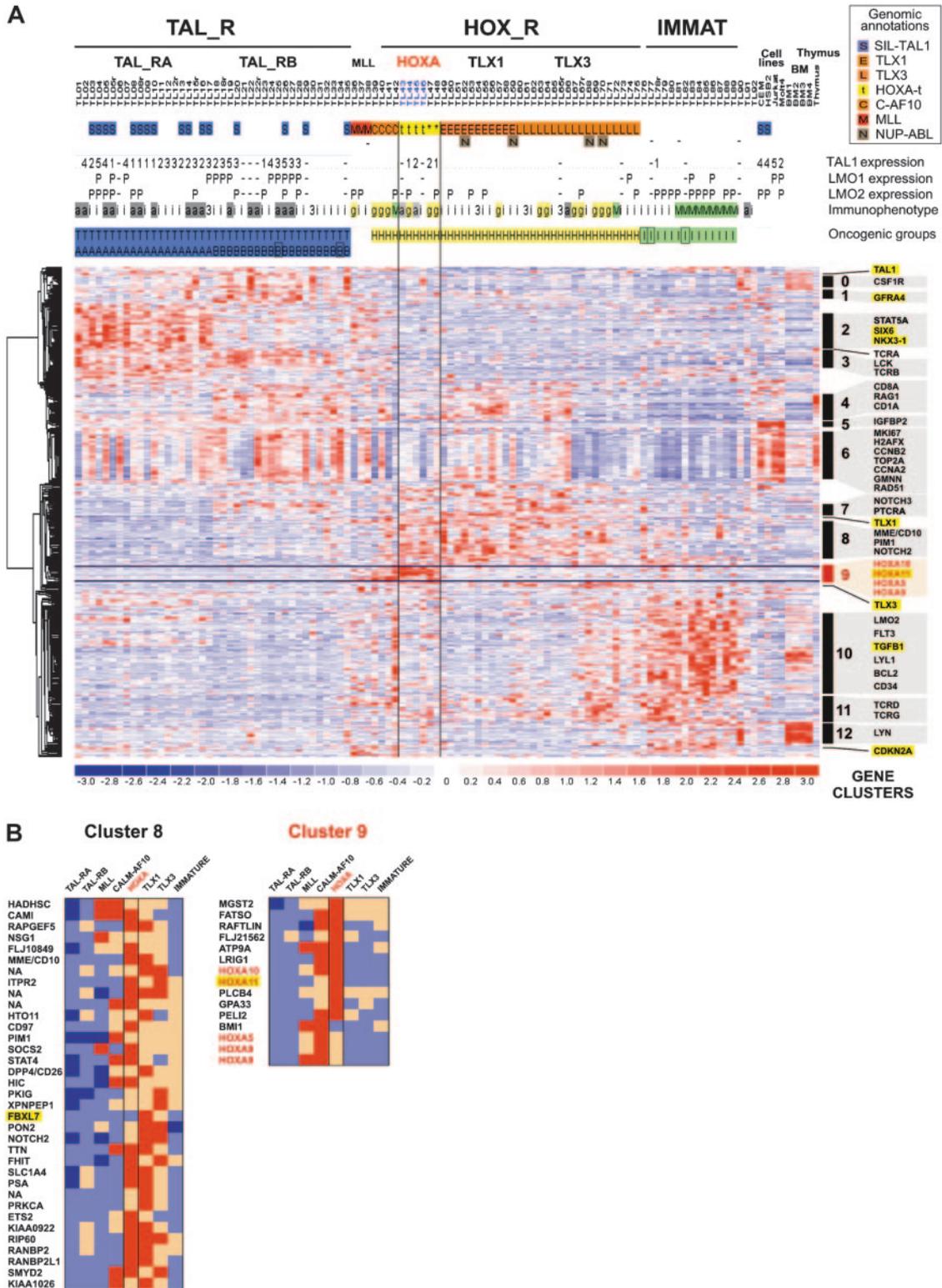
By comparison with the HOXA\_R samples, and more generally with the HOX\_R group, samples from the immature group strongly express the *LMO2*, *LYL1*, *CD34*, *BCL2*, *FLT3*, and *TGFB1* genes (cluster C10, "hematopoiesis") but have low expression of genes from other clusters, particularly from cluster C6, which appears linked to "proliferation and mitosis," whereas there is a variable expression of this cluster throughout samples in the other groups (Figure 4A). Half the TLX3-expressing cases coexpress genes characteristic of both the immature and HOX\_R groups, and accordingly these cases cluster with immature cases. Samples from the TAL\_R group express genes involved in TCRB signaling (cluster C3), particularly in the TAL\_RB samples. In addition, the TAL\_RB samples share with HOX\_R cases (particularly TLX1

cases) a high expression of T-cell differentiation genes (cluster C4 and C7), including *NOTCH3* and *PTCRA*. Samples from the TAL\_RA subgroup specifically express the signaling genes *STAT5A* and *STAM1* (cluster C2). Interestingly, numerous genes not normally expressed in thymus have ectopic expression in TAL\_R samples, including 2 homeobox genes, *NKX3-1* and *SIX6*.

**Analysis of oncogenic T-ALL subgroups with respect to normal T-cell differentiation**

At this point, we analyzed gene expression in T-ALL samples with respect to normal thymic differentiation. For this purpose, 11 distinct thymic subpopulations were purified from normal human thymus, and global gene expression analysis using U133A microarrays was performed. The 469 probe sets differentially expressed between oncogenic groups (Figure 4A) were filtered for thymic expression and proliferation genes were excluded, resulting in 168 genes (see "Microarray experiments and data analysis" in the Supplemental Materials and methods).

First, an unsupervised hierarchy of the thymic subpopulations and the 168 genes was performed. Importantly, the sample classification recapitulates the sequential steps of thymocyte differentiation (Figure 5A). Moreover, examination of the gene clusters of normal thymic subpopulations frequently shows similar patterns compared to gene clusters of T-ALL (compare Figures 4A and 5A), pinpointing networks whose coordinated expression is conserved in leukemic cells (ie, *FLT3-LMO2-CD34-LYL1* and *CD1A-CD8A-RAG1*). In rarer cases, different gene clustering suggests that abnormal coexpression in T-ALL samples could be important in the



**Figure 4. Supervised clustering of genes differentially expressed among major sample groups.** (A) This hierarchical classification was performed using 469 probe sets detecting differential expression among sample groups (Figure 2A). Samples were preclassified according to oncogenic group assignment as defined in Figure 2B. The same conventions for sample annotations are used. The HOXA\_R subgroup is indicated (*HOXA*-rearranged,  $n = 4$ , and *HOXA*-related,  $n = 2$ ). Clusters of genes defined by this supervised clustering are indicated (clusters 0-12). Genes of particular interest are labeled using Locuslink symbol, and genes without significant thymic expression are highlighted in yellow (ectopic thymic expression). The cluster C9, which is defined by *HOXA*-expressing samples and contains *HOXA* genes, is highlighted in red. (B) List of genes from clusters C8 and C9 and relative expression in oncogenic subgroups. Genes from cluster C8 are frequently highly expressed in subsets from the HOXA\_R group; cluster C9 is defined mainly by the *HOXA*-expressing samples (HOXA\_R, CALM-AF10, and MLL) and includes *HOXA* genes (in red); "ectopic" genes with respect to thymus expression are highlighted in yellow. NA indicates Locuslink symbol not available. Relative expression in each T-ALL group is shown by color scale.

oncogenic process. For instance, *PTCRA* and *NOTCH3* are coexpressed in subgroups of T-ALL samples, including HOX\_R samples (Figure 4A), whereas in normal thymic subpopulations *PTCRA* is expressed at the highest level before the  $\beta$ -selection process and *NOTCH3* afterward (Figure 5A).

Then, by analyzing leukemic samples using the 168 “thymic genes,” we found that a consistent hierarchy of T-ALL could be obtained (Figure 5B; compare with Figure 2B). Therefore, to define the differentiation stages of leukemic cells according to biologic pathways, T-ALL and thymic subpopulations data were combined. Hierarchical analysis and multidimensional scaling visualization were performed and informative coclustering of T-ALL subgroups and thymic subsets was observed (Figure 5C-D). The HOX\_R samples cluster with the Pre-T1 and Pre-T2/Pre-T3(E) subpopulations, together with the *TLX1*- and half the *TLX3*-expressing samples. Therefore, these samples appear arrested at a stage corresponding to T-lineage engagement preceding the  $\beta$ -selection process. By contrast, the TAL\_R samples cluster with subpopulations corresponding to steps following  $\beta$ -selection. Samples from the immature group, as well as the *CALM-AF10*, *MLL*, and half the *TLX3*-expressing leukemic samples, cluster with the most immature ETP and Pro-T subpopulations, which correspond to the lineage choice process. Notably, further examination of gene expression in leukemic samples with respect to normal subpopulations reveals numerous abnormalities in the differentiation process (in addition to the differentiation arrest itself). For instance, a few HOX\_R samples, including 2 *TCRB-HOXA* cases, unexpectedly coexpress *CD10/MME* and TCR $\alpha\beta$  heterodimer (Figure 4B). Strikingly, numerous T-cell differentiation-associated genes are expressed at dramatically lower levels in T-ALL samples compared with the relevant normal thymic subpopulations, such as *CD8A*, *RAG1*, *CD1s*, *PTCRA*, and *TCRA* in the TAL-R cases (Figure 5C; *CD8A* and *RAG1* expression was confirmed in thymic subsets and T-ALL samples using RQ-PCR, not shown). Conversely, several genes are expressed at higher levels in T-ALL than in normal thymic subsets, such as *STAT5A* in TAL\_RA cases.

#### Oncogenic groups of T-ALL can be classified using “ectopic” gene expression

Because subgroups of T-ALL can be easily classified by genes normally expressed during thymic differentiation, we wondered whether the unsupervised clustering of T-ALL samples defined in this work merely reflects the differentiation status of the cells, or whether it may also specifically highlight oncogenic events. The 469 probe sets differentially expressed between oncogenic groups (Figure 4A) were filtered for ectopic expression with reference to expression in normal thymic subpopulations and bone marrow (considering that a fraction of T-ALL cells coexpress immature and myeloid-associated genes), resulting in a working list of 23 “ectopic” genes (see “Microarray experiments and data analysis” in the Supplemental Materials and methods). An unsupervised classification of the leukemic samples was performed using these genes. Strikingly, the resulting hierarchy properly classifies the major T-ALL oncogenic subgroups (Figure 6).

Importantly, this gene-filtering strategy was able to select the *TLX1* and *TLX3* genes and a *HOXA* gene (*HOXA11*), all genes that can be targeted by chromosomal translocations as demonstrated here for the *HOXA* locus. Note, however, that some genes linked to oncogenesis were excluded due to their expression in certain thymic subpopulations or in bone marrow, such as *TAL1*, *LMO2*, and the *HOXA* genes except *HOXA11*. Interestingly, 2 homeobox genes, *NKX3-1* and *SIX6*, are included in the list of ectopic genes,

and they are expressed in TAL\_R cases. Notably, a gene related to *NKX3-1*, namely *NKX2-5*, has been involved in chromosomal translocations in T-cell lines.<sup>41</sup> Additional genes involved in development or morphogenesis were also selected, including *TFAP2C* and *GDF10* (BMP family). Expression level of *NKX3-1*, *SIX6*, and *TFAP2C* was confirmed by specific RQ-PCR in representative T-ALL samples. It therefore appears that ectopic expression of at least one developmental gene is found in most leukemic cases from the major HOX\_R and TAL\_R groups.

These results demonstrate that the subgroups of T-ALL are not only defined by their differentiation status but also by ectopic expression of genes, including developmental genes.

## Discussion

### The *HOXA* gene cluster is an oncogenic target in T-ALL

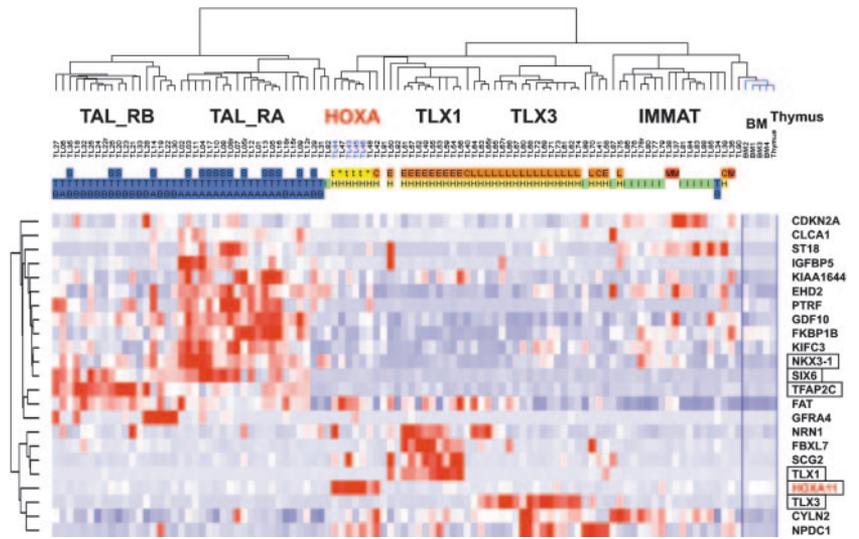
By combining molecular cytogenetic and expression analysis approaches, we have identified and characterized a new recurrent molecular rearrangement involving the *HOXA* and *TCRB* loci in T-ALL. Interphasic FISH patterns and breakpoint sequencing were consistent with chromosomal inversion inv(7)(p15q34) in cases TL44, TL45 and TL46, and translocation t(7;7)(p15;q34) in case TL43 (no karyotypic data were available). Notably, such chromosomal rearrangements have been recurrently reported in T-ALL,<sup>42</sup> (see also <http://cgap.nci.nih.gov/Chromosomes/Mitelman>), but it was probably assumed that they would always reproduce in T-ALL the *TCRG-TCRB* rearrangements frequently found in T cells in patients with ataxia-telangiectasia.<sup>43,44</sup>

The *HOXA* gene cluster is one of the 4 major homeobox gene clusters. Each cluster displays 9 to 11 paralogous *HOX* genes, with a conserved structural organization and a tightly regulated expression.<sup>45-48</sup> We found that the disruption of the *HOXA* locus by *TCRB* chromosomal rearrangements was associated with a global overexpression of the *HOXA* genes (Figure 3B). This may be due to the juxtaposition of *TCRB* regulatory elements in the vicinity of the *HOXA* genes<sup>7,49</sup> and also to the separation of the locus into 2 parts, which may disrupt the normal regulatory elements of the cluster. Alternatively, nonconventional deregulation mechanisms involving miRNA might also be involved<sup>50</sup> or, at a higher level, the new chromosome structure resulting from the rearrangement might induce global changes in the nucleus and chromatin organization.<sup>51</sup> Interestingly, the 4 *HOXA* breakpoints cluster tightly at the vicinity of the *HOXA10* and *HOXA9* genes (Figure 3A). Whereas we could not detect a *TCRB-HOXA* fusion transcript, strong expression of an alternative transcript known as the *HOXA10B* transcript was detected in addition to the global *HOXA* gene overexpression (Figure 3A-B). This transcript, also called the *HOXA10* short transcript, has been found in embryonic cells and myeloid cell lines but not in normal hematopoiesis.<sup>32,33,52</sup> It encodes a short protein that contains the *HOXA10* homeobox DNA-binding domain but lacks the N-terminal regulatory region and that may have distinct functional properties.<sup>33,34</sup> The *TCRB-HOXA* translocation is, to our knowledge, the first example of a recurrent chromosomal rearrangement associated with such a direct and major dysregulation of a multigenic locus.

*MLL*-translocated cases of the present series demonstrate a strong expression of the *HOXA* genes, consistent with previous reports.<sup>30,31</sup> The direct targeting of the *HOXA* locus by chromosomal rearrangement strongly supports the hypothesis that dysregulation of *HOXA* gene expression, triggered by *MLL* fusion proteins, is causal in *MLL*-related oncogenesis, consistent with in vitro properties of *MLL* proteins.<sup>53-56</sup> The fact that *MLL* and HOX\_R groups do not cluster in unsupervised



**Figure 6. Unsupervised clustering of T-ALL samples based on 'ectopic' genes.** The 469 probe sets differentially expressed between oncogenic groups (Figure 4A) were filtered for ectopic expression with reference to expression in normal thymic subpopulations and bone marrow, resulting in a working list of 23 ectopically expressed genes. Genes known to be involved in development are boxed. The main T-ALL oncogenic subgroups and annotations are indicated as in other figures.



hierarchy may suggest that additional genes targeted by the *MLL* fusions could play an important role in oncogenesis. A major oncogenic role of *HOXA* genes in T-ALL with *CALM-AF10* translocation is also suggested by our results.

The *HOX* gene network is involved in embryogenesis, stem cell renewal, cell fate, and control of cell differentiation in many tissue types including hematopoietic system and thymus.<sup>45,57-60</sup> The evidence increasingly suggests that *HOX* genes might play a central role in oncogenesis, including in leukemia.<sup>2,20,58,61-65</sup> In myeloid leukemias, several *HOX* genes including *HOXA9* can be involved in gene fusion with *NUP98*, leading to chimeric proteins.<sup>66-68</sup> The present identification of the *TCRB-HOXA* rearrangement in human leukemia, involving the prototypical homeobox genes *HOXA*, strongly reinforces the view that abnormal expression of developmental genes can be causal in human cancer.

**Large-scale expression analysis defines complete typology of T-ALL**

Unsupervised classifications allow definition of the overall typology of samples, without a priori knowledge of the sample annotations.<sup>69</sup> New homogeneous subgroups can be identified using these methods, as shown here in the *HOXA*-rearranged cases. Importantly, comparison of the sample annotations with the distinct oncogenic groups in our series of 92 cases suggests that a complete model of T-ALL has been here delineated. Major stable groups have been defined by unsupervised analysis on a large number of samples and genes, and this classification makes sense because it pinpoints links between these groups and genetic rearrangements, ectopic expression of genes including

known oncogenes, and critical stages of thymus differentiation. It is therefore likely that potential new subgroups in T-ALL would represent discreet variations within these major groups or rare distinct subgroups.

Using a supervised approach based on oncogene expression, Ferrando et al identified a set of genes differentially expressed in T-ALL samples.<sup>19</sup> An unsupervised classification of our samples using a list of genes based on this set properly classifies the major oncogenic groups (Figure S4). It is likely that the *LYL1* group defined in this previous work partially overlaps with our provisionally named immature group (note, however, that *LYL1* is expressed in the corresponding normal thymic immature subpopulation). Importantly, 3 cases here identified as *HOXA*-translocated cocluster using this former set of genes, within a branch containing the vast majority of *TLX1* and *TLX3* cases (Figure S4). Comparison of these 2 studies demonstrates the reliability of the gene expression approach and emphasizes the view that *TCRB-HOXA*-rearranged cases represent a homogeneous T-ALL subgroup.

**HOXA genes are included in biologic networks that connect oncogenesis, thymic differentiation pathways, and ectopic gene expression**

Important data were obtained by analyzing expression profiles of leukemic samples with respect to those of normal human thymocyte subpopulations. Although oncogenic processes dramatically alter levels of gene expression, we found that conserved gene expression networks allow pertinent coclustering of leukemic and thymic samples (Figure 5C). T-ALL cells of the major groups (*TAL\_R*, *HOX\_R*, and immature) appear arrested at distinct stages of differentiation, as previously suggested,<sup>19,28,29,70</sup> corresponding

**Figure 5. Unsupervised clustering of human thymocyte subpopulations and leukemic cells.** (A) Unsupervised classification of 11 subpopulations purified from normal human thymus and subsequently analyzed by global gene expression on the U133A array. For human thymic subsets, ETP indicates early thymocyte precursor with lymphomyeloid potential<sup>25,26</sup>; the Pro-T, Pre-T1, Pre-T2/Pre-T3(E) (ie, Pre-T2 and Pre-T3(E), not distinguished here), Pre-T3(L), immature DP (immature double-positive cells), DP-TCR $\alpha\beta$ lo, DP-TCR $\alpha\beta$ int and DP-TCR $\alpha\beta$ br, CD4-SP (CD4 single-positive cells), and CD8-SP, have been described previously (for a review, see Carrasco et al<sup>4</sup>). The lineage choice,  $\beta$ -selection, and selection processes as reported in normal thymus differentiation are indicated. The 469 probe sets differentially expressed between oncogenic groups (Figure 4A) were filtered for variable and reliable thymic expression. In addition, "proliferation genes" defined by the "proliferation-mitosis" cluster (cluster C6 in Figure 4A) were discarded from this analysis to avoid spurious sample clustering due to proliferation rate. The resulting final list of 168 "thymic" genes is shown in Table S3. Genes of particular interest are shown, and *HOXA* genes are indicated in red. (B) Unsupervised classification of T-ALL samples using the same 168 "thymic" gene list. The same conventions are used as in Figure 2. The main oncogenic subgroups are indicated. The same genes of interest as in Figure 5A are shown. (C) Unsupervised analysis of combined T-ALL samples, normal human thymic subpopulations, and bone marrow (BM) cells. A selection of genes with a strongly altered expression level compared with the corresponding normal thymic subsets is indicated (*RAG1* and *CD8A* gene expression was also directly analyzed by RQ-PCR in normal thymic samples and T-ALL, giving consistent data; not shown). (D) Three-dimensional representation of a multidimensional scaling analysis performed using normal thymic subsets, normal bone marrow, and T-ALL samples of the main oncogenic subgroups (*TAL\_RA*, *TAL\_RB*, *HOXA*, *TLX1*, *TLX3*, immature). The lineage choice,  $\beta$ -selection, and selection processes are indicated.

to critical events in thymic maturation. Specifically, the *HOXA*-rearranged and *TLX1*-expressing cases closely match differentiation stages of Pre-T1 to Pre-T2/Pre-T3(E) subsets, whereas *TLX3*-expressing cases extend from ETP to Pre-T2/Pre-T3(E), all stages preceding the  $\beta$ -selection process. This correlation between oncogene expression and arrest at precise stages of differentiation may suggest that expression of distinct oncogenes is associated with the inhibition of distinct differentiation pathways. For instance, it has been reported that ectopic expression of TAL1 oncoprotein may perturb T-cell differentiation by disturbing the expression of E2A target genes.<sup>71-73</sup> Consistently, we found in TAL\_R samples that many genes characteristic of steps later than  $\beta$ -selection have dramatic underexpression, although these samples cocluster with thymic subpopulations corresponding to these late stages (Figure 5C). It therefore appears that these samples do not display a complete  $\beta$ -selection pattern. One can hypothesize that protection mechanisms have been selected in evolution to avoid the generation of tumors at this stage normally associated with pre-TCR signaling and a resulting high proliferation rate. For instance, RAG2 phosphorylation inhibits V(D)J recombination in cycling thymocytes,<sup>74</sup> and this may protect highly proliferative cells in  $\beta$ -selection from a major mechanism of oncogenic chromosomal rearrangement.

Once arrested, leukemic cells might continuously activate certain biologic pathways normally transiently used at these stages, like antiapoptotic pathways. In addition, we found a distortion in expression of certain genes critical for thymic differentiation. For instance, *NOTCH2*, *NOTCH3*, and *PTCRA* genes are coexpressed in most human T-ALL cases (Figure 4A), as reported.<sup>75</sup> Interestingly, these genes cocluster in T-ALL samples but not in normal thymic subpopulations, suggesting that an abnormal coexpression of these genes could be specifically associated with oncogenesis. Consistent with a major oncogenic role of the activated *NOTCH* genes, *NOTCH1* is frequently targeted by activating mutations (or rare chromosomal translocation) in human T-ALL,<sup>11,13</sup> and expression of a constitutively active intracellular domain of Notch3 favors murine T-cell leukemia in a *Ptcra*-dependent manner.<sup>75</sup> Interestingly, 2 of the 4 *HOXA*-rearranged cases here identified also have prototypical *NOTCH1* DNA mutation (leading to L1601P and H1602L+1603insP mutated proteins in cases TL43 and TL45, respectively, data not shown), demonstrating that *HOXA* and *NOTCH1* genomic abnormalities can be associated and probably cooperate in these cases.

Finally, it appears from our data that a number of genes, including developmental genes, have abnormal, frequently ectopic, expression in certain T-ALL samples compared with normal thymic subpopulations. These include *HOXA*, *TLX1*, and *TLX3* genes, which can be targeted by recurrent chromosomal translocation as demonstrated here for the *HOXA* genes, and *CDKN2A/INK4A/ARF*, frequently inactivated by DNA mutation or deletion as a tumor suppressor gene.<sup>17,18</sup> Ectopic expression in T-ALL samples of other developmental genes, such as *NKX3-1*, *SIX6*, and *TFAP2C* suggests that these genes might also be involved in T-ALL oncogenesis (Figure 6). Interestingly, expression of some of these genes has been previously linked to cancer.<sup>76,77</sup> Notably, both the *NKX2-5* gene (homeobox gene paralogous to *NKX3-1*) and the *TALI* gene are targeted by chromosomal rearrangements in the T-ALL cell line CEM-CCRF, suggesting potential oncogenic cooperation.<sup>41</sup> Altogether, our results are consistent with a model in which developmental genes, including the prototypical *HOXA* genes in some leukemia subgroups, play a central role in T-ALL oncogenesis.<sup>63</sup>

## Acknowledgments

This paper reports data of the pilot project of the Carte d'Identité des Tumeurs program from the Ligue Nationale Contre le Cancer.

We thank all physicians who sent leukemic samples to the Hematology Laboratory, and especially Thierry Leblanc and Emmanuel Raffoux. We gratefully thank F.A.B. members Georges Flandrin and Marie-Thérèse Daniel for expert morphologic evaluation of leukemic samples. We thank the Xavier Fund for RNA extraction and contribution to oncogene expression analysis, Marie Romeo for CALM-AF10 and BMI1 transcript evaluation, Claire Pichereau for contribution to FISH analysis, and Antoine Crinquette for contribution to Southern blot experiments. We thank Philippe Dessen (Institut Gustave Roussy, Villejuif, France) for updates of gene annotations. We also thank Christelle Thibault, Philippe Kastner, and Doulaye Dembelé (Strasbourg Genopole National Facility, Illkirch, France), Fabien Patel (CIT program, Ligue Nationale contre le Cancer), and Elisabeth Savariau and Robin Nancel (Service d'Infographie, Institut Universitaire d'Hématologie, Paris). We thank Catharine M. Green, Jacques Haiech, Alain Aurias, and Didier Auboeuf for critical comments on the manuscript.

## References

- Boucheix C, David B, Sebban C, et al. Immunophenotype of adult acute lymphoblastic leukemia, clinical parameters, and outcome: an analysis of a prospective trial including 562 tested patients (LALA87). French Group on Therapy for Adult Acute Lymphoblastic Leukemia. *Blood*. 1994;84:1603-1612.
- Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. *N Engl J Med*. 2004;350:1535-1548.
- Spits H, Blom B, Jaleco AC, et al. Early stages in the development of human T, natural killer and thymic dendritic cells. *Immunol Rev*. 1998;165:75-86.
- Carrasco YR, Navarro MN, de Yébenes VG, Ramiro AR, Toribio ML. Regulation of surface expression of the human pre-T cell receptor complex. *Semin Immunol*. 2002;14:325-334.
- Spits H. Development of alphabeta T cells in the human thymus. *Nat Rev Immunol*. 2002;2:760-772.
- Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100:57-70.
- Rabbitts TH. Chromosomal translocations in human cancer. *Nature*. 1994;372:143-149.
- Ferrando AA, Look AT. Clinical implications of recurring chromosomal and associated molecular abnormalities in acute lymphoblastic leukemia. *Semin Hematol*. 2000;37:381-395.
- Bernard OA, Busson-LeConiat M, Ballerini P, et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. *Leukemia*. 2001;15:1495-1504.
- Rabbitts TH. LMO T-cell translocation oncogenes typify genes activated by chromosomal translocations that alter transcription and developmental processes. *Genes Dev*. 1998;12:2651-2657.
- Ellisen LW, Bird J, West DC, et al. TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell*. 1991;66:649-661.
- Tycko B, Smith SD, Sklar J. Chromosomal translocations joining LCK and TCRB loci in human T cell leukemia. *J Exp Med*. 1991;174:867-873.
- Weng AP, Ferrando AA, Lee W, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*. 2004;306:269-271.
- Rubnitz JE, Camitta BM, Mahmoud H, et al. Childhood acute lymphoblastic leukemia with the MLL-ENL fusion and t(11;19)(q23;p13.3) translocation. *J Clin Oncol*. 1999;17:191-196.
- Asnafi V, Radford-Weiss I, Dastugue N, et al. CALM-AF10 is a common fusion transcript in T-ALL and is specific to the TCR $\gamma$ madelta lineage. *Blood*. 2003;102:1000-1006.
- Graux C, Cools J, Melotte C, et al. Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. *Nat Genet*. 2004;36:1084-1089.
- Cayuela JM, Madani A, Sanhes L, Stern MH, Sigaux F. Multiple tumor-suppressor gene 1 inactivation is the most frequent genetic alteration in

- T-cell acute lymphoblastic leukemia. *Blood*. 1996; 87:2180-2186.
18. Gardie B, Cayuela JM, Martini S, Sigaux F. Genomic alterations of the p19ARF encoding exons in T-cell acute lymphoblastic leukemia. *Blood*. 1998;91:1016-1020.
  19. Ferrando AA, Neuberg DS, Staunton J, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell*. 2002;1:75-87.
  20. Abate-Shen C. Deregulated homeobox gene expression in cancer: cause or consequence? *Nat Rev Cancer*. 2002;2:777-785.
  21. Gabert J, Beillard E, van der Velden VH, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia—a Europe Against Cancer program. *Leukemia*. 2003; 17:2318-2357.
  22. Delabesse E, Bernard M, Meyer V, et al. TAL1 expression does not occur in the majority of T-ALL blasts. *Br J Haematol*. 1998;102:449-457.
  23. Ramiro AR, Trigueros C, Marquez C, San Millan JL, Toribio ML. Regulation of pre-T cell receptor (pT alpha-TCR beta) gene expression during human thymic development. *J Exp Med*. 1996;184: 519-530.
  24. Carrasco YR, Trigueros C, Ramiro AR, de Yebenes VG, Toribio ML. Beta-selection is associated with the onset of CD8beta chain expression on CD4(+)CD8alpha(+) pre-T cells during human intrathymic development. *Blood*. 1999;94: 3491-3498.
  25. de Yebenes VG, Carrasco YR, Ramiro AR, Toribio ML. Identification of a myeloid intrathymic pathway of dendritic cell development marked by expression of the granulocyte macrophage-colony-stimulating factor receptor. *Blood*. 2002; 99:2948-2956.
  26. Garcia-Peydro M, de Yebenes VG, Toribio ML. Sustained Notch1 signaling instructs the earliest human intrathymic precursors to adopt a gamma-delta T-cell fate in fetal thymus organ culture. *Blood*. 2003;102:2444-2451.
  27. Marculescu R, Le T, Simon P, Jaeger U, Nadel B. V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J Exp Med*. 2002;195:85-98.
  28. Macintyre EA, Smit L, Ritz J, Kirsch IR, Strominger JL. Disruption of the SCL locus in T-lymphoid malignancies correlates with commitment to the T-cell receptor alpha beta lineage. *Blood*. 1992;80:1511-1520.
  29. Asnafi V, Beldjord K, Libura M, et al. Age-related phenotypic and oncogenic differences in T-acute lymphoblastic leukemias may reflect thymic atrophy. *Blood*. 2004;104:4173-4180.
  30. Armstrong SA, Staunton JE, Silverman LB, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet*. 2002;30:41-47.
  31. Ferrando AA, Armstrong SA, Neuberg DS, et al. Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of HOX dysregulation. *Blood*. 2003; 102:262-268.
  32. Lawrence HJ, Sauvageau G, Ahmadi N, et al. Stage- and lineage-specific expression of the HOXA10 homeobox gene in normal and leukemic hematopoietic cells. *Exp Hematol*. 1995;23:1160-1166.
  33. Benson GV, Nguyen TH, Maas RL. The expression pattern of the murine Hoxa-10 gene and the sequence recognition of its homeodomain reveal specific properties of abdominal B-like genes. *Mol Cell Biol*. 1995;15:1591-1601.
  34. Eklund EA, Jalava A, Kakar R. Tyrosine phosphorylation of HoxA10 decreases DNA binding and transcriptional repression during interferon gamma-induced differentiation of myeloid leukemia cell lines. *J Biol Chem*. 2000;275:20117-20126.
  35. Bash RO, Hall S, Timmons CF, et al. Does activation of the TAL1 gene occur in a majority of patients with T-cell acute lymphoblastic leukemia? A Pediatric Oncology Group study. *Blood*. 1995;86: 666-676.
  36. Kees UR, Heerema NA, Kumar R, et al. Expression of HOX11 in childhood T-lineage acute lymphoblastic leukaemia can occur in the absence of cytogenetic aberration at 10q24: a study from the Children's Cancer Group (CCG). *Leukemia*. 2003;17:887-893.
  37. Ferrando AA, Herblot S, Palomero T, et al. Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. *Blood*. 2004;103:1909-1911.
  38. Alkema MJ, Wiegant J, Raap AK, Berns A, van Lohuizen M. Characterization and chromosomal localization of the human proto-oncogene BMI-1. *Hum Mol Genet*. 1993;2:1597-1603.
  39. Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature*. 2003;423:255-260.
  40. Hanson RD, Hess JL, Yu BD, et al. Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc Natl Acad Sci U S A*. 1999;96: 14372-14377.
  41. Nagel S, Kaufmann M, Drexler HG, MacLeod RA. The cardiac homeobox gene NKX2-5 is deregulated by juxtaposition with BCL11B in pediatric T-ALL cell lines via a novel t(5;14)(q35.1;q32.2). *Cancer Res*. 2003;63:5329-5334.
  42. Schneider NR, Carroll AJ, Shuster JJ, et al. New recurring cytogenetic abnormalities and association of blast cell karyotypes with prognosis in childhood T-cell acute lymphoblastic leukemia: a Pediatric Oncology Group report of 343 cases. *Blood*. 2000;96:2543-2549.
  43. Stern MH, Lipkowitz S, Aurias A, Griscelli C, Thomas G, Kirsch IR. Inversion of chromosome 7 in ataxia telangiectasia is generated by a rearrangement between T-cell receptor beta and T-cell receptor gamma genes. *Blood*. 1989;74: 2076-2080.
  44. Bernard O, Groettrup M, Mugneret F, Berger R, Azogui O. Molecular analysis of T-cell receptor transcripts in a human T-cell leukemia bearing a t(1;14) and an inv(7); cell surface expression of a TCR-beta chain in the absence of alpha chain. *Leukemia*. 1993;7:1645-1653.
  45. Krumlauf R. Hox genes in vertebrate development. *Cell*. 1994;78:191-201.
  46. Kondo T, Duboule D. Breaking colinearity in the mouse HoxD complex. *Cell*. 1999;97:407-417.
  47. Brend T, Gilthorpe J, Summerbell D, Rigby PW. Multiple levels of transcriptional and post-transcriptional regulation are required to define the domain of Hoxb4 expression. *Development*. 2003;130:2717-2728.
  48. Spitz F, Gonzalez F, Duboule D. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell*. 2003; 113:405-417.
  49. Whitehurst CE, Chattopadhyay S, Chen J. Control of V(D)J recombinational accessibility of the D beta 1 gene segment at the TCR beta locus by a germline promoter. *Immunity*. 1999;10:313-322.
  50. Mansfield JH, Harfe BD, Nissen R, et al. MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet*. 2004;36:1079-1083.
  51. Chambeyron S, Bickmore WA. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*. 2004;18:1119-1130.
  52. Lowney P, Corral J, Detmer K, et al. A human Hox 1 homeobox gene exhibits myeloid-specific expression of alternative transcripts in human hematopoietic cells. *Nucleic Acids Res*. 1991;19: 3443-3449.
  53. Milne TA, Briggs SD, Brock HW, et al. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell*. 2002;10:1107-1117.
  54. Nakamura T, Mori T, Tada S, et al. ALL-1 is a histone methyltransferase that assembles a super-complex of proteins involved in transcriptional regulation. *Mol Cell*. 2002;10:1119-1128.
  55. Yokoyama A, Wang Z, Wysocka J, et al. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol*. 2004;24:5639-5649.
  56. Zeisig BB, Milne T, Garcia-Cuellar MP, et al. Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol Cell Biol*. 2004;24:617-628.
  57. Sauvageau G, Thorsteinsdottir U, Eaves CJ, et al. Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. *Genes Dev*. 1995;9:1753-1765.
  58. Lawrence HJ, Sauvageau G, Humphries RK, Largman C. The role of HOX homeobox genes in normal and leukemic hematopoiesis. *Stem Cells*. 1996;14:281-291.
  59. Izon DJ, Rozenfeld S, Fong ST, Komuves L, Largman C, Lawrence HJ. Loss of function of the homeobox gene Hoxa-9 perturbs early T-cell development and induces apoptosis in primitive thymocytes. *Blood*. 1998;92:383-393.
  60. So CW, Karsunky H, Wong P, Weissman IL, Cleary ML. Leukemic transformation of hematopoietic progenitors by MLL-GAS7 in the absence of Hoxa7 or Hoxa9. *Blood*. 2004;103: 3192-3199.
  61. Nakamura T, Largaespada DA, Shaughnessy JD Jr, Jenkins NA, Copeland NG. Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nat Genet*. 1996;12:149-153.
  62. Thorsteinsdottir U, Sauvageau G, Hough MR, et al. Overexpression of HOXA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. *Mol Cell Biol*. 1997;17:495-505.
  63. Look AT. Oncogenic transcription factors in the human acute leukemias. *Science*. 1997;278: 1059-1064.
  64. Lawrence HJ, Rozenfeld S, Cruz C, et al. Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias. *Leukemia*. 1999;13:1993-1999.
  65. Owens BM, Hawley RG. HOX and non-HOX homeobox genes in leukemic hematopoiesis. *Stem Cells*. 2002;20:364-379.
  66. Borrow J, Shearman AM, Stanton VP Jr, et al. The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nat Genet*. 1996;12:159-167.
  67. Nakamura T, Largaespada DA, Lee MP, et al. Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nat Genet*. 1996;12:154-158.
  68. Fujino T, Suzuki A, Ito Y, et al. Single-translocation and double-chimeric transcripts: detection of NUP98-HOXA9 in myeloid leukemias with HOXA11 or HOXA13 breaks of the chromosomal translocation t(7;11)(p15;p15). *Blood*. 2002;99:1428-1433.
  69. Ebert BL, Golub TR. Genomic approaches to hematologic malignancies. *Blood*. 2004;104:923-932.
  70. Chiaretti S, Li X, Gentleman R, et al. Gene expression profile of adult T-cell acute lymphocytic

- leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood*. 2004;103:2771-2778.
71. Chervinsky DS, Zhao XF, Lam DH, Ellsworth M, Gross KW, Aplan PD. Disordered T-cell development and T-cell malignancies in SCL LMO1 double-transgenic mice: parallels with E2A-deficient mice. *Mol Cell Biol*. 1999;19:5025-5035.
72. Herblot S, Steff AM, Hugo P, Aplan PD, Hoang T. SCL and LMO1 alter thymocyte differentiation: inhibition of E2A-HEB function and pre-T alpha chain expression. *Nat Immunol*. 2000;1:138-144.
73. O'Neil J, Shank J, Cusson N, Murre C, Kelliher M. TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. *Cancer Cell*. 2004;5:587-596.
74. Lin WC, Desiderio S. Regulation of V(D)J recombination activator protein RAG-2 by phosphorylation. *Science*. 1993;260:953-959.
75. Bellavia D, Campese AF, Checquolo S, et al. Combined expression of pTalpha and Notch3 in T cell leukemia identifies the requirement of preTCR for leukemogenesis. *Proc Natl Acad Sci U S A*. 2002;99:3788-3793.
76. Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, et al. Roles for Nkx3.1 in prostate development and cancer. *Genes Dev*. 1999;13:966-977.
77. Boshier JM, Williams T, Hurst HC. The developmentally regulated transcription factor AP-2 is involved in c-erbB-2 overexpression in human mammary carcinoma. *Proc Natl Acad Sci U S A*. 1995;92:744-747.