blood®
Regular Article

LYMPHOID NEOPLASIA

Concurrent *CDX2 cis*-deregulation and *UBTF::ATXN7L3* fusion define a novel high-risk subtype of B-cell ALL

Marie Passet,[1,2] Rathana Kim,[1,2,*] Stéphanie Gachet,[1,*] François Sigaux,[1,2,*] Julie Chaumeil,[3] Ava Galland,[4] Thomas Sexton,[5] Samuel Quentin,[2] Lucie Hernandez,[1] Lise Larcher,[1,2] Hugo Bergugnat,[2] Tao Ye,[5] Nezih Karasu,[5] Aurélie Caye,[6] Beate Heizmann,[5] Isabelle Duluc,[4] Patrice Chevallier,[7] Philippe Rousselot,[8] Françoise Huguet,[9] Thibaut Leguay,[10] Mathilde Hunault,[11] Françoise Pflumio,[12] Jean-Noël Freund,[4] Camille Lobry,[1] Véronique Lhéritier,[13] Hervé Dombret,[14,15] Claire Domon-Dell,[4] Jean Soulier,[1,2] Nicolas Boissel,[14,15] and Emmanuelle Clappier[1,2]

[1]Université Paris Cité, Institut de Recherche Saint-Louis, INSERM U944, Centre national de la recherche scientifique Unité mixte de recherche 7212 GenCellDi, Paris, France; [2]Laboratoire d'Hématologie, Hôpital Saint-Louis, Assistance Publique-Hôpitaux de Paris, Paris, France; [3]Université Paris Cité, Institut Cochin, INSERM U1016, Centre national de la recherche scientifique Unité mixte de recherche 8104, Paris, France; [4]Université de Strasbourg, INSERM, Interface de recherche fondamentale et appliquée en cancérologie/Unité mixte de recherche-S1113, Fédération Hospitalo-Universitaire ARRIMAGE, ITI InnoVec, Fédération de Médecine Translationnelle de Strasbourg, Strasbourg, France; [5]Université de Strasbourg, Institute of Genetics and Molecular and Cellular Biology, Centre national de la recherche scientifique Unité mixte de recherche 7104, INSERM U1258, Illkirch, France; [6]Département de Génétique, Hôpital Robert Debré, Assistance Publique-Hôpitaux de Paris, Université de Paris, Institut de Recherche Saint-Louis, Paris, France; [7]Department of Hematology, Centre Hospitalo-Universitaire Nantes, INSERM Unité mixte de recherche1232 & Centre national de la recherche scientifique ERL6001 Centre de Recherche en Cancérologie et Immunologie Intégrée Nantes Angers, Nantes, France; [8]Department of Hematology, Centre Hospitalier de Versailles, Unité mixte de recherche 1184 Commissariat à l'Energie Atomique, University Paris-Saclay, Le Chesnay, France; [9]Institut Universitaire du Cancer, Toulouse, France; [10]Department of Hematology, Centre Hospitalo-Universitaire de Bordeaux, Hôpital du Haut-Levêque, Pessac, France; [11]Département des Maladies du sang, Centre Hospitalo-Universitaire Angers, Fédération Hospitalo-Universitaire Grand Ouest Against Leukemia, Université d'Angers, Université de Nantes, INSERM, Centre national de la recherche scientifique, Centre de Recherche en Cancérologie et Immunologie Nantes-Angers, Structure Fédérative de Recherche Interactions Cellulaires et Applications Thérapeutiques, Angers, France; [12]Laboratoire des Cellules Souches Hématopoïétiques et des Leucémies, INSERM U1274, Commissariat à l'Energie Atomique, Université de Paris-Université Paris-Saclay, Fontenay-aux-Roses, France; [13]Coordination du Groupe Group for Research in Adult Acute Lymphoblastic Leukemia, Hospices Civils de Lyon, Hôpital Lyon Sud, Lyon, France; [14]Département d'Hématologie Clinique, Hôpital Saint-Louis, Assistance Publique-Hôpitaux de Paris, EA-3518, Paris, France; and [15]Institut de Recherche Saint-Louis, Université Paris Cité, Paris, France

KEY POINTS

- ***CDX2 cis*-deregulation and *UBTF::ATXN7L3* fusion driven by focal deletions define a novel subtype of B-ALL.**
- **CDX2/UBTF::ATXN7L3 is a high-risk B-ALL subtype in young adults, which warrants improved therapeutic strategies.**

Oncogenic alterations underlying B-cell acute lymphoblastic leukemia (B-ALL) in adults remain incompletely elucidated. To uncover novel oncogenic drivers, we performed RNA sequencing and whole-genome analyses in a large cohort of unresolved B-ALL. We identified a novel subtype characterized by a distinct gene expression signature and the unique association of 2 genomic microdeletions. The 17q21.31 microdeletion resulted in a *UBTF::ATXN7L3* fusion transcript encoding a chimeric protein. The 13q12.2 deletion resulted in monoallelic ectopic expression of the homeobox transcription factor *CDX2*, located 138 kb in *cis* from the deletion. Using 4C-sequencing and CRISPR interference experiments, we elucidated the mechanism of *CDX2 cis*-deregulation, involving *PAN3* enhancer hijacking. CDX2/UBTF ALL (n = 26) harbored a distinct pattern of additional alterations including 1q gain and *CXCR4* activating mutations. Within adult patients with Ph$^-$ B-ALL enrolled in GRAALL trials, patients with CDX2/UBTF ALL (n = 17/723, 2.4%) were young (median age, 31 years) and dramatically enriched in females (male/female ratio, 0.2, *P* = .002). They commonly presented with a pro-B phenotype ALL and moderate blast cell infiltration. They had poor response to treatment including a higher risk of failure to first induction course (19% vs 3%, *P* = .017) and higher post-induction minimal residual disease (MRD) levels (MRD ≥ 10^{-4}, 93% vs 46%, *P* < .001). This early resistance to treatment translated into a significantly higher cumulative incidence of relapse (75.0% vs 32.4%, *P* = .004) in univariate and multivariate analyses. In conclusion, we discovered a novel B-ALL entity defined by the unique combination of *CDX2 cis*-deregulation and *UBTF::ATXN7L3* fusion, representing a high-risk disease in young adults.

Introduction

Acute leukemia is a paradigm for the oncogenic role of deregulated transcription factors (TF).[1] Most acute lymphoblastic leukemias (ALL) involve the activation of an aberrant transcriptional program as the initiating oncogenic event, which can be caused by inappropriate expression of a genuine TF or expression of a corrupted TF. Both situations occur mainly by chromosomal translocations, involving either immunoglobulin or T-cell receptor regulatory sequences that deregulate expression of the targeted gene or fusion genes encoding chimeric proteins that harbor altered transcriptional properties.

B-cell precursor ALL (B-ALL) can be subclassified according to recurrent cytogenetic abnormalities that drive specific transcriptional programs and are associated with distinct clinical presentation, response to treatment, and prognosis.[2,3] Consequently, cytogenetic and/or molecular characterization at diagnosis is the cornerstone of risk-adapted treatment stratification in children, along with minimal residual disease assessment.[4] This is the result of extensive annotation of large cohorts of patients and prognostic correlations in the context of clinical trials. In adults, with the exception of Philadelphia chromosome (Ph)-positive ALL that benefit from distinct treatment protocols including tyrosine kinase inhibitors, few genomic abnormalities are widely recognized as useful for treatment stratification.[4] This is partly related to the poorer knowledge of genomic abnormalities defining clinically relevant subgroups of patients, as compared with B-ALL in children.

In order to improve our understanding of adult B-ALL, we performed transcriptome profiling of a large cohort of adult B-ALL harboring no known cytogenetic abnormalities. We identified a novel subtype of B-ALL, elucidated its underlying genomic alterations, and described its clinical features. This novel B-ALL entity affects young women, is characterized by poor treatment outcome, and relies on the unique combination of 2 genomic alterations resulting in *CDX2 cis*-deregulation and *UBTF::ATXN7L3* fusion.

Methods

Patients and samples

Patients with Ph-negative B-ALL treated in the GRAALL-2005,[5] GRAALL-2014, and EWALL-INO trials (clinicaltrials.gov numbers NCT00327678, NCT02617004, and NCT03249870, respectively) were included. Screening for recurrent cytogenetic aberrations was performed by standard cytogenetic procedures in local laboratories. Targeted molecular analyses were performed centrally, including screening for recurrent fusion transcripts by reverse transcription multiplex ligation-dependent probe amplification,[6] screening for *ERG* and *IKZF1* intragenic deletions by breakpoint-specific PCR,[7,8] and quantitative reverse transcription PCR (qRT-PCR) for *CRLF2* and *EPOR* transcripts. All cases negative for the major classifying chromosomal abnormalities (ie, *KMT2A* rearrangements, *TCF3::PBX1*, high hyperdiploidy, low hypodiploidy, and *ETV6::RUNX1*) or for *BCR::ABL1*–like alteration or *ERG* intragenic deletion were subjected to RNA sequencing (RNA-seq). Blast percentage was assessed by cytometry analysis of mononuclear cells subjected to nucleic acid extractions, and cases with <20% blasts were excluded. The studied cohort had higher white blood cells (WBC) and cumulative incidence of relapse (CIR) as compared with non-studied patients (supplemental Table 1). Additional cases allocated to the major adult B-ALL subtypes were included in the RNA-seq cohort (ie, *KMT2A*-rearranged, low hypodiploidy and high hyperdiploidy; n = 16, 10, and 8 cases, respectively). Nine additional cases referred to our laboratory for molecular investigation in the context of patient care and for whom a CDX2/UBTF ALL was identified, including 3 relapses, were included in the cohort for characterization of CDX2/UBTF ALL. All patients or relatives gave informed consent according to the Declaration of Helsinki, and the study was approved by the local ethics committee. A flowchart describing the study cohort is provided in supplemental Figure 1.

High-throughput sequencing

RNA-seq libraries were prepared using the Truseq stranded mRNA library preparation kit (Illumina, San Diego, CA), and 2x150bp paired-end sequencing was performed using a NextSeq 500 (Illumina). Raw, unfiltered RNA-seq reads were aligned to human reference genome hg19 using STAR 2.5.2b.[9] Fusions transcripts were detected using FusionCatcher 0.99.7c β.[10] For *UBTF::ATXN7L3* fusion, visual inspection of each BAM file was done because it could be occasionally overlooked by bioinformatic algorithms. Details of gene expression analysis and mutation detection are provided in supplemental Methods.

Whole-genome long-read sequencing was performed using the PacBio platform. High-molecular-weight DNA was extracted and purified using the MagAttract HMW DNA kit (Qiagen). DNA shearing was performed using g-TUBE (Covaris) with a targeted 20 kb fragment size, and fragments below 5 kb were removed using AMPure PB beads at 3.1X. Library preparation was conducted according to the manufacturer protocol (SMRTbell Express Template Prep Kit 2.0; Pacific Biosciences), and samples were processed using sequencing primer v2 and Sequel II Binding Kit 2.0. Samples were sequenced on a Sequel II instrument in consensus circular sequencing (CCS) mode using 2 SMRTcells per sample. CCS sequencing reads were generated using SMRTlink software and aligned against the human genome (hg38) using pbmm2 of PacBio SMRTtools v.8.0. Structural variant calling was performed using SMRT Analysis. Structural variants of a minimum of 10 bp were retained if covered by at least 2 CCS reads.

Experimental procedures regarding RNA-DNA fluorescent in situ hybridization (FISH), chromatin immunoprecipitation sequencing (ChIP-seq), Assay for Transposase-Accessible Chromatin with highthroughput sequencing (ATAC-seq), 4C-seq,CRISPR interference (CRISPRi), targeted next-generation sequencing (NGS), and array-CGH are provided in supplemental Methods.

Results

A novel B-ALL subtype characterized by a unique gene expression signature and 2 distinct genomic alterations pointing to candidate oncogenic drivers: *CDX2* gene and *UBTF::ATXN7L3* fusion

We performed comprehensive transcriptome and genome analyses of a cohort of 302 B-ALL cases, most of which were negative for the common classifying cytogenetic aberrations, formerly named "B-other" (268/302, 89%). By combining 2 unsupervised methods, weighted gene coexpression network analysis and hierarchical bi-clustering, several clusters were identified (supplemental Figure 2; supplemental Tables 2-6). Aggregation with genomic and fusion transcript data enabled us to allocate major clusters to previously described subtypes within B-other ALL, namely Ph-like,[11,12] DUX4/ERG,[7,13-15] ZNF384,[13] MEF2D,[16] and PAX5 P80R[17] as visualized on *t*-distributed stochastic neighbor embedding analysis (Figure 1A). In addition, we identified a novel cluster comprising 23 B-other cases defined by a unique transcriptional signature.

To investigate the molecular basis of this novel B-ALL subtype, we performed long-read WGS in 2 cases. Integrated analysis of long-read WGS and RNA-seq data uncovered 2 relevant somatic structural variants, both observed in the 2 cases. First, short deletions

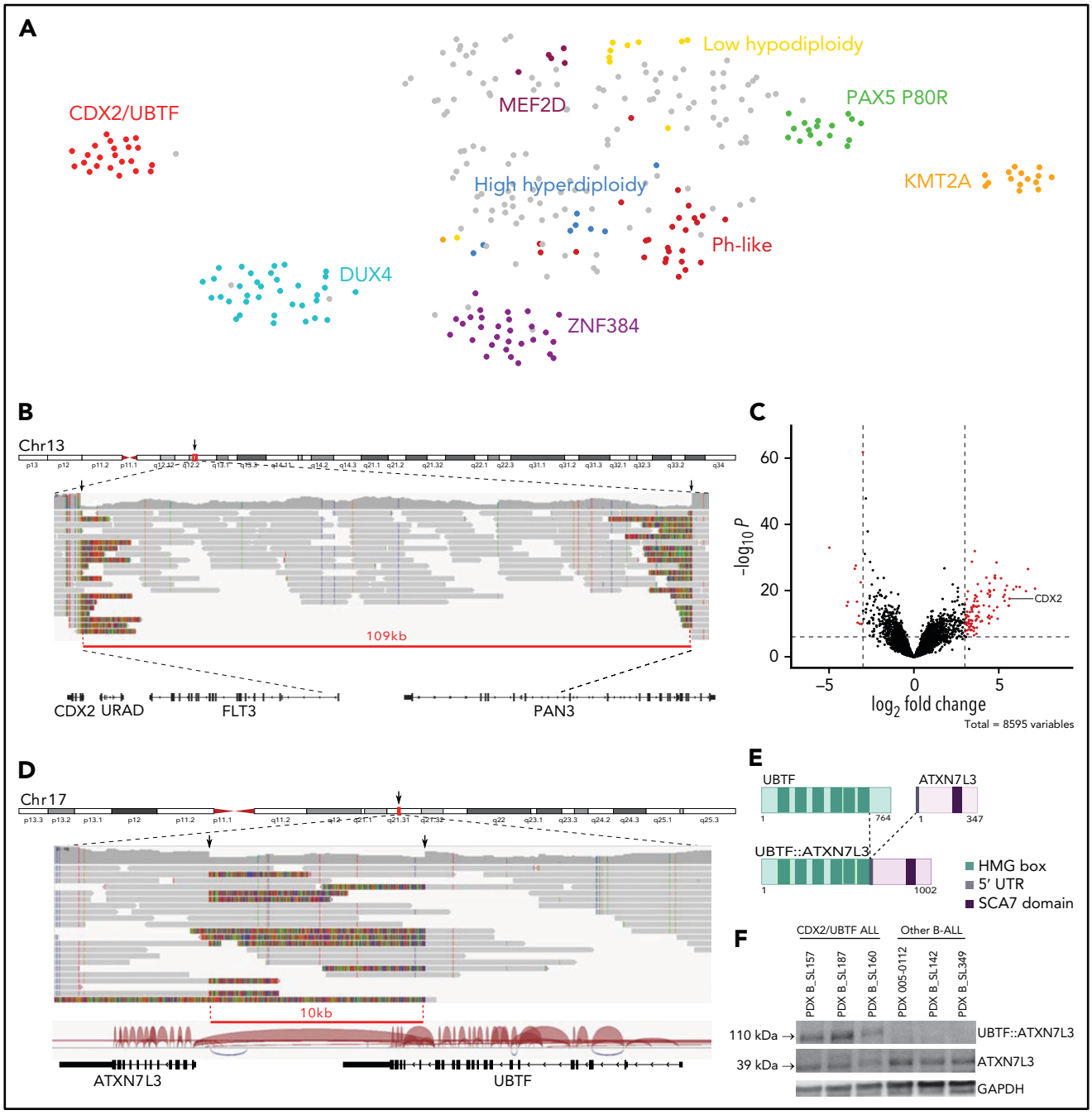


Figure 1. Identification of a novel gene expression B-ALL cluster characterized by *CDX2* high expression and *UBTF::ATX7NL3* fusion. (A) Gene expression profiling of 302 B-ALL cases shown in a 2-dimensional tSNE plot. This analysis was performed with 235 selected genes listed in supplemental Table 6. Major B-ALL subtypes are highlighted in different colors, and gray dots indicate other B-ALL. A novel cluster of cases with distinct gene expression profiling was identified, thereafter named CDX2/UBTF ALL. (B) WGS long-reads visualized on IGV software in 1 CDX2/UBTF case (B_SL160) showing a deletion at 13q12.2 locus close to the *CDX2* gene. The red line underlines the deleted region. (C) Volcano plot of differentially expressed genes between CDX2/UBTF ALL (n = 23) and other B-ALL cases (n = 279). Genes with fold change greater than 3 and -log10 *P* value >7 are shown in red. *CDX2* is highlighted as 1 of the 10 most highly expressed genes in the novel cluster. (D) WGS long-reads and RNA-seq splice junctions visualized on IGV software in 1 CDX2/UBTF case (B_SL160) showing a deletion at 17q21.31 locus involving the *UBTF* and *ATXN7L3* genes. The red line underlines the deleted region. (E) Scheme of the predicted UBTF::ATXN7L3 chimeric protein. The resulting in-frame fusion transcript incorporated 60 nucleotides upstream the canonical translation start codon of *ATXN7L3*. (F) Western-blot of CDX2/UBTF cases using ATXN7L3 antibody shows detection of normal ATXN7L3 protein (39 kDa) in all samples and another protein at a size corresponding to the predicted UBTF::ATXN7L3 chimeric protein (110 kDa) in CDX2/UBTF cases only. Western blot performed on PDX from 3 CDX2/UBTF cases (B_SL157, B_SL187, and B_SL160), 2 PAX5 P80R cases (005-0112 and B_SL142), and 1 KMT2A-r case (B_SL349). IGV, integrative genomics viewer; PDX, patient-derived xenograft; tSNE, t-distributed stochastic neighbor embedding; WGS, whole-genome sequencing.

were observed at 13q12.2 locus (83 and 109 kb) (Figure 1B). Strikingly, those were located 138 kb upstream of the *CDX2* gene, one of the most highly expressed genes in the novel cluster (Figure 1C; supplemental Table 7). Second, short deletions (10 kb) at 17q21.31 involved the *UBTF* and *ATXN7L3* genes, with the loss of the 4 last exons of *UBTF* (exons 18-21) and intergenic

sequence, bringing the 5′ portion of *UBTF* closer to the adjacent gene *ATXN7L3* (Figure 1D). This resulted in an in-frame *UBTF::ATXN7L3* fusion transcript detected in RNA-seq data and confirmed by qRT-PCR (Figure 1D-E).

CDX2 is a member of the caudal-type family (Cdx) of homeodomain TFs. It plays essential roles in development during early stages of embryogenesis, including primitive hematopoiesis.[18,19] By contrast, *CDX2* is not expressed in hematopoietic cells in postnatal life at any stage of differentiation[20] but is restricted to the gut, where it is critical for homeostasis and behaves as a tumor suppressor in colorectal cancer.[21] Intriguingly, ectopic *CDX2* expression has been reported in most myeloid and lymphoblastic acute leukemias,[20,22] suggesting a role in leukemogenesis. Further works have shown that ectopic Cdx2 is indeed able to induce acute myeloid leukemia (AML) and myelodysplastic syndrome in mice.[23-25] To understand the relevance of our finding with respect to these observations, we analyzed *CDX2* expression in a panel of AML, T-, and B-ALL and normal blood and bone marrow samples (supplemental Figure 3). Strikingly, *CDX2* expression in B-ALL cases from the novel cluster was 1000-fold higher than in all other leukemia cases, suggesting a major driving role in this subtype.

UBTF (upstream binding transcription factor) is expressed ubiquitously and encodes a HMG-box DNA-binding protein involved in ribosome biogenesis. ATXN7L3 is a component of the histone deubiquitination module of the SAGA complex, which activates transcription by chromatin remodeling.[26] UBTF was previously found involved in oncogenic fusions with ETS4 in prostate cancer[27] and with MAML3 in neuroendocrine tumors,[28] where it promotes expression of its partner genes. In agreement with this model, *ATXN7L3* was slightly but significantly more expressed in cases from the novel cluster (supplemental Figure 4). In addition, tandem duplication of *UBTF* was recently reported in pediatric AML.[29,30] The predicted UBTF::ATXN7L3 fusion protein retained the major part of UBTF, including all high mobility group box domains and the entire ATXN7L3. Western blot analysis confirmed the presence of an ATXN7L3-containing protein at the expected size in cases harboring the alteration (Figure 1F). Importantly, *UBTF::ATXN7L3* fusion transcripts were found in each case of the cluster and were absent in all but 1 other B-ALL (Figure 2), suggesting a role as a codriver in this subtype, further named CDX2/UBTF.

13q12.2 deletion drives *cis*-deregulation of *CDX2* expression

Using array-CGH and/or targeted NGS, focal 13q12.2 deletions were identified in every CDX2/UBTF case (Figure 3A; supplemental Table 8). Fine mapping of those deletions using capture-based sequencing of the genomic region identified a common minimal deleted region of 41 kb length, which included the intergenic region between 2 contiguous genes, *FLT3* and *PAN3*, and the promoter and first exon of both genes. CDX2/UBTF cases had lower expression of *FLT3* and *PAN3*, consistent with gene dosage (Figure 3B). Interestingly, our data also revealed a correlated expression of *FLT3* and *CDX2* in all the other, non-CDX2/UBTF, B-ALL (Figure 3C), suggesting coregulation of the 2 genes within the locus.

By screening additional patients' samples and B-ALL cell lines for *CDX2* expression, we identified 3 patients and the NALM16 cell line as having high *CDX2* expression (supplemental Figure 5). The 3 patients presented 13q12.2 deletion and *UBTF::ATXN7L3* fusion and were thus included in the CDX2/UBTF group. NALM16 harbored a short duplication (38 kb) at 13q12.2 within the *PAN3* gene but no *UBTF::ATXN7L3* fusion (Figure 3A).

Analysis of breakpoint sequences revealed recombination signal sequence–like motifs and nontemplated sequence nucleotides at junction (supplemental Figure 6), indicating that the deletion most likely occurred through off-target cleavage by the RAG1/2 recombinase in a differentiating B-cell precursor. Such deletions usually occur in regions of open chromatin around active genes.[31] To understand if those deletions were causative of *CDX2* deregulated expression rather than merely related to active transcription of the locus, we analyzed allele-specific expression of *CDX2* using single-nucleotide polymorphisms located in its untranslated regions. In every CDX2/UBTF case with informative single-nucleotide polymorphisms (n = 10 cases), a clear mono-allelic expression pattern was observed (supplemental Table 9). We performed sequential RNA and DNA FISH on PDX of CDX2/UBTF ALL combining a probe to visualize the *CDX2* nascent transcripts (RNA FISH part)

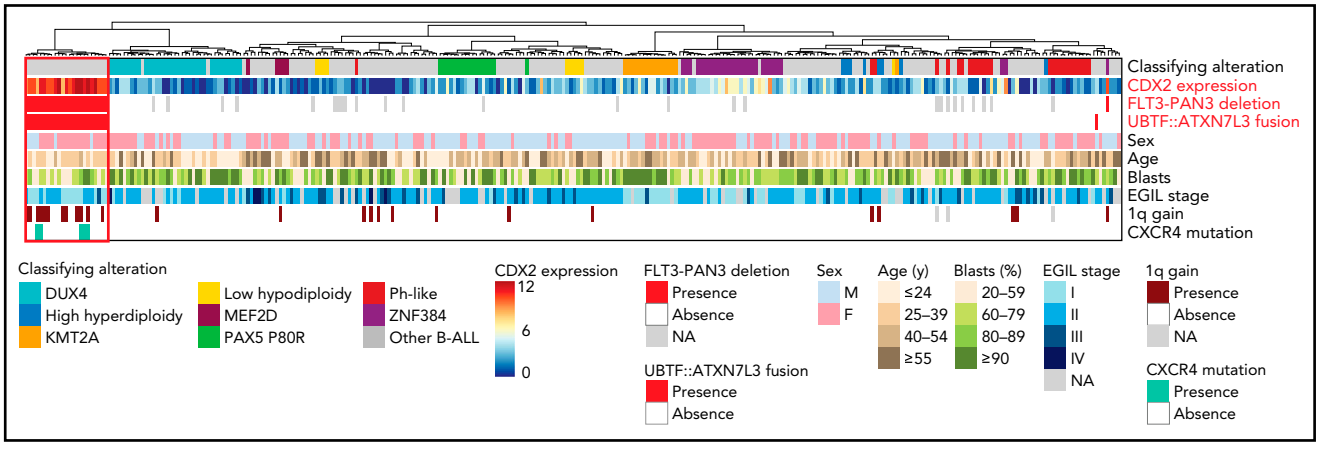


Figure 2. Summary oncoplot highlighting the main features of CDX2/UBTF ALL within the whole RNA-seq cohort. Classifying alterations, sex, age, blast percentage, and immunophenotypic classification according to European Group for the Immunological Classification of Leukaemia (EGIL); 1q gain and *CXCR4* mutations are indicated. *CDX2* expression is indicated in color scale from RNA-seq data. FLT3-PAN3 deletion corresponds to the 13q12.2 deletion involving *FLT3* and *PAN3* promoters and exon 1 (excluding 13q12.2 deletion sparing *FLT3*). UBTF::ATXN7L3 fusion indicates the presence of the fusion transcript as detected on RNA-seq data.

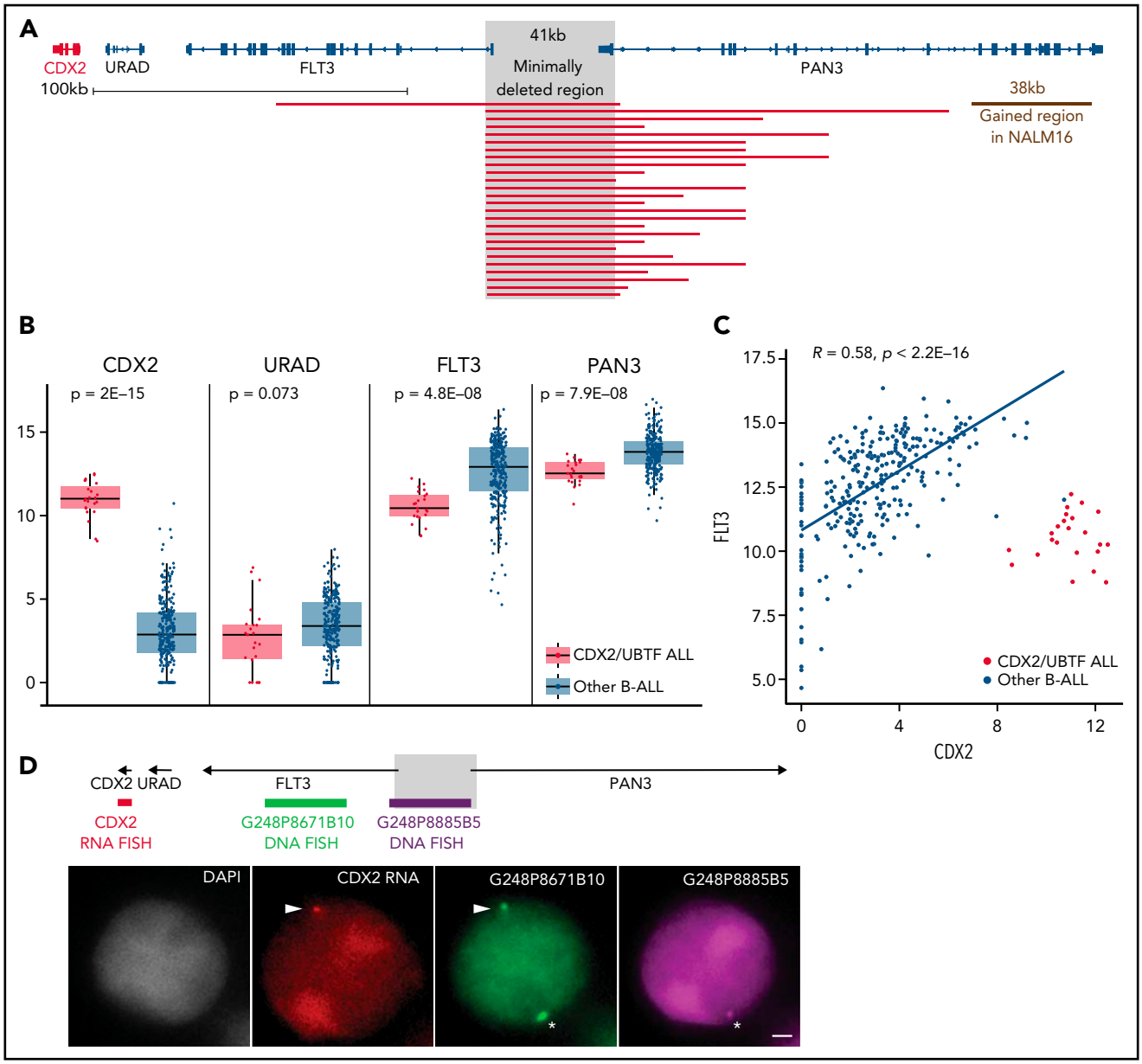


Figure 3. The 13q12.2 microdeletion upstream *CDX2* is associated with deregulation of *CDX2* in *cis*. (A) Capture-based sequencing of 13q12.2 locus identified focal deletion upstream of *CDX2* in all CDX2/UBTF cases. The gray zone indicates the minimally deleted region. (B) Expression analysis of *CDX2, URAD, FLT3*, and *PAN3* in CDX2/UBTF cases as compared with other B-ALL. Gene expression was quantified by RNA-seq (log2 of counts normalized using DESeq2's median of ratios), and comparison analysis was performed by the Wilcoxon rank-sum test. (C) Dot-plot of expression data of *CDX2* and *FLT3* genes based on RNA-seq data for the whole cohort (n = 302) and Spearman's rank correlation analysis of non-CDX2/UBTF ALL. (D) Representative experiment (B_SL157) of sequential RNA-DNA FISH showing *CDX2* primary transcripts (red signal) expressed from the allele located on the chromosome 13 exhibiting the genomic deletion (green G248P8671B10 signal and missing purple G248P8885B5 signal), shown with arrowheads. Wild-type chromosome 13 (green G248P8671B10 and purple G248P8885B5 signals) is shown by asterisks. DNA counterstained with DAPI (gray). Scale bar: 1 μm. DAPI, 4′,6-diamidino-2-phenylindole.

and 2 probes mapping the deleted and nondeleted regions at 13q12.2 (DNA FISH part). RNA-DNA FISH showed that *CDX2* was uniquely transcribed from the allele in *cis* from the deletion (Figure 3D; supplemental Table 10), demonstrating that *CDX2* deregulated expression was mediated by the genomic deletion in *cis*.

Recurrent 13q12.2 deletions were recently reported in B-ALL and shown to deregulate *FLT3*.[32] However, those deletions were different because they spared the *FLT3* gene. In our cohort, we identified 10 cases having such deletions, which did not cluster with CDX2/UBTF ALL but spread across other B-ALL. Indeed, our expression data show distinct transcriptional effects on the locus, drastic deregulation of *CDX2* expression on one hand vs moderate *FLT3* increase on the other hand (supplemental Figure 7). Of note, 1 case had a deletion involving *FLT3* with high *CDX2* expression but no *UBTF::ATXN7L3* fusion, and it did not cluster with CDX2/UBTF cases.

13q12 deletion alters 3D DNA topology with hijacking of a *PAN3* enhancer

Using ChIP-seq for histone modifications, we mapped *cis*-regulatory elements within the 13q12.2 locus in 2 PDX of CDX2/

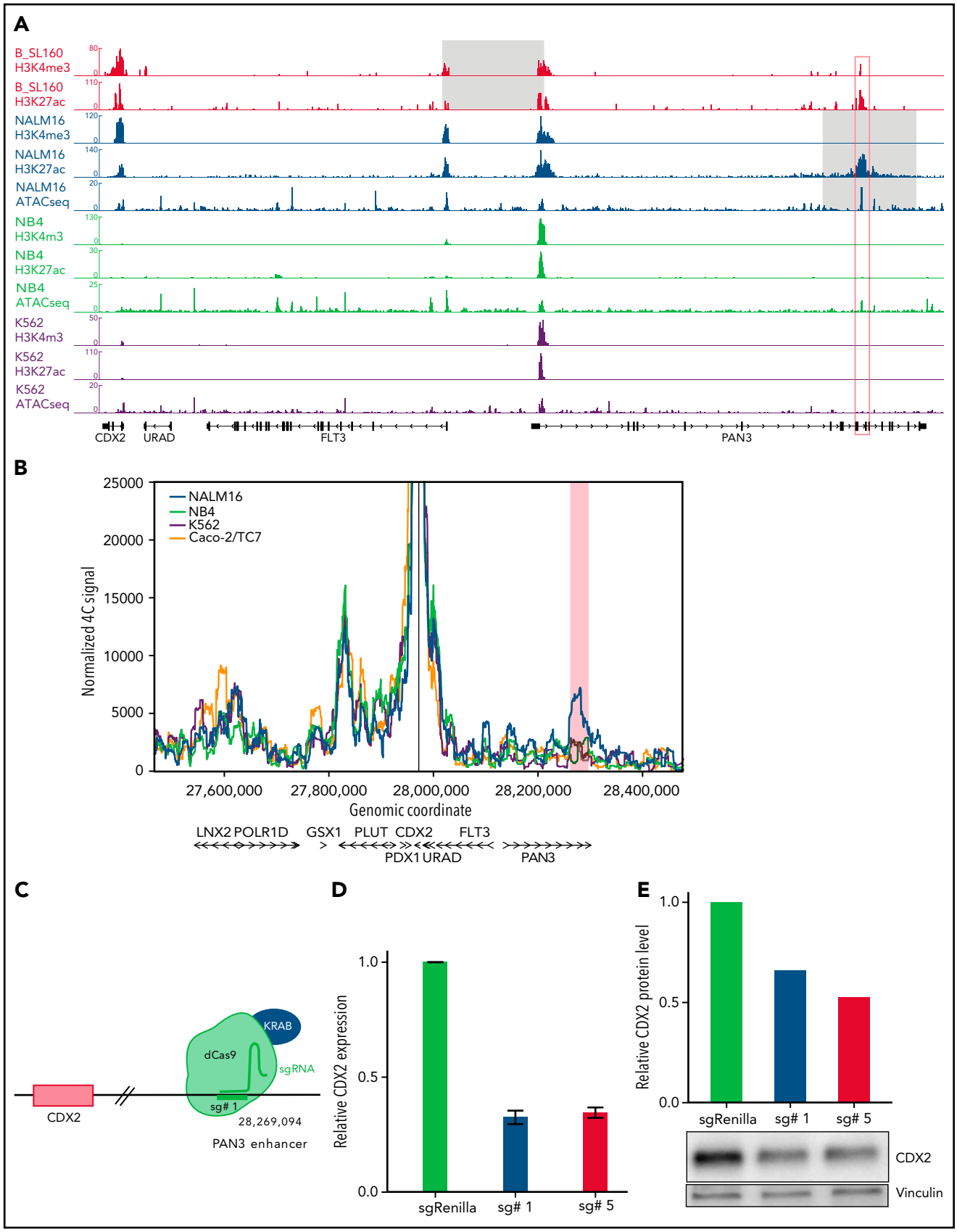


Figure 4.

UBTF cases and in NALM16 and control cell lines (Figure 4A). H3K4me3 and H3K27ac peaks were observed at promoter regions of *CDX2*, *FLT3*, and *PAN3* in CDX2/UBTF ALL and NALM16, indicating active promoters. In addition, a strong H3K27ac peak was observed within the *PAN3* gene, indicative of an active enhancer. Importantly, this element is spared in both types of 13q12.2 deletions, and Yang et al[32] showed that it drives *cis*-deregulation of *FLT3*. In addition, the short duplication found in NALM16 encompasses this region, suggesting that this enhancer may be also involved in *CDX2* deregulation. We performed 4C-seq on NALM16 and control cell lines using the *CDX2* promoter as the viewpoint (Figure 4B). In addition to a conserved interaction downstream *CDX2* seen in all cell lines, an interaction between the *CDX2* promoter and the *PAN3* enhancer region was observed in NALM16 only. Finally, we designed a CRISPRi experiment using single-guide RNAs targeting the *PAN3* enhancer, coexpressed with a transcriptional repressor domain fused to catalytically dead CRISPR-Cas9 (Figure 4C). Upon *PAN3* enhancer repression in NALM16, we observed marked reduction of *CDX2* expression at the transcript and protein levels (Figure 4D-E). Collectively, these results definitively demonstrate the role of the *PAN3* enhancer on *CDX2* aberrant expression due to a genomic alteration within the 13q12.2 locus and strongly suggest that *CDX2* deregulation in CDX2/UBTF ALL patients is the result of hijacking of the *PAN3* enhancer.

Gene expression signature of CDX2/UBTF ALL

To explore the deregulated gene expression in CDX2/UBTF ALL, we analyzed the core set of genes most differentially expressed (Figure 5A; supplemental Table 7). This signature comprised a small number of genes highly expressed in CDX2/UBTF ALL and, to a lesser extent, in most other B-ALL, like *CD34* (cluster 4), and also genes shared with some subtypes, like *MEIS1*, also expressed in *KMT2A*-rearranged cases (cluster 3). Of note, some HOX genes were also moderately expressed in CDX2/UBTF ALL, in agreement with the role of CDX2 as an upstream factor (supplemental Figure 8).[33,34] A small group of genes had much lower expression in CDX2/UBTF ALL, like *CD9* (cluster 5). The vast majority of the signature comprised genes not expressed in other B-ALL and ectopic with regard to hematopoietic cells (annotated "not expressed in $CD34^+$"), a number of which being expressed in brain or blood vessels. A limited number of CDX2 direct target genes in the context of colon cancer (Caco-2/TC7 cell line) were identified, including *CD9*, *CDH4*, *NR6A1*, *PLD5*, and *SPON1*. GSEA showed enrichment in genes regulated by the polycomb repressive complex 2 (PRC2) in embryonic stem cells, expressed in cord blood–derived hematopoietic stem cells, and related to homophilic cell adhesion (Figure 5B). GSEA also showed negative enrichment in genes related to cell cycle and mTOR signaling, suggestive of a lower metabolic state.

Genomic landscape of CDX2/UBTF ALL

We characterized additional alterations using array-CGH (n = 20, Figure 5C) and targeted sequencing of a custom panel of genes recurrently altered in B-ALL (n = 26, Figure 5D). A number of were recurrent in CDX2/UBTF ALL, the most frequent being duplication of the 1q arm, seen in 50% (13/26) of patients, often related to unbalanced chromosomal translocations as shown by cytogenetic data (supplemental Table 12). Other nonfocal copy-number alterations involved loss of 6q (n = 6), 9q (n = 6), 13q including *DLEU* (n = 6) and *RB1* (n = 4), and 3q including *TBL1XR1* (n = 3) and *TP63* (n = 3). Focal deletions involving *PAX5* 5′ or gene promoter were particularly frequent (n = 13/26, 50%); other genes targeted by recurrent deletions were *IKZF1* (n = 5/26, 19%), *VPREB1*, *ETV6*, *CDKN1B*, *KDM6A*, *CTCF*, and *PTPRD*.

Targeted sequencing showed frequent mutations of RAS pathway and *CREBBP* genes (27% and 19%, respectively) (Figure 5D; supplemental Tables 13-14). In addition, RNA-seq analysis for variant calling uncovered recurrent mutations of *CXCR4*, further confirmed by Sanger sequencing and present in 5 cases (22%). Interestingly, *CXCR4* mutations were uniquely detected in CDX2/UBTF ALL. Those mutations are predicted to truncate the C-terminus, like described in WHIM syndrome and in Waldenström macroglobulinemia and shown to confer gain-of-function.[35-37]

Clinical presentation and outcome of CDX2/UBTF ALL

The 26 CDX2/UBTF ALL patients included 20 females, demonstrating highly skewed sex ratio (male/female ratio, 0.3), and comprised 4 adolescents and 22 adults. No CDX2/UBTF ALL patient was identified in adults aged 55 or older.

The immunophenotype of CDX2/UBTF ALL was frequently classified as pro–B/B-I stage according to EGIL (16/26, 62%), with negative or partial expression of CD10 (n = 22/26, 85%) and negative expression of CD20 (n = 20/22, 91%), whereas CD34 and CD38 were highly expressed (supplemental Table 15).

Seventeen CDX2/UBTF ALL patients were included in the GRAALL-2005/2014 protocols prospectively enrolling adults aged 18 to 59 years (Table 1). Thus, the prevalence of this subtype in adult Ph-negative ALL is 2.4% (17/723). Female predominance was significant in this cohort (83% vs 44% in other B-ALL, $P = .002$), and patients tended to be younger (median age, 31 years vs 38 years, $P = .13$). Blast infiltration in the bone marrow was significantly lower in CDX2/UBTF ALL (91% vs 81%, $P = .04$), and their immunophenotype was skewed to pro–B/B-I EGIL stage (12/17, 71%, $P < .001$).

Patients with CDX2/UBTF ALL displayed a slow response to induction phase. Peripheral blood response to steroid prophase

Figure 4. CDX2 *cis*-deregulation is driven by hijacking of an enhancer located in the *PAN3* gene. (A) ChIP-seq signals of histone modifications (H3K4me3 and H3K27ac) of 1 representative CDX2/UBTF case (B_SL160) and ChIP-seq and ATAC-seq signals of NALM16, NB4, and K562 cell lines at the *CDX2-PAN3* locus. *CDX2* expression levels of cell lines are provided in supplemental Figure 5. The gray zone on CDX2/UBTF ALL and NALM16 traces indicate respectively the minimally deleted region and the duplicated region. The red box corresponds to the position of the enhancer identified in *PAN3*. (B) 4C-seq normalized signals of NALM16, NB4, K562, and Caco-2/TC7 cell lines with *CDX2* promoter as view point. The red zone highlights the NALM16-specific called interaction (C) CRISPRi experimental scheme. In order to inactivate the enhancer in *PAN3*, we used the catalytically inactive CRISPR-associated protein 9 (dCas9) fused to the repressor KRAB with specific single-guide RNA. Representative example of a CRISPRi experiment showing *CDX2* expression decrease in NALM16 after CRISPRi targeting the *PAN3* enhancer, measured by qRT-PCR (D), and western blot (E).

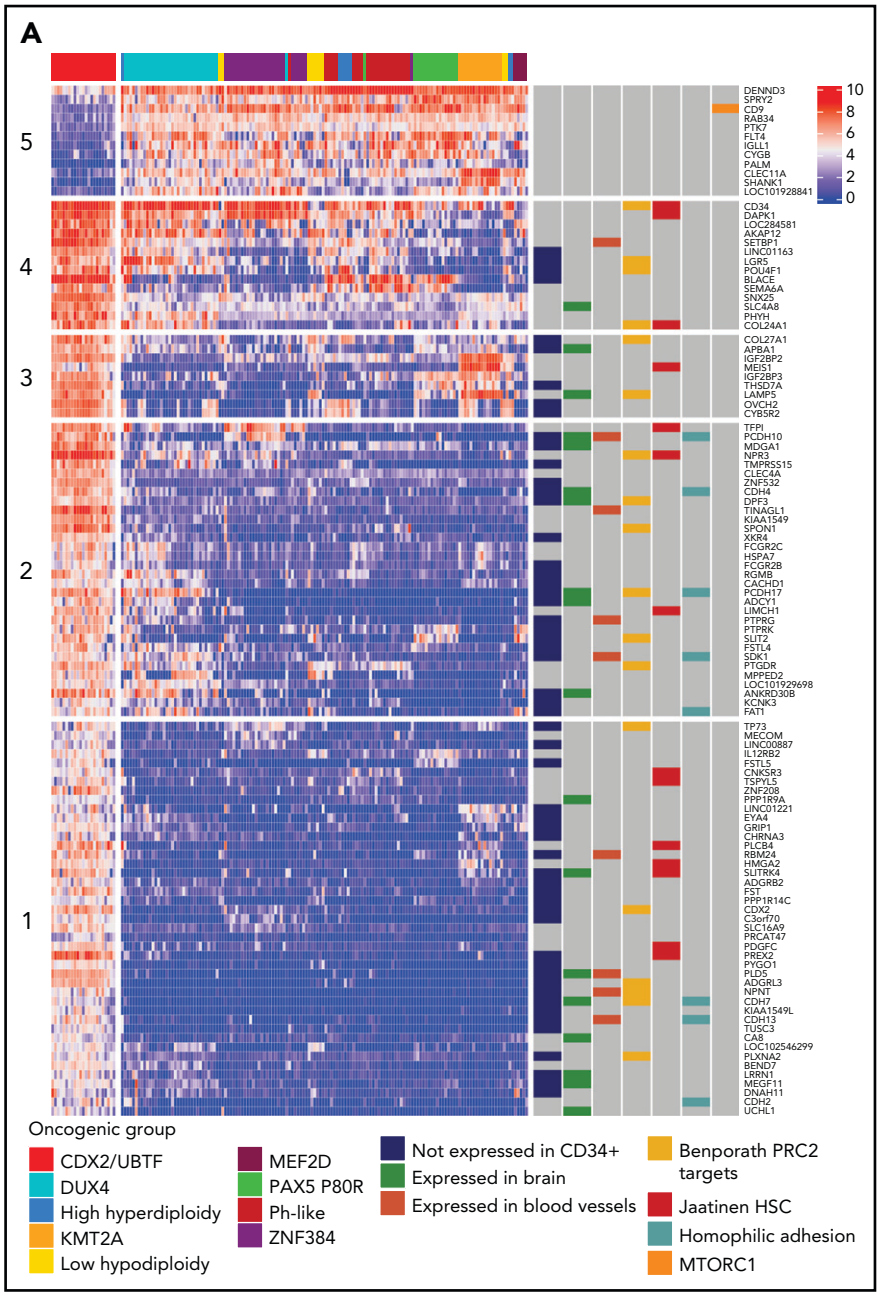


Figure 5. Transcriptional signature and genomic landscape of CDX2/UBTF ALL. (A) Heatmap representing biclustering of gene expression from the major subtypes in the RNA-seq cohort (excluding unclassified cases). Genes were selected after comparison between expression means (fold change >3, $P < 10^{-7}$) resulting in a list of 110 genes (supplemental Table 7). Annotations on the right are derived from selected gene set enrichment analyses (GSEA) shown in (B) and expression data from The Human Cell Atlas Bone Marrow Single-Cell Interactive Web Portal (http://www.altanalyze.org/ICGS/HCA/splash.php) and the TSEA atlas (genetics.wustl.edu/jdlab/tsea/) (B) Selected GSEA for genes expressed in CDX2/UBTF ALL as compared with other cases, showing significant enrichment in genes regulated by the Polycomb repressive complex 2 (PRC2), genes expressed in cord blood–derived $CD133^+$ cells enriched in hematopoietic stem cells, genes involved in homophilic cell-cell adhesion, and genes expressed in a subset of embryonic neurons. The differential gene list is negatively enriched in genes upregulated through mTORC1 signaling and E2F target genes. GSEA core genes are listed in supplemental Table 11. (C) Copy-number alterations in CDX2/UBTF ALL as determined by array-CGH analysis (n = 20). (D) Heatmap of recurrent alterations in CDX2/UBTF cases as determined by array-CGH and targeted sequencing (n = 26), with frequencies of altered cases indicated on the left. *Cases with low blast infiltration (ie, <30%), with possibly focal copy-number alterations overlooked. TSEA, tissue specific expression analysis.

at day 8 and bone marrow blast clearance at day 15 (M1 bone marrow) were observed in only 50% (vs 84% in other B-ALL, $P = .002$) and 13% (vs 60%, $P < .001$) of CDX2/UBTF cases, respectively. One patient died during induction, and 3 patients achieved a late CR after second induction (19%, vs 3%, $P = .017$). All but 1 evaluable patient (13/14, 93%) had a high MRD ($\geq 10^{-4}$) at CR, as compared with 46% in others ($P < .001$). Therefore, CDX2/UBTF ALL is a subtype with very poor response to chemotherapy.

Patients with poor response at day 8 or at day 15 were eligible to allo-SCT in GRAALL-2005, whereas patients with high postinduction MRD were eligible in GRAALL-2014. Accordingly, patients with CDX2/UBTF ALL received more frequent allo-SCT (10/16,

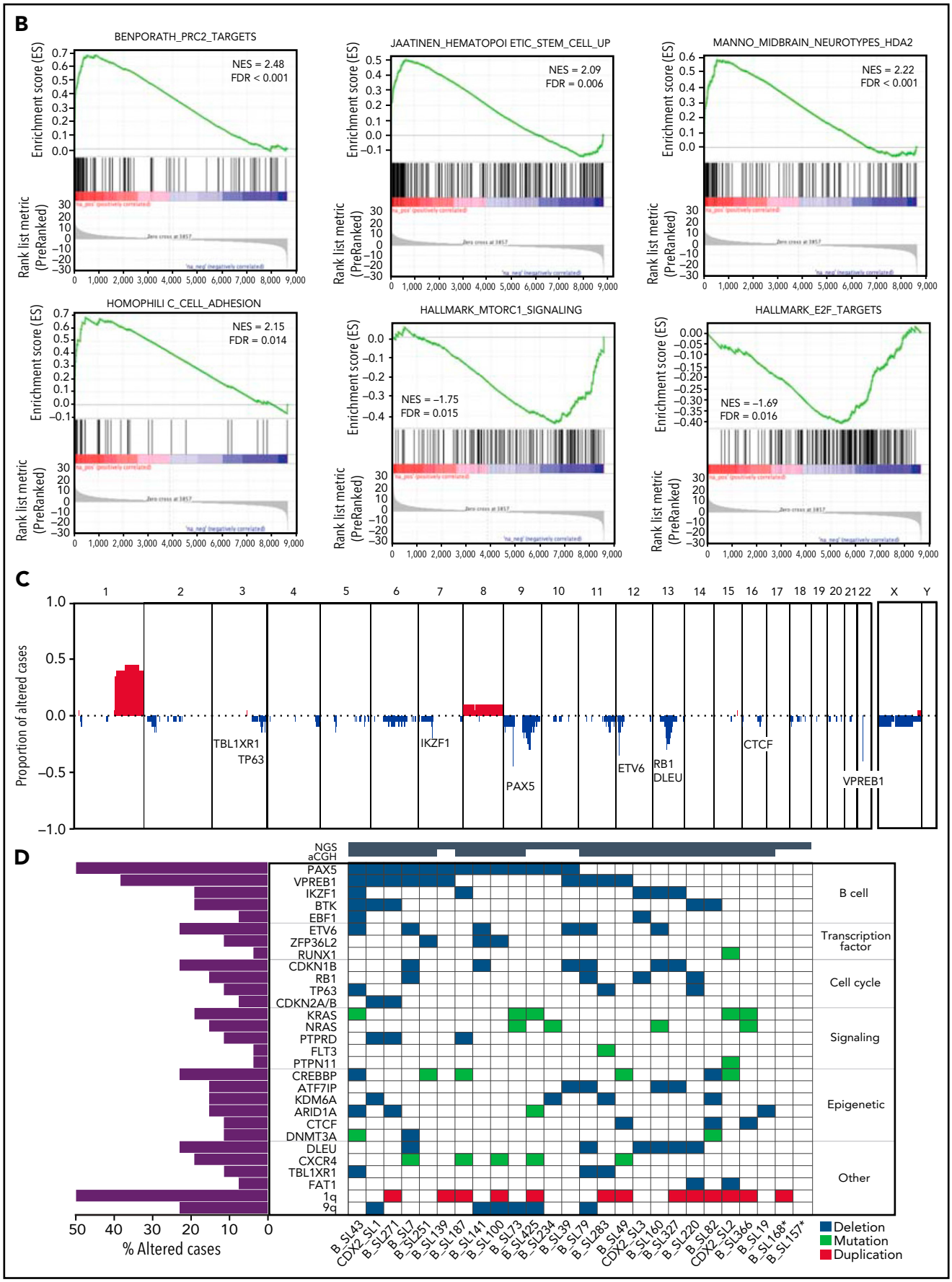


Figure 5. (continued)

Table 1. Study patient characteristic

	All patients (n = 723)	Other B-ALL (n = 706)	CDX2/UBTF ALL (n = 17)	*P* value
Patient-related characteristics				
Median age, y (range)	38 (18-60)	38 (18-60)	31 (19-50)	.13
Gender, M/F (ratio)	400/323 (1.2)	397/309 (1.3)	3/14 (0.2)	.002
Disease-related characteristics				
Median WBC, G/L (range)	9 (0-712)	9 (0-712)	9 (1-63)	.72
BM blast percentage (range)	91 (21-100)	91 (21-100)	81 (35-98)	.04
CNS involvement	57/713 (8%)	57/696 (8%)	0/17 (0%)	.39
Immunophenotypic classification (EGIL)				<.001
Pro-B (I)	148/622 (24%)	136/605 (22%)	12/17 (71%)	
Common (II)	364/622 (59%)	361/605 (60%)	3/17 (18%)	
Pre-B (III)	102/622 (16%)	100/605 (17%)	2/17 (12%)	
Mature (IV)	8/622 (1%)	8/605 (1%)	0/17 (0%)	
Response-related characteristics				
Good PB prednisone response at day 8	582/701 (83%)	574/685 (84%)	8/16 (50%)	.002
Good BM response at day 15	397/668 (59%)	395/653 (60%)	2/15 (13%)	<.001
Late CR (achieved after induction 2)	24/672 (4%)	21/656 (3%)	3/16 (19%)	.017
CR (after induction 1 and 2)	672/723 (93%)	656/706 (93%)	16/17 (94%)	1.0
Postinduction MRD ≥10-4	255/542 (47%)	242/528 (46%)	13/14 (93%)	<.001
Postinduction MRD ≥10-3	151/542 (28%)	144/528 (27%)	7/14 (50%)	.073
Postremission outcome				
Allo-SCT in first CR	184/672 (27%)	174/656 (27%)	10/16 (63%)	.003
3-y CIR, % (95% CI)	33.5 (29.8-37.6)	32.4 (28.7-36.5)	75.0 (50.6-93.5)	.004
3-y OS, % (95% CI)	65.9 (62.0-69.5)	65.9 (61.9-69.5)	67.9 (38.8-85.4)	.86

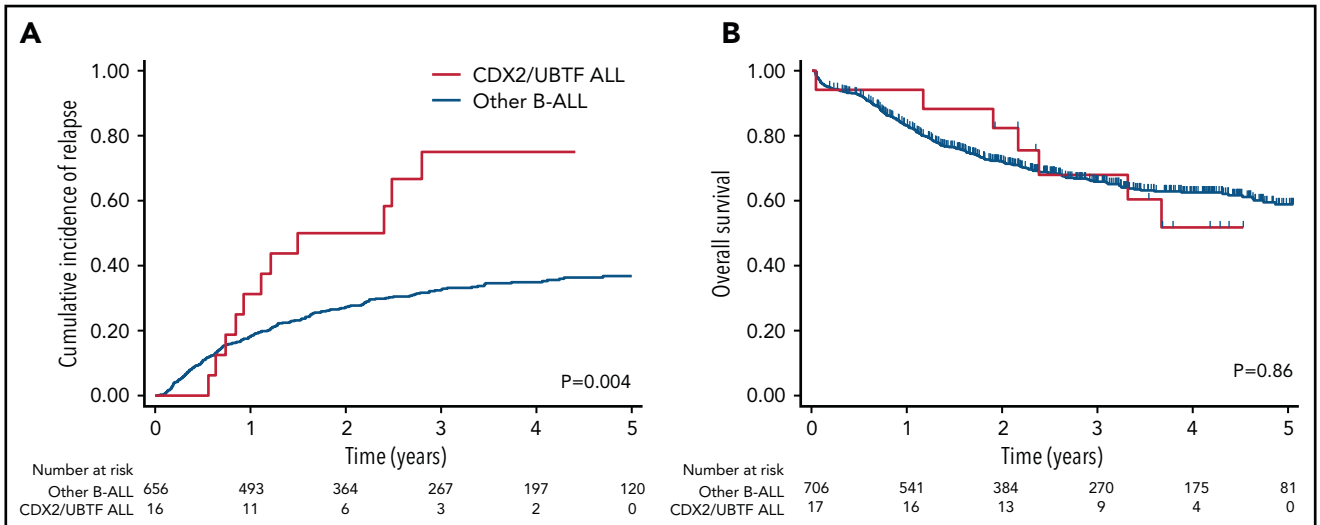
For BM blast percentage, data were available for 423 patients. Good PB prednisone response at day 8 is defined by less than 1 G/L blast. Good BM response at day 15 is defined by <5% blast. MRD evaluation was performed by PCR quantification of immunoglobulin or T-cell receptor clono-specific rearrangements according to EuroMRD guidelines.

Allo-SCT, allogeneic stem cell transplantation; BM, bone marrow; CNS, central nervous system; CR, complete remission; EGIL, European Group for the Immunological Characterization of Leukemias; F, female; M, male; MRD, minimal residual disease; OS, overall survival; PB, peripheral blood.

63% vs 74/656, 27%, $P = .003$). The median follow-up of the cohort was 3.4 years. At 3 years, the CIR estimates were 75.0% (95% CI, 50.6-93.5) in CDX2/UBTF ALL patients and 32.4% (95% CI, 28.7-36.5) in other B-ALL (specific hazard ratio, 2.43; 95% CI, 1.32-4.46; $P = .004$, Figure 6; supplemental Figure 9). CDX2/UBTF ALL was still associated with a significantly higher CIR in multivariate analysis including age, WBC counts, and postinduction MRD as covariates (specific hazard ratio, 2.41; 95% CI, 1.20-4.84; $P = .013$, Figure 6C). However, the 3-year OS did not differ between CDX2/UBTF ALL (67.9%; 95% CI, 38.8-85.4) and other B-ALL patients (65.9%; 95% CI, 61.9-69.5), which may reflect a favorable response to current salvage procedures in B-ALL.

Discussion

Aiming to elucidate the oncogenic bases of unresolved cases of adult B-ALL, we uncovered a novel leukemia entity with unique genomic and phenotypic features. Most intriguing is the systematic combination of 2 genomic alterations pointing to distinct candidate oncogenic drivers, namely *CDX2* and *UBTF::ATXN7L3*. CDX2 is a homeobox TF well known for its role during embryogenesis and in adult intestinal homeostasis. Inappropriate reactivation of CDX2 has been widely observed in leukemias, and an oncogenic role was proposed.[20,22] However, the mechanism of ectopic deregulation has remained an enigma and, even though *CDX2* expression in the hematopoietic compartment is able to induce leukemia in mice,[23-25] there was no direct evidence of its contributing role in human leukemogenesis. Our study, demonstrating the targeting and strong deregulation of *CDX2* by a recurrent alteration in a subset of B-ALL, thus provides definitive support for its oncogenic role in human leukemia. A recent article from Yasuda et al[38] reported the CDX2-expressing B-ALL subtype but did not elucidate the mechanism of its deregulation. Interestingly, we uncovered focal alterations within the CDX2-FLT3-PAN3 locus in all CDX2/UBTF ALL patients and in the NALM16 cell line and proved *CDX2 cis*-deregulation mediated by enhancer hijacking. We also provide evidence that those deletions likely result from RAG-mediated inappropriate recombination events, which are commonly seen in ALL. To be targeted by RAG enzymes, a locus should be in an open chromatin state and, accordingly, translocations and deletions are enriched in B-cell TFs and other genes actively transcribed during B-cell development.[31] Because *CDX2* is not expressed in normal hematopoietic cells, one can speculate that recombination events close to *FLT3* and *PAN3* promoters could occur, owing to their active transcriptional state. Hence, *FLT3* expression is initiated in multipotent progenitors and plays a key role in the maintenance of lymphoid-primed multipotent and common lymphoid progenitors. Expression pattern and conditional knockout mouse studies showed that *FLT3* expression is maintained and critical at the pro-B stage, where RAG expression is induced.[39] Therefore, our data highlight how neighbor



C

	Univariate analysis			Multivariate analysis		
	Hazard ratio	95% CI	P value	Hazard ratio	95% CI	P value
Cumulative incidence of relapse						
Age*	1.01	1.00–1.02	0.16	1.01	1.00–1.03	0.026
Log(WBC)*	1.25	1.15–1.36	<0.001	1.32	1.19–1.46	<0.001
Post-induction MRD ≥ 10^{-4}	2.02	1.46–2.78	<0.001	1.61	1.15–2.25	0.005
CDX2/UBTF	2.43	1.32–4.46	0.004	2.41	1.20–4.84	0.013
Overall survival						
CDX2/UBTF	1.07	0.51–2.28	0.86			

Figure 6. Outcome analyses of CDX2/UBTF ALL patients. Cumulative incidence of relapse (A) and OS (B) Kaplan-Meier curves for B-ALL patients enrolled in GRAALL-2005/2014 trials according to CDX2/UBTF ALL status. (C) Univariate and multivariate outcome analyses for CDX2/UBTF ALL patients. *Continuous variables. WBC, white blood cell count.

genes can be involved in a deleterious loop linking gene expression, chromatin structure, and oncogenic activation, suggesting a mechanistic framework for deregulation of ectopic genes by RAG-mediated alterations.

The functional consequences of both *CDX2* deregulation and *UBTF::ATXN7L3* chimeric fusion remain to be investigated, as well as the likely oncogenic cooperation suggested by the unique combination of those 2 alterations. CDX2 is known to act as an upstream regulator of HOX genes,[33,34] the expression of which was observed in CDX2/UBTF ALL, although at a lesser extent than in *KMT2A*-rearranged ALL. However, this probably does not recapitulate CDX2-mediated oncogenesis, and identifying the relevant targets of CDX2 in B-ALL will require dedicated work, taking into account its context-dependent function.[40,41] Not less complex may be the role of UBTF::ATXN7L3. Upstream binding transcription factor, the product of *UBTF*, plays a critical role in ribosomal RNA transcription.[42] Interestingly, alterations targeting *UBTF* (ie, internal tandem duplications) were reported as a recurrent lesion in pediatric AML.[29,30] Because it is ubiquitously expressed and interacts with DNA as a dimer, its role in the fusion may also fit with the common model for chimeric fusions where the 5′ partner allows deregulated expression and dimerization of the 3′ partner, altering its properties of DNA binding or cofactors recruitment.[43]

ATXN7L3 is an adaptor protein that orchestrates activities of multiple deubiquitinating enzymes and acts as a global facilitator for histone 2B (H2B) deubiquitination.[26] Misregulation of H2B monoubiquitination has been observed in cancers and shown to promote oncogenesis.[44] Hence, H2B deubiquitination is required for optimal expression of many target genes of transcription activators.[45,46] One may speculate that altered activity of ATXN7L3 factor fused to UBTF might enable CDX2 to drive an aberrant transcriptional program and reveal its oncogenic potential. The core signature of CDX2/UBTF ALL comprises a number of ectopic genes normally expressed in brain or blood vessels, the chromatin state of which in lymphoid progenitors may not allow activation by CDX2 alone. In support of a unique oncogenic cooperation are the 2 outlier cases from our cohort, 1 having only the *CDX2 cis*-deregulating deletion and the other only the *UBTF::ATXN7L3* fusion and neither presenting the typical and complete signature of the subtype. However, functional analyses will be required to decipher the role of these drivers and their possible functional interaction.

By analyzing biological and patients' characteristics as well as additional genetic alterations of CDX2/UBTF ALL patients, we highlighted several salient features. First, patients were strikingly enriched in females. The basis for this bias is still unknown and deserves investigation. Second, the immunophenotype of

leukemic cells was remarkable, with negativity of CD10 and CD20 (pro–B/B-I EGIL stage), a common feature of *KMT2A*-rearranged ALL but otherwise uncommon. CD34 high expression and CD9 negativity were also noted in both immunophenotypic data and gene expression signature. Third, the pattern of additional genomic alterations exhibited several notable features. They comprised a number of large chromosomal imbalances, including recurrent duplications in 1q arm, also observed by Yasuda et al,[38] and deletions in 3q, 6q, 9q, and 13q, which is not common in B-ALL and suggests underlying genomic instability. Of note, CDX2 has been previously involved in inhibition of DNA double-strand break repair activity in intestinal context.[47] Focal deletions commonly observed in B-ALL involving *IKZF1*, *PAX5*, and *ETV6* deletions were also observed in CDX2/UBTF ALL, with a notable high frequency of deletions of *PAX5* 5′ or promoter regions. *CREBBP* and RAS pathway alterations were relatively frequent in CDX2/UBTF ALL. Interestingly, this association has been reported in high-hyperdiploid B-ALL and linked to treatment resistance and relapse.[48] Finally, we identified recurrent mutations of *CXCR4*, which had never been reported to date in acute leukemias. *CXCR4* encodes a chemokine receptor broadly expressed in hematopoietic cells and critical for their interaction with the bone marrow niche and migration. Mutations observed in CDX2/UBTF ALL were somatically acquired and were seen in a minor leukemia fraction in some patients. Of note, 2 distinct *CXCR4* mutations were detected in 2 patients, which suggests specific constraints or cooperation in oncogenesis paths of CDX2/UBTF ALL. Further work is warranted to elucidate the significance of *CXCR4* mutations in CDX2/UBTF ALL and, possibly, the relevance of the therapeutic use of a CXCR4 antagonist to mobilize leukemic cells.

Prognosis of CDX2/UBTF ALL was analyzed within the GRAALL-2005 and 2014 trials, comprising 17 cases out of 723 with available molecular and clinical data. CDX2/UBTF ALL patients exhibited criteria of poor response to treatment at every stage of evaluation (ie, response at day 8, day 15, and end of induction). Of note, poor sensitivity to chemotherapy may be related to a low-proliferating condition of these leukemia cells, as suggested by normal WBC counts at diagnosis and negative enrichment in cell cycle genes and mTOR signaling. Three patients failed to reach CR after 1 induction phase, and all but 1 had positive postinduction MRD. Accordingly, a very high relapse rate was observed, beyond those of known poor-risk B-ALL subtypes like KMT2A rearrangement and low hypodiploidy. A poor prognosis was also reported in Yasuda et al, with lower disease-free and overall survival.[38] By contrast, OS of CDX2/UBTF ALL was not lower in the present study, likely because of efficient salvage therapies currently available. Overall, our study reveals a novel high-risk B-ALL subtype, which requires novel therapeutic developments.

Acknowledgments

The authors thank the patients and all the GRAALL investigators and biologists who contributed samples and data for this study. The authors thank the Saint-Louis Molecular Hematology laboratory for technical support, especially Agnès Pérus, Hélène Boyer, and Emilie Gaudas, and Loïc Maillard from the research team. The authors thank Pierre Lemaire and Clémentine Chauvel for expert advice on immunophenotypic data. The authors thank Delphine Ndiaye-Lobry and Alexandra da Silva Babinet for the help with A-DNA FISH experiments. They thank the IMAG'IC facility (member of the National Infrastructure France BioImaging [ANR-10-INBS-04]) of Institut Cochin. The authors thank the Platform of Biological Resources from Robert Debré hospital, Assistance Publique-Hôpitaux de Paris, Paris, France. Long-read WGS was performed on the Meary third-generation sequencing platform of Université de Paris. Sequencing of 4C-seq was performed by the GenomEast platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). UMR-944, EA-3518, UMR-1274, Meary platform, and GRAALL are members of the Institut Carnot OPALE.

This study was supported by grants from Association Laurette Fugain, La Ligue régionale du Haut-Rhin Contre le Cancer (2019-20), and Institut National du Cancer (PLBIO 19-289). M.P. was a recipient of a grant from ITMO Cancer of Aviesan Cancer Control Strategy, on funds administered by INSERM. This work was supported by THEMA, the national center for precision medicine in leukemia.

Authorship

Contribution: M.P. performed experiments, analyzed data, and wrote the manuscript; R.K. and S.G. performed experiments and contributed to data analysis and study design; F.S. analyzed gene expression data and contributed to study design and manuscript writing; J.C., A.G., N.K., L.H., and B.H. performed experiments and analyzed data; S.Q., T.S., and T.Y. performed bioinformatics analyses; L.L. and H.B. contributed to patients' annotations and diagnostic analyses; A.C. contributed to patient's sample and annotation; C.L., F.P., and J-N.F. contributed to study design; T.L., M.H., F.H., V.L., and H.D. contributed to patients' samples and clinical annotations and led the clinical trials; C.D.-D. contributed to data analysis and study design; J.S. contributed to data analysis, study design, and manuscript writing; N.B. contributed to patients' samples and clinical annotations, led the clinical trials, performed clinical statistical analyses, and contributed to manuscript writing; E.C. designed research, analyzed data, and wrote the manuscript; and all authors reviewed the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: M.P., 0000-0003-4620-8782; R.K., 0000-0002-9987-7237; S.G., 0000-0003-0491-9949; J.C., 0000-0002-8976-2890; T.Y., 0000-0002-3394-2083; N.K., 0000-0003-3568-3705; A.C., 0000-0001-5996-3037; F.H., 0000-0002-9576-8578; T.L., 0000-0002-3123-4948; M.H., 0000-0001-7777-5216; F.P., 0000-0001-8995-596X; J-N.F., 0000-0002-0971-3774; C.L., 0000-0003-0550-4921; H.D., 0000-0002-5454-6768; J.S., 0000-0002-8734-5356.

Correspondence: Emmanuelle Clappier, Laboratoire d'Hématologie, Hôpital Saint-Louis, 1 Ave Claude Vellefaux, 75010 Paris, France; e-mail: emmanuelle.clappier@aphp.fr.

Footnotes

Submitted 4 January 2022; accepted 12 March 2022; prepublished online on *Blood* First Edition 22 March 2022. DOI 10.1182/blood.2021014723.

*R.K., S.G., and F.S. contributed equally to this study.

RNA-seq, WGS, and ChIPseq data files from B-ALL patients have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession numbers E-MTAB-11364, E-MTAB-11400, and E-MTAB-11500. Files can be made available upon comprehensive request after internal approval according to the Saint-Louis Research Institute scientific policy and to French laws.

Send data sharing requests via e-mail to the corresponding author.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

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 USC section 1734.

REFERENCES

1. Look AT. Oncogenic transcription factors in the human acute leukemias. *Science.* 1997; 278(5340):1059-1064.
2. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet.* 2013; 381(9881):1943-1955.
3. Yeoh E-J, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell.* 2002;1(2):133-143.
4. Moorman AV. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica.* 2016; 101(4):407-416.
5. Huguet F, Chevret S, Leguay T, et al; Group of Research on Adult ALL (GRAALL). Intensified therapy of acute lymphoblastic leukemia in adults: report of the randomized GRAALL-2005 clinical trial. *J Clin Oncol.* 2018;36(24):2514-2523.
6. Ruminy P, Marchand V, Buchbinder N, et al. Multiplexed targeted sequencing of recurrent fusion genes in acute leukaemia. *Leukemia.* 2016;30(3):757-760.
7. Clappier E, Auclerc MF, Rapion J, et al. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia.* 2014;28(1):70-77.
8. Caye A, Beldjord K, Mass-Malo K, et al. Breakpoint-specific multiplex polymerase chain reaction allows the detection of IKZF1 intragenic deletions and minimal residual disease monitoring in B-cell precursor acute lymphoblastic leukemia. *Haematologica.* 2013;98(4):597-601.
9. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
10. Nicorici D, Satalan M, Edgren H, et al. FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *bioRxiv.* 2014
11. Den Boer ML, van Slegtenhorst M, De Menezes RX, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol.* 2009;10(2): 125-134.
12. Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med.* 2014;371(11):1005-1015.
13. Yasuda T, Tsuzuki S, Kawazu M, et al. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults [published correction appears in *Nat Genet.* 2016;48(12):1591]. *Nat Genet.* 2016;48(5):569-574.
14. Zhang J, McCastlain K, Yoshihara H, et al; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat Genet.* 2016;48(12):1481-1489.
15. Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nat Commun.* 2016;7(1): 11790.
16. Gu Z, Churchman M, Roberts K, et al. Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nat Commun.* 2016;7:13331.
17. Passet M, Boissel N, Sigaux F, et al; Group for Research on Adult ALL (GRAALL). *PAX5* P80R mutation identifies a novel subtype of B-cell precursor acute lymphoblastic leukemia with favorable outcome. *Blood.* 2019; 133(3):280-284.
18. Davidson AJ, Zon LI. The caudal-related homeobox genes cdx1a and cdx4 act redundantly to regulate hox gene expression and the formation of putative hematopoietic stem cells during zebrafish embryogenesis. *Dev Biol.* 2006;292(2):506-518.
19. Wang Y, Yabuuchi A, McKinney-Freeman S, et al. Cdx gene deficiency compromises embryonic hematopoiesis in the mouse. *Proc Natl Acad Sci USA.* 2008;105(22): 7756-7761.
20. Scholl C, Bansal D, Döhner K, et al. The homeobox gene CDX2 is aberrantly expressed in most cases of acute myeloid leukemia and promotes leukemogenesis. *J Clin Invest.* 2007;117(4):1037-1048.
21. Bonhomme C, Duluc I, Martin E, et al. The Cdx2 homeobox gene has a tumour suppressor function in the distal colon in addition to a homeotic role during gut development. *Gut.* 2003;52(10):1465-1471.
22. Riedt T, Ebinger M, Salih HR, et al. Aberrant expression of the homeobox gene CDX2 in pediatric acute lymphoblastic leukemia. *Blood.* 2009;113(17):4049-4051.
23. Rawat VPS, Cusan M, Deshpande A, et al. Ectopic expression of the homeobox gene Cdx2 is the transforming event in a mouse model of t(12;13)(p13;q12) acute myeloid leukemia. *Proc Natl Acad Sci USA.* 2004; 101(3):817-822.
24. Galland A, Gourain V, Habbas K, et al. CDX2 expression in the hematopoietic lineage promotes leukemogenesis via TGFβ inhibition. *Mol Oncol.* 2021;15(9):2318-2329.
25. Vu T, Straube J, Porter AH, et al. Hematopoietic stem and progenitor cell-restricted Cdx2 expression induces transformation to myelodysplasia and acute leukemia. *Nat Commun.* 2020;11(1):3021.
26. Atanassov BS, Mohan RD, Lan X, et al. ATXN7L3 and ENY2 coordinate activity of multiple H2B deubiquitinases important for cellular proliferation and tumor growth. *Mol Cell.* 2016;62(4):558-571.
27. Barros-Silva JD, Paulo P, Bakken AC, et al. Novel 5′ fusion partners of ETV1 and ETV4 in prostate cancer. *Neoplasia.* 2013;15(7): 720-726.
28. Fishbein L, Leshchiner I, Walter V, et al; Cancer Genome Atlas Research Network. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell.* 2017;31(2):181-193.
29. Umeda M, Ma J, Huang BJ, et al. Integrated genomic analysis identifies UBTF tandem duplications as a recurrent lesion in pediatric acute myeloid leukemia [published online ahead of print 16 February 2022]. *Blood Cancer Discov.* 2022;bloodcandisc.BCD-21-0160-A.2021. doi:10.1158/2643-3230.BCD-21-0160.
30. Stratmann S, Yones SA, Mayrhofer M, et al. Genomic characterization of relapsed acute myeloid leukemia reveals novel putative therapeutic targets. *Blood Adv.* 2021;5(3): 900-912.
31. Papaemmanuil E, Rapado I, Li Y, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet.* 2014;46(2):116-125.
32. Yang M, Safavi S, Woodward EL, et al. 13q12.2 deletions in acute lymphoblastic leukemia lead to upregulation of FLT3 through enhancer hijacking. *Blood.* 2020; 136(8):946-956.
33. Lengerke C, Daley GQ. Caudal genes in blood development and leukemia. *Ann N Y Acad Sci.* 2012;1266(1):47-54.
34. McKinney-Freeman SL, Lengerke C, Jang I-H, et al. Modulation of murine embryonic stem cell-derived CD41+c-kit+ hematopoietic progenitors by ectopic expression of Cdx genes. *Blood.* 2008;111(10):4944-4953.
35. Hernandez PA, Gorlin RJ, Lukens JN, et al. Mutations in the chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease. *Nat Genet.* 2003;34(1):70-74.
36. Balabanian K, Lagane B, Pablos JL, et al. WHIM syndromes with different genetic anomalies are accounted for by impaired CXCR4 desensitization to CXCL12. *Blood.* 2005;105(6):2449-2457.
37. Hunter ZR, Xu L, Yang G, et al. The genomic landscape of Waldenstrom macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood.* 2014;123(11):1637-1646.
38. Yasuda T, Sanada M, Kawazu M, et al. Two novel high-risk adult B-cell acute lymphoblastic leukemia subtypes with high expression of CDX2 and IDH1/2 mutations. *Blood.* 2022;139(12):1850-1862.
39. Zriwil A, Böiers C, Kristiansen TA, et al. Direct role of FLT3 in regulation of early lymphoid progenitors. *Br J Haematol.* 2018; 183(4):588-600.
40. Kumar N, Tsai Y-H, Chen L, et al. The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development. *Development.* 2019;146(5):dev172189.
41. Mahony S, Edwards MD, Mazzoni EO, et al. An integrated model of multiple-condition ChIP-Seq data reveals predeterminants of Cdx2 binding. *PLOS Comput Biol.* 2014; 10(3):e1003501.

42. Bell SP, Learned RM, Jantzen HM, Tjian R. Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science*. 1988;241(4870):1192-1197.

43. So CW, Cleary ML. Dimerization: a versatile switch for oncogenesis. *Blood*. 2004;104(4): 919-922.

44. Jeusset LM, McManus KJ. Characterizing and exploiting the many roles of aberrant H2B monoubiquitination in cancer pathogenesis [published online ahead of print 22 December 2021]. *Semin Cancer Biol*. doi: 10.1016/j.semcancer.2021.12.007.

45. Zhao Y, Lang G, Ito S, et al. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol Cell*. 2008; 29(1):92-101.

46. Stanek TJ, Gennaro VJ, Tracewell MA, et al. The SAGA complex regulates early steps in transcription via its deubiquitylase module subunit USP22. *EMBO J*. 2021;40(16): e102509.

47. Soret C, Martin E, Duluc I, et al. Distinct mechanisms for opposite functions of homeoproteins Cdx2 and HoxB7 in double-strand break DNA repair in colon cancer cells. *Cancer Lett*. 2016;374(2):208-215.

48. Malinowska-Ozdowy K, Frech C, Schönegger A, et al. KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. *Leukemia*. 2015;29(8): 1656-1667.