Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia

Bing Chen1,2,*†, Lu Jiang3,†, Meng-Ling Zhong1,†, Jian-Feng Li4,†, Ben-Shang Li5,1, Li-Jun Peng5,†, Yu-Ting Dai5, Bo-Wen Cui5,3, Tian-Qi Yan5, Wei-Na Zhang5, Xiang-Qin Weng5, Yin-Yin Xie5, Jing Lu5, Rui-Bao Ren5, Su-Ning Chen6, Jian-Da Hu7, De-Pei Wu8, Zhu Chen9,2, Jing-Yang Tang9,2, Jin-Yan Huang9,2, Jian-Qing Mi10,2, and Sai-Juan Chen9,2

1State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; 2Department of Hematology and Oncology, Shanghai Institute of Hematology, Shanghai Children’s Medical Center, Key Laboratory of Pediatric Hematology and Oncology, Ministry of Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China; 3Jiangsu Institute of Hematology, The First Affiliated Hospital of Soochow University, Soochow 215006, China; and 4Fujian Institute of Hematology, Fujian Medical University Union Hospital, Fuzhou 350001, China

Contributed by Zhu Chen, December 1, 2017 (sent for review September 29, 2017; reviewed by Hervé Dombret and Arnold Ganser)

T-cell acute lymphoblastic leukemia (T-ALL) is a clonal malignancy of immature T cells. Recently, the next-generation sequencing approach has allowed systematic identification of molecular features in pediatric T-ALL. Here, by performing RNA-sequencing and other genome-wide analysis, we investigated the genomic landscape in 61 adult and 69 pediatric T-ALL cases. Thirty-six distinct gene fusion transcripts were identified, with SET-NUP214 being highly related to adult cases. Among 18 previously unknown fusions, ZBTB16-ABL1, TRA-SALL2, and involvement of NXX2-1 were recurrent events. ZBTB16-ABL1 functioned as a leukemogenic driver and responded to the effect of tyrosine kinase inhibitors. Among 48 genes with mutation rates >3%, 6 were newly found in T-ALL. An aberrantly overexpressed short mRNA transcript of the SLC17A9 gene was revealed in most cases with overexpressed TAL1, which predicted a poor prognosis in the adult group. Up-regulation of HOXA, MEFC2, and LYL1 was often present in adult cases, while TAL1 overexpression was detected mainly in the pediatric group. Although most gene mutations were mutually exclusive, they coexisted with gene mutations. These genetic abnormalities were correlated with deregulated gene expression markers in three subgroups. This study may further enrich the current knowledge of T-ALL molecular pathogenesis.

Significance

To get more insights into the disease mechanism of T-cell acute lymphoblastic leukemia (T-ALL), particularly in an adult group, we addressed the genomic landscape in 130 patients, including 61 cases of adult T-ALL. A number of new genetic aberrations were identified using integrated transcriptome and genomic analysis. Distinct T-ALL subgroups were defined according to the interplay among different genetic abnormalities and gene transcription patterns. Characterization of genomic features of T-ALL is valuable not only for a better understanding of leukemogenesis, but also for patient stratification and tailored therapy.

T-ALL | transcriptome | fusion gene | ZBTB16-ABL1 | gene mutation

RAS-MAPK pathway genes and CDKN2A/2B deletions (2, 5, 12–15).

In recent years, next-generation sequencing techniques, including whole-genome sequencing, whole-exome sequencing (WES), and RNA sequencing (RNA-seq), have extended the list of genetic abnormalities in T-ALL to epigenetic factors and translation/RNA stability pathways (16–21). Of note, a subgroup of T-ALL with a characteristic immunophenotype (CD5+, CD1a−/CD8−, CD8‘weak’) has been designated as early T-cell precursor ALL (ETP-ALL), with special gene mutation patterns (16, 22). The integrated analysis of gene alterations in two recent series of pediatric T-ALL provided an enlarged view of the genomic landscape in this heterogenous disease (20, 23). However, complex interplay of gene fusions, sequence abnormalities, and transcriptional expression profiles, especially in adult cases, needs to be further addressed to refine the current model of T-ALL leukemogenesis and to reveal potential new biomarkers and therapeutic targets.

In this study, RNA-seq, WES, and other genomic analyses were performed in 61 adult and 69 pediatric T-ALL cases. A number of previously undescribed gene fusions, mutations, aberrant transcripts, and gene expression patterns were identified. Specifically, all authors declared no conflict of interest.


PNAS license:

Data deposition: The datasets reported in this paper have been deposited in the Chinese Leukemia Genotype-Phenotype Archive, bioinfo.rfh.com.cn/cga (accession nos. CGA500000000002).

(C)Licence

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1717125115/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1717125115
we also addressed the leukemogenic potential of the recurrent fusion gene ZBTB16-ABL1.

Results
Clinical Characteristics of Patients. Consecutive samples were enrolled in ChiCTR-ONRC-14004968 (Shanghai Institute of Hematology) and ChiCTR-ONC-14005003 (Shanghai Children’s Medical Center) trials. The clinical and hematological information is listed in Table 1 and Dataset S1. In the pediatric group, a complete remission (CR) rate of 92.8% (64/69) and the 3-y overall survival (OS) rate of 72.8% (95% CI 67.3–78.3) were achieved. In contrast, in adults, the CR rate was 75.4% (46/61) while the 3-y OS rate was 29.0% (95% CI 22.4–35.6) (SI Appendix, Fig. S1), revealing an inferior prognosis and the need for further improved treatment strategy.

Overview of Gene Fusion Transcripts. Based on RNA-seq data and reverse transcription PCR (RT-PCR) confirmation, 69 cases (53.08%) harbored a fusion gene. Thirty-six different fusions were found, including 18 known fusions covering common types such as STIL-TAL1, SET-NUP214, and NUP214-ABL1 (24–26), and 18 previously unknown fusions, of which five were identified in adults (Fig. 1, SI Appendix, Fig. S2, and Datasets S2 and S3). Although similar frequencies of gene fusions were found in the two groups (50.82% in adults vs. 55.07% in children), the types of fusions seemed less diverse in adults than in children (16 vs. 26) (Dataset S2). Of note, newly discovered ZBTB16-ABL1 and TRA-SALL2 fusions, and the involvement of NXX2-1, were recurrent events (Dataset S3). Four types of gene abnormalities were identified among previously unknown fusions. In ZBTB16-ABL1, the N-terminal moiety of ZBTB16 was fused to the body region of ABL1, defining a typical chimeric ORF (type 1). The type 2 fusion was represented by case C47 where the 5′ untranslated region (UTR) of EVL, a gene highly expressed in T-ALL cells found in this work (SI Appendix, Fig. S3a), was fused to the entire ORF of homeobox gene NXX2-1. This fusion led to an aberrantly high expression of NXX2-1 (SI Appendix, Fig. S3b). In the same case, an EVL-SFTA3 fusion was also detected, probably caused by a splicing between EVL and SFTA3, the latter being located downstream from NXX2-1 (Fig. 1). Type 3 fusion was characterized by genes fused to TCR-alpha, -beta, or -delta (TRA/ TRB/TRD) loci, leading to overexpression of SALL2 [a TF deregulated in various cancers (27)] and NXX2-1 (SI Appendix, Fig. S3c). Although the initiation ATG codon of SALL2 in cases C22 and C38 was deleted in the fusion mRNA, an alternative initiation ATG codon was found at amino acid 63. The major functional domains of SALL2 could thus be maintained. One case (C60) was found to carry a type 4 fusion, with a short transcript of PPP4R3A fused to IGH. Since PPP4R3A is a known tumor suppressor gene (28), the disruption of its ORF could cause gene inactivation. It is worth pointing out that in most cases (92.8%, 64/69), gene fusions were mutually exclusive.

Leukemogenic Power of the Recurrent Fusion Gene ZBTB16-ABL1. ZBTB16 (also known as promyelocytic leukemia zinc finger PLZF) contains one BTB domain and nine zinc fingers. The ZBTB16-ABL1 chimeric protein maintained the BTB and two or three zinc fingers (Fig. 2A). The ZBTB16-ABL1 occurred in two young children (case C11, 1 y old, accompanied by NOTCH1 and ZEB2 mutations and case C23, 2 y old, with mutations of PTEF, MYCN, and PIK3CD; Figs. 3 and 4). Both cases died within 1 y after diagnosis. When the proliferation rates of Jurkat cells transfected with ZBTB16-ABL1 or vehicle were compared, a stronger stimulatory effect on cellular proliferation by ZBTB16-ABL1 was observed (Fig. 2B). Moreover, ZBTB16-ABL1 promoted cell-cycle progression and DNA replication (Fig. 2C). We then examined protein tyrosine kinase (PTK) activity of ZBTB16-ABL1. The in vitro PTK activity of ZBTB16-ABL1 was fourfold that of ABL1, comparable to that of BCR-ABL1, which was fivefold the wild-type protein (Fig. 2D). Both fusion kinases had similar sensitivity to PTK inhibitors including imatinib and dasatinib (Fig. 2E).

In retrovirus-mediated bone marrow (BM) transplantation experiments, all 10 mice carrying ZBTB16-ABL1 developed chronic myeloid leukemia-like myeloproliferative disease and quickly succumbed to death (Fig. 1).
after transplantation (Fig. 2F and SI Appendix, Fig. S4). Massive elevation of maturing myeloid cells in BM, peripheral blood, and spleen was revealed. BM examination confirmed a significant increase in Mac-1+Gr-1+ myeloid elements (SI Appendix, Fig. S4).

The ZBTB16-ABL1 BMT mice treated with imatinib or dasatinib showed significantly prolonged survival (Fig. 2G).

Identification of Distinct Gene Expression Groups and Ablerrant RNA Splicing/Transcripts. Unsupervised clustering methods were applied to the classification of gene expression from T-ALL cases, and three distinct subgroups were identified by using a set of the 1,484 most differentially expressed genes (SI Appendix, Figs. S5 and S6). The well-known expression marker genes in T-ALL, such as TAL1, LMO1, LYL1, HOXA, TLX1, and TLX3, showed distinct clustering patterns. TLX1/TLX3 overexpressions were found in one subgroup (G1), and LYL1 expression was enriched mainly in another subgroup (G2), whereas TAL1 and LMO1 overexpressions were clustered in the third subgroup (G3). HOXA overexpression was found in both G1 and G2. We also searched for aberrant mRNA splicing products or transcripts. An unusual transcript of SLC17A9, encoding a transmembrane transporter for adenosine triphosphate (ATP) and other small molecules, was highly expressed in 58 of 130 cases. Different from well-annotated SLC17A9 transcripts containing 13 exons, this short transcript had only exon 9–13, with a 5’ UTR from intron 8 and a putative initiation ATG codon at the 314th amino acid in exon 9, leading to an ORF of 123 amino acids and a truncation of the major facilitator superfamily (MFS) domain (SI Appendix, Fig. S7). Karyotype analysis revealed chromosomal translocations in one case with TLX3 overexpression and in two cases with LMO2 up-regulation. Among cases with overexpressed LMO1/2, TAL1/2, NXX2-1, and LYL1, amplification of TAL1, NXX2-1, or LYL1 was detected by single-nucleotide polymorphism (SNP) array in four cases, whereas insertion of an enhancer region in exon 13 was found in eight other cases (Fig. 4).

Gene Mutation Profiling. RNA-seq data-based mutation detection was used in all 130 cases according to a recently published highly stringent procedure (29, 30). In total, 119 genes were found mutated at least twice (Datasets S4 and S5). Attention was given to 48 genes with mutation rates over 3% (4/130 cases), including 6 newly identified mutated genes [CELSR3, PAK4, MINK1, NR4A1, BOD1L1, and VCP] (Fig. 3 and SI Appendix, Fig. S8), two of them (PAK4 and BOD1L1) being mainly involved in adults. Abnormalities of NOTCH1, FBXW7, PHF6, JAK3, Pten, and JAK1 exhibited high mutation rates (74.6–100%). In addition, some gene families had multiple members affected by mutations, such as the histone-lysine methyltransferase family (KMT) and the ribosomal protein family (RPL) (Fig. 3). Mutated genes (>3%) were functionally divided into seven categories (C1–C7) (3): NOTCH1 pathway, signaling pathways, epigenetic factors, TFs, cell-cycle regulators, translation, and RNA stability-associated molecules, as well as others (Fig. 3). In the great majority (95.7%) of cases with gene fusions, at least one of the C1–C7 mutations coexisted, whereas FBXW7 and DNMT3A mutations occurred more frequently in cases without fusions (Dataset S6). Of note, the frequencies of mutations in adults were much higher than those in children (P = 0.001, SI Appendix, Fig. S9 and Dataset S7), especially among C2 and C3 genes. A significant correlation between the age and the number of gene mutations was found (R² = 0.1147, P = 0.001; SI Appendix, Fig. S9B). The status of CDKN2A/2B was investigated in 115 cases, and CDKN2A and CDKN2B deletions were found in 75 and 64 cases, respectively (Fig. 4).

Correlation Among Gene Fusions/Mutations/Ablerrantly Overexpressed Transcripts and Distinct Gene Expression Subgroups. A close correlation was found among gene fusions/mutation profiles/aberrant transcripts and gene expression patterns, enabling a synthetic view of genomic abnormalities in T-ALL (Fig. 4 and SI Appendix, Table S1). In the G1 subgroup [28 cases (21.5%), 14 adults and 14 children], 21 cases bearing overexpressed TLX1/TLX3 shared a similar transcriptome profile to 7 cases with high expression of HOXA family genes. The G2 subgroup (37 cases, 28.5%), including 91.7% of our ETP cases, was composed mainly of adults (31/37 cases). One feature of G2 was a close association with NUP family genes containing fusions (15/37 cases). An array of core TFs for early hematopoiesis (HOXA, LYL1, MEF2C, SP1I, RARA, ELK3, BLNK, STAP1, ZBTB46, BTK, and NFKBIE) was also highly expressed in G2. The fact that MEF2C is overexpressed in 30/37 (81.8%) cases (SI Appendix, Fig. S10) suggested a common disease mechanism (31). In addition, ETP cases had higher mutation rates of epigenetic factors, IDH2, DNMT3A, and EZH2 in particular,
than in other T-ALL cases (Dataset S8). Notably, all 20 cases with STIL-TAL1 and 4 cases with LMO12-TRA were clustered in G3 (65 cases [50%], 17 adult and 48 children), as were cases with variant fusions involving TAL1 or STIL. Several genes related to STIL-TAL1 (TAL1, RUNX1, MYB, ETS1, and BCL2L1B) were found overexpressed (Fig. 4), while up-regulated SIX6 was detected in 53 of 65 cases (81.5%). Interestingly, the aberrantly overexpressed SLC17A9 short transcript was grouped in G3 (55/65 cases), contributing to its unique signature. Scrutiny of gene abnormalities further divided G3 into two subgroups (G3a, 20 cases including 17 children; G3b, 45 cases comprising 14 adults and 31 children). The NOTCH1 (20 vs. 86.67%, P < 0.001) and FBXW7 (10 vs. 53.33%, P = 0.002) mutations were found significantly lower in G3a than in G3b (Dataset S9). Compared with G1/G2, G3 displayed lower gene mutation rates of JAK-STAT (4.62 vs. 46.15%, P < 0.001) and RAS (6.15 vs. 32.31%, P < 0.001) pathways, but a higher mutation frequency of the PI3K pathway (21.54 vs. 3.08%, P = 0.003) and the PTEN gene (18.46 vs. 1.54%, P = 0.003).

**Genetic Characteristics of T-ALL Patients According to Immunophenotypes.** In ETP cases, the genetic feature was characterized by more SET-NUP214 fusions, overexpression of HOX4/MEF2C/LYL1, and high frequencies of mutations of RAS pathway/epigenetic factors. The genetic characteristics of early precursor (pro) cohorts were similar to those of ETP. However, among groups of precursor (pre), cortical, and medullary T-ALL, the frequencies of STIL-TAL1 and aberrantly overexpressed SLC17A9 short transcript were much higher (Fig. 4). Among 18 newly discovered gene fusions, ZBTB16-ABL1, TRA-SALL2 fusions, and involvement of NXX2-1 were recurrent events. It is well known that ZBTB16 functions as a hematopoietic regulator and is a fusion partner to retinoic acid receptor alpha (RARα) in a subset of acute promyelocytic leukemia (APL) (32).

In adults, cases with the aberrantly overexpressed short SLC17A9 transcript exhibited very poor outcome, with 3-y OS and 3-y EFS rates of 10.3% (95% CI 1.0–19.6) and 10.5% (95% CI 1.1–19.9), respectively. Conversely, eight cases with SET-NUP214 showed relatively good prognosis, with a 3-y OS rate of 87.5% (95% CI 75.8–99.2) and a 3-y EFS rate of 70.0% (95% CI 51.8–88.2) (Fig. 5).

**Discussion**

In this study, by performing genomic landscape analysis, we identified a series of previously unknown fusion genes, gene mutations, and aberrant mRNA transcripts in 130 Chinese T-ALL cases. These findings were made through scrutiny of the molecular data, and it is possible that some features could be ascribed to the ethnic genetic background of T-ALL from the Chinese population. The genomic information of T-ALL in adults was also compared with that of pediatric cases. Taken as a whole, each of the 61 adult and 69 child T-ALL cases in our series harbored at least one major genetic lesion, including gene fusions, mutations of C1-C7 genes, and aberrant expression of genes key to leukemogenesis. This dataset further enriched our understanding of T-ALL and may be translated into useful clinical stratification markers or drug targets. Indeed, molecular markers predicted three distinct risk groups in our adult T-ALL cases.
disturbed hematopoiesis. More importantly, ZBTB16-ABL1 also
had SH3, SH2, and the tyrosine kinase domain of ABL1 and pos-
sessed constitutive PTK acti-
vity with strong transforming
power (33). The fact that leukemia mice bearing
ZBTB16-ABL1
responded to the therapeutic effect of PTK inhibitors suggested
a potential clinical application of the drugs in relevant cases. In
the future, patient-derived xenograft animal models will be used
to further investigate the leukemogenesis and drug response.

An interesting finding of this work is that the number of gene
mutations per case was higher in adult than in child T-ALL. The
gene fusions/mutations and transcriptome features in our T-ALL
series clearly showed an interplay of essential genetic abnormalities.

Fusion genes rarely coexisted in the same case, suggesting that they
play driver roles. However, fusions were often accompanied by at
least one of the C1–C7 mutations, suggesting possible cooperative
effects. Another dimension of genomic abnormalities lies at the level
of transcription. Among 62 cases without fusions, 41 (66.1%) har-
bored deregulated expression of TAL1, LYL1, LMO1/2, TLX1,
TX3, HOX4, and NKX2-1 genes, at least in part due to intergenic
structural aberrations or enhancer mutations. With regard to tran-
scriptome analyses, SPI1 overexpression is noteworthy, since it serves
as a regulator for MEF2C expression in normal lymphoid develop-
ment and is highly expressed in prethymic progenitors (34). Recently,
SPI1 was found to form a fusion gene with STMN1 or TCF7 (35).

Fig. 4. Integrated phenotypic and molecular architecture of leukemic cells from 61 adult and 69 pediatric T-ALL patients. (Top) Age, gender, and immu-
nophenotype for each sample within each of the three unique gene expression groups: TLX1/3/HOXA (G1), ETP/LYL1/HOXA (G2), and TAL1/LMO1 (G3). The
second vertical panel presents aberrant gene overexpression known to be associated with chromosome rearrangements and/or fusion transcripts. White
labels on red rectangle are used to indicate cases with chromosomal translocations possibly involving TLX3 [t (1, 5) (p11;q35)] and LMO2 [t (11, 14) (p13;q11)]
revealed by karyotype analysis (circles; Dataset S1), with amplification of TAL1, NKX2-1, or LYL1 genes revealed by SNP array (triangles) or with insertion of
the enhancer regions of TAL1, LMO1, or LMO2 genes (plus signs), respectively. The third vertical panel (heatmap) shows the abnormal expression pattern of a
group of genes potentially involved in early T-cell differentiation and hematopoietic regulation. Gene names with an aberrant expression pattern revealed in
this work are in red. In the fourth vertical panel, aberrant transcript and fusion genes are presented. (Bottom) CDKN2A/2B deletions and mutations from C1 to
C4 categories.
Although no such fusion was identified in our cohort, simultaneous high expression of SPI1 and ME22C in a subset of our cases might be essential in leukemogenesis. In addition, most cases harboring TIL-TAL1/TAL1 overexpression presented an aberrantly overexpressed SLC17A9 transcript with truncation of the MFS functional domain, and overexpression of this transcript predicted a poor outcome in adult patients. Further investigation is warranted to explore the role of this aberrantly overexpressed transcript in T-ALL pathogenesis. Moreover, larger cohort studies are needed to validate the prognostic significance of the overexpression of SLC17A9 short transcript and the SET-NUP214 fusion in adult T-ALL.

Materials and Methods

Patients. The study cohort was composed of 61 adults and 69 pediatric patients who were followed from 2007 to 2016. The patients’ clinical study was approved by the ethical boards of the participating centers (Institutes of Hematology in Shanghai, Soochow, and Fuzhou). All patients gave informed consent for treatment and cryopreservation of BM and peripheral blood samples according to the Declaration of Helsinki. Details of treatment protocols are available in SI Appendix, SI Materials and Methods.

RNA-Seq, WES Data, and CDKNA2A/2B Analyses. RNA-seq was performed according to a previously described method (36). WES was performed in 36 patients, 16 individuals having their own normal control samples (blood samples in CR). Reading pairs were aligned to the human reference genome hg19. Procedures of reading pairs alignment, mutation calling from RNA-seq or WES data, and the gene expression/pathway analysis are listed in SI Appendix, SI Materials and Methods. RT-PCR was used to confirm the newly identified fusion genes. A SNP array was performed to detect the intergenic rearrangements. Detection of CDKNA2A/2B deletion is described in SI Appendix, SI Materials and Methods.

Functional Study of the ZBTB16-ABL1 Fusion Gene. The wild-type ZBTB16 and ABL1 CDNA clones were kindly provided by OriGene. Procedures for plasmid construction, proliferation assay, cell-cycle analysis, PTyr assay, and the drug inhibition test are described in SI Appendix, SI Materials and Methods. All animal experiments were approved by the Animal Care and Use Committee of Shanghai Jiao Tong University School of Medicine. Detailed information is available in SI Appendix, SI Materials and Methods.

Data Availability. The datasets have been deposited in the Chinese Leukemia Genotype-Phenotype Archive (bioinfo.rjh.com.cn/cga) under accession no. CGA00000000000002.

Acknowledgments. This work was supported by the Chinese National Key Basic Research Project 973 (Grant 2013CB636800); the Chinese Ministry of Health Project (Grant 2012ZX09503-002); the Mega-Projects of Scientific Research for the 12th Five-Year Plan (2013ZX09303002); the National Natural Science Foundation of China (Grants 81470311, 81670137, 81570122, 81670147); the National Key Research and Development Program (Grant 2016YFC0902800); the Samuel Waxman Cancer Research Foundation; and the Center for High Performance Computing at Shanghai Jiao Tong University.