PNAS RESEARCH ARTICLE | MEDICAL SCIENCES

Transcriptome-wide subtyping of pediatric and adult T cell acute lymphoblastic leukemia in an international study of 707 cases

Yu-Ting Dai[a,1], Fan Zhang[a,1], Hai Fang[a,1], Jian-Feng Li[a,b], Gang Lu[a], Lu Jiang[a], Bing Chen[a], Dong-Dong Mao[a], Yuan-Fang Liu[a], Jin Wang[a], Li-Jun Peng[a], Chong Feng[a,b], Hai-Feng Chen[c], Jun-Xi Mu[c], Qun-Ling Zhang[d], Hao Wang[e], Hany Ariffin[f], Tan Ah Moy[g], Jing-Han Wang[h], Yin-Jun Lou[h], Su-Ning Chen[i,j], Qian Wang[i], Hong Liu[i], Zhe Shan[i], Itaru Matsumura[k], Yasushi Miyazaki[l], Takahiko Yasuda[m], Li-Ping Dou[e], Xiao-Jing Yan[n], Jin-Song Yan[o], Allen Eng-Juh Yeoh[p,q,r], De-Pei Wu[i,j], Hitoshi Kiyoi[s], Fumihiko Hayakawa[t], Jie Jin[h], Sheng-Yue Wang[a], Xiao-Jian Sun[a,2], Jian-Qing Mi[a,2], Zhu Chen[a,2], Jin-Yan Huang[u,v,2], and Sai-Juan Chen[a,2]

Contributed by Zhu Chen; received November 18, 2021; accepted February 28, 2022; reviewed by Mark Lathrop and Bin-Bing Zhou

T cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy of T cell progenitors, known to be a heterogeneous disease in pediatric and adult patients. Here we attempted to better understand the disease at the molecular level based on the transcriptomic landscape of 707 T-ALL patients (510 pediatric, 190 adult patients, and 7 with unknown age; 599 from published cohorts and 108 newly investigated). Leveraging the information of gene expression enabled us to identify 10 subtypes (G1–G10), including the previously undescribed one characterized by *GATA3* mutations, with *GATA3*^{R276Q} capable of affecting lymphocyte development in zebrafish. Through associating with T cell differentiation stages, we found that high expression of *LYL1/LMO2/SPI1/HOXA* (G1–G6) might represent the early T cell progenitor, pro/precortical/cortical stage with a relatively high age of disease onset, and lymphoblasts with *TLX3/TLX1* high expression (G7–G8) could be blocked at the cortical/postcortical stage, while those with high expression of *NKX2-1/TAL1/LMO1* (G9–G10) might correspond to cortical/postcortical/mature stages of T cell development. Notably, adult patients harbored more cooperative mutations among epigenetic regulators, and genes involved in JAK-STAT and RAS signaling pathways, with 44% of patients aged 40 y or above in G1 bearing *DNMT3A/IDH2* mutations usually seen in acute myeloid leukemia, suggesting the nature of mixed phenotype acute leukemia.

T-ALL | RNA sequencing | molecular subtyping | mutation | T cell differentiation stages

T cell acute lymphoblastic leukemia (T-ALL) is characterized by malignant transformation and proliferation of T cell progenitors (1, 2), accounting for 10 to 15% of pediatric and 20 to 25% of adult ALL cases (2, 3). Therapeutic progress has led to a gradual improvement in clinical outcomes, with a curable rate achieving up to 90% in children but much lower rate (60%) in adults (4). Recent advances in high-throughput genomic technologies have spurred the cohort-scale genetic analysis for identifying recurrent genetic abnormalities in T-ALL (5, 6). Recurrent genetic events in lymphoid neoplasia have been reported to cooperatively induce malignant transformation of normal thymocytes through transcriptional deregulation, leaving traces in the form of specific expression patterns (7–9). In most T-ALL patients, genetic abnormalities often involve genes encoding transcription factors, as highlighted by lymphoblastic leukemia-associated factors (e.g., *LYL1*, *LMO1/2*, *TLX1/3*, *NKX2-1*, and *TAL1/2*) (10). These genes may exhibit aberrant expression levels, when structural variations involve the T cell receptor gene (*TCR*) or enhancer regions from other partner genes: for example, the transcription factor gene *TAL1* in the *STIL-TAL1* deletion (11). Most commonly seen in early T cell precursor ALL (ETP-ALL) patients are the *SET-NUP214* fusion and fusions involving the gene *NUP98* (12). In addition to gene fusions and rearrangements, mutations are observed in over 90% of T-ALLs (9). Driver mutations can occur in genes essential for regulating T lymphocyte development, to name but a few: the NOTCH signaling (*NOTCH1*, *FBXW7*, *NOTCH3*), the JAK-STAT signaling (*JAK3*, *STAT5B*), the PI3K-AKT-mTOR (mammalian target of rapamycin) signaling (*PIK3R1*, *PIK3CD*), epigenetic regulators (*PHF6*, *USP7*), PRC2 complexes (*EZH2*, *SUZ12*), and transcription factors (*BCL11B*, *ETV6*, *GATA3*). Mutations can also occur in genes essential for modulating cell proliferation and differentiation, such as the RAS signaling (*KRAS*, *NRAS*), cell cycle, or apoptosis-related factors (*CDKN2A*, *CCDN3*), and translational regulators (*RPL10*, *RPL5*). In staging of T cell maturation, T-ALL can be arrested at different stages of T cell development. Based on the clinical immunophenotypes, T-ALL has been

Significance

We provide transcriptomic insights into differences between pediatric and adult T cell acute lymphoblastic leukemia (T-ALL) patients through an international collaborative effort integrating RNA-sequencing data of 707 patients. Ten subtypes were identified, each characterized by distinct gene mutation profiles and dysregulated expression signatures of leukemogenic factors, and associated with T cell development stages. Adult T-ALL tends to have characteristics of early T cell precursor ALL, mostly corresponding to the mixed phenotype acute leukemia, whereas pediatric T-ALL shows a wide spectrum of aberrant molecular features, from early T cell precursor to mature T cell compartments. Our findings have important implications for disease mechanism of T-ALL that differs between pediatric and adult patients, facilitating further refined targeted therapy.

Reviewers: M.L., McGill University and Génome Québec Innovation Centre; and B-.B.Z., Key Laboratory of Pediatric Hematology & Oncology Ministry of Health, Shanghai Children's Medical Center.

The authors declare no competing interest.

[1]Y.-T.D., F.Z., and H.F. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: xjsun@sibs.ac.cn, jianqingmi@shsmu.edu.cn, zchen@stn.sh.cn, huangjinyan@zju.edu.cn, or sjchen@stn.sh.cn.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2120787119/-/DCSupplemental.

Published April 6, 2022.

subclassified into ETP-ALL, pro/precortical, cortical, postcortical, and mature T-ALL (5, 13). Notably, the recurrent genetic alterations show a degree of association with T cell development stages (14, 15), while combining both information from genetic alterations and differentiation arrests for improved subtyping has not yet been reported.

It is well recognized that T-ALL display a large variability of clinical and genetic features between pediatric and adult patients (9). Previously we described the landscape of B cell precursor-ALL molecular abnormalities, proposing 14 distinct subtypes based on RNA-sequencing (RNA-seq) data of 1,223 patients (7). By analogy, we here aimed to elucidate genetic/transcriptomic alterations alteration in T-ALL, in particular genomic insights into differences between pediatric and adult patients. We aggregated the evidence from RNA-seq data of 707 T-ALL patients (510 pediatric, 190 adult, and 7 with unknown age), including 599 obtained from six published international public cohorts (cohorts 1 to 6) (5, 9, 16–19) and 108 newly contributed from two more centers of excellence in China (cohorts 7 and 8) (*SI Appendix*, Table S1 and Dataset S1). All RNA-seq data were uniformly preprocessed and subjected to integrative analysis using a well-established pipeline (Fig. 1*A*) (7, 20, 21), generating resources on gene expression and genetic alterations.

Results

Discovery of Gene Fusions and Characterization of T-ALL Subtypes Based on RNA-seq Data from 707 Cases. We carried out an integrated analysis combining RNA-seq data from multiple

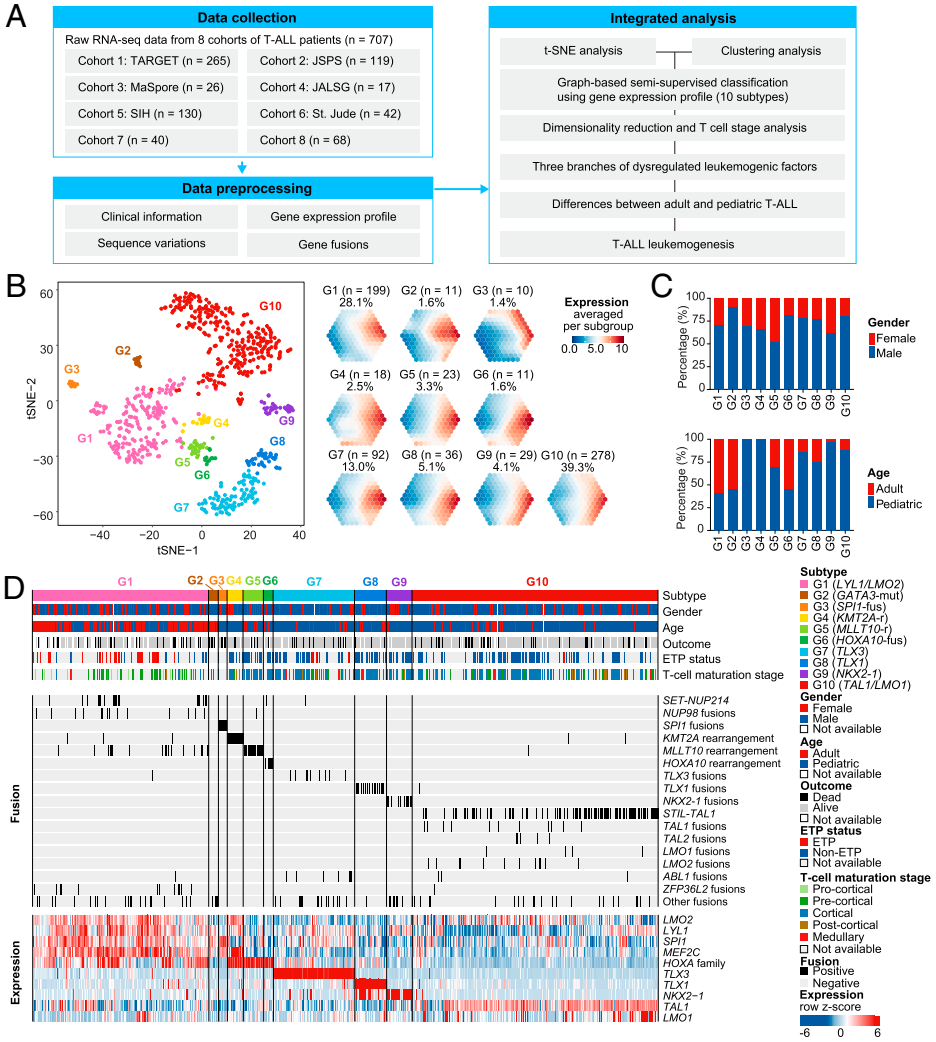

Fig. 1. Overview of molecular subtypes of T-ALL. (*A*) Overview of the T-ALL study workflow. RNA-seq data of T-ALL patients from eight cohorts are collected and integrated. After quality control, the gene-expression profile, sequence variations, and gene fusions identified from RNA-seq data are subjected to further analysis. tSNE analysis and hierarchical clustering methods are applied to determine the subtypes of T-ALL. (*B*) Two-dimensional tSNE plot and suprahexagonal map of 707 T-ALL patients. On the tSNE plot, each dot represents one T-ALL patient. The top 5% of genes demonstrating variance (with a perplexity score of 15 and a θ-value of 0.2) are subjected to tSNE analysis. Patient samples are colored according to the subtypes. Shown, *Right*, are illustrations of subtype-specific expression using a suprahexagonal map. (*C*) Bar plot of the percentage of patients based on age and gender in each subtype. (*D*) Profiling of clinical characteristics and genetic features identified in 707 T-ALL patients. Columns indicate T-ALL patients, and rows represent three panels: clinical information panel (subtypes, age, gender, clinical outcome, ETP status, T-cell maturation stage), fusion panel (gene fusions, including fusions reported in the original study from public cohorts and identified in RNA-seq), and expression panel (gene-expression level of dysregulated leukemic factors). Patient samples are ordered according to the unsupervised hierarchical clustering within each subtype. For the gene-expression panel, up- and down-regulated genes are shown in the heatmap in red and blue, respectively. Ten subtypes are defined according to their molecular features: G1 (*LYL1/LMO2* overexpression, *LYL1/LMO2*), G2 (*GATA-3* mutation, *GATA-3* mut), G3 (*SPI1*-fusion, *SPI1*-fus), G4 (*KMT2A*-rearrangement, *KMT2A*-r), G5 (*MLLT10*-rearrangement, *MLLT10*-r), G6 (*HOXA10*-fusion, *HOXA10*-fus), G7 (*TLX3* overexpression probably due to fusion to *TCR*, *TLX3*), G8 (*TLX1* overexpression probably due to fusion to *TCR*, *TLX1*), G9 (*NKX2-1* overexpression, *NKX2-1*), and G10 (*TAL1/LMO1* overexpression, *TAL1/LMO1*).

cohorts totaling 707 T-ALL patients (Fig. 1*A*). Based on the information of gene expression with batch effects corrected across cohorts (*SI Appendix*, Fig. S1 *A* and *B*), 10 distinct subtypes were identified using a graph-based semisupervised classification approach (*SI Appendix*, *SI Materials and Methods*). The classification of each subtype was evident in both visual inspection and robust analysis using random forest (*SI Appendix*, Fig. S1 *C*–*G*). The tightness of patient groupings was illustrated using the two-dimensional representation determined by *t*-distributed stochastic neighbor embedding (tSNE), complemented with subtype-specific expression illustrations using a suprahexagonal map, collectively revealing the (dis)similarity among subtypes (Fig. 1*B*).

We next characterized subtypes regarding demographic information, gene fusion events, and dysregulated expression of leukemogenic factors (Fig. 1 *C* and *D* and Dataset S2). In G1, G2, and G6, over 50% of patients were adults, representing the highest proportion of adult patients among all subtypes (Fig. 1*C*). G1 was associated with *SET-NUP214* and *NUP98* fusions and the elevated expression of *LYL1* (participating in lymphomagenesis) (22), *LMO2* (perturbing T-cell differentiation) (23), *SPI1* (expressing in ETPs and essential for the normal hematopoiesis) (24), *MEF2C* (activated in T-ALL) (25), and *HOXA* family genes (Fig. 1*D*, expression panel). Notably, ETP-ALL patients were mainly classified into G1 as compared to other subtypes (42 of 59 vs. 7 of 215, $P = 1.7e\text{-}32$, χ^2 test) (Fig. 1*D* and Dataset S2). The G2, tightly located close to G1 (Fig. 1*B*), was identified as a subtype, which was represented by *GATA3* mutations, while no significant fusion transcripts were detected (Fig. 1*D*). The G3, another subtype close to G1, contained patients all harboring *SPI1* fusions (*TCF7-SPI1* and *STMN1-SPI1*). Patients in G4, G5, and G6 exhibited overexpression of *HOXA* family genes (Fig. 1*D*). All patients in G4 harbored *KMT2A* fusions and *MLLT10* rearrangements were regarded as key fusion events in G5, whereas *HOXA10* fusions mainly occurred in G6 (Fig. 1*D*). T-ALLs in G7 and G8 were, respectively, characterized by *TLX3* and *TLX1* overexpression (Fig. 1*D*). The G9 subtype was unique because of the overexpression of *NKX2-1* (interfering T cell differentiation by ectopic expression) (26). The G10 subtype was featured by diverse fusion events (such as *STIL-TAL1*, *TAL2* fusions, *LMO2* fusions, and *LMO1* fusions), and also by the overexpression of *TAL1* (participating in an oncogenic transcriptional program) (27) and *LMO1* (altering T-cell differentiation together with *TAL1*) (28) in most patients (Fig. 1*D*).

In summary, transcriptome-driven molecular subtyping is biologically relevant, with each subtype associated with unique molecular abnormalities, namely: G1 (*LYL1/LMO2* overexpression), G3 (*SPI1* fusion), G4 (*KMT2A* rearrangement), G5 (*MLLT10* rearrangement), G6 (*HOXA10* fusion), G7 (*TLX3* overexpression probably involving *TCR* fusion), G8 (*TLX1* overexpression probably involving *TCR* fusion), G9 (*NKX2-1* overexpression), and G10 (*TAL1/LMO1* overexpression). The identification of subtype G2 motivated further investigation into its molecular and functional mechanism.

Evidence for Subtypes from Nonsilent Gene Mutations in T-ALL. Using the previously established workflow (7, 9), we sought to explore evidence from our RNA-seq dataset, allowing the identification of nonsilent gene mutations with high sensitivity and specificity (*SI Appendix*, Fig. S1*H*). The number of nonsilent mutations detected in patients was significantly correlated with age (Spearman correlation coefficient $R = 0.44$, $P < 0.0001$) (*SI Appendix*, Fig. S1*I*). We identified a total of 2,380 candidate mutated genes, with 78 recurring in >1% of T-ALL patients. These 78 genes were broadly grouped into 9 functional categories (C1 to C9): NOTCH signaling (C1), epigenetic regulators (C2), transcription factors (C3), PI3K-AKT-mTOR signaling (C4), JAK-STAT signaling (C5), RAS signaling pathway (C6), translation (C7), proliferation/apoptosis (C8), and others (C9) (Datasets S3–S5). Mutations distributed among 10 subtypes (G1 to G10) are detailed in *SI Appendix*, Fig. S2 and summarized based on functional categories (C1 to C9) (Fig. 2 *A*, *Middle*). The most frequently mutated genes included *NOTCH1* (492 of 707, 69.6%), followed by *FBXW7*, *PHF6*, *PTEN*, and others (Fig. 2 *A*, *Bottom*). Most mutations in *FBXW7*, *PHF6*, and *PTEN*, three well-known tumor suppression genes, were loss-of-function in nature (29–31). Though with a much lower frequency (3.5%), *GATA3* mutations were highly enriched ($P = 8.7e\text{-}18$, Fisher's exact test) in G2 (Fig. 2*A*), thus supporting the notation of G2 (*GATA3*-mut) (Fig. 1*D*).

Mutated genes were functionally diverse (Fig. 2*A*), possibly acting as a coherent network. To support this, we integrated the knowledge of gene interactions (defined by Kyoto Encyclopedia of Genes and Genomes [KEGG] pathways). A network of mutated genes emerged, which was useful to explore relationships between mutated genes and subtypes (Fig. 2*B*). We next explored the percentage of mutations in each subtype and compared the frequency of the given mutations to that in other subtypes (Fig. 2*C* and Dataset S6). For example, mutations of *NOTCH1* were found in all subtypes, but the rates were relatively lower in G1 (111 of 199, 55.78%), G4 (10 of 18, 55.56%), and G10 (183 of 278, 65.83%), suggestive of an tendency of independence, and higher in other subtypes, including G6 (11 of 11, 100%) and G2 (10 of 11, 90.9%), suggestive of cooccurrence (Fig. 2*C*). As expected, the *GATA3* mutations most significantly clustered in G2 (Fig. 2*C*). For other subtypes, key findings are summarized: among patients in G1 (*LYL1/LMO2*), mutated genes in epigenetic regulators (*PHF6*, *ASXL1*, *CHD4*, *EZH2*, *SETD2*, *DNMT3A*, and *IDH2*), transcription factors (*WT1*, *RUNX1*, *ETV6*, *MED12*, *IKZF1*), JAK-STAT signaling (*JAK3*, *JAK1*), RAS signaling (*NRAS*, *KRAS*), and spliceosome complex (*U2AF1*) (P value < 0.05) (Dataset S6) showed concurrent tendency (Fig. *2C*). Gene mutations in the PI3K-AKT-mTOR signaling were enriched in G10 (*TAL1/LMO1*) (C4, $P = 3.5e\text{-}24$) (Dataset S6). We also examined overall correlations between gene–gene (Dataset S6) and category–category correlations (Dataset S6). The results revealed mutations in PI3K-AKT-mTOR signaling (C4) were mutually exclusive with those in the JAK-STAT signaling (C5, $P = 1.2e\text{-}5$) and the RAS signaling (C6, $P = 1.5e\text{-}3$) (*SI Appendix*, Fig. S3 and Dataset S6). Of note, mutations in isocitrate dehydrogenases (*IDH2*) significantly cooccurred with the epigenetic modifier *DNMT3A* and *NRAS* mutations (*SI Appendix*, Fig. S3 and Dataset S6). We also noticed that $RPL10^{R98S}$ hotspot mutation formed two subclusters (G9 and G10) (*SI Appendix*, Fig. S2), which was reported to be associated with young age T-ALL, and altered T cell development by enhancing JAK-STAT signaling (32, 33).

***GATA3*^{R276Q} Functions as a Driver Gene in T-ALL Leukemogenesis.** Given that *GATA3* point mutations were identified to signify subtype G2, we next addressed their functional features. GATA3 is a key transcriptional regulator for T cell development through binding to the DNA consensus sequence GATA (34). The elevation of *GATA3* gene expression was found in T cell lymphoproliferative disorders (35, 36) and solid

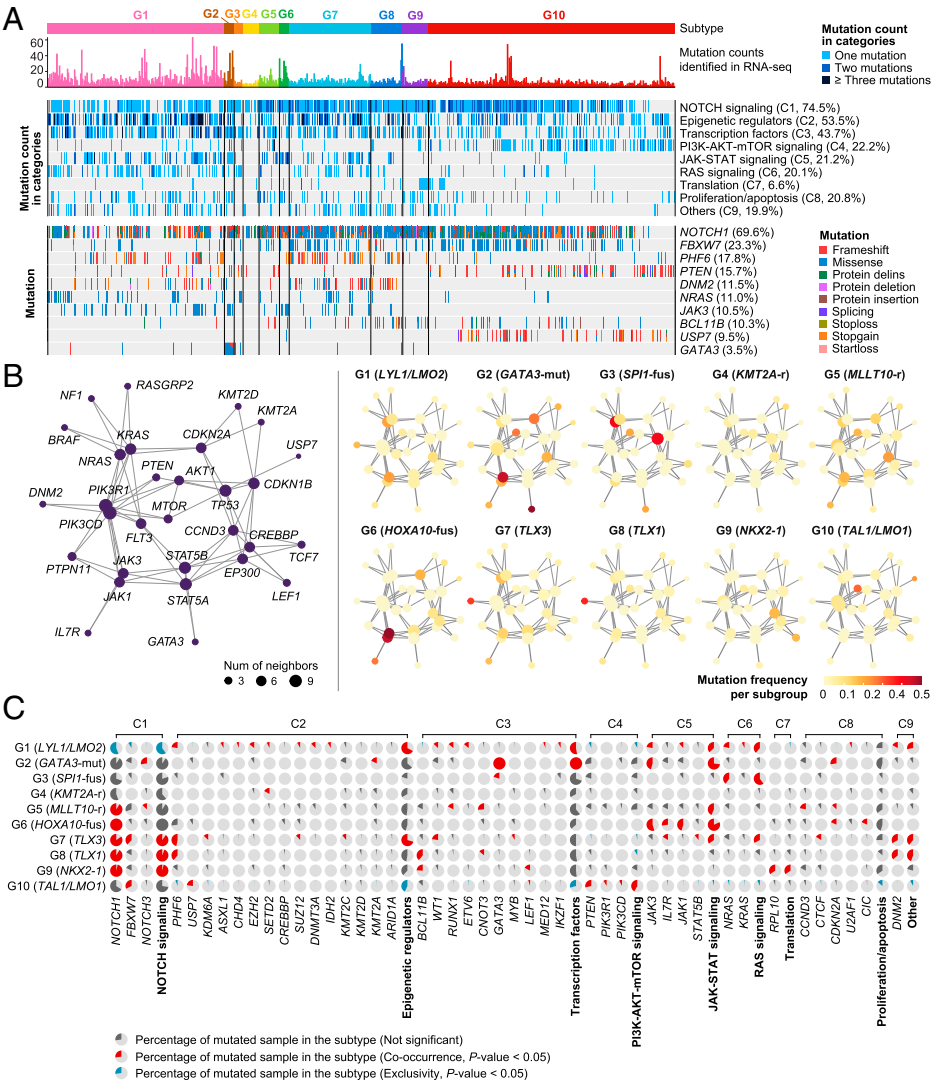

Fig. 2. The landscape of molecular interaction and pairwise relationship between nonsilent gene mutations. (*A*) Profiling of nonsilent gene mutations identified in 707 T-ALL RNA-seq. Mutation counts, gene mutations with high frequencies, and mutations in different categories are illustrated in three panels. In the *Top*, the number of mutations identified in RNA-seq data are illustrated as a barplot. In the *Middle*, genes with over 10% mutation frequency in T-ALL, as well as *USP7* (9.5%) and *GATA3* (used to discriminate *GATA*-mut subtype), are visualized. In the *Bottom*, mutation events in different categories are summarized using a blue label. (*B*, *Left*) Network visualization of mutated genes with edges defined by the knowledge of gene interactions from KEGG pathways. (*Right*) The same network but with nodes color-coded by the subtype-specific mutation frequency. (*C*) Comparison of the percentage of mutations in each subtype. Tendencies of cooccurrence and independence/exclusivity between gene mutations and subtypes are calculated, respectively. Red pies represent statistically significant cooccurrence, blue ones indicate statistically significant exclusivity, while gray ones show tendencies of gene mutation relationship that does not reach to statistical significance. Statistical significance of cooccurrence and exclusivity is calculated by comparing the mutations frequency in this subtype with other subtypes using χ^2 test (when cases in all conditions >5) or Fisher's exact test. Due to the limited sample sizes in some subtypes, some tendencies of relationship between gene mutations and subtypes could not always reach to statistical significance. Statistical results between mutations and subtypes are listed in Dataset S6.

tumors, such as breast cancer (*SI Appendix*, Fig. S4 *A* and *B*). Indeed, most mutations were clustered on the DNA binding domain with an intact open reading frame (ORF), in contrast to some cases of *GATA3* mutations distributed in other subtypes with a truncated ORF of the genes (Fig. 3*A*). Previously, the lack of or aberrant gene expression of *GATA3* was linked to cancerogenesis, especially in leukemia (37). Despite limited sample size, the prognosis of G2 cases seemed to be poor (using available survival data, 5 dead and 2 alive in G2, 112 dead and 455 alive in other subtypes, $P = 4.8e\text{-}3$ by Fisher's exact test). The mutants identified in patients in G2 (*GATA3*-mut) were all located on the N-finger domain of GATA3 protein, such as R276Q ($n = 5$) (Fig. 3*A*). The expression level of *GATA3* in G2 was significantly elevated, further supporting the importance of *GATA3* in leukemogenesis (Fig. 3*B*). We also noticed that the expression levels of altered *GATA3* were significantly higher in G2 (Fig. 3*C*), which was not observed in other subtypes (*SI Appendix*, Fig. S4 *C* and *D*). Although one case in G2 had lower *GATA3* expression (*SI Appendix*, Fig. S3*B*), the variant allele frequency of $GATA3^{R276Q}$ was 0.891 (Dataset S5) and this case harbored $NRAS^{G13D}$. Based on the crystal structure of GATA3, there exist two types of GATA3/DNA complexes (38): one for the protein binding to a palindromic DNA site (the wrapping structure) (*SI Appendix*, Fig. S4*E*), the other where GATA3 targets two separate DNA molecules (the bridging structure) (*SI Appendix*, Fig. S4*F*). The residue R276, located on the zinc core module of the N-finger, likely interacts with the DNA binding sites (*SI Appendix*, Fig. S4 *E* and *F*).

To evaluate the binding affinity, we built two different GATA3/DNA complex sets (wrapping vs. bridging conformations),

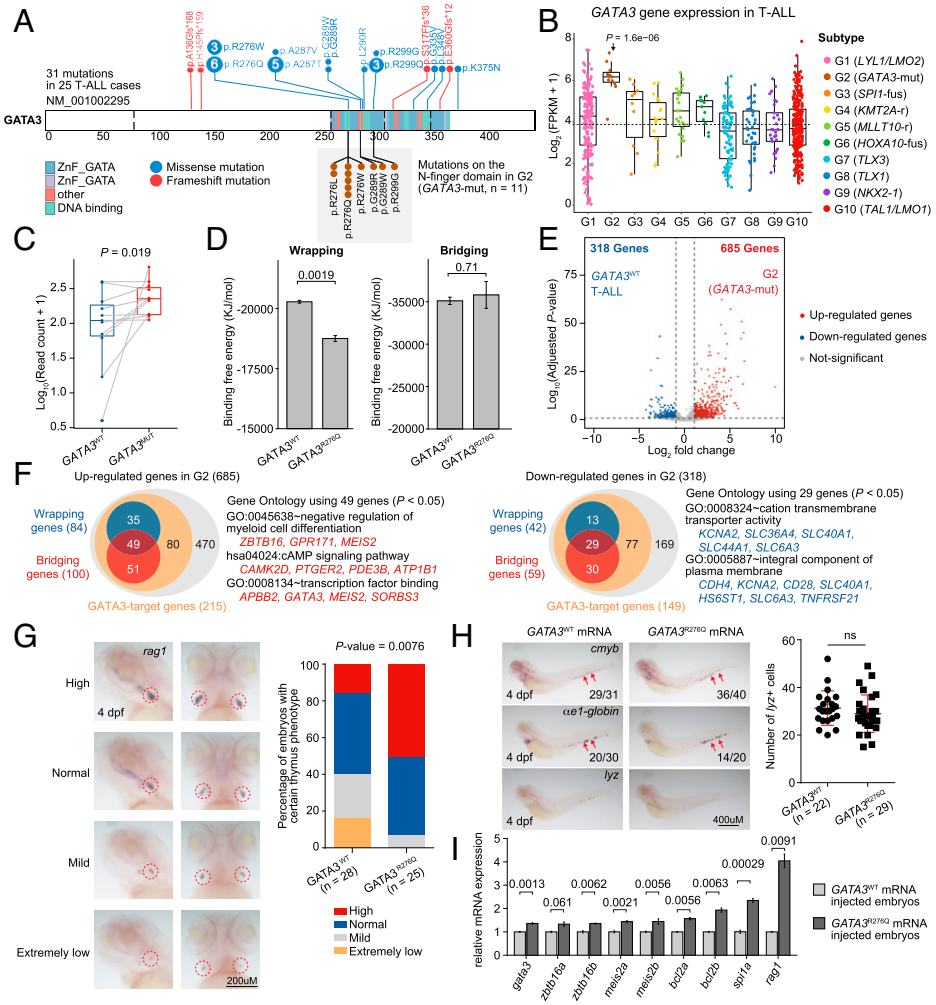

Fig. 3. Schematic representation of *GATA3* point mutations. (*A*) Protein structure of GATA3 and its mutations in T-ALL. All mutations on GATA3 are visualized on the upper area of the protein structure, and mutations in the N-finger domain identified in G2 are visualized in the lower area. (*B*) Boxplot of gene expression of *GATA3* in each subtype. The dashed line represents the mean value of *GATA3* in 707 T-ALLs. The *P* values are calculated by comparing with the mean gene expression of *GATA3* using Wilcoxon rank-sum test. (*C*) Boxplot of count of *GATA3*^{WT} and altered *GATA3*^{MUT} reads in G2. *P* value is calculated using paired Wilcoxon rank-sum test. (*D*) Binding free energy (KJ/mol) reveals binding affinities of GATA3 (wild type, R276Q) protein and wrapping and bridging DNA sequence. (*E*) Volcano plot shows the differentially expressed genes between G2 (*GATA3*-mut) and *GATA3*^{WT} T-ALL. Each dot represents one gene. Genes significantly up-regulated in G2 (*GATA3*-mut) are colored in red, and significantly down-regulated in G2 (*GATA3*-mut) in blue. (*F*, *Left*) Gene-expression level of predicted bridging genes in T-ALL patients with different *GATA3* genotypes; (*Right*) the gene ontology results using up-regulated bridging genes in *GATA3*^{R276Q}. (*G*) WISH results of *rag1* RNA probes between *GATA3*^{WT}- and *GATA3*^{R276Q}-mRNA injected embryos at 4 dpf. The phenotypes are defined as four groups: high, normal, mild, and extremely low according to the *rag1* RNA^{+} area in thymus. The percent is quantified (*Right*). *P* value is calculated using Fisher's exact test. (*H*) WISH results of *cmyb*, *αe-globin*, and *lyz* RNA probes between *GATA3*^{WT}- and *GATA3*^{R276Q} -injected embryos at 4 dpf. *P* value is calculated using Wilcoxon rank-sum test. (*I*) qRT-PCR analysis of mRNA expression of the GATA3 downstream genes and *rag1* in both *GATA3*^{WT}- and *GATA3*^{R276Q}-mRNA injected embryos at 4 dpf. The relative mRNA expressions are normalized to human *GATA3*. *P* values are calculated using Student's *t* test.

each containing the wild-type GATA3 (GATA3^{WT}) and mutant (GATA3^{R276Q}), and calculated the binding free energy to estimate the GATA3/DNA interaction (*SI Appendix*, *SI Materials and Methods*). Our model simulations showed that for the wrapping complex, the binding free energy was significantly higher in GATA3^{R276Q}/DNA (thus lower DNA-binding affinity) than in GATA3^{WT}/DNA, and no difference was observed for the bridging structure (Fig. 3*D*). To predict target genes that might be affected by wrapping/bridging motifs, we employed HOMER (39), combined with chromatin immunoprecipitation-sequencing data of GATA3^{WT} in human Jurkat cell lines (40) (Dataset S7), to define three sets of target genes: genes affected by wrapping motifs (wrapping targets), bridging motifs (bridging targets), and those insusceptible to wrapping/bridging motifs (other targets). Meanwhile, we identified 685 significantly up-regulated genes and 318 down-regulated genes in G2 (*GATA3*-mut), but not in *GATA3*^{MUT} T-ALL cases falling into other subtypes (Fig. 3*E* and *SI Appendix*, Fig. S4*G*). These data supported the unique role of GATA3 in defining G2 as a distinct subtype. Among the 685 up-regulated genes in G2, 215 (31%) were GATA3 target genes, including 84 wrapping targets, 100 bridging targets, and 49 wrapping/bridging targets (Fig. 3*F*). These latter 49 genes were of functional relevance to negative regulation of myeloid cell differentiation (*ZBTB16* and *MEIS2*), the cAMP signaling pathway (*ATP1B1*), and transcription factor binding (*GATA3*) (Fig. 3*F*). Notably, *ZBTB16*, also known as promyelocytic leukemia zinc finger *PLZF*, was reportedly to be involved in T cell lineage development (41) (*SI Appendix*, Fig. S4 *H* and *I*).

To examine the in vivo effect of *GATA3*^{R276Q} in hematopoiesis, we tried to use the zebrafish as a model. In this regard, we tested the leukemogenic role of the above-mentioned *RPL10*^{R98S}, an established leukemogenic mutation (32, 33), in a zebrafish experiment. Indeed, the results showed that *cmyb*, a

hematopoietic stem/progenitor cell (HSPC) marker, was aberrantly up-regulated in the caudal hematopoietic tissue of *RPL10*^{R98S} mRNA-injected embryos compared to *RPL10*^{WT}, as revealed by whole-mount in situ hybridization (WISH), reflecting an abnormal proliferation of HSPCs (*SI Appendix*, Fig. S5). Now that the system was feasible, we carried out the overexpression assay for both *GATA3*^{WT} and *GATA3*^{R276Q} in zebrafish. *GATA3*^{WT} and *GATA3*^{R276Q} mRNAs were injected into embryos for transient expression, followed by WISH examining the definitive hematopoiesis (Fig. 3 *G* and *H*). The expression of *rag1* (a lymphocyte marker in zebrafish) was significantly up-regulated in *GATA3*^{R276Q} mRNA-injected embryos in the thymus at 4 d postfecundation (dpf, $P = 0.0076$) (Fig. 3*G*), but no such altered expression was found with *cmyb*, *αe1-globin*, and *lyz* (Fig. 3*H*), the latter two being markers of erythroid cells and neutrophils, respectively. qRT-PCR confirmed the up-regulation of GATA3 downstream genes (including *gata3*, *zbtb16a/b*, *meis2a/b*, *bcl2a/b*, and *spi1a*) and *rag1* mRNA in the *GATA3*^{R276Q} mRNA-injected group (Fig. 3*I* and Dataset S8). Taken together, our results confirmed *GATA3*^{R276Q} as a driver for thymocyte proliferation in zebrafish embryos through either enhancing the effect of hematopoiesis-associated transcription factors (*GATA3*, *ZBTB16*, *MEIS2*) or activating target genes involved in T cell development pathways (including TGF-β, NOTCH, and Wnt/β-catenin signaling) (*SI Appendix*, Fig. S6*A*). These signaling cascades might collectively affect T cell proliferation and differentiation, eventually contributing to the pathogenesis of T-ALL.

Association of Molecular Subtypes with T Cell Development Stages. Given that genetic alterations distributed differently among subtypes, particularly ETP-ALL patients mainly found in G1 (*LYL1/LOM2*), we hypothesized that the subtypes might be inherently relevant to the maturation stages of T cells. In support of this, patients in G1 had low expression levels of T cell-related markers—such as *CD1A*, *CD2*, *CD3E*, *CD4*, and *CD8A* (immunophenotype-related genes)—but higher hematopoietic stem cell-related markers, such as *CD34* (*SI Appendix*, Fig. S6*B*). Notably, the myeloid marker *CD33* was highly expressed in G1 (*SI Appendix*, Fig. S6*B*). Furthermore, the expression patterns for the hematopoietic-related features in G2 and G3 were similar to those in G1, suggesting that T-lineage elements in G2 and G3 might be primitive (*SI Appendix*, Fig. S6*B*). Leukemic cells in G4 to G10 subtypes were likely at relatively late-stage T cell development, considering the decreased expression of HSPC-related markers and the increased expression of T cell-related markers (*SI Appendix*, Fig. S6*B*). To systematically associate subtypes with T cell development stages, we used the diffusion maps (42) for dimensionality reduction of 707 T-ALL, yielding three distinct branches (Fig. 4*A*): branch 1 for patients mainly from G1 to G6, branch 2 (G7 and G8), and branch 3 (G9 and G10). These three branches differed in ETP status and T cell maturation stages, showing that ETP-ALL with pro/precortical immunophenotypes were enriched in branch 1, whereas postcortical and medullary patients were enriched in branch 3 (*SI Appendix*, Fig. S6*C*).

Next, we used available public data of T cell expression functional clusters (43) (Dataset S9) to characterize subtypes and branches. The T cell differentiation stages in branch 1 patients were the earliest, branch 2 patients at an intermediate stage, and branch 3 patients at the late stage (the mature T-ALL stage) (Fig. 4*B*). Together with dysregulated leukemogenic factors and subtypes, T-ALL patients thus could be divided into three differentiation arrest branches: *LYL1/LMO2/SPI1/HOXA* high expression (branch 1, G1 to G6), *TLX3/TLX1* high expression (branch 2, G7 to G8), and *NKX2-1/TAL1/LMO1* high expression (branch 3, G9 to G10). Using both ETP signature (ETP status) and precortical signature (T cell development status), we generated the signature-specific enrichment score for each patient (Fig. 4*C* and Dataset S10). Patients with *LYL1/LMO2/SPI1/HOXA* tended to have higher scores for both ETP and precortical signatures, whereas patients with *NKX2-1/TAL1/LMO1* had lower scores for both signatures (Fig. 4*C*). We also compared the age and gender composition between the three branches and found that the *LYL1/LMO2/SPI1/HOXA* branch had the highest percentage of adult patients ($P = 1.3e-26$, χ^2 test) (Fig. 4*D*) and harbored more mutations (Fig. 4*E*). Functional categories were distributed differently among the three branches: mutations in NOTCH signaling (C1) were enriched in the *TLX3/TLX1* patients; mutations in epigenetic regulators (C2), transcription factors (C3), JAK-STAT signaling (C5), RAS signaling (C6), and proliferation/apoptosis (C8) were in the *LYL1/LMO2/SPI1/HOXA* and *TLX3/TLX1*; and mutations in PI3K-AKT-mTOR signaling (C4) were specifically concentrated in the *NKX2-1/TAL1/LMO1* branch (Fig. 4*F*).

In light of the prevailing perspectives (1, 5) and the knowledge obtained in this study, a working model of T-ALL leukemogenesis was proposed, with four key points (Fig. 4*G*): 1) the accumulation of genetic abnormalities, such as gene fusions and cancer driver mutations, could cause the dysregulation of different leukemogenic factors, ultimately blocking normal T cell development; 2) for T-ALL patients with the *LYL1/LMO2/SPI1/HOXA* (G1 to G6), the immunophenotype of blasts might represent the population blocked at a near HSPC or very early T cell development stage (i.e., ETP-ALL or near ETP, pro/precortical or cortical), and the age-onset for leukemogenesis tended to be higher (with over 50% adult patients); 3) for T-ALL patients with the *TLX3/TLX1* (G7 and G8), leukemic blasts could be blocked at an intermediate stage, with the cortical or postcortical immunophenotype of T-ALL cells; and 4) the immunophenotype of T-ALL patients with the *NKX2-1/TAL1/LMO1* (G9 and G10) might correspond to cortical, postcortical, or mature T cell counterparts at the late stage.

Exploring Genomic, Expression, and Cellular Correlates likely Explaining Differences in Adult and Pediatric T-ALL Patients. Correlating age with mutations in T-ALL, we found that *DNMT3A* and *IDH2* mutations tended to occur in the relatively elderly patients, with the mean age of 53 y for *IDH2* and 48 y for *DNMT3A* (Fig. 5*A* and *SI Appendix*, Fig. S7*A*). When the G1 group, which contained the majority of ETP-ALL in this series, was further scrutinized, the mutations rates of *DNMT3A* and *IDH2*, two genes with high lesion frequencies in acute myeloid leukemia (AML), were concentrated in about 44% of the G1 cases with the age of over 40 y (*SI Appendix*, Fig. S7*B*). The mutations tended to be frameshift and stop-gain for *DNMT3A*, suggesting loss-of-function (*SI Appendix*, Fig. S7*C*) (44), whereas hotspot missense mutations R140Q were observed in *IDH2*, in agreement with a gain-of-function alteration (*SI Appendix*, Fig. S7*D*) (45). It was reported that *DNMT3A* and *IDH2* mutations could cooperate to induce AML (44), but such cooperation was not yet identified in T-ALL. Mutated genes in epigenetic regulators (*IDH2*, *DNMT3A*, *CHD4*, *ASXL1*, *CREBBP*, and *EZH2*), transcription factors (*IKZF1*, *ETV6*, and *RUNX1*), JAK-STAT signaling (*JAK3* and *JAK1*), and RAS signaling (*NRAS*) were enriched in

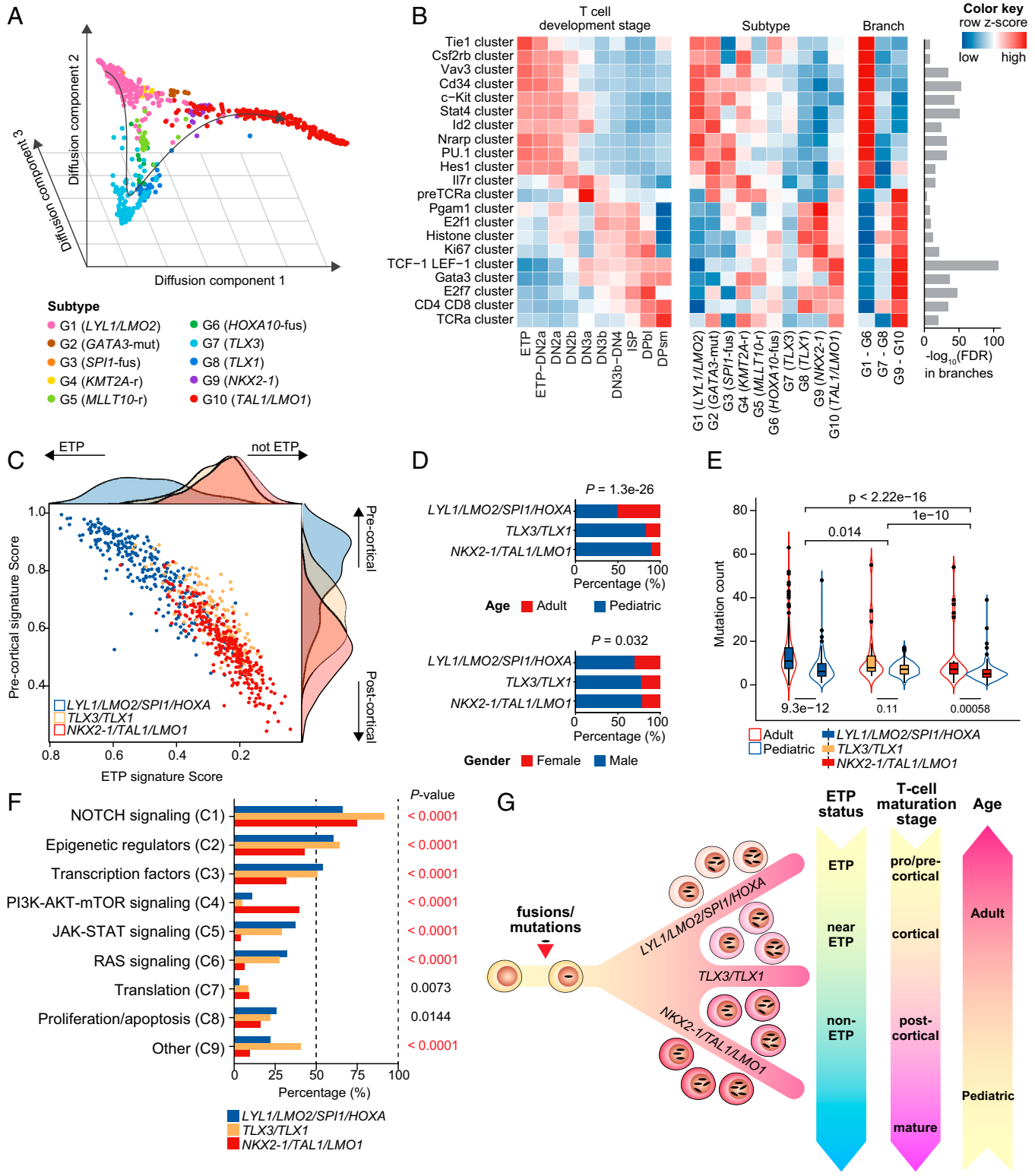

Fig. 4. Dimensionality reduction analysis revealing T cell development in different subtypes. (*A*) Visualization of the dimensions calculated by diffusion map using 707 T-ALL patients. The top 5% variance genes in RNA-seq data are subjected to diffusion map analysis and the first three diffusion components are visualized using three-dimensional plots. Each point represents one sample. (*B*) Gene-expression patterns of signatures of different functional clusters. These clusters were differentially expressed in different T cell stages. The *Left* heatmap shows the expression levels of functional clusters in different T cell stages, the *Middle* heatmap shows the expression levels in different subtypes, while the *Right* heatmap shows the expression levels in different branches. Expression is calculated using the mean value of the genes and then scaled as the row *z*-score. (*C*) Scatter and density plot of enrichment score (ES) for ETP and precortical signatures in different dysregulated leukemogenic factor branches. (*D*) Bar plot of the percentage of patients according to age (*Upper*) and gender (*Lower*) in each dysregulated leukemogenic factor branch. *P* values are calculated using a χ^2 test. (*E*) Violin plot of mutation counts identified in RNA-seq of each branch. The outline color represents age information and the internal boxplot represents the three branches. *P* values are calculated using Wilcoxon rank-sum test. (*F*) Comparisons of different functional categories of mutations in the three branches. *P* values are calculated using Fisher's exact test. (*G*) Model of the association between the accumulation of genetic abnormalities, the dysregulation of leukemogenic factors, T cell stages, and age in T-ALL leukemogenesis.

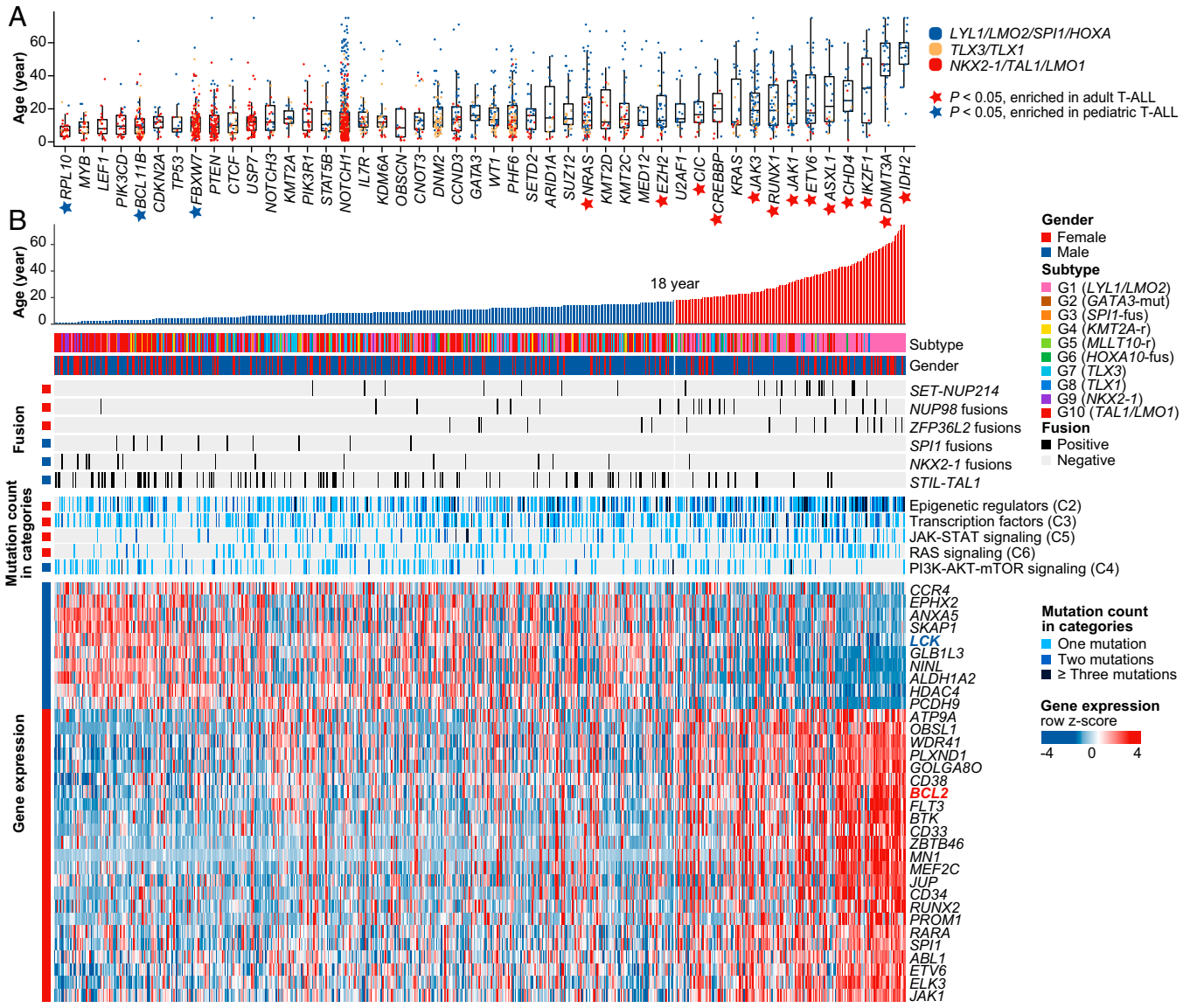

Fig. 5. Profiling of genetic abnormalities in pediatric and adult T-ALL. (*A*) Boxplot of age distribution of mutated genes (>2%) in T-ALL. Patient samples are colored based on the three dysregulated leukemogenic factor branches. Genes are ordered according to the mean age of patients. Mutations significantly enriched in adult T-ALL are marked with red stars, while those significantly enriched in pediatric T-ALL are marked with blue stars. *P* value is calculated using Fisher's exact test. (*B*) Profiling of gene fusions, mutation categories, and vulnerable pathways for potential therapeutic targets in pediatric and adult T-ALL. Columns indicate T-ALL patients, and rows represent four panels: clinical information panel (age, subtypes, gender), fusion panel (gene fusions, only $P < 0.05$ between pediatric and adult patients were illustrated), mutation counts in categories ($P < 0.05$), and gene-expression panel (potential therapeutic targets significantly correlated with age, false-discovery rate < 0.05). The gene-expression levels are normalized by *z*-score transformation.

adult T-ALL, while those in *FBXW7*, *BCL11B*, and *RPL10* were more likely to occur in pediatric T-ALL ($P < 0.05$) (Fig. 5*A*). Regarding gene fusions, we sorted the patients by their age at diagnosis (Fig. 5*B*). *SET-NUP214*, *NUP98* fusions and *ZFP36L2* fusions were significantly enriched in adult patients, whereas *SPI1* fusions, *NKX2-1* fusions, and *STIL-TAL1* were more likely to occur in pediatric patients ($P < 0.05$) (Fig. 5*B*). Since differences in genetic abnormalities/expression did exist between pediatric and adult patients, we searched for genes that were significantly correlated with age, especially those in pathways of vulnerability for potential therapeutic targets (Fig. 5*B*). Indeed, the expression levels of two therapeutic targets, *BCL2* (participating in apoptosis) and *LCK* (involved in preTCR activation) (46), were found oppositely correlated with age in T-ALL (*SI Appendix*, Fig. S7*E*). Other targets (*JAK1*, *ABL1*, and *FLT3*) were mainly up-regulated in adult T-ALL (Fig. 5*B*). Of note, *FLT3* and its related gene signatures (such as the AML pathway and *BCL2*) was highly expressed in G1, G2, G3, and G7, especially in ETP and procortical T-ALLs (*SI Appendix*, Fig. S8 *A–C*), indicating FLT3 inhibitors as a potential therapeutic target for future study.

It was recently established that ETP-ALL share features with mixed phenotype acute leukemia (MPAL) (47). We thus used AML gene signatures for enrichment analysis in an ETP-ALL subset and found a significantly augmented AML enrichment score with age (*SI Appendix*, Fig. S8*B*). Applying the xCell algorithm (48) and the 17-gene stemness score (49), we were able to infer the proportion of hematopoietic stem cell, common lymphoid progenitor, and common myeloid progenitor, and show a much higher likelihood of adult patients than pediatric ones to exhibit MPAL (*SI Appendix*, Fig. S8 *D* and *E*). Extending the leukemogenesis model of MPAL, we deduced that genetic lesions might occur at an earlier HSPC stage in adult T-ALL, resulting in abnormal proliferation/differentiation/

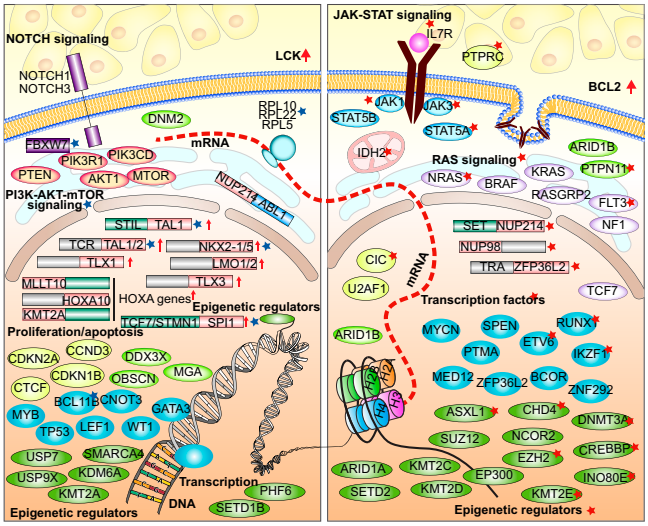

Fig. 6. Schematic presentation of gene-expression alterations and gene lesions identified in T-ALL. (*Left*) Genomic aberrations in pediatric T-ALL, and (*Right*) illustration of those in adult T-ALL. Gene fusions and mutations (>1%) and their subcellular localizations from cell surface membrane through cytosolic compartments to cell nucleus are represented. Mutations are illustrated in the ellipse and in different colors. Fusions and mutations that are significantly enriched in pediatric T-ALL are marked with blue stars, while those significantly enriched in adult T-ALL are marked with red stars. Genes whose overexpression is most likely due to fusions are marked with red arrows. LCK and BCL2 overexpression may represent drug targets and are labeled with arrowheads.

apoptosis of myeloid and lymphoid lineages, which might be significant in explaining the difference in prognosis and treatment responses observed between pediatric and adult T-ALL (*SI Appendix*, Fig. S8 *F* and *G*, respectively illustrating pediatric and adult leukemic hematopoiesis models).

Discussion

The transcriptomic landscape of 707 T-ALLs unveiled 10 subtypes in this study, which was superior to previous classifications of T-ALL that depended on the single/double combination of dysregulated leukemogenic factors, such as *LOM2*, *LYL1*, *HOXA* family genes, *TLX3/1*, *NKX2-1*, *LMO1*, and *TAL1/2*. Limited by the small sample size, we previously classified patients into three parts, namely *TLX1/3/HOXA*, *ETP/LYL1/HOXA*, *TAL1/LMO1* (9). The classification of ALL based on gene-expression patterns in larger ALL cohorts could help to identify rare subtypes (7). In the present study, *GATA3* mutations (G2) were first clarified as a subtype of T-ALL with a poor prognosis. Because previous studies have reported *GATA3* in solid tumors, including breast cancer (35, 36), how the role of *GATA3* dysregulation participated in T cell commitment would be further investigated in T-ALL and other T cell disorder diseases. Additionally, three subtypes with elevated expression of *HOXA* family genes were revealed—namely G4 (*KMT2A*-r), G5 (*MLLT10*-r), and G6 (*HOXA10*-fus)—each representing a small number of T-ALL patients, similar to G2, which might be the reason why they were not found as independent subtypes in a previous study (9).

Meanwhile, the expanded sample size of adult T-ALL allowed us to conduct a comparison of abnormal genome/transcriptome landscapes between pediatric and adult patients (Fig. 6). We illustrated that some fusions/mutations differed significantly in frequencies between the two age groups. Adult patients tended to harbor more nonsilent mutations than pediatric ones, especially in epigenetic regulators—with *DNMT3A* and *IDH2* being the most representative ones—JAK-STAT signaling, and RAS signaling pathways. These sequence mutations could cooperate with aberrant gene expression and exert an effect on clinical outcomes. Our results provide evidence that the more complex genetic abnormalities in leukemic cells in adults than in children may contribute to the unfavorable prognosis in the former age group. In addition, adult patients, particularly those aged over 40 y, are more likely to bear the features of MPAL, which render the malignant cells less sensitive to the current therapeutic agents. In this regard, it may be interesting to note the emergence of some potential therapeutic drug targets in adult T-ALL, such as *BCL-2* and *FLT3*.

There are some limitations to note in our study. First, the number of patients in some subtypes is still limited. Second, using RNA-seq alone to identify copy number variations or deletions, such as of *CDKN2A/CDKN2B*, is challenging. Fusions involving *TLX1*/*TLX3* and *LMO1/LMO2* are hardly identified from RNA-seq data, although their high expression levels are suggestive of the existence of fusion to *TCR*. Third, without samples from normal tissues as control and genomic DNA sequencing data in most cohorts, nonsilent mutations were inferred based on the previously reported mutation profiles in T-ALL and other leukemia (5, 7, 9, 16–21) and improvement of mutation-calling pipelines (described in *SI Appendix, SI Materials and Methods*). Despite these limitations, this work can facilitate an in-depth understanding of the biological nature of T-ALL. The dimensionality reduction analysis proves useful to determine the associations between molecular subtypes and phenotypes according to the stages of blockage in T cell development.

In conclusion, we identified 10 subtypes of T-ALL, characterized their genetic alteration patterns, and investigated the associations between these subtypes and T cell development stages. These results revealed that the involvement of T cell differentiation stage was earliest in G1 and latest in G10. Based on the dysregulated leukemogenic factors, we have revealed relative mutation/abnormal expression features in adult and pediatric T-ALL patients. Furthermore, our study lends support to the feasibility of RNA-seq as a clinical platform for the classification of T-ALL.

Materials and Methods

Patients. Patients in cohorts 1 to 6 were obtained from public database. Patients in cohort 7 were from the Hematological Biobank, Jiangsu Biobank of Clinical Resources during 2016 to 2019. Patients in cohort 8 were from a multicenter study under the coordination of the Shanghai Institute of Hematology, including Chinese People's Liberation Army General Hospital and the First Affiliated Hospital, Zhejiang University College of Medicine, and Second Hospital of Dalian Medical University, these cohorts being followed from 2016 to 2020. Informed consent for cohorts 7 and 8 patients in the study was obtained by the participating centers. This research was approved by Ruijin Hospital Ethics Committee. Detailed information of the patients is listed in *SI Appendix, SI Materials and Methods* and Dataset S1.

RNA-seq Analysis, Mutations and Fusions Calling, Zebrafish Experiment, and T Cell Differentiation Stage Analysis. Detailed materials and methods are provided in *SI Appendix, SI Materials and Methods*. All animal experiments were approved by the Committee of Animal Use for Research at Shanghai Jiao Tong University School of Medicine (China).

Data Availability. RNA sequencing data generated in this study are deposited at the National Omics Data Encyclopedia (NODE) (accession code OEP002748). Previously published data used for this work were Liu et al. (5), Seki et al. (17), Qian et al. (16), Yasuda et al. (18), Chen et al. (9), and Autry et al. (19).

ACKNOWLEDGMENTS. We thank the Center for High Performance Computing of Shanghai Jiao Tong University for providing computing support; and Prof. Jinghui Zhang and Prof. Jun J. Yang from the Department of Pharmaceutical

Sciences, St. Jude Children's Research Hospital, for their helpful advice on data analyses. This work was supported by the National Natural Science Foundation of China General Program (81770205, 32170663, 81670147, 81861148030, Antrag M-0377); the State Key Laboratory of Medical Genomics; the Double First-Class Project (WF510162602) from the Ministry of Education; the Shanghai Collaborative Innovation Program on Regenerative Medicine and Stem Cell Research (2019CXJQ01); the Overseas Expertise Introduction Project for Discipline Innovation (111 Project; B17029); the Shanghai Shenkang Hospital Development Center (SHDC2020CR5002); the Shanghai Major Project for Clinical Medicine (2017ZZ01002); and the Innovative Research Team of High-level Local Universities in Shanghai.

Author affiliations: [a]Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, 200025 Shanghai, China; [b]School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 200240 Shanghai, China; [c]State Key Laboratory of Microbial metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 200240 Shanghai, China; [d]Department of Medical Oncology, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, 200032 Shanghai, China; [e]Department of Hematology, Chinese People's Liberation Army General Hospital, 100853 Beijing, China; [f]University of Malaya Cancer Research Institute, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia; [g]Department of Paediatrics, KK Women's & Children's Hospital, 229899 Singapore; [h]Department of Hematology, The First Affiliated Hospital, Zhejiang University College of Medicine, 310058 Hangzhou, People's Republic of China; [i]National Clinical Research Center for Hematologic Diseases, Jiangsu Institute of Hematology, The First Affiliated Hospital of Soochow University, 215006 Suzhou, People's Republic of China; [j]Institute of Blood and Marrow Transplantation, Collaborative Innovation Center of Hematology, Soochow University, 215006 Suzhou, People's Republic of China; [k]Division of Hematology and Rheumatology, Kindai University Faculty of Medicine, 5778502 Osaka, Japan; [l]Department of Hematology, Atomic Bomb Disease Institute, Nagasaki University, 8528521 Nagasaki, Japan; [m]Clinical Research Center, Nagoya Medical Center, National Hospital Organization, 4600001 Nagoya, Japan; [n]Department of Hematology, The First Affiliated Hospital of China Medical University, 110001 Liaoning, China; [o]Department of Hematology, Second Hospital of Dalian Medical University, 116027 Leaoning, People's Republic of China; [p]Centre for Translational Research in Acute Leukaemia, Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, 117597 Singapore; [q]Viva-University Children's Cancer Centre, Khoo Teck Puat-National University Children's Medical Institute, National University Hospital, 119074 Singapore; [r]Cancer Science Institute of Singapore, National University of Singapore, 117599 Singapore; [s]Department of Hematology and Oncology, Nagoya University Graduate School of Medicine, 4668550 Nagoya, Japan; [t]Department of Integrated Health Sciences, Division of Cellular and Genetic Sciences, Nagoya University Graduate School of Medicine, 4610047 Nagoya, Japan; [u]Center for Biomedical Big Data, The First Affiliated Hospital, School of Medicine, Zhejiang University, 310003 Hangzhou, China; and [v]Cancer Center, Zhejiang University, 310058 Hangzhou, China

Author contributions: X.-J.S., J.-Q.M., Z.C., J.-Y.H., and S.-J.C. designed research; F.Z., H.F., G.L., L.J., B.C., D.-D.M., Y.-F.L., J.W., L.-J.P., C.F., H.-F.C., J.-X.M., T.A.M., S.-Y.W., X.-J.S., J.-Q.M., Z.C., J.-Y.H., and S.-J.C. performed research; J.-F.L., Q.-L.Z., H.W., H.A., J.-H.W., Y.-J.L., S.-N.C., Q.W., H.L., Z.S., I.M., Y.M., T.Y., L.-P.D., X.-J.Y., J.-S.Y., A.E.-J.Y., D.-P.W., H.K., F.H., and J.J. contributed new reagents/analytic tools; Y.-F.L., J.W., L.-J.P., H.W., H.A., T.A.M., J.-H.W., Y.-J.L., S.-N.C., Q.W., H.L., Z.S., I.M., Y.M., T.Y., L.-P.D., X.-J.Y., J.-S.Y., A.E.-J.Y., D.-P.W., H.K., F.H., J.J., S.-Y.W., and J.-Q.M. contributed data and provided clinical information; Y.-T.D., H.F., and J.-Y.H. analyzed data; and Y.-T.D., Z.C., J.-Y.H., and S.-J.C. wrote the paper.

1. T. Girardi, C. Vicente, J. Cools, K. De Keersmaecker, The genetics and molecular biology of T-ALL. *Blood* **129**, 1113–1123 (2017).
2. C. H. Pui, M. V. Relling, J. R. Downing, Acute lymphoblastic leukemia. *N. Engl. J. Med.* **350**, 1535–1548 (2004).
3. S. Chiaretti, R. Foà, T-cell acute lymphoblastic leukemia. *Haematologica* **94**, 160–162 (2009).
4. C. H. Pui, W. E. Evans, Treatment of acute lymphoblastic leukemia. *N. Engl. J. Med.* **354**, 166–178 (2006).
5. Y. Liu *et al.*, The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1211–1218 (2017).
6. I. Iacobucci, C. G. Mullighan, Genetic basis of acute lymphoblastic leukemia. *J. Clin. Oncol.* **35**, 975–983 (2017).
7. J. F. Li *et al.*, Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11711–E11720 (2018).
8. Y. F. Liu *et al.*, Genomic profiling of adult and pediatric B-cell acute lymphoblastic leukemia. *EBioMedicine* **8**, 173–183 (2016).
9. B. Chen *et al.*, Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 373–378 (2018).
10. L. Belver, A. Ferrando, The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat. Rev. Cancer* **16**, 494–507 (2016).
11. Q. Chen *et al.*, Coding sequences of the tal-1 gene are disrupted by chromosome translocation in human T cell leukemia. *J. Exp. Med.* **172**, 1403–1408 (1990).
12. P. Van Vlierberghe *et al.*, The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood* **111**, 4668–4680 (2008).
13. E. P. Noronha *et al.*; Brazilian Collaborative Study Group of Acute Leukemia, The profile of immunophenotype and genotype aberrations in subsets of pediatric T-cell acute lymphoblastic leukemia. *Front. Oncol.* **9**, 316 (2019).
14. E. Coustan-Smith *et al.*, Early T-cell precursor leukaemia: A subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol.* **10**, 147–156 (2009).
15. U. Koch, F. Radtke, Mechanisms of T cell development and transformation. *Annu. Rev. Cell Dev. Biol.* **27**, 539–562 (2011).
16. M. Qian *et al.*, Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. *Genome Res.* **27**, 185–195 (2017).
17. M. Seki *et al.*, Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1274–1281 (2017).
18. T. Yasuda *et al.*, Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. *Nat. Genet.* **48**, 569–574 (2016).
19. R. J. Autry *et al.*, Integrative genomic analyses reveal mechanisms of glucocorticoid resistance in acute lymphoblastic leukemia. *Nat. Can.* **1**, 329–344 (2020).
20. L. Jiang *et al.*, Multidimensional study of the heterogeneity of leukemia cells in t(8;21) acute myelogenous leukemia identifies the subtype with poor outcome. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20117–20126 (2020).
21. J. Xiong *et al.*, Genomic and transcriptomic characterization of natural killer T cell lymphoma. *Cancer Cell* **37**, 403–419.e6 (2020).
22. Y. Zhong, L. Jiang, H. Hiai, S. Toyokuni, Y. Yamada, Overexpression of a transcription factor LYL1 induces T- and B-cell lymphoma in mice. *Oncogene* **26**, 6937–6947 (2007).
23. D. J. Curtis, M. P. McCormack, The molecular basis of Lmo2-induced T-cell acute lymphoblastic leukemia. *Clin. Cancer Res.* **16**, 5618–5623 (2010).
24. A. Champhekar *et al.*, Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. *Genes Dev.* **29**, 832–848 (2015).
25. S. Nagel *et al.*, MEF2C is activated by multiple mechanisms in a subset of T-acute lymphoblastic leukemia cell lines. *Leukemia* **22**, 600–607 (2008).
26. I. Homminga *et al.*, Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* **19**, 484–497 (2011).
27. T. K. Tan, C. Zhang, T. Sanda, Oncogenic transcriptional program driven by TAL1 in T-cell acute lymphoblastic leukemia. *Int. J. Hematol.* **109**, 5–17 (2019).
28. S. Herblot, A. M. Steff, P. Hugo, P. D. Aplan, T. Hoang, SCL and LMO1 alter thymocyte differentiation: Inhibition of E2A-HEB function and pre-T alpha chain expression. *Nat. Immunol.* **1**, 138–144 (2000).
29. M. Sanchez-Martin, A. Ferrando, The NOTCH1-MYC highway toward T-cell acute lymphoblastic leukemia. *Blood* **129**, 1124–1133 (2017).
30. H. M. McRae *et al.*, PHF6 regulates hematopoietic stem and progenitor cells and its loss synergizes with expression of TLX3 to cause leukemia. *Blood* **133**, 1729–1741 (2019).
31. A. M. Martelli *et al.*, The key roles of PTEN in T-cell acute lymphoblastic leukemia development, progression, and therapeutic response. *Cancers (Basel)* **11**, 629 (2019).
32. T. Girardi *et al.*, The T-cell leukemia-associated ribosomal RPL10 R98S mutation enhances JAK-STAT signaling. *Leukemia* **32**, 809–819 (2018).
33. K. R. Kampen *et al.*, The ribosomal RPL10 R98S mutation drives IRES-dependent BCL-2 translation in T-ALL. *Leukemia* **33**, 319–332 (2019).
34. I. C. Ho, T. S. Tai, S. Y. Pai, GATA3 and the T-cell lineage: Essential functions before and after T-helper-2-cell differentiation. *Nat. Rev. Immunol.* **9**, 125–135 (2009).
35. K. Kataoka *et al.*, Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat. Genet.* **47**, 1304–1315 (2015).
36. C. Murga-Zamalloa, R. A. Wilcox, GATA-3 in T-cell lymphoproliferative disorders. *IUBMB Life* **72**, 170–177 (2020).
37. L. Fransecky *et al.*, Silencing of GATA3 defines a novel stem cell-like subgroup of ETP-ALL. *J. Hematol. Oncol.* **9**, 95 (2016).
38. Y. Chen *et al.*, DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep.* **2**, 1197–1206 (2012).
39. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
40. D. Hnisz *et al.*, Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
41. E. V. Rothenberg, J. Ungerbäck, A. Champhekar, Forging T-lymphocyte identity: Intersecting networks of transcriptional control. *Adv. Immunol.* **129**, 109–174 (2016).
42. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
43. M. Mingueneau *et al.*; Immunological Genome Consortium, The transcriptional landscape of $\alpha\beta$ T cell differentiation. *Nat. Immunol.* **14**, 619–632 (2013).
44. X. Zhang *et al.*, Dnmt3a loss and Idh2 neomorphic mutations mutually potentiate malignant hematopoiesis. *Blood* **135**, 845–856 (2020).
45. F. Wang *et al.*, Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622–626 (2013).
46. Y. Gocho *et al.*, Network-based systems pharmacology reveals heterogeneity in LCK and BCL2 signaling and therapeutic sensitivity of T-cell acute lymphoblastic leukemia. *Nat. Can.* **2**, 284–299 (2021).
47. T. B. Alexander *et al.*, The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* **562**, 373–379 (2018).
48. D. Aran, Z. Hu, A. J. Butte, xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
49. S. W. Ng *et al.*, A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).