Published in final edited form as:

Nat Genet. 2017 August ; 49(8): 1211–1218. doi:10.1038/ng.3909.

THE GENOMIC LANDSCAPE OF PEDIATRIC AND YOUNG ADULT T-LINEAGE ACUTE LYMPHOBLASTIC LEUKEMIA

Yu Liu[1], **John Easton**[1], **Ying Shao**[1,2], **Jamie Maciaszek**[3], **Zhaoming Wang**[1], **Mark R. Wilkinson**[2], **Kelly McCastlain**[2], **Michael Edmonson**[1], **Stanley B. Pounds**[4], **Lei Shi**[4], **Xin Zhou**[1], **Xiaotu Ma**[1], **Edgar Sioson**[1], **Yongjin Li**[1], **Michael Rusch**[1], **Pankaj Gupta**[1], **Deqing Pei**[4], **Cheng Cheng**[4], **Malcolm A. Smith**[5], **Jaime Guidry Auvil**[6], **Daniela S. Gerhard**[6], **Mary V. Relling**[7], **Naomi J. Winick**[8], **Andrew J. Carroll**[9], **Nyla A. Heerema**[10], **Elizabeth Raetz**[11], **Meenakshi Devidas**[12], **Cheryl L. Willman**[13], **Richard C. Harvey**[13], **William L. Carroll**[14], **Kimberly P. Dunsmore**[15], **Stuart S. Winter**[16], **Brent L Wood**[17], **Brian P. Sorrentino**[3], **James R. Downing**[2], **Mignon L. Loh**[18], **Stephen P Hunger**[19,*], **Jinghui Zhang**[1,*], and **Charles G. Mullighan**[2,*]

[1]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, United States

[2]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee, United States

[3]Department of Hematology, St. Jude Children's Research Hospital, Memphis, Tennessee, United States

Address for correspondence: Stephen P. Hunger, Children's Hospital of Philadelphia, CTRB #3060, 3501 Civic Center Boulevard, Philadelphia, PA 19104, hungers@email.chop.edu; Jinghui Zhang, St. Jude Children's Research Hospital, Department of Computational Biology, 262 Danny Thomas Place, Mail Stop 1135, Memphis, TN 38105, T: 1-901-595- 6829, jinghui.zhang@stjude.org; Charles G. Mullighan, St. Jude Children's Research Hospital, Department of Pathology, Mail Stop 342, 262 Danny Thomas Place, Memphis, TN 38105, T: 1-901-595-3387, F: 1-901-595-5947, Charles.mullighan@stjude.org.

Accession codes

Therapeutically Applicable Research to Generate Effective Treatments, dbGaP (http://www.ncbi.nlm.nih.gov/gap) phs000218 and phs000464.

AUTHOR CONTRIBUTIONS

Y.Liu., Z.W., M.E., X.M., Y.Li, R.C.H. analyzed genomic data.
M.R.W., M.R., P.G. managed genomic data and databases
X.Z., E.S. prepared data visualization in PeCan.
J.E., Y.S. performed genomic assays
J.M., K. McC., B.S. performed experiments
S.B.P., L.S., D.P., C.C., M.D. performed statistical analysis
M.A.S., J.G.A., D.S.G. oversaw the NCI TARGET project
M.V.E., N.J.W., E.R., W.L.C. K.P.D., S.S.W. provided patient data
B.L.W. performed immunophenotyping of leukemia samples
A.J.C., N.A.H. performed cytogenetic analysis
J.R.D. oversaw genomic analyses
M.L.L. and S.P.H. led Children's Oncology Group ALL studies and the ALL TARGET project
J.Z. supervised genomic analysis
C.G.M. analyzed genomic data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

[4]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee, United States

[5]Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, Maryland, United States

[6]Office of Cancer Genomics, National Cancer Institute, Bethesda, Maryland United States

[7]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, Tennessee, United States

[8]University of Texas Southwestern Medical Center, Dallas, Texas, United States

[9]Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, United States

[10]Department of Pathology, College of Medicine, The Ohio State University, Columbus, Ohio, United States

[11]Department of Pediatrics, Huntsman Cancer Institute and Primary Children's Hospital, University of Utah, Salt Lake City, Utah, United States

[12]Department of Biostatistics, Colleges of Medicine, Public Health & Health Profession, University of Florida, Gainesville, Florida, United States

[13]Department of Pathology, The Cancer Research and Treatment Center, University of New Mexico, Albuquerque, New Mexico, United States

[14]Department of Pediatrics, Perlmutter Cancer Center, New York University Medical Center, New York, New York, United States

[15]Health Sciences Center, University of Virginia, Charlottesville, Virginia, United States

[16]Department of Pediatrics, University of New Mexico, Albuquerque, New Mexico, United States

[17]Seattle Cancer Care Alliance, Seattle, Washington, United States

[18]Department of Pediatrics, Benioff Children's Hospital, University of California at San Francisco, San Francisco, California, United States

[19]Department of Pediatrics and the Center for Childhood Cancer Research, Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, United States

Abstract

Genetic alterations activating NOTCH1 signaling and T cell transcription factors, coupled with inactivation of the INK4/ARF tumor suppressors are hallmarks of T-ALL, but detailed genome-wide sequencing of large T-ALL cohorts has not been performed. Using integrated genomic analysis of 264 T-ALL cases, we identify 106 putative driver genes, half of which were not previously described in childhood T-ALL (e.g. *CCND3, CTCF, MYB, SMARCA4, ZFP36L2* and *MYCN*). We described new mechanisms of coding and non-coding alteration, and identify 10 recurrently altered pathways, with associations between mutated genes and pathways, and stage or subtype of T-ALL. For example, NRAS/FLT3 mutations were associated with immature T-ALL,

JAK3/STAT5B mutations in *HOX1* deregulated ALL, PTPN2 mutations in *TLX1* T-ALL, and PIK3R1/PTEN mutations in *TAL1* ALL, suggesting that different signaling pathways have distinct roles according to maturational stage. This genomic landscape provides a logical framework for the development of faithful genetic models and new therapeutic approaches.

T-lineage ALL comprises up to 25% of ALL cases,[1,2] and is sub classified according to stages of thymic maturation, including early cortical, late cortical and mature T-cell stages. Early T-cell precursor T-ALL (ETP ALL) is an additional high-risk subtype that lacks expression of several T cell surface markers, and exhibits aberrant expression of myeloid and stem cell markers.[3,4] Cytogenetic and candidate gene mutational analyses have identified recurring genetic alterations in T-ALL, several of which show association with tumor cell maturational stage or clinical features. Activating mutations of NOTCH1 are present in at least 60% of ALL cases.[5,6] Activation of oncogenic transcription factors is also a hallmark of T-ALL, commonly from rearrangement to T-cell receptor loci. These include the basic helix-loop-helix transcription factor genes *TAL1*,[7] *TAL2*,[8] *LYL1*[9] and *OLIG2* (*BHLHB1*);[10] *TLX1* (*HOX11*),[11] *TLX3* (*HOX11L2*),[12] *NKX2-1*, *NKX2-2* and *NKX2-5*;[13,14] the LIM-only domain genes *LMO1* and *LMO2*;[15,16] the *HOXA* homeobox genes; *MYC*[17] and *MYB*.[18,19] Additional rearrangements result in the expression of chimeric fusion genes involving *KMT2A (MLL), HOXA* and tyrosine kinase genes such as *ABL1*.[20,21] Deletion of the *CDKN2A/CDKN2B* tumor suppressor loci is observed in over 70% of T-ALL,[22] with deletion of genes encoding the cell cycle regulators RB1[18] and CDKN1B ($p27^{Kip1}$)[23] in approximated 15% of cases.

Additional targets of mutation include loss-of-function alterations of hematopoietic transcription factor genes including *BCL11B*, *ETV6*, *GATA3*, *LEF1, RUNX1* and *WT1*.[24–28] Several genes encoding epigenetic regulators and chromatin modifiers are also mutated in T-ALL, including *EED, EZH2* and *SUZ12* that encode core components of the polycomb repressor complex 2 (PRC2) that mediates the repressive histone 3 lysine 27 trimethylation mark, the histone demethylase *KDM6A*[29,30] and the deubiquitinating enzyme *USP7*.[31] Several intracellular signaling pathways are activated by mutation, including the PI3K-AKT-mTOR pathway, most commonly from *PTEN* deletion[18,32], *PTPN2* deletion[33] or *AKT1* mutation;[34] the JAK-STAT pathway via activating mutations in the interleukin 7 receptor alpha chain (*IL7RA*) gene,[35,36] *JAK1*,[37] *JAK3*[38,39] or *STAT5B*;[40] the Ras-MAPK signaling through *KRAS* and *NF1* mutation,[24,41] and chimeric ABL1 fusion proteins such as *NUP214-ABL1* and *ETV6-ABL1*.[20,21] Mutations in ribosomal biogenesis processing have recently been identified, including *RPL5, RPL10* and *RPL11*[42] and the plant homeodomain factor gene *PHF6* that has a putative role in chromatin modification.[43]

Although these data have provided important insights into the oncogenic pathways that drive leukemogenesis in T-ALL there are relatively few data from unbiased, genome-wide sequencing approaches, and those available have usually examined relatively small cohorts,[24,44] gene panels[31,45] or a single modality of sequencing,[42,46]. Here we performed an integrated, multimodality genomic analysis of a uniformly treated cohort of childhood and young adult T-ALL to identify the spectrum and constellations of genetic alterations in this disease.

Molecular classification of T-ALL

In a collaborative study of the Children's Oncology Group (COG), the National Cancer Institute Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative and the St Jude–Washington University Pediatric Cancer Genomic Project (PCGP), we studied 264 children and young adults with newly diagnosed T-ALL consecutively enrolled on the COG trial AALL0434[47] between 2007–2011 with available tumor and remission material (Supplementary Tables 1 – 2). DNA extracted from tumor and remission samples was subjected to whole-exome sequencing (WES) and single nucleotide polymorphism (SNP) microarray genotyping, and tumor RNA to transcriptome sequencing (RNA-Seq) (Supplementary Tables 3 – 5). A subset of novel or complex sequence mutations and mutations of the *TAL1* enhancer[48,49] were verified by amplicon-based next generation sequencing (Supplementary Table 6). Immunophenotypic data were available for all cases and 189 (71.6%) cases included data necessary to enable classification into ETP-ALL (19 cases), near-ETP (24 cases fulfilling criteria for ETP ALL, except with normal expression CD5 expression)[4] and non-ETP ALL (146 cases).

Based on genetic alterations and dysregulated expression of transcription factor genes, and hierarchical clustering of RNA-seq gene expression data, 242 (91.7%) patients were classified into 8 subgroups: deregulation of *TAL1* (N=87), *TAL2* (N=8), *TLX1* (N=26), *TLX3* (N=46), *HOXA* (N=33), *LMO1/2* (N=10), *LMO2-LYL1* (N=18) and *NKX2-1* (N=14) (Supplementary Table 1 and Supplementary Figure 1). ETP ALL cases commonly exhibited *LMO2/LYL1* deregulation (7 of 19 cases, 36.8%), near- and non-ETP ALL cases were enriched for *TAL1* deregulation (41.7% and 27.4% respectively), *TLX3* deregulation was most commonly observed in near-ETP cases, and *TLX1* deregulation in non-ETP ALL (χ^2 P=0.0004, Supplementary Table 7).

Sequence mutations in pediatric T-ALL

We identified 4165 non-synonymous single nucleotide (SNV), insertion/deletion (indel) or internal tandem duplication (ITD) sequence mutations in 2694 genes by exome or RNA sequencing in the cohort of 264 samples, with a mean of 15.8 mutations (range 2–50) per case (s). We used the mutation significance detection tools MutSigCV[50], MuSiC[51] and Medal Ceremony, and mutant allele expression data to identify potential driver mutations. Importantly, 24 genes harboring somatic mutations of known functional significance were mutated in only a single case, including *EP300* (p300), *KIT* and *TP53*. Overall, 106 driver genes were identified (Supplementary Table 9), including known targets of mutation in T-ALL: *NOTCH1* (N=197 cases; 74.6%, including 5 cases with intragenic *NOTCH1* deletions), *FBXW7* (63; 23.9%), *PHF6* (50; 18.9%), *PTEN* (37; 14.0%), *USP7* (31; 11.7%), *DNM2* (29; 11.0%); *BCL11B* (26; 9.8%), *WT1* (24; 9.1%), *JAK3* (20; 7.6%), *NRAS* (20; 7.6%), *IL7R* (18; 6.8%), *PIK3R1* (17; 6.4%) and *RPL10* (16; 6.1%) (Figure 1 and Supplementary Figure 2). Fifty-three (51%) of these 106 genes, and 39 (48%) of 82 recurrently mutated genes had not previously been described to be mutated in pediatric T-ALL.[24,31,34,42,44] These included mutations in the transcriptional regulatory genes *MYB* (N=13), *CNOT3* (N=9), *MED12* (N=7), *MYCN* (N=6) and *TSPYL2* (N=4); epigenetic regulators, including *CTCF* (N=13), *SMARCA4* (N=7) and *CREBBP* (N=5); cell cycle

genes, including *CCND3* (N=15) and *CDKN1B* (N=4); the deubiquitination gene *USP9X* (N=7); RNA processing genes, including *DDX3X* (N=4) and *U2AF1* (N=4); mutations predicted to activate PI3K-AKT signaling, including *PIK3CD* (N=5); and the RNA-binding protein and suppressor of *NOTCH1* expression *ZFP36L2*[52,53] (N=5, 3 mutations and three fusions). Four genes (*CNOT3, KMT2D, RPL5* and *FAT1*) were previously identified as targets of mutation in adult, but not pediatric T-ALL.

Sub-clonal mutations (defined as those with a mutant allele fraction (MAF) of less than 0.3) were frequently observed in many driver genes (Figure 1), indicating sub-clonal evolution is a hallmark of T-ALL. Moreover, many cases harbored multiple subclonal mutations in the same gene, indicating that these mutations are secondary events in leukemogenesis but exert driver functions in individual subclones. For example, *NOTCH1* was the most frequently mutated gene with 264 sequence mutations identified in 196 cases, with majority of mutations in the heterodimerization (HD) domain (62.9%, 166/264) or PEST domain (31.4%, 83/264). Although subclonal *NOTCH1* mutations have been previously reported, the extent of multi- and subclonality revealed in this study is substantially greater that previously described.[54] Of the 264 mutations, 116 (43.9%) were sub-clonal. Fifty-eight (29.6%) samples harbored multiple *NOTCH1* mutations (49 samples with 2 mutations and 9 samples with 3 mutations).

To determine if germline sequence mutations may influence the risk of developing T-ALL, we analyzed 89 genes (Supplementary Table 10), including 60 associated with autosomal dominant cancer predisposition syndromes,[55] and the 35 most frequent targets of somatic mutation in this study. Only 4 potentially deleterious germline mutations, in *BRAF, BRCA1, BRCA2* and *RUNX1* were identified (Supplementary results).

Gene rearrangements in pediatric T-ALL

Analysis of transcriptome sequencing data identified transcripts derived from 255 gene rearrangements in 191 (72.3%) samples (Supplementary Table 11). Eighty-three chimeric in-frame fusions were identified, 54 of which arose from inter-chromosomal rearrangements, 29 from intra-chromosomal events. The most frequent were *MLLT10* fusions (6 unique rearrangements in 13 cases); fusions of *KMT2A* (6 fusions in 12 cases); *ABL1* fusions (4 fusions in 5 cases); and *NUP98* fusions (4 fusions in 5 cases) (Table 1). Multiple novel fusions involving signaling pathways, transcriptional regulation or RNA processing were identified, including *ETV6-CTNNB1*, *IL9R-VAMP7*, *PCM1-JAK2*, *CD99-JAK2*, *CSTF3-LMO2*, *STMN1-SPI1* and *NCOR1-NFIA* (1 case each). The second class of rearrangements identified by RNA-seq were those deregulating oncogenes, including *TAL1* (N=58), *TLX1* (N=17), *TLX3* (N=13), and *TAL2* (N=6). Importantly, this analysis identified involving novel rearrangement partners and/or mechanisms of deregulation for known targets of rearrangement, but also showed that transcriptome sequencing does not identify all rearrangements, notably cases with *TLX3* deregulation (Supplementary Results; Supplementary Table 12 and Supplementary Figure 3). Whole genome sequencing (WGS) of 25 cases lacking a rearrangement identified 5 additional cases with rearrangements of core T-ALL transcription factor genes that were not identified by analysis of transcriptome sequencing data as breakpoints were located in intergenic regions, and the rearrangements

did not result in expression of a chimeric fusion product: *TCRA-LMO2, TCRB-TAL2, TCRA-LMO2, BCL11B-TLX3 and NKX2-1* (Supplementary Table 1).

The transcription factor gene *MYB* was altered by amplification (N=33), rearrangement (N=11) or mutation (N=13) in 49 (18.6%) cases (Supplementary Figure 4). We identified multiple mechanisms of rearrangement, including juxtaposition of the 5′ region of *MYB* to promoter/enhancer regions of partner genes (*TCRB*, N=2) and *SLC12A9* (N=1). Rearrangements of the 3′ region of *MYB* resulting in loss of a negative regulatory region were identified in 3 cases (involving *PLAGL1, BDP1 and CHMP1A*, N=1 each). Fourteen sequence mutations in 13 samples were identified, including a novel hot spot of missense and in-frame insertion mutations at codon 14 in 7 cases (Supplementary Figure 2). This residue lies in an unstructured region of MYB adjacent to a nuclear localization sequence that may perturb the intracellular localization, and thus activity of MYB.

Recurrently targeted pathways in pediatric T-ALL

We integrated sequence mutation, structural variant/rearrangement and DNA copy number alteration data (Supplementary Tables 13–14) and identified 10 functional pathways recurrently mutated in T-ALL: transcriptional regulation (91% of cases), cell cycle regulation and tumor suppression (84%), NOTCH1 signaling (79%), epigenetic regulation (68%), PI3K-AKT-mTOR signaling (29%), JAK-STAT signaling (25%), Ras signaling (14%), ribosomal function (13%), ubiquitination (9%) and RNA processing (9%) (Figure 2 and Supplementary Tables 15 – 16). Most commonly mutated were transcriptional regulators including those oncogenes defining T-ALL subgroups (e.g. *TAL1*, *TLX1*, *TLX3*, *LMO2*, and *NKX2-1*), known recurrently mutated genes (*MYB*, *LEF1*, *BCL11B*) as well as 16 genes mutated in less than 10 cases (Supplementary Figure 5). Genes encoding regulators of cell cycle progression and/or tumor suppressors were mutated in 83.7% of cases, the recurrent targets of alteration being *CDKN2A/CDKN2B* (78.4%), *CDKN1B* (12.9%) *RB1* (9.5%), *CCND3* (6.1%), with single mutations in *CASP8*, *CDKN2C* and *TP53.* In addition, broad deletions of chromosome 6q14-q23 were present in 19.3% of cases, and were enriched in *TAL1* cases (Supplementary Results, Supplementary Table 17 and Supplementary Figures 6 – 8).

Mutations activating NOTCH1 signaling were observed in 79% of cases, including *NOTCH1* mutations (74.6%; 149 HD, 78 PEST), *FBXW7* mutations (25.4%) and mutations in both genes in 20.8% cases (N=55) (Supplementary Figure 9).

We observed significant association between mutations and T-ALL subtypes, and variation in the frequency of genes and pathways mutated across T-ALL subtypes, consistent with the notion that such subtype-enriched genomic alterations reflect genes critical to specific stages of T-cell development (Figure 3, Supplementary Tables 18 – 19 and Supplementary Figure 10). Notably, the NOTCH1 signaling pathway was nearly always mutated in *LMO1/2*, *NKX2-1*, and *TLX1* cases, but was less frequently mutated in *TAL1*/2 and *LMO2-LYL1* cases. Epigenetic mutations were particularly common in *TLX3* cases (93.5%), and ribosomal processing mutations were highly enriched in *NKX2-1* cases (50.0%).

Excluding *CDKN2A*, *NOTCH1* and *FBXW7*, which are mutated in multiple T-ALL subtypes, we identified the 37 most recurrently mutated genes that were significantly enriched in at least one T-ALL subtype. We constructed a network linking mutations to T cell developmental stages considering both gene-subgroup and gene-gene correlation (Figure 3). Mutation of the JAK-STAT or Ras signaling pathways (e.g. *FLT3*, *NRAS* and *JAK3*) were enriched in the *LMO2/LYL1* and *HOXA* subgroups while those of PI3K signaling pathway (commonly *PTEN* and *PIK3R1*) were enriched in the *TAL1* subgroup, suggesting a transition of the predominant signaling pathway from early T cell development to later stages. *USP7* alterations were significantly enriched in the *TAL1* subgroup, *CTCF* in *TLX3* cases, and *DNM2* and *PHF6* in *TLX1/3* cases, suggesting that different epigenetic regulators are involved in the early versus late cortical stages of T cell development. Alteration of the transcription factor *LEF1* was most common in *NKX2-1* cases and *BCL11B, MYC* in *TLX1* cases. Mutations in *RPL5* and *RPL10* were enriched in *TLX3* and *NKX2-1* subgroups, respectively, suggesting that these two ribosome biogenesis genes may have distinct roles in early and late cortical development. *MYB* (n=49), *IL7R* (n=18) and *ABL1* (n=18, including 3 cases with NUP214-ABL1, one with NUP214-ABL1 and ETV6-ABL1, one with MBNL1-ABL1, one with an in frame ABL1 deletion and the remainder with broad amplification of 9q, see Supplementary Results and Supplementary Figure 11) were the most frequently mutated genes that were not associated with a subtype, suggesting that these may drive signaling pathways in multiple stages of T-cell maturation.

Somatic alterations targeting transcriptional regulators

In addition to alterations of transcription factor genes that define T-ALL subgroups, 26 transcription factor genes were mutated, 24 in more than one case (Supplementary Figure 5).[56] A novel mutation in T-ALL was the MYCN p.Pro44Leu (P44L) mutation, identified in 6 cases (Figure 4a). This mutation has previously been identified in neuroblastoma and is associated with elevated levels of *MYCN* gene expression, suggesting a role in oncogenesis.[57] The mutated residue lies adjacent to the conserved phospho-degron recognized by E3 ubiquitin ligases FBXW7, SKP2 and HUWE1, consistent with the possibility that this mutation may perturb MYCN protein levels or protein-protein interactions.[58–60] To examine the functional consequences of this mutation, we transduced *Arf*$^{-/-}$ mouse thymocytes with lentiviral vectors expressing either wild-type or P44L MYCN. Transduced cells were cultured on OP9-DL1 cells and transplanted after ten days into irradiated *Rag*$^{-/-}$ recipient mice.[61] Both wild-type and mutant MYCN vectors induced a highly penetrant T-lineage leukemia, but with a modest but significantly shorter latency in cells expressing the mutant P44L MYCN vector (Figure 4b–d). To explore the hypothesis that this mutation may perturb ubiquitin-mediated degradation, we transfected NIH3T3 cells with wild-type or mutant HA-tagged MYCN lentiviral vectors, treated cells with cycloheximide to block protein translation, and quantified MYCN protein levels over time by immunoblotting. We observed significantly slower decay of MYCN P44L levels compared to wild-type protein (Figure 4e–f), consistent with the notion that this mutation impairs protein degradation, thus enhancing oncogenicity.

TAL1-deregulated ALL was the most common subtype of T-ALL in this study (87/33.0% cases). Fifty-nine cases had structural alterations identified by RNA-seq or SNP array

analysis (51 with deletions resulting *STIL-TAL1* fusion, 8 with other *TAL1* rearrangements) but a substantial minority lacked such an alteration. Recently, a non-coding insertion mutation upstream of the *TAL1* locus was reported that creates a MYB binding site, which results in recruitment of CREBBP (CBP), acetylation of histone 3 lysine 27 and recruitment of RUNX1, GATA3 and TAL1.[48] We identified 15 cases (5.7%) with identical or similar *TAL1* mutations that result in acquisition of the MYB binding motif. Notably, we identified a distinct indel mutation in four additional cases that results acquisition of TCF1/2 motifs, and loss of a GMEB1 binding motif, suggesting a distinct mechanism for TAL1 deregulation in these cases (Supplementary Figure 12). The remaining 11 TAL1 cases lack a known genomic mechanism for gene deregulation.

Additional transcription factor mutations included *MED12* (N=7) and *TSPYL2* (N=4, Supplementary Figure 2). *MED12* encodes part of the kinase module of the large MEDIATOR complex that bridges a range of transcriptional pathways to RNA polymerase 2 and transcription, and is mutated in multiple tumors.[62] The majority of mutations were frame-shift or nonsense mutations, suggesting these are tumor suppressor genes in T-ALL. Thirty-two cases had focal deletions or gains of chromosome 8q distal to *MYC* at or adjacent to the NOTCH1 MYC Enhancer (NME) region that interacts with the *MYC* locus and facilitates MYC expression (Supplementary Figure 13).[63,64] These alterations were most common in the *HOXA* and *TLX1/3* subgroups, and included deletions in addition to the previously reported amplifications, both of which were associated with elevated *MYC* expression (data not shown).

Signaling pathway mutations in T-ALL

Mutations activating signaling pathways were observed in 65% of cases. These included PI3K-AKT (28.8% of cases), JAK-STAT (24.6%), Ras signaling (13.6%) and rearrangements involving *ABL1*, *GPR128, IL2RB* and *IL9R* (Figure 5). Exclusivity was observed between PI3K-AKT pathway and JAK-STAT or Ras pathway alterations (Fisher's exact test, P<0.0001), with PI3K-AKT mutations prevalent in *TAL1* cases, and JAK-STAT or Ras mutations in *TLX1/3* and *HOXA* cases. Although *PTEN*, *PIK3R1* and *AKT1* were the most commonly mutated genes in the PI3K-AKT pathway, multiple novel targets of mutation in T-ALL were identified. These included a recurrent somatic PIK3CD p.Glu1021Lys mutation observed in five cases (four *TAL1* and one *LMO2/LYL1* case). This has been reported to be a dominant AKT activating germline mutation.[65] Other recurring alterations resulting in activation of signaling pathways included rearrangement of ABL1 and *TFG-GPR128*, a rearrangement also observed in high-risk B-ALL and lymphoma (N=3).[66]

Fifty-five (20.8%) cases had multiple signaling mutations, with 13 cases carrying at least 3 mutations. The most common patterns were a JAK-STAT activating mutation and concomitant mutations involving additional JAK-STAT (N=19), Ras (N=11), PI3K-AKT (N=6) or other mutations (N=6), or PI3K-AKT mutations with second PI3K-AKT mutations (N=13), Ras (N=4) or other mutations (N=3). This raised that question as to whether these concomitant mutations reside in the same or distinct clones. This is of importance for the selection of targeted therapeutic approaches, and the potential for emergence of subclones

resistant to agents targeting a single pathway. To address this, we examined the MAF of signaling pathway mutations in 14 cases with JAK3 sequence mutations. Thirteen cases carried JAK3 variant with MAF >0.3 (30%) indicating the mutations were present in a predominant/major clone (Supplementary Figure 14). In 10 of these cases, the signaling mutation of the next highest MAF (including JAK-STAT and Ras pathways) was also clonal. In three cases the JAK3 mutation was clonal, but the second signaling mutation was subclonal, indicating that both mutations coexist in a subclone. In one case, both mutations were subclonal, and may exist in the same or different clones. Thus, in the majority of cases, the most common JAK-STAT and second most common signaling mutations coexist in the same clone.

Patterns of epigenetic mutation in T-ALL

Epigenomic regulators were recurrently mutated in all T-ALL sub-groups with the exception of *TAL2*-rearranged ALL (0 of 8 cases), with a prevalence ranging from 10 of 18 (55.6%) *LMO2/LYL1* cases to 43 of 46 (93.5%) *TLX3* cases (Supplementary Table 19 and Supplementary Figure 15). In addition to the known association between *PHF6* mutations and *TLX3* rearrangement,[43] several additional associations between mutation and subtype were observed. Alterations of *USP7* were almost exclusively observed in TAL1 cases (29 of 33 USP7 mutations). *USP7* encodes a ubiquitin-specific protease that stabilizes MDM2 and TP53.[67] The majority of mutations (76.5%) were heterozygous out-of-frame indel (N=25) or nonsense (N=1) mutations in or proximal to the catalytic domain indicating the mutations result in loss-of-function (Supplementary Figure 2). Additional associations between mutations and T-ALL subtypes included *CTCF* (56.0% of 25 mutations) and *KDM6A* (42.9% of 21 mutations) in *TLX3* cases, and *KMT2A* (77.8% of 18 mutations) in HOXA cases (Supplementary Table 19).

Mutational spectra of ETP and non-ETP T-ALL

ETP ALL has previously been described to be a distinct form of leukemia with characteristic immunophenotypic, transcriptomic, genetic, and clinical features compared to other T-ALL cases.[3] Prior studies have reported a mutational spectrum enriched for mutations in transcriptional, epigenetic and signaling genes more characteristic of myeloid malignancies.[3,68] As expected, many (7 of 19) ETP ALL cases fell in the *LMO2/LYL1* overexpressing group, but with additional cases in the HOXA (N=3), TAL1 (N=2) and TLX3 (N=4) subgroups (Supplementary Table 6). Alterations in transcriptional regulators were nearly universal, occurring in 89.5% of ETP ALL, 87.5% of near-ETP and 89.0% of non-ETP cases). Alterations in JAK-STAT signaling (47.4% c.f. 25.0%/21.9%, P=0.05), Ras signaling (36.8% c.f. 20.8%/10.3%, P=0.005) and epigenetic regulation (84.2% c.f. 83.3%/65.8%, P=0.08) were more common, and cell cycle alterations (31.6% c.f. 87.5%/87.0%, P<0.001) less common (Supplementary Table 19). Genes more frequently altered in ETP ALL included *FLT3* (5 of 9 in ETP cases, P<0.001), *WT1* (8 of 21 in ETP, P<0.001), *IKZF1* (3 of 8 ETP, P=0.029) and *MED12* (4 of 6 in ETP, P<0.001) and rearrangement of *NUP98* (3 of 4 cases in ETP ALL, P<0.001). FLT3 was also overexpressed in ETP versus near-/non-ETP cases, with five of 8 cases with highest FLT3 expression having concomitant mutations (data not shown). In contrast, genes less commonly mutated in ETP ALL were *CDKN2A/B*,

CNOT3, *DDX3X*, *FBXW7*, *LEF1*, *MYCN* and *RPL10*, alterations of which were exclusively observed in non-ETP cases.

Genomic determinants of outcome

The incidence of relapse was only 7.5% (95% CI ±1.7%) in this cohort, and the only univariable associations with relapse identified were with *AKT1*, *MLLT10, CNOT3* and *PTEN* alterations (see supplementary results and Supplementary Tables 20 – 23).

Discussion

Although many of the most frequent mutations in T-ALL have previously been described, by performing integrated genomic analysis in a large cohort of T-ALL cases, we have identified a large number of unrecognized targets of mutation, and demonstrated stage- and subtype-specific associations between genes, cellular pathway and outcome. While many mutated genes reside in a “long tail” of mutations, in silico functional analysis shown that many of these mutations are likely deleterious and contribute to the pathogenesis of ALL. Moreover, many such genes fall in pathways known to be mutated in T-ALL (including JAK-STAT, Ras and PTEN-PI3K signaling pathways), thereby increasing the proportion of T-ALL cases with mutations in these pathways that are potentially amenable to therapeutic approaches with signal transduction inhibitors. We also show that many of the known targets of mutation are altered at greater frequency than previously recognized (e.g. NOTCH1), and that subclonal and multiclonal mutations are more prevalent than previously recognized. This is consistent with the notion that many of these mutations are acquired as secondary events subsequent to the founding alterations deregulating hematopoietic transcription factors. This is also a finding of potential therapeutic relevance, as we show that multiple signaling mutations commonly co-occur in the same clone. This supports the rationale for targeting these pathway(s) in such cases.

In addition to identifying new targets of mutation, the integrated analysis identified multiple new modalities of mutation in known T-ALL driver genes, including sequence mutations and enhancer alterations deregulating *MYCN*, novel rearrangements and mutations of *MYB*, rearrangements and complex mutations of *ZFP36L2*, and novel *TAL1* enhancer mutations. As proof of principal, we show that the *MYCN* mutations accelerate leukemogenesis, in part by stabilizing protein expression. This integrated analysis also suggests that genetic alterations, such as the common broad deletion of 6q, is likely to promote tumorigenesis by disrupting multiple genes, although this will require formal testing in models that engineer the deletion or disrupt the activity of multiple genes in the region simultaneously.

A limitation of this study is that whole genome sequencing was not performed on all cases. Analysis of RNA-seq, exome and DNA copy number data failed to identify lesions activating key transcription factors in a proportion of cases. Using whole genome sequencing in a subset of cases, we identified additional rearrangements in a subset of cases. Thus, comprehensive definition of the genomic landscape of T-ALL will require this approach.

The multi-modality genomic analysis performed in all cases enabled a detailed analysis of associations between mutated gene and subtypes of ALL that are considered to arise at defined stages of thymic development. We show that while several pathways are mutated in the majority of cases, mutations in individual genes (e.g. JAK-STAT, Ras or *PTEN* mutations; *ETV6* or *RUNX1* transcription factor mutations; and *PHF6* and *USP7* epigenetic mutations) are significantly enriched in different ALL subtypes. This suggests that not only are these alterations leukemogenic cooperative driver events with the founding alterations deregulating transcription factors, but that these genes may have central roles in distinct stages of thymic ontogeny, and are "primed" for mutational deregulation following acquisition of a sentinel rearrangement in specific progenitors. Consequently, this detailed portrait of the T-ALL genomic landscape offers a rational foundation for the design of genetic models of T-ALL that faithfully mirror the human disease.

ONLINE METHODS

Patients and samples

264 consecutively recruited children and young adults with available banked diagnostic and remission material on the Children's Oncology Group Study of T-lineage ALL (NCT00408005) were studied. The study was approved by the St Jude Children's Research Hospital Institutional Review Board. Informed consent or assent was obtained from all subjects and/or their legal guardians.

Exome sequencing

Exome sequencing was performed by Beckman Coulter Genomics (Danvers, MA) or at St Jude Children's Research Hospital. In general, samples with at least 1μg each of tumor and remission DNA were sequenced at Beckman Coulter (216 samples), and those with limiting material at St Jude. Details are as follows.

Beckman Coulter Genomics—DNA shearing was performed using Covaris instrument library construction a done on a Beckman Coulter Biomek FXP using Beckman Coulter Genomics SPRIworks HT Kit (https://www.beckmancoulter.com/wsrportal/wsr/research-and-discovery/products-and-services/next-gen-library-preparation/spriworks-ht/index.htm). First-round PCR (4–8 cycles) was performed using primers appropriate for Illumina (GA and HiSeq) sequencers, and clean-up steps with BC/Agencourt AMPure XP beads. Target capture utilized SureSelectXT Human All Exon V5 (Agilent Technologies) and supplied hybridization and associated reagents. Second-round PCR (10–16 cycles) used TruSeq index adapters. Library quality control was performed using a TapeStation (Agilent) and qPCR using Kapa standard curves. Sequencing was performed using Illumina HiSeq 2000 and 2500 sequencers in high-output (TruSeq SBS v3), rapid-high-output (HiSeq SBS v4) and rapid run (HiSeq Rapid SBS v1) modes. All runs were 100 nucleotide paired end reads, analyzed with on-board software RTA v1.18 and HCS v2.2.

St. Jude Children's Research Hospital—Library construction utilized DNA tagmentation (fragmentation and adapter attachment) performed using the reagent provided in the Illumina Nextera rapid exome kit, and was performed using the Caliper Biosciences

(Perking Elmer) Sciclone G3. First-round PCR (10 cycles) was performed using Illumina Nextera kit reagents, and clean-up steps employ BC/Agencourt AMPure XP beads. Target capture utilized Illumina Nextera rapid capture exome kit and supplied hybridization and associated reagents. The pre-hybridization pool size was 12 samples, and second round PCR (10 cycles) performed with Nextera kit reagents. Library quality control was performed using a Victor fluorescence plate reader with Quant-it dsDNA reagents for pre-pool quantitation, and Agilent Bio-analyzer 2200 for final library quantitation. Sequencing was performed using Illumina HiSeq 2500 instruments in high output mode using TruSeq SBS v3 chemistry. All runs were 100-nt paired end reads, and data analyzed using on-board software RTA v1.18 and HCS v2.2.

A detailed comparison of coverage using the two sequencing approaches is available upon request.

Whole Exome Sequencing Analysis

Paired-end reads were aligned to the reference human genome assembly GRCh37-lite using BWA[69] (version 0.5.10), and analyzed as previously described[24], including coverage and quality assessment, single nucleotide variation (SNV) and indel detection, annotation and prediction of deleterious effects for sequence mutations. High C>A/G>T mutation rate was observed in a subset of 43 samples that had shorter library insert size (ranging from 88bp to 141bp, with a median of 103bp) than the expected median library insert size of 190bp. Variants detected in these samples shown enrichment for subclonal C>A or G>T mutations, which matched the profile of previously described 8-oxoG mutations caused by sequencing artifacts introduced by DNA shearing during library construction[57]. For these samples, we applied a filter to remove the low MAF (≤ 0.2) C>A or G>T mutations except for those that were detected in samples with expected library insert size.

A subset (N=271) of novel or complex sequence mutations was selected for verification by amplicon-based next generation sequencing. We included 21 of potential artifactual 8-oxoG variants (Supplementary Table 6), none of which were verified.

Driver mutations were identified based on the analysis of MutSigCV[50], MuSiC[51] and Medal Ceremony, a novel algorithm that we developed to assess mutation pathogenicity based on match to known hotspots in oncogenes or loss-of-function in known tumor suppressor genes (M.E. et al., manuscript in preparation). Expression of the mutant allele is required to consider a missense mutation a candidate driver if it was not identified by MutSigCV or MuSIC. The sequence mutations were visualized in PeCan Data Portal[70] (https://pecan.stjude.org/home).

RNA-sequencing

TRIzol was used to extract RNA from acute lymphoblastic leukemia samples of bone marrow or peripheral blood. 500ng DNaseI treated total RNA had ribosomal RNA removed using Ribo-Zero rRNA removal beads and was converted into a cDNA library using Illumina TruSeq stranded Total RNA Kit with Ribo-Zero Gold. After adapter ligation each cDNA library was purified and enriched by PCR amplification. Each library underwent 100-cycle paired-end sequencing on the Illumina HiSeq 2000. Four samples were multiplexed into

each flow cell lane for sequencing. Base calls and quality scores were produced by CASAVA 1.8.

RNA Sequencing Analysis

RNA-seq data were mapped using StrongArm (M.R. et al., manuscript in preparation) and rearrangements were identified using CICERO as previously described[71]. Gene level read count was generated using HTseq-count[72] and FPKM value was calculated based on the transcript models in GENCODE v19. Quantile normalization was applied to the $\log_2$ transformed FPKM matrix. Only gene expressed (defined as FPKM > 1) in more than 30% of samples were further analyzed[73]. Median absolute deviation (MAD) was calculated for each gene across the cohort and top 500 genes with largest MAD were selected for cluster analysis using Ward's minimum variance method.

Vector constructs and lentiviral production

A hemagglutinin (HA)-tagged wild-type MYCN cDNA in the mouse stem cell virus-internal ribosome entry site-yellow fluorescent protein (MSCV-IRES-YFP) retroviral vector was provided by Dr. Gerard Grosveld[74]. The MSCV-HA-MYCN region was excised from the retroviral vector with XbaI and DraIII and cloned into a self-inactivating CL20 lentiviral vector backbone to yield a CL20-HA-NMYC-IRES-GFP vector. The introduction of a point mutation at codon 44 changing the native amino acid from proline to leucine was performed with the QuikChange II XL site-directed mutagenesis kit (Stratagene). Ecotropic lentivirus was produced by transient transfection of 293T cells with helper plasmids and then concentrated by centrifugation.

Thymocyte transduction and transplantation

Thymi were explanted from 4–8 week old $Arf^{-/-}$ mice[75] and single cell suspensions prepared. Thymocytes were then stained with CD4-phycoerythrin (PE), CD8-PE, and Ter119-PE antibodies (BD), incubated at 4°C for 10 min, washed with MACS buffer (PBS, MACS BSA Stock Solution (Miltenyi Biotec), and 2 mM EDTA (Sigma), and incubated with anti-PE microbeads (Miltenyi Biotec) for 15 min. Cells washed with MACS buffer were centrifuged, filtered, and placed on a pre-rinsed LS column (Miltenyi Biotec). Unattached CD4-/CD8-/Ter119- cells were collected and the purity of the "double-negative" thymocyte population was analyzed by flow cytometry.

For viral vector transduction, the CD4-/CD8- thymocytes were plated at a density of 0.5×10^6 cells/well on OP9-DL1 stromal cells (gift from J.C. Zuniga-Pflucker, University of Toronto, Toronto, Canada) in α-MEM (Invitrogen) containing 20% FBS, penicillin/streptomycin, glutamine, sodium pyruvate, and supplemented with 5 ng/ml each of FLT-3 (R&D Systems) and IL-7 (PeproTech). After 24 hours, polybrene was added and the cells were transduced with 1ml of vector supernatant collected from 293T cells, and spun at 2000 RPM for 1h at room temperature. This transduction was repeated again at 48h. Every 2–3 days, thymocytes were replated onto fresh OP9-DL1 cells and on day 8 were immunophenotyped by flow cytometric analysis.

Vector-transduced thymocytes were harvested from OP9-DL1 cell cultures on day 10 and transplanted by tail vein injection into sublethally irradiated (6Gy) female $\gamma c^{-/-}Rag2^{-/-}$ mice. In the initial primary transplant, 1×10^6 bulk cells containing 92% and 88% vector positive cells for wild-type or mutant MYCN, respectively were transplanted. In the second primary transplant, 0.5×10^6 bulk cells containing 70% and 53% vector positive cells for wild-type or mutant MYCN, respectively were transplanted. No randomization or blinding was performed. All experimental procedures were reviewed and approved by the Institutional Animal Care and Use Committee of St. Jude Children's Research Hospital.

Histology and immunophenotypic analysis

Tissues were fixed in 10% buffered formalin and immunohistochemistry was performed using standard procedures. Slides of 4–6μm sections were cut from formalin-fixed paraffin embedded tissues, and sections were stained with hematoxylin and eosin or by the avidin-biotin-peroxidase method using anti-Myeloperoxidase, CD3, B220, CD117, GATA-1, RUNX1, and TdT antibodies. Peripheral blood was collected in 10 mmol/L EDTA for standard CBCs and smears stained with Wright-Giemsa. Expression of cell surface markers and GFP was measured by flow cytometry (BD Biosciences Fortessa) using fluorophore-conjugated antibodies to CD4, CD8, CD25, CD44, CD3, B220, Gr1, Mac1, Ter119, and (BD Pharminogen, Franklin Lakes, NJ).

Immunoblot analysis

Protein lysates were extracted from unsorted splenic tumor samples using fresh RIPA lysis and extraction buffer (Thermo Fisher Scientific, Waltham, MA) prepared with cOmplete protease inhibitor and PhosSTOP phosphotase inhibitor (Roche Diagnostics, Indianapolis, IN). Protein concentration was quantified by the BCA assay (Thermo Fisher Scientific). Tumor samples were separated by electrophoresis in MOPS buffer on 4–12% Bis-Tris NuPAGE gels (Invitrogen), transferred to polyvinylidene fluoride membranes (Invitrogen), blocked with 5% milk, and probed using a rabbit polyclonal antibody against HA (Y-11, Santa Cruz Biotechnology, Dallas, TX) at a 1/200 dilution. The secondary antibody utilized was a donkey anti-rabbit (GE Healthcare) at a dilution of 1/500 and an antibody to beta actin (C-4, Santa Cruz Biotechnology, Dallas, TX) was used at a dilution of 1/1000 to control for protein loading.

Cycloheximide stability assays

NIH 3T3 Cells were transduced with HA-tagged constructs expressing either MYC-WT or the MYCN P44L mutant, treated with cycloheximide (20ug/mL), and collected at baseline, 30, 60, 90, 120 and 180 minutes for protein analysis. Cells were lysed in RIPA buffer (Sigma) and 30μg of protein was loaded on 4–12% NuPage Bis-Tris gels (Life Technologies) and electrophoresed at 200V for 1 hour. Blots were simultaneously probed with anti-HA-Tag (1:1,000; SIGMA #SAB4300603) and actin (1:1,000; S. Cruz #sc-1615) antibodies, subsequently detected using fluorescently conjugated secondary antibodies, and quantified using Odyssey Imaging System Studio Software (v4.0, LI-COR Biosciences). The HA signal was log-transformed and curves fitted using the least squares method, and curves compared by F test in Prism (GraphPad).

Statistical analysis

Associations between categorical values were examined using Fisher's exact test. These included associations of T-ALL subtype and gene/pathway, and subtype/gene/pathway with level of minimal residual disease (MRD). Data are reported as *P* values without correction for multiple comparisons. Associations between T-ALL subtype and treatment outcome (event-free survival and overall survival) were performed using the Kaplan-Meier estimator with Peto's estimator of standard deviation and the log-rank test[76–78]. An event was defined as a failure to achieve remission, a relapse after remission, or the development of a second malignancy. Analyses were performed using Prism v6.0 (GraphPad), R (www.r-project.org)[79], and SAS (v9.1.2, SAS Institute, Cary, NC). Integrated analysis and statistical modeling of genomic data is described in detail in Supplementary Results.

Data availability

Data may be accessed through the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) website at https://ocg.cancer.gov/programs/target. The NCI TARGET initiative specifies and supports the long-term deposition and maintenance of the data files, methods and quality control steps involved in the comprehensive genomic analysis of TARGET samples. The sequencing BAM and FASTQ files from whole exome and RNA-seq are accessible through the database of genotypes and phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap) under the accession numbers phs000218 (TARGET) and substudy specific accession phs000464 (TARGET ALL Expansion Phase 2). Gene expression, chromosome segmental copy number, SNV/indel, structural variants, and the clinical information are available through the TARGET Data Matrix (https://ocg.cancer.gov/programs/target/data-matrix). These are annotated within MIAME-compliant MAGE-TAB metadata files fully describing the methods, the specimen processing details and the quality control parameters. The genomic landscape reported in this study could be explored at the St. Jude PeCan Data Portal available at http://pecan.stjude.org/proteinpaint/study/target-tall.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Genome Sequencing Facility, Hartwell Center for Bioinformatics and Biotechnology, and flow cytometry and cell sorting core facility of St Jude Children's Research Hospital.

This work was supported in part by the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital; by a St. Baldrick's Foundation Scholar Award (to C.G.M.); by National Cancer Institute Grants P30 CA021765 (St. Jude Cancer Center Support Grant), U01 CA157937 (to C.L.W. and S.P.H.), and grants to the Children's Oncology Group: U10 CA98543 (Chair's grant and supplement to support the COG ALL TARGET project), U10 CA98413 (Statistical Center), and U24 CA114766 (Specimen Banking). This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E (to C.G.M.). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

1. Hunger SP, Mullighan CG. Acute Lymphoblastic Leukemia in Children. N Engl J Med. 2015; 373:1541–52. [PubMed: 26465987]
2. Aifantis I, Raetz E, Buonamici S. Molecular pathogenesis of T-cell leukaemia and lymphoma. Nat Rev Immunol. 2008; 8:380–90. [PubMed: 18421304]
3. Coustan-Smith E, et al. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. Lancet Oncology. 2009; 10:147–156. [PubMed: 19147408]
4. Inukai T, et al. Clinical significance of early T-cell precursor acute lymphoblastic leukaemia: results of the Tokyo Children's Cancer Study Group Study L99-15. Br J Haematol. 2012; 156:358–65. [PubMed: 22128890]
5. Weng AP, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004; 306:269–71. [PubMed: 15472075]
6. Pear WS, et al. Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. J Exp Med. 1996; 183:2283–91. [PubMed: 8642337]
7. Begley CG, et al. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc Natl Acad Sci U S A. 1989; 86:2031–5. [PubMed: 2467296]
8. Xia Y, et al. TAL2, a helix-loop-helix gene activated by the (7;9)(q34;q32) translocation in human T-cell leukemia. Proc Natl Acad Sci U S A. 1991; 88:11416–20. [PubMed: 1763056]
9. Mellentin JD, Smith SD, Cleary ML. lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell. 1989; 58:77–83. [PubMed: 2752424]
10. Wang J, et al. The t(14;21)(q11.2;q22) chromosomal translocation associated with T-cell acute lymphoblastic leukemia activates the BHLHB1 gene. Proc Natl Acad Sci U S A. 2000; 97:3497–502. [PubMed: 10737801]
11. Hatano M, Roberts CW, Minden M, Crist WM, Korsmeyer SJ. Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. Science. 1991; 253:79–82. [PubMed: 1676542]
12. Bernard OA, et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. Leukemia. 2001; 15:1495–504. [PubMed: 11587205]
13. Homminga I, et al. Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. Cancer Cell. 2011; 19:484–97. [PubMed: 21481790]
14. Nagel S, Kaufmann M, Drexler HG, MacLeod RA. The cardiac homeobox gene NKX2-5 is deregulated by juxtaposition with BCL11B in pediatric T-ALL cell lines via a novel t(5;14)(q35.1;q32.2). Cancer research. 2003; 63:5329–34. [PubMed: 14500364]
15. Royer-Pokora B, Loos U, Ludwig WD. TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t(11;14)(p13;q11). Oncogene. 1991; 6:1887–93. [PubMed: 1923511]
16. McGuire EA, et al. The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. Mol Cell Biol. 1989; 9:2124–32. [PubMed: 2501659]
17. Erikson J, et al. Deregulation of c-myc by translocation of the alpha-locus of the T-cell receptor in T-cell leukemias. Science. 1986; 232:884–6. [PubMed: 3486470]
18. Mullighan CG, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007; 446:758–64. [PubMed: 17344859]
19. Clappier E, et al. The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. Blood. 2007; 110:1251–61. [PubMed: 17452517]
20. Graux C, et al. Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. Nat Genet. 2004; 36:1084–9. [PubMed: 15361874]

21. Van Limbergen H, et al. Molecular cytogenetic and clinical findings in ETV6/ABL1-positive leukemia. Genes Chromosomes Cancer. 2001; 30:274–82. [PubMed: 11170285]
22. Hebert J, Cayuela JM, Berkeley J, Sigaux F. Candidate tumor-suppressor genes MTS1 (p16INK4A) and MTS2 (p15INK4B) display frequent homozygous deletions in primary cells from T- but not from B-cell lineage acute lymphoblastic leukemias. Blood. 1994; 84:4038–44. [PubMed: 7994022]
23. Remke M, et al. High-resolution genomic profiling of childhood T-ALL reveals frequent copy-number alterations affecting the TGF-beta and PI3K-AKT pathways and deletions at 6q15-16.1 as a genomic marker for unfavorable early treatment response. Blood. 2009; 114:1053–62. [PubMed: 19406988]
24. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012; 481:157–63. [PubMed: 22237106]
25. Gutierrez A, et al. The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. Blood. 2011; 118:4169–73. [PubMed: 21878675]
26. Gutierrez A, et al. Inactivation of LEF1 in T-cell acute lymphoblastic leukemia. Blood. 2010; 115:2845–51. [PubMed: 20124220]
27. Tosello V, et al. WT1 mutations in T-ALL. Blood. 2009; 114:1038–45. [PubMed: 19494353]
28. Van Vlierberghe P, et al. ETV6 mutations in early immature human T cell leukemias. J Exp Med. 2011; 208:2571–9. [PubMed: 22162831]
29. Ntziachristos P, et al. Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. Nature. 2014; 514:513–7. [PubMed: 25132549]
30. Van der Meulen J, et al. The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. Blood. 2015; 125:13–21. [PubMed: 25320243]
31. Huether R, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. Nat Commun. 2014; 5:3630. [PubMed: 24710217]
32. Palomero T, et al. Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. Nat Med. 2007; 13:1203–10. [PubMed: 17873882]
33. Kleppe M, et al. Deletion of the protein tyrosine phosphatase gene PTPN2 in T-cell acute lymphoblastic leukemia. Nat Genet. 2010; 42:530–5. [PubMed: 20473312]
34. Gutierrez A, et al. High frequency of PTEN, PI3K, and AKT abnormalities in T-cell acute lymphoblastic leukemia. Blood. 2009; 114:647–50. [PubMed: 19458356]
35. Shochat C, et al. Gain-of-function mutations in interleukin-7 receptor-alpha (IL7R) in childhood acute lymphoblastic leukemias. J Exp Med. 2011; 208:901–8. [PubMed: 21536738]
36. Zenatti PP, et al. Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. Nat Genet. 2011; 43:932–9. [PubMed: 21892159]
37. Flex E, et al. Somatically acquired JAK1 mutations in adult acute lymphoblastic leukemia. J Exp Med. 2008; 205:751–8. [PubMed: 18362173]
38. Jeong EG, et al. Somatic mutations of JAK1 and JAK3 in acute leukemias and solid cancers. Clin Cancer Res. 2008; 14:3716–21. [PubMed: 18559588]
39. Bains T, et al. Newly described activating JAK3 mutations in T-cell acute lymphoblastic leukemia. Leukemia. 2012; 26:2144–6. [PubMed: 22425895]
40. Kontro M, et al. Novel activating STAT5B mutations as putative drivers of T-cell acute lymphoblastic leukemia. Leukemia. 2014; 28:1738–42. [PubMed: 24573384]
41. Balgobind BV, et al. Leukemia-associated NF1 inactivation in patients with pediatric T-ALL and AML lacking evidence for neurofibromatosis. Blood. 2008; 111:4322–8. [PubMed: 18172006]
42. De Keersmaecker K, et al. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. Nat Genet. 2013; 45:186–90. [PubMed: 23263491]
43. Van Vlierberghe P, et al. PHF6 mutations in T-cell acute lymphoblastic leukemia. Nat Genet. 2010; 42:338–42. [PubMed: 20228800]
44. Bandapalli OR, et al. The activating STAT5B N642H mutation is a common abnormality in pediatric T-cell acute lymphoblastic leukemia and confers a higher risk of relapse. Haematologica. 2014; 99:e188–92. [PubMed: 24972766]

45. Vicente C, et al. Targeted sequencing identifies associations between IL7R-JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia. Haematologica. 2015; 100:1301–10. [PubMed: 26206799]
46. Atak ZK, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. PLoS Genet. 2013; 9:e1003997. [PubMed: 24367274]
47. Winter SS, et al. Safe integration of nelarabine into intensive chemotherapy in newly diagnosed T-cell acute lymphoblastic leukemia: Children's Oncology Group Study AALL0434. Pediatr Blood Cancer. 2015; 62:1176–83. [PubMed: 25755211]
48. Mansour MR, et al. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science. 2014; 346:1373–7. [PubMed: 25394790]
49. Navarro JM, et al. Site- and allele-specific polycomb dysregulation in T-cell leukaemia. Nat Commun. 2015; 6:6094. [PubMed: 25615415]
50. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–8. [PubMed: 23770567]
51. Dees ND, et al. MuSiC: identifying mutational significance in cancer genomes. Genome Res. 2012; 22:1589–98. [PubMed: 22759861]
52. Galloway A, et al. RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence. Science. 2016; 352:453–9. [PubMed: 27102483]
53. Hodson DJ, et al. Deletion of the RNA-binding proteins ZFP36L1 and ZFP36L2 leads to perturbed thymic development and T lymphoblastic leukemia. Nat Immunol. 2010; 11:717–24. [PubMed: 20622884]
54. Mansour MR, et al. Notch-1 mutations are secondary events in some patients with T-cell acute lymphoblastic leukemia. Clin Cancer Res. 2007; 13:6964–9. [PubMed: 18056171]
55. Zhang J, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med. 2015; 373:2336–46. [PubMed: 26580448]
56. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009; 10:252–63. [PubMed: 19274049]
57. Pugh TJ, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet. 2013; 45:279–84. [PubMed: 23334666]
58. Popov N, Schulein C, Jaenicke LA, Eilers M. Ubiquitylation of the amino terminus of Myc by SCF(beta-TrCP) antagonizes SCF(Fbw7)-mediated turnover. Nat Cell Biol. 2010; 12:973–81. [PubMed: 20852628]
59. Zhao X, et al. The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. Dev Cell. 2009; 17:210–21. [PubMed: 19686682]
60. Zhao X, et al. The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. Nat Cell Biol. 2008; 10:643–53. [PubMed: 18488021]
61. Treanor LM, et al. Interleukin-7 receptor mutants initiate early T cell precursor leukemia in murine thymocyte progenitors with multipotent potential. J Exp Med. 2014; 211:701–13. [PubMed: 24687960]
62. Taatjes DJ. The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends Biochem Sci. 2010; 35:315–22. [PubMed: 20299225]
63. Yashiro-Ohtani Y, et al. Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. Proc Natl Acad Sci U S A. 2014; 111:E4946–53. [PubMed: 25369933]
64. Herranz D, et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat Med. 2014; 20:1130–7. [PubMed: 25194570]
65. Angulo I, et al. Phosphoinositide 3-kinase delta gene mutation predisposes to respiratory infection and airway damage. Science. 2013; 342:866–71. [PubMed: 24136356]
66. Roberts KG, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. N Engl J Med. 2014; 371:1005–15. [PubMed: 25207766]

67. Ye M, et al. STIP is a critical nuclear scaffolding protein linking USP7 to p53-Mdm2 pathway regulation. Oncotarget. 2015; 6:34718–31. [PubMed: 26460617]
68. Neumann M, Greif PA, Baldus CD. Mutational landscape of adult ETP-ALL. Oncotarget. 2013; 4:952–3. [PubMed: 23807748]

METHODS ONLY REFERENCES

69. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–60. [PubMed: 19451168]
70. Zhou X, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat Genet. 2016; 48:4–6. [PubMed: 26711108]
71. Parker M, et al. C11orf95-RELA fusions drive oncogenic NF-kappaB signalling in ependymoma. Nature. 2014; 506:451–5. [PubMed: 24553141]
72. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–9. [PubMed: 25260700]
73. Hoadley KA, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014; 158:929–44. [PubMed: 25109877]
74. Kawagoe H, Kandilci A, Kranenburg TA, Grosveld GC. Overexpression of N-Myc rapidly causes acute myeloid leukemia in mice. Cancer Res. 2007; 67:10677–85. [PubMed: 18006809]
75. Kamijo T, et al. Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19ARF. Cell. 1997; 91:649–59. [PubMed: 9393858]
76. Mullighan CG, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009; 360:470–80. [PubMed: 19129520]
77. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep. 1966; 50:163–70. [PubMed: 5910392]
78. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc. 1999; 94:496–509.
79. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2009.

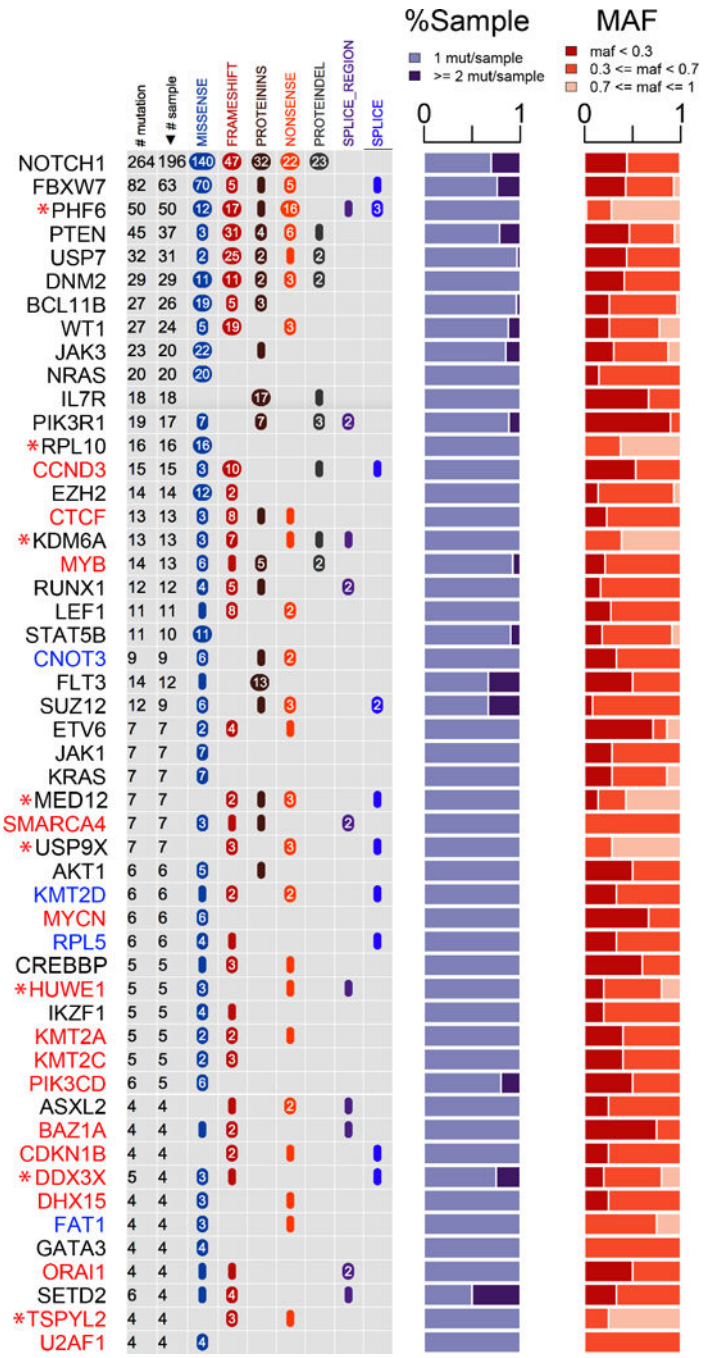

Figure 1. The fifty-five most common targets of sequence mutation in T-ALL

The left panel shows genes with mutations ordered by recurrence, and stratified by mutation type. Gene symbols in blue denote those previously found to be mutated in adult T-ALL cases, but not children. Those in red have not previously been reported in T-ALL. Genes on chromosome X are indicated with asterisks. The center panel depicts the number of mutations for each gene per case. The right panel indicates mutant allele fraction (MAF) of sequence mutation across the cohort.

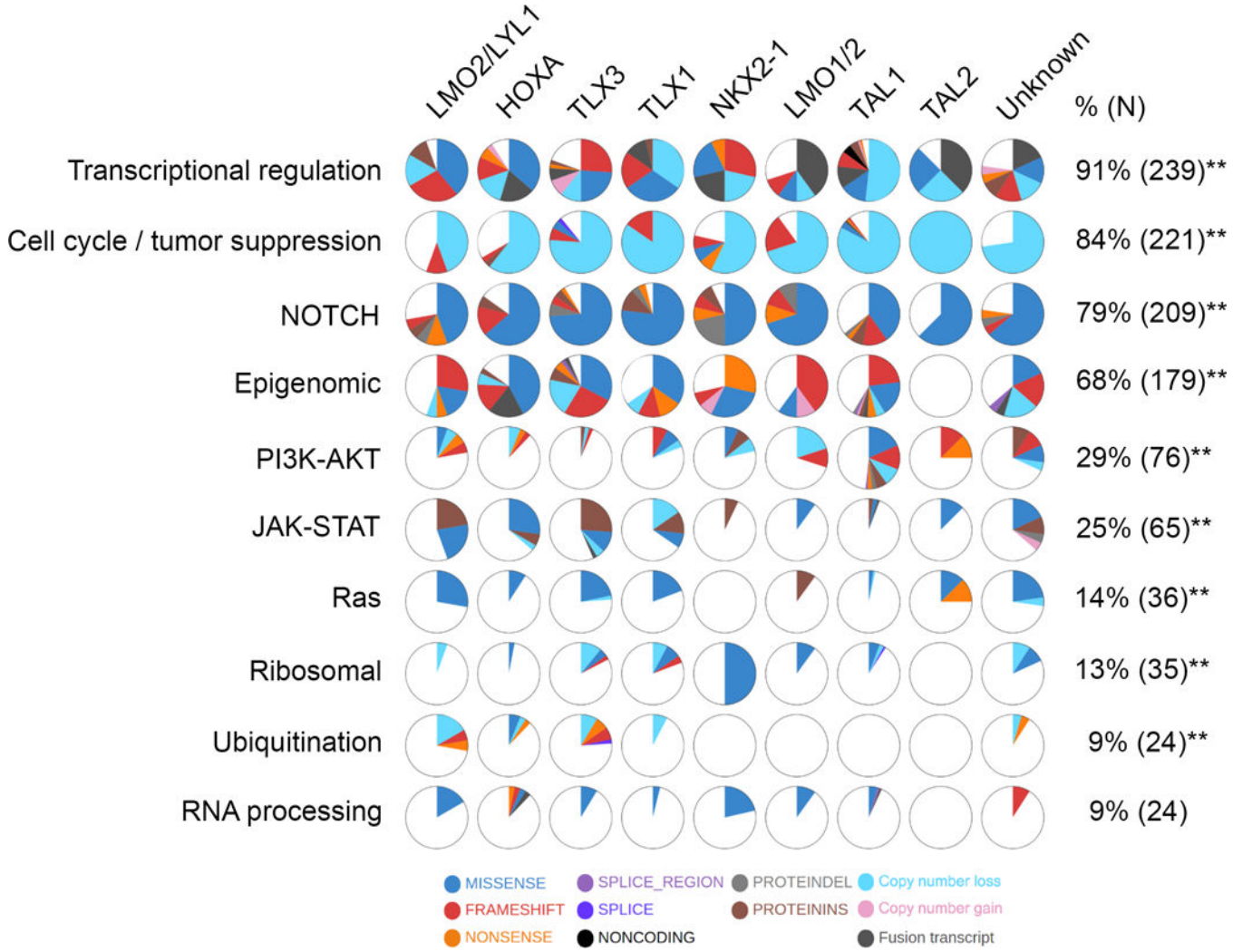

Figure 2. Recurrently mutated pathways in T-ALL

Ten recurrently mutated pathways and the proportion of cases in each T-ALL subgroup shown as pie diagrams. The majority of pathways showed significant associations between prevalence of alteration and T-ALL subgroup, **, $P<0.01$.

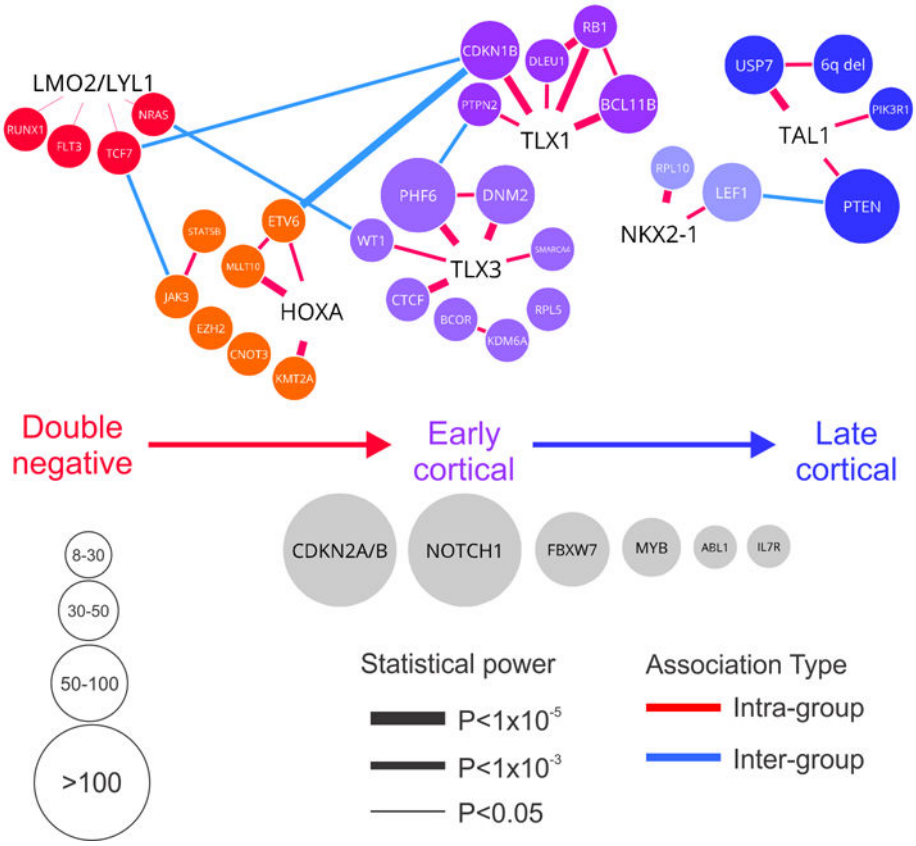

Figure 3. A network depicting association between genetic alterations and T-ALL subgroups and stages of T-cell development

Nodes represent genetic alterations and are color-coded by the T-ALL subtype in which they are most enriched. The size of each node represents the frequency of the genetic alteration across the entire cohort. Lines connecting the nodes indicate statistically significant co-occurrence of a gene-gene pair or a gene-subtype pair. Intra- and inter- subtype co-occurrence is shown in red and blue, respectively with the thickness of each line represents the significance of *P* value. Nodes in gray are genetic alterations that are not specifically enriched in a T-ALL subtype.

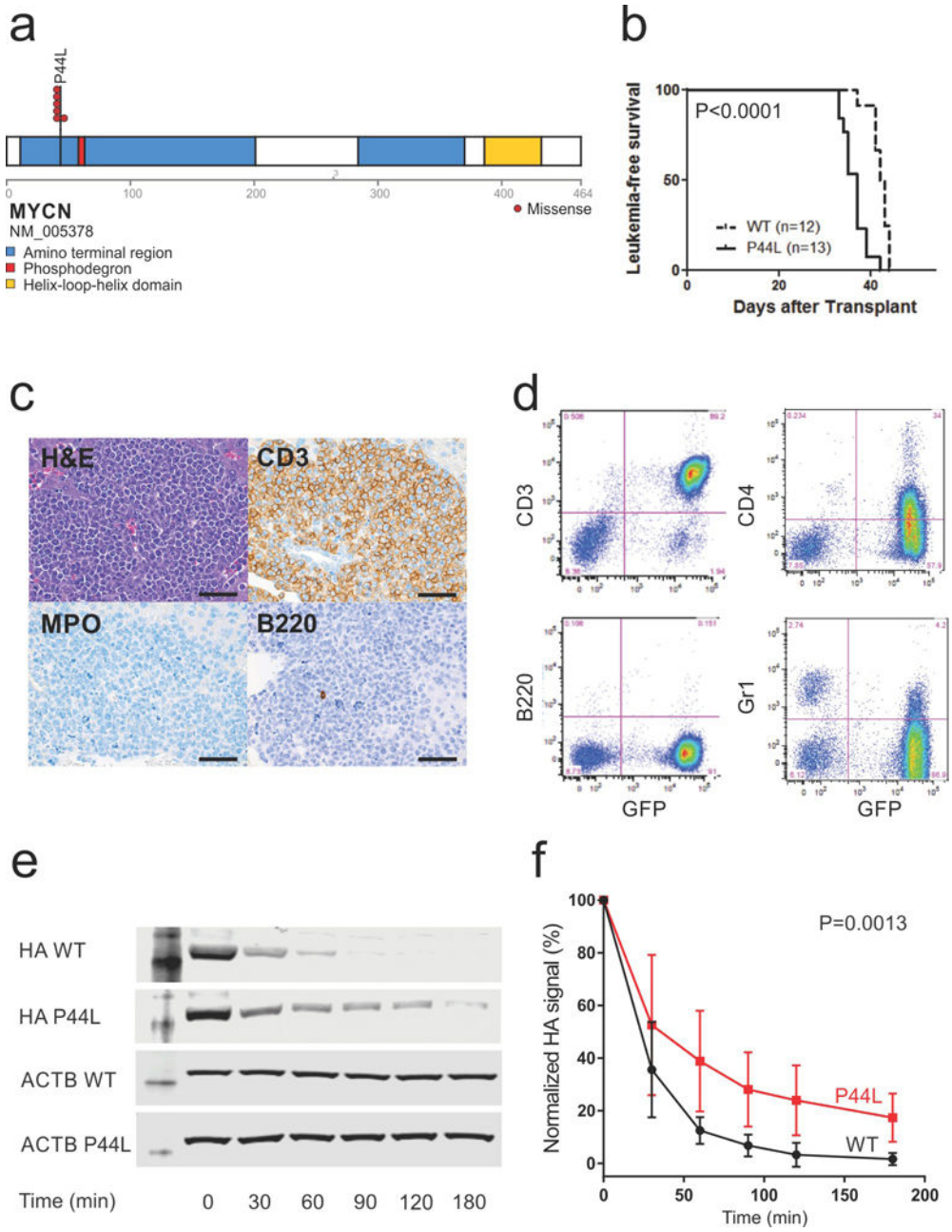

Figure 4. Accelerated leukemogenesis in mutant MYCN driven T-ALL

a, MYCN mutations in T-ALL. The putative MYCN phosphodegron is showin in red. **b**, Kaplan-Meier curves from 1 of two experiments showing leukemia-free survival of mice transplanted with thymocytes transduced with the wild-type P44L MYCN vectors, N=10 per arm. **c**, Leukemic infiltrates were present in the liver (shown), kidney, spleen, thymus, and bone marrow of all animals. Tumors were of T cell lineage (CD3+, RUNX1+, MPO−, GATA1−, TdT−). Bars, 75μm. **d**, Expression of CD3, CD4, B220 and Gr1 in leukemic cells. **e–f**. Immunoblot analysis of MYCN expression (via the HA epitope tag) in NIH3T3 cells showing increased stability following cycloheximide treatment for MYCN P44L. The estimated half-life for wild-type MYCN was 0.47 hours (95% CI −.38–0.6) compared to

0.96 hours (0.77–1.27) for MYCN P44L (P=0.0013). Two independent cycloheximide treatment experiments were performed, with one representative experiment shown in **e** and data in **f** showing mean ± SD. Immunoblots in **e** have been cropped.

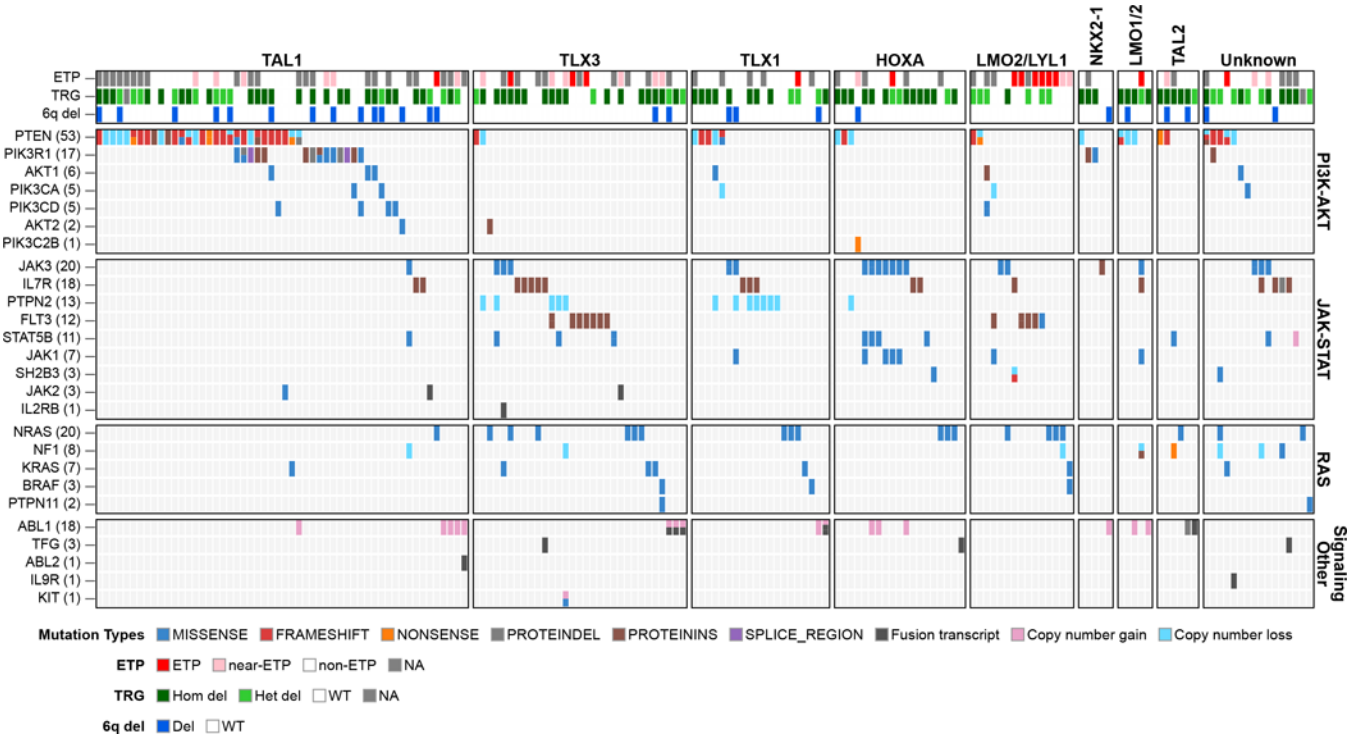

Figure 5. Signaling mutations in T-ALL

Heatmap showing the enrichment of PI3K mutations in TAL1-deregulated ALL cases, and the enrichment of JAK-STAT and Ras mutations in TLX1/3 deregulated cases.

Table 1

Rearrangement resulting in deregulated expression of target gene		
Target Gene	**N**	**Rearrangement partner**
TLX1	17	*TRB* (7), *TRA* (5), *TLX1*-upstream (4), *LINC00502-TLX1* (1)
TLX3	13	*BCL11B-TLX3* (8), *CDK6-TLX3* (2), *TRB* (1), *TLX3-CACS15* (1), *TLX3-TARBP1* (1)
TAL1	58	*STIL* (50), *TRA* (5), *DHX9-TAL1* (1), *GNPAT-TAL1* (1), *TAL1-TARP* (1)
TAL2	6	*TRB* (6)
LMO1	3	*TRB* (2), *TRA* (1)
LMO2	13	*TRA* (6), *TRB* (5), *CSTF3-LMO2* (1), *FOXJ3-LMO2* (1)
LYL1	1	*TRB* (1)
NKX2-1	9	*TRA* (6), *NKX2*-1-*BCL11B* (1), *CDK6-NKX2-1* (1), *NKX2-1-DIO2* (1)
HOXA	7	*TRB* (4), *HOXA* insertion (1), *LINC01260-HOXA* (1), *POLR2E-HOXA* (1)
MYB	11	*TRB* (2), *SLC12A9-MYB* (1), *MYB-PLAGL1* (1), *MYB-BDP1* (1), *MYB-CHMP1A* (1)
		AHI1-MYB (5)
b. Rearrangement resulting in chimeric Fusions		
Gene	N	Partner Gene
MLLT10	15	*PICALM* (9), *DDX3X* (2), *KMT2A* (1), *FAM171A1* (1), *NAP1L1* (1), *CAPS2* (1)
KMT2A	12	*MLLT1* (7), *ELL* (1), *MLLT10* (1), *MLLT4* (1), *MLLT6* (1), *CT45A4* (1)
ABL1	7	*NUP214* (4), *SLC9A3R1* (1), *ETV6* (1), *MBNL1* (1)
NUP98	5	*RAP1GDS1* (2), *CCDC28A* (1), *LNP1* (1), *PSIP1* (1)
TFG	3	*GPR128* (3)
JAK2	2	*PCM1* (1), *CD99* (1)
ETV6	2	*ABL1* (1), *CTNNB1* (1)
ZC3HAV1	2	*ABL2* (1), *AKAP11* (1)
LMAN2	2	*NSD1* (1), *PAPOLA* (1)