nature genetics

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] & Charles G Mullighan[2]

Genetic alterations that activate NOTCH1 signaling and T cell transcription factors, coupled with inactivation of the INK4/ARF tumor suppressors, are hallmarks of T-lineage acute lymphoblastic leukemia (T-ALL), but detailed genome-wide sequencing of large T-ALL cohorts has not been carried out. Using integrated genomic analysis of 264 T-ALL cases, we identified 106 putative driver genes, half of which had not previously been described in childhood T-ALL (for example, *CCND3*, *CTCF*, *MYB*, *SMARCA4*, *ZFP36L2* and *MYCN*). We describe new mechanisms of coding and noncoding alteration and identify ten 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 *HOXA1* deregulated ALL, *PTPN2* mutations in *TLX1* deregulated T-ALL, and *PIK3R1*/*PTEN* mutations in *TAL1* deregulated ALL, which suggests 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-ALL constitutes up to 25% of acute lymphoblastic leukemia (ALL) cases[1,2] and is subclassified according to the stage of thymic maturation, such as the early cortical, late cortical or mature T cell stage. Early T cell precursor T-ALL (ETP ALL) is a high-risk subtype that lacks expression of several T cell surface markers and includes 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* (ref. 7), *TAL2* (ref. 8), *LYL1* (ref. 9) and *OLIG2* (*BHLHB1*)[10]; *TLX1* (*HOX11*)[11], *TLX3* (*HOX11L2*)[12], *NKX2-1*, *NKX2-2* and *NKX2-5* (refs. 13,14); the LIM-only domain genes *LMO1* and *LMO2* (refs. 15,16); the *HOXA* homeobox genes; and *MYC*[17] and *MYB*[18,19]. Additional rearrangements result in the expression of chimeric fusion genes involving *KMT2A* (*MLL*), *HOXA* genes, and tyrosine kinase genes such as *ABL1* (refs. 20,21). Deletion of the *CDKN2A/CDKN2B* tumor suppressor loci is observed in over 70% of T-ALL cases[22], and deletion of genes encoding the cell cycle regulators RB1 (ref. 18) and CDKN1B ($p27^{Kip1}$)[23] is noted in approximately 15% of cases.

Additional mutations include loss-of-function alterations of hematopoietic transcription factor genes, including *BCL11B*, *ETV6*, *GATA3*, *LEF1*, *RUNX1* and *WT1* (refs. 24–28). Several genes that

[1]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [2]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [3]Department of Hematology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [4]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [5]Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, Maryland, USA. [6]Office of Cancer Genomics, National Cancer Institute, Bethesda, Maryland, USA. [7]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [8]University of Texas Southwestern Medical Center, Dallas, Texas, USA. [9]Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA. [10]Department of Pathology, College of Medicine, The Ohio State University, Columbus, Ohio, USA. [11]Department of Pediatrics, Huntsman Cancer Institute and Primary Children's Hospital, University of Utah, Salt Lake City, Utah, USA. [12]Department of Biostatistics, Colleges of Medicine, Public Health & Health Profession, University of Florida, Gainesville, Florida, USA. [13]Department of Pathology, The Cancer Research and Treatment Center, University of New Mexico, Albuquerque, New Mexico, USA. [14]Department of Pediatrics, Perlmutter Cancer Center, New York University Medical Center, New York, New York, USA. [15]Health Sciences Center, University of Virginia, Charlottesville, Virginia, USA. [16]Department of Pediatrics, University of New Mexico, Albuquerque, New Mexico, USA. [17]Seattle Cancer Care Alliance, Seattle, Washington, USA. [18]Department of Pediatrics, Benioff Children's Hospital, University of California at San Francisco, San Francisco, California, USA. [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, USA. Correspondence should be addressed to S.P.H. (hungers@email.chop.edu), J.Z. (jinghui.zhang@stjude.org) or C.G.M. (charles.mullighan@stjude.org).

Received 16 August 2016; accepted 9 June 2017; published online 3 July 2017; doi:10.1038/ng.3909

encode epigenetic regulators and chromatin modifiers are also mutated in T-ALL, including *EED*, *EZH2* and *SUZ12*, which encode core components of polycomb repressor complex 2 (PRC2). PRC2 mediates the repressive histone H3 lysine 27 trimethylation mark; *KDM6A*, which encodes a histone demethylase[29,30]; and *USP7*, which encodes a deubiquitinating enzyme[31]. Several intracellular signaling pathways are activated by mutation, including the PI3K–AKT–mTOR pathway, most commonly affected by *PTEN* deletion[18,32], *PTPN2* deletion[33] or *AKT1* mutation[34]; the JAK–STAT pathway via activating mutations in the interleukin 7 receptor α-chain gene (*IL7R*)[35,36], *JAK1* (ref. 37), *JAK3* (refs. 38,39) or *STAT5B*[40]; Ras–MAPK signaling through *KRAS* and *NF1* mutation[24,41]; and chimeric *ABL1* fusion genes such as *NUP214–ABL1* and *ETV6–ABL1* (refs. 20,21). Mutations in ribosomal biogenesis processing have recently been identified, including in *RPL5*, *RPL10* and *RPL11* (ref. 42) and in the plant homeodomain factor gene *PHF6*, which 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 encompassed relatively small cohorts[24,44], gene panels[31,45] or a single modality of sequencing[42,46]. Here we carried out an integrated, multimodality genomic analysis of a uniformly treated cohort of pediatric and young adult subjects with T-ALL to identify the spectrum and constellations of genetic alterations in this disease.

RESULTS

Molecular classification of T-ALL

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

On the basis of genetic alterations and dysregulated expression of transcription factor genes, as well as hierarchical clustering of RNA-seq gene expression data, we classified 242 (91.7%) subjects into eight subgroups with deregulation of one of the following genes: *TAL1* ($n = 87$), *TAL2* ($n = 8$), *TLX1* ($n = 26$), *TLX3* ($n = 46$), *HOXA* ($n = 33$), *LMO1/LMO2* ($n = 10$), *LMO2/LYL1* ($n = 18$) and *NKX2-1* ($n = 14$) (**Supplementary Table 1** and **Supplementary Fig. 1**). ETP ALL cases commonly showed *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 was most common in the non-ETP ALL group ($\chi^2 P = 0.0004$; **Supplementary Table 7**).

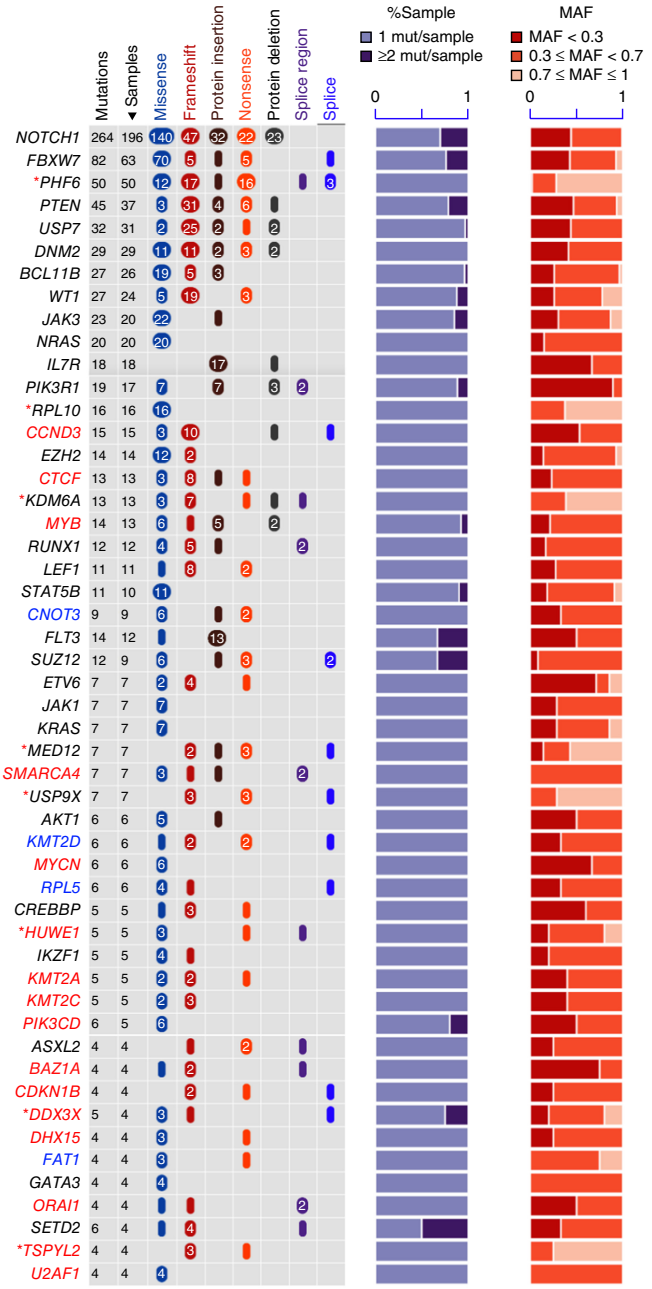

Figure 1 The 55 most common targets of sequence mutation in T-ALL. Left, genes with mutations listed in order of recurrence and stratified by mutation type. Gene symbols in blue are those previously found to be mutated in adult T-ALL cases but not in children. Those in red have not previously been reported in T-ALL. Asterisks indicate genes on chromosome X. The column labeled "%Sample" represents the number of mutations for each gene per case. The far-right column shows the MAF of sequence mutation across the cohort.

Sequence mutations in pediatric T-ALL

We identified 4,165 nonsynonymous single-nucleotide variant (SNV), insertion/deletion (indel) or internal tandem duplication sequence mutations in 2,694 genes by exome sequencing or RNA-seq in the cohort of 264 samples, with a mean of 15.8 mutations (range: 2–50) per case (**Supplementary Table 8**). 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, we identified 106 driver genes (**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%) (**Fig. 1** and **Supplementary Fig. 2**). Fifty-three (51%) of these 106 genes, and 39 (48%) of 82 recurrently mutated genes, had not previously been observed as mutated in pediatric T-ALL[24,31,34,42,44]. These included 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* ($n = 4$); genes in which mutations are predicted to activate PI3K–AKT signaling, such as *PIK3CD* ($n = 5$); and *ZFP36L2*, which encodes an RNA-binding protein and suppressor of *NOTCH1* expression ($n = 5$; three mutations and three fusions)[52,53]. Four genes (*CNOT3*, *KMT2D*, *RPL5* and *FAT1*) were previously identified as mutated in adult, but not pediatric, T-ALL.

We observed subclonal mutations (defined as those with a mutant allele fraction (MAF) of <0.3) frequently in many driver genes (**Fig. 1**), which indicated that subclonal 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, and most mutations in the heterodimerization domain (62.9%; 166/264) or PEST domain (31.4%; 83/264). Although subclonal *NOTCH1* mutations have been reported previously, the extent of multi- and subclonality identified in this study is substantially greater that previously described[54]. Of the 264 mutations, 116 (43.9%) were subclonal. Fifty-eight (29.6%) samples harbored multiple *NOTCH1* mutations (49 samples with two mutations, and 9 samples with three mutations).

To determine whether germline sequence mutations may influence a person's 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. We identified only four potentially deleterious germline mutations, in *BRAF*, *BRCA1*, *BRCA2* and *RUNX1* (**Supplementary Note 1**).

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**). We identified 83 chimeric in-frame fusions, 54 of which arose from interchromosomal rearrangements, and 29 from intrachromosomal events. The most frequent were fusions of *MLLT10* (six unique rearrangements in 13 cases), *KMT2A* (six fusions in 12 cases), *ABL1* (four fusions in 5 cases) and *NUP98* (four fusions in 5 cases) (**Table 1**). We also identified several novel fusions involving signaling pathways, transcriptional regulation or RNA processing, including *ETV6–CTNNB1*, *IL9R–VAMP7*, *PCM1–JAK2*, *CD99–JAK2*, *CSTF3–LMO2*, *STMN1–SPI1* and *NCOR1–NFIA* (one case each). The second class of rearrangements identified by RNA-seq was those that deregulate oncogenes, including *TAL1* ($n = 58$), *TLX1* ($n = 17$), *TLX3* ($n = 13$) and *TAL2* ($n = 6$). Importantly, this analysis not only identified previously unreported 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 Note 1**, **Supplementary Table 12** and **Supplementary Fig. 3**). Whole-genome sequencing (WGS) of 25 cases that lacked 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 the breakpoints were located in intergenic regions and the rearrangements did not result in the expression of a chimeric fusion product: *TRA–LMO2*, *TRB–TAL2*, *BCL11B–TLX3* and *NKX2-1* (**Supplementary Table 1**).

Table 1 Rearrangements in T-ALL samples

Rearrangements resulting in deregulated expression of a target gene

Target gene	*n*	Rearrangement partner (number of cases)
TLX1	17	*TRB* (7), *TRA* (5), *TLX1*-upstream (4), *LINC00502–TLX1* (1)
TLX3	13	*BCL11B–TLX3* (8), *CDK6–TLX3* (2), *TRB* (1), *TLX3–CASC15* (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)

Rearrangements 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), *AFDN* (1), *MLLT6* (1), *CT45A3* (1)
ABL1	7	*NUP214* (4), *SLC9A3R1* (1), *ETV6* (1), *MBNL1* (1)
NUP98	5	*RAP1GDS1* (2), *CCDC28A* (1), *LNP1* (1), *PSIP1* (1)
TFG	3	*ADGRG7* (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)

The transcription factor gene *MYB* was altered by amplification ($n = 33$), rearrangement ($n = 11$) or mutation ($n = 13$) in 49 (18.6%) cases (**Supplementary Fig. 4**). We identified multiple mechanisms of rearrangement, including juxtaposition of the 5′ region of *MYB* to promoter/enhancer regions of partner genes (*TRB*; $n = 2$) and *SLC12A9* ($n = 1$). We identified rearrangements of the 3′ region of *MYB* that resulted in the loss of a negative regulatory region in three cases (involving *PLAGL1*, *BDP1* and *CHMP1A*; $n = 1$ each). We identified 14 *MYB* mutations in 13 samples, including a novel hot spot of missense and in-frame insertion mutations at codon 14 in seven cases (**Supplementary Fig. 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** and **14**)

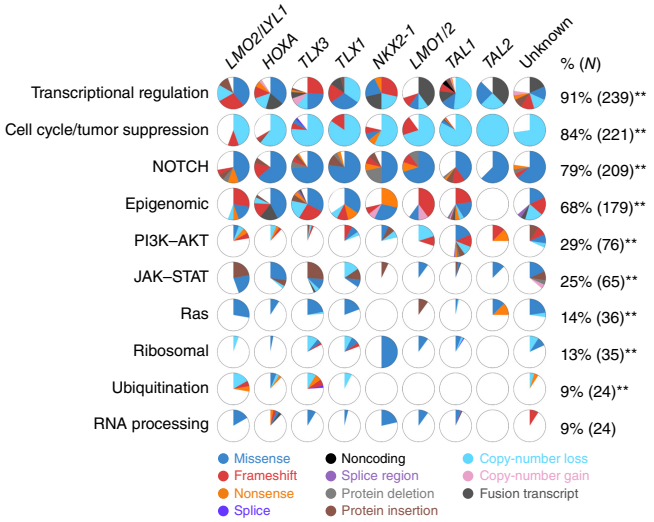

Figure 2 Recurrently mutated pathways in T-ALL. The diagram shows ten recurrently mutated pathways and the proportion of cases in each T-ALL subgroup. The majority of pathways showed significant associations between prevalence of alteration and T-ALL subgroup (** $P < 0.01$, X^2).

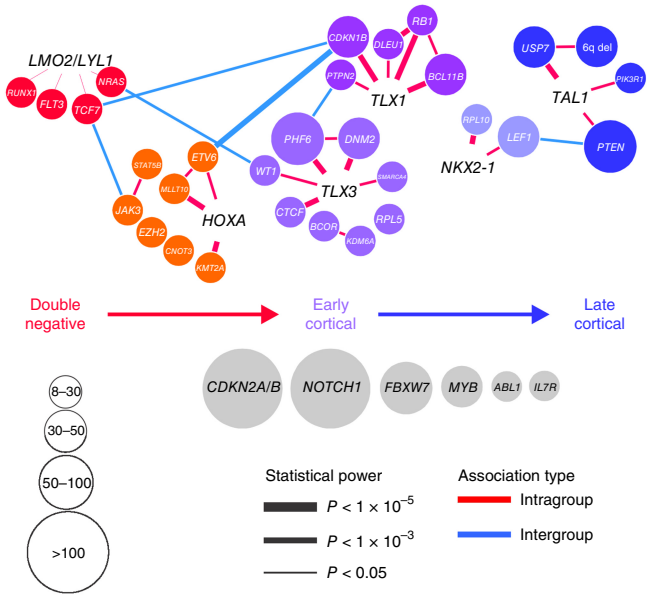

Figure 3 A network depicting associations between genetic alterations and T-ALL subgroups and stages of T cell development. Nodes represent genetic alterations and are color-coded according to the T-ALL subtype in which they are most enriched. The size of each node represents the frequency of the genetic alteration across the entire cohort (see the key in the lower left corner). The thickness of lines connecting the nodes indicates the statistical significance of co-occurrence of a gene–gene or gene–subtype pair. Intra- and intersubtype co-occurrence are indicated by red and blue lines, respectively. Nodes in gray represent genetic alterations that were not specifically enriched in a T-ALL subtype.

and identified ten functional pathways that were 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%) (**Fig. 2** and **Supplementary Tables 15** and **16**). Most commonly mutated were transcriptional regulators, including those oncogenes that define T-ALL subgroups (for example, *TAL1*, *TLX1*, *TLX3*, *LMO2* and *NKX2-1*), known recurrently mutated genes (*MYB*, *LEF1* and *BCL11B*) and 16 genes that were mutated in fewer than ten cases (**Supplementary Fig. 5**). Genes encoding regulators of cell cycle progression and/or tumor suppressors were mutated in 83.7% of cases, with the recurrent targets of alteration being *CDKN2A*/*CDKN2B* (78.4%), *CDKN1B* (12.9%), *RB1* (9.5%) and *CCND3* (6.1%), and single mutations observed 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 Note 1**, **Supplementary Table 17** and **Supplementary Figs. 6–8**).

We observed mutations that activate NOTCH1 signaling in 79% of cases, including mutations in *NOTCH1* (74.6%; 149 heterodimerization, 78 PEST), *FBXW7* (25.4%) or both (20.8% cases) (n = 55) (**Supplementary Fig. 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 (**Fig. 3**, **Supplementary Tables 18** and **19**, and **Supplementary Fig. 10**). Notably, we observed mutations in the NOTCH1 signaling pathway in nearly all *LMO1*/*LMO2*, *NKX2-1* and *TLX1* cases, but less frequently in *TAL1*/*TAL2* 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 to link mutations to T cell developmental stages, considering both gene–subgroup and gene–gene correlation (**Fig. 3**). Mutations of genes in the JAK–STAT or Ras signaling pathway (for example, *FLT3*, *NRAS* and *JAK3*) were enriched in the *LMO2*/*LYL1* and *HOXA* subgroups, whereas those of the 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. We noted significant enrichment of *USP7* alterations in the *TAL1* subgroup, *CTCF* alterations in *TLX3* cases, and *DNM2* and *PHF6* alterations in *TLX1*/*TLX3* cases, suggesting that different epigenetic regulators are involved in the early versus late cortical stages of T cell development. Alteration of the transcription factor gene *LEF1* was most common in *NKX2-1* cases, and changes in *BCL11B* or *MYC* were most common in *TLX1* cases. Mutations in *RPL5* and *RPL10* were enriched in the *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 three 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; **Supplementary Note 1** and **Supplementary Fig. 11**) were the most frequently mutated genes that were not associated with a subtype, which suggests that they 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, we observed mutation of 26 transcription factor genes, with 24 mutated in more than one case[56] (**Supplementary Fig. 5**). A previously unreported mutation in T-ALL was the *MYCN*

mutation encoding p.Pro44Leu (P44L), which we identified in six cases (**Fig. 4a**). This mutation has previously been identified in neuroblastoma and is associated with elevated levels of *MYCN* expression, suggesting a role in oncogenesis[57]. The altered 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 *Cdkn2a*$^{-/-}$ mouse thymocytes with lentiviral vectors expressing either wild-type or P44L MYCN. We cultured transduced cells on OP9-DL1 cells and transplanted them after 10 d into irradiated *Rag2*$^{-/-}$ recipient mice[61]. Both wild-type and mutant MYCN vectors induced a highly penetrant T-lineage leukemia, but we observed a modest but significantly shorter latency in cells that expressed the mutant P44L MYCN vector (**Fig. 4b–d**). To explore the hypothesis that this mutation may perturb ubiquitin-mediated degradation, we transfected NIH3T3 cells with wild-type or mutant hemagglutinin (HA)-tagged MYCN lentiviral vectors, treated the cells with cycloheximide to block protein translation, and quantified MYCN protein levels over time by immunoblotting. We observed significantly slower decay of P44L MYCN compared with that of the wild-type protein (**Fig. 4e**,**f**), consistent with the notion that this mutation impairs protein degradation and thereby enhances oncogenicity.

TAL1-deregulated ALL was the most common subtype of T-ALL in this study (87 cases (33.0%)). Fifty-nine cases had structural alterations identified by RNA-seq or single-nucleotide polymorphism array analysis (51 with deletions that resulted in *STIL–TAL1* fusion, and 8 with other *TAL1* rearrangements), but a substantial minority lacked such an alteration. Recently, a noncoding insertion mutation upstream of the *TAL1* locus was reported that creates a MYB-binding site, resulting in recruitment of CREBBP (CBP), acetylation of histone 3 on lysine 27, and recruitment of RUNX1, GATA3 and TAL1 (ref. 48). We identified 15 cases (5.7%) with identical or similar *TAL1* mutations that resulted in acquisition of the MYB-binding motif. Notably, we identified a distinct indel mutation in four additional cases that resulted in the acquisition of TCF1/TCF2 motifs and loss of a GMEB1 binding motif, suggesting a distinct mechanism for TAL1 deregulation in these cases (**Supplementary Fig. 12**). The remaining 11 TAL1 cases lacked a known genomic mechanism for gene deregulation.

Additional mutated transcription factor genes included *MED12* ($n = 7$) and *TSPYL2* ($n = 4$) (**Supplementary Fig. 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 several different types of tumors[62]. The majority of mutations that we observed were frame-shift or nonsense mutations, which suggests that 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 region that interacts with the *MYC* locus and facilitates MYC expression[63,64] (**Supplementary Fig. 13**). These alterations were most common in the *HOXA* and *TLX1*/*TLX3* subgroups, and included deletions in addition to the previously reported amplifications, both of which were associated with elevated *MYC* expression (data not shown).

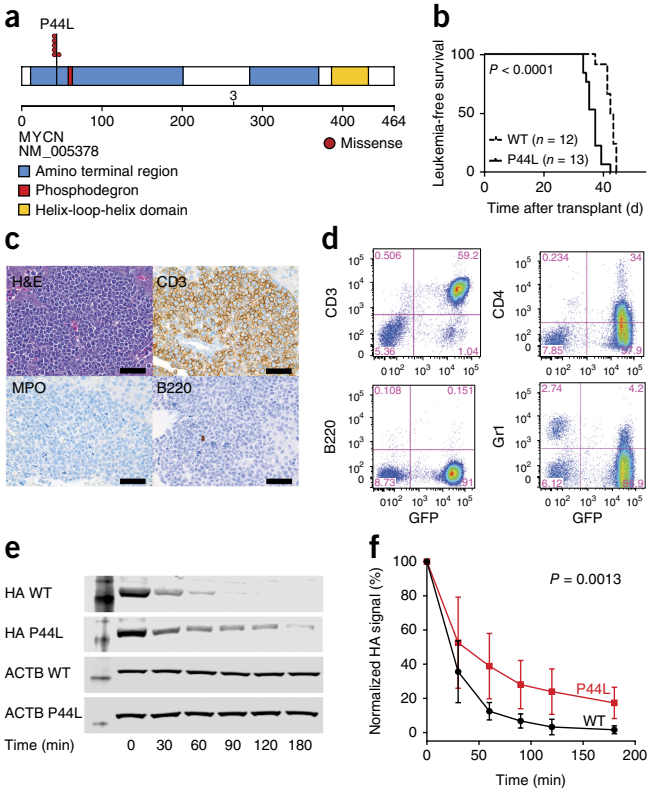

Figure 4 Accelerated leukemogenesis in mutant-MYCN-driven T-ALL. (**a**) *MYCN* mutations in T-ALL. The putative MYCN phosphodegron is shown in red. (**b**) Kaplan–Meier curves from one of two experiments, showing the leukemia-free survival of mice transplanted with thymocytes transduced with wild-type (WT) and P44L MYCN vectors ($n = 10$ per arm; P values determined by log-rank test). (**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$^-$). Scale bars, 75 µm. (**d**) Expression of CD3, CD4, B220 and Gr1 in leukemic cells. Numbers in corners indicate the percentage of cells in each quadrant. (**e**,**f**) Immunoblot analysis of MYCN expression (via the HA epitope tag) in NIH3T3 cells, showing increased stability of P44L MYCN after cycloheximide treatment. The estimated half-life for wild-type MYCN was 0.47 h (95% CI: 0.38–0.6), compared with 0.96 h (0.77–1.27) for P44L MYCN ($P = 0.0013$ extra sum-of-squares F test). We carried out two independent cycloheximide-treatment experiments; one representative experiment is shown in **e**, and the data in **f** show the mean ± s.d. of both experiments. The immunoblots shown in **e** are cropped; full blots are presented in **Supplementary Data Set 1**.

Signaling pathway mutations in T-ALL

We observed mutations that activated signaling pathways in 65% of cases. These included the PI3K–AKT (28.8% of cases), JAK–STAT (24.6%) and Ras signaling (13.6%) pathways, and rearrangements involving *ABL1*, *ADGRG7* (*GPR128*), *IL2RB* and *IL9R* (**Fig. 5**). We observed exclusivity between PI3K–AKT pathway and JAK–STAT or Ras pathway alterations ($P < 0.0001$, Fisher's exact test), with PI3K–AKT mutations prevalent in *TAL1* cases, and JAK–STAT or Ras mutations in *TLX1*/*TLX3* and *HOXA* cases. Although *PTEN*, *PIK3R1* and *AKT1* were the most commonly mutated genes in the PI3K–AKT pathway, we also identified several mutations previously unreported in T-ALL. These included a recurrent somatic *PIK3CD* mutation encoding p.Glu1021Lys, which we observed in five cases (four *TAL1* and one *LMO2*/*LYL1*). This has been reported to be a dominant AKT-activating germline mutation[65]. Other recurring alterations resulting in the activation of signaling pathways included rearrangement of *ABL1* and *TFG–ADGRG7* (*TFG–GPR128*), a rearrangement also observed in high-risk B-lineage ALL and lymphoma ($n = 3$)[66].

Fifty-five cases (20.8%) had multiple signaling mutations, with 13 cases carrying at least three 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

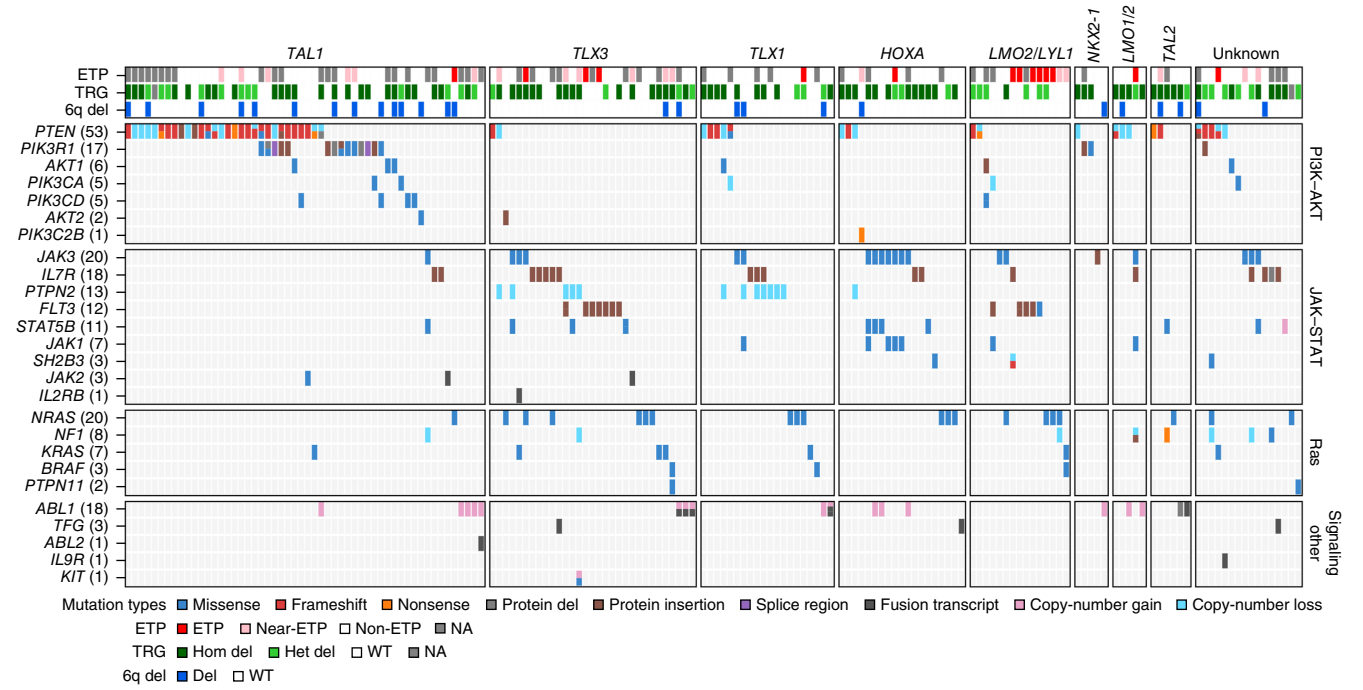

Figure 5 Signaling mutations in T-ALL. The heat map shows the enrichment of PI3K mutations in *TAL1*-deregulated ALL cases, and the enrichment of JAK–STAT and Ras mutations in *TLX1*/*TLX3*-deregulated cases. TRG, T cell receptor-γ locus; Del, deletion; Hom, homozygous; Het, heterozygous; WT, wild-type; NA, data not available.

secondary PI3K–AKT ($n = 13$), Ras ($n = 4$) or other mutations ($n = 3$). This raised the question of whether these concomitant mutations reside in the same or distinct clones. This is of importance for the selection of targeted therapeutic approaches and for the potential for the emergence of subclones resistant to agents that target a single pathway. To address this, we examined the MAF of signaling pathway mutations in 14 cases with *JAK3* sequence mutations. Thirteen of the cases carried a *JAK3* variant with MAF > 0.3 (30%), indicating that the mutations were present in a predominant/major clone (**Supplementary Fig. 14**). In ten of these cases, the signaling mutation with the next highest MAF (including the 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 coexisted in a subclone. In one case, both mutations were subclonal, and may have existed in the same or different clones. Thus, in the majority of cases, the most common JAK–STAT and second most common signaling mutations coexisted in the same clone.

Patterns of epigenetic mutation in T-ALL

Epigenomic regulators were recurrently mutated in all T-ALL subgroups, 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 Fig. 15**). In addition to the known association between *PHF6* mutations and *TLX3* rearrangement[43], we observed several additional associations between mutation and subtype. We observed alterations of *USP7* almost exclusively in *TAL1* cases (29 of 33 *USP7* mutations). *USP7* encodes a ubiquitin-specific protease that stabilizes MDM2 and TP53 (ref. 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 that these mutations result in loss of function (**Supplementary Fig. 2**). Additional gene mutations associated with 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 been described as a distinct form of leukemia with characteristic immunophenotypic, transcriptomic, genetic and clinical features compared with other T-ALL types[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) of our ETP ALL cases fell in the *LMO2*/*LYL1*-overexpressing group, but there were 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% versus 25.0%/21.9%; $P = 0.05$), Ras signaling (36.8% versus 20.8%/10.3%; $P = 0.005$) and epigenetic regulation (84.2% versus 83.3%/65.8%; $P = 0.08$) were more common, and cell cycle alterations (31.6% versus 87.5%/87.0%; $P < 0.001$) were less common (**Supplementary Table 19**). Genes more frequently altered in ETP ALL included *FLT3* (5 of 9 ETP cases; $P < 0.001$), *WT1* (8 of 21 ETP cases; $P < 0.001$), *IKZF1* (3 of 8 ETP cases; $P = 0.029$) and *MED12* (4 of 6 ETP cases; $P < 0.001$); we also noted frequent rearrangement of *NUP98* (3 of 4 ETP ALL cases; $P < 0.001$). *FLT3* was overexpressed in ETP compared with near-/non-ETP cases, with five of the eight cases with the highest *FLT3* expression having concomitant mutations (data not shown). In contrast, genes less commonly mutated in ETP ALL were *CDKN2A*/*CDKN2B*, *CNOT3*, *DDX3X*, *FBXW7*, *LEF1*, *MYCN* and *RPL10*, alterations of which were observed exclusively 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 (**Supplementary Note 1** and **Supplementary Tables 20–23**).

DISCUSSION

Although many of the most frequent mutations in T-ALL have been described previously, our integrated genomic analysis of a large cohort of T-ALL cases identified a large number of unrecognized targets of mutation and demonstrated stage- and subtype-specific associations among genes, cellular pathways and outcomes. Although many mutated genes reside in a 'long tail' of mutations, *in silico* functional analysis shows that many of these mutations are probably deleterious and contribute to the pathogenesis of ALL. Moreover, many such genes fall in pathways known to be affected in T-ALL (including the JAK–STAT, Ras and PTEN–PI3K signaling pathways), which increases 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 frequencies than previously recognized (for example, *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 founding alterations that deregulate hematopoietic transcription factors. This is also a finding of potential therapeutic relevance, as we observed that multiple signaling mutations commonly co-occurred in the same clone. This provides a rationale for targeting these pathways in such cases.

In addition to identifying previously unrecognized targets of mutation, the integrated analysis identified multiple new modalities of mutation in known T-ALL driver genes, including sequence mutations and enhancer alterations that deregulate *MYCN*, novel rearrangements and mutations of *MYB*, rearrangements and complex mutations of *ZFP36L2*, and novel *TAL1* enhancer mutations. As proof of principle, we demonstrated that the *MYCN* mutations accelerated leukemogenesis, in part by stabilizing protein expression. This integrated analysis also suggests that genetic alterations, such as the common broad deletion of 6q, are 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 we did not carry out WGS for all cases. Analysis of RNA-seq, exome and DNA copy-number data did not identify lesions that activated key transcription factors in a proportion of cases. Using WGS, 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 conducted for all cases enabled a detailed analysis of associations between mutated genes and subtypes of ALL that are considered to arise at defined stages of thymic development. We show that although several pathways contained mutations in the majority of cases, mutations in certain individual genes (for example, JAK–STAT, Ras or *PTEN* gene mutations; *ETV6* or *RUNX1* transcription factor gene mutations; and *PHF6* and *USP7* epigenetic mutations) were significantly enriched in different ALL subtypes. This suggests not only that these alterations are leukemogenic cooperative driver events in which the founding alterations deregulate transcription factors, but also that these genes may have central roles in distinct stages of thymic ontogeny and be 'primed' for mutational deregulation after the 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.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

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. MSCV-IRES-YFP retroviral vector was provided by G. Grosveld. (St. Jude Children's Research Hospital, Memphis, Tennessee, USA) OP9-DL1 stromal cells were a gift from J.C. Zuniga-Pflucker (University of Toronto, Toronto, Ontario, Canada). This work was supported in part by the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital, St. Baldrick's Foundation (Scholar Award to C.G.M.), the National Cancer Institute (grants P30 CA021765 (St. Jude Cancer Center Support Grant), U01 CA157937 (to C.L.W. and S.P.H.), U10 CA98543 (to the Children's Oncology Group (COG); Chair's grant and supplement to support the COG ALL TARGET project), U10 CA98413 (to the COG Statistical Center) and U24 CA114766 (to COG; Specimen Banking), Outstanding Investigator Award R35 CA197695 (to C.G.M.), and 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 US Government.

AUTHOR CONTRIBUTIONS

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

1. Hunger, S.P. & Mullighan, C.G. Acute lymphoblastic leukemia in children. *N. Engl. J. Med.* **373**, 1541–1552 (2015).
2. Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat. Rev. Immunol.* **8**, 380–390 (2008).
3. Coustan-Smith, E. *et al.* Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol.* **10**, 147–156 (2009).
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.* **156**, 358–365 (2012).
5. Weng, A.P. *et al.* Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
6. Pear, W.S. *et al.* Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. *J. Exp. Med.* **183**, 2283–2291 (1996).
7. Begley, C.G. *et al.* Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor δ-chain diversity region and results in a previously unreported fusion transcript. *Proc. Natl. Acad. Sci. USA* **86**, 2031–2035 (1989).
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. USA* **88**, 11416–11420 (1991).
9. Mellentin, J.D., Smith, S.D. & Cleary, M.L. *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* **58**, 77–83 (1989).
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. USA* **97**, 3497–3502 (2000).
11. Hatano, M., Roberts, C.W., Minden, M., Crist, W.M. & Korsmeyer, S.J. Deregulation of a homeobox gene, *HOX11*, by the t(10;14) in T cell leukemia. *Science* **253**, 79–82 (1991).

12. Bernard, O.A. *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* **15**, 1495–1504 (2001).
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* **19**, 484–497 (2011).
14. Nagel, S., Kaufmann, M., Drexler, H.G. & MacLeod, R.A. The cardiac homeobox gene *NKX2-5* is deregulated by juxtaposition with *BCL11B* in pediatric T-ALL cell lines via a novel t(5;14)(q35.1;q32.2). *Cancer Res.* **63**, 5329–5334 (2003).
15. Royer-Pokora, B., Loos, U. & Ludwig, W.D. *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* **6**, 1887–1893 (1991).
16. McGuire, E.A. *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.* **9**, 2124–2132 (1989).
17. Erikson, J. *et al.* Deregulation of c-*myc* by translocation of the α-locus of the T-cell receptor in T-cell leukemias. *Science* **232**, 884–886 (1986).
18. Mullighan, C.G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
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* **110**, 1251–1261 (2007).
20. Graux, C. *et al.* Fusion of *NUP214* to *ABL1* on amplified episomes in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **36**, 1084–1089 (2004).
21. Van Limbergen, H. *et al.* Molecular cytogenetic and clinical findings in *ETV6/ABL1*-positive leukemia. *Genes Chromosom. Cancer* **30**, 274–282 (2001).
22. Hebert, J., Cayuela, J.M., 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* **84**, 4038–4044 (1994).
23. Remke, M. *et al.* High-resolution genomic profiling of childhood T-ALL reveals frequent copy-number alterations affecting the TGF-β and PI3K–AKT pathways and deletions at 6q15-16.1 as a genomic marker for unfavorable early treatment response. *Blood* **114**, 1053–1062 (2009).
24. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
25. Gutierrez, A. *et al.* The *BCL11B* tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. *Blood* **118**, 4169–4173 (2011).
26. Gutierrez, A. *et al.* Inactivation of *LEF1* in T-cell acute lymphoblastic leukemia. *Blood* **115**, 2845–2851 (2010).
27. Tosello, V. *et al. WT1* mutations in T-ALL. *Blood* **114**, 1038–1045 (2009).
28. Van Vlierberghe, P. *et al. ETV6* mutations in early immature human T cell leukemias. *J. Exp. Med.* **208**, 2571–2579 (2011).
29. Ntziachristos, P. *et al.* Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513–517 (2014).
30. Van der Meulen, J. *et al.* The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13–21 (2015).
31. Huether, R. *et al.* The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat. Commun.* **5**, 3630 (2014).
32. Palomero, T. *et al.* Mutational loss of *PTEN* induces resistance to NOTCH1 inhibition in T-cell leukemia. *Nat. Med.* **13**, 1203–1210 (2007).
33. Kleppe, M. *et al.* Deletion of the protein tyrosine phosphatase gene *PTPN2* in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **42**, 530–535 (2010).
34. Gutierrez, A. *et al.* High frequency of PTEN, PI3K, and AKT abnormalities in T-cell acute lymphoblastic leukemia. *Blood* **114**, 647–650 (2009).
35. Shochat, C. *et al.* Gain-of-function mutations in interleukin-7 receptor-α (*IL7R*) in childhood acute lymphoblastic leukemias. *J. Exp. Med.* **208**, 901–908 (2011).
36. Zenatti, P.P. *et al.* Oncogenic *IL7R* gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. *Nat. Genet.* **43**, 932–939 (2011).
37. Flex, E. *et al.* Somatically acquired *JAK1* mutations in adult acute lymphoblastic leukemia. *J. Exp. Med.* **205**, 751–758 (2008).
38. Jeong, E.G. *et al.* Somatic mutations of *JAK1* and *JAK3* in acute leukemias and solid cancers. *Clin. Cancer Res.* **14**, 3716–3721 (2008).
39. Bains, T. *et al.* Newly described activating *JAK3* mutations in T-cell acute lymphoblastic leukemia. *Leukemia* **26**, 2144–2146 (2012).
40. Kontro, M. *et al.* Novel activating *STAT5B* mutations as putative drivers of T-cell acute lymphoblastic leukemia. *Leukemia* **28**, 1738–1742 (2014).
41. Balgobind, B.V. *et al.* Leukemia-associated *NF1* inactivation in patients with pediatric T-ALL and AML lacking evidence for neurofibromatosis. *Blood* **111**, 4322–4328 (2008).
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.* **45**, 186–190 (2013).
43. Van Vlierberghe, P. *et al. PHF6* mutations in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **42**, 338–342 (2010).
44. Bandapalli, O.R. *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* **99**, e188–e192 (2014).
45. Vicente, C. *et al.* Targeted sequencing identifies associations between IL7R–JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia. *Haematologica* **100**, 1301–1310 (2015).
46. Atak, Z.K. *et al.* Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. *PLoS Genet.* **9**, e1003997 (2013).
47. Winter, S.S. *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* **62**, 1176–1183 (2015).
48. Mansour, M.R. *et al.* An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
49. Navarro, J.M. *et al.* Site- and allele-specific Polycomb dysregulation in T-cell leukaemia. *Nat. Commun.* **6**, 6094 (2015).
50. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
51. Dees, N.D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
52. Galloway, A. *et al.* RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence. *Science* **352**, 453–459 (2016).
53. Hodson, D.J. *et al.* Deletion of the RNA-binding proteins ZFP36L1 and ZFP36L2 leads to perturbed thymic development and T lymphoblastic leukemia. *Nat. Immunol.* **11**, 717–724 (2010).
54. Mansour, M.R. *et al.* Notch-1 mutations are secondary events in some patients with T-cell acute lymphoblastic leukemia. *Clin. Cancer Res.* **13**, 6964–6969 (2007).
55. Zhang, J. *et al.* Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
56. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).
57. Pugh, T.J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279–284 (2013).
58. Popov, N., Schülein, C., Jaenicke, L.A. & Eilers, M. Ubiquitylation of the amino terminus of Myc by $SCF^{\beta\text{-}TrCP}$ antagonizes SCF^{Fbw7}-mediated turnover. *Nat. Cell Biol.* **12**, 973–981 (2010).
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* **17**, 210–221 (2009).
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.* **10**, 643–653 (2008).
61. Treanor, L.M. *et al.* Interleukin-7 receptor mutants initiate early T cell precursor leukemia in murine thymocyte progenitors with multipotent potential. *J. Exp. Med.* **211**, 701–713 (2014).
62. Taatjes, D.J. The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends Biochem. Sci.* **35**, 315–322 (2010).
63. Yashiro-Ohtani, Y. *et al.* Long-range enhancer activity determines *Myc* sensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl. Acad. Sci. USA* **111**, E4946–E4953 (2014).
64. Herranz, D. *et al.* A NOTCH1-driven *MYC* enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130–1137 (2014).
65. Angulo, I. *et al.* Phosphoinositide 3-kinase δ gene mutation predisposes to respiratory infection and airway damage. *Science* **342**, 866–871 (2013).
66. Roberts, K.G. *et al.* Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N. Engl. J. Med.* **371**, 1005–1015 (2014).
67. Ye, M. *et al.* STIP is a critical nuclear scaffolding protein linking USP7 to p53–Mdm2 pathway regulation. *Oncotarget* **6**, 34718–34731 (2015).
68. Neumann, M., Greif, P.A. & Baldus, C.D. Mutational landscape of adult ETP-ALL. *Oncotarget* **4**, 954–955 (2013).

ONLINE METHODS

Patients and samples. We studied 264 consecutively recruited children and young adults with available banked diagnostic and remission material from the Children's Oncology Group Study of T-ALL (NCT00408005). 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 or at St. Jude Children's Research Hospital. In general, samples with at least 1 µg each of tumor and remission DNA were sequenced at Beckman Coulter (216 samples), and those with limiting amounts of material were sequenced at St. Jude.

Beckman Coulter Genomics. DNA shearing was done with Covaris instrument library construction on a Beckman Coulter Biomek FXP with the Beckman Coulter Genomics SPRIworks HT kit (https://www.beckmancoulter.com/wsr-portal/wsr/research-and-discovery/products-and-services/next-gen-library-preparation/spriworks-ht/index.htm). First-round PCR (4–8 cycles) was done with primers appropriate for Illumina (GA and HiSeq) sequencers, and clean-up steps were done with BC/Agencourt AMPure XP beads. For target capture we used SureSelectXT Human All Exon V5 (Agilent Technologies) and supplied hybridization and associated reagents. Second-round PCR (10–16 cycles) used TruSeq index adaptors. Library quality control was done with a TapeStation (Agilent) and qPCR with Kapa standard curves. Sequencing was done on 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-nt paired-end reads analyzed with the on-board software RTA v1.18 and HCS v2.2.

St. Jude Children's Research Hospital. Library construction used DNA tagmentation (fragmentation and adaptor attachment) with the reagent provided in the Illumina Nextera rapid exome kit, and was done with the Caliper Biosciences (PerkinElmer) Sciclone G3. First-round PCR (ten cycles) was done with Illumina Nextera kit reagents, and clean-up steps used BC/Agencourt AMPure XP beads. For target capture we used the Illumina Nextera rapid-capture exome kit and supplied hybridization and associated reagents. The pre-hybridization pool size was 12 samples, and second-round PCR (ten cycles) was done with Nextera kit reagents. Library quality control was done with a Victor fluorescence plate reader with Quant-it dsDNA reagents for pre-pool quantitation, and an Agilent Bio-analyzer 2200 for final library quantitation. Sequencing was done with Illumina HiSeq 2500 instruments in high-output mode with TruSeq SBS v3 chemistry. All runs were 100-nt paired-end reads, and data were analyzed with the on-board software RTA v1.18 and HCS v2.2.

A detailed comparison of coverage with 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 with BWA[69] (version 0.5.10) and analyzed as previously described[24], including coverage and quality assessment, SNV and indel detection, annotation, and prediction of deleterious effects for sequence mutations. We observed a high C>A/G>T mutation rate in a subset of 43 samples that had a shorter library insert size (range, 88–141 bp; median, 103 bp) than the expected median library insert size of 190 bp. Variants detected in these samples showed 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 the expected library insert size.

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

Driver mutations were identified by MutSigCV[50], MuSiC[51] and Medal Ceremony, a novel algorithm that we developed to assess mutation pathogenicity on the basis of matches to known hot spots in oncogenes or loss of function in known tumor suppressor genes (M.E. *et al.*, manuscript in preparation). Expression of the mutant allele was required in order for a missense mutation to be considered as 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-seq. TRIzol was used to extract RNA from ALL bone marrow or peripheral blood samples. We removed ribosomal RNA from 500 ng of DNase I–treated total RNA with Ribo-Zero rRNA removal beads and converted it into a cDNA library with the Illumina TruSeq stranded Total RNA kit with Ribo-Zero Gold. After adaptor 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-seq analysis. RNA-seq data were mapped with StrongArm (M.R. *et al.*, manuscript in preparation), and rearrangements were identified with CICERO as previously described[71]. The gene-level read count was generated with HTseq-count[72], and the number of fragments per kilobase of transcript per million mapped reads (FPKM) was calculated on the basis of the transcript models in GENCODE v19. Quantile normalization was applied to the $\log_2$ transformed FPKM matrix. Only genes expressed (defined as those with FPKM > 1) in more than 30% of samples were further analyzed[73]. The median absolute deviation was calculated for each gene across the cohort, and the top 500 genes with the largest median absolute deviation values were selected for cluster analysis by Ward's minimum variance method.

Vector constructs and lentiviral production. An MSCV-IRES-YFP (HA-tagged wild-type MYCN cDNA in the mouse stem cell virus–internal ribosome entry site–yellow fluorescent protein) 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. A point mutation at codon 44 that changed the native amino acid from proline to leucine was introduced 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 three 9-week-old female *Cdkn2a*$^{-/-}$ mice[75] housed in ventilated racks in an AAALAC-accredited facility, and single-cell suspensions were prepared. Thymocytes were then stained with CD4–phycoerythrin (PE; BD Pharmingen; 553653), CD8–PE (BD Pharmingen; 553033), and Ter119–PE antibodies (BD Pharmingen; 553673). These antibodies were validated for flow cytometry with reactivity to mouse antigens by the manufacturer. Thymocytes were then 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 per well on OP9-DL1 stromal cells (a gift from J.C. Zuniga-Pflucker, University of Toronto) in α-MEM (Invitrogen) containing 20% FBS, penicillin–streptomycin, glutamine, and sodium pyruvate, and supplemented with 5 ng/ml each of FLT-3 (R&D Systems) and IL-7 (PeproTech). After 24 h, polybrene was added and the cells were transduced with 1 ml of vector supernatant collected from 293T cells, and spun at 2,000 r.p.m. for 1 h at room temperature. This transduction was repeated again at 48 h. Every 2–3 d, thymocytes were replated onto fresh OP9-DL1 cells, and on day 8 the cells were immunophenotyped by flow cytometric analysis. GFP$^+$ OP9-DL1 cells were independently authenticated for the presence of the Notch ligand Delta-like 1 via flow cytometry with PE-conjugated anti-mouse DLL1 (BioLegend; 128307). OP9-DL1 cells that were positive for both GFP and PE were used for coculture. Cells were not tested for mycoplasma contamination.

Vector-transduced thymocytes were harvested from OP9-DL1 cell cultures on day 10 and transplanted by tail vein injection into sublethally irradiated (6 Gy) female *Il2rg*$^{-/-}$*Rag2*$^{-/-}$ mice. In the initial primary transplant, 1×10^6 bulk cells containing 92% or 88% vector-positive cells for wild-type (n = 13 mice; 6 weeks old) or mutant *MYCN* (n = 13 mice; 6 weeks old), respectively, were transplanted. In the second primary transplant, 0.5×10^6 bulk cells containing 70% or 53% vector-positive cells for wild-type (n = 10 mice; 12 weeks old) or mutant *MYCN* (n = 10 mice; 12 weeks old), respectively, were transplanted. No randomization or blinding was performed. All experimental procedures involving mice 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 done with standard procedures. Slides of 4–6-µm sections were cut from formalin-fixed paraffin-embedded tissues, and sections were stained with hematoxylin and eosin or by the avidin–biotin–peroxidase method with myeloperoxidase (Dako; A0398), CD3 (Santa Cruz Biotechnology; sc-1127), B220/CD45R (BD Pharmingen; 553084), CD117 (R&D; AF1356), GATA-1 (Abcam; ab131456), RUNX1 (Abcam; ab92336), and TdT (Supertechs; 004) antibodies. Peripheral blood was collected in 10 mmol/L EDTA for standard complete blood counts and smears stained with Wright–Giemsa. Expression of cell surface markers and GFP was measured by flow cytometry (BD Biosciences Fortessa) with fluorophore-conjugated antibodies to CD4 (552051), CD8 (557959), CD25 (557192), CD44 (560569), CD3 (553066), B220 (561102), Gr1 (557979), Mac1 (553311) and Ter119 (557853), all from BD Pharmingen.

Immunoblot analysis. Protein lysates were extracted from unsorted splenic tumor samples with fresh RIPA lysis and extraction buffer (Thermo Fisher Scientific) prepared with cOmplete protease inhibitor and PhosSTOP phosphatase inhibitor (Roche Diagnostics). Protein concentration was quantified by 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 with a rabbit polyclonal antibody to HA (Y-11; Santa Cruz Biotechnology) at a 1/200 dilution. The secondary antibody used was donkey anti-rabbit (GE Healthcare) at a dilution of 1/500, and an antibody to β-actin (C-4; Santa Cruz Biotechnology) was used at a dilution of 1/1,000 to control for protein loading.

Cycloheximide stability assays. NIH3T3 cells were transduced with HA-tagged constructs expressing either wild-type MYCN or the P44L MYCN mutant, treated with cycloheximide (20 µg/mL), and collected at baseline and 30, 60, 90, 120 and 180 min for protein analysis. Cells were lysed in RIPA buffer (Sigma), and 30 µg of protein was loaded on 4–12% NuPage Bis-Tris gels (Life Technologies) and electrophoresed at 200 V for 1 h. Blots were simultaneously probed with anti-HA-tag (1:1,000; SAB4300603; Sigma) and anti-actin (1:1,000; sc-1615; Santa Cruz Biotechnology), subsequently detected with fluorescently conjugated secondary antibodies, and quantified with Odyssey Imaging System Studio software (v4.0, LI-COR Biosciences). The HA signal was log-transformed, and curves were fitted via the least-squares method and compared by F test in Prism (GraphPad).

Statistical analysis. Associations between categorical values were examined by Fisher's exact test. These included associations of T-ALL subtype with gene/pathway, and subtype/gene/pathway with level of minimal residual disease. Reported P values were not subjected to correction for multiple comparisons. Associations between T-ALL subtype and treatment outcome (event-free survival and overall survival) were done with the Kaplan–Meier estimator with Peto's estimator of s.d. 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 done with Prism v6.0 (GraphPad), R (http://www.r-project.org)[79], and SAS (v9.1.2, SAS Institute, Cary, NC, USA). Integrated analysis and statistical modeling of genomic data are described in detail in **Supplementary Note 1**.

Data availability. Data may be accessed through the 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 sequencing and RNA-seq are accessible through the database of genotypes and phenotypes (dbGaP; http://www.ncbi.nlm.nih.gov/gap) under accession number phs000218 (TARGET) and substudy specific accession phs000464 (TARGET ALL Expansion Phase 2). Gene expression, chromosome segmental copy number, SNV/indel, structural variant, and clinical information is 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 can be explored at the St. Jude PeCan Data Portal, available at http://pecan.stjude.org/proteinpaint/study/target-tall.

69. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
70. Zhou, X. *et al.* Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat. Genet.* **48**, 4–6 (2016).
71. Parker, M. *et al. C11orf95-RELA* fusions drive oncogenic NF-κB signalling in ependymoma. *Nature* **506**, 451–455 (2014).
72. Anders, S., Pyl, P.T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
73. Hoadley, K.A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944 (2014).
74. Kawagoe, H., Kandilci, A., Kranenburg, T.A. & Grosveld, G.C. Overexpression of N-Myc rapidly causes acute myeloid leukemia in mice. *Cancer Res.* **67**, 10677–10685 (2007).
75. Kamijo, T. *et al.* Tumor suppression at the mouse *INK4a* locus mediated by the alternative reading frame product p19^{ARF}. *Cell* **91**, 649–659 (1997).
76. Mullighan, C.G. *et al.* Deletion of *IKZF1* and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* **360**, 470–480 (2009).
77. Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* **50**, 163–170 (1966).
78. Fine, J.P. & Gray, R.J. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* **94**, 496–509 (1999).
79. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2009).

doi:10.1038/ng.3909