The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia

Yu Liu1, John Easton1, Ying Shao1,2, Jamie Maciaszka3, Zhaoming Wang1, Mark R Wilkinson2, Kelly McCastlain2, Michael Edmonson1, Stanley B Pounds1,4, Lei Shi4, Xin Zhou1, Xiaotu Ma1, Edgar Sioson1, Yongjin Li1, Michael Rusch1, Pankaj Gupta1, Deqing Pei4, Cheng Cheng4, Malcolm A Smith5, Jaime Guidry Auvil6, Daniela S Gerhard6, Mary V Relling7, Naomi J Winick8, Andrew J Carroll9, Nyla A Heerema10, Elizabeth Raetz11, Meenakshi Devidas12, Cheryl I Willman13, Richard C Harvey13, William L Carroll14, Kimberly P Dunsmore15, Stuart S Winter16, Brent L Wood17, Brian P Sorrentino3, James R Downing2, Mignon L Loh18, Stephen P Hunger19, Jinghui Zhang1 & Charles G Mullighan2

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 TLI1 deregulated T-ALL, and PIK3R1/PTEN mutations in TALI1 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) cases1,2 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 markers3,4. 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 cases5,6. 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 (BHLHBI1)10, TLI1 (HOX11)11, TLI3 (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 MYC17 and MYB18,19. 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 cases22, and deletion of genes encoding the cell cycle regulators RB1 (ref. 18) and CDKN1B (p27Kip1)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...
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,32,33 PTPN2 deletion33 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 STAT5B40; Ras–MAPK signaling through KRAS and NF1 mutation;41,42; 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 cohorts44,45, gene panels46,47 or a single modality of sequencing.48,46 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 enhancer48,49 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) 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 (χ²P = 0.0004; Supplementary Table 7).

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 MutSigCV50, MuSiC51 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 been previously observed as mutated in pediatric T-ALL.\textsuperscript{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)\textsuperscript{32,33}. 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\textsuperscript{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 patient’s risk of developing T-ALL, we analyzed 89 genes (Supplementary Table 10), including 60 associated with autosomal dominant cancer predisposition syndromes\textsuperscript{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, CTF3–LMO2, STMN1–SPI1 and NCOA1–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, TRA–TALL2, BCL11B–TLX3 and NDX2–1 (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 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 5′ region of MYB that resulted in the loss of a negative regulatory region in three cases (involving PLAGL1, BDF1 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)
and identified ten functional pathways that were recurrently mutated in T-ALL: transcriptional regulation (91% of cases), cell cycle regulation and tumor suppression (84%), NOTCH 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 NXX2-1), 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 TALI 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, NXX2-1 and TLX1 cases, but less frequently in TALI/TAL2 and LMO2/LYL1 cases. Epigenetic mutations were particularly common in TLX3 cases (93.5%), and ribosomal processing mutations were highly enriched in NXX2-1 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 TALI 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 TALI 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 NXX2-1 cases, and changes in BCL11B or MYC were most common in TLX1 cases. Mutations in RPL5 and RPL10 were enriched in the TLX3 and NXX2-1 subgroups, respectively, suggesting that these two ribosomal 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 MBL1–ABL1, one with an in-frame ABL1 deletion, and the remainder with broad amplification of 9q4; 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 (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 oncogenesis57. The altered residue lies adjacent to the conserved phospho-degron recognized by E3 ubiquitin ligases FBXL7, SKP2 and HUWE1, consistent with the possibility that this mutation may perturb MYCN protein levels or protein–protein interactions58–60. 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 mice61. 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 lentivectors, 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 TSPY2L2 (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 tumors62. 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 expression63,64 (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).

**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 ILC9 (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 mutation65. 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...
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 rearrangement48, 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 types3. Prior studies have reported a mutational spectrum enriched for mutations in transcriptional, epigenetic and signaling genes more characteristic of myeloid malignancies3,68. 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 multiclusal 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. Grosfeld. (St. Jude Children’s Research Hospital, Memphis, Tennessee, USA) OP9-DL1 stromal cells were a gift from J.C. Zumiga-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 CA021763 (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; Spectrum 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. J.G.A. and D.S.G. oversaw the NCI TARGET project. M.A.S., J.G.A. and D.S.G. oversaw the NCI TARGET project. J.G.A. and D.S.G. 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 analyses. C.G.M. analyzed genomic data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Hunger, S.P. & Mullighan, C.G. Acute lymphoblastic leukemia in children. N. Engl. J. Med. 373, 1541–1552 (2015).
2. Altfiński, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukemia and lymphoma. Nat. Rev. Immunol. 8, 380–390 (2008).
3. Coustan-Smith, E. et al. Early T-cell precursor leukemia: a subtype of very high-risk acute lymphoblastic leukemia. 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 leukaemic stem-cell line disrupts the T-cell antigen receptor 6-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. Melilentin, J.D., Smith, S.D. & Cleary, M.L. Jk–I, 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)(c1.2;1.2q22) 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. Dereactivation 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(15;14)(q26;p32), is associated with expression of the Bak1L2 gene in T acute lymphoblastic leukaemia. *Leukemia* 15, 1495–1504 (2001).

13. Homminga, I. et al. Integrated transcript and genome analyses reveal NXXK-2 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukaemia. *Cancer Cell* 19, 484–496 (2011).

14. Nagel, S., Kaufmann, M., Drexler, H.G. & MacLeod, R.A. The cardiac homeobox gene NXXK-2 is deregulated by juxtaposition with BCL11B in pediatric TALL cell lines via a novel t(11;14)(q11.2;q32.2). *Cancer Res.* 63, 5329–5334 (2003).

15. Royer-Pokora, B., Loos, U. & Ludwig, W.D. The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Nat. Genet.* 42, 338–342 (2010).

16. Fischel-Ghodsian, N. et al. The (11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia. *J. Exp. Med.* 200, 1738–1742 (2014).

17. 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).
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 SPRWorks HT kit (https://www.beckmancoulter.com/wsr-portal/wsr/research-and-discovery/products-and-services/next-gen-library-preparation/sprworks-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) was 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 fragmentation (tagmentation 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 BWA69 (version 0.5.10) and analyzed as previously described24, 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-oxyG mutations caused by sequencing artifacts introduced by DNA shearing during library construction52. 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 artificial 8-oxyG variants (Supplementary Table 6), none of which were verified.

Driver mutations were identified by MutSigCV50, MuSiC51 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 Portal29 (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 cDNA library with the 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-seq analysis. RNA-seq data were mapped with StrongArm (M.R. et al., manuscript in preparation), and rearrangements were identified with CicERO as previously described31. The gene-level read count was generated with HTseq-count22, 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-transformed FPKM matrix. Only genes expressed (defined as those with FPKM > 1) in more than 30% of samples were further analyzed25. 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 Grosveld74. 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-yfp 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). Ectotropic 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 C57BL/6 mice23 housed in ventilated racks in an AAALAC-accredited facility, and single-cell suspensions were prepared. Thymocytes were then stained with CD4–phycocerythrin (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 × 106 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 rpm 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-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<sup>−/−</sup> Rag2<sup>−/−</sup> mice. In the initial primary transplant, 1 × 10<sup>6</sup> 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<sup>6</sup> 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 cuts 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; AF136), GATA-1 (Abcam; ab131456), RUNX1 (Abcam; ab92336), and TdT (Superwechsels; 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 (530666), 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; sc-1127), B220/CD45R (BD Pharmingen; 553084), CD117 (560569), CD3 (530666), B220 (561102), Gr1 (557979), Mac1 (553311) and Ter119 (557853), all from BD Pharmingen.

**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% Bis-Tris NuPage 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 fluoroconjugated 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<sup>76–78</sup>. 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)<sup>79</sup>, 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.