EUROPEAN
HEMATOLOGY
ASSOCIATION

haematologica
Journal of the European Hematology Association

Targeted sequencing identifies association between IL7R-JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia

by Carmen Vicente, Claire Schwab, Michaël Broux, Ellen Geerdens, Sandrine Degryse, Sofie Demeyer, Idoya Lahortiga, Alannah Elliott, Lucy Chilton, Roberta La Starza, Cristina Mecucci, Peter Vandenberghe, Nicholas Goulden, Ajay Vora, Anthony V. Moorman, Jean Soulier, Christine J. Harrison, Emmanuelle Clappier, and Jan Cools

Targeted sequencing identifies association between IL7R-JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia

Carmen Vicente[1,2], Claire Schwab[3], Michaël Broux[1,2], Ellen Geerdens[1,2], Sandrine Degryse[1,2], Sofie Demeyer[1,2], Idoya Lahortiga[1,2], Alannah Elliott[3], Lucy Chilton[3], Roberta La Starza[4], Cristina Mecucci[4], Peter Vandenberghe[1], Nicholas Goulden[5], Ajay Vora[6], Anthony V. Moorman[3], Jean Soulier[7], Christine J. Harrison[3,8], Emmanuelle Clappier[7,8], Jan Cools[1,2,8]

Affiliations:
[1]Center for Human Genetics, KU Leuven, Leuven, Belgium
[2]Center for the Biology of Disease, VIB, Leuven, Belgium
[3]Leukaemia Research Cytogenetics Group, Northern Institute for Cancer Research, Newcastle University, Newcastle-upon-Tyne, United Kingdom
[4]Hematology Unit, University of Perugia, Polo Unico S.M. Misericordia, Perugia, Italy
[5] Department of Haematology, Great Ormond Street Hospital, London, UK
[6] Department of Haematology, Sheffield Children's Hospital, Sheffield, UK
[7]U944 INSERM and Hematology laboratory, St-Louis Hospital, APHP, Hematology University Institute, University Paris-Diderot, Paris, France

Short title: Targeted sequencing in T-ALL

[8]**Corresponding authors:**
Jan Cools (jan.cools@cme.vib-kuleuven.be),
Christine Harrison (christine.harrison@newcastle.ac.uk),
Emmanuelle Clappier (emmanuelle.clappier@sls.aphp.fr)

Abstract

T-cell acute lymphoblastic leukemia is caused by the accumulation of multiple oncogenic lesions, including chromosomal rearrangements and mutations. To determine the frequency and co-occurrence of mutations in T-cell acute lymphoblastic leukemia, we performed targeted re-sequencing of 115 genes across 155 diagnostic samples (45 adult and 110 childhood cases). *NOTCH1* and *CDKN2A/B* were mutated/deleted in more than half of the cases, while an additional 37 genes were mutated/deleted in 4 to 20% of cases. We identified the IL7R-JAK pathway to be mutated in 27.7% of cases, with *JAK3* mutations being the most frequent event in this group. Copy number variations were also detected, including deletions of *CREBBP* or *CTCF* and duplication of *MYB*. *FLT3* mutations were rare, but a novel extracellular mutation in FLT3 was detected and confirmed to be transforming. Furthermore, we identified complex patterns of pairwise associations, including a significant association between mutations in *IL7R-JAK* genes and epigenetic regulators (*WT1, PRC2, PHF6*). Our analyses showed that *IL7R-JAK* genetic lesions did not confer adverse prognosis in T-cell acute lymphoblastic leukemia cases enrolled in the UK ALL2003 trial. Overall, these results identify interconnections between the T-cell acute lymphoblastic leukemia genome and disease biology, and suggest a potential clinical application for JAK inhibitors in a significant proportion of T-cell acute lymphoblastic leukemia patients.

Keywords: targeted sequencing, T-ALL, signaling, tyrosine kinase, epigenetic modulators

Introduction

T-cell acute lymphoblastic leukemia (T-ALL) comprises a group of aggressive hematological tumors accounting for 10%-15% of pediatric and 25% of adult ALL cases, which are more frequent in males than females[1-3]. T-ALL development is a multi-step process in which different genetic lesions accumulate and alter the mechanisms controlling proliferation, survival, cell cycle and differentiation of T-cells. Loss of the *CDKN2A (p16)* locus and aberrant *NOTCH1* signaling constitute the most predominant oncogenic lesions involved in the pathogenesis of T-ALL. Deletions of the *CDKN2A* locus in chromosome band 9p21 are present in up to 70% of T-ALL,[4] while up to 60% carry *NOTCH1* activating mutations.[1] *NOTCH1* mutations lead to ligand independent cleavage and activation of the intracellular NOTCH1 part (ICN) and/or to stabilization of active protein.[2] NOTCH1 is an essential protein for T-cell development and aberrant activation of NOTCH1 in T-ALL affects many different pathways including the cell cycle, NFκB and PI3K/AKT pathways. In addition, T-ALL cases harbor chromosomal rearrangements that result in aberrant expression of transcription factor genes, such as *TLX1*, *TLX3*, *TAL1*, *LMO2*, and *HOXA*.[5-10] These chromosomal aberrations are often used to classify T-ALL into subclasses associated with expression of one of these transcription factors, which in many cases also resembles a specific block in differentiation.[11]

The mutational landscape of T-ALL also includes somatic mutations of *IL7R*, *JAK3, JAK1*, *PTEN*, and *NRAS* signaling proteins.[12-15] Activating mutations in *JAK1*, *JAK3* or *IL7R* lead to activation of the JAK/STAT pathway, resulting in stimulation of the proliferation and survival pathways in the leukemia cells. Recent next generation sequencing studies identified further recurrent mutations in proteins involved in mRNA degradation and translation (*CNOT3* and *RPL10*) and in proteins implicated in the regulation of chromatin structure, including histone demethylases (*KDM6A/UTX*), and members of the polycomb repressive complex 2 (PRC2: genes *EZH2*, *SUZ12* and *EED*). [16-19] Sequencing studies have suggested that on average 10 to 20 protein altering mutations can be detected in T-ALL cells,[18-20] but their exact frequency and patterns of co-occurrence have not been investigated in detail in large T-ALL cohorts.

Here, we used Haloplex targeted DNA capture followed by Illumina massive parallel sequencing to investigate the coding sequence of 115 recurrently mutated genes in a cohort of 155 T-ALL samples. Our results revealed that 40 genes had genetic alterations (combining sequence mutations with copy number variations) in more than 4% of cases. Our comprehensive sequence analysis identified mutations/copy number variations of the IL7R-JAK signaling pathway members in 27.7% of T-ALL samples screened; an observation with therapeutic potential. Statistically significant pairwise associations were found between

different mutations, indicating the presence of functional interactions among different pathways in T-ALL pathogenesis. Of significance, we found a mutually exclusive relationship between IL7R-JAK mutations and the presence of *TAL1/LMO2* rearrangements. We also identified positive correlations among IL7R-JAK mutations and mutations/deletions in *PHF6* and members of the PRC2 complex. Our findings begin to unravel the diversity of genetic lesions that are implicated in T-ALL development.

Methods

DNA samples

T-ALL patient samples (n=155: 111 children, 44 adults) were collected from various institutions (**Supplementary Table 1**). Diagnosis of T-ALL was based on morphology, cytochemistry and immunophenotyping according to the World Health Organization criteria. Genomic DNA was isolated from bone marrow (either fixed or fresh bone marrow cells). [5,21,22] To investigate the prognostic relevance of *IL7R*, *JAK1* and *JAK3* mutations, screening for mutations in these three genes by Sanger sequencing was performed in an independent cohort of 78 T-ALL patients. Those patients were all enrolled into the United Kingdom (UK)_ Children's Cancer and Leukaemia Group (CCLG) ALL2003 trial.[23] This study was approved byt the ethics committees of the institutes involved and informed consent was obtained. Samples and clinical data were stored in accordance with the declaration of Helsinki.

Capture design

The SureDesign software was used to design two slightly different Haloplex capture assays (**Table 1**). The total amplicon number for design A was 23127 with a region size of 472.006 kbp and a predicted target coverage of >99%. For design B the total amplicon number was 19694 with a region size of 418.373kbp and a predicted target coverage of >99%. For this study, 80 samples were processed with Design A and 75 samples with Design B. In both assays, the coding exons of selected genes (based on RefSeq, CCDS and VEGA databases) were targeted with an extra 10 bases upstream and downstream. Targeted regions comprised the coding sequence of genes that were either recently identified as recurrently mutated in ALL or other hematological malignancies (known driver genes) or were similar to known oncogenes (candidate driver genes) to be sequenced[18,19,24,25]. For statistical analyses we only considered the 115 genes that were sequenced in both Haloplex designs (**Supplementary Table 2**). Library preparation and sequencing was performed as described in the **Supplementary material**.

Data analyses

In the NextGENe software (v2.2.1, Softgenetics, State College, PA, USA), we performed the following steps: (1) the fastQ output file was converted into a FASTA file to eliminate reads that were not "paired" and that did not meet the criteria of the default settings; (2) reads from the converted unique FASTA file were aligned to the reference genome (Human_v37.2). After alignment a *.pjt file was created and opened in the NextGENe Viewer; (3) a mutation report was created using the coordinates from the targeted enrichment kit as a *.bed file to enable calling of single nucleotide variants (SNVs) and small insertion/deletions (indels) in the regions of interest; (4) an expression report was created from which the mean, minimal and maximal coverage per target and targeted nucleotide was calculated. The coverage was defined as the average number of reads representing a given nucleotide in the reconstructed sequence. To interpret the data, additional custom-filtering criteria were imposed to minimize false-positive rates (**Supplementary material**). Polymorphisms annotated in dbSNP138 or 1000 Genomes databases were excluded from the analyses. For variant calling we also required as a minimum a read depth of 20 and an allele frequency of at least 15%.

In vitro cell experiments

JAK3 wild type cDNA and mutants were generated by GenScript, and were cloned into the MSCV-GFP vector. Viral vector production, retroviral transduction and culture of Ba/F3 cells were performed as previously described.[15] Western blot analyses were performed as described in the Supplementary material.

Results

Sequencing metrics and validation of the gene panel

Next-generation sequencing studies have contributed significantly to our understanding of the genomic landscape of T-ALL. We recently profiled two cohorts of T-ALL using exome sequencing (67 cases)[18] and RNA-sequencing (31 cases),[24] while Zhang and colleagues profiled immature T-ALL cases (also known as early thymic or T-cell precursor T-ALL, ETP-ALL) ETP-ALL using whole genome sequencing (12 cases)[19] and targeted re-sequencing in 94 childhood T-ALL cases. To determine the spectrum of mutations present in both adult and childhood T-ALL, we selected 155 T-ALL cases (44 adult/111 childhood) to be analyzed by targeted re-sequencing. These cases were derived from two different cohorts from France (n= 80) and UK (n= 75). Based on all available sequence data, we selected 115 genes that were recurrently mutated in ALL or other hematological malignancies.

We used Haloplex enrichment to capture all exons of the selected genes, followed by

Illumina sequencing. Two slightly different designs were used to profile the two T-ALL cohorts, which yielded a total of 12.9 Gb of sequence, with capture efficiencies of 45 to 60%. The average coverage per sample was 209X. All genes were covered at >20X for at least 95% of their coding regions (**Table 1**). After excluding sequencing/mapping errors and known polymorphisms, 685 single nucleotide variants (SNVs)-small insertions/deletions (indels) were identified in 103 genes as high-probability changes.

In order to validate the accuracy of our Haloplex assay and bioinformatics analyses (to which we will refer as Haloplex analyses), we sequenced 158 variants across 33 genes using Sanger sequencing (**Supplementary Table 3**). We confirmed 137 of the 158 predicted SNVs and indels (86.7%). Furthermore, we also found that 124 additional SNVs and indels (although not validated here by Sanger sequencing), were already described and documented in the COSMIC database.[26] To determine the sensitivity of Haloplex for detection of SNV and indels, we examined sequence mutations in *CNOT3* (exon 5, n= 3/79) and *RPL10* (exon 4, n= 2/79) and indels in *IL7R* (n= 12/155) and *FLT3* (n= 1/155) genes, for which Sanger sequencing data were available. Haloplex identified all known point mutations in *CNOT3* and *RPL10*, but failed to detect 6 of 12 *IL7R* indels and 1 of 1 *FLT3-ITD* (**Supplementary Table 3**).

Data from Haloplex target enrichment can be also used to identify copy number variations (CNVs) at the captured loci.[27] To this end, we normalized the coverage of all genes against the on-target mapped nucleotides across samples. First we looked at X-chromosome genes (*MTMR8, KDM6A, PHF6, MAGEC3, RPL10* and *USP9X*). As shown in **Figure 1A**, all samples except one (TL91) showed read depths for X chromosome genes consistent with patient gender, with females showing a ~2 fold higher coverage than males. No copy number aberrations of *MAGEC3* or *USP9X* genes were found. There was evidence for deletion of *KDM6A* (1 sample), *MTMR8* (1 sample) and *PHF6* (4 samples), with possible duplication of *RPL10* (2 samples) or *PHF6* (1 sample) genes (**Figure 1A**, **Supplementary Table 3**).

To determine the accuracy of Haploplex data for the detection of CNVs, we examined CNVs at the *MYB* (n= 155), *CDKN2A* (n= 95) and *PTPN2* (n= 80) loci for which FISH, MLPA or arrayCGH data were available. Overall, 64% (9/14) of cases with *MYB* duplication, 93% (69/74) of cases with *CDKN2A* deletion and 100% (6/6) of cases with *PTPN2* deletion were identified by our Haloplex copy number analysis (**Figure 1B**, **Supplementary Table 4**). Haloplex data predicted four additional cases with *MYB* duplication that have a lower normalized coverage when compared with those samples with true *MYB* duplications. However, as they were not confirmed by any other method, we concluded that these were

false positive results. Together, our data indicated that Haloplex re-sequencing data can be used for the identification of SNVs and CNVs with high confidence, with deletions being easier to detect than duplications. The detection of larger indels is difficult, especially for longer insertions.

Landscape of genetic alterations in T-ALL

In total, 153 of the 155 patient samples (98.7%) harbored at least one sequence mutation/copy number change, with a median number of 5 (1–65) genetic lesions per sample. Known T-ALL tumor suppressor genes such as *CDKN2A/B*, *PHF6*, *PTEN* and *PTPN2* were frequently identified as common targets of gene deletions. From the 115 genes analyzed, 40 genes showed genetic alterations (SNV, indel or CNV) in more than 4% of cases (**Figure 2**). Only *NOTCH1*, *CDKN2A* and *CDKN2B* aberrations were found in more than 50% of cases. Mutation frequencies of *PHF6* (19.4%), *FBXW7* (17.4%), *WT1* (14.8%), *PTEN* (12.3%), *BCL11B* (12.3%) and *PTPN2* (7.7%) were in the range of previously reported frequencies.[2] Our data confirmed that *JAK3* was frequently mutated in T-ALL (16.1%).[19] Other recently identified mutations were also found including *CREBBP* (7.7%),[28] *CNOT3* (5.2%)[18] and *RPL10* (5.2%).[18] In contrast to AML, *FLT3* mutations were rare among the T-ALL cases, although we identified one new mutation in the extracellular domain of FLT3 (S471C), which was confirmed to be a transforming mutation (**Supplementary Figure 1**).

Genes with identified genomic lesions were grouped into functional pathways linked to T-ALL pathogenesis (**Supplementary Table 2**). The most frequently affected pathway was transcriptional regulation, with mutations observed in as many as 81.9% of cases, followed by genes associated with cell cycle regulation (74.2%), chromatin modification (38.1%), genes encoding kinases (29.7%), phosphatases (26.5%) and DNA repair pathways (18.7%).

The mean number and spectrum of some of the identified genetic lesions closely correlated with patient age and T-ALL subtype. Alterations in *PTPN2* (p=0.005; 18.2% vs. 3.6%), *MYB* (p=0.008; 22.7% vs. 6.3%) and *CREBBP* (p=0.039; 15.9% vs. 4.5%) were more prevalent among adult cases. In contrast, *EZH2* (p=0.006, 13.5%) and *RPL10* (p=0.106, 7.2%) aberrations were exclusively found in children, although the trend did not reach significance for *RPL10*.

We also found new and previously reported correlations between genetic lesions and T-ALL subgroups. Immature T-ALL cases had a lower incidence of *CDKN2A* (p<0.001, 13.3% vs. 81.9%) and *CDKN2B* (p<0.001, 13.3% vs. 66.4%) deletions,[19] while the incidence of mutations in *JAK3* (p=0.001, 53.3% vs. 12.9%)[19], *JAKMIP2* (p=0.010, 26.7% vs. 4.3%),

RUNX1 (p=0.020, 20% vs. 2.6%)[19] and *IL7R* (p=0.043, 26.7% vs. 7.8%)[19] was significantly higher in those cases. Aberrations of *PHF6* (p=0.006, 53.8% vs. 16.8%),[29] *PTPN2* (p=0.018, 30.8% vs. 6.5%),[22] and *CDKN2B* (p=0.014, 92.3% vs. 55.1%) were more frequently found among *TLX1* positive cases. *TLX3* positive cases had significantly more lesions in *CDKN2A* (p=0.001, 92.9% vs. 68.8%), *CDKN2B* (p=0.001, 85.7% vs. 52%), *DNM2* (p=0.006, 28.6% vs. 8%), *WT1* (p<0.001, 42.9% vs. 8.8%)[30] and *JAK1* mutations (p=0.001, 17.9% vs. 0.8%). *PHF6* (p=0.005, 7.4% vs. 26%),[29] *JAK3* (p=0.002, 3.7% vs. 23%) *DNM2* (p<0.001, 18%), *PTPN2* (p=0.009, 12%) and *WT1* (p=0.001, 1.9% vs. 22%) lesions were underrepresented in *TAL1/LMO2* positive patients (**Figure 3**).

The IL7R-JAK axis is an important oncogenic pathway in T-ALL

The IL7R-JAK signaling cascade is an essential signaling pathway in hematopoiesis, and somatic mutations in *IL7R*, *JAK1*, *JAK3* and *STAT5* have been reported previously in T-ALL.[12,15,18,31] However, the exact frequency of mutations for all members of the signaling pathway is unclear. In our series, sequence mutations/copy number variations of the *IL7R*, *JAK1*, *JAK3* and *STAT5* genes were found in 27.7% of cases, confirming that this signaling pathway is an important oncogenic axis in T-ALL (**Figure 4A**). Incidences of cases with genetic lesions of *IL7R*, *JAK1*, *JAK3* and *STAT5* genes were 9.7%, 4.5%, 16.1% and 4.5%, respectively. We found that 5.8% of cases carried combinations of these mutations (**Supplementary table 2**).

We identified *JAK3* as the most common targeted member of the pathway (16.1% of T-ALL cases). Most mutations were located within the pseudokinase domain, which has a regulatory function on kinase activity (**Figure 4A**). The M511I mutation was the most frequent. In 34.7% of JAK3 mutant cases the M511I mutation was detected together with a second mutation of *JAK3* (**Figure 4B**). Seven patients had mutations/duplications of *IL7R* (n=2), *JAK1* (n=3) or *STAT5* (n=2) in addition to a *JAK3* mutation.

The majority of *JAK3* mutations identified in T-ALL have been confirmed as activating mutations,[15] but the transforming capacity of four JAK3 mutations identified in this study (Q283H, V678L, Q865E and E958K) was not studied previously. We expressed these mutants in Ba/F3 cells and confirmed that JAK3 V678L and Q865E were capable of transforming the cells to IL3 independent growth (Figure 4C). The JAK3 Q283H and Q865E mutants were not transforming (**Figure 4C**), suggesting that these may be passenger mutations. In summary, our results show that the IL7R-JAK signaling pathway is frequently mutated in T-ALL, with *JAK3* being the most frequently altered gene (14.9% of T-ALL cases harbor transforming JAK3 mutations).

Mutations of the IL7R-JAK axis and epigenetic modulators are associated in T-ALL

Having observed that the components of the *IL7R-JAK* axis were frequently mutated in T-ALL, we analyzed the possible associations between these mutations and the different T-ALL subgroups (**Table 2**). The cases having *IL7R-JAK* mutations were more frequently found among *TLX3*+ cases (31% vs. 13.5, p=0.013), immature T-ALL cases (25% vs. 5.5%, p=0.002),[19] or *HOXA*+ cases (20.9% vs. 5.4%, p=0.006), but underrepresented among *TAL1/LMO2* positive cases (p=0.001, 14% vs. 43.2%). Of importance, we observed that mutations of *IL7R-JAK* were positively associated with mutations in epigenetic factors (p=0.001, 42.4% vs. 18.8%). More detailed analyses revealed specific associations between mutations in *IL7R-JAK* and *PHF6* (p=0.002, 34.9% vs. 13.4%), *WT1* (p=0.001, 30.2% vs. 8.9%), and members of the PRC2 complex (*EZH2, SUZ12* and *EED*) (p=0.033, 27.9% vs. 13.4%) (**Table 2**). If we take into account the two *JAK3* mutations (Q283H and Q865E) that lack transforming potential, the significance of these associations is not altered (Supplementary material). Together, these results indicate that functional interactions are present between the *IL7R-JAK* signaling pathway and epigenetic modifiers (WT1, PHF6, PRC2).

Co-occurring mutations and analysis of clonality in cases with IL7R-JAK mutations

Clonal evolution has been documented in a range of hematological malignancies.[32] The proportion of sequencing reads reporting a given mutation can be used to identify whether mutations are present in the major clone or a subclone at the time of diagnosis.

Overall, we did not observe any specific genes that were more frequently mutated in subclones. As we had detected a high frequency of *JAK3* mutations, we analyzed the 23 cases with *JAK3* mutations in more detail (**Figure 5**). Of these cases, 82.6% showed the presence of the *JAK3* mutation in more than 35% of the cells, indicating that *JAK3* mutations occur predominantly in the major clone at diagnosis. One of these cases carried a *JAK3* M511I mutation at an allele frequency of 96.4%. As there were no copy number variations and the allele frequency for the SNPs present in that genomic region was around 100%, we conclude that copy number neutral loss of heterozygosity was responsible for this homozygous *JAK3* mutation. In cases with two *JAK3* mutations or with *JAK3* and *JAK1* mutations together, one *JAK3* mutation was typically present in the major clone (>40% VAF in 9 out of 12 cases), with a minor clone having the second mutation (or both mutations). These data confirm that a single *JAK3* mutation can contribute to leukemia development, while the acquisition of an additional *JAK3* or *JAK1* mutation is likely to be of benefit for driving the growth of subclones during disease progression.

Mutations of IL7R, JAK3 and JAK1 genes do not have an adverse prognostic impact

To gain insight into the clinical relevance of *IL7R*, *JAK3* and *JAK1* mutations, we analyzed their prognostic significance in a cohort of 78 T-ALL cases treated on the UKALL2003 protocol.[23] Sanger sequencing analyses revealed that 10 of the 78 cases (12.8%) carried mutations in IL7R-JAK genes. The seven males and three females were aged 2 to 22 years old and 7 were classified as NCI high risk. All patients achieved complete remission within 28 days but 4/8 were minimal residual disease (MRD) positive at this time point. Eight of the 10 patients (80%) remain in first complete remission 3.7-6.9 years from diagnosis. The other two patients (both carrying *IL7R* indels) relapsed. One male who presented before his second birthday and with a very high white cell count (497×10^9/L) experienced an isolated central nervous system relapse after 9 months and subsequently died. The second patient, a 13-year-old male, experienced an isolated bone marrow relapse after 18 months, but remains alive 3.3 years after an HLA-matched unrelated donor transplant despite failing induction after relapse. In conclusion, this cohort of T-ALL patients fared quite well, and there is no evidence that IL7R-JAK mutations are associated with a poor prognosis.

Discussion

We have performed targeted sequencing of a large cohort of T-ALL cases, coupled with a comprehensive study of their genetic background, in order to elucidate molecular lesions and their co-occurrence in patient samples. We used Haloplex enrichment to sequence the entire coding regions of 115 genes (known T-ALL driver genes or candidate driver genes). We showed that Haloplex sequencing for the detection of SNVs is highly specific and sensitive. Deletions of entire genes were also detected with a high degree of accuracy. In contrast, the detection of gene duplications and indels, such as *MYB* duplications and insertions into the *IL7R* and *FLT3* genes was less reliable, leading to false positive detection of duplications and false negative detection of indels. Similar findings were recently shown for targeted re-sequencing in acute myeloid leukemia, and it was shown that increased coverage could improve the identification of copy number alterations in Haloplex assays.[27] Haloplex analyses should not be preferred over aCGH or MLPA techniques for detection of copy number variations, especially for the detection of duplications.

Visual inspection of the reads for *IL7R* and *FLT3* identified mutant alleles with indels, which had been missed by the analysis software. However, in most cases the mutant reads were missing, suggesting that the mutant exons had not been adequately captured during the target enrichment step. These observations highlight the need for further refinements to the

alignment algorithms and the capture design for cancer samples, which will greatly increase the ability to detect large indels such as those present in *IL7R* and *FLT3*.

Once the Haloplex method was successfully validated, we used this approach to characterize the mutational profile of a cohort of 155 T-ALL cases. Targeted re-sequencing was performed in absence of germline/remission matched DNA and this should be considered. However, the fact that we identified known gene mutations/copy number variations at the same frequencies in our T-ALL cohort as reported in the literature, confirms both the validity of our approach and our cohort. Similar to observations in other tumors, only a few genes (*NOTCH1*, *CDKN2A/B*) were mutated or deleted in >50% of the T-ALL cases, while a large number were mutated in <20% of the cases. These observations corroborate the complexity and variation of events underlying T-ALL malignant transformation.

We observed that the IL7R-JAK signaling pathway was targeted in 27.7% of T-ALL. These findings have potential therapeutic implications, as JAK kinase inhibitors are known to target these alterations. [15,33-38] *JAK3* is the most frequently mutated gene (16.1% of T-ALL cases) within the IL7R-JAK signaling pathway. Based on the results described in this manuscript and our previously published work,[15] the transforming capacities of a total of 16 JAK3 mutants have been tested using the Ba/F3 in vitro cell system. Six of these 16 JAK3 mutants (R272H, Q283H, R403H, Q865E, R925S and E1106G) lacked transforming potential, illustrating that results from sequencing need to be confirmed by functional assays to distinguish driver mutations from passenger mutations. If we exclude the *JAK3* mutations that lack transforming potential, then the frequency of cases with *JAK3* mutations in our series is 14.9% (instead 16.1%).

Some T-ALL cases harbored two transforming *JAK3* mutations, which suggests that those T-ALL cells that are dependent on a *JAK3* mutation may gain proliferative advantage by mutating the second *JAK3* allele. In line with this hypothesis, our clonality analysis of *JAK3* mutated cases showed that in those with two JAK3 mutations or a *JAK3* and *JAK1* mutation together, one *JAK3* mutation was typically present in the major clone, with the second mutation found in a lower percentage of the cells, most likely representing a minor clone.

We also investigated the prognostic impact of mutations of *IL7R*, *JAK1* and *JAK3*, using data from the UKALL2003 trial, in particular because these mutations occur more frequently in immature T-ALL, a subgroup of T-ALL initially associated with a poor outcome in some studies.[39]

Despite the small subset of *IL7R*, *JAK1* or *JAK3* mutant T-ALL patients (n=10) in this cohort, it appeared that in general these patients fared well. It should be noted that the outcome of immature T-ALL patients has improved since initial reports of this subtype, and that more recent data from UKALL2003 and COG AALL0434 suggest immature T-ALL patients may not have worse outcome with intensified therapy.[23,40] In fact, seven of the 10 patients with mutations within the IL7R-JAK pathway received very intensive chemotherapy based on risk stratification. Our data provide no evidence that *IL7R*, *JAK1* and *JAK3* mutations are a consistent marker of poor prognosis. Nevertheless, it might be that immature T-ALL cases harboring IL7R-JAK mutations would benefit from the use of JAK inhibitors in T-ALL trials, since it might reduce the toxic side effects from high-dose chemotherapy treatments. The efficacy of ruxolitinib (JAK1/2 inhibitor) was recently demonstrated in murine xenograft models of immature T-ALL.[41] Of interest, both JAK/STAT pathway activation and ruxolitinib efficacy were independent of the presence of JAK/STAT pathway mutations, raising the possibility that the therapeutic potential of ruxolitinib in T-ALL extends beyond those cases with JAK mutations.[41] Further studies are warranted to clarify this issue.

Importantly, we have found statistically significant associations between gene mutations and signaling pathway-groups or T-ALL patient characteristics. These include previously described associations between immature T-ALL and *JAK3* mutations[19], and between TLX1 positive cases and *PHF6* mutations/deletions.[29] Our current study points out that mutations in *PHF6*, *WT1* and PRC2 complex members are particularly prevalent in T-ALL cases harboring mutations of the IL7R-JAK signaling pathway. We speculate that such epigenetic mutations could result in chromatin changes that will affect the accessibility of STAT5 target genes for transcription, cooperating in this way with the activation of the *IL7R-JAK-STAT* signaling pathway. Functional studies are needed to understand the underlying mechanism of this association.

Authorship contribution

CV, CS, MB, EG, SD, SoD, IL, AE, LC, RLS, CM, PV and AVM performed experiments and analyzed data; CS, CJH, EC and JS contributed samples; AV and NG ran the ALL2003 trial; CJH, EC, CV, JC wrote the manuscript; CJH, EC and JC supervised the study.

Conflict of interest disclosure

The authors declare no competing financial interests.

References

1. Weng AP, Weng AP, Ferrando AA, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004;306(5694):269–271.

2. Van Vlierberghe P, Ferrando A. The molecular basis of T cell acute lymphoblastic leukemia. J Clin Invest. 2012;122(10):3398–3406.

3. Aifantis I, Raetz E, Buonamici S. Molecular pathogenesis of T-cell leukaemia and lymphoma. Nat Rev Immunol. 2008;8(5):380–390.

4. Hebert J, Cayuela JM, Berkeley J, Sigaux F. Candidate tumor-suppressor genes MTS1 (p16INK4A) and MTS2 (p15INK4B) display frequent homozygous deletions in primary cells from T- but not from B-cell lineage acute lymphoblastic leukemias. Blood. 1994;84(12):4038–4044.

5. Soulier J, Clappier E, Cayuela J-M, et al. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood. 2005;106(1):274–286.

6. Ferrando AA, Ferrando AA, Neuberg DS, et al. Prognostic importance of TLX1 (HOX11) oncogene expression in adults with T-cell acute lymphoblastic leukaemia. Lancet. 2004;363(9408):535–536.

7. Bernard OA, Bernard OA, Busson-LeConiat M, et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. Leukemia. 2001;15(10):1495–1504.

8. Aplan PD, Aplan PD, Lombardi DP, et al. Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science. 1990;250(4986):1426–1429.

9. Van Vlierberghe P, Van Vlierberghe P, van Grotel M, et al. The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. Blood. 2006;108(10):3520–3529.

10. Speleman F, Cauwelier B, Dastugue N, et al. A new recurrent inversion, inv(7)(p15q34), leads to transcriptional activation of HOXA10 and HOXA11 in a subset of T-cell acute lymphoblastic leukemias. Leukemia. 2005;19(3):358–366.

11. Dadi S, Le Noir S, Payet-Bornet D, et al. TLX Homeodomain Oncogenes Mediate T Cell Maturation Arrest in T-ALL via Interaction with ETS1 and Suppression of TCRα Gene Expression. Cancer Cell. 2012;21(4):563–576.

12. Zenatti PP, Ribeiro D, Li W, et al. Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. Nat Genet. 2011;43(10):932–939.

13. Palomero T, Palomero T, Sulis ML, et al. Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. Nat Med. 2007;13(10):1203–1210.

14. Bar-Eli M, Ahuja H, Foti A, Cline MJ. N-RAS mutations in T-cell acute lymphocytic leukaemia: analysis by direct sequencing detects a novel mutation. Br J Haematol. 1989;72(1):36–39.

15. Degryse S, Degryse S, de Bock CE, et al. JAK3 mutants transform hematopoietic cells through JAK1 activation causing T-cell acute lymphoblastic leukemia in a bone marrow transplant mouse model. Blood. 2014.

16. Van der Meulen J, Sanghvi V, Mavrakis K, et al. The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. Blood. 2015;125(1):13–21.

17. Ntziachristos P, Ntziachristos P, Tsirigos A, et al. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nat Med.

2012;18(2):298–301.

18. De Keersmaecker K, De Keersmaecker K, Atak ZK, et al. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. Nat Genet. 2013;45(2):186–190.

19. Zhang J, Zhang J, Ding L, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012;481(7380):157–163.

20. Holmfeldt L, Holmfeldt L, Wei L, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat Genet. 2013;45(3):242–252.

21. Clappier E, Clappier E, Cuccuini W, et al. The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. Blood. 2007;110(4):1251–1261.

22. Kleppe M, Lahortiga I, Chaar El T, et al. Deletion of the protein tyrosine phosphatase gene PTPN2 in T-cell acute lymphoblastic leukemia. Nat Genet. 2010;42(6):530–535.

23. Patrick K, LaFave LM, Wade R, et al. Outcome for children and young people with Early T-cell precursor acute lymphoblastic leukaemia treated on a contemporary protocol, UKALL 2003. Br J Haematol. 2014;166(3):421–424.

24. Atak ZK, Gianfelici V, Hulselmans G, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. PLoS Genet. 2013;9(12):e1003997.

25. Dagklis A, Pauwels D, Lahortiga I, et al. Hedgehog pathway mutations in T-cell acute lymphoblastic leukemia. Haematologica. 2015;100(3):e102–5.

26. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2014.

27. Bolli N, Manes N, McKerrell T, et al. Characterization of gene mutations and copy number changes in acute myeloid leukemia using a rapid target enrichment protocol. Haematologica. 2015;100(2):214–222.

28. Mullighan CG, Hanna DMT, Zhang J, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011;471(7337):235–239.

29. PHF6 mutations in T-cell acute lymphoblastic leukemia. 2010;42(4):338–342.

30. Tosello V, Mansour MR, Barnes K, et al. WT1 mutations in T-ALL. Blood. 2009;114(5):1038–1045.

31. Bandapalli OR, Bandapalli OR, Schuessele S, et al. The activating STAT5B N642H mutation is a common abnormality in pediatric T-cell acute lymphoblastic leukemia and confers a higher risk of relapse. Haematologica. 2014.

32. Landau DA, Carter SL, Getz G, Wu CJ. Clonal evolution in hematological malignancies and therapeutic implications. Leukemia. 2014;28(1):34–43.

33. Flanagan ME, Blumenkopf TA, Brisette WH, et al. Discovery of CP-690,550: a potent and selective Janus kinase (JAK) inhibitor for the treatment of autoimmune diseases and organ transplant rejection. J Med Chem. 2010;53(24):8468–8484.

34. Quintás-Cardama A, Vaddi K, Liu P, et al. Preclinical characterization of the selective JAK1/2 inhibitor INCB018424: therapeutic implications for the treatment of myeloproliferative neoplasms. Blood. 2010;115(15):3109–3117.

35. LaFave LM, Levine RL. JAK2 the future: therapeutic strategies for JAK-dependent malignancies. Trends Pharmacol Sci. 2012;33(11):574–582.

36. Wilkins BS, Wilkins BS, Radia D, et al. Resolution of bone marrow fibrosis in a patient

receiving JAK1/JAK2 inhibitor treatment with ruxolitinib. Haematologica. 2013;98(12):1872–1876.

37. Hanna DM, Fellowes A, Vedururu R, Mechinaud F, Hansford JR. A unique case of refractory primary mediastinal B-cell lymphoma with JAK3 mutation and the role for targeted therapy. haematologica. 2014;99(9):e156–158.

38. Chase A, Bryant C, Score J, et al. Ruxolitinib as potential targeted therapy for patients with JAK2 rearrangements. Haematologica. 2013;98(3):404–408.

39. Coustan-Smith E, Van Vlierberghe P, Mullighan CG, et al. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. Lancet Oncol. 2009;10(2):147–156.

40. Wood BL, Winter SS, Dunsmore KP, et al. T-Lymphoblastic Leukemia (T-ALL) Shows Excellent Outcome, Lack of Significance of the Early Thymic Precursor (ETP) Immunophenotype, and Validation of the Prognostic Value of End-Induction Minimal Residual Disease (MRD) in Children's Oncology Group (COG) Study AALL0434. Blood. 2014;124(21):1.

41. Maude SL, Dolai S, Delgado-Martin C, et al. Efficacy of JAK/STAT pathway inhibition in murine xenograft models of early T-cell precursor (ETP) acute lymphoblastic leukemia. Blood. 2015;125(11):1759–1767.

Table 1. Targeted sequencing summary of 155 T-ALL samples

	Design A (n=80 T-ALL cases)	Design B (n=75 T-ALL samples)
Design information		
Number of genes in design	149	141
Region size of the design (kbp)	472.006	418.373
Total amplicons in design	23127	19694
Minimum sequencing required (Mbp)	247.7	229.35
Sequence information		
Reads passing filtration (Gb)	**8.7** (3.0 - 25.0)	**4.2** (1.4 - 14.5)
Reads mapped (%)	**98.1** (90.4 - 99.1)	**79.6** (41.4 - 96.2)
Capture efficiency (%)*	**60.8** (55 - 66.3)	**45.5** (23.2 - 56.5)
Mean depth	**2312** (733 - 6580)	**959** (199 - 3586)
≥ 10x coverage (%)†	**97.9** (94.9 - 99.0)	**97.0** (91.2 - 98.6)
≥ 20x coverage (%)†	**97.1** (92.6 - 98.8)	**95.6** (85.6 - 98.2)
≥ 50x coverage (%)†	**94.9** (87.0 - 98.0)	**91.8** (73.1 - 97.2)
≥ 100x coverage (%)†	**91.9** (79.7 - 94.3)	**85.3** (57.5 - 96.2)
≥ 200x coverage (%)†	**86.6** (66.7 - 94.3)	**73.5** (31.1 - 94.4)

Data are presented median (0.25~0.75quartile).

*Capture Efficiency (%), the ratio of the read number that mapped to the targeted region to the total read number.

†≥ 10x, 20x, 50x, 100x, 200x coverage (%), the ratio of target regions covered 10, 20, 50, 100 or 200x to the total number of target regions.

Table 2. Overall clinical, immunophenotypic and molecular cytogenetic characteristics of T-ALL patients with mutations in the IL7R-JAK signaling pathway

	IL7R-JAK signaling pathway			Type of Association
	Wt (n=112)	Mut (n=43)	*P*	
Clinical				
Gender			0.586	
Male	83 (74.1%)	30 (69.8%)		
Female	29 (25.9%)	13 (30.2%)		
Median age (range)	11 (1-63)	12 (2-66)	0.767†	
T-ALL clusters	**Wt**	**Mut**	***P***	
HOXA+ (n=15)	6 (5.4%)	9 (20.9%)	**0.006***	Positive
TLX1+ (n=13)	10 (11.2%)	3 (9.7%)	1.000*	
TLX3+ (n=28)	15 (13.5%)	13 (31%)	**0.013**	Positive
TAL1/LMO2+ (n=54)	48 (43.2%)	6 (14%)	**0.001**	Negative
Immature T-ALL+ (n=14)	4 (4.4%)	10 (25%)	**0.001***	Positive
Del9p21 status	**Wt**	**Mut**	***P***	
Wild-type	27 (24.1%)	13 (30.2%)		
Mutant	85 (75.9%)	30 (69.8%)	0.435	
PHF6 status	**Wt**	**Mut**	***P***	
Wild-type	97 (86.6%)	28 (65.1%)		
Mutant	15 (13.4%)	15 (34.9%)	**0.002**	Positive
PRC2 status	**Wt**	**Mut**	***P***	
Wild-type	97 (86.6%)	31 (72.1%)		
Mutant	15 (13.4%)	12 (27.9%)	**0.033**	Positive
WT1 status	**Wt**	**Mut**	***P***	
Wild-type	102 (91.1%)	30 (69.8%)		
Mutant	10 (8.9%)	13 (30.2%)	**0.001**	Positive

Wt: wild-type; Mut: mutant; P: P value. Median age indicated in years.
Significant P values are indicated in bold; all P values were calculated by using Pearson's χ2 test, unless indicated otherwise: *Fisher's exact test; †Mann-Whitney-U test.
The *HOXA+*, *TLX1+* and *TLX3+* and *TAL1/LMO2+* groups were based on the presence of *HOXA*, *TLX1*, *TLX3*, *TAL1* or *LMO2* rearrangements or by having a *HOXA*, *TLX1*, *TLX3 or TAL1/LMO2* expression signature. The immature T-ALL group was defined immunophenotypically or based on gene expression cluster analysis.

Figure 1. Haloplex target enrichment analyses can be used to identify copy number variations. For *MAGEC3, KDM6A, MYB* and *PTPN2* genes, the normalized coverage against the on-target mapped nucleotides is plotted for all patients on an arbitrary linear scale. In order to find out real copy number variations, for each gene we calculated the first quartile (Q1), third quartile (Q3) and the interquartile range (IQR= Q1 – Q3) from the normalized coverage values. Samples having a normalized coverage for a particular gene greater than Q3+1.5*IQR were considered to have a duplication/amplification in that gene. Conversely, samples having a normalized coverage less than Q1-1.5*IQR, were considered to carry a deletion. The calculated values for each gene are represented in the graphs as dotted lines. For X-chromosome genes, we did those calculations taking along only male or female samples. **(A)** Read depths for *MAGEC3* are consistent with patient gender (except case number TL91), with females showing a ~2 fold higher coverage than males. While no copy number variations were found in *MAGEC3*, there was evidence for deletion of *KDM6A* in case number 4139. **(B)** By using Haloplex we were able to confirm the presence of *PTPN2* deletions in all (6/6) *PTPN2* positive cases (shown as blue dots in the graphs). *MYB* duplications were confirmed in 64% (9/14) of *MYB* positive cases (orange dots). Haloplex analyses predicted four additional cases carrying *MYB* duplication (pink dots), which were false positive results.

Figure 2. Overview of the genetic lesions identified in the most frequently mutated genes in T-ALL. (A) Single nucleotide variants, indels and copy number variations are shown for the 40 genes we found mutated in more than 4% of the cases in our series. Each type of mutation is indicated by different color. Age, gender and T-ALL characteristics are also shown for each patient (bottom table). **(B)** Spectrum of genetic lesions in IL7R-JAK signaling members (*IL7R*, *JAK3*, *JAK1* and *STAT5* genes), *PHF6*, PRC2 complex members (*SUZ12*, *EZH2* and *EED*) and *WT1*. The cases having *PHF6*, PRC2 and *WT1* mutations were more frequently found among *IL7R-JAK* positive cases.

Figure 3. The spectrum of some of the identified genetic lesions closely correlated with patient characteristics and T-ALL subtypes. Associations between the most frequently mutated genes (top 40 genes, ranked from the most frequently targeted gene) and patient characteristics-T-ALL subtypes are shown. Associations with p-values lower than 0.05, 0.01 and 0.001 are indicated with circles of different sizes. Positive associations are shown in orange color, negative associations are shown in blue color. Associations that could not be considered as positive or negative are shown in gray color. A = adult group, P =

pediatric group, F = female group.

Figure 4. Overview of the identified mutations in the IL7R-JAK signaling pathway members. **(A)** Schematic view of the IL7R, JAK3, JAK1, STAT5A and STAT5B proteins and their main protein domains. The mutations found on each of those proteins are shown. **(B)** The variant allele frequencies for *JAK3* mutations are shown in the graph. In eight out of the 23 cases carrying JAK3 mutations (case numbers indicated with asterisks), the M511I mutation was detected together with a second mutation of *JAK3*. **(C)** Proliferation curve of Ba/F3 cells expressing various *JAK3* mutants or the empty vector, in the absence of cytokines. Mutations that did not stimulate proliferation more than the empty vector were considered as not transforming mutants. **(D)** Western blot analysis of whole cell lysates of Ba/F3 cells transformed by the JAK3 mutants V678L and E958K, or empty vector. A protein lysate of Ba/F3 cells transformed by the JAK3 M511I mutant (previously described as transforming) was included as a positive control. Phosphorylation of JAK1 and JAK3 was detected for all JAK3 mutants. JAK3 protein expression was detected with a human specific antibody, not recognizing the endogenous Jak3 expression. The transforming JAK3 mutants were able to phosphorylate STAT5.

Figure 5. *JAK3* **mutations occur predominantly in the major clone at diagnosis.** Variant allele fractions of the mutated genes identified in the 23 cases with *JAK3* mutations. For clarity reasons, only those genes mutated in more than 4% of T-ALL cases are shown. *JAK3* mutations are predominantly present in the major clone when two *JAK3* mutations or *JAK3* and *JAK1* mutations occur together, while a minor clone has the second mutation.
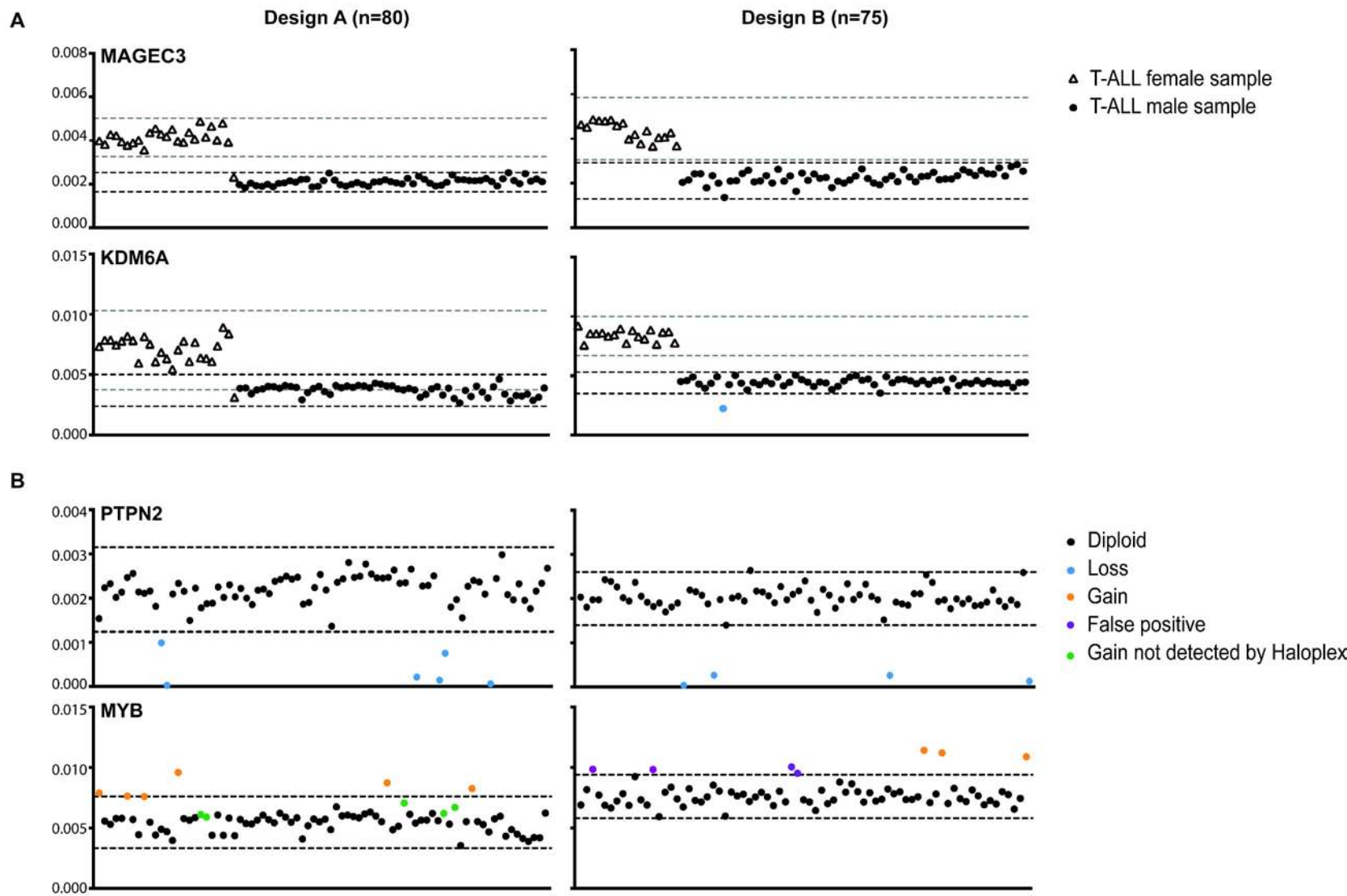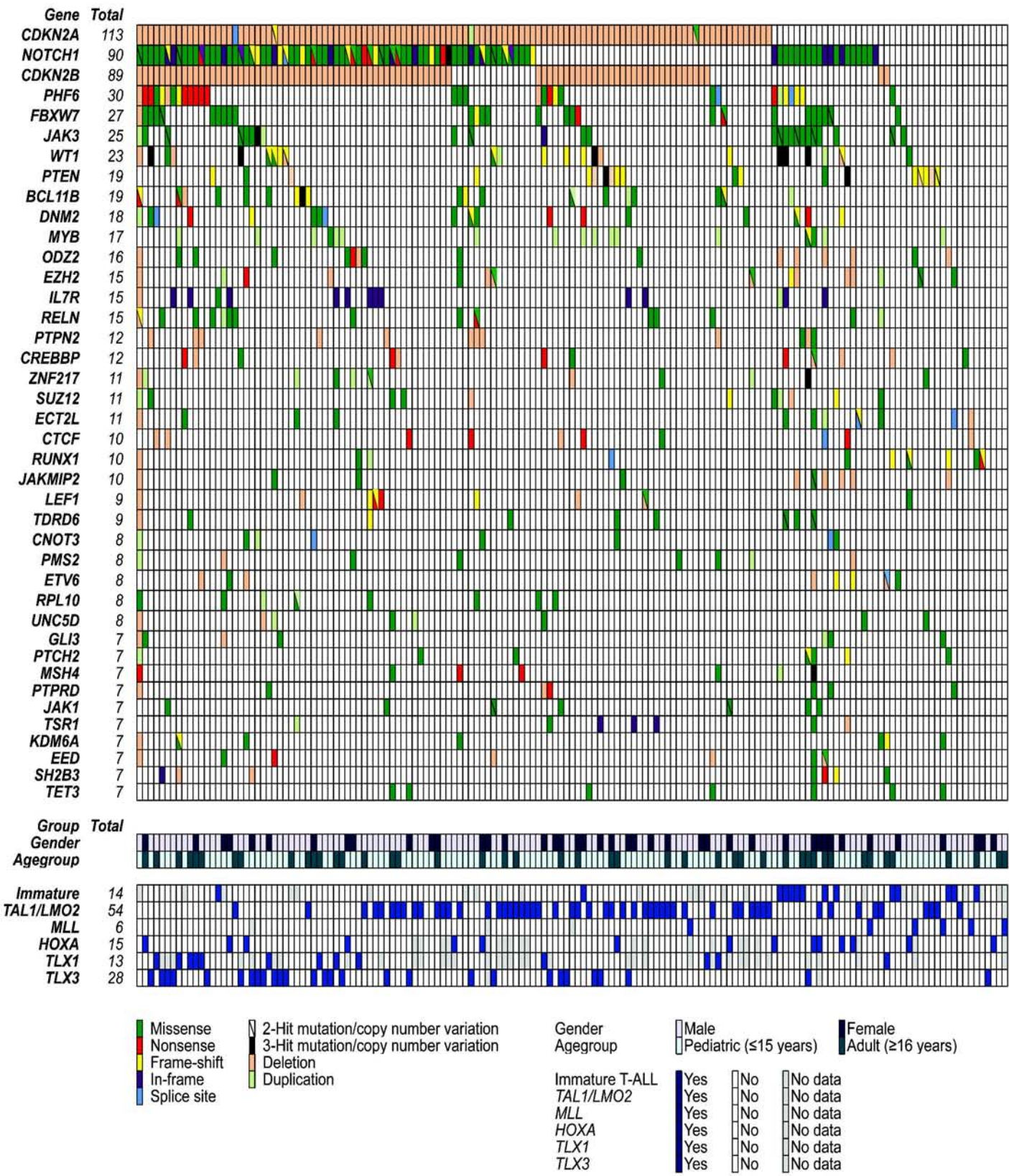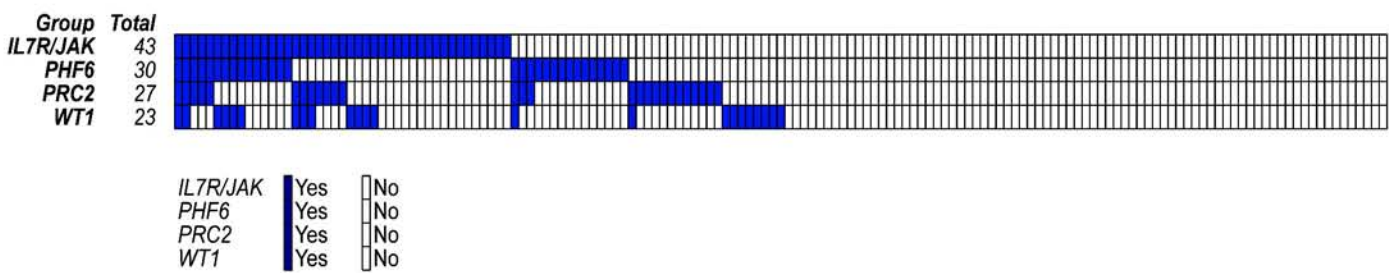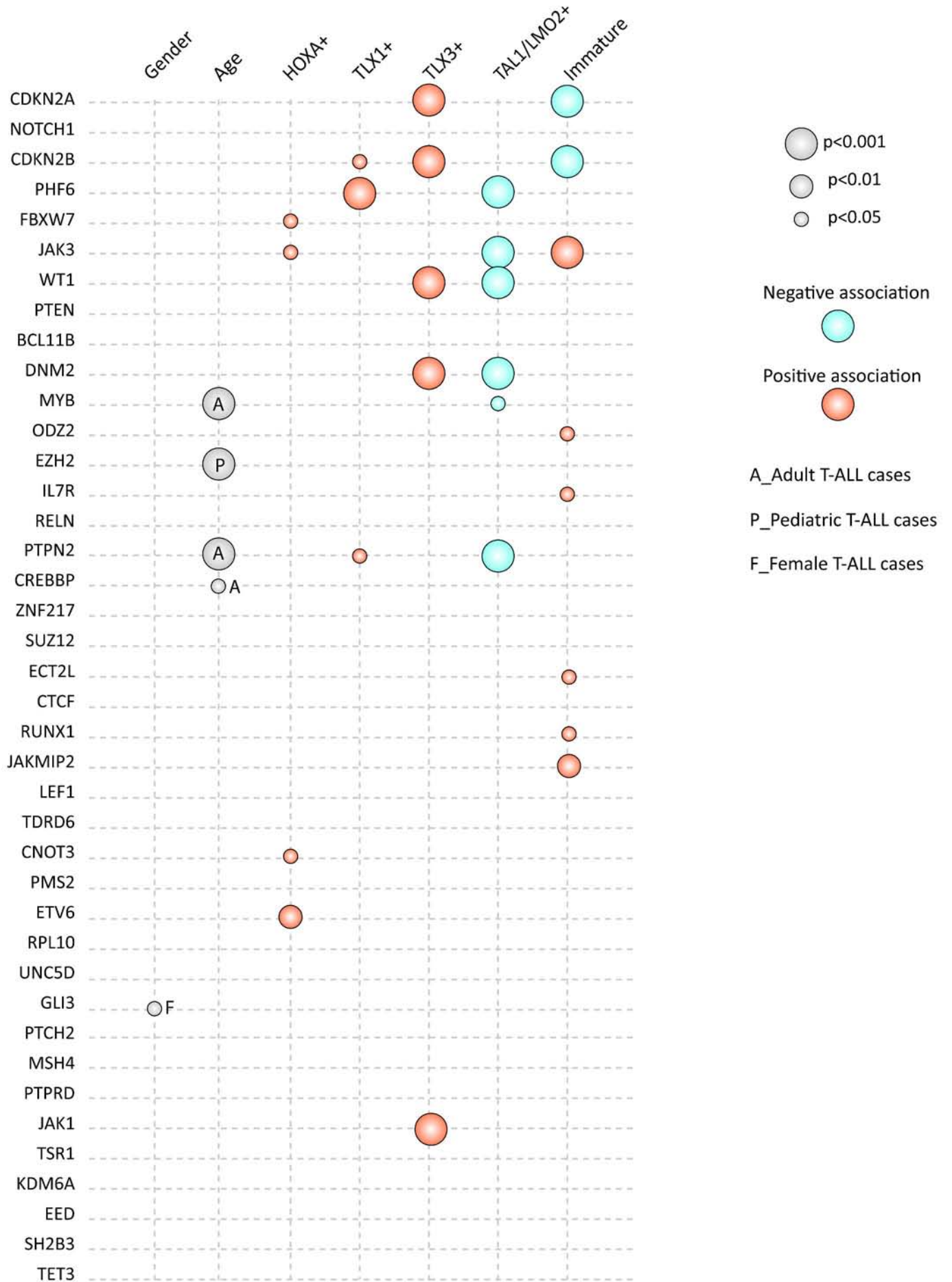
Figure 1

Figure 2
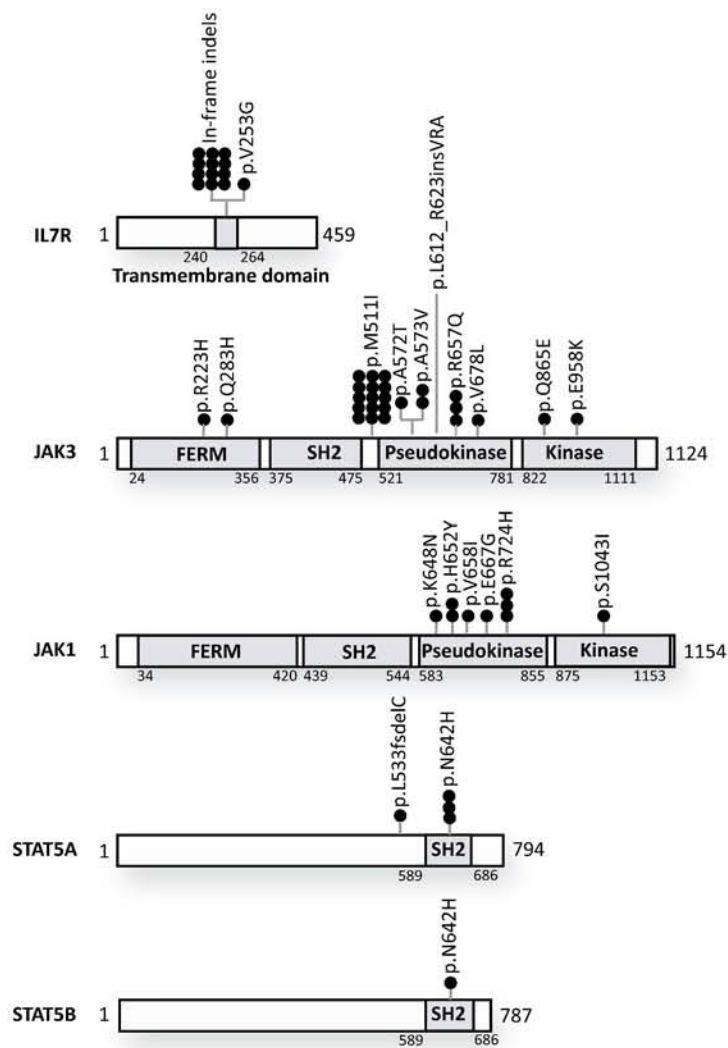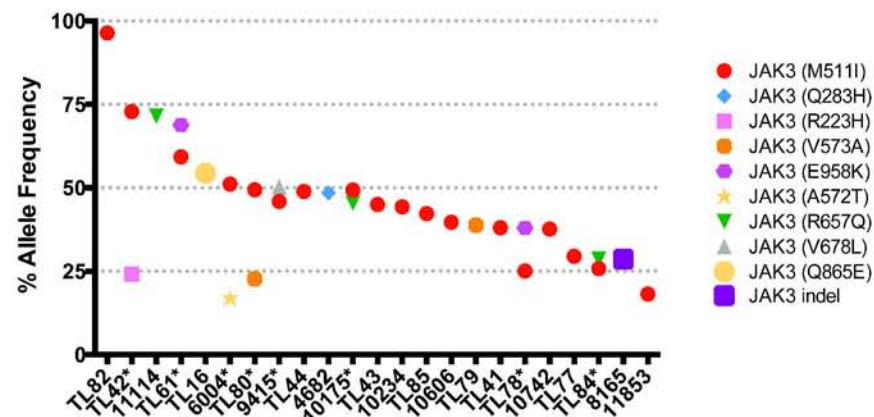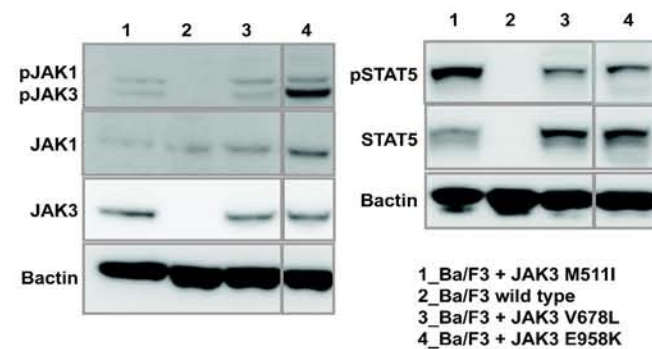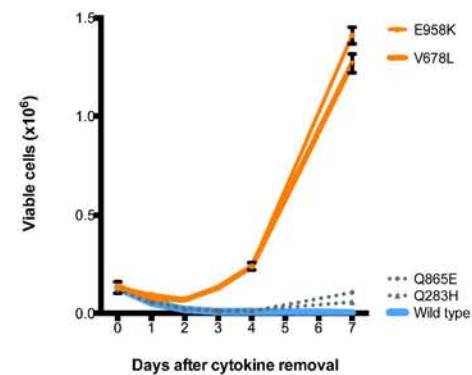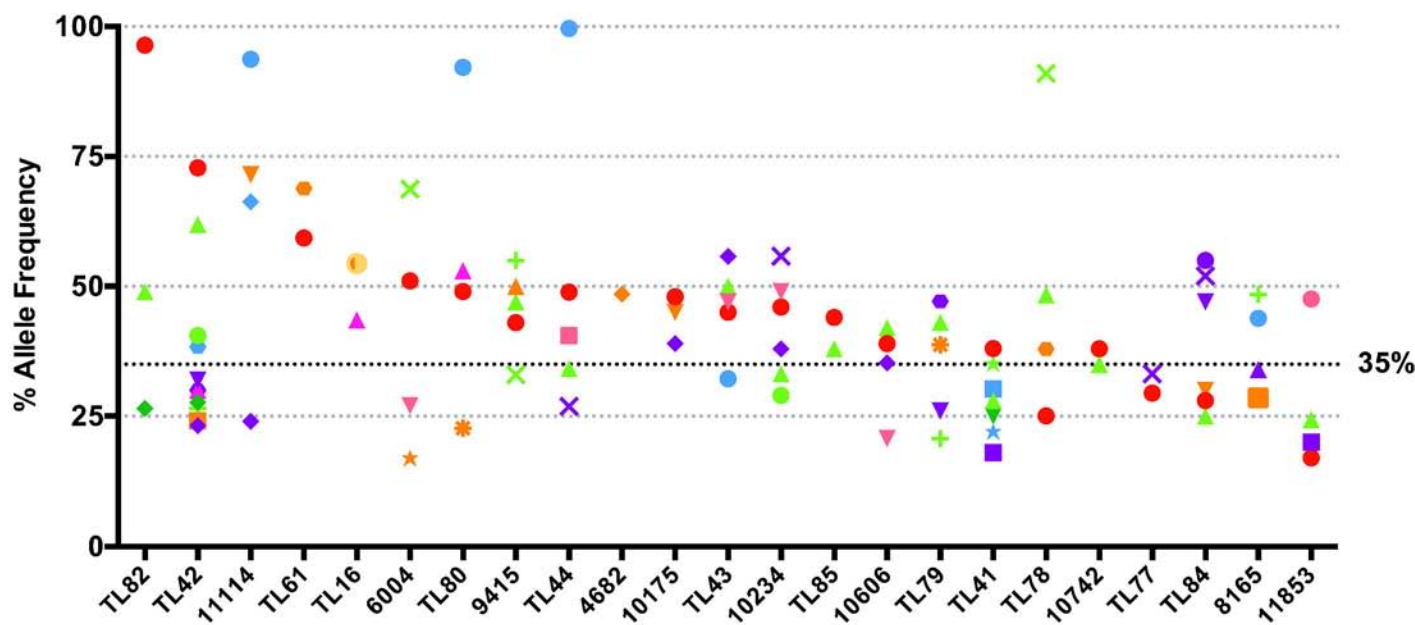
Figure 3

Figure 4

Figure 5

SUPPLEMENTARY INFORMATION

Targeted sequencing identifies association between IL7R-JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia

Carmen Vicente[1,2], Claire Schwab[3], Ellen Geerdens[1,2], Michaël Broux[1,2], Sandrine Degryse[1,2], Sofie Demeyer[1,2], Idoya Lahortiga[1,2], Alannah Elliott[3], Lucy Chilton[3], Roberta La Starza[4], Cristina Mecucci[4], Peter Vandenberghe[1], Nicholas Goulden[5], Ajay Vora[6], Anthony V. Moorman[3], Jean Soulier[7], Christine J. Harrison[3,8], Emmanuelle Clappier[7,8], Jan Cools[1,2,8]

[8]Corresponding authors:
Jan Cools (jan.cools@cme.vib-kuleuven.be),
Christine Harrison (christine.harrison@newcastle.ac.uk),
Emmanuelle Clappier (emmanuelle.clappier@sls.aphp.fr)

This file contains:

The following supplementary files are available as a separate .xls file:

Supplementary tables 1, 2, 3 and 4

Jan Cools 8
Deleted: 4

Jan Cools 8
Deleted: 5

Jan Cools 8
Deleted: 7

1. Supplementary Methods

DNA samples

T-ALL patients (n=155: 111 children, 44 adults) were collected from various institutions. Samples had been previously characterized for oncogene expression and classified into oncogenic subgroups according to combined FISH (fluorescence in situ hybridization) analyses, immunophenotypic data and oncogene expression. Copy number changes of *CDKN2A* (n=95), *MYB* (n=115) and *PTPN2* (n=80) had been previously identified by FISH, MLPA (Multiplex Ligation-dependent Probe Amplification) or arrayCGH. The *HOXA+*, *TLX1+* and *TLX3+* and *TAL1/LMO2+* groups were based on the presence of *HOXA*, *TLX1*, *TLX3*, *TAL1* or *LMO2* rearrangements or by having a *HOXA*, *TLX1*, *TLX3* or *TAL1/LMO2* expression signature. The immature T-ALL group (also known as early thymic or T-cell precursor T-ALL) was defined immunophenotypically or based on gene expression cluster analysis.

Haloplex library preparation and sequencing

Enrichment of the region of interest was performed using the Haloplex Target Enrichment System-Fast Protocol. Briefly, 200 ng of DNA per sample was aliquoted into 8 digestion reactions, each containing 2 restriction enzymes. DNA from the 8 reactions was then pooled and hybridized to Haloplex probes, allowing for purification using magnetic beads. Purified fragments were ligated, amplified and barcoded through 19 cycles of PCR and samples were sequenced on a HiSeq2000 instrument using a 100 bp paired-end protocol. The HiSeq Paired End Cluster Generation Kit was used to generate the clusters and the TruSeq SBS Kit v3 was used for sequencing. Image analysis and base calling was performed using the Illumina RTA software version 1.13.48.

The additional custom filtering criteria refers to the fact that we excluded five variants that were identified in nearly all samples analyzed. Since those variants were not found in dbSNP138, 1000 Genomes or the COSMIC databases, we believe these are false positives of data analysis. These are the variants that have been excluded from the analyses:

Position	Gene	Nucleotide change	Amino acid change
6037058	PMS2	delAA	
103141302	RELN	C>CT	2853G>GS
90645514	IDH2	G>GT	37R>SR
29508805	NF1	T>GT	Splice
103629804	RELN	insGCCGCC	Splice

Validation of identified genomic lesions

We sequenced 158 selected variants (**Supplementary Table 3**) using Sanger sequencing. Analysis of the chromatograms from Sanger sequencing was performed using CLC Main Workbench 6 (CLC Bio).

Western blot analyses

Cells were lysed in cold lysis buffer containing 5 mM Na_3VO_4 and protease inhibitors (Complete tablets, Roche). The proteins were separated on NuPAGE NOVEX Bis-Tris 4%–12% gels (Invitrogen) and transferred to PVDF membranes. Subsequent Western blot analysis was performed using primary antibodies directed against JAK1 (Millipore), JAK3, phospho-STAT5, (Cell Signaling), phospho-JAK1, STAT5 (Santa Cruz Biotech.) and β-actin (Sigma). Anti-phospho-JAK1 antibody was used to detect both phosphorylated JAK1 and JAK3. Western blot detection was performed using secondary antibodies conjugated with horseradish peroxidase (GE Healthcare) and western blot lightning plus-ECL (PerkinElmer).

Statistical analyses

Statistical analyses were carried out using SPSSstatistics v21. Pearson's X^2 Fisher's exact tests were performed to test significance levels for nominal data distributions, whereas the Mann-Whitney U test was used for continuous data.

Validation of the newly identified FLT3 S471C mutation

The FLT3-S471C variant was generated by GenScript, and was cloned into the MSCV-GFP vector. Viral vector production, retroviral transduction and culture of Ba/F3 cells were performed as previously described.[1] Ba/F3 cells expressing the FLT3-ITD (W51 mutation, Kelly et al.[2]) and FLT3-D835Y were previously described.[3] Western blot analyses were performed as indicated above using antibodies directed against FLT3, STAT5 (Santa Cruz Biotech.), anti-phosphoFLT3 (Tyr591), phospho-STAT5 (Cell Signaling) and β-actin (Sigma). Immunoprecipitations was not

performed prior western blot analyses. In order to test the sensitivity of the FLT3-S471C variant to FLT3 inhibition, Ba/F3 cells expressing FLT3-S471C, FLT3-ITD and FLT3-D835Y were treated with 1000nM of the inhibitor AC220 for 30 minutes, and analyzed by western blotting. For dose-response experiments we used 8 different AC220 concentrations (50, 100, 200, 400, 800, 1600, 3200 and 6400 nM) each in triplicate with DMSO as a negative control. The number of viable cells was counted after the treatment 24h with ATP-lite 1step reagent (PerkinElmer). Luminescence was measured with the multilabel plate readers Envision and Victor X4 (PerkinElmer). IC50 values were calculated with GraphPad Prism.

2. Supplementary Results

In our association analyses, the putative *JAK3* passenger mutants (JAK3 Q283H and Q865E mutants that were unable to transform Ba/F3 cells) were initially considered as mutants. We performed the same association analyses excluding those JAK3 passenger mutant cases (scored now as wild type).

Then, the frequency of cases with *IL7R-JAK* mutations in our series is 26.5% (instead 27.7%). All the associations we found with the IL7R-JAK positive cases remained significant as shown in the table below.

	IL7R-JAK signaling pathway			
	Wt (n=114)	Mut (n=41)	*P*	Type of Association
Clinical				
Gender			0.539	
Male	85 (74.6%)	28 (68.3%)		
Female	29 (25.4%)	13 (31.7%)		
Median age (range)	11 (1-63)	12 (2-66)	0.849†	
T-ALL clusters	**Wt**	**Mut**	***P***	
HOXA+ (n=15)	6 (5.3%)	9 (22%)	**0.004‡**	Positive
TLX1+ (n=13)	10 (11.0%)	3 (10.3%)	1.000‡	
TLX3+ (n=28)	15 (13.3%)	13 (32.5%)	**0.007**	Positive
TAL1/LMO2+ (n=54)	49 (43.4%)	5 (12.2%)	**<0.001**	Negative
Immature+ (n=14)	4 (4.3%)	10 (25.6%)	**0.001‡**	Positive
Del9p21 status	**Wt**	**Mut**	***P***	
Wild-type	28 (24.6%)	12 (29.3%)		
Mutant	86 (75.4%)	29 (70.7%)	0.541	
PHF6 status	**Wt**	**Mut**	***P***	
Wild-type	99 (86.8%)	26 (63.4%)		
Mutant	15 (13.2%)	15 (36.6%)	**0.001**	Positive
PRC2 status	**Wt**	**Mut**	***P***	
Wild-type	99 (86.8%)	29 (70.7%)		
Mutant	15 (13.2%)	12 (29.3%)	**0.020**	Positive
WT1 status	**Wt**	**Mut**	***P***	
Wild-type	104 (91.2%)	30 (68.3%)		
Mutant	10 (8.8%)	13 (31.7%)	**0.001**	Positive

3. Supplementary Figures

Figure S1. The newly identified mutation in the extracellular domain of FLT3 (S471C) is a transforming mutation

A

Reference sequence   TCAGACAAGTCTCCC
                     S   D   K   S   P

Consensus mutated   TCAGACAAGTNTCCC
sequence   S   D   K   C   P

Sanger sequence

TCAGACAAGTSTCCC

TCAGACAAGTSTCCC

B

- ••• IL3
- ── no IL3
- ── FLT3 ITD
- ── D835Y
- ···▼·· S471C

C

* pFLT3
* FLT3
*
b-actin

pSTAT5
STAT5
b-actin

D

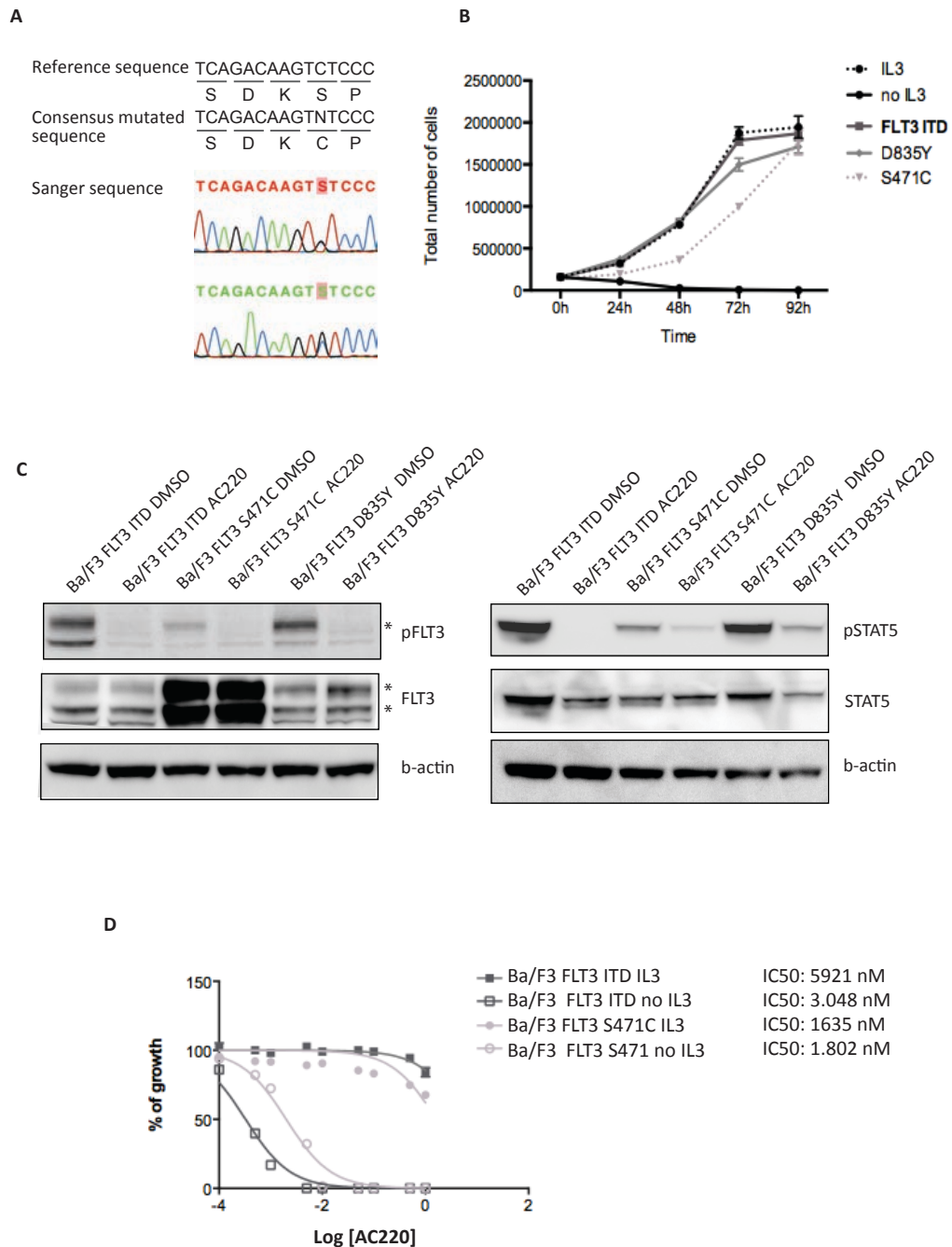|	
■ Ba/F3 FLT3 ITD IL3	IC50: 5921 nM
▫ Ba/F3 FLT3 ITD no IL3	IC50: 3.048 nM
Ba/F3 FLT3 S471C IL3	IC50: 1635 nM
○ Ba/F3 FLT3 S471 no IL3	IC50: 1.802 nM

Figure S1. The newly identified mutation in the extracellular domain of FLT3 (S471C) is a transforming mutation. (A) Sanger sequencing confirmation of the FLT3-S471C variant **(B)** Proliferation curve of Ba/F3 cells expressing FLT3-ITD, FLT3-D835Y and FLT3-S471C variants. Cells expressing FLT3-S471C were able to grow in the absence of IL3. Ba/F3 cells expressing FLT3-ITD, FLT3-D835Y or Ba/F3 wild type supplemented with IL-3 were used as a positive control. Cells expressing Ba/F3 cells wild type without IL-3 were used as a negative control. **(C)** Western blotting showing the phosphorylation and expression of signaling proteins. The expression in Ba/F3 cells of the variant FLT3-S471C resulted in constitutive phosphorylation of FLT3 and STAT5. The level of phosphorylation was lower compared to Ba/F3 cells carrying FLT3-ITD and FLT3-D835Y that correlates with the slower proliferation detected also in comparison with cells expressing FLT3-ITD or FLT3-D835Y. The phosphorylation of FLT3 was reduced for all the Ba/F3 upon treatment with AC220. However, the phosphorylation of STAT5 could be detected after the treatment with AC220 for the cells carrying the S471S and the D835Y mutations. **(D)** We calculated the IC50 values for the AC220 inhibitor treating the cells with FLT3-S741C or FLT3-ITD with increasing concentrations of the inhibitor and measuring the proliferation after 24 hours of inhibition. Both cell lines were highly sensitive to the treatment with AC220 with IC50 values lower than 5 nM.

4. Additional references

1. Degryse S, de Bock CE, Cox L, et al. JAK3 mutants transform hematopoietic cells through JAK1 activation causing T-cell acute lymphoblastic leukemia in a bone marrow transplant mouse model. *Blood*. 2014.

2. Kelly LM, Liu Q, Kutok JL, et al. FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. *Blood*. 2002;99(1):310–318.

3. Pauwels D, Sweron B, Cools J. The N676D and G697R mutations in the kinase domain of FLT3 confer resistance to the inhibitor AC220. *Haematologica*. 2012;97(11):1773–1774.