**SUPPORTING INFORMATION**

**METHODS**

**Patient materials and cell line**

We recruited a cohort of 13 T-PLL cases from collaborating centers for this study. Moreover, we included four previously published T-PLL cases.1 Among these four previously published cases, case #8a and from the 13 T-PLL cases case 8 were obtained from the same patient. Thus, this sample was analysed two times independently. For 13 of the 16 T-PLL cases the presence of the hallmark T-PLL aberrations resulting in TRA-*TCL1A* or TRA-*MTCP1* fusion had been determined by fluorescent in situ hybridization (FISH) and/or conventional cytogenetics. The immunophenotypic characterization by immunohistochemistry and/or flow cytometry (FACS) of the cases was performed by the individual centers submitting the cases. Supporting Information Table 1 gives an overview of materials available and techniques performed for each case.

In addition to these 16 T-PLL cases, we included the cell line SUP-T11 in our analysis, which has been described as a model cell line for studying aberrant *TCL1A* activation.2,3 SUP-T11 was obtained from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Culture. The cell line was derived from the bone marrow of a 74-year-old man diagnosed with a T‑ALL.2,3 SUP-T11 cells are described to express high levels of *TCL1A* and *TCL1B* due to their juxtaposition from 14q32.1 to the regulatory element of the T-cell receptor alpha/delta locus at 14q11.2,3 The cell line was tested negative for mycoplasma contamination and its authenticity was confirmed by STR analysis using StemElite ID System (Promega).

**Conventional cytogenetics and molecular cytogenetics**

If not obtained from previous publications or from the original data from the centers, conventional cytogenetics and FISH analysis were performed as previously described.4 Karyotypes were described according to ISCN guidelines (2016).5 FISH was done on 9/13 cases using frozen tissue and 4/13 cases using FFPE material. Moreover, for the SUP-T11 cell line, FISH was done using fixed cells from cell culture. For the detection of TRA-*TCL1A* or TRA-*MTCP1* rearrangements besides a commercial break-apart for the TRA/D locus (LSI TRA/D Dual Color, Break Apart Rearrangement Probe, Vysis/Abbott, Wiesbaden, Germany) non-commercial probes were used as published.4,6,7 At least 100 nuclei were examined for each probe whenever possible. Digital image acquisition, processing, and evaluation were performed using ISIS digital image analysis version 5.0 (MetaSystems, Altlußheim, Germany).

**Flow Cytometry**

In cases #1 to #5 the DNA was extracted from whole peripheral blood and the immunophenotyping for these was done by the respective centers. In cases #6 to #13 the tumor cells were enriched by FACS sorting. To this end, at the day of sample collection, PBMCs were enriched with Ficoll from whole blood. Mononuclear fraction was viably frozen in 10% DMSO in RPMI and 40% FBS. Banked patient cells were thawn in IMDM + 20% FBS, and washed twice with PBS supplemented with 3% BSA. FACS staining was performed for T-PLL-specific T-cell surface markers labelled with specific fluorophores: anti-CD45- V500, anti-CD19-APC, anti-CD2-FITC, anti-CD3-APCCy7, anti-CD7-V450, anti-CD5-PerCPCy5.5, anti-CD4-PECy7 and anti-CD8-PE. CD45+ CD19- CD3+ cells fraction were was sorted for CD5+ and CD7+. Thus, the T-PLL cells were sorted to be CD45+ CD19- CD3+CD5+CD7+.

The immunophenotype of the SUP-T11 cell line was assessed using FACS sorting for T‑ALL-specific markers using fluorophore labelled antibodies: anti-CD45-PerCP-Cy5-5, anti-CD1a-PE, anti-CD3-APC, anti-TCR alpha/beta-PE, anti-CD5-APC, anti-CD10-PE, anti-CD34-APC, anti-CD8-PE, anti-CD4-PE, anti-CD7-FITC, anti-TCR gamma/delta-FITC. The results of the FACS of SUP-T11 are depicted in the Supporting Information Figure 1.

To obtain T-cells from apparently healthy donors the following strategy was applied. Buffy coats from three different age-matched healthy donors were obtained. Ficoll-hypaque density gradient assay was performed to extract the PBMCs, which were viably frozen in freezing media consisting of FCS supplemented with 10% DMSO. B-cell depletion was performed using Magnetic activated cell sorting (MACS). In brief, the frozen PBMCs were thawed and magnetically labeled with human anti-CD19 coated MicroBeads (Miltenyi Biotec). The cell suspension was loaded onto MACS LS columns placed on MACS Separator according to the manufacturer's instructions. The CD19-depleted (CD19-) fraction was collected from the flow through while the CD19+ cells were retained in the MACS columns. The magnetically retained CD19+ fraction was eluted upon removal of the columns from the MACS separator and used as a control for fluorescence-activated cell sorting (FACS). The CD19-depleted cell fractions were stained with anti-CD45-V500, anti-CD19-FITC, anti-CD3-APC, anti-CD8-PE and anti-CD4-PE-Cy7 antibodies (BD Biosciences).To obtain CD4+ T-cells, the cells were sorted to be CD45+CD19-CD3+CD4+CD8- and for CD8+ T-cells to be CD45+CD19-CD3+CD4-CD8+. The FACS gating strategy for sorting of CD4+ and CD8+ T-cells is shown in Supporting Information Figure 2.

**DNA and RNA extraction**

For the cases #1 to #5, DNA and RNA was extracted from whole peripheral blood using the Gentra Puregene Blood Kit (Qiagen, Hilden, Germany) and RNAeasy kit (Qiagen), respectively. For cases #6 to #13 DNA and RNA was extracted from the sorted T-PLL cells using the Qiagen All Prep kit (Qiagen) according to manufacturer´s protocols. The RNA from the benign CD4+ and CD8+ T-cells each from three separate healthy donors was extracted using the RNAeasy plus kit (Qiagen). For the SUP-T11 cell line the DNA and RNA was extracted using FlexiGene DNA kit (Qiagen) and RNAeasy plus kit (Qiagen), respectively.

To test the quality of the DNA, we performed High Sensitivity DNA Assay on the 2100 TapeStation (Agilent Technologies) and only the samples with a DIN value > 8 were considered for WGS. The quantity of the DNA was measured using Qubit fluorometer together with the Quant-iT dsDNA BR Assay Kit (Life Technologies). The quality of the RNA was assessed using the High Sensivity RNA Assay on the 2200 TapeStation and only the samples with a RIN value > 8 were considered for RNA sequencing. The quantification of the RNA was done using the Quant-iT RNA BR Assay Kit on a Qubit System.

**Whole genome sequencing and analysis**

Whole genome sequencing was performed on 14/17 T-PLL cases (including SUP-T11) with at least 60% tumor cell content, estimated based on FISH results for the T-PLL hallmark alterations and the SUP-T11 cell line. Custom WGS was performed by the service provider Centogene AG. Genomic DNA was fragmented by sonication and adapters were ligated to generate fragments using the TruSeq DNA PCR-free LT Sample preparation Kit (Illumina) for subsequent sequencing on a HiSeqXTen platform (Illumina) to yield an average coverage depth of 30 reads.

Bioinformatic analysis of the WGS data included the following: Read pairs were mapped to the human reference genome (build 37, version hs37d5, (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz>), using bwa-mem (version 0.7.15 with minimum base quality threshold set to zero [-T 0] and remaining settings left at default values).8 Reads were coordinate-sorted using bamsort from the biobambam package (version 0.0.148) and duplicates marked using sambamba (version 0.6.5). Structural variant calling was executed by the SOPHIA algorithm (<https://bitbucket.org/utoprak/sophia>; v35) run with the DKFZ-ODCF SOPHIA workflow <https://github.com/DKFZ-ODCF/SophiaWorkflow> using the no-control mode.

**Detection of TRA-*TCL1A* breakpoint junctions**

The TRA-*TCL1A* translocations and inversions were called using SOPHIA. Using these breakpoint positions the aligned reads of the WGS data of the T-PLL cases were inspected in the IGV viewer (IGV 2.4.9).9 The WGS data were mined for the occurrence of mate reads at the breakpoint junctions which would indicate a TRA-*TCL1A* fusion. By this approach the reads covering the breakpoint junctions were retrieved which were used for further analysis.

**Breakpoint verification by Sanger sequencing**

The TRA-*TCL1A* breakpoints identified by WGS were verified using Sanger sequencing with primers flanking the breakpoints. PCR primers and PCR conditions for Sanger sequencing are listed in Supporting Information Tables 3 and 4. PCR products were sequenced using the Big Dye Terminator v1.1 Cycle Sequencing Kit (Life Technologies). Sequence analysis was performed using an ABI PRISM 3130 Genetic Analyzer.

**RNA sequencing and analysis**

Stranded RNA sequencing of 11 T-PLL cases (including SUP-T11) and the three benign CD4+ T‑cells and CD8+ T-cell samples was performed at the Cologne Center for Genomics. In the first step the removal of ribosomal RNA using biotinylated target-specific oligos combined with Ribo-Zero gold rRNA removal beads from 1 ug total RNA input was done. The Ribo-Zero Human/Mouse/Rat kit was used for depletion of cytoplasmic and mitochondrial rRNA from the samples. Following purification, the RNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. These cDNA fragments had the addition of a single'A' base and subsequent ligation of the adapter. The products were purified and enriched with PCR (20 μl template, 15 cycles) to create the final cDNA library. After library validation and quantification (Agilent 2100 Bioanalyzer), equimolar amounts of library were pooled. Pools were quantified using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System and sequenced on an Illumina NovaSeq S2 Flowcell using the PE100 protocoll. Bioinformatic analysis: RNA was sequenced at a depth of 100 million paired-end reads per case. Raw read data quality was trimmed using trim-galore (version 0.4.4) (<http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/>) with cutadapt version 1.8.3.10 Quality was controlled using fastqc. Paired end reads were 75 bp in length. Reads were aligned with segemehl (version 0.2.0)11 against the hg19 genome. Coverage (bigwig) was normalized to number of mapped reads (rpm). List of normalized fragments per kilobase of transcript per million mapped reads (fpkm) for all the genes were generated which were further used to analyze the gene expression.

**Analysis of breakpoint junctions at TRA/Dand *TCL* gene locus of the translocated/inverted allele**

The site of recombination is directed by recombination signal sequences (RSS) which flank each receptor gene segment and consist of a conserved heptamer (consensus 5´-CACAGTG-3´) and nonamer (consensus 5´-ACAAAAACC-3´) separated by either 12 or 23 nucleotides of more highly variable sequence.12 For the present study, the TRA-*TCL1A* fusion sequence analysis was done using the sequences extracted from the WGS data. RSS sequences at the V(D)J gene junctions were manually retrieved by screening the proximity of the breakpoint regions for occurrence of conserved heptamer and nonamer sequences.

**Analysis of the normal/germline TRA gene VJ recombination**

In order to identify the V and J gene segments of the TRAloci which have been used during the VJ recombination to form the functional T-cell receptor, the expression of those gene segments was analyzed using RNA sequencing data. The expression values of the V and J gene segments forming a TRA were obtained from the RNA sequencing data of 11 T-PLL cases (including SUP-T11 cell line) and only those where the fpkm values were > 1 fpkm were taken into consideration for the analysis. In the next step, the WGS data were mined for the occurrence of mate reads at those which would indicate a fusion of those V and J expressed gene segments. By this approach we could identify the gene segments which were recombined to form the exon of the TRA gene. In a next step, we used the IMGT/V-QUEST sequence alignment software to check the recombined sequences for presence of single nucleotide variants as well as insertion/deletions and to analyze if the rearrangement yielded a functional sequence (Supporting Information Table 7).

**RESULTS**

**Analysis of the non-RSS associated breakpoints** **using the RSSsite web interface to identify and score cryptic RSS.**

To detect the presence of cryptic RSS in the cases with the non-RSS-associated breakpoints located between TRAV41 and TRADV2, we used the RSS database CNR-ITB for the analysis (<https://www.itb.cnr.it/rss/index.html>). Using this database we analyzed the chromosomal region between TRAV41 and TRADV2, i.e., chr14:22852856-22855418 (GENCODE version 19). In the current version of the system, pass/fail RIC thresholds are set accordingly to the work of Cowell and colleagues.13,14 According to the database RSS-23 are scored as functional (pass) with RIC score ≥ −58.45. Based on the RIC score, we detected a functional (pass) cryptic RSS-23 at the breakpoint in cases #1 and #2 and another cryptic RSS-23 located 85 bps from the breakpoint in case #4 based on the RIC score.

**Analysis of T-cell receptor gene expression**

In order to identify if the T-PLL cases produced a functional VJ rearranged T-cell alpha receptor, we analyzed the VJ alpha and C alpha transcript expression. Based on the RNA-seq and WGS data we were able to identify the exact transcripts which were recombined to produce the T-cell receptor alpha. The gene segments encoding for the T-cell receptor alpha spread over a genetic region in 14q11 covering 1500 Kb. This region includes around 100 variable alpha (VA) and 60 joining alpha (JA) gene segments as well as a single constant alpha (CA) gene segment. To constitute a functional T-cell receptor alpha chain, one of the V and one of the J gene segments are rearranged during T-cell development to form a single exon.15,16 We could show a differential usage of the VA and JA gene segments in the T-cell receptor alpha genes on the unaffected allele and the translocated/inverted allele 14q11. The TRA gene of the unaffected allele were formed by V segments distal to J gene segments and J gene segments distal to V gene segments. In contrast, the TRA gene of the translocated/inverted allele is formed by V segments proximal to J gene segments and J gene segments proximal to V gene segments (Supporting Information Tables 6 and 7). In the next step, the fusion sequences of the VJ segments of the unaffected allele were analyzed for the presence of mutations. No mutations were detected in the sequences of all 11 T-PLL cases and the rearranged gene segments were found to be in frame using the IMGT/V-QUEST sequence alignment software (Supporting Information Table 7). This data indicated that the unaffected allele in 14q11 encodes for a functional T-cell receptor which is expressed in those T-PLL cells.

**REFERENCES**

1. Schrader A, Crispatzu G, Oberbeck S, et al. Actionable perturbations of damage responses by TCL1/ATM and epigenetic lesions form the basis of T-PLL. *Nat Commun*. 2018;9(1):697.

2. Virgilio L, Narducci MG, Isobe M, et al. Identification of the TCL1 gene involved in T-cell malignancies. *Proc Natl Acad Sci USA*. 1994;91(26):12530-12534.

3. Sugimoto K-J, Shimada A, Wakabayashi M, et al. T-cell lymphoblastic leukemia/lymphoma with t(7;14)(p15;q32) [TCRγ-TCL1A translocation]: a case report and a review of the literature. *Int J Clin Exp Pathol*. 2014;7(5):2615-2623.

4. Schlegelberger B, Zwingers T, Harder L, et al. Clinicopathogenetic significance of chromosomal abnormalities in patients with blastic peripheral B-cell lymphoma. Kiel-Wien-Lymphoma Study Group. *Blood*. 1999;94(9):3114-3120.

5. International Standing Committee on Human Cytogenomic Nomenclature, McGowan-Jordan J, Simons A, Schmid M, eds. *ISCN: An International System for Human Cytogenomic Nomenclature (2016)*. Basel ; New York: Karger; 2016.

6. López C, Bergmann AK, Paul U, et al. Genes encoding members of the JAK-STAT pathway or epigenetic regulators are recurrently mutated in T-cell prolymphocytic leukaemia. *British Journal of Haematology*. 2016;173(2):265-273.

7. Ventura RA, Martin-Subero JI, Jones M, et al. FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. *J Mol Diagn*. 2006;8(2):141-151.

8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.

9. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26.

10. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j*. 2011;17(1):10.

11. Hoffmann S, Otto C, Doose G, et al. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol*. 2014;15(2):R34.

12. Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol*. 2000;18:495-527.

13. Cowell LG, Davila M, Kepler TB, Kelsoe G. Identification and utilization of arbitrary correlations in models of recombination signal sequences. *Genome Biol*. 2002;3(12):RESEARCH0072.

14. Cowell LG, Davila M, Ramsden D, Kelsoe G. Computational tools for understanding sequence variability in recombination signals. *Immunol Rev*. 2004;200:57-69.

15. Thuderoz F, Simonet M-A, Hansen O, et al. Numerical modelling of the V-J combinations of the T cell receptor TRA/TRD locus. *PLoS Comput Biol*. 2010;6(2):e1000682.

16. Weaver D, Boubnov N, Wills Z, Hall K, Staunton J. V(D)J recombination: double-strand break repair gene products used in the joining mechanism. *Ann N Y Acad Sci*. 1995;764:99-111.

**FIGURE LEGENDS**

**Supporting Information Figure 1** Examples of FACS results of selected markers of the SUP‑T11 cell line. SUP-T11 cells expressed CD2 (A), CD3 (B) and CD7 (C) and TRA/B (J) T-cell surface markers and were negative for TdT (D), CD34 (E), CD1a (F), CD4 (G), CD8 (H) and TRG/D (I). The percentage of positive SUP-T11 cells for every marker is indicated above the single plots. The results suggest the diagnosis of T-PLL to be more likely than that of a T-ALL.

**Supporting Information Figure 2** Examples of FACS results of the sorting strategy of the benign CD4+ and CD8+ T-cells. First, the cell population being CD19- and CD45+ was selected (A). This cell population was sorted for being CD3 positive (B). In a last step, these CD19-CD45+CD3+ T‑cells were further sorted for CD4+ T-cells and CD8+T-cells (C).

**Supporting Information Figure 3** Location of the breakpoints of the hallmark alterations within the TRA/D locus (14q11). Within the UCSC genome browser (GENCODE version 19), track bars indicate the breakpoint locations in the TRA/D gene locus in 17 T-PLL cases (including SUP-T11) with inv(14) (red bars), t(14;14) (blue bars), or t(X;14) (green bar).

**Supporting Information Figure 4** Schematic representation of the breakpoints of T-PLL cases carrying inv(14) and t(14;14). The arrows in red indicate the breakpoints of inv(14) cases which are located centromeric of *TCL1A*. The arrows in blue indicate the breakpoints of t(14;14) cases which are located telomeric of *TCL1A*.The arrows with asterisk (\*) indicate regions where more than one breakpoint was found.

**Supporting Information Figure 5** *TCL1A* expression in 10 T-PLL cases with inv(14)/t(14;14) and one T-PLL case with t(X;14) and SUP-T11 cell line in comparison to benign CD4+ and CD8+ T‑cells. Depicted is the strand-specific expression of the negative strand of the *TCL1A* gene locus. In the top of the figure the location of the exons and introns of the *TCL1A* gene are indicated based on the UCSC gene annotation. Below are the Sashimi plots depicting the RNA expression levels of the T-PLL as well as benign T-cells. Of note is that for each plot a different scale has been used. The benign T-cells and case #13 carrying t(X;14) do not express *TCL1A* whereas the T-PLL cases with inv(14)/t(14;14) display different levels of *TCL1A* expression.

**Supporting Information Figure 6** Expression of *MTCP1* and *CMC4* in the T-PLL cases carrying the translocation t(X;14), inv(14) and t(14;14). The upper part of the figure shows the breakpoints in the Xq28 locus for the T-PLL case 13 with t(X;14) as well as the UCSC gene annotations The breakpoints are located downstream of the *MTCP1* gene locus, with the first breakpoint being located within the *BRCC3* gene. The lower part of the figure shows the magnification of that region focusing on the *MTCP1* and *CMC4* gene loci as indicated by the UCSC gene annotation. Below are the plots of the negative strands of the RNA-seq data of the benign T-cells as well as of the T-PLL case 13 harboring the t(X;14) and the remaining 10 cases including SUP-T11 carrying inv(14) and t(14;14). These plots show a low expression or no expression at all of *MTCP1* (expression 5.5 fpkm) and a low expression of *CMC4* (mean expression 4.5 fpkm) in the benign T-cells and the T-PLL cases carrying inv(14) and t(14;14). The T-PLL case with t(X;14) has in comparison to the benign T-cells a higher *MTCP1* (43.6 fpkm) as well as *CMC4* (38.1 fpkm) expression.