**Patients and tissue samples**

DNA was isolated from frozen tissue with >80% tumor cells by phenol-chloroform extraction and ethanol precipitation. RNA was extracted using PerfectPure (5 PRIME). Paraffin-based studies were performed on blocks or tissue microarrays described previously.[1-3] Tumors were classified by World Health Organization criteria.[4] The study was approved by the Mayo Clinic Institutional Review Board.

**Mate-pair library construction and sequencing**

A mate-pair library was prepared using the Mate-pair Library Prep Kit (Illumina), following the manufacturer's instructions. Briefly, 10 μg genomic DNA was fragmented using nebulization to ~2-5 kb. The ends were repaired and a 3' adenine added. Fragments were 3'-biotinylated and separated on an agarose gel. DNA fragments of ~5-5.5 kb were excised, purified, and blunt-end ligated using circularization ligase (Illumina). Non-circularized fragments were eliminated with DNA exonuclease. Circularized DNA was fragmented on a Covaris E210 to ~300-600 bp. Fragments were immobilized on M-280 streptavidin beads (Dynal). The ends were repaired and a 3' adenine and paired end DNA adaptors (Illumina) were added. Fragments were enriched by 18 cycles of PCR using primers PE 1.0 and PE 2.0 (Illumina). The supernatant was run on an agarose gel and fragments ~400-600 bp were excised, purified, and analyzed on an Agilent Bioanalyzer DNA 1000 chip. The library was loaded onto one lane of an Illumina flow cell at a concentration of 9 pM generating an average of 215,000 clusters/tile using the Illumina cluster station and paired end cluster kit v4. The flow cell

was sequenced as a 76x2 paired-end read on an Illumina GAIIx using SBS sequencing kit v4 and SCS v2.5. Base-calling was performed using Illumina Pipeline v1.5.

**Bioinformatic analysis**

Mate-pair sequencing data were mapped to the human genome (NCBI GRCh37 assembly, accessed through ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/Assembled_chromosomes/) using our previously published binary indexing algorithm.[5] Briefly, both the reference genome and sequencing data were assigned binary representations, wherein each nucleotide was identified by two binary digits, as follows:

| Nucleotide | G | A | C | T |
|---|---|---|---|---|
| 1st binary bit | 1 | 1 | 0 | 0 |
| 2nd binary bit | 1 | 0 | 1 | 0 |

Each 32 consecutive nucleotides then were converted into two 32-bit binary numbers, referred to as the base and check arrays, as in the following example:

| 32-nucleotide sequence | GAGCCCCAAA TGCCTTCTTT GGTTTTCTTA GA |
|---|---|
| 1st 32-bit binary number: base array | 1110000111 0100000000 1100000001 01 |
| 2nd 32-bit binary number: check array | 1011111000 0111001000 1100001000 10 |

This binary representation converts a 3.2 billion nucleotide genome to ~100 million, non-overlapping 32-bit numbers. At 8 bytes per 32 nucleotides, the memory requirement for the entire genome is 800 megabytes (1.6 gigabytes including the reverse complement).

The reference genome then was stored in RAM in sequential segments. For each segment, an index table was created with one column containing all possible 27-bit binary numbers (0-111111111111111111111111111) and the other column containing the first position in the genome where the corresponding sequence is found. A second "look-up" table, corresponding to the number of bases in the segment stored in RAM, recorded successive positions in the genome for each 27-bit number, identified all possible 32-bit sequences, and checked for overlapping numbers. The length of the look-up table is the length of the genome segment (number of positions) being stored into RAM.

Each 27-bit number from the sequencing data then was mapped by identifying the corresponding matching position(s) in the index and look-up tables. Finally, the binary representation of the sequence data was compared to that of the genome using exclusive OR operations to identify mismatches, of which up to 2 were permitted. Sequences with exactly one mapped read were written to an output file for further analysis. The following example shows the binary representation of a segment of the genome and its corresponding index table:

**Segment of the Genome**

```
   ↓ Position 1104701                ↓ Position 1104733

...111000011101000110000000101 01001 111000011101000110000000110...

   ←          27 bits        →      ←          27 bits        →
```

**Index table**

| *All possible 27-bit numbers* | *Genome Position* |
|---|---|
| 000000000000000000000000000 | 0 |
| 000000000000000000000000001 | 0 |
| 000000000000000000000000010 | 0 |
| 000000000000000000000000011 | 0 |
| . | . |
| . | . |
| . | . |
| 111000011101000110000000101 | 1104701 |
| 111000011101000110000000110 | 1104733 |
| . | . |
| . | . |
| . | . |
| 111111111111111111111111111 | 0 |

Fragments pointing to candidate translocations were found by selecting all fragments whose mate-pair end tags mapped uniquely to two distant loci (defined as different chromosomes or separated by >25,000 bp on the same chromosome). Random ligations of two different fragments could occur at the mate pair step when fragments are circularized or at the adaptor ligation step before the library construction. However, such locus pairs are likely to be represented by only a single mate-pair, whereas the number of non-identical mate-pairs representing a translocation is expected to be approximately equal to the bridged genomic coverage. To find possible translocations supported by multiple fragments, first we devised a search algorithm in R. Each fragment representing

a possible translocation was grouped with other fragments with similar mapping results: these fragments were referred to as associates. Two fragments were confirmed to be associates when the sum of the pair-wise distances between corresponding tags was <15,000 bp. Many of these associates were false positives, and appeared when one tag mapped to one gene and the other tag mapped to a homologous gene at a distant genomic locus. Therefore, one representative fragment from groups of associates with 5 or more fragments was re-mapped to the reference genome using BLAT (accessed through http://genome.ucsc.edu/). BLAT results were used to reject candidate translocations if: (1) there were multiple matches for one or both tags of a fragment with "Score" >50 and "Identity" >95% (this accounted for the majority of rejections); or (2) both tags of the fragment mapped close to each other (~5 kb) on the same chromosome with good scores. Candidate translocations that were not rejected were advanced to the next stage of validation by experimental means.

**Polymerase chain reaction (PCR) and conventional sequencing**

PCR was performed on genomic DNA using combinations of primers listed in Supplemental Table 1 using HotStar Taq DNA polymerase kit (Qiagen, Valencia, CA) on a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA) with the following parameters: 95°Cx15 min, 30 cycles of (95°Cx30 sec; 60°Cx30 sec; 72°Cx5 min), 72°Cx12 min, and 4°Cx∞. Reaction products were visualized on a 0.8% agarose gel and reactions with single bands of expected sizes were sequenced bidirectionally using the appropriate primers on a 3730xl DNA Analyzer (Applied Biosystems). Sequences were aligned to the human genome (February 2009 build; GRCh37/hg19) using BLAT.

**Fluorescence *in situ* hybridization (FISH)**

Human bacterial artificial chromosome (BAC) clones flanking the relevant areas of 6p25.3 and 7q32.3 were identified using the University of California Santa Cruz Genome Browser (accessed through http://www.genome.ucsc.edu) and obtained from Invitrogen (Carlsbad, CA). BAC DNA was isolated using the Qiagen Plasmid Maxi Kit and directly labeled using the Vysis® Nick Translation Kit (Abbott Molecular, Des Plaines, IL) and either Vysis SpectrumGreen d-UTP or Vysis SpectrumOrange d-UTP (Abbott Molecular). Each individual clone was tested for accuracy and specificity by hybridization to both normal metaphases and sections of paraffin-embedded normal tissue. FISH probes for 6p25.3 are shown in supplemental Figure 1. For the 7q32.3 BAP probe, BAC RP11-244M6 was labeled with SpectrumOrange and BAC RP11-36B6 was labeled with SpectrumGreen. Both BACs were labeled with SpectrumGreen for use in the t(6;7)(p25.3;q32.3) dual fusion (D-) FISH probe.

Five-micron paraffin sections were deparaffinized in CitriSolv, dehydrated in 100% ethanol, and air dried. Slides then were pre-treated in 1 M Tris/0.5 M EDTA followed by NaCl protease treatment to remove proteins and non-DNA cellular components. Slides were dehydrated in an ethanol series and air dried. The DNA probe was applied and slides were coverslipped and sealed with a continuous bead of rubber cement. The slide and probe were co-denatured and hybridized overnight. Slides then were washed and counterstained with 6-diamidino-2-phenylindole dihydrochloride. FISH signals were visualized using a fluorescent microscope (DM5000B, Leica, Wetzlar,

Germany) and pictures were captured at a magnification of 630x using a MOD camera (Applied Imaging, Santa Clara, CA) and CytoVysion software (Applied Imaging).

**Real-time Quantitative PCR (RQ-PCR)**

RQ-PCR was performed in duplicate on a 7900HT Fast Real-Time PCR System (Applied Biosystems) on a cohort of 31 ALCLs (9 cutaneous ALK-negative, 16 systemic ALK-negative, 6 systemic ALK-positive) occurring in 29 patients.

For gene expression, cDNA was prepared from total RNA using the High Capacity Reverse Transcription Kit (Applied Biosystems). Primer/probe sets were obtained from Applied Biosystems (5' *DUSP22*, Hs00169616_m1; 3' *DUSP22*, Hs00414885_m1; *IRF4*, Hs01056534_m1) or Integrated DNA Technologies (Coralville, IA; *GAPDH*: forward, GAAGGTGAAGGTCGGAGTC; reverse, GAAGATGGTGATGGGATTTC; probe, /56-FAM/CAAGCTTCCCGTTCTCAGCC/3IAbRQSp/). Expression values of unknowns were derived from standard curves generated from serial dilutions of a reference standard for each gene. Expression levels of genes of interest were normalized to expression of *GAPDH* and shown as fold change relative to the mean value from non-translocated cases.

For miRNA expression, miRNA-specific cDNAs were prepared from total RNA using the TaqMan microRNA Reverse Transcription Kit (Applied Biosystems) and expression levels were quantified using assay kits from Applied Biosystems according to the manufacturer's instructions. Expression levels of miRNAs of interest were derived

from comparison to U6 expression using the ΔΔCT method and shown as fold change relative to the mean value from non-translocated cases.

**Statistics**

Gene and miRNA expression values were compared between groups using Student's *t*-test; p values <.05 were considered significant.

**Table S1. PCR primers used for validation and sequencing of t(6;7)(p25.3;q32.3)**

**Validation of der(6)**

| Reaction | Chromosome 6 Primer | Chromosome 7 Primer |
|---|---|---|
| 1 | TATTAGTGCTGGACAATGCTG | GACCTTTCCCCCTACAGAC |
| 2 | " | GGGAACCTTGTTCACATGATA |
| 3 | " | AGGCTGAGTCTCTACTTCGTG |
| 4 | " | TGATGTAGCGTGAGGATCTTA |
| 5 | " | TCAGAATGAAAGCCATACATC |
| 6 | " | AGCTTTCTGTTCACGTGTCT |

**Validation of der(7)**

| Reaction | Chromosome 6 Primer | Chromosome 7 Primer |
|---|---|---|
| 1 | GCTCCCAAACTACGCTCTT | CATCTCTGTCTCAGCGGAACT |
| 2 | " | GGTACCATGAGCACGGATG |
| 3 | " | GCCAAATCACTGTCTTACCAG |

**Table S2. Results of 7q32.3 breakapart FISH in 142 PTCLs without translocations involving 6p25.3**

| | 7q32.3 FISH Result | | |
|---|---|---|---|
| **PTCL Type** | **Rearranged** | **Not Rearranged** | **Hybridization Failure** |
| ALCL, ALK negative (systemic) | 0 | 13 | 7 |
| ALCL, ALK negative (cutaneous) | 0 | 3 | 1 |
| ALCL, ALK positive | 0 | 11 | 3 |
| PTCL, not otherwise specified* | 0 | 42 | 15 |
| Angioimmunoblastic TCL | 0 | 19 | 3 |
| Extranodal NKTL | 0 | 7 | 2 |
| Mycosis fungoides | 0 | 4 | 1 |
| T-cell LGL | 0 | 4 | 0 |
| Enteropathy-type TCL | 0 | 2 | 0 |
| Hepatosplenic TCL | 0 | 1 | 1 |
| Subcutaneous panniculitis-like TCL | 0 | 1 | 1 |
| Lymphomatoid papulosis† | 0 | 1 | 0 |
| Total | 0 | 108 | 34 |

*CD30: positive in 6, negative in 49, not tested in 2.

†One case of lymphomatoid papulosis, a cutaneous T-cell lymphoproliferative disorder with similarities to cutaneous ALCL,[6] was examined for a 7q32.3 rearrangement in the present study. We previously examined a series of 32 cases of lymphomatoid papulosis for 6p25.3 rearrangements and found one positive case (3%);[7] however the partner locus in that case is unknown.

Abbreviations: FISH, fluorescence *in situ* hybridization; PTCL, peripheral T-cell lymphoma; ALCL, anaplastic large cell lymphoma; ALK, anaplastic lymphoma kinase; TCL, T-cell lymphoma; NKTL, natural killer-/T-cell lymphoma, nasal type; LGL, large granular lymphocytic leukemia.

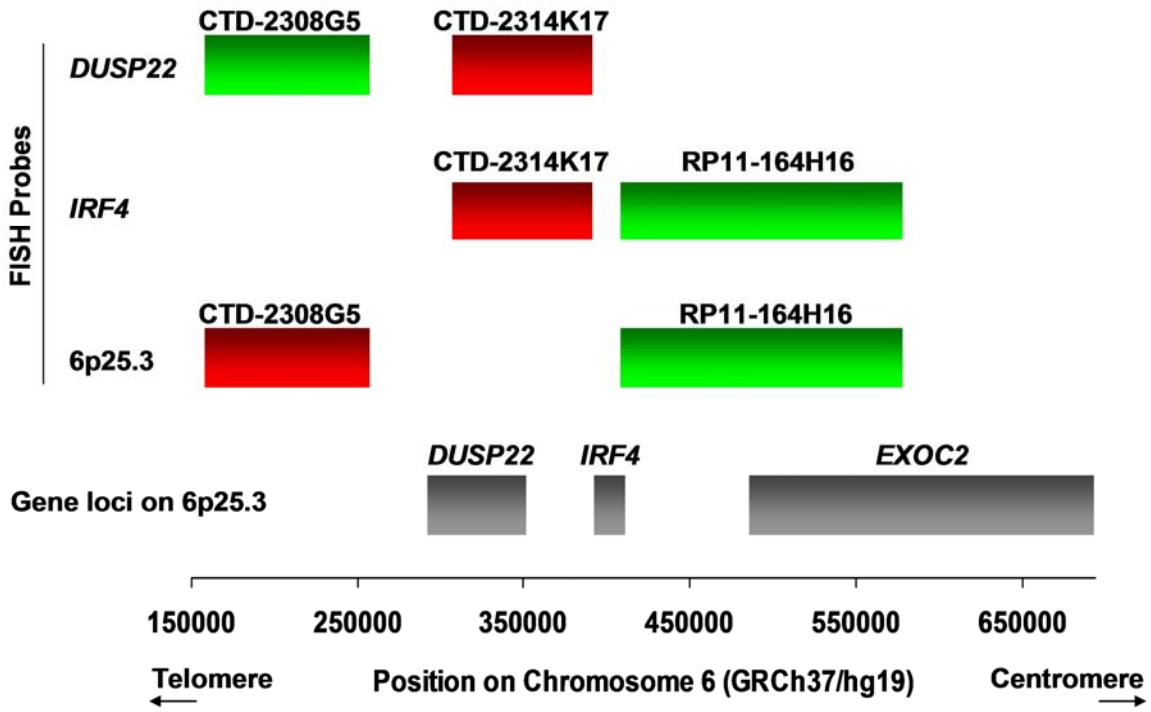


**Figure S1. FISH probes for 6p25.3.** Schematic diagram indicating the bacterial artificial chromosomes (BACs) utilized to construct the FISH probes, their alignment to chromosome 6, and their relationship to genes in this region. Red and green colors indicate labeling with SpectrumOrange and SpectrumGreen, respectively for breakapart probes (see supplemental Methods). For t(6;7)(p25.3;q32.3) dual-fusion FISH, both BACs of the 6p25.3 probe were labeled with SpectrumOrange.

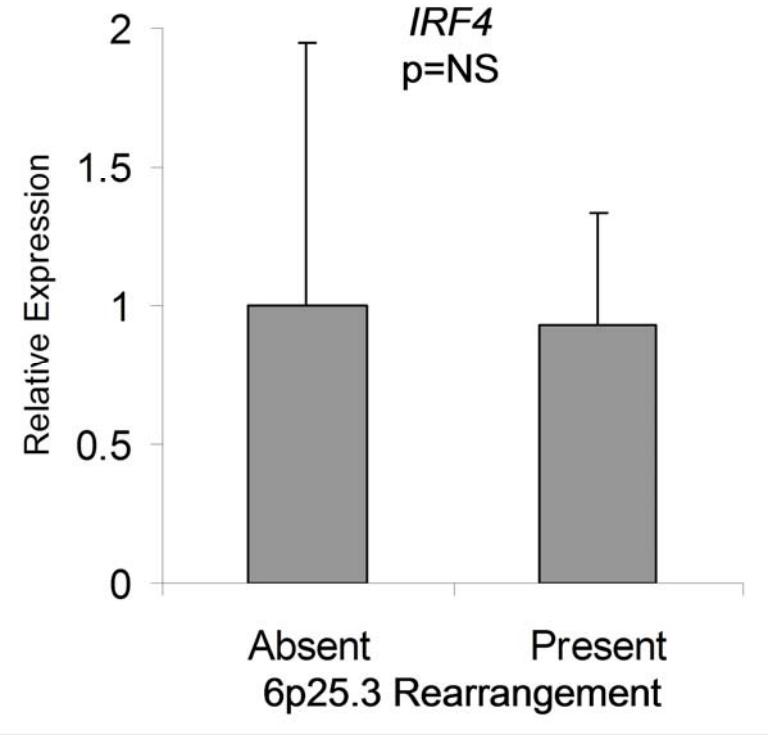


**Figure S2. Expression of *IRF4*.** *IRF4* expression in ALCLs without and with 6p25.3 rearrangements (real-time quantitative PCR, shown as expression relative to the mean value of the non-translocated cases - means ±SDs: 1.00±0.95 vs. 0.93±0.40).

# REFERENCES


1. Feldman AL, Law M, Grogg KL, et al. Incidence of TCR and TCL1 gene translocations and isochromosome 7q in peripheral T-cell lymphomas using fluorescence in situ hybridization. Am J Clin Pathol. 2008;130:178-185.

2. Feldman AL, Sun DX, Law ME, et al. Overexpression of Syk tyrosine kinase in peripheral T-cell lymphomas. Leukemia. 2008;22:1139-1143.

3. Feldman AL, Law M, Remstein ED, et al. Recurrent translocations involving the IRF4 oncogene locus in peripheral T-cell lymphomas. Leukemia. 2009;23:574-580.

4. Swerdlow S, Campo E, Harris N, et al. eds. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. In: Bosman F, Jaffe E, Lakhani S, Ohgaki H eds. World Health Organization Classification of Tumours (ed 4). Lyon: International Agency for Research on Cancer; 2008.

5. Vasmatzis G, Klee EW, Kube DM, Therneau TM, Kosari F. Quantitating tissue specificity of human genes to facilitate biomarker discovery. Bioinformatics. 2007;23:1348-1355.

6. Ralfkiaer E, Willemze R, Paulli M, Kadin ME. Primary cutaneous CD30-positive T-cell lymphoproliferative disorders. In: Swerdlow S, Campo E, Harris N, et al., eds. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (ed 4). Lyon: International Agency for Research on Cancer; 2008.

7. Wada D, Law M, Hsi ED, et al. Specificity of IRF4 translocations for primary cutaneous anaplastic large cell lymphoma: a multicenter study of 204 skin biopsies. Mod Pathol. In press.