

Frequent *CTLA4-CD28* gene fusion in diverse types of T-cell lymphoma

Hae Yong Yoo,^{1,2*} Pora Kim,^{3*} Won Seog Kim,^{2,4*} Seung Ho Lee,^{1*} Sangok Kim,^{3,5*} So Young Kang,⁶ Hye Yoon Jang,⁷ Jong-Eun Lee,⁷ Jaesang Kim,⁸ Seok Jin Kim,^{2,4} Young Hye Ko,^{2,6†} and Sanghyuk Lee^{3,5,8†}

¹Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University; ²Samsung Biomedical Research Institute, Research Institute for Future Medicine, Samsung Medical Center; ³Ewha Research Center for Systems Biology (ERCSB), Ewha Womans University; ⁴Division of Hematology and Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine; ⁵Department of Bio-Information Science, Ewha Womans University; ⁶Department of Pathology, Samsung Medical Center, Sungkyunkwan University School of Medicine; ⁷DNA Link Inc.; and ⁸Department of Life Science, Ewha Womans University, Seoul, Korea

**These authors contributed equally to this work.*

†These authors are co-corresponding authors of this article.

©2016 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2015.139253

Received: November 10, 2015.

Accepted: January 22, 2016.

Pre-published: January 27, 2016.

Correspondence: yhko310@skku.edu or sanghyuk@ewha.ac.kr

Supplementary Information for

Frequent CTLA4-CD28 gene fusion in diverse types of T cell lymphoma

Hae Yong Yoo^{1,2*}, Pora Kim^{3*}, Won Seog Kim^{2,4*}, Seung Ho Lee^{1*}, Sangok Kim^{3,5*}, So Young Kang⁶, Hye Yoon Jang⁷, Jong-Eun Lee⁷, Jaesang Kim⁸, Seok Jin Kim^{2,4}, Young Hye Ko^{2,6†}, and Sanghyuk Lee^{3,5,8†}

¹Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul 06351, Korea

²Samsung Biomedical Research Institute, Research Institute for Future Medicine, Samsung Medical Center, Seoul 06351, Korea

³Ewha Research Center for Systems Biology (ERCSB), Ewha Womans University, Seoul 03760, Korea

⁴Division of Hematology and Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea

⁵Department of Bio-Information Science, Ewha Womans University, Seoul 03760, Korea

⁶Department of Pathology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea

⁷DNA Link Inc., Seoul 03759, Korea

⁸Department of Life Science, Ewha Womans University, Seoul 06351, Korea

*These authors contributed equally to this work.

†These authors are co-corresponding authors of this article.

Corresponding Authors:

Young Hye Ko, Department of pathology, Samsung medical Center, Sungkyunkwan University School of Medicine, and Samsung Biomedical Research Institute, Research Institute for Future Medicine, Samsung Medical Center, Seoul 06351, Korea, Phone: +82-2-3140-2762 Fax: +82-2-3410-0025 E-mail: yhko310@skku.edu

Sanghyuk Lee, Ewha Research Center for Systems Biology (ERCSB), Ewha Womans

University, Seoul 03760, Korea, Phone: +82-2-3277-2888 Fax: +82-2-3277-6809 E-mail:
sanghyuk@ewha.ac.kr

Table of Contents

Supplementary Methods

Quantitative real time PCR	4
Cell culture and transfection	4
Antibodies and immunoblotting	4
Library preparation and targeted deep sequencing	5
Bioinformatics analysis of targeted deep sequencing data.	6

Supplementary Tables

Supplementary Table S1. List of genes for targeted deep sequencing	8
Supplementary Table S2. Study subjects for sequencing and their clinical information..	10
Supplementary Table S3. List of gene-specific PCR primers.....	13
Supplementary Table S4. Summary statistics of targeted deep sequencing	14

Supplementary Figures

Supplementary Fig. S1. Validation of the <i>CTLA4-CD28</i> fusion transcript	16
Supplementary Fig. S2. Validation of the <i>CTLA4-CD28</i> fusion gene in patients	17
Supplementary Fig. S3. Functional analyses of the <i>CTLA4-CD28</i> fusion gene in H9 cell line	18
Supplementary Fig. S4. Surface expression levels of CTLA4 and the CTLA4-CD28 fusion	19
Supplementary Fig. S5. Fluorescence <i>in situ</i> hybridization analysis for the fusion.....	20
Supplementary Fig. S6. Copy number analysis of <i>CD28</i> and <i>CTLA4</i> genes	21
Supplementary Fig. S7. Sanger sequencing of genomic DNA for mapping exact fusion points	22
Supplementary Fig. S8. Computational pipeline for analyzing targeted deep sequencing data	23
Supplementary Fig. S9. Mutation profile of <i>TET2</i> gene from targeted sequencing.....	24

Supplementary Files

Supplementary File S1. List of known mutations in lymphoma	
--	--

Supplementary Methods

Quantitative real-time PCR. For quantitative analysis of genomic DNA and mRNA in patient samples, 30 ng genomic DNA and 2 μ g mRNA were used, respectively. mRNA was reverse-transcribed, using Super Script 2 (Invitrogen) with random primer. qRT-PCR was carried out with SYBR Green PCR master mix (Applied Biosystems) and gene specific primers listed (Supplementary Table 3) on ABI PRISM 7900HT. We compared the normalized the C_T values using the housekeeping gene β -actin.

Cell culture and transfection. The Jurkat E6.1 (human T cell acute lymphoblastic leukemia), H9 (human cutaneous T lymphocyte lymphoma) and HUT78 (human cutaneous T lymphocyte lymphoma) cell lines obtained from the American Type Culture Collection (ATCC) were maintained in RPMI-1640 medium containing 10% FBS, 100 U/ml penicillin, 100 μ g/ml streptomycin and 250 ng/ml amphotericin B at 37 °C and 5% CO₂. Cultured cells were regularly tested for mycoplasma infection using the MycoAlert mycoplasma detection kit (Lonza). To generate expression vector plasmids for CTLA4 and CTLA4-CD28 fusion, cDNA encoding CTLA4 and CD28 were amplified by PCR from appropriate sources and inserted into LZRSpMBN linker IRES EGFP vector. Plasmids expressing CTLA4 and CTLA4-CD28 fusion were transfected into cells using the Nucleofector I device (Amaxa) with Nucleofector solution V and program X-001 (Jurkat) and G-014 (H9). Typically, one million cells were transfected with 2 μ g of plasmid. A GFP-expressing plasmid was used to measure transfection efficiency. Expression of CTLA4 and CTLA4-CD28 fusion proteins was analyzed by flow cytometry. Transfected cells were stained with PE-Cy7 CD28 (CD28.2, BD pharmingen) and PE-Cy5 CTLA4 (BNI3, BD pharmingen) for 30 min and analyzed using FACS Verse (BD bioscience).

Antibodies and immunoblotting. Antibodies used for protein blot were phospho-AKT(S473) (Cell Signaling Technology), AKT antibody (Cell Signaling Technology), phospho-ERK(T202/Y204) antibody (Cell Signaling Technology), ERK antibody (Cell Signaling Technology), α -tubulin antibody (Santa Cruz Biotechnology). Horseradish peroxidase (HRP) conjugated secondary antibodies (Bio-Rad) were used to detect primary antibodies. Equal protein loading was assessed by immunoblotting for α -tubulin. For immunoblotting analysis of CTLA4 and CTLA4-CD28 fusion signaling, cells transfected with indicated plasmids were treated with 2 μ g/ml mouse anti-human CD28 (BD Pharmingen), 2 μ g/ml mouse anti-human CTLA4 (BD

Pharmingen) or normal 2 $\mu\text{g/ml}$ rabbit IgG for 10 min on ice, followed by crosslinking with 5 $\mu\text{g/ml}$ goat anti-mouse IgG or goat anti-rabbit IgG (Ab frontier) for 10 min on ice. Then cells were incubated at a 37°C for 1 hour and lysed. Equal amounts of cell lysates were subjected to SDS-PAGE, transferred and probed with antibody. Equal protein loading was assessed by immunoblotting for α -tubulin.

Library preparation and targeted deep sequencing. We produced the deep sequencing data for 70 target genes on Ion PI v2 chip following the Ion AmpliSeq protocol of Life Technologies. A set of primer pairs were designed by the Ion AmpliSeq Designer in order to cover the coding DNA sequences and adjacent intronic regions (5 bp) of target genes. We achieved the design coverage of 98.9% for the target region of 227 kbp with 2,965 amplicons. Genomic DNA was pre-treated with uracil-DNA glycosylase (UDG) that had been reported to reduce high number of artifactual single nucleotide changes in FFPE samples (Do H *et al.* Clinical Chemistry 2013). Sequencing libraries were prepared according to the manufacturer's protocol. Briefly, 20 ng of UDG-treated FFPE DNAs were amplified with the Ion AmpliSeq Library kit 2.0 for each of 74 lymphoma patient cases. Primers were partially digested, and each sample was barcoded with the Ion Xpress Barcode Adapters kit. Adapter-ligated and barcoded libraries were purified with AMPure XP reagent (Beckman Coulter) and PCR-amplified for 5 cycles. Resulting products were quantified by Agilent 2100 BioAnalyzer and Agilent Bioanalyzer DNA High-Sensitivity LabChip (Agilent Technologies). We pooled 20 uniquely barcoded libraries at the equimolar concentration for multiplexed sequencing on a single Ion PI v2 chip. Emulsion PCR was performed for the pooled sample on Ion OneTouch system using Ion PI Template OT2 200 Kit v3. Template-positive ISPs were sequenced on Ion PI v2 chip using Ion PI Sequencing 200 Kit v3 in the Ion Proton System. Mapping of sequencing data to the human genome (see below) indicated that the depths of reading ranged from 728X to 2578X with the average depth of 1204X. Summary of amplicon sequencing is available in Supplementary Table 4. Library preparation and sequencing procedures were performed by DNA Link Inc. in Korea.

Bioinformatics analysis of targeted deep sequencing data. The overview of the computational pipeline for analyzing deep sequencing data is shown in Supplementary Fig. S8. Human genome assembly (GRCh37, hg19) of repeat-masked version was downloaded from the UCSC genome browser. The sequencing data were mapped to the human genome using the Torrent Mapping Alignment Program (TMAP ver. 3.4.1) with the option of ‘mapall -g 2 -a 0 -y stage1 map1’. Genome Analysis Toolkit (GATK ver. 1.6.7)¹⁹ was used for local realignment and base quality score recalibration. Next, we called SNPs and indels using the GATK HaplotypeCaller with the option of ‘-dontUseSoftClippedBases -stand_emit_cof 20 -stand_call_cof 20’. Finally, ANNOVAR (ver. 2013-05-20) was used for functional annotation of genetic variants with ljb23_all for hg19_refGene release 66 transcriptome model²⁰. Mutation analysis by HaplotypeCaller yielded 64,062 variations including 9,183 exonic ones. Filtering out nonfunctional candidates (synonymous mutations and in-frame indels), we obtained 9,104 mutations of functional significance comprising 745 nonsynonymous point mutations, 155 stop gain mutations, 2 stop loss mutations, 8,202 frame-shift indels. The large number of frame-shift indels is presumably due to the well-known sequencing error for homopolymer repeats in Ion Proton sequencing. For known point mutations from the literature survey (Supplementary File S1), we simply checked their presence in the list of point mutations, and confirmed the presence of 8 missense mutations (RHOA_G17V, IDH2_R172K, DNMT3A_R882H, CD28_T195P, FYN_R176C, VAV1_E524D, RHOA_T19I, PLCG1_S345F) and 2 nonsense mutations in TET2 gene (TET2_R544X, TET2_R1404X). For novel mutations, we applied several stringent filters to reduce false positives. Firstly, SNPs with population diversity in the dbSNP database (ver. 143) were removed. We further filtered out mutations of germline origin using exome sequencing data for paired samples of tumor and matched normal samples from 5 AITL patients, 100 lung cancer patients, 50 gastric cancer patients in Korean population (in-house data). Finally, we confirmed the mutation by Sanger sequencing for cases when patient samples were available. We have thus identified 3 novel cases of point mutations (CD28_F51I, TCF3_G302S, and KMT2D_S4251F). Loss-of-function mutations in TET2 gene are prevalent in myeloid cancers and lymphoma. Thus, a careful analysis of indels leading to loss-of-function is essential, especially for the sequencing data with homopolymer problem. Through a careful evaluation, we determined reliable indels as (i) no. of alternate reads ≥ 10 (i.e. minor allele frequency > 0.01 at 1000X sequencing depth) and (ii) frequency of patient cases ≥ 7 (i.e. observed in less than 10% of lymphoma patients). Indels with patient recurrency over 10% turned out to be false positives in most cases. Among 182 indels from the GATK HaplotypeCaller predictions on TET2 gene, we have obtained 7 indels satisfying

both conditions. In addition, 24 missense mutations were predicted in TET2 gene, 6 of which were filtered out as SNPs with population diversity and additional 11 candidates were non-damaging according to ANNOVAR annotation, leaving 7 missense mutations of functional consequences. Including 10 nonsense mutations, the number of loss-of-function mutations in TET2 gene was 24 (7 missense mutations, 10 nonsense mutations, 7 indels), 19 of which had been previously reported in hematological tumors (Supplementary Fig. S9). Five mutations (E1151X, 1715-1717del, 1747-1751del, Y1148D, G1861V) were novel from our study, and Y1148D mutation was confirmed by Sanger sequencing. Deep sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) database under accession ID SRP056397 and BioProject ID PRJNA278986.

Supplementary Table S1. List of genes for targeted deep sequencing

Gene Symbol	Lymphoma Type*	Publication	First Author	Ref. No.
ACTB	DLBCL	PNAS 2012	Lohr JG	14
ANKRD11	TP53-related			
ARID1A	DLBCL	PNAS 2013	Zhang J	18
ATM	PTCL	Nat. Genet. 2014	Palomero T	3
B2M	PTCL	Nat. Genet. 2014	Palomero T	3
BCL10	FL	Nat. Genet. 2014	Okosun J	15
BCL2	BL	Nature 2012	Schmitz R	17
BRAF	DLBCL	PNAS 2012	Lohr JG	14
BTG1	DLBCL	PNAS 2012	Lohr JG	14
CARD11	FL	Nat. Genet. 2014	Okosun J	15
CCND3	BL	Nature 2012	Schmitz R	17
CD28	AITL	Nat. Genet. 2014	Yoo HY	5
CD58	PTCL	Nat. Genet. 2014	Palomero T	3
CD79B	FL	Nat. Genet. 2014	Okosun J	15
CDKN2A	PTCL	Nat. Genet. 2014	Palomero T	3
CREBBP	FL	Nat. Genet. 2014	Okosun J	15
DNMT3A	TCL	NEJM 2012	Couronné L	26
EBF1	FL	Nat. Genet. 2014	Okosun J	15
EZH2	AITL	Nat. Genet. 2014	Yoo HY	5
FYN	PTCL	Nat. Genet. 2014	Palomero T	3
GNA13	BL	Nature 2012	Schmitz R	17
HIST1H1C	FL	Nat. Genet. 2014	Okosun J	15
HIST1H1E	FL	Nat. Genet. 2014	Okosun J	15
ID3	BL	Nature 2012	Schmitz R	17
IDH1	DLBCL	PNAS 2013	Zhang J	18
IDH2	AITL	BLOOD 2012	Cairns RA	1
IRF8	DLBCL	PNAS 2013	Zhang J	18
JAK2	JAK-STAT pathway			
JAK3	JAK-STAT pathway			
KLHL6	FL	Nat. Genet. 2014	Okosun J	15
KMT2C	AITL	Nat. Genet. 2014	Yoo HY	5
KMT2D	AITL	Nat. Genet. 2014	Yoo HY	5
KRAS	DLBCL	PNAS 2012	Lohr JG	14
LILRB1	AITL	Nat. Genet. 2014	Yoo HY	5
MEF2B	FL	Nat. Genet. 2014	Okosun J	15
MKI67	BL	Nature 2012	Schmitz R	17

MTOR	DLBCL	PNAS 2013	Zhang J	18
MUC2	AITL	Nat. Genet. 2014	Yoo HY	5
MYC	BL	Nature 2012	Schmitz R	17
MYD88	FL	Nat. Genet. 2014	Okosun J	15
NOTCH1	DLBCL	PNAS 2012	Lohr JG	14
P2RY8	DLBCL	PNAS 2012	Lohr JG	14
PCLO	DLBCL	PNAS 2012	Lohr JG	14
PIK3CD	DLBCL	PNAS 2013	Zhang J	18
PIK3R1	DLBCL	PNAS 2013	Zhang J	18
PIM1	DLBCL	PNAS 2012	Lohr JG	14
PLCG1	AITL	Nat. Genet. 2014	Yoo HY	5
POU2F2	DLBCL	PNAS 2013	Zhang J	18
PRKD2	PTCL	Nat. Genet. 2014	Palomero T	3
PTPN1	PMBCL	Nat. Genet. 2014	Gunawardana J	13
RHOA	AITL	Nat. Genet. 2014	Yoo HY	5
RHOT2	PTCL	Nat. Genet. 2014	Palomero T	3
SGK1	BL	Nature 2012	Schmitz R	17
SMARCAL1	PTCL	Nat. Genet. 2014	Palomero T	3
SMARCD1	AITL	Nat. Genet. 2014	Yoo HY	5
SOCS1	FL	Nat. Genet. 2014	Okosun J	15
STAT1	JAK-STAT pathway			
STAT2	JAK-STAT pathway			
STAT3	JAK-STAT pathway			
STAT6	FL	Nat. Genet. 2014	Okosun J	15
TCF3	BL	Nature 2012	Schmitz R	17
TET2	TCL	NEJM 2012	Coutronné L	26
TET3	PTCL	Nat. Genet. 2014	Palomero T	3
TNFAIP3	FL	Nat. Genet. 2014	Okosun J	15
TNFRSF14	FL	Nat. Genet. 2014	Okosun J	15
TP53	TP53-related			
TP63	TP53-related			
VAV1	AITL	Nat. Genet. 2014	Yoo HY	5
WIF1	DLBCL	PNAS 2013	Zhang J	18
WWOX	TP53-related			

*AITL = angioimmunoblastic T cell lymphoma, BL = Burkitt lymphoma, DLBCL = diffuse large B cell lymphoma, FL = Follicular lymphoma, PMBCL = primary mediastinal large B cell lymphoma, PTCL = peripheral T cell lymphoma, TCL = T cell lymphoma

Supplementary Table S2. Study subjects for sequencing and their clinical information

case no.	Diagnosis	Age	Gender	Tissue of origin	CD28-CTLA4 fusion	Targeted resequencing
1	AITL	66	M	Lymph node	N	no
2	AITL	64	M	Lymph node	N	yes
3	AITL	28	F	Lymph node	P	yes
4	AITL	59	M	Lymph node	P	yes
5	AITL	76	F	Lymph node	P	yes
6	AITL	72	M	Lymph node	P	yes
7	AITL	67	M	Lymph node	P	no
8	AITL	59	M	Lymph node	P	yes
9	AITL	52	M	Lymph node	P	yes
10	AITL	42	M	Lymph node	P	yes
11	AITL	60	F	Lymph node	N	yes
12	AITL	53	F	Lymph node	P	yes
13	AITL	67	M	Lymph node	N	yes
14	AITL	57	F	Lymph node	P	yes
15	AITL	58	M	Lymph node	N	no
16	AITL	62	M	Lymph node	P	no
17	AITL	51	M	Lymph node	P	yes
18	AITL	60	F	Lymph node	P	yes
19	AITL	57	F	Lymph node	P	yes
20	AITL	47	M	Lymph node	P	yes
21	AITL	50	F	Lymph node	N	yes
22	AITL	58	M	Lymph node	N	no
23	AITL	67	M	Lymph node	N	no
24	AITL	75	M	Lymph node	P	no
25	AITL	64	M	Lymph node	N	yes
26	AITL	74	M	Lymph node	P	no
27	AITL	72	M	Lymph node	N	no
28	AITL	75	M	Lymph node	N	yes
29	AITL	62	M	Lymph node	P	yes
30	AITL	38	M	Lymph node	N	yes
31	AITL	54	M	Lymph node	P	yes
32	AITL	57	M	Lymph node	P	yes
33	AITL	65	M	Lymph node	P	yes
34	AITL	41	F	Lymph node	N	no
35	AITL	77	M	Lymph node	N	no
36	AITL	78	M	Lymph node	N	yes
37	AITL	79	M	Lymph node	N	no
38	AITL	57	M	Lymph node	P	no
39	AITL	73	F	Lymph node	N	yes

40	AITL	70	M	Lymph node	P	yes
41	AITL	63	M	Lymph node	P	no
42	AITL	60	F	Lymph node	P	yes
43	AITL	77	M	Lymph node	N	no
44	AITL	56	M	Lymph node	N	no
45	AITL	48	M	Lymph node	N	yes
46	ALCL. ALK-	62	M	Lymph node	N	yes
47	ALCL. ALK+	9	M	Mediastinum	N	yes
48	EATL	52	F	Small intestine	N	no
49	EATL	53	F	Jejunum	N	yes
50	EATL	50	M	Ileum	N	yes
51	ENKL	45	M	Nasal cavity	P	yes
52	ENKL	58	F	Lymph node	P	yes
53	ENKL	31	M	Cecum	N	yes
54	ENKL	64	M	Nasal cavity	P	yes
55	ENKL	54	M	Soft palate	N	yes
56	ENKL	50	M	Nasal cavity	N	yes
57	ENKL	46	M	Nasal cavity	N	yes
58	ENKL	66	F	Nasal cavity	N	yes
59	ENKL	57	M	Testis	N	yes
60	ENKL	34	M	Nasal cavity	N	yes
61	ENKL	51	F	Nasal cavity	N	no
62	ENKL	41	F	Lymph node	N	no
63	ENKL	48	M	Nasal cavity	P	yes
64	ENKL	48	M	Paranasal sinus	N	no
65	ENKL	47	M	Testis	N	yes
66	ENKL	43	M	Ileum	P	yes
67	ENKL	25	M	Jejunum	N	yes
68	ENKL	45	M	Testis	N	yes
69	ENKL	83	F	Nasal cavity	N	yes
70	ENKL	54	F	Nasal cavity	N	yes
71	ENKL	40	M	Colon	P	yes
72	ENKL	75	M	Nasal cavity	N	yes
73	ENKL	55	F	Nasal cavity	N	no
74	ENKL	60	M	Nasal cavity	P	yes
75	ENKL	44	F	Nasal cavity	N	yes
76	ENKL	32	F	Nasal cavity	N	yes
77	ENKL	73	M	Nasal cavity	ND	yes
78	ENKL	34	M	Nasal cavity	ND	yes
79	ENKL	76	F	Soft tissue	N	no
80	ENKL	43	F	Nasal cavity	N	no
81	ENKL	62	M	Nasal cavity	P	no
82	ENKL	44	M	Nasopharynx	P	yes

83	ENKL	24	M	Skin	N	no
84	ENKL*	55	M	Lymph node	ND	yes
85	PTCL	74	M	Lymph node	P	no
86	PTCL	29	F	Mediastinum	N	yes
87	PTCL	51	M	Lymph node	N	yes
88	PTCL	62	M	Stomach	N	no
89	PTCL?	64	F	Tonsil	P	no
90	PTCL	54	F	Lymph node	N	no
91	PTCL	67	M	Lymph node	N	no
92	PTCL	46	F	Lymph node	P	yes
93	PTCL	48	F	Lymph node	P	yes
94	PTCL	15	M	Lymph node	N	no
95	PTCL	70	M	Lymph node	N	no
96	PTCL	78	M	Larynx	N	no
97	PTCL	69	F	Jejunum	N	no
98	PTCL	49	M	Soft tissue	N	yes
99	PTCL	57	M	Lymph node	N	yes
100	PTCL	66	M	Lymph node	P	yes
101	PTCL	60	F	Lymph node	N	no
102	PTCL	73	F	Lymph node	P	yes
103	PTCL	60	F	Lymph node	N	yes
104	PTCL	49	F	Lymph node	N	no
105	PTCL	47	F	Lymph node	N	no
106	PTCL	51	F	Lymph node	P	no
107	PTCL	77	M	Lymph node	N	no
108	PTCL	24	M	Lymph node	N	no
109	PTCL	38	M	Skin	N	yes
110	PTCL	42	F	Colon	N	no
111	PTCL	71	F	Lymph node	N	no
112	PTCL	55	F	Colon	P	no
113	PTCL	75	F	Nasopharynx	ND	yes
114	PTCL	65	F	Tonsil	P	no
115	PTCL	68	M	Lymph node	N	yes
116	PTCL	53	M	Lymph node	N	yes
117	PTCL	69	M	Lymph node	N	no
118	PTCL	76	M	Lymph node	P	yes
119	PTCL	39	F	Lymph node	N	yes

AITL: Angioimmunoblastic T cell lymphoma, ENKL: Extranodal NK/T cell lymphoma, * ENKL, nodal, ALCL; Anaplastic large cell lymphoma, EATL: Enteropathy-associated T cell lymphoma, PTCL: Peripheral T cell lymphoma, NOS (not otherwise specified)

N: negative, P:positive, ND: not done

Supplementary Table S3. List of gene-specific PCR primers

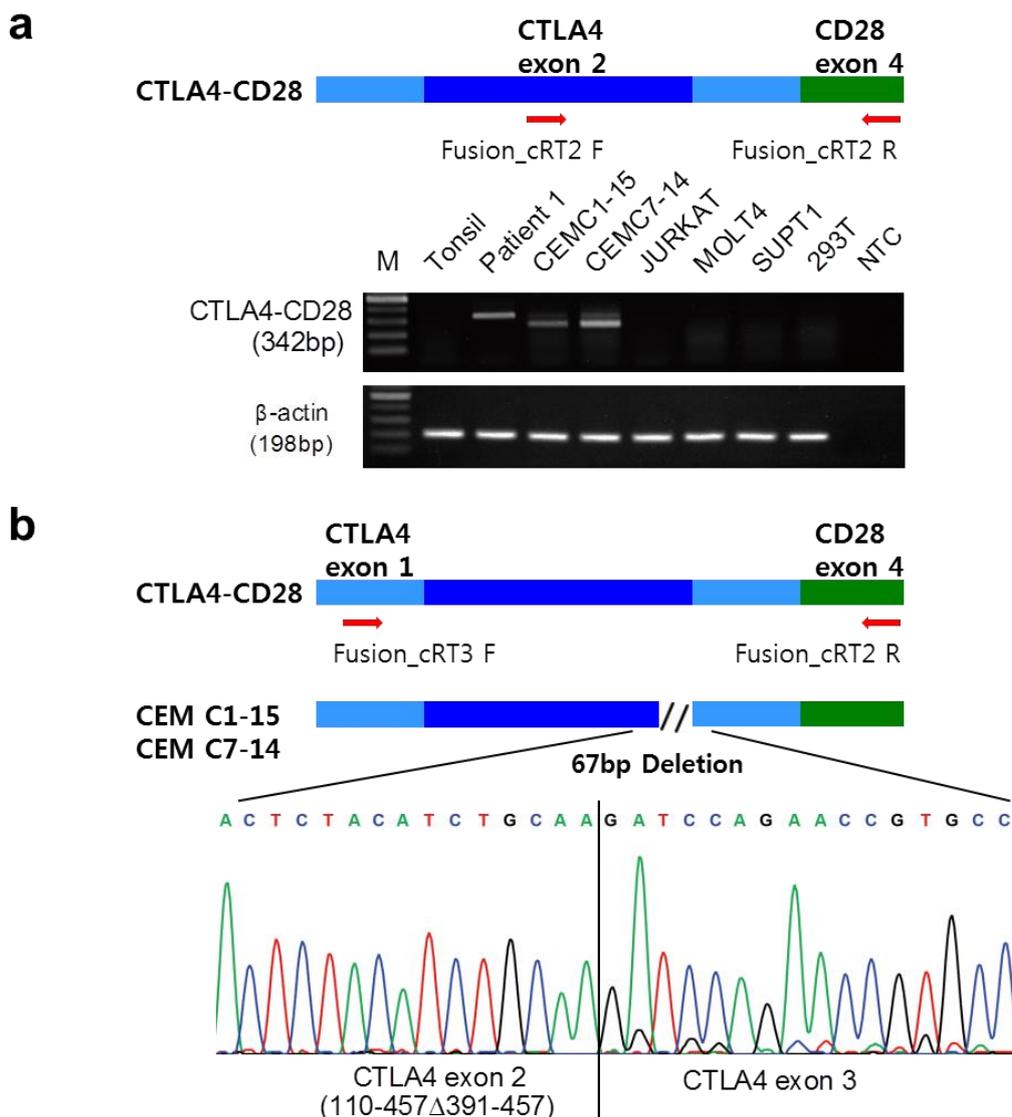
PCR Type	Figure Reference	Primer Name	Sequence
cDNA RT-PCR primer	Fig. 2A, Fig. S2	Fusion_cRT1 F Fusion_cRT1 R	5'- GATCCTTGCAGCAGTTAGTTCGGGG-3' 5'- GGGCTGGTAATGCTTGCGGGTGGGC-3'
	Fig. S1A	Fusion_cRT2 F Fusion_cRT2 R	5'- GGA CTGAGGGCCATGGACACGGGAC-3' 5'- GGAGCGATAGGCTGCGAAGTCGC-3'
	Fig. S1B	Fusion_cRT3 F Fusion_cRT2 R	5'-AGCTGAACCTGGCTACCAGG-3' 5'-GGAGCGATAGGCTGCGAAGTCGC-3'
	Fig. 2A, Figs. S1,S2	β -actin_cRT F β -actin_cRT R	5'-CCAACCGCGAGAAGATGACC-3' 5'-GGTCCAGACGCAGGATGGC-3'
cDNA qPCR primer	Fig. 3C	CD28_cQ F CD28_cQ R	5'-ACAATGCGGTCAACCTTAGC-3' 5'-ACCTGAAGCTGCTGGGAGTA-3'
	Fig. 3C	CTLA4_cQ F CTLA4_cQ R	5'-GTGCCCAGATTCTGACTTCC-3' 5'-CTGGCTCTGTTGGGGGCATTTTC-3'
	Fig. 3C	Fusion_cQ F Fusion_cQ R	5'-CAGCAGTTAGTTCGGGGTTG-3' 5'-GCGGGGAGTCATGTTTCATGT-3'
	Fig. 3C	β -actin_cQ F β -actin_cQ R	5'-CCAACCGCGAGAAGATGACC-3' 5'-GGTCCAGACGCAGGATGGC-3'
gDNA PCR primer	Fig. 3E	Fusion_G1 F Fusion_G1 R	5'- CTGTCTCAGGGAGGCTCTGC-3' 5'- CGGCTGGCTTCTGGATAGG-3'
	Fig. S8A	Fusion_G1 F Fusion_G2 R	5'-CTGTCTCAGGGAGGCTCTGC-3' 5'-GGCGGTCATTTCTATCCAG-3'
gDNA qPCR primer	Fig. S6	CD28_gQ1 F CD28_gQ1 R	5'-AGGCATTGATGAGGATACGC-3' 5'-TTCTATCCCTTGCCATGACC-3'
	Fig. S6	CD28_gQ2 F CD28_gQ2 R	5'-AGGGATGGGTTACAGCACAG-3' 5'-GAGTTCGAGGAAGCCAGTTG-3'
	Fig. S6	CD28_gQ3 F CD28_gQ3 R	5'-CGGTGAGCAAGCAGAATACA-3' 5'-GGAAGAGCAACCAACTCCAG-3'
	Fig. S6	CD28_gQ4 F CD28_gQ4 R	5'-GGCCCACATTCCAACCTTACC-3' 5'-GGGAAGAGGCTCCCAGAATC-3'
	Fig. S6	CD28_gQ5 F CD28_gQ5 R	5'-TCCAATCAGACCAGGTAGGAGC-3' 5'-CCACAACCCACTTTGGATCTCC-3'
	Fig. 3B, Fig. S6	CD28_gQ6 F CD28_gQ6 R	5'-CACAGGCATGTTCTACCTCAGG-3' 5'-GGACCTGAAGGGTGACGAGG-3'
	Fig. 3B, Fig. S6	CTLA4_gQ1 F CTLA4_gQ1 R	5'-CTCACTATCCAAGGACTGAGGGC-3' 5'-CTGGGTTCCGTTGCCTATGC-3'
	Fig. S6	CTLA4_gQ2 F CTLA4_gQ2 R	5'-TGCAATTTAGGGGTGGACCT-3' 5'-AGAATCTGGGCACGGTTCTGGAT-3'
	Fig. S6	CTLA4_gQ3 F CTLA4_gQ3 R	5'-CCATCACCTGGAAGTCACCT-3' 5'-CCCAGTCAAGCAAACCTGGAT-3'
	Fig. S6	CTLA4_gQ4 F CTLA4_gQ4 R	5'-CAGGGAAGTTTTGTGGAGGA-3' 5'-CACAAATCCACGCAATCAAG-3'
Fig. 3B, Fig. S6	β -actin_gQ F β -actin_gQ R	5'-TGAGCCTCATCTCCACGTA-3' 5'-TCTCAGCCAGCACCATGACT-3'	

Supplementary Table S4. Summary statistics of targeted deep sequencing

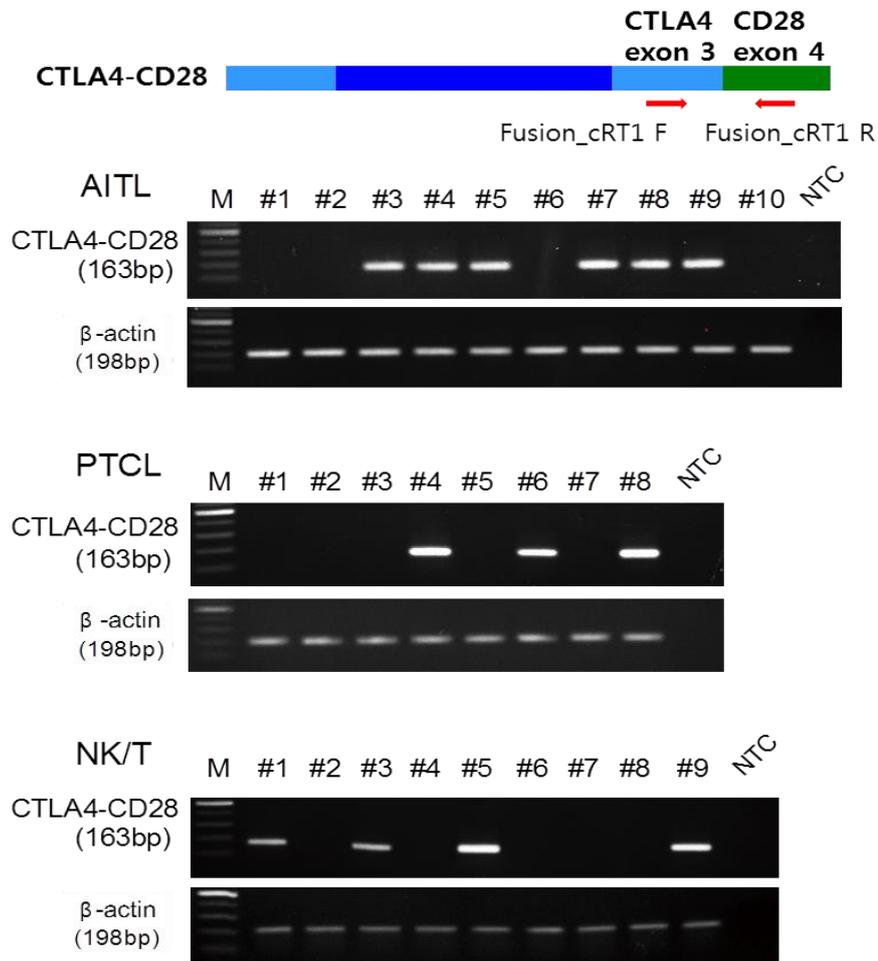
Gene	No. Exons	No. Amplicons	CDS (bp)	Target (bp)	design coverage		experimental coverage	
					bp	%	bp	%
ACTB	5	14	1128	1178	1178	100.0	1178	100.0
ANKRD11	11	80	7992	8102	7650	94.4	7679	94.8
ARID1A	20	82	6858	7058	6987	99.0	6898	97.7
ATM	62	152	9171	9791	9747	99.6	9771	99.8
B2M	3	7	360	390	390	100.0	390	100.0
BCL10	3	10	702	732	732	100.0	732	100.0
BCL2	2	8	753	773	773	100.0	715	92.5
BRAF	18	39	2301	2481	2446	98.6	2449	98.7
BTG1	2	8	516	536	536	100.0	533	99.4
CARD11	24	54	3465	3705	3705	100.0	3705	100.0
CCND3	6	13	927	987	987	100.0	959	97.2
CD28	4	11	663	703	703	100.0	703	100.0
CD58	6	13	757	817	817	100.0	817	100.0
CD79B	6	11	693	753	753	100.0	753	100.0
CDKN2A	5	11	912	962	962	100.0	962	100.0
CREBBP	31	96	7329	7639	7638	100.0	7497	98.1
DNMT3A	24	46	2864	3104	3104	100.0	3104	100.0
EBF1	16	31	1776	1936	1936	100.0	1936	100.0
EZH2	19	38	2256	2446	2446	100.0	2421	99.0
FYN	12	27	1770	1890	1890	100.0	1890	100.0
GNA13	4	15	1134	1174	1174	100.0	1166	99.3
HIST1H1C	1	7	642	652	652	100.0	652	100.0
HIST1H1E	1	7	660	670	640	95.5	670	100.0
ID3	2	5	360	380	380	100.0	380	100.0
IDH1	8	19	1245	1325	1325	100.0	1325	100.0
IDH2	11	20	1359	1469	1464	99.7	1469	100.0
IRF8	8	18	1281	1361	1353	99.4	1361	100.0
JAK2	23	56	3399	3629	3629	100.0	3629	100.0
JAK3	23	51	3375	3605	3601	99.9	3605	100.0
KLHL6	7	24	1866	1936	1936	100.0	1841	95.1
KMT2C	59	199	14736	15326	15103	98.5	9903	99.9
KMT2D	54	199	16614	17154	17148	100.0	1154	97.2
KRAS	5	11	687	737	737	100.0	15212	99.3
LILRB1	14	22	1967	2107	1713	81.3	16726	97.5
MEF2B	8	16	1107	1187	1171	98.7	737	100.0

MKI67	14	103	9771	9911	9819	99.1	1770	84.0
MTOR	57	116	7650	8220	8215	99.9	8159	99.3
MUC2	50	108	8442	8942	8191	91.6	7769	86.9
MYC	3	16	1365	1395	1395	100.0	1395	100.0
MYD88	5	14	954	1004	1004	100.0	1004	100.0
NOTCH1	34	99	7668	8008	7901	98.7	7930	99.0
P2RY8	1	10	1080	1090	1090	100.0	0	0
PCLO	25	181	15446	15696	15571	99.2	15601	99.4
PIK3CD	22	44	3135	3355	3355	100.0	3355	100.0
PIK3R1	17	33	2297	2467	2462	99.8	2467	100.0
PIM1	6	16	1215	1275	1275	100.0	1219	95.6
PLCG1	32	64	3876	4196	4196	100.0	4196	100.0
POU2F2	14	22	1444	1584	1584	100.0	1558	98.4
PRKD2	18	39	2637	2817	2817	100.0	2794	99.2
PTPN1	10	20	1308	1408	1408	100.0	1308	92.9
RHOA	4	9	582	622	622	100.0	622	100.0
RHOT2	19	32	1857	2047	1986	97.0	2028	99.1
SGK1	17	32	1935	2105	2105	100.0	2105	100.0
SMARCAL1	16	42	2865	3025	3025	100.0	3025	100.0
SMARCD1	13	27	1548	1678	1678	100.0	1678	100.0
SOCS1	1	6	636	646	646	100.0	588	91.0
STAT1	23	44	2257	2487	2487	100.0	2487	100.0
STAT2	23	40	2556	2786	2782	99.9	2657	95.4
STAT3	23	40	2313	2543	2521	99.1	2543	100.0
STAT6	21	42	2544	2754	2714	98.6	2754	100.0
TCF3	19	34	2192	2382	2382	100.0	2382	100.0
TET2	9	70	6098	6188	6188	100.0	6111	98.8
TET3	11	61	5388	5498	5498	100.0	5498	100.0
TNFAIP3	8	28	2373	2453	2453	100.0	2453	100.0
TNFRSF14	8	15	852	932	932	100.0	932	100.0
TP53	12	21	1263	1383	1383	100.0	1328	96.0
TP63	16	36	2200	2360	2360	100.0	2360	100.0
VAV1	27	44	2538	2808	2808	100.0	2808	100.0
WIF1	10	17	1140	1240	1240	100.0	1240	100.0
WWOX	10	20	1303	1403	1403	100.0	1403	100.0
Total	1105	2965	216,353	227,403	224,902	98.9	222,449	97.82

Supplementary Fig. S1. Validation of the *CTLA4-CD28* fusion transcript. (a) Validation of the *CTLA4-CD28* fusion transcript by RT-PCR using samples from patients and cell lines. Arrows indicate the approximate positions of oligonucleotide primers on the *CTLA4-CD28* fusion transcript. PCR products were amplified from patient 1, CEMC1-15 cells and CEMC7-14 cells and validated by Sanger sequencing. β -actin was used as an internal control, and NTC indicates the no template control. Normal tonsil tissue and 293T cells were used as negative controls. (b) The fusion products from CEMC1-15 and CEMC7-14 cells were shorter due to the partial deletion in the exon 2 of *CTLA4* gene (391-457 bp region). This is a frame-shift deletion that would lead to loss-of-function of the fusion transcript. Thus, the fusion transcript is expected to play no active functional role in these cell lines.



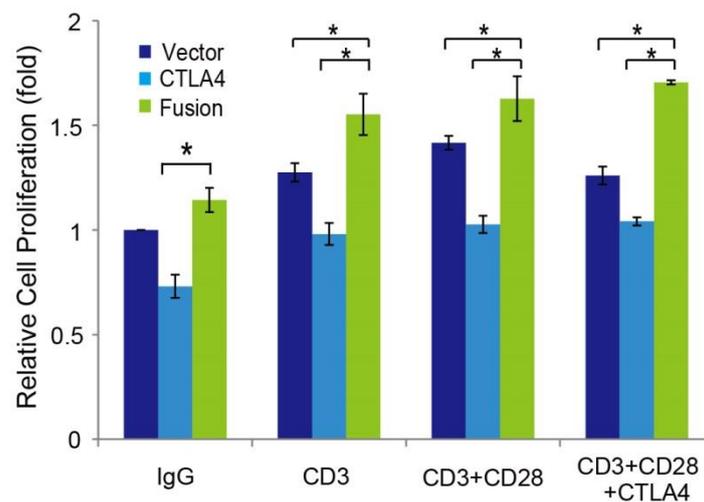
Supplementary Fig. S2. Validation of the *CTLA4-CD28* fusion gene in patients. Fusion transcripts were validated by RT-PCR using FFPE samples from TCL patients. Arrows indicate the approximate positions of oligonucleotide primers on the *CTLA4-CD28* fusion transcript. PCR products were validated by Sanger sequencing. β -actin was used as an internal control, and NTC indicates the no template control.



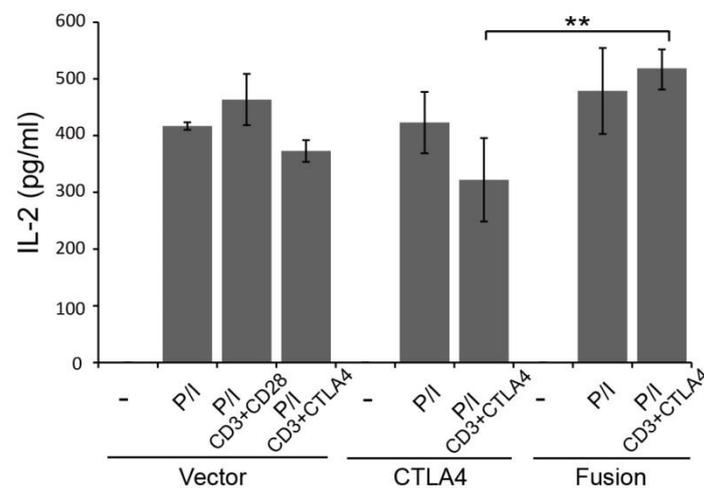
Supplementary Fig. S3. Functional analyses of the *CTLA4-CD28* fusion gene in H9 cell line.

(a) *In vitro* proliferation assay in H9 cells. H9 cells expressing the *CTLA4-CD28* fusion show enhanced cell proliferation (* $P < 0.05$ compared with cells expressing wild-type CTLA4) after co-stimulation with anti-CD3/CTLA4 antibodies. Each experiment was repeated three times with five replicates, and the data are expressed as the mean \pm standard deviation. **(b)** Expression of the *CTLA4-CD28* fusion enhanced interleukin 2 (IL-2) production after CTLA4 activation (** $P < 0.01$ compared with cells expressing wild-type CTLA4). H9 cells were stimulated with PMA/Ionomycin (P/I) without or with anti-CD3/CTLA4 antibodies to activate CTLA4. IL-2 measurement was carried out three independent times, and the data are expressed as the mean \pm standard deviation.

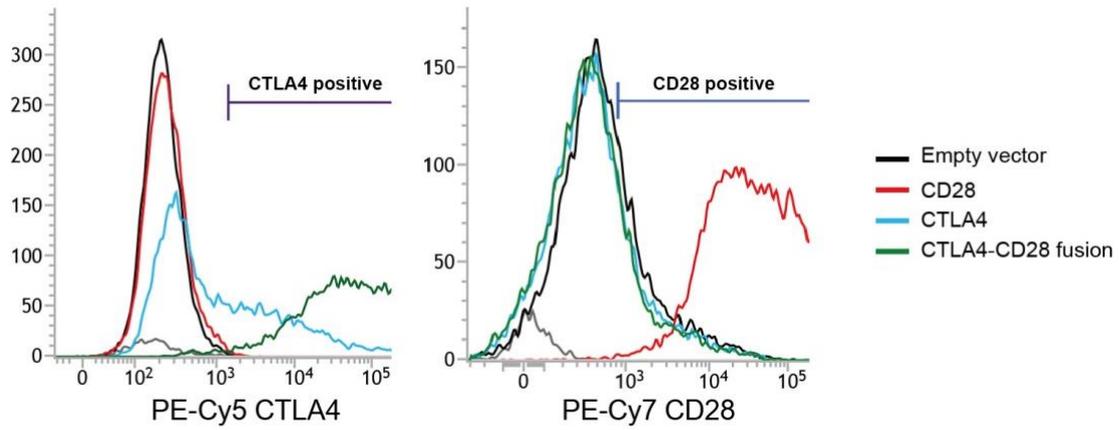
a



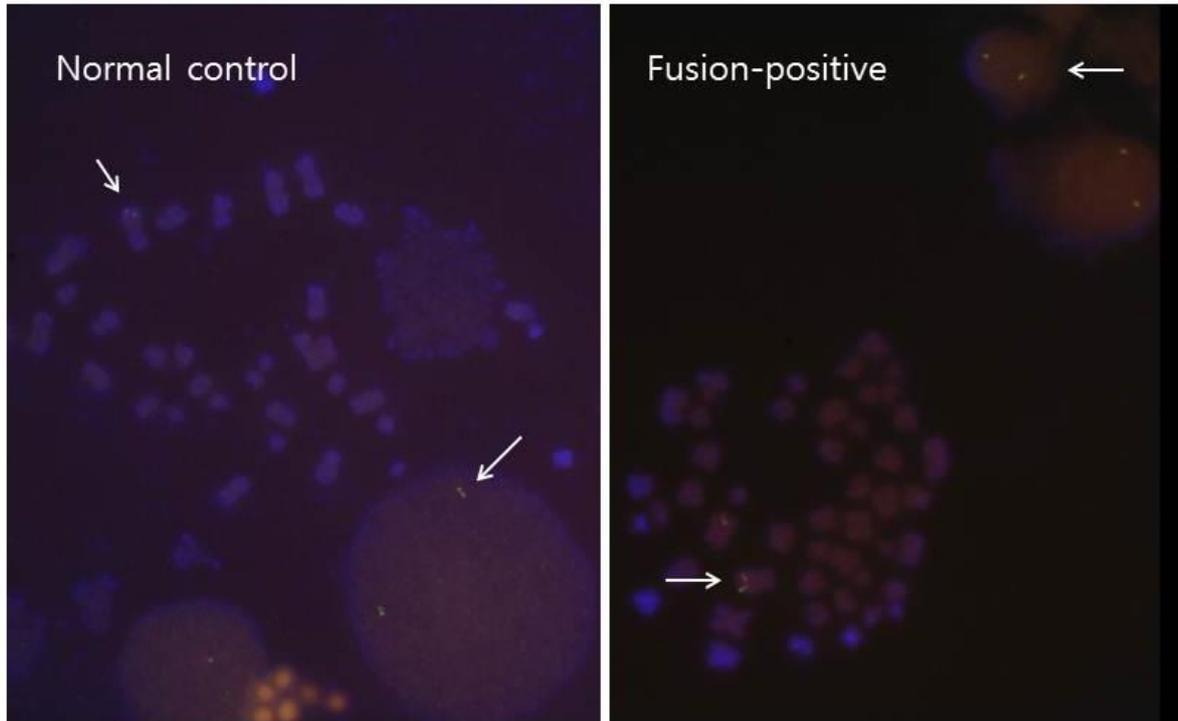
b



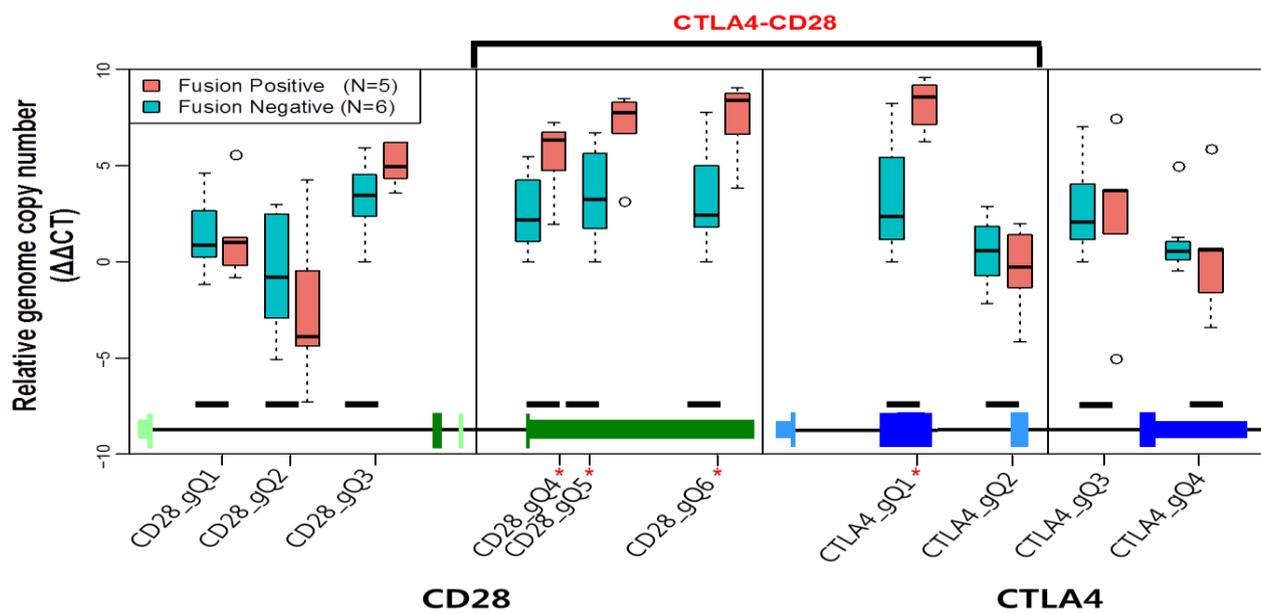
Supplementary Fig. S4. Surface expression levels of CTLA4 and the CTLA4-CD28 fusion. Transfected Jurkat cells were stained with PE-Cy5 labeled anti-CTLA4 and PE-Cy7 labeled anti-CD28. GFP-positive cells were gated and analyzed. Only the cells expressing CTLA4 and CTLA4-CD28 fusion transcripts have positive signals on PE-Cy5 labeled anti-CTLA4 staining.



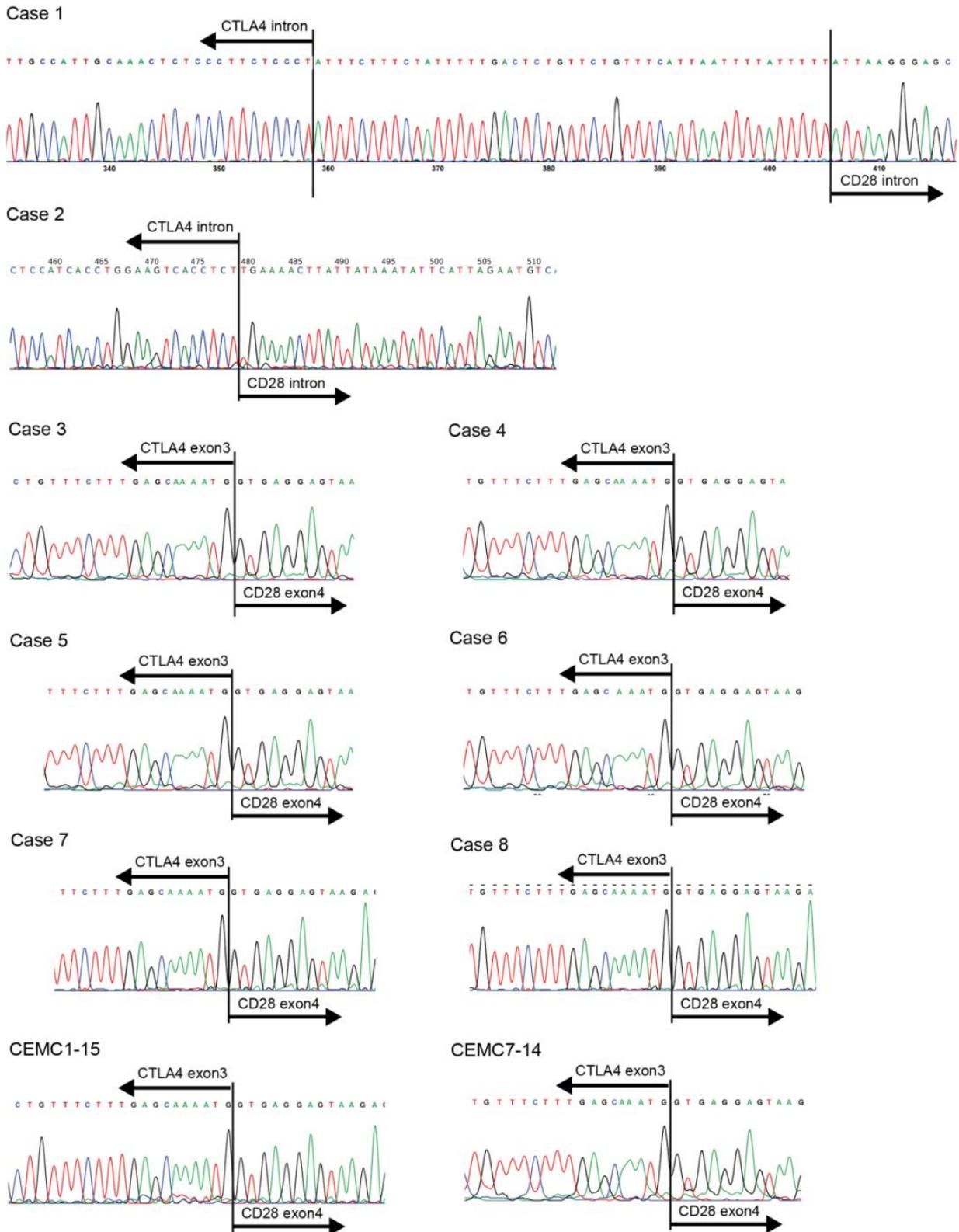
Supplementary Fig. S5. Fluorescence *in situ* hybridization analysis for the fusion. Metaphase chromosome of normal cells without fusion showed two orange signals which indicated merged green and red signal derived from closely located CD28 and CTLA genes, respectively. The cells with fusion showed two orange signals as well. There was no separate signal indicating the presence of episomal fusion gene. This does not rule out the possibility of episome formation in all fusion-positive cases.



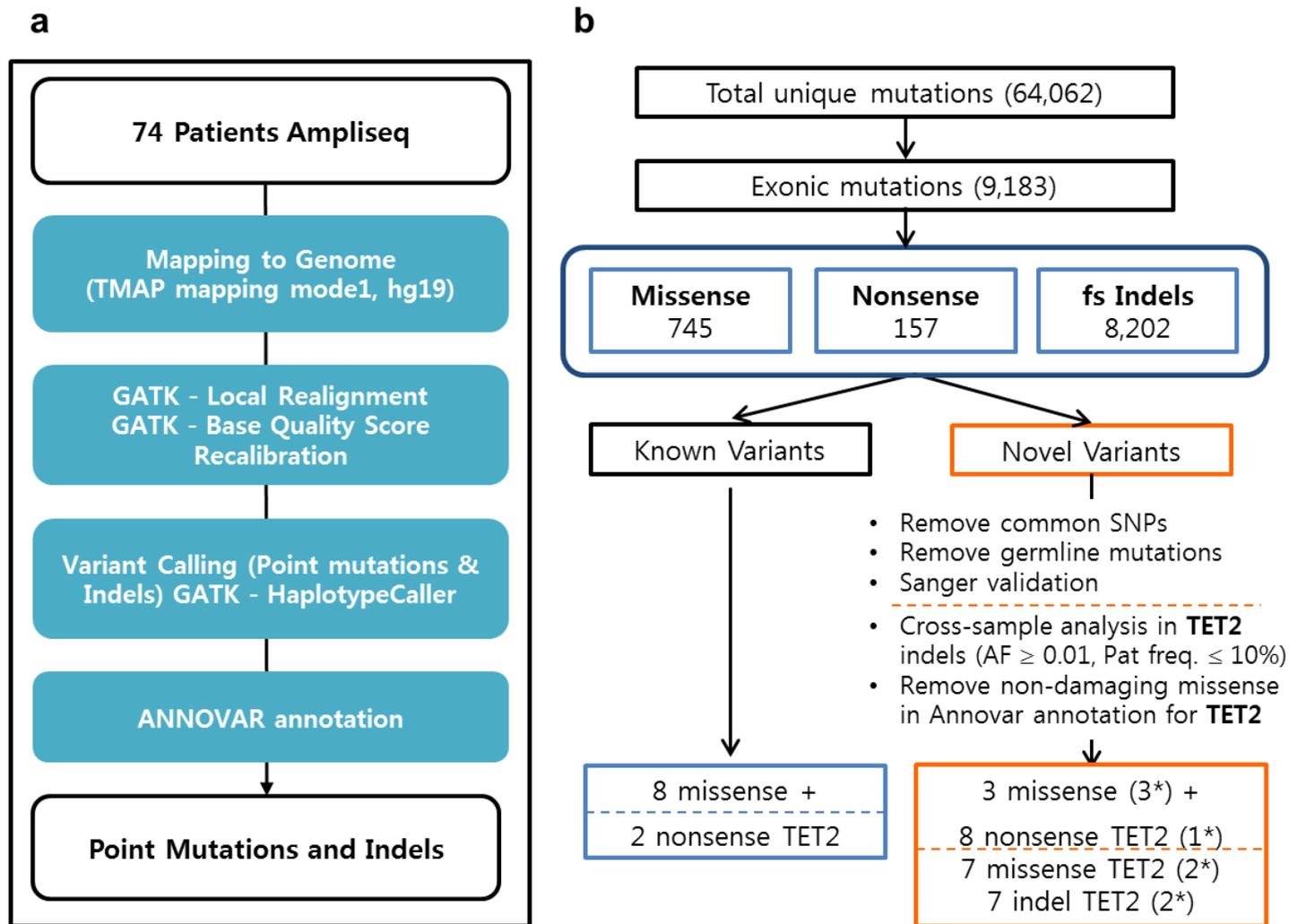
Supplementary Fig. S6. Copy number analysis of *CD28* and *CTLA4* genes in fusion negative and positive TCL patients. Copy number change was estimated relative to that in the peripheral blood cells from a normal individual. Quantitative real time PCR was carried out using multiple primer pairs covering genomic loci of *CD28* and *CTLA4* genes. β -actin was used as a normalizing control. Four out of five primer pairs within the fusion region showed significant copy number gains in fusion-positive patients. Results from two representative primer pairs (*CD28_gQ5* and *CTLA4_gQ1*) were shown in Fig. 3b. The fusion regions are indicated by the bracket at the top, and the primer positions are indicated by the bars at the bottom. Cases of statistically significant copy number gain are indicated by the asterisks (P-value < 0.05).



Supplementary Fig. S7. Sanger sequencing of genomic DNA for mapping exact fusion points. In Case 1, a LINE element of 47 bp long was identified between the fusion introns. Cases 3-6 and two cell lines are examples of direct joining of two exons. Vertical bars indicate the break points.



Supplementary Fig. S8. Computational pipeline for analyzing targeted deep sequencing data



Supplementary Fig. S9. Mutation profile of *TET2* gene from targeted sequencing

