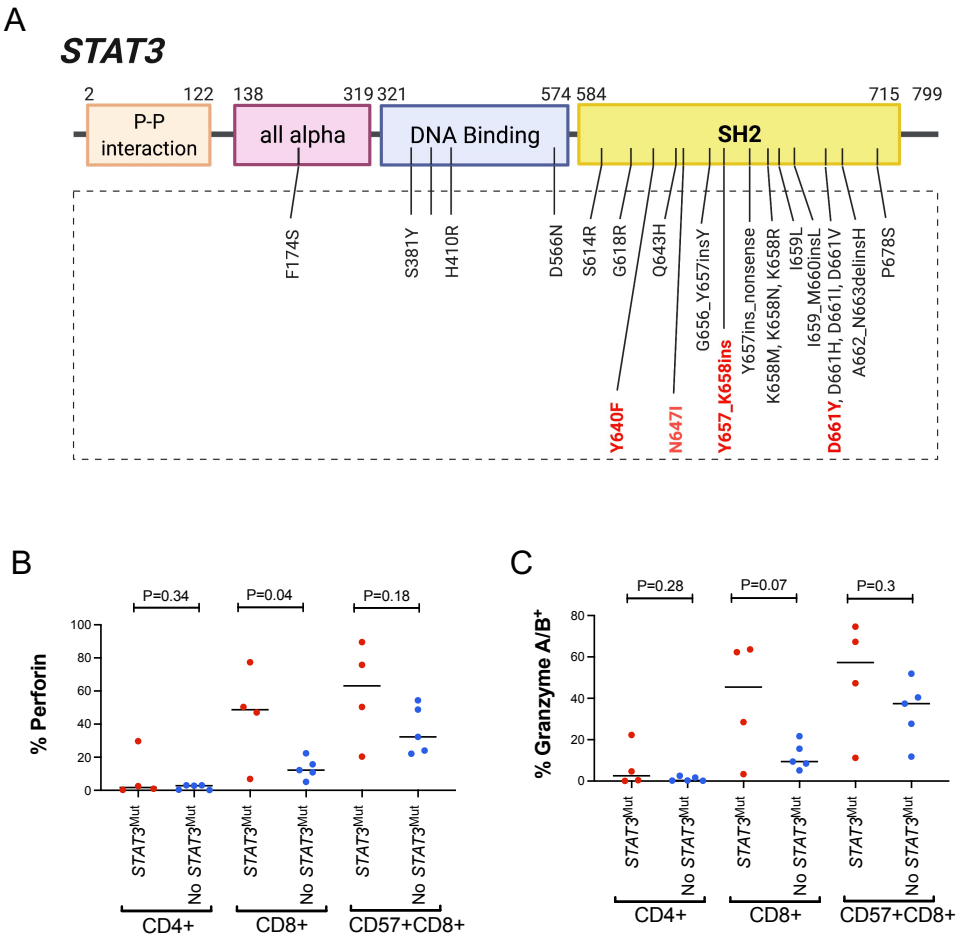
Somatic *STAT3* mutations in CD8+ T cells of healthy blood donors carrying human T-cell leukemia virus type 2

Daehong Kim,[1,2] Mikko Myllymäki,[1,2] Matti Kankainen[1-4] Timo Järvinen,[1,2] Giljun Park,[1,2] Roberta Bruhn,[5,6] Edward L. Murphy[5,6] and Satu Mustjoki[1,2,4]

[1]Hematology Research Unit Helsinki, Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland; [2]Translational Immunology Research Program and Department of Clinical Chemistry and Hematology, University of Helsinki, Helsinki, Finland; [3]Department of Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [4]iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [5]Vitalant Research Institute, San Francisco, CA, USA and [6]University of California San Francisco, San Francisco, USA

Correspondence: SATU MUSTJOKI - satu.mustjoki@helsinki.fi

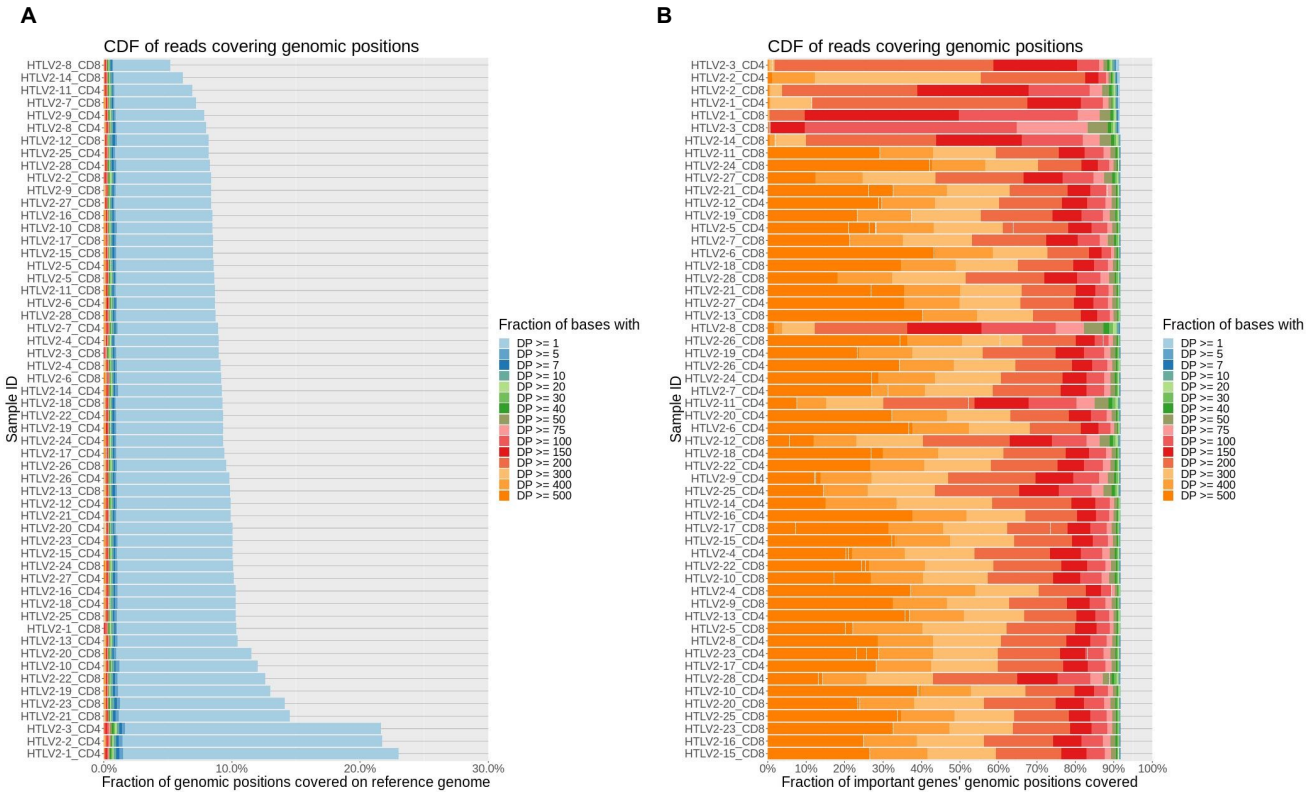
doi:10.3324/haematol.2021.279140



Supplementary Figure S1. Detection of *STAT3* mutations and immunophenotyping in HTLV-2 positive blood donors.

(A) Detection of *STAT3* mutations by ultra-deep targeted amplicon sequencing. The sequencing was performed with CD4+ and CD8+ T cells of HTLV-2 positive cases (n = 30) and HTLV-2 negative cases (n = 35). *STAT3* mutations discovered in CD8+ T cells of HTLV-2 positive blood donors are shown in red. All these mutations have also previously been found in LGL leukemia. Additional *STAT3* mutations reported in LGL leukemia are presented with black letters. We designed 23 primer pairs to identify *STAT3* mutations in the mutational hotspot exons of the *STAT3* gene. Index primer 1 (AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTTCCCTACACGACGCTCTTC CGATC*T), index primer 2 (CAAGCAGAAGACGGCATACGAGATNNNNN NNNGTGACTGGAGTTCA GACGTGTGCTCTTCCGATC*T). Amplicons from different samples were indexed with sample specific index primers to make pooling of several samples together possible. Following overhangs were added to locus specific primers to make them compatible with the index primers: Adapter1 (before locus specific forward primer 5'-3') ACACTCTTTCCCTACACGACGCTCTTCCGATCT, Adapter2 (before locus specific reverse primer 5'-3') AGACGTGTGCTCTTCCGATCT. Oligo primer sequences were adapted from Illumina adapter sequences document #1000000002694. Amplicons were amplified with 1-step PCR protocol in a volume of 20 µL containing ~10 ng of sample DNA, 10 µL of 2× Phusion High-Fidelity PCR Master Mix (Cat: 10402678, Thermo Fisher Scientific, USA), 0.025µM of each locus-specific primer, 0.5µM of index primer 1 (P5) and 0.5µM of index primer 2 (P7). The reaction mix was brought to a final volume with water. In this reaction, the locus-specific primers are present in limiting quantities. Adapter primers amplify these products further. G-Storm GS4 (Somerton) thermal cyclers were used to cycle the 1-step PCR samples according to the program: initial denaturation at 98°C 30

s, 30 cycles at 98°C for 10 s, at 59°C for 30 s, and at 72°C for 15 s, and the final extension at 72°C for 10 minutes. The amplification of the samples was checked with FlashGel™ System (Lonza, Basel, Switzerland). Amplified samples were pooled together without exact quantification and the pool was purified with Agencourt AMPure XP beads (A63881, Beckman Coulter, USA) twice, using 0,8x volume of beads compared to the sample pool volume. Purification with 0,8x bead volume removed effectively primer dimers from the sample pool. Agilent 2100 Bioanalyzer (Agilent Genomics, USA) was used to quantify amplification performance and yield of the purified sample pools. Sample pools were sequenced with Illumina MiSeq System using Illumina MiSeq Reagent Kit v2500 cycles kit (MS-102-2003, Illumina, USA). Amplicon alignment was performed with Bowtie2 and GATK IndelRealigner. Variant calling was done if the variant count was >5 and base frequency was 0.5% of all reads. Variants with a frequency ratio (<0.9) were excluded. (B-C) The graphs represent the flow-cytometry based immunophenotyping analysis for cytotoxic markers. Percentage of perforin (B) and granzyme (C) in CD4+, CD8+, and CD8+CD57+ T cells in HTLV-2 positive blood donors harboring *STAT3* mutations (n = 4, $STAT3^{Mut}$) and without *STAT3* mutations (n = 5, No $STAT3^{Mut}$). Each dot represents one individual, and horizontal lines indicate median values. Statistically significant difference was evaluated using Mann-Whitney U test. For immunotyping, peripheral blood mononuclear cells (PBMCs) were immunostained with the antibody panel including anti-CD45_V500, anti-CD3_APC, anti-CD8_PE-Cy7, anti-CD57_PE, granzyme A/B_Alexa700, perforine_PerCP-Cy5.5. Antibodies were purchased from BD bioscience (NJ, USA).



Supplementary Figure S2. Immunogene panel sequencing.

Somatic mutations in the coding regions of 2,533 immune-related genes in a custom designed panel from 28 HTLV-2 positive blood donors were investigated using Illumina NovaSeq 6000 system in S1 flow cell (Illumina, CA, USA). Read length was 100+15+8+100. KAPA Hyper Plus Library preparation and SeqCap EZ Enrichment were processed according to SeqCapEZ HyperCap Workflow User's Guide Version 2.1 (Roche Nimblegen, Madison, WI, USA) using KAPA Hyper Plus library preparation kit and xGen® Dual Index UMI Adapters. 7 cycles were used for pre capture amplification. Libraries were pooled to 5-plex reactions according to concentration. The enrichment was performed using Nimblegen Immunopanel v2 probes. 10 cycles were used for post capture amplification. Library pools was quantified for sequencing using 2100 Bioanalyzer High sensitivity kit (Agilent Technologies). Variants specific for CD8+ T cells and found only in CD8+ T cells with the help of a matched normal sample (CD4+ T cells) were accepted. Variant filtering included exclusion of variants not passing all MuTect2 filters, TLOD <6.30 or TLOD <5.00 and not supported by ≥5 independent COSMIC variants, present in CD4+ and CD8+ panel of normal samples from 21 healthy individuals, not locating in RefSeq exonic or splice-site (+/-5 bp) regions, minor allele frequency (MAF) >0.01 in ExAC, gnomAD exome (v2.1.1), or gnomAD genome (v2.1.1) ALL, FIN, NFE, EAS, AMR or AFR databases, MAF >0.01 in gnomAD genome (v3.0) ALL, AFR, AMI, AMR, ASJ, EAS, FIN, NFE, OTH, or SAS male, female, or male and female databases, MAF >0.01 in gnomAD genome (v3.0) RAW database, MAF >0.01 in esp6500 database, MAF >0.01 in 1000g database, quality value <40, SNV strand-odd-ratio value >3, indel strand-odd-ratio value >11, CADD score <3, LR score <2, indel length >5, total read depth <10, variant allele frequency <0.02, variant allele frequency difference between tumor and normal <0.02, ALT_F1R2

depth <1, or ALT_F2R1 depth <1. Sequencing coverage was analysed using SAMtools (Version: 3.7.6). With the computed information including read depths of genomic positions, a cumulative sum of the percentage was created. The analysis was performed to the whole genome and for selected genome regions. The analysis of the selected genome regions was restricted to the immunopanel gene areas, which were padded by ± 5 bases. For whole genome, the analysis revealed a mean DP (coverage) of 1.23 (min: 0.53, max: 1.72, IQR: 0.48) for CD8+ samples and a mean DP of 1.38 (min: 0.87, max: 1.72, IQR: 0.18) for CD4+ samples. For the selected genome regions, a mean DP of 355 (min: 105, max: 490, IQR: 146) for CD8+ samples and a mean DP of 373 (min: 189, max: 472, IQR: 120) for CD4+ samples. The mean coverage was obtained for any genomic position. The histogram illustrates the fraction of bases covered with a given or higher read depth in CD4+ and CD8+ samples. Colours correspond to the different read depth bins. (A) Sequencing coverage for whole genome and (B) sequencing coverage for the selected genome regions. CDF, Cumulative Distribution Function; DP, Coverage (the number of reads). The histogram was drawn using R (Version: 4.0.3).

Supplementary Table S1. Variants in HTLV-2 positive subjects' CD8+ T cells detected by immunopanel sequencing

HTLV2+	Cells	Gene	Chr	Position	Ref	Var	ExonicFunc.refGene	Amino acid change	VAF (%)
HTLV2+_1	CD8+	PLEKHG5	1	6473093	G	A	nonsynonymous SNV	R370C	3.1
HTLV2+_1	CD8+	MPL	1	43340080	G	A	stopgain	W269X	2.4
HTLV2+_1	CD8+	NR1H4	12	100534961	G	A	nonsynonymous SNV	V224M	2.3
HTLV2+_1	CD8+	IL17D	13	20721779	G	A	nonsynonymous SNV	R145H	3.5
HTLV2+_1	CD8+	DIDO1	20	62906068	T	A	nonsynonymous SNV	R469S	4.9
HTLV2+_1	CD8+	EP300	22	41152296	C	T	nonsynonymous SNV	P1030S	3.2
HTLV2+_1	CD8+	FGF18	5	171456597	C	T	nonsynonymous SNV	T139M	2.3
HTLV2+_1	CD8+	MAPK14	6	36028224	C	T	nonsynonymous SNV	R23C	2.6
HTLV2+_1	CD8+	GPSM1	9	136336066	C	T	nonsynonymous SNV	R131W	3.3
HTLV2+_2	CD8+	KRAS	12	25245348	C	A	nonsynonymous SNV	G13C	2.8
HTLV2+_2	CD8+	ASXL1	20	32433553	G	T	nonsynonymous SNV	G452V	2.1
HTLV2+_3	CD8+	BCL10	1	85270774	T	C	nonsynonymous SNV	R64G	3.2
HTLV2+_3	CD8+	SLC19A2	1	169477691	C	A	nonsynonymous SNV	A91S	3.6
HTLV2+_3	CD8+	STAP2	19	4327371	C	A	nonsynonymous SNV	R202L	4.3
HTLV2+_3	CD8+	FYB	5	39134954	T	C	nonsynonymous SNV	K526E	3.7
HTLV2+_5	CD8+	NT5E	6	85450158	C	G	nonsynonymous SNV	R7G	25.5
HTLV2+_11	CD8+	NCOR2	12	124335503	T	C	nonsynonymous SNV	E2082G	2.0
HTLV2+_11	CD8+	MTA1	14	105445493	C	G	nonsynonymous SNV	L58V	2.6
HTLV2+_11	CD8+	BCL6	3	187729243	G	A	stopgain	Q388X	6.6
HTLV2+_11	CD8+	ADCY8	8	131039574	T	C	nonsynonymous SNV	T254A	2.6
HTLV2+_11	CD8+	RPS6KA3	X	20167701	T	C	nonsynonymous SNV	K497R	8.7
HTLV2+_12	CD8+	CD248	11	66315502	G	A	nonsynonymous SNV	P509L	2.4
HTLV2+_12	CD8+	PLA2R1	2	160016666	A	T	nonsynonymous SNV	I500N	2.4
HTLV2+_12	CD8+	INPP5D	2	233147580	A	G	nonsynonymous SNV	Y338C	6.6
HTLV2+_12	CD8+	P4HTM	3	48990489	T	C	nonsynonymous SNV	L78P	5.3
HTLV2+_12	CD8+	FSCN1	7	5593579	G	C	nonsynonymous SNV	E215Q	9.1
HTLV2+_13	CD8+	VWF	12	6018949	G	C	nonsynonymous SNV	P1490R	14.2
HTLV2+_13	CD8+	KMT2D	12	49026710	G	A	stopgain	R5086X	4.0
HTLV2+_13	CD8+	NFKBIA	14	35403707	G	A	stopgain	Q107X	7.9
HTLV2+_13	CD8+	CTCF	16	67629413	C	G	nonsynonymous SNV	H573D	15.0
HTLV2+_13	CD8+	PIK3R5	17	8888723	T	C	nonsynonymous SNV	H355R	8.2
HTLV2+_13	CD8+	STAT3	17	42322464	T	A	nonsynonymous SNV	Y640F	16.2
HTLV2+_14	CD8+	PSAP	10	71831202	G	A	nonsynonymous SNV	P100L	2.8
HTLV2+_14	CD8+	COLEC12	18	346954	A	T	nonsynonymous SNV	L223Q	5.0
HTLV2+_14	CD8+	VCAN	5	83540916	G	C	nonsynonymous SNV	G2638A	4.2
HTLV2+_14	CD8+	MXRA5	X	3318004	C	A	splicing		2.4
HTLV2+_15	CD8+	IL6ST	5	55954876	C	G	nonsynonymous SNV	D462H	4.8
HTLV2+_17	CD8+	FANCA	16	89762005	C	A	nonsynonymous SNV	W932C	3.7
HTLV2+_17	CD8+	FOXN1	17	28524926	G	A	nonsynonymous SNV	G183R	2.9
HTLV2+_17	CD8+	SHANK1	19	50703758	G	-	frameshift deletion	P432fs	8.1
HTLV2+_17	CD8+	C6	5	41176516	T	C	nonsynonymous SNV	D376G	2.3
HTLV2+_19	CD8+	IFNAR1	21	33352806	C	G	nonsynonymous SNV	P398A	3.0
HTLV2+_19	CD8+	LRBA	4	150849456	A	C	nonsynonymous SNV	I1375R	3.4
HTLV2+_20	CD8+	VWF	12	5971682	C	T	nonsynonymous SNV	E2489K	6.0
HTLV2+_20	CD8+	IL20RA	6	137011387	A	T	nonsynonymous SNV	L97H	5.1
HTLV2+_21	CD8+	TNFRSF18	1	1203945	G	A	nonsynonymous SNV	P202S	4.8
HTLV2+_21	CD8+	GBP6	1	89385456	A	G	nonsynonymous SNV	K630R	5.9
HTLV2+_21	CD8+	CD3D	11	118339455	T	C	nonsynonymous SNV	Y149C	4.0
HTLV2+_21	CD8+	VWF	12	5952401	A	T	nonsynonymous SNV	L2702Q	7.0
HTLV2+_21	CD8+	CAMK1	3	9767666	C	T	splicing		4.1
HTLV2+_21	CD8+	ABCA13	7	48520274	G	A	stopgain	W4677X	5.8
HTLV2+_22	CD8+	ETNK1	12	22625222	C	A	nonsynonymous SNV	A20D	4.9
HTLV2+_22	CD8+	TYRO3	15	41569016	C	T	nonsynonymous SNV	R416C	4.0
HTLV2+_22	CD8+	KIF23	15	69444826	C	T	nonsynonymous SNV	R806C	5.6
HTLV2+_22	CD8+	MXRA5	X	3322805	T	G	stopgain	Y960X	4.4
HTLV2+_23	CD8+	SMAD7	18	48948425	G	-	frameshift deletion	P209fs	2.6
HTLV2+_24	CD8+	CACNB2	10	18150935	G	A	nonsynonymous SNV	S58N	8.4
HTLV2+_24	CD8+	LRP1	12	57201154	G	A	splicing		2.7
HTLV2+_24	CD8+	NBEA	13	35667500	G	T	nonsynonymous SNV	R2843L	2.5
HTLV2+_24	CD8+	BCL11B	14	99257729	T	A	nonsynonymous SNV	T57S	2.0
HTLV2+_24	CD8+	TUFM	16	28845004	C	T	nonsynonymous SNV	A156T	5.6
HTLV2+_24	CD8+	MMP2	16	55488695	T	G	nonsynonymous SNV	Y329D	8.7
HTLV2+_24	CD8+	SIGLEC11	19	49958319	C	T	nonsynonymous SNV	G539R	3.2
HTLV2+_26	CD8+	SMAD7	18	48949961	A	G	nonsynonymous SNV	L155P	2.1
HTLV2+_27	CD8+	STAT3	17	42322402	C	A	nonsynonymous SNV	D661Y	4
HTLV2+_28	CD8+	RTEL1	20	63674075	G	A	nonsynonymous SNV	A325T	2.2

Immunopanel sequencing was performed on CD8+ T cells from 28 HTLV-2 positive subjects. Two HTLV-2 positive subjects had somatic *STAT3* mutations (Y640F and D661Y) as discovered in deep amplicon seuquencing. *STAT3* mutations (N647I and Y657_K658insY) detected in amplicon sequencing did not pass all filtering in immunopanel sequencing analysis due to lower coverage. Chr, Chromosome; Ref, reference base; Var, variant base; ExonicFunc.refGene, exonic variant function; VAF, variant allele frequency; Freq_Ratio.