SUPPLEMENTARY APPENDIX

The analysis of clonal diversity and therapy responses using *STAT3* mutations as a molecular marker in large granular lymphocytic leukemia

Hanna L. M. Rajala,[1] Thomas Olson,[2] Michael J. Clemente,[3] Sonja Lagström,[4] Pekka Ellonen,[4] Tuija Lundan,[5] David E. Hamm,[6] Syed Arshi Uz Zaman,[4] Jesus M. Lopez Marti,[4] Emma I. Andersson,[1] Andres Jerez,[3,7] Kimmo Porkka,[1] Jaroslaw P. Maciejewski,[3] Thomas P. Loughran,[2] and Satu Mustjoki[1]

[1]Hematology Research Unit, Department of Hematology, University of Helsinki and Helsinki University Central Hospital Cancer Center, Helsinki, Finland; [2]University of Virginia Cancer Center, Charlottesville, VA, USA; [3]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA; [4]Institute for Molecular Medicine (FIMM), University of Helsinki, Finland; [5]Department of Clinical Chemistry and TYKSLAB, University of Turku and Turku University Central Hospital, Finland; [6]Adaptive Biotechnologies Corp, Seattle, WA, USA; and [7]Hematology and Medical Oncology Department, Hospital Universitario Morales Meseguer, Universidad de Murcia, IMIB-Arrixaca, Murcia, Spain

 doi:10.3324/haematol.2014.113142
Manuscript received on July 8, 2014. Manuscript accepted September 29, 2014.
Correspondence: satu.mustjoki@helsinki.fi

The analysis of clonal diversity and therapy responses using *STAT3* mutations as a molecular marker in large granular lymphocytic leukemia

Supplementary appendix

Table of contents

Supplementary methods 2
Deep targeted sequencing and capillary sequencing of *STAT3* exon 21 2
High-throughput T cell receptor sequencing 2
Allele-specific oligonucleotide real-time quantitive PCR 3
Supplementary tables 4
Supplementary table 1 4
Supplementary table 2 5
Supplementary table 3 6
Supplementary table 4 7
Supplementary table 5 8
Supplementary table 6 9
Supplementary table 7 10
Supplementary references 11
Supplementary figures 12
Supplementary Figure 1 12
Supplementary Figure 2 13

Supplementary methods

Deep targeted sequencing and capillary sequencing of *STAT3* exon 21

STAT3 exon 21 was sequenced using Illumina Miseq Platform as described previously [1]. Each amplicon was amplified in multiplexed PCR reaction containing locus-specific PCR primers carrying Illumina compatible adapter sequences, Illumina adapter primers carrying Illumina P5/P7 grafting sequences and index sequences, resulting in Illumina -compatible paired-end sequencing templates. The number of PCR cycles was 30, and the input of DNA was 10-20 ng, although it could be less in the case of sorted lymphocyte subpopulations. The genome equivalent, 1ng DNA≈330 genomes, corresponds to 165 cells with dlploid genomes. The DNA input of 10-20ng equals 3300-6600 genomes, meaning that theoretically all of these would be sequenced once when coverage is over 3300-6600. PCR amplicons were sequenced as 151 base pair (bp) or 251 bp paired end reads and two 8 bp index reads using the Illumina MiSeq instrument (Illumina, San Diego, CA, US). Primer sequences are in Supplementary Table 6.

The data was analyzed using an in-house bioinformatics pipeline developed specifically for reporting somatic variants from amplicon sequencing data. Low-quality reads were not filtered out before alignment: after alignment, quality values (i.e. Phred scores) were used to exlude error bases with low quality from further analysis. First, previously reported *STAT3* mutations were verified in all *STAT3*-mapped amplicons: all variants with variant allele count over 5 and variant allele frequency (VAF) over 0.5% were taken into consideration. Second, all variants with variant allele count over 5 and variant allele frequency (VAF) over 0.5% were called from *STAT3*-mapped amplicons. From these candidates false positives were initially filtered out based on the noise (estimated error rate) level from control sample in every run making use of a binomial distribution to compute p-value for the event that more than the frequency of alternative alleles were observed when the null hypothesis is true. However, variants with VAF over 2% were called independent of the noise. A specific frequency ratio was used to filter out false positive by dividing the ratio of variant calls/number of all the bases (at a position) by the ratio of variant allele quality sum/quality sum of all the bases. All samples with a frequency ratio ≥0.80 were considered to be true mutations, and the variants from both scripts with a borderline frequency ratio between 0.75-0.79 were verified with Integrative Genomics Viewer (IGV, Broad Institute, Cambridge, UK).

High-throughput T cell receptor sequencing

TCRB complementarity determining regions (CDR3β) were amplified and sequenced by Adaptive Biotechnologies Corp (Seattle, WA, US) using the ImmunoSEQ assay. A multiplex PCR system was used to amplify CDR3β sequences from flow cytometry-sorted T-LGL leukemia DNA samples using 52 forward primers for the Vβ gene segment and 13 reverse primers for the Jβ segment. This approach generates a 60 base-pair fragment capable of identifying the VDJ region spanning each unique CDR3β [2]. Amplicons were sequenced using the Illumina HiSeq platform (Illumina). Using a baseline developed from a suite of synthetic templates, primer concentrations and computational corrections are used to correct for the primer bias common to multiplex PCR reactions. Raw sequence data was filtered based on the

TCRβ V, D and J gene definitions provided by the IMGT database (www.imgt.org) and binned using a modified nearest-neighbor algorithm to merging closely related sequences and remove both PCR and sequencing errors. Data was analyzed using the ImmunoSEQ analyzer toolset (Adaptive Biotechnologies). The analysis was performed using the Survey level of the ImmunoSEQ assay, which is capable of detecting one cell in 40,000 T cells [3].

Allele-specific oligonucleotide real-time quantitive PCR

In real-time PCR the mutational load was quantified in relation to a known *STAT3* Y640F or D661V positive reference sample with high variant allele frequency (VAF) in amplicon sequencing. The reference sample was used for preparing a log-linear dilution series with pooled normal buffy coat DNA and a standard curve was constructed based on the real-time PCR results. The mutational load of samples examined was quantified from the standard curve and data was normalized with an albumin reference gene qPCR assay. ASO-qPCR primer sequences are in Supplementary Table 7.

Supplementary tables

Supplementary table 1

Comparison of ASO-PCR and amplicon sequencing of *STAT3* mutations Y640F and D661V. ASO-PCR results are reported in relation to the positive control sample,and the normalized sensitivity level varied between 0.03%-0.1%. Patient number in brackets refers to Figure 1.

Patient	STAT3 amplicon result, VAF [1]	Y640F ASO-PCR	D661V ASO-PCR
Positive control 1 (Patient 57)	Y640F, 27%	Pos	Neg
Positive control 2 (Patient 9)	D661V, 37%, Y640F, 2%	Pos (3% of Control 1)	Pos
Test 1 (Patient 54)	Y640F, 30%	Pos (75% of Control 1)	Neg (<0.05% of Control 2)
Test 2 (Patient 7)	Y640F, 2%, D661V, 5%	Pos (1% of Control 1)	Pos (1.5% of Control 2)
Test 3 (Patient 14)	D661V, 15%	Pos (0.03% of Control 1)	Pos (35% of Control 2)
Test 4 (Patient 21)	I659L, 5%	Neg (<0.03% of Control 1)	Neg (<0.1% of Control 2)
Test 5 (Not in Figure 1)	No mutations	Neg (<0.06% of Control 1)	Neg (<0.08% of Control 2)
Test 6 (Not in Figure 1)	No mutations	Neg (<0.06% of Control 1)	Neg (<0.08% of Control 2)

[1]Abbreviations: VAF, variant allele frequency; pos, positive; neg, negative.

Supplementary table 2

Dilution series and *STAT3* amplicon sequencing analysis of a CD8+ T-LGL sample with *STAT3* D661V (chr17:40474419 T>A) and Y640F (chr17:40474482 T>A) mutations. The in-house bioinformatics pipeline does not filter low quality reads prior to alignment, but uses quality values (QV, i.e. Phred scores) to exlude low quality variants (see the methods section of Supplementary appendix). Low average QVs (under 20) are marked in red and those variants are considered as errors.

The D661V mutation is under detection limit with the dilution of 1:100, at which point the number of error calls with low QV probably interferes with the detection of the variant. The Y640F mutation with lower VAF is detected with the dilution of 1:2. Frequency ratio under 0.8 and QVs under 20 are marked in red.

Mutation	Dilution[1]	Freq.ratio[2]	Total depth[3]	**T**	T, QV[4]	**A**	A, QV[4]	G	G, QV[4]	C	C, QV[4]	N[4]	N, QV[4]
D661V	1:1	1,01	11878	**56,5%**	35	**37,3%**	33	1,0%	17	0,7%	15	4,5%	2
	1:2	1,01	5497	**78,0%**	35	**16,6%**	34	0,8%	15	0,3%	14	4,3%	2
	1:20	0,95	8132	**93,6%**	36	**1,6%**	32	0,4%	16	0,2%	17	4,2%	2
	1:50	0,90	35526	**94,3%**	36	**0,6%**	31	0,4%	17	0,2%	16	4,4%	2
	1:100	0,79	26770	**94,5%**	36	**0,4%**	27	0,4%	17	0,2%	16	4,4%	2
Y640F	1:1	0,98	15787	**97,7%**	36	**1,4%**	35	0,6%	18	0,3%	17	0,1%	2
	1:2	0,86	7207	**98,5%**	35	**0,6%**	30	0,6%	18	0,3%	16	0,0%	2
	1:20	0,71	10643	**99,0%**	36	**0,3%**	25	0,5%	19	0,2%	16	0,0%	2
	1:50	0,53	46134	**98,9%**	36	**0,2%**	19	0,5%	18	0,3%	17	0,0%	2
	1:100	0,55	34932	**99,0%**	36	**0,2%**	19	0,4%	18	0,3%	17	0,1%	2

Abbreviations: Freq, frequency; N, unrecognized base; QV, quality value

[1]Dilution series made from a sample with both *STAT3* D661V and Y640F mutations.
[2]The calculation of frequency ratio is explained in the Methods section of this manuscript.
[3]Number of reads mapped to *STAT3* exon 21.
[4]The average QV (Phred score) of bases: when FASTQ files are formed, a non-negative quality value (QV) is assigned to each called base using a logged transformation of the error probability. For example, a QV of 30 means that the error probability is 0.1%, whereas a QV of 20 equals 1% error rate. A base with QV under 20 is widely considered as a sequencing error.
[4]Unrecognized bases with low quality values.

Supplementary table 3

Amplicon sequencing coverage and VAFs of *STAT3* mutations (VAF of the largest variant was used in the analysis in case of multiple mutations).

	VAF, median (range)	All mapped reads, median (range)[1]	Coverage at the variant allele location[2]
All amplicon sequencing positive samples (n=82)	17% (0.6-51)	14196 (840-565170)	4103 (264-94038)
Positive by capillary sequencing (n=50)	27% (9-51)	14472 (840-565170)	4060 (264-94038)
Positive only by amplicon sequencing (n=32)	3% (0.6-25)	12616 (1255-309580)	5281 (395-21839)
All amplicon sequencing-negative samples (n=131)	NA	6931 (1270-804232)	NA

[1]The total number of amplicon reads aligned to *STAT3* exon 21 (both forward and reverse reads).
[2]The number of amplicon reads at the location of the variant allele. Most of the *STAT3* mutations detected are in the area covered only by the reverse reads, thus the coverage is approximately half of the total number of *STAT3*-aligned reads.
Abbreviations: NA, not applicable; VAF, variant allele frequency.

Supplementary table 4

The number of TCRB CDR3 sequencing reads. Patient numbers are from Figure 1.

Patient	Lymphocyte fraction	Total number of reads	Productive sequences	Unique sequences
8	CD8+Vb3+	2515757	2466380	1134
	CD3+Vb3neg	2493453	2417798	869
48	CD8+Vb21.3+	2475907	2418665	956
	CD8+Vb21.3+neg	2575304	2226166	14386
14	CD8+Vb17+	2768966	2687142	598
	CD3+	2908398	2465234	9777

Supplementary table 5

Deep TCRB sequencing results of three T-LGL leukemia patients.

Patients[1]	STAT3 mutations (VAF)	CRD3 nucleotide sequence	CDR3 amino acid sequence	Frequency (%)	Count	V-gene name	D-gene name	J-gene name	Healthy[3]
Patient 8, Vb3+ fraction (TCRBV 28) (Fig. 4A)	Y640F 45%	GCCAGCACCAACCAGACATCTATGTACCTCTGTGCCAGCAGTTCCCTTCGCTCAGGGCCCATGAACACTGAAGCTTTCTTTGGACAA	CASSSLRSGPMNTEAFF	73	1834805	TCRBV28-01	TCRBD01-01	TCRBJ01-01	0/586 (0%)
		CAGCGCACACAGCAGGAGGACTCCGCCGTGTATCTCTGTGCCAGCAGCTTAGCGTGGGGCATAAATTCACCCCTCCACTTTGGGAAC	CASSLAWGINSPLHF	13	337759	TCRBV07-08	TCRBD01-01	TCRBJ01-06	1/586 (0.2%)
Patient 8, Vb3^{neg} fraction (Fig. 4A)	I659L 30%	CAGCGCACACAGCAGGAGGACTCCGCCGTGTATCTCTGTGCCAGCAGCTTAGCGTGGGGCATAAATTCACCCCTCCACTTTGGGAAC	CASSLAWGINSPLHF	74	1844935	TCRBV07-08	TCRBD01-01	TCRBJ01-06	1/586 (0.2%)
Patient 48, Vb21.3+ (TCRBV 11-02) fraction (Fig. 4B)	N647I 34%	CAGCCTGCAAAGCTTGAGGACTCGGCCGTGTATCTCTGTGCCAGCAGCTTAAGAGCGGGGGGGCCCAATGAGCAGTTCTTCGGGCCA	CASSLRAGGPNEQFF	64	1572352	TCRBV11-02	TCRBD02-01	TCRBJ02-01	6/586 (1%)
		GTGACCAGTGCCCATCCTGAAGACAGCAGCTTCTACATCTGCAGTGCTAGCGACCGCAATAGCAATCAGCCCCAGCATTTTGGTGAT	CSASDRNSNQPQHF	26	647660	TCRBV20-01	TCRBD01-01	TCRBJ01-05	33/586 (6%)
Patient 48, Vb21.3^{neg} fraction (Fig. 4B)	Y640F 4%, D661V 0.8%, D661Y 0.5%	TCGGCTGCTCCCTCCCAAACATCTGTGTACTTCTGTGCCAGCAGTCGGACAGGGATCCTAGCCAAAAACATTCAGTACTTCGGCGCC	CASSRTGILAKNIQYF	4	97873	TCRBV06-02 or TCRBV06-03	TCRBD01-01	TCRBJ02-04	4/586 (0.7%)
		GTCCACGGAGTCAGGGGACACAGCACTGTATTTCTGTGCCAGCAGCAAAGTGGGTCGACAGGGGACGGAAAACTGTTTTTTGGCAGT	Unproductive TCRB rearrangement	3	67700	TCRBV21-01	TCRBD01-01	TCRBJ01-04	NA
		CTGCAGCCAGAAGACTCGGCCCTGTATCTCTGCGCCAGCAGCCAATTGGACAGGGTGTATGGGGCCAACGTCCTGACTTTCGGGGCC	CASSQLDRVYGANVLTF	2	58461	TCRBV04-03	TCRBD01-01	TCRBJ02-06	0/586 (0%)
		GCCCTGCAGCCAGAAGACTCAGCCCTGTATCTCTGCGCCAGCAGCCAGACTAGCGGGGGGGAATACAATGAGCAGTTCTTCGGGCCA	CASSQTSGGEYNEQFF	2	42250	TCRBV04-01	TCRBD02-01	TCRBJ02-01	6/586 (1%)
Patient 14, Vb17+ (TCRBV 19) fraction (Supplementary Fig. 1)	D661V 41%	TCGGCCCAAAAGAACCCGACAGCTTTCTATCTCTGTGCCAGTAGTCCCAGGACAGGGTATAGCAATCAGCCCCAGCATTTTGGTGAT	CASSPRTGYSNQPQHF	83	2307767	TCRBV19-01	TCRBD01-01	TCRBJ01-05	114/586 (20%)
Patient 14, CD3+ fraction, contains Vb13.6+ cells (TCRBV 06-06) (Supplementary Fig. 1)	D661V 8%[2]	TCGGCCCAAAAGAACCCGACAGCTTTCTATCTCTGTGCCAGTAGTCCCAGGACAGGGTATAGCAATCAGCCCCAGCATTTTGGTGAT	CASSPRTGYSNQPQHF	10	300013	TCRBV19-01	TCRBD01-01	TCRBJ01-05	114/586 (20%)
		AGGCTGGAGTTGGCTGCTCCCTCCCAGACATCTGTGTACTTCTGTGCCAGCAGCCCCCTAGGGTTGGAGACCCAGTACTTCGGGCCA	CASSPLGLETQYF	4	124889	TCRBV06-06	TCRBD01-01	TCRBJ02-05	71/586 (12%)

[1]Patient numbers are from Figure 1. The TCRBV gene name according to the International ImMunoGeneTics information system (IMGT) nomenclature is written in brackets after each Vbeta clone.
[2]D661Y mutation seen in the smaller Vb13.6+ expansion was seen with manual inspection in the CD3+ fraction with VAF of 0.45%, which is slightly under the detection treshold of the amplicon bioinformatics pipeline.
[3]The existence of the amino acid sequence in a cohort of 586 healthy individuals, provided by the Adaptive Biotechnologies company.
Abbreviations: CDR3, complementarity determining region 3; D, diversity; J, joining; V, variable; VAF, variant allele frequency

Supplementary table 6

Locus specific amplicon primer sequences carrying tails corresponding to the Illumina adapter sequences. Locus specific primer sequences are underlined.

Primer name	Primer sequence
STAT3_exon21-F	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCCAAAAATTAAATGCCAGGA-3'
STAT3_exon21-R	5'-AGACGTGTGCTCTTCCGATCTGGTTCCATGATCTTTCCTTCC-3'

Supplementary table 7

STAT3 Y640F- and D661V-specific primer sequences used in ASO-PCR.

Primer/Probe name	Primer / Probe sequence
Y640F fwd	5′-ATCCAGTCCGTGGAACCcTT-3′
D661V fwd	5′-TCATCATGGGCTATAAGATCAaGGT-3′
STAT3x21-Rev	5′-TCTCTGGCCGACAATACTTTCC-3′
STAT3x21-Pr	FAM- CCTGACATTCCCAAGGAGGAGGCA –BHQ

Supplementary references

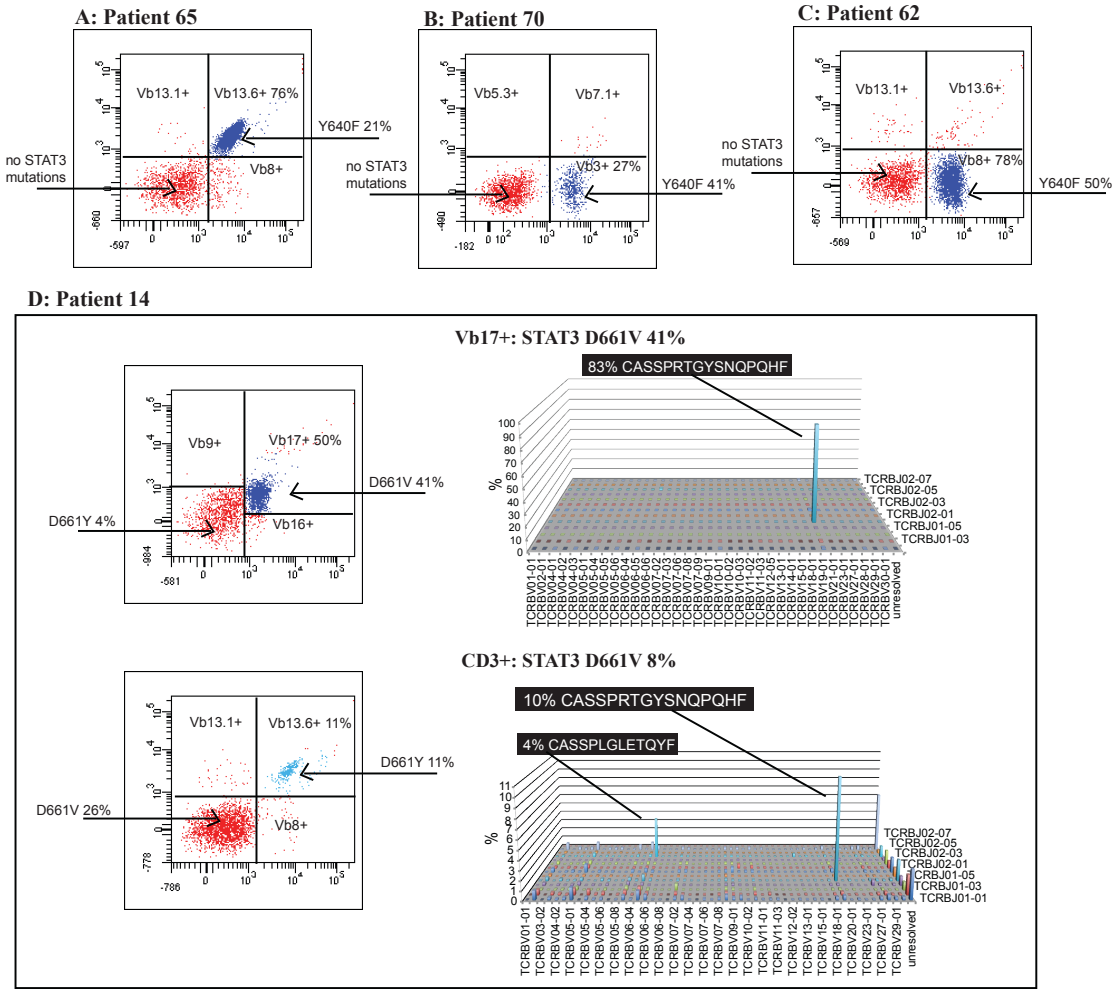
1. Rajala HL, Eldfors S, Kuusanmaki H, van Adrichem AJ, Olson T, Lagstrom S, et al. Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia. *Blood*. 2013;**121**(22):4541-50.

2. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009;**114**(19):4099-107.

3. Robins H, Desmarais C, Matthis J, Livingston R, Andriesen J, Reijonen H, et al. Ultra-sensitive detection of rare T cell clones. *J Immunol Methods*. 2012;**375**(1-2):14-9.

Supplementary figures

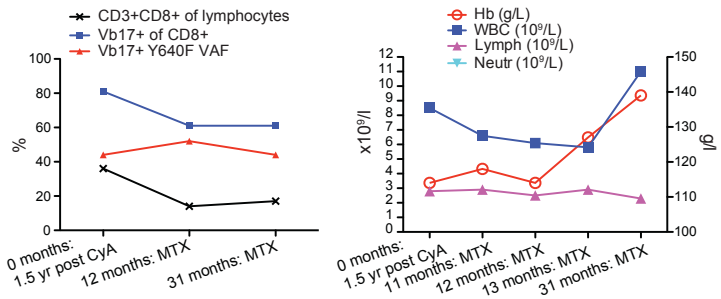
Supplementary Figure 1



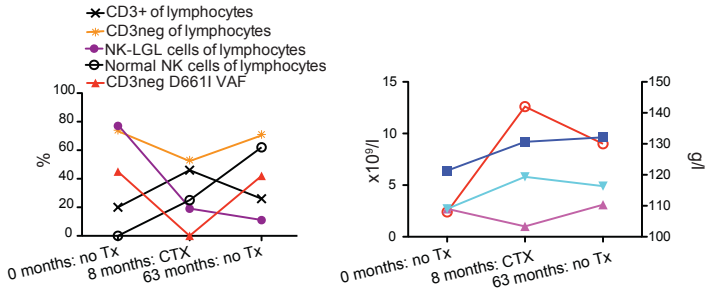
Supplementary Figure 1. Results of STAT3 amplicon sequencing from sorted lymphocyte populations.
(A-C) STAT3 sequencing results from three monoclonal T-LGL leukemia patients. (D) STAT3 sequencing results of patient 14, who had two Vbeta expansions, Vb17+ (50% of CD8+ cells) and Vb13.6+ (11% of CD8+). Deep TCRB sequencing was done from Vb17+ cells, and the percentage of largest clone (83%) was in good accordance with the VAF of D661V mutation (VAF 41%). Whole CD3+ fraction was sequenced in parallel, and it contained the D661Y-mutated Vb13.6+ clone (4% of reads) in addition to the immunodominant expansion (10% of reads): the percentage of immunodominant Vb17+ cells from total CD3+ fraction was 18% in the Vbeta analysis. If all unproductive sequences are excluded from the TCRB analysis, the proportion of the TCRBV19-01 (Vb17+) clone is 12% and matches with Vbeta results.

Supplementary Figure 2

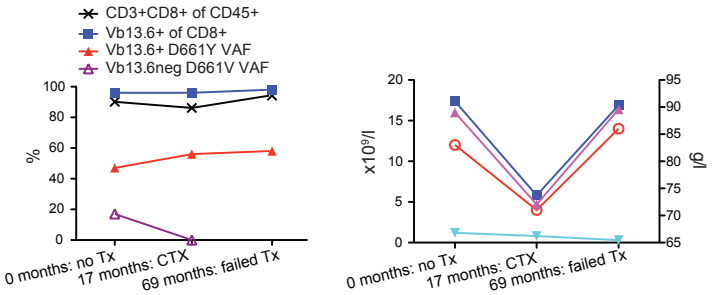
A: Patient 69, MTX



B: Patient 1, CTX



C: Patient 32, CTX



Supplementary Figure 2. STAT3 mutation analysis and lymphocyte populations at different timepoints during methotrexate (MTX) or cyclophosphamide (CTX) treatment.

(A) Patient 69 was treated with methotrexate, and although the patient achieved partial response (improved Hb values), the size of the *STAT3*-mutated clone dimished only slightly.

(B) Patient 1 had aberrant CD3negCD16/56neg NK-LGL expansion, which was D661I-mutated. NK-LGL cells were suppressed and normal NK cells were restored during CTX treatment.

(C) Patient 32 was treated with CTX, and failed to respond to treatment. CTX administration was stopped because of developing cytopenias, and leukemic D661Y-mutated clone (over 90% of CD8+) cells was unchanged. The amount of Vb13.6neg cells was too low for sorting at the last timepoint.

Abbreviations: CTX, cyclophosphamide; CyA, cyclosporine; MTX, methotrexate; Tx, therapy; VAF, variant allele frequency