LYMPHOID NEOPLASIA

Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia

Hanna L. M. Rajala,[1] Samuli Eldfors,[2] Heikki Kuusanmäki,[2] Arjan J. van Adrichem,[2] Thomas Olson,[3] Sonja Lagström,[2] Emma I. Andersson,[1] Andres Jerez,[4] Michael J. Clemente,[4] Yiyi Yan,[5] Dan Zhang,[3] Andy Awwad,[3] Pekka Ellonen,[2] Olli Kallioniemi,[2] Krister Wennerberg,[2] Kimmo Porkka,[1] Jaroslaw P. Maciejewski,[4] Thomas P. Loughran Jr,[3] Caroline Heckman,[2] and Satu Mustjoki[1]

[1]Hematology Research Unit Helsinki, Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; [2]Institute for Molecular Medicine Finland, Helsinki, Finland; [3]Penn State Hershey Cancer Institute, Pennsylvania State College of Medicine, Hershey, PA; [4]The Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH; and [5]York Hospital, Department of Medicine, York, PA

Key Points

- Somatic mutations were discovered for the first time in the SH2 domain of the *STAT5b* gene in LGL leukemia.
- The mutations are activating and lead to increased phosphorylation and transcriptional activity of STAT5b.

Large granular lymphocytic (LGL) leukemia is characterized by clonal expansion of cytotoxic T cells or natural killer cells. Recently, somatic mutations in the signal transducer and activator of transcription 3 (*STAT3*) gene were discovered in 28% to 40% of LGL leukemia patients. By exome and transcriptome sequencing of 2 STAT3 mutation-negative LGL leukemia patients, we identified a recurrent, somatic missense mutation (Y665F) in the Src-like homology 2 domain of the *STAT5b* gene. Targeted amplicon sequencing of 211 LGL leukemia patients revealed 2 additional patients with *STAT5b* mutations (N642H), resulting in a total frequency of 2% (4 of 211) of *STAT5b* mutations across all patients. The Y665F and N642H mutant constructs increased the transcriptional activity of STAT5 and tyrosine (Y694) phosphorylation, which was also observed in patient samples. The clinical course of the disease in patients with the N642H mutation was aggressive and fatal, clearly different from typical LGL leukemia with a relatively favorable outcome. This is the first time somatic *STAT5* mutations are discovered in human cancer and further emphasizes the role of STAT family genes in the pathogenesis of LGL leukemia. (*Blood*. 2013;121(22):4541-4550)

Introduction

Large granular lymphocytic (LGL) leukemia is a chronic lymphoproliferative disorder characterized by the expansion of $CD3^+CD8^+CD57^+$ cytotoxic T cells or $CD3^{neg}CD16/56^+$ natural killer (NK) cells.[1-3] It is often associated with severe cytopenias and other comorbidities such as rheumatoid arthritis, all thought to be autoimmune-based and mediated by cytotoxic LGL lymphocytes.[4] Leukemic LGLs are resistant to Fas-mediated apoptosis, and both interleukin-15 (IL-15) and platelet-derived growth factor receptor have been shown to promote their survival.[5]

In a previous study, we showed that 40% of T cell LGL (T-LGL) leukemia patients harbored a somatic signal transducer and activator of transcription 3 (*STAT3*) mutation in the Src-like homology 2 (SH2) domain of STAT3.[6] The amino acid changes increased the phosphorylation and transcriptional activity of STAT3, suggesting that in 40% of cases a mutated *STAT3* gene causes aberrant STAT3 signaling seen in LGL leukemia. A similar spectrum of *STAT3* mutations and prevalence of 30% were also verified in NK-LGL leukemia.[7] However, it should be noted that the expression of many STAT3 target genes was also increased in *STAT3* mutation-negative LGL leukemia cases, suggesting that in these patients other mechanisms resulting in the activation of the Janus kinase–STAT pathway may operate.[6]

To understand the molecular pathogenesis of LGL leukemia and discover the genetic basis of disease development in patients without a *STAT3* mutation, we performed whole exome sequencing of 2 T-LGL leukemia patients lacking a *STAT3* mutation, and validated the results in a large cohort of T- and NK-LGL leukemia patients.

Methods

Study patients

The study was undertaken in compliance with the principles of the Helsinki declaration and was approved by the ethics committees in the Helsinki University Central Hospital (Helsinki, Finland), the Cleveland Clinic (Cleveland, OH), and the Penn State Hershey Cancer Institute (Hershey, PA). All patients and healthy controls gave written informed consents.

The study population consisted of 211 patients with LGL leukemia: 173 patients had T-LGL and 38 patients NK-LGL leukemia. One hundred one

Submitted December 20, 2012; accepted April 7, 2013. Prepublished online as *Blood* First Edition paper, April 17, 2013; DOI 10.1182/blood-2012-12-474577.

S.E., H.K., and A.J.v.A. contributed equally to this study.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

samples were from the Penn State Hershey Cancer Institute, 87 from the Cleveland Clinic, and 23 from Finland. All patients met the criteria of LGL leukemia as defined by the World Health Organization (2008). The main characteristics of the patients are summarized in supplemental Table 1 (available on the *Blood* website).

Sample collection and DNA/RNA extraction

Mononuclear cells (MNCs) were separated from peripheral blood (PB) using Ficoll gradient separation (GE Healthcare), and $CD4^+$ and $CD8^+$ T cells were further separated with magnetic bead separation: PB MNCs were labeled either with CD4 or CD8 MicroBeads (Miltenyi Biotec) according to the instructions of the manufacturer and separated with an AutoMACS cell sorter (Miltenyi Biotec). The purity of the sorted fractions was confirmed by flow cytometry (FACSAria; Becton Dickinson). In NK-LGL cases, whole blood or the MNC fraction was used for analysis.

Genomic DNA was isolated from fresh or frozen sorted MNCs or from whole blood samples using the Genomic DNA NucleoSpin Tissue kit (Macherey-Nagel). RNA was extracted using the miRNAeasy kit (QIAGEN). DNA and RNA concentration and purity were measured with the Nanodrop (Thermo Fisher), Qubit 2.0 Fluorometer (Life Technologies), and Agilent 2100 Bioanalyzer (Agilent Technologies).

TCR Vβ analysis

T-cell receptor (TCR) Vβ families were analyzed from T-LGL leukemia patients' whole blood or PB MNC samples by combining CD4 and CD8 antibodies with the panel of TCR Vβ antibodies corresponding to 24 members of variable regions of the TCR β chain (~70% coverage of normal human TCR Vβ repertoire) (Beckman-Coulter Immunotech). Samples were analyzed using FACSAria and FACSDiva Software (Becton Dickinson).

Exome sequencing

Exome sequence library preparation, Agilent exome capture, and the bioinformatics pipeline for identification and annotation of candidate somatic mutations were performed as previously described.[6]

Microarray analysis

We used the Illumina Human HT-12 v4 BeadChip expression array which targets >47 000 probes in the human genome (cover content from NCBU RefSeq Release 38 and legacy UniGene content). The data were read with an iScan instrument (Illumina) and primary analysis was done with Genome Studio software (version 2011.1; Illumina). Further normalization and analysis was performed with the Chipster open source platform.[8] Microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-1611.

Validation and screening of candidate somatic mutations by capillary sequencing

The candidate somatic mutations were validated with capillary sequencing, and specific primers were designed using the Primer-Blast search (National Center for Biotechnology Information; http://blast.ncbi.nlm.nih.gov/). Capillary polymerase chain reaction (PCR) products were either separated by gel electrophoresis and extracted from the gel using the QIAquick Gel Extraction kit (QIAGEN) or purified using the ExoSAP-IT for PCR Product Clean-Up kit (Affymetrix). The purified PCR products were then sequenced with the BigDye v.1.1 Cycle Sequencing kit and the ABI PRISM 3730xl DNA analyzer. Sequences were analyzed using Chromas Pro and the Blast search. The primer sequences are in supplemental Table 2.

Amplicon sequencing of STAT5b

Locus-specific primers were designed to cover STAT5a exon 17 and STAT5b exon 16 using Primer3 with user-defined parameters (http://primer3.wi.mit.edu/). Primers were designed for a maximum 300-nucleotide-long PCR amplicons. The locations of single-nucleotide polymorphisms (SNPs) were masked using dbSNP-build 135. Repeat masking was not performed. After designing the locus-specific primer sequences, sequence tails corresponding to the Illumina adapter sequences, 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT and 5′-AGACGTGTGCTCTTCCGATCT, were added to the 5′ end of the forward and reverse locus-specific primers, respectively (primer sequences in supplemental Table 3). All oligonucleotides were ordered from Sigma-Aldrich. Sample preparation was performed according to an in-house targeted PCR amplification protocol. Each amplicon was amplified in multiplexed PCR-containing locus-specific PCR primers carrying Illumina-compatible adapter sequences, Illumina TruSeq Universal Adapter primer, and Illumina TruSeq Adapter primer with a sample-specific 6-bp index sequence (primer sequences in supplemental Table 4). The locus-specific primers are present in limiting quantities and are thus consumed in initial cycles of amplification forming intermediate amplification products. In subsequent amplification cycles, the adapter primers further amplify these intermediate products forming an Illumina-compatible paired-end sequencing template.

The PCR was performed in a volume of 20 μL containing 10 ng of sample DNA, 10 μL of 2× Phusion High-Fidelity PCR Master Mix (Thermo Scientific Inc), 0.025μM of each locus-specific primer, 0.5μM Illumina TruSeq Universal Adapter primer1, 0.5μM Illumina Truseq Adapter primer1, and the reaction mix was brought to a final volume with water (Sigma-Aldrich). Samples were cycled in a DNA Engine Tetrad 2 (Bio-Rad Laboratories) or G-Storm GS4 (Somerton) thermal cycler as follows: initial denaturation at 98°C for 30 seconds, 30 cycles at 98°C for 10 seconds, at 59°C for 30 seconds, and at 72°C for 15 seconds, final extension at 72°C for 10 minutes, followed by hold at 10°C. Following PCR amplification, samples were purified using a Performa V3 96-Well Short Plate (EdgeBio) and QuickStep 2 SOPE Resin (EdgeBio) and then pooled together without exact quantification. Purified sample pools were analyzed on an Agilent 2100 Bioanalyzer (Agilent Technologies) to quantify amplification performance and yield. Sequencing of PCR amplicons was performed using the Illumina MiSeq instrument with MiSeq Control Software (version 1.2; Illumina). Samples were sequenced as 151-bp paired-end reads and two 8-bp index reads using amplicon workflow. Data analysis was done on Illumina MiSeq Reporter Software version 1.3 or an earlier version.

STAT5b mutagenesis

An expression plasmid pCMV6-XL6 containing the wild-type coding sequence of *STAT5b* (OriGene) was modified to include the Y665F and N642H mutations. The mutant constructs were created using the GENEART Site-Directed Mutagenesis System (Invitrogen) according to the manufacturer's protocol. The primer sequences used for the mutagenesis are in supplemental Table 4. The mutations were confirmed by capillary sequencing of the plasmid DNA (primer sequences in supplemental Table 5).

STAT5b reporter assay

HeLa cells were plated and transiently transfected as described previously[6] with pCMV6-XL6 STAT5b plasmids (110 ng/well) and pGL4.52 (Luc2P/STAT5RE/Hygro) (90 ng; Promega) using Lipofectamine 2000 (0.3 μL/well; Invitrogen). The following day One-Glo luciferase detection reagent (Promega) was used to determine luciferase activity according to the manufacturer's recommendations.

Western blot analysis

Transiently transfected HeLa cells were lysed in SDT lysis buffer 6 and ~6 μg of protein was loaded per sample. The proteins were transferred to a polyvinylidene difluoride (PVDF) LF membrane, after which the membrane was blocked with 5% bovine serum albumin (BSA) for 1 hour. Primary antibodies for STAT5 were obtained from Cell Signaling Technology: STAT5 (124H6, mouse monoclonal antibody [mAb]) dilution 1:500 and phospho-STAT5 (Tyr705, D3A7, rabbit mAb) dilution 1:500, anti-α-Tubulin was obtained from Sigma Aldrich (T9026 mouse mAb) dilution 1:500. Membranes were incubated with primary antibodies diluted to Tris-buffered saline and Tween 20 (TBS-T) + 5% BSA overnight at +4°C. The membranes were

Table 1. Clinical characteristics of exome-sequenced T-LGL leukemia patients and patients with STAT5b mutation

No.	LGL leukemia type	Mutations	Sex	Age at diagnosis, y	WBC count at diagnosis, 10^9/L	Lymph count at diagnosis, 10^9/L	Lymphocyte count, 10^9/L*	$CD8^+$ Vβ †	Concomitant disorders	Therapy‡
1	$CD3^+CD8^+$ $CD56^+$ T cell	STAT5b Y665F	F	71	11.8	7.0	8.5	Vb22, 91%	Collagenosis	No
2	$CD3^+CD8^+$ $CD56^+$ T cell	STAT5b Y665F	F	49	16.1	12.9	13.7	Vb17, 94%	None	No
3	$CD3^+CD8^+$ $CD56^+$ T cell	STAT5b N642H	M	74	90.0	85.5	85.5	Vb21, 87%§	Pancytopenia, neutropenia, splenomegaly	Yes
4	$CD3^{neg}CD56^+$ NK cell	STAT5b N642H	M	75	164.7	131.8	131.8	Vb7, 27%§ Vb3, 16%§	Hemolytic anemia, neutropenia, splenomegaly, MGUS	Yes

F, female; M, male; MGUS, monoclonal gammapathy of unknown significance; WBC, white blood cell.
*Lymphocyte count at the time point when the sample was taken.
†Proportion of Vβ clone of $CD8^+$ cells.
‡Pharmacological treatment at any time point.
§Proportion from total T-cell fraction.

incubated with a secondary infrared antibodies anti-mouse IRDye 680 (Odyssey; LI-COR Biosciences) and anti-rabbit IRDye 800CW (Odyssey; LI-COR Biosciences) diluted 1:15 000 in 5% BSA TBS-T for 1 hour at room temperature (RT). The proteins were visualized with the Odyssey imaging system (LI-COR Biosciences).

Subcellular fractionation of patient samples

Subcellular fractionation and western blotting have been performed as described previously.[6] The different fractions of normalized protein aliquots were separated on either an 8% or 15% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel and next transferred to a PVDF membrane. The membranes were incubated with the following primary antibodies: rabbit anti-phospho (Tyr694) STAT5b (Cell Signaling Technology) 1:500, rabbit anti-STAT5 (Cell Signaling Technology), rabbit anti-phospho (Y705) STAT3 (Cell Signaling Technology) 1:2000, rabbit anti-phospho (Ser10) histone H3 (Cell Signaling) 1:3000, mouse anti–glyceraldehyde-3-phosphate dehydrogenase (GAPDH; Novus Biologicals) 1:5000 in 5% BSA in TBS-T or mouse anti-STAT3 (Cell Signaling Technology) 1:1000 in 5% milk in TBS-T. As secondary antibodies, goat anti-mouse IRDye 680 (LI-COR Biosciences) and goat anti-rabbit IRDye 800CW (LI-COR Biosciences) 1:15 000 in 5% BSA in TBS-T were used to match the mouse and rabbit primary antibodies, respectively.

Phospho-STAT5 and total STAT5 ELISA analysis

Enzyme-linked immunosorbent assay (ELISA) analysis was performed using the PhosphoTracer STAT5a/b pTyr693/699 + Total ELISA kit (Abcam). MNCs stored as cell pellets in −70°C were lysed according to the kit instructions and protein concentrations were measured with the Qubit 2.0 fluorometer (Life Technologies). All samples were analyzed as duplicates. Fluorescence intensity was measured with the PHERAstar FS reader (BMG Labtech).

STAT5b protein structure analysis

A homology model of the human STAT5b monomer (Uniprot sequence ID P51692) was retrieved from the SWISS-MODEL repository.[9] The STAT5b dimer structure was modeled by aligning the STAT5b monomer to 2 monomers in the mouse STAT3 dimer crystal structure template (RCDB ID: 1bg1)[10] and structural coordinates of the bound DNA fragment where copied from 1bg1. Structural alignments were generated by Dali pairwise comparison.[11]

Microarray data analysis

Microarray data were normalized and log2-transformed with Chipster as described previously.[12,13] Similarity of expression profiles was determined using pairwise Pearson correlation and a distance dendrogram was constructed using the average linkage method. Differentially expressed genes were detected using the empirical Bayes test with *P* value cutoff of .05.[14] Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for overexpressed genes was performed with a hypergeometric test.

Results

STAT5b mutation identified by exome sequencing

Two patients with untreated T-LGL leukemia and no *STAT3* mutation were chosen for exome sequencing (patients 1 and 2, Table 1). They both had a $CD3^+CD8^+CD57^+TCR\alpha\beta^+$ phenotype and a single monoclonal expansion detected by Vβ analysis (Figure 1A-C,F,H). $CD8^+$ and $CD4^+$ T cells were purified and used as a source of tumor and matched control DNA in exome sequence analysis. Exome sequencing yielded ~61 787 600 reads (range, 52 890 514-67 739 599) that could be mapped to the reference genome. The paired-end read length in both cases was 101 nucleotides.

Both patients had the same novel somatic heterozygous missense mutation in the *STAT5b* gene leading to a Y665F mutant protein (Table 2; Figures 1D-E and 2). The codon that encodes for the Y665F mutation is located in exon 16 of the SH2 domain of *STAT5b*, which mediates dimerization and activation of STAT5b by *trans*-phosphotyrosine binding (Figure 2). In the case of patient 1, the mutation was only found in $CD3^+CD8^+$ leukemic cells, whereas $CD3^+CD4^+$ T cells, $CD19^+$ B cells, and $CD3^{neg}CD56^+$ NK cells showed wild-type *STAT5b* (Figure 1E-G; supplemental Figure 1). The sequencing results of patient 2 were otherwise similar, but also $CD3^{neg}CD56^+$ NK cells were *STAT5b*-mutation positive (supplemental Figure 1).

The other genes, which harbored validated somatic mutations in addition to *STAT5b*, included bone morphogenetic protein receptor type II (*BMPR2*) in patient 1, and sphingomyelin synthase 2 (*SGMS2*), and early B-cell factor 3 (*EBF3*) in patient 2 (Table 2).

Screening of STAT5b and STAT5a mutations in LGL leukemia patient cohort

We designed deep next-generation sequencing primers for *STAT5b* exon 16 containing the Y665F mutation, and standard capillary sequencing primers for adjacent exons 14-15 and 17-18. The screening cohort consisted of 171 LGL leukemia patients with T cell phenotype

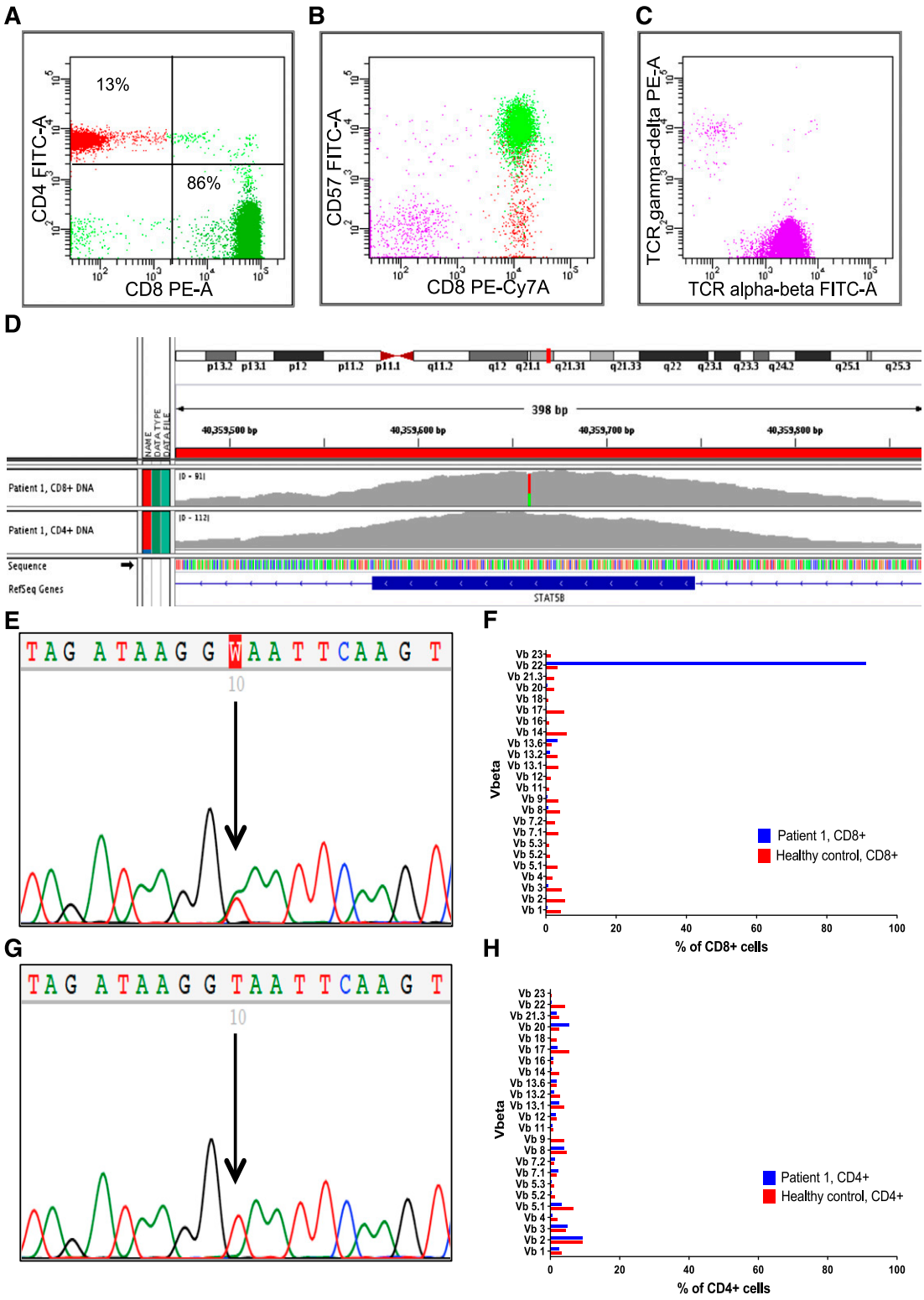

Figure 1. Sequencing and flow cytometry results of the patient 1 with *STAT5b* Y665F mutation. (A) The majority of the cells in the CD45-positive lymphocyte population were CD3 and CD8 positive. (B) Gated from $CD3^+$ cells, $CD8^+$ cells also expressed CD57 antigen and (C) TCR α-β on the cell surface. (D) IGV visualization of exon 16 of the *STAT5b* gene sequence. The tracks of both $CD8^+$ tumor DNA and $CD4^+$ control DNA are shown. In the $CD8^+$ DNA, 30 of 73 (41%) of the reads supported the variant allele (green), and 43 of 73 (59%) the normal allele (red). In the $CD4^+$ DNA, only the normal allele is seen (gray). (E) The mutation T>A (Y665F) in capillary sequencing of $CD8^+$ DNA. (F) Ninety-one percent of the CD8-positive cells consisted of a single Vβ21 clone when analyzed by flow cytometry. (G) Only the normal allele T is visible in the capillary sequencing of $CD4^+$ DNA. (H) The Vβ analysis of $CD4^+$ cells presents normal polyclonal lymphocyte population. Y, tyrosine; F, phenylalanine; PE, phycoerythrin; FITC, fluorescein isothiocyanate; PE-Cy7, phycoerythrin-cyanine 7; T, thymidine; A, adenosine; IGV, Integrative Genomics viewer.

Table 2. Somatic mutations of patients 1 and 2 identified by exome sequencing and confirmed by capillary sequencing

No.	Gene	Chromosome	Position in chromosome	Mut. type	Variant base	Reference base	CD8 variant, (n)*	CD8 reference, (n)†	CD4 variant, (n)‡	CD4 reference, (n)§	Protein	Somatic *P*‖	GERP ¶
1	STAT5b	17	40359659	Missense	A	T	30	43	0	99	Y665F	9.34×10^{-14}	4.82
	BMPR2	2	203379664	Missense	A	G	42	39	1	113	E195K	7.10×10^{-19}	5.57
2	STAT5b	17	40359659	Missense	A	T	41	38	0	63	Y665F	5.98×10^{-14}	4.82
	SGMS2	4	108820744	Missense	T	C	89	123	0	212	R157C	1.20×10^{-32}	5.52
	EBF3	10	131761204	Splice-5	G	A	41	73	0	93	NA	4.42×10^{-13}	4

C, cysteine; E, glutamic acid; F, phenylalanine; K, lysine; NA, not applicable; R, arginine; Y, tyrosine.
*Exome sequencing reads supporting variant allele in tumor sample.
†Exome sequencing reads supporting reference allele in tumor sample.
‡Exome sequencing reads supporting variant allele in control sample.
§Exome sequencing reads supporting reference allele in control sample.
‖Somatic *P* value for somatic/loss of heterozygosity events.
¶Rejected-substitution score describing the conservation of the amino acid from the program GERP (34 mammalian species, range of −11.6 to 5.82, with 5.82 being the most conserved).

and 38 patients with NK phenotype (supplemental Table 1). Two additional LGL leukemia patients (patients 3 and 4 in Table 1) were found to have a different somatic missense mutation N642H in the *STAT5b* SH2 domain (Figures 2-3; supplemental Table 6), while no *STAT5b* mutations were detected in the rest of the patients. Patient number 3 had a $CD3^+$ T-cell phenotype, but the LGL cells also expressed CD56 antigen (Figure 3A). The mutation was homozygous, as the results both from amplicon and Sanger sequencing showed only variant allele (Figure 3D-E; supplemental Table 6). The mutant allele was only found in $CD56^+$ T-LGL cells, and not in the CD56-negative normal T-cell fraction (Figure 3A-F). Patient number 4 had a heterozygous N642H mutation and the majority (80%) of leukemic cells were $CD3^-CD56^+$ NK cells (Figure 3G). Interestingly though, this patient had a separate aberrant $CD3^+CD56^+$ T-cell population, and sorting results confirmed that the mutation could be found in both of these cell populations (Figure 3G-L). In accordance with these results, the quantitative TCR rearrangement PCR of the PB MNC fraction showed an abnormal T-cell expansion (supplemental Figure 2).

As STAT5a and STAT5b protein structures are very similar, and they are capable of homo- and heterodimerization with each other,[15,16] we next designed amplicon sequencing primers to cover exon 17 of *STAT5a*, corresponding to the exon 16 of *STAT5b*. Adjacent exons (15-16 and 18-19) were sequenced by capillary sequencing. Among the 209 patients screened, none harbored mutations in the *STAT5a* gene.

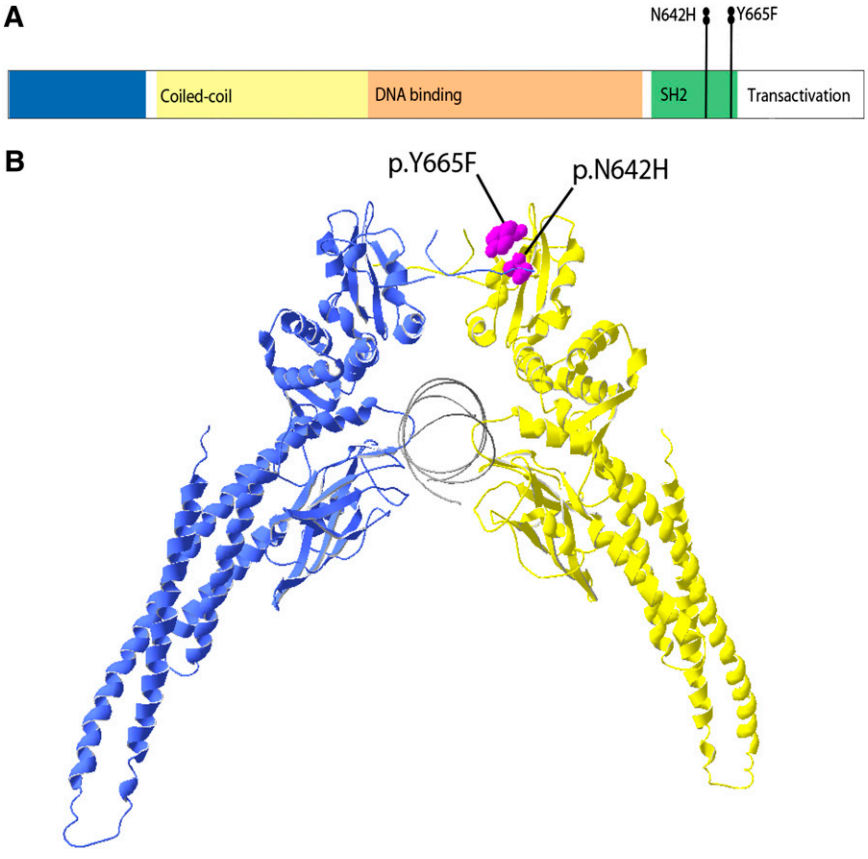

Figure 2. Linear representation of STAT5b mutation locations and structure model of STAT5b homodimer. (A) The STAT5b protein domains and localizations of the mutations. Both STAT5b mutations Y665F and N642H are located in the SH2 domain of STAT5b protein. (B) A 3-dimensional model of STAT5b protein structure. The mutations Y665F and N642H are located on the surface of SH2 domain and are colored in purple. F, phenylalanine; H, histidine; N, aspartic acid; and Y, tyrosine.

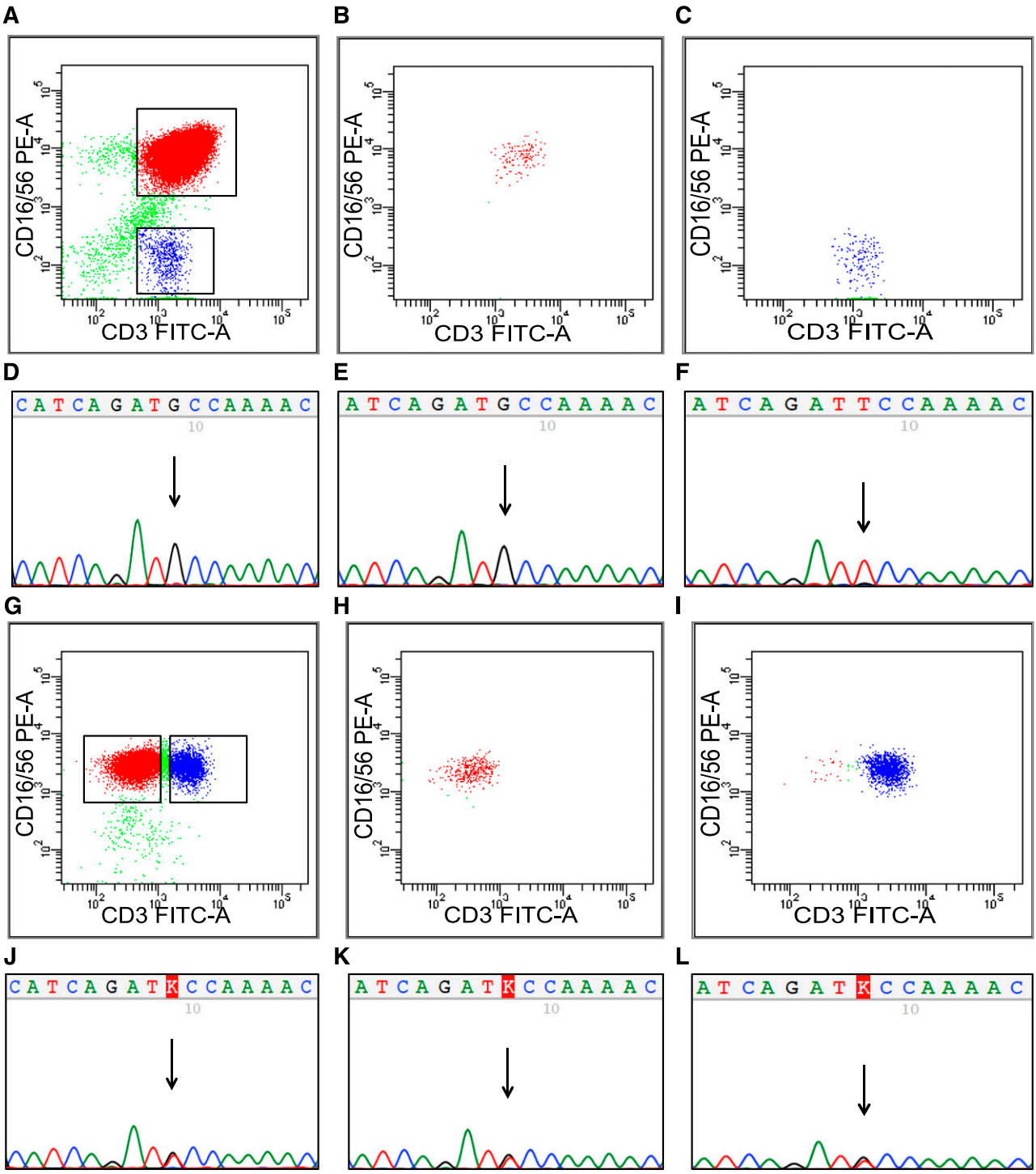

Figure 3. *STAT5b* sequencing from sorted fractions of patients 3 and 4. (A) In patient 3 (Table 1) with STAT5b N642H mutation, leukemic LGLs expressed CD3 and CD56. (B) Leukemic $CD3^+CD56^+$ cells and (C) $CD3^+CD16/56^{neg}$ control cells were sorted with flow cytometry. (B-C) The purity of sorted fractions. By capillary sequencing, the mutation N642H is visible in (D) the MNC fraction and in (E) the $CD3^+CD16/56^+$ fraction, but not in (F) $CD3^+CD16/56^{neg}$ cells. (G) The phenotype of leukemic cells in patient number 4 (Table 1) was $CD3^{neg}CD16/56^+$ NK cell, but also $CD3^+$ T cells aberrantly expressed CD56 antigen. Both (H) $CD3^{neg}CD16/56^+$ and (I) $CD3^+CD16/56^+$ cells were sorted by flow cytometry, and the panels show purity of sorted fractions. By capillary sequencing, the mutation N642H was present in (J) MNCs, but also in (K-L) both of the sorted fractions, suggesting that the aberrant $CD3^+CD56^+$ cells also belong to the leukemic clone.

In addition, 12 patients with reactive, nonmalignant LGL lymphocytosis and oligoclonal TCR rearrangement[17,18] and 2 T-cell acute lymphoblastic leukemia cell lines (Jurkat and MOLT-4) were screened, and no *STAT5a* or *STAT5b* mutations were found. The 1000 Genomes (www.1000genomes.org/data) and dbSNP (www.ncbi.nlm.nih.gov/projects/SNP) population-variation data sets were also screened, but the identified mutations were not present in either database.

The *BMPR2*, *SGMS2*, and *EBF3* mutations found by exome sequencing in 2 patients with STAT5b mutations were also screened by capillary sequencing from 77 LGL leukemia patients, but none harbored these mutations.

STAT5 is phosphorylated and translocated into the nucleus

One of the crucial steps in STAT5b activation is phosphorylation by tyrosine kinases, dimerization by *trans*-phosphotyrosine binding, and translocation of the dimers to the nucleus.[16] The localization of total and phosphorylated STAT5 was analyzed by western blotting of lysed MNCs and cytosolic and nuclear lysates. Samples were available from both patients with *STAT5b* mutation Y665F and from a healthy control. Both *STAT5b* mutated patients expressed pSTAT5b in the nuclear fraction, whereas no pSTAT5b was detected in cells derived from healthy control (Figure 4A). In addition, nuclear fractions were extracted from 3 LGL leukemia patients without *STAT5b* mutations and from 5 healthy controls and no clear pSTAT5 expression was detected in those samples (supplemental Figure 3).

The western blot results were confirmed with pSTAT5 and total STAT5 ELISA and the results were congruent: the 2 patients (1 and 2) with *STAT5b* mutation Y665F expressed a higher amount of pSTAT5 when compared with all other samples (Figure 4B).

Gene expression analysis of LGL leukemia patients based on STAT3/STAT5b mutation status

Gene expression analysis was performed using the HumanHT-12 v4 Expression BeadChip array. RNA was extracted from $CD8^+$ cells from 2 patients with *STAT5b* mutation Y665F, 1 patient with *STAT5b* mutation N642H, 3 patients with *STAT3* mutation, 2 LGL patients without mutation in *STAT3/STAT5b*, and 2 healthy controls. In addition, NK cells from 2 healthy controls were sorted and used in the analysis.

The clustering analysis based on the gene expression data showed that both patients with *STAT5b* Y665F mutation were grouped together and were further clustered with other LGL leukemia patients (both *STAT3* mutation-positive and *STAT3/STAT5b* mutation-negative patients) (Figure 4C). However, patient 4 with *STAT5b* mutation N642H and aggressive NK-LGL leukemia diverged from all other patients.

The comparison of the gene expression profile between patients with *STAT5b* mutation Y665F and healthy controls showed that 4894 genes were differentially expressed ($P < .05$). Fifty-six genes with a fold change of either >2.5 or <-2.5 ($\log_{10}$ scale) are presented in supplemental Table 7. Many of the genes upregulated in patients with the *STAT5b* mutations were also upregulated in other LGL leukemia patients (Figure 4D), suggesting that related signaling pathways operate despite the mutation status of LGL cells. Interestingly, the KEGG pathway analysis identified that in patients with *STAT5b* mutations, the chronic myeloid leukemia pathway is enriched (supplemental Table 8), which well supports STAT5 activation. In addition, protein synthesis, cell cycle, and metabolic pathways were overrepresented (supplemental Table 8).

Clinical characteristics of the patients with STAT5b mutations

Both patients with the *STAT5b* mutation Y665F had $CD3^+CD8^+$ $CD56^+TCR\alpha\beta^+$ lymphocytosis in PB and bone marrow, and $>90\%$ of $CD3^+CD8^+$ T cells consisted of a single Vβ expansion (Table 1; Figure 1A-C; supplemental Figure 1). During 18 months of follow-up, no treatment has been needed thus far and blood differential counts have repeatedly been normal besides lymphocytosis.

In contrast, in both patients with the *STAT5b* N642H mutation, LGL leukemia manifested as an aggressive form of the disease. Patient 3 had a massive splenomegaly, the phenotype of leukemic LGLs was $CD3^+CD56^+$, and quantitative TCR rearrangement PCR was positive (Table 1; Figure 3A; supplemental Figure 2). The patient died despite receiving systemic chemotherapy, and the autopsy indicated extensive lymphocyte infiltration in the bone marrow, stomach, small bowel, lungs, liver, spleen, and lymph nodes. The other patient with a *STAT5b* N642H mutation (patient 4) had NK-LGL leukemia with normal karyotype, and symptoms included neutropenia, hemolytic anemia, and splenomegaly (Table 1; Figure 3G). The patient was treated by splenectomy, which induced hematologic remission, but after 7 years of follow-up he developed chemorefractory LGL leukemia and eventually died due to his disease. The 2 patients with *STAT5b* N642H mutation were the only ones with aggressive LGL leukemia in our patient cohort.

Both mutants Y665F and N642H increase the transcriptional activity of STAT5b

For the analysis of the transcriptional activity of mutated STAT5b, the Y665F and N642H mutation constructs were transiently transfected to HeLa cells simultaneously with a luciferase sequence under the control of a STAT5-binding element. The analysis was done with unstimulated cells, and both mutants clearly induced increased luciferase activity when compared with the wild-type STAT5b plasmid (Figure 5A). The cells with the N642H construct exhibited stronger activation of transcription.

After the luciferase experiment, transfected HeLa cells were lysed and both the total STAT5 and pSTAT5 were measured by western blot analysis. The results were congruent with the luciferase assay: especially the N642H mutant which showed increased phospho-STAT5 expression when compared with the wild-type STAT5 (Figure 5B-C).

Discussion

To our knowledge, this is the first description of somatic *STAT5* mutations in human cancer patients, although STAT5 plays an important role in many hematologic malignancies and solid tumors. Exome sequencing followed by targeted Sanger and amplicon sequencing revealed that the *STAT5b* gene is recurrently mutated in LGL leukemia. Similar to the previously described *STAT3* mutations, *STAT5b* mutations were located in the SH2 domain and increased transcriptional activity and phosphorylation of the mutant protein.

STAT5b and *STAT5a* are closely related and they share a sequence identity of over 90% in complementary DNA and protein level, but their biological functions differ slightly.[19] STAT5 was first described as a mammary gland factor.[20] Consequently, $Stat5a^{-/-}$ mice display impaired mammary gland development, whereas $Stat5b^{-/-}$ mice had impaired pituitary growth hormone secretion and body growth.[21,22] Similarly in humans, homozygous deletion in *STAT5b* causes growth hormone insensitivity and short stature.[23] In comparison with $Stat3^{-/-}$ which leads to complete embryonic lethality in mice, $Stat5a/b^{-/-}$ double deficiency results in $>99\%$ perinatal death due to severe combined immunodeficiency and a lack normal lymphoid development.[24] Considering the immunologic aspects of *STAT5* genes, Stat5a deficiency in mice leads to decreased T- and NK-cell proliferation in response to antigenic stimulation.[25] Furthermore, both basal and IL-2–induced cytolytic activity of NK cells is considerably diminished in $STAT5b^{-/-}$ mice, suggesting that STAT5b plays a major role in NK-cell activation.[25] There are also a few case

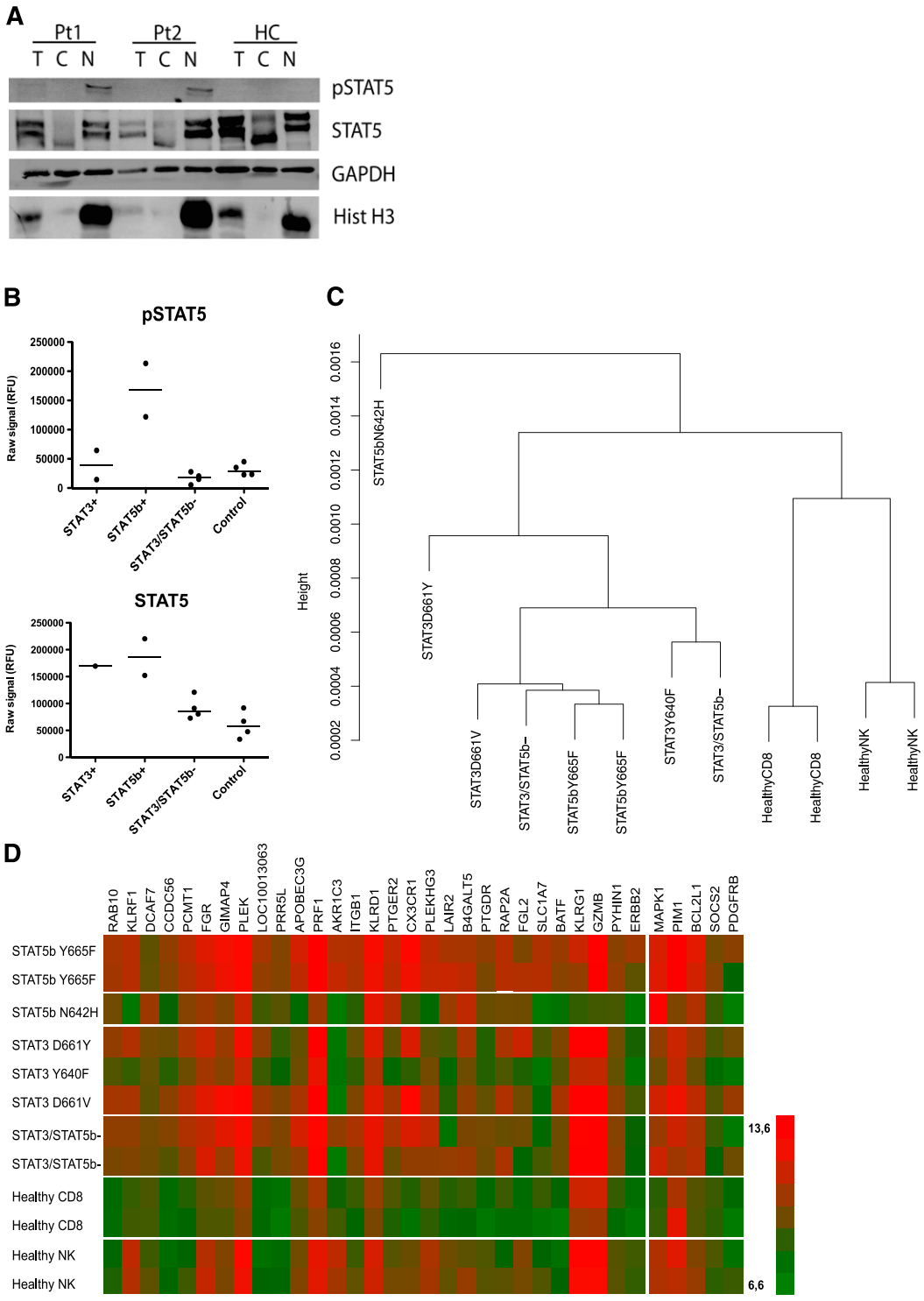

Figure 4. STAT5b is phosphorylated and localized in the nucleus leading to increased expression of STAT5 target genes in patients with Y665F mutation. (A) MNC samples from LGL patients 1 and 2 (Table 1) carrying the Y665F *STAT5b* mutation, and a healthy donor were fractionated. Normalized aliquots of total cell (T), cytosolic (C), and nuclear (N) fractions were separated on an SDS-PAGE gel, transferred to PVDF membrane, and western blot analysis was performed using anti-pSTAT5, anti-STAT5 anti-GAPDH (protein loading control), and anti-Histone H3 antibodies (nuclear localization control). (B) Phospho-STAT5 and STAT5 expression were analyzed by ELISA using PB MNC samples from 2 LGL leukemia patients with *STAT3* mutation (Y640F and D661V), 2 patients with *STAT5b* Y665F mutation, 4 patients without *STAT3/STAT5b* mutation, and 4 healthy controls. (C) A dendrogram showing the clustering analysis of LGL leukemia patients and healthy controls based on the gene expression profile. (D) A heatmap reflecting gene expression profile of 2 LGL leukemia patients with *STAT5b* mutation Y665F, 1 patient with *STAT5b* mutation N642H, 3 patients with *STAT3* mutation, 2 patients with *STAT3/STAT5b* WT, and 4 healthy controls. Shown are 29 upregulated genes (log_2 scale) based on the comparison between *STAT5b* Y665F mutated patients and healthy controls ($CD8^+$ cells). On the right, data from 5 known STAT5b target genes are displayed.

reports of STAT5b-deficient patients who have reduced numbers of NK cells.[26] Interestingly, in our study, the leukemic T-LGL cells of all patients with *STAT5B* mutations expressed the CD56 NK marker, and 1 patient had a NK phenotype suggesting that activating *STAT5b* mutations might be associated with transformation of cells related to the NK-cell lineage.

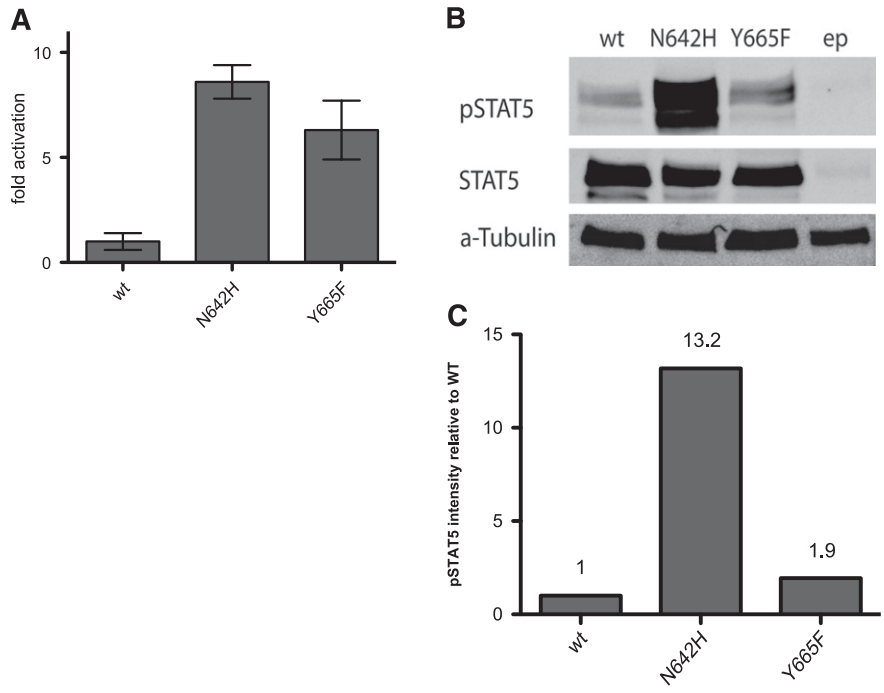

Figure 5. STAT5b mutations Y665F and N642H increase the transcriptional activity of STAT5. (A) HeLa cells were transfected simultaneously with a wild-type STAT5b vector or STAT5b mutants Y665F or N642H, and a luciferase reporter construct with a STAT5-responsive element. Each condition was tested in triplicate and the fold increase compared with wild-type vector is presented. (B) HeLa cells transfected with empty vector, wild-type STAT5b or mutant construct Y665F or N642H were lysed, and STAT5 and phospho-STAT5 levels were analyzed by western blot. (C) The western blot results shown as a fold change of the intensity of phospho-STAT5 expression compared with total STAT5 protein, normalized to the phospho-STAT5/STAT5 expression levels of HeLa cells with the wild-type STAT5 construct (Odyssey Imaging System; LI-COR Biosciences). wt, wild type; ep, empty plasmid.

The STAT5b mutations described in this article were located in the SH2 domain, close to the phosphotyrosine-binding loop (N642H) and transactivation domain (Y665F).[10] Based on the 3-dimensional structure, we hypothesize that the mutations stabilize the dimer structure, thus causing STAT5b to bind constitutively to the DNA and activate a cascade of downstream target genes as shown with the RNA expression data of patient samples. In vitro results with the mutant constructs also demonstrated that the N642H and Y665F mutations increase the transcriptional activity and phosphorylation of STAT5b protein. In many other hematologic malignancies, constitutive STAT5 activation is a secondary event: in chronic myeloid leukemia BCR/ABL expression leads to STAT5 activation,[27] and in myeloproliferative diseases STAT5 is activated by mutated Janus kinase 2.[28] In a network model of LGL leukemia signaling pathways, Zhang et al demonstrated that constitutive presence of IL-15 and PDGF, both upstream regulators of STAT5, is adequate to sustain leukemic LGLs.[29] Similarly, in a mouse model, prolonged IL-15 overexpression led to malignant transformation of LGL cells.[30] Furthermore, in a previous study, which first described activated STAT3 in LGL leukemia, 2 of 12 patients also expressed concurrently activated STAT5.[31] Our microarray data suggest that gene expression patterns in LGL leukemia patients with different mutation status resemble each other. This is in accordance with recent findings showing that in T cells STAT3 and STAT5 share many downstream target genes.[32]

According to published in vitro studies, mutations in *STAT5* itself seem to be able to act as a primary event in malignant transformation.[33] The oncogenic potential of *STAT5b* mutation N642H, established now in 2 LGL leukemia patients, was demonstrated previously by Ariyoshi et al as they showed that the transduction of the N642H mutant to mouse Ba/F3 cells led to the IL-3–independent growth of these cells causing constitutive phosphorylation and even prolonged hyperphosphorylation of STAT5 when Ba/F3 cells were stimulated with IL-3.[33] Despite these biologic properties, no mutations in *STAT5a* or *STAT5b* were observed in a cohort of 49 patients with acute leukemia.[34]

Although the number of patients with STAT5 mutations is small and no definitive conclusions between mutations and clinical characteristics can be made, it is intriguing that the N642H mutation seems to be associated with more severe disease. Both patients with the N642H mutation had atypical, aggressive chemorefractory LGL leukemia and died due to their disease. This corresponds well with the in vitro findings showing that N642H is more active than the Y665F mutation. Specific STAT5 inhibitors have not entered in clinical trials yet, but it is tempting to speculate that they may have therapeutic potential in LGL patients with STAT5 mutations. However, already there are now some molecules which inhibit the function of STAT5, such as the psychotropic drug pimozide, which was discovered in cell-based screens.[35] In a chronic myeloid leukemia model it was shown to induce apoptosis without inhibiting the BCR-ABL1 kinase.[35]

In conclusion, this study identified somatic, activating mutations in the *STAT5b* gene, further emphasizing the role of STAT family genes in the pathogenesis of LGL leukemia. In particular, *STAT5b* N642H mutations were related to aggressive clinical disease and studies evaluating the therapeutic potential of STAT5 inhibitors are warranted.

Acknowledgments

Gene expression analysis was performed by the Institute for Molecular Medicine Finland (FIMM) Technology Centre, University of Helsinki. Personnel at the Hematology Research Unit Helsinki and FIMM are acknowledged for their expert clinical and technical assistance.

This work was supported by the Academy of Finland, the Finnish Cancer Societies, Sigrid Juselius Foundation, the National Clinical Graduate School, Blood Disease Foundation, Finnish Association of Haematology, the Finnish Funding Agency for Technology and Innovation and the European Regional Development Fund, the Jane and Aatos Erkko Foundation, and the Gyllenberg Foundation. Work of J.P.M. and M.J.C. was supported in part by the National

Institutes of Health (grants 2K24HL077522, R01 CA127264A, and R01AI085578).

Authorship

Contribution: H.L.M.R. and S.M. designed the study, coordinated the project, analyzed the data, and wrote the manuscript; H.K., A.J.v.A., and H.L.M.R. designed and performed the functional laboratory studies and contributed to the writing of the manuscript; H.L.M.R., S.L., E.I.A., and P.E. performed sequence analysis and validated mutations; S.E. designed and performed bioinformatics analysis; T.O., A.J., M.J.C., Y.Y., D.Z., and A.A. provided patient samples and participated in the laboratory studies; C.H. and K.W. designed and supervised the functional laboratory studies and contributed to the writing of the manuscript; K.P., O.K., J.P.M., and T.P.L. participated in the study design, data analysis, and contributed to the writing of the manuscript; and all authors read and approved the final manuscript.

Conflict-of-interest disclosure: H.L.M.R. has received honoraria from Novartis. K.P. has received research funding and honoraria from Novartis and Bristol-Myers Squibb. S.M. has received honoraria from Novartis and Bristol-Myers Squibb. The remaining authors declare no competing financial interests.

Correspondence: Satu Mustjoki, Hematology Research Unit Helsinki, Helsinki University Central Hospital, Haartmaninkatu 8, PO Box 700, FIN-00029 Helsinki, Finland; e-mail: satu.mustjoki@helsinki.fi.

References

1. Loughran TP Jr, Kadin ME, Starkebaum G, et al. Leukemia of large granular lymphocytes: association with clonal chromosomal abnormalities and autoimmune neutropenia, thrombocytopenia, and hemolytic anemia. *Ann Intern Med.* 1985;102(2):169-175.
2. Loughran TP Jr. Clonal diseases of large granular lymphocytes. *Blood.* 1993;82(1):1-14.
3. Sokol L, Loughran TP Jr. Large granular lymphocyte leukemia. *Oncologist.* 2006; 11(3):263-273.
4. Burks EJ, Loughran TP Jr. Pathogenesis of neutropenia in large granular lymphocyte leukemia and Felty syndrome. *Blood Rev.* 2006; 20(5):245-266.
5. Leblanc F, Zhang D, Liu X, Loughran TP. Large granular lymphocyte leukemia: from dysregulated pathways to therapeutic targets. *Future Oncol.* 2012;8(7):787-801.
6. Koskela HL, Eldfors S, Ellonen P, et al. Somatic STAT3 mutations in large granular lymphocytic leukemia. *N Engl J Med.* 2012;366(20): 1905-1913.
7. Jerez A, Clemente MJ, Makishima H, et al. STAT3 mutations unify the pathogenesis of chronic lymphoproliferative disorders of NK cells and T-cell large granular lymphocyte leukemia. *Blood.* 2012;120(15):3048-3057.
8. Kallio MA, Tuimala JT, Hupponen T, et al. Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics.* 2011;12:507.
9. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.* 2009;37(Database issue):D387-D392.
10. Becker S, Groner B, Müller CW. Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature.* 1998;394(6689):145-151.
11. Hasegawa H, Holm L. Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol.* 2009;19(3):341-348.
12. Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods.* 2003;31(4):265-273.
13. Ritchie ME, Silver J, Oshlack A, et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics.* 2007; 23(20):2700-2707.
14. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article3.
15. Ambrosio R, Fimiani G, Monfregola J, et al. The structure of human STAT5A and B genes reveals two regions of nearly identical sequence and an alternative tissue specific STAT5B promoter. *Gene.* 2002;285(1-2):311-318.
16. Leonard WJ, O'Shea JJ. Jaks and STATs: biological implications. *Annu Rev Immunol.* 1998; 16:293-322.
17. Kreutzman A, Juvonen V, Kairisto V, et al. Mono/oligoclonal T and NK cells are common in chronic myeloid leukemia patients at diagnosis and expand during dasatinib therapy. *Blood.* 2010; 116(5):772-782.
18. Mustjoki S, Ekblom M, Arstila TP, et al. Clonal expansion of T/NK-cells during tyrosine kinase inhibitor dasatinib therapy. *Leukemia.* 2009;23(8): 1398-1405.
19. Mui AL, Wakao H, Harada N, O'Farrell AM, Miyajima A. Interleukin-3, granulocyte-macrophage colony-stimulating factor, and interleukin-5 transduce signals through two forms of STAT5. *J Leukoc Biol.* 1995;57(5):799-803.
20. Gouilleux F, Wakao H, Mundt M, Groner B. Prolactin induces phosphorylation of Tyr694 of Stat5 (MGF), a prerequisite for DNA binding and induction of transcription. *EMBO J.* 1994;13(18): 4361-4369.
21. Liu X, Robinson GW, Wagner KU, Garrett L, Wynshaw-Boris A, Hennighausen L. Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.* 1997;11(2): 179-186.
22. Udy GB, Towers RP, Snell RG, et al. Requirement of STAT5b for sexual dimorphism of body growth rates and liver gene expression. *Proc Natl Acad Sci U S A.* 1997;94(14):7239-7244.
23. Hwa V, Camacho-Hübner C, Little BM, et al. Growth hormone insensitivity and severe short stature in siblings: a novel mutation at the exon 13-intron 13 junction of the STAT5b gene. *Horm Res.* 2007;68(5):218-224.
24. Yao Z, Cui Y, Watford WT, et al. Stat5a/b are essential for normal lymphoid development and differentiation. *Proc Natl Acad Sci U S A.* 2006; 103(4):1000-1005.
25. Imada K, Bloom ET, Nakajima H, et al. Stat5b is essential for natural killer cell-mediated proliferation and cytolytic activity. *J Exp Med.* 1998;188(11):2067-2074.
26. Kanai T, Jenks J, Nadeau KC. The STAT5b pathway defect and autoimmunity. *Front Immunol.* 2012;3:234.
27. Carlesso N, Frank DA, Griffin JD. Tyrosyl phosphorylation and DNA binding activity of signal transducers and activators of transcription (STAT) proteins in hematopoietic cell lines transformed by Bcr/Abl. *J Exp Med.* 1996;183(3):811-820.
28. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005; 352(17):1779-1790.
29. Zhang R, Shah MV, Yang J, et al. Network model of survival signaling in large granular lymphocyte leukemia. *Proc Natl Acad Sci U S A.* 2008; 105(42):16308-16313.
30. Mishra A, Liu S, Sams GH, et al. Aberrant overexpression of IL-15 initiates large granular lymphocyte leukemia through chromosomal instability and DNA hypermethylation. *Cancer Cell.* 2012;22(5): 645-655.
31. Epling-Burnette PK, Liu JH, Catlett-Falcone R, et al. Inhibition of STAT3 signaling leads to apoptosis of leukemic large granular lymphocytes and decreased Mcl-1 expression. *J Clin Invest.* 2001;107(3):351-362.
32. Kang K, Robinson GW, Hennighausen L. Comprehensive meta-analysis of signal transducers and activators of transcription (STAT) genomic binding patterns discerns cell-specific cis-regulatory modules. *BMC Genomics.* 2013;14:4.
33. Ariyoshi K, Nosaka T, Yamada K, et al. Constitutive activation of STAT5 by a point mutation in the SH2 domain. *J Biol Chem.* 2000; 275(32):24407-24413.
34. Yamada K, Ariyoshi K, Onishi M, et al. Constitutively active STAT5A and STAT5B in vitro and in vivo: mutation of STAT5 is not a frequent cause of leukemogenesis. *Int J Hematol.* 2000; 71(1):46-54.
35. Nelson EA, Walker SR, Weisberg E, et al. The STAT5 inhibitor pimozide decreases survival of chronic myelogenous leukemia cells resistant to kinase inhibitors. *Blood.* 2011;117(12):3421-3429.