Hematology Research Unit Helsinki,
Department of Hematology,
University of Helsinki and Helsinki University Central Hospital Cancer Center,
Helsinki, Finland

The National Graduate School of Clinical Investigation (CLIGS)

MOLECULAR PATHOGENESIS OF LARGE GRANULAR LYMPHOCYTIC LEUKEMIA

Hanna Rajala

ACADEMIC DISSERTATION

To be presented, with the permission of the Medical Faculty, University of Helsinki, for public examination in Biomedicum lecture hall 2, Haartmaninkatu 8, on June 6th 2014, at 12 o'clock noon.

Helsinki 2014

Supervised by:
Docent Satu Mustjoki, M.D., Ph.D.
Department of Hematology
University of Helsinki and Helsinki University Central Hospital Cancer Center
Helsinki, Finland

Professor Kimmo Porkka, M.D., Ph.D.
Department of Hematology
University of Helsinki and Helsinki University Central Hospital Cancer Center
Helsinki, Finland

Reviewed by:
Professor Sirpa Leppä, M.D., Ph.D.
Department of Oncology
Helsinki University Central Hospital
Helsinki, Finland

Docent Marko Pesu, M.D., Ph.D.
BioMediTech
University of Tampere
Tampere, Finland

Opponent:
Professor Olli Silvennoinen, M.D., Ph.D.
School of Medicine
University of Tampere
Tampere, Finland

To Jukka

TABLE OF CONTENTS

ABBREVIATIONS

AA	Aplastic anemia
ABL	Abelson murine leukemia viral oncogene homolog
AICD	Activation-induced cell death
AIHA	Autoimmune hemolytic anemia
AKT	V-akt murine thymoma viral oncogene homolog
ALL	Acute lymphoblastic leukemia
ALPS	Autoimmune lymphoproliferative syndrome
AML	Acute myeloid leukemia
ARMS	Amplification-refractory mutation system
ASO-qPCR	Allele-specific oligonucleotide quantitive PCR
BCL	B cell leukemia/lymphoma
BCR	Breakpoint cluster region
BM	Bone marrow
BMPR2	Bone marrow morphogenetic receptor 2
bp	Base pair
CAPN13	Calpain-13
CCL5	Chemokine (C-C) motif ligand 5
CCR	Chemokine receptor
CD	Cluster of differentiation
CDR	Complementarity determining regions
ChIP-seq	Chromatin immunoprecipitation-sequencing
CLL	Chronic lymphoid leukemia
CLPD-NK	Chronic lymphoproliferative disorder of NK cells
CML	Chronic myeloid leukemia
CMR	Complete molecular response
CMV	Cytomegalovirus
CR	Complete response
CRM1	Chromosome region maintenance 1
CTLA-4	Cytotoxic T lymphocyte antigen 4
D	Diversity
DISC	Death-inducing silencing complex
DLBCL	Diffuse large B cell lymphoma
DMSO	Dimethyl sulfoxide
DOCK8	Dedicator of cytokinesis 8
EBF3	Early B cell factor 3
EBV	Ebstein-Barr virus
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
FADD	FAS-associated protein with death domain
FAS	Fas cell surface death receptor
FASL	FAS ligand
FFPE	Formalin-fixed paraffin-embedded
FLT3	FMS-like tyrosine kinase-3
FOXP3	Forkhead box P3
G-CSF	Granulocyte colony-stimulating factor
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GAS	Gamma-activated site or IFN-gamma activated sequence
HCMV	Human cytomegalovirus
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HOXC9	Homeobox C9
HTLV	Human T lymphotrophic virus
IFN	Interferon
Ig	immunoglobulin
IGV	Integrative genomics viewer
IL	Interleukin
IL6R	Interleukin-6 receptor
IL6ST	Interleukin-6 signal transducer
J	Joining
JH1	Jak homology 1
JAK	Janus kinase
KIR	Killer immunoglobulin-like receptor
KLRC	Killer cell lectin-like receptor

LGL	Large granular lymphocyte/lymphocytic
MAP2K	Mitogen-activated protein kinase kinase
MBL	Monoclonal B cell lymphocytosis
MCL-1	Myeloid cell leukemia-1
MDS	Myelodysplastic syndrome
MGG	May-Grünwald-Giemsa
MGUS	Monoclonal gammapathy of unknown significance
MHC	Major histocompatibility complex
MIP-1	Macrophage inflammatory protein-1
MNC	Mononuclear cell
MPD	Myeloproliferative disorder
MSR1	Macrophage scavenger receptor 1
MTX	Methotrexate
MYC	V-myc avian myelocytomatosis viral oncogene homolog
NCR	Natural cytotoxic receptor
NFκB	Nuclear factor kappaB
NGS	Next-generation sequencing
NK	Natural killer
NOE	N-oleoylethanolamine
ORR	Overall response rate
PAK	P21 protein (Cdc42/Rac)-activated kinase
PB	Peripheral blood
PD-1	Programmed cell death-1
PDGF	Platelet-derived growth factor
PE	Paired end
PHA	Phytohaemagglutin
PI3K	Phosphatidylinositol-4,5-biphosphate 3-kinase
PIAS	Protein inhibitor of activated STAT
PLL	Prolymphocytic leukemia
PR	Partial response
PRCA	Pure red cell aplasia
PTLD	Post-transplant lymphoproliferative disorder
PTP	Protein tyrosine-phosphatase
PTPRT	Protein tyrosine phosphatase, receptor type
QV	Quality value
RA	Rheumatoid arthritis
RAC	Ras-related C3 botulinum toxin substrate
S1P	Sphingosine-1-phosphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SGMS2	Sphingomyelin synthase 2
SH2	SRC-like homologue 2
SIE	Sis-inducible element
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOCS	Suppressor of cytokine signaling
STAT	Signal transducer and activator of transcription
SYK	Spleen tyrosine kinase
TAD	Transactivating domain
TCR	T cell receptor
TCRB	T cell receptor beta chain
Th	Effector T helper cell
Th17	IL17-producing effector T helper cell
TNF	Tumor necrosis factor
TNFAIP3	Tumor necrosis factor alpha-induced protein 3
TP53	Tumor protein p53
Treg	Regulatory T cell
TYK	Tyrosine kinase
V	Variable
VAF	Variant allele frequency
Vbeta	T cell receptor beta chain variable region
WES	Whole-exome sequencing
WHO	World Health Organisation
wt	wild type

ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

I.   **Koskela H (Rajala H)***, Eldfors S*, Ellonen P, van Adrichem AJ, Kuusanmäki H, Andersson EI, Lagström S, Clemente MJ, Olson T, Jalkanen SE, Majumder MM, Almusa H, Edgren H, Lepistö M, Mattila P, Guinta K, Koistinen P, Kuittinen T, Penttinen K, Parsons A, Knowles J, Saarela J, Wennerberg K, Kallioniemi O, Porkka K, Loughran TP Jr, Heckman CA, Maciejewski JP, Mustjoki S. Somatic *STAT3* mutations in large granular lymphocytic leukemia. *N Engl J Med* 2012; 366(20):1905-13.

II.   Jerez A, Clemente MJ, Makishima H, **Koskela H (Rajala H)**, Leblanc F, Peng Ng K, Olson T, Przychodzen B, Afable M, Gomez-Segui I, Guinta K, Durkin L, Hsi ED, McGraw K, Zhang D, Wlodarski MW, Porkka K, Sekeres MA, List A, Mustjoki S, Loughran TP, Maciejewski JP. *STAT3* mutations unify the pathogenesis of chronic lymphoproliferative disorders of NK cells and T-cell large granular lymphocyte leukemia. *Blood.* 2012 11;120(15):3048-57

III.   **Rajala H**, Eldfors S*, Kuusanmäki H*, van Adrichem AJ*, Olson T, Lagström S, Andersson EI, Jerez A, Clemente MJ, Yan Y, Zhang D, Awwad A, Ellonen P, Kallioniemi O, Wennerberg K, Porkka K, Maciejewski JP, Loughran TP Jr, Heckman C, Mustjoki S. Discovery of somatic *STAT5b* mutations in large granular lymphocytic leukemia. *Blood.* 2013 30;121(22):4541-50.

IV.   **Rajala H**, Olson T, Clemente M, Lagström S, Ellonen P, Lundan T, Uz Zaman A, Lopez Marti J, Andersson E, Jerez A, Porkka K, Maciejewski J, Loughran T, Mustjoki S. The analysis of clonal hierarchy and therapy responses using *STAT3* mutations as a molecular marker in large granular lymphocytic leukemia. Submitted.

*The authors contributed equally to this work.

ABSTRACT

Large granular lymphocytic (LGL) leukemia is a chronic incurable disorder characterized by expansion of cytotoxic T- or natural killer (NK)-lymphocytes infiltrating blood and bone marrow. The diagnostic criteria include persistent lymphocytosis in peripheral blood, and in the case of T-LGL leukemia, detection of a clonal rearrangement of the T cell receptor. Typical clinical and hematological characteristics, including anemia, neutropenia, and autoimmune disorders, further support the diagnosis. Current immunosuppressive treatment options, such as methotrexate, induce remission in only 50% of cases and no targeted therapies exist. The aim of this project was to characterize the molecular pathogenesis of LGL leukemia and find molecular markers that could be used for diagnosis and novel therapeutic approaches.

In the first study, the molecular background of LGL leukemia was analyzed by exome sequencing of a T-LGL leukemia patient. The index patient carried a novel D661V mutation in the Signal transducer and activator of transcription 3 (*STAT3*) gene. STAT3 is a known oncoprotein and a transcription factor. In the subsequent screening, the prevalence of somatic *STAT3* mutations in T-LGL cases was 40%. Based on functional studies, the mutations located in the SRC-like homology 2 (SH2) domain increased phosphorylation and transcriptional activity of STAT3.

A larger LGL-leukemia patient cohort, including both T-LGL leukemia and chronic lymphoproliferative disorder of NK cells (CLPD-NK) patients, was analyzed in the second study. Activating *STAT3* mutations were seen in 27% and 30% of cases, respectively. The analysis of clinical characteristics showed that patients with *STAT3* mutations were more likely to have neutropenia and rheumatoid arthritis and also required more frequent treatment when compared with unmutated patients.

The third study concentrated on the analysis of *STAT3* mutation-negative patients by exome sequencing. Two T-LGL leukemia patients carried Y665F mutation in *STAT5B* gene. In the screening of over 200 T-LGL leukemia and CLPD-NK cases, two patients presented with additional N642H mutation in *STAT5B*. The mutations were situated in the SH2 domain and led to increased phosphorylation and transcriptional activity of STAT5B. The *in vitro* effects of N642H mutation were more prominent, and both patients harboring N642H mutation had an untypically aggressive clinical presentation.

In the fourth study, the clonal hierarchy and *STAT3* mutation spectrum during immunosuppressive treatment was analyzed by deep amplicon-sequencing of *STAT3* and simultaneous deep sequencing analysis of a T cell receptor beta chain (TCRB) repertoire. A total of 22% of STAT3-mutated patients harbored multiple *STAT3* mutations mostly in separate lymphocyte clones. Mutated clones did not share common T cell receptor beta chain (TCRB) sequences by deep sequencing method in three T-LGL leukemia cases analyzed. Complete remission after immunosuppressive therapy resulted in the eradication of the *STAT3*-mutated clone in most patients, whereas partial responses reflected only modest changes in the leukemic clone burden.

In conclusion, the activating *STAT3* and *STAT5B* mutations provide a basis for understanding the molecular pathogenesis of leukemic LGL expansion. LGL leukemia can result from chronic antigen exposure in combination with immunogenetic factors such as *STAT* mutations. The detection of *STAT3* and *STAT5B* mutations can be included in the diagnostic criteria of LGL leukemia, and facilitate development of novel therapeutics.

INTRODUCTION

Large granular lymphocytic (LGL) leukemia was first described in the literature in the 1980s, but the ultimate factors driving leukemic expansion remain unknown despite several studies that have noted marked disturbances in the cytokine profiles of LGL leukemia patients and dysregulated immune system.[1-3] Consequently, therapeutic approaches have focused on the use of non-targeted immunosuppressive drugs, but these have proved to be ineffective in most patients and have marked side effects in long-term use.

Leukemic T-large granular lymphocytes (T-LGLs) are terminally differentiated effector memory cells that have encountered their antigen. Phenotypically similar but transient lymphocytosis is often seen during viral infections or autoimmune diseases. In these disorders, expanded lymphocytes undergo programmed apoptosis by activation-induced cell death (AICD). Based on previous studies, LGL leukemia cells are resistant to Fas cell surface death receptor (FAS)-mediated AICD. Several anti-apoptotic pathways are active in LGL leukemia, the most important being interleukin-6 (IL-6)-Janus kinase (JAK)-Signal transducer and activator of transcription 3 (STAT3) pathway, which controls the transcription of different proteins that are connected to cell cycle control and proliferation. STAT3 inhibition restores the FAS-sensitivity of leukemic LGLs.[4] STAT3 is a pleiotropic transcription factor and an oncogene, and its activation has been described in multiple malignancies as an important secondary event caused by activation of cancer-specific oncoproteins. Chromosomal aberrations are extremely uncommon in LGL leukemia and disease-specific genetic changes have not been observed. In recent years, the rapid development of novel sequencing methods has enabled fast and relatively cheap analysis of somatic genetic variants in different cancers. These methods could be used to sequence the exome or genome of leukemic LGLs using tumor cell populations and their normal counterparts, which can be collected and sorted from the bloodstream.

The diagnosis of LGL leukemia often requires follow-up and no specific molecular markers exist. The analysis of T-cell receptor (TCR) rearrangement in T-LGL leukemia proves the clonality of lymphocyte expansion, but clonal cells are a normal physiological phenomenon after antigen-exposure and thus are seen also in non-malignant conditions. Similar methods are not available for the analysis of clonality in chronic lymphoproliferative disorder of NK cells (CLPD-NK). In many malignancies, specific molecular markers are used in diagnostics and they also represent a subject for targeted therapy: for example, chronic myeloid leukemia (CML) is characterized by the Breakpoint cluster region (BCR)-Abelson murine leukemia viral oncogene homolog 1 (ABL1) fusion gene, which is targeted by tyrosine kinase inhibitors. Indeed, modern rational drug design is based on the hypothesis that a disease has a specific molecular abnormality, which can be addressed with targeted therapy.

The aim of this study was to characterize the molecular pathogenesis of LGL leukemia using second-generation sequencing methods. In addition, the clonal hierarchy of leukemic cells was analyzed in detail using flow cytometry-based sorting and deep targeted sequencing.

REVIEW OF THE LITERATURE

1. History of LGL leukemia

Large granular lymphocytes (LGL) belong to the lymphoid lineage of hematopoietic cells and are morphologically recognized as mature lymphocytes by mononuclear cell type and azurophilic granules (Figure 1).[5] Several diseases are caused by LGLs; for example, a report of patients with neutropenia and increased number of circulating LGLs was published in 1977.[6] Thomas P. Loughran and colleagues demonstrated the proof of clonality in 1985, based on the detection of clonal cytogenetic abnormalities in two patients with increased numbers of LGLs and chronic neutropenia.[1] The clonal cells were able to infiltrate different tissues, so the disease entity was named LGL leukemia and the term is still used.[1] In its history, the disease was also called various names such as T cell chronic lymphocytic leukemia (CLL), lymphoproliferative disorder of granular lymphocytes, or Tγ lymphocytosis.[7] The FAB classification in 1990 established LGL leukemia as the preferred term and in 1993, a distinction was made between T-LGL and natural killer (NK) cell LGL leukemia.[7,8] In the latest World Health Organization (WHO) 2008 classification, NK-LGL leukemia was further divided into two subcategories: chronic lymphoproliferative disorder of NK-cells (CLPD-NK) and rare aggressive NK-LGL leukemia.[9] In the following text, the term LGL leukemia includes both T-LGL leukemia and CLPD-NK.
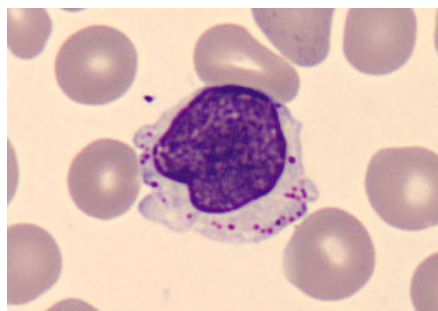
Figure 1. *MGG staining of a large granular lymphocyte in peripheral blood. Courtesy of Kirsi Latvala.*

2. Clinical characteristics and diagnosis of LGL leukemia

2.1. Common features of T-LGL leukemia and CLPD-NK

The diagnosis of T-LGL leukemia and CLPD-NK is based on persistent (lasting over six months) peripheral blood (PB) lymphocytosis with expression of characteristic immunophenotype, positive clonality testing (in case of T-LGL leukemia), and typical clinical presentation (Table 1).[3,9]

The cells have a distinct morphology: large size (15-18 μm), abundant cytoplasm with a variable number of azurophilic granules and a relatively round nucleus with condensed chromatin (Figure 1).[10] The normal LGL count in PB is 0.2-0.4 $x10^9$/l, while in LGL leukemia patients it is usually over 2 $x10^9$/l, although the diagnosis can be established with lower values (>0.5 $x10^9$/l) if other diagnostic criteria are met.[2,11,12] LGLs comprise 5% to 15% of peripheral blood mononuclear cells in normal individuals. They mediate direct cytotoxic activity and can be divided into two

subsets based on CD3 expression on their surface: CD3[neg] NK cells and CD3+ T cells.[13] Some patients have both T and NK cell expansions.[14] The number of bone marrow (BM) infiltrating lymphocytes varies in LGL leukemia, and the LGLs resemble morphologically granulocytic or monocytic precursor cells, making it difficult to distinguish them without immunohistochemical staining.[10] LGLs can infiltrate the red pulp of the spleen, causing mild or moderate splenomegaly, but lymphadenopathy is rare.[10] Involvement of the liver is also common, although rarely examined by histological samples. [10]

In addition to persistent lymphocytosis and morphological analysis, the diagnosis of LGL leukemia requires usually immunophenotypic examination and clonality testing, and is further supported by typical hematological and autoimmune manifestations.

2.2. Laboratory diagnostics of T-LGL leukemia

Morphological analysis alone cannot separate leukemic LGLs from their normal counterparts, but the phenotype of the cells is almost always aberrant.[15] Modern laboratories analyze the surface marker expression by flow cytometry. Leukemic LGLs in T-LGL leukemia are mature postthymic lymphocytes. The phenotype of the cells is usually CD3+ T-cell receptor (TCR)-α/β+ CD4[neg] CD8[+] CD5[dim] CD16+ CD27[neg] CD28[neg] CD57+ CD62L[neg], which defines them as activated effector lymphocytes (Table 1).[10] The expression of common T-cell antigens such as CD5 and CD7 is often lost or diminished.[15] The CD8+ TCR- α/β^{+} phenotype comprises over 90% of cases, whereas CD4[+] or TCR-γ/δ^{+} LGL leukemia is rare.[7,16] CD57 and CD16 are commonly NK cell antigens, but their aberrant expression is typical for T-LGL leukemia.[7] Based on the expression of CD45RA, and the lack of CD62L and chemokine-receptor 7 (CCR7) expression, leukemic LGLs have the phenotype of terminal-effector memory cells.[17,18] Expression of CD56 in T-LGL is connected to atypical aggressive T-LGL leukemia.[19]

The clonality of leukemic cells is analyzed by sequencing of TCR gamma and delta chain (TCR-γ/δ) rearrangement. Standardized protocols have been developed by the BIOMED-2 consortium.[20,21] Although T-LGL leukemia cells usually express TCR-α/β, the previously rearranged gamma chain is not deleted during the subsequent rearrangement process of the α/β genes and thus can be used in the DNA-based clonality assay. The TCR β-chain variable (V) region repertoire can be assessed using monoclonal flow cytometry antibodies, which recognize about 75% of the total spectrum of Vβ.[22] Unlike PCR methods, this assay gives also a quantitative estimate of the number and size of T cell expansions.

2.3. Laboratory diagnostics of CLPD-NK

CLPK-NK cells are usually CD2[+] CD3[neg] CD4[neg] CD8[neg] CD16[+] CD56[+], and often also CD57[+]; however, the expression of the common NK cell marker CD16 can vary (Table 1).[2,7,15] TCR clonality testing is not applicable in CLPK-NK, and no definite method for clonality analysis exists, unless leukemic cells harbor cytogenetic abnormalities. NK-type LGLs express killer immunoglobulin-like receptors (KIR) on their surfaces: in CLPD-NK, possible monotypic or skewed expression pattern may be observed by using monoclonal antibodies to CD158a, CD158b, and CD158e, but testing is not widely used outside of research settings.[23]

2.4. Hematological aberrations

The largest cohort describing the clinical characteristics of LGL leukemia patients included 229 French patients (209 T-LGL, 27 CLPD-NK, and 1 aggressive NK-LGL).[12] In this cohort, 51% of T-LGL and 54% of CLPD-NK patients had absolute lymphocyte counts >$4x10^9$/l, but the absolute LGL count was <$1x10^9$/l in 36% and 47% of the study patients, respectively.[12] The mean LGL count in the whole series was $1.7x10^9$/l.[12] Neutropenia (<$1.5x10^9$/l) was the most common cytopenia observed (61% and 48% in T-LGL and CLPD-NK, respectively), and a marked proportion of the patients (26% and 16%) presented with severe neutropenia (<$0.5x10^9$/l), whereas anemia (24% and 28%) and thrombocytopenia (19% and 8%) were rarer.[12] Recurrent infections related to neutropenia were commonly seen in both T-LGL (23%) and CLPD-NK (18%).[12] The results from the French cohort were concordant with other large previously published series.[7,24]

The mechanism of neutropenia is probably multi-factorial and related to both impaired myeloid proliferation and increased peripheral destruction. The level of circulating FAS ligand (FASL) is elevated in LGL leukemia patients, and the serum of LGL leukemia patients triggers apoptosis of neutrophils, as normal neutrophil survival is regulated by the FAS-FASL system.[25] Anti-neutrophil antibodies in LGL leukemia patients' sera could also cause neutropenia in an immunocomplex-mediated mechanism.[2] The infiltration of BM in LGL leukemia is often subtle—even non-existent—in some patients, and cannot explain neutropenia in LGL leukemia patients when compared with acute leukemias in which the massive infiltration of leukemic blast cells causes profound cytopenias.[26] However, the BM microenvironment seems to be dysregulated in LGL leukemia: severe fibrosis was connected to LGL BM infiltration and associated with the presence of cytopenias.[27]

The most common hematological aberrations and even malignant conditions associated with LGL leukemia are different B cell dyscrasias ranging from polyclonal hypergammaglobulinemia to monoclonal gammapathy of unknown significance (MGUS), and lymphoma.[28,29] In a cohort of 63 patients, coexisting B cell abnormalities were identified in 43% of the patients.[29] These included both nonclonal conditions such as hypergammaglobulinemia/hypogammaglobulinemia (16%) and clonal disorders including MGUS (19%), chronic lymphocytic leukemia (CLL) (8%), and follicular lymphoma (2%). In another study, abnormal monotypic B cell populations (monoclonal B cell lymphocytosis, MBL) were detected by flow cytometry in 28% of 57 LGL leukemia patients.[28] MBL can also be seen in healthy elderly individuals without apparent lymphocytosis, with a prevalence of 5%.[30] When associated with elevated lymphocyte counts, MBL is a known risk factor for CLL.[30] The French cohort showed different gammapathies in 47% of patients, and the most common neoplasms were lymphomas (7%).[12] The pathogenetic mechanism leading to concurrent LGL and B cell expansion is not known: LGL leukemia possibly has either developed in response to antigenic stimulation caused by abnormal B cell population or vice versa, or both populations have developed independently in response to the same or different antigens.[28] The relationship of other neoplasms to LGL leukemia has not been confirmed to the same extent.[12]

2.5. Autoimmune manifestations

The incidence of autoimmune diseases is higher in LGL leukemia patients than in the normal population and the patients can present with signs and symptoms of autoimmunity before LGL leukemia is diagnosed; 34% of LGL leukemia patients manifest with some type of autoimmune disorder.[12,31] The most common concomitant disorder is rheumatoid arthritis (RA), which is more common in T-LGL leukemia than in CLPD-NK, and seen in 11-36% of patients, depending on the study cohort (Table 1).[7,12,24] Felty's syndrome is a rare complication of RA, characterized by the clinical triad of RA, splenomegaly, and neutropenia. Felty's syndrome and LGL leukemia have similar clinical and laboratory manifestations, with the Human leukocyte antigen-DR4 (HLA-DR4) haplotype as a predisposing factor.[32] Both have aberrations in the T cell population, such as a skewed T cell receptor rearrangement repertoire, and they probably share the same pathogenetic mechanism.[33,34]

Autoimmune cytopenias can occur concomitantly with LGL leukemia: in the French, cohort the prevalences of pure red cell aplasia (PRCA) and autoimmune hemolytic anemia (AIHA) were 3% and 4%, respectively.[12] However, PRCA is remarkably more common in Asian populations, with prevalence of 47% in LGL leukemia patients, whereas neutropenia, which is common in Caucasian patients, is almost nonexistent in Asian cohorts.[35] The reason for this difference is not known.

Other autoimmune disorders such as Sjögren's syndrome, different vasculitis, endocrinopathies, and chronic inflammatory bowel disease are sometimes associated with LGL leukemia.[12,36]

2.6. Differential diagnosis of LGL leukemia

The diagnosis of LGL leukemia is based on many factors, and even clonality of expanded lymphocytes is not a definitive marker for the diagnosis. Several benign conditions can mimic LGL leukemia; the incidence of oligoclonal or small monoclonal asymptomatic T cell populations is increased with age, and in one previous study clonal T cells were seen in all (14/14) subjects older than 65 years.[37,38] Transient or even persistent oligo/monoclonal T lymphocytosis is sometimes seen during viral infections. For example, some HIV infected patients have circulating clonal CD8+ LGL expansions, which show cytotoxic properties against viral core protein.[39] Human cytomegalovirus (CMV) stays in the body as a latent form, and immunocompetent individuals often have clearly detectable amounts of CMV antigen-specific T and NK cells, both during primary infection and at later timepoints.[40] Clonal T cells and increased amounts of NK cells can also be seen in chronic myeloid leukemia (CML) patients during tyrosine-kinase inhibitor therapy.[41,42] While repeated examination of clinical characteristics, laboratory values, and clonality can differentiate most of these conditions of T cell lineage, similar evaluation of NK-lineage processes is more complicated because of the lack of good clonality marker.[3]

Different T and NK neoplasms also need to be taken into consideration while establishing the diagnosis of LGL leukemia. These include, for example, T cell prolymphocytic leukemia (T-PLL), hepatosplenic T cell lymphoma, extranodal NK/T cell lymphoma of the nasal type, and aggressive NK cell leukemia.[10] T-PLL and hepatosplenic T cell lymphoma both have more aggressive clinical courses, marked hepatosplenomegaly, and cytogenetic abnormalities, and the immunophenotype of the leukemic cells differs from LGL leukemia (usually CD4+ or CD4+CD8+ in T-PLL and CD4^{neg}CD8^{neg} in hepatosplenic T cell lymphoma).[43,44] NK cell-originating

malignancies are aggressive neoplasms, which are thought to associate with Epstein-Barr virus (EBV). They are rare in Western countries, in contrary to Asian and South American populations.[45] In particular, NK cell leukemia has an extremely fatal course, with poor median survival of less than two months.[46] A rare subtype of T-LGL leukemia, separate from the indolent T-LGL form, is characterized by the expansion of CD56-expressing T-LGLs, B symptoms (fever, night sweat, and weight loss), lympadenopathy, and splenomegaly.[19] It is resistant to immunosuppressive therapy.[19] Unlike aggressive NK-LGL leukemia, no connection to EBV infection has been detected.[19]

Post-transplant lymphoproliferative disorders (PTLD) are potentially fatal conditions that develop after either solid organ or hematopoietic stem cell transplantation, and their clinical presentation varies from polyclonal lymphoid hyperplasia to aggressive lymphoma.[47] PTLD originates from B cells, and is in most cases associated with EBV.[48]

Autoimmune lymphoproliferative syndrome (ALPS), which is caused by genetic defects in FAS-mediated apoptosis (usually loss-of-function mutations in the intracellular domain of FAS), presents typically as chronic lymphoproliferation, hepatosplenomegaly, autoimmune diseases, autoimmune cytopenias, and increased incidence of cancer. The mutations can be either germline or somatic, but the disease is usually diagnosed in childhood and the phenotype of the massive lymphoproliferation is $CD4^{neg} CD8^{neg}$.[49,50]

Table 1. *Clinical characteristics of T-LGL leukemia and CLPD-NK.*

Type	Hematological findings	Flow cytometry markers	Clonality analysis	Clinical features	Auto-immune conditions
T-LGL	PB LGL count $>0.5 \times 10^9$/l. In appropriate clinical context PB LGL count can be $<0.5 \times 10^9$/l.	CD3+ CD8+ CD16+ CD57+ TCR-α/β+	PCR method (TCR-γ) or flow cytometry (Vbeta)	Splenomegaly, cytopenias, recurrent infections, splenomegaly, LGL infiltrates in BM and liver, B cell dyscrasias	Rheumatoid arthritis common
CLPD-NK		$CD3^{neg}$ CD16+ CD56+	Flow cytometry KIR analysis (not common in the clinic)		Rheumatoid arthritis rare

Abbreviations: BM, bone marrow; KIR, killer cell immunoglobulin-like receptor; PB, peripheral blood; TCR, T cell receptor; Vbeta, T cell receptor beta chain variable region.

3. Treatment of LGL leukemia

The pharmacological therapy in LGL leukemia is connected to the related cytopenias and concomitant autoimmune disorders: the indications for treatment include severe neutropenia (neutrophil count repeatedly $<0.5 \times 10^9$) or moderate neutropenia with infections, symptomatic anemia, and related autoimmune diseases (most often RA).[51] Prospective large trials concerning the selection of treatment have not been published and no standardized therapy guidelines exist. Current treatment with immunosuppressive agents is based on data from retrospective series. A schematic suggestion for LGL leukemia treatment is presented in Figure 2.

3.1. Immunosuppressive treatment

The most commonly administered first-line treatment is low-dose oral methotrexate (MTX), which has also been used to treat RA.[52] The first report of MTX in LGL leukemia (including 10 patients) described overall response rate (ORR) of 60%, and responses were durable during follow-up (range 1.3-9.6 years).[53] In the French cohort, ORR was also 55%.[12] However, relapse rates were marked: 12/18 (67%) of the patients who were followed for more than one year did not maintain remission.[12] If a response is achieved, continuous MTX is recommended.[51]

The alkylating agent cyclophosphamide has been especially used as a second-line treatment: ORR in the French cohort was 66%, and 11/15 patients responded even after failure to first-line MTX.[12] It was first used to treat LGL leukemia with concomitant PRCA, in which it was effective, with 75% (6/8) ORR.[54] In the first-line setting, the results are also promising: ORR was 71% (32/45) and 47% achieved hematologic CR.[55] ORR was similar in T-LGL leukemia (72%) versus CLPD-NK (68%) and in neutropenic patients (72%) versus anemic (67%) patients.[55] The median treatment duration for responders was 8.5 months: long cyclophosphamide-monotherapy regimens are not recommended because of the risk of secondary malignancies such as myelodysplastic syndrome (MDS) or acute myeloid leukemia (AML).[56] The relapse rate was low (13%) and grade 1 and 2 toxicities were seen in 18% of patients.[55]

Cyclosporine A is also used to treat LGL leukemia, as an alternate therapy to MTX and cyclophosphamide.[51] ORR in LGL leukemia with PRCA is comparable to cyclophosphamide, but relapse is common if treatment is discontinued.[54] In another study, the ORR was much poorer: only 21% after three months of treatment.[12]

3.2. Other therapies

The effects of glucocorticoids such as prednisone are not remarkable in the treatment of LGL leukemia: in a group of 22 LGL leukemia patients, only two achieved PR or CR with first-line steroid therapy.[12]

Neutropenia is associated with recurrent, even fatal, bacterial infections in LGL leukemia.[12] Granulocyte colony-stimulating factor (G-CSF) has some efficacy in elevating neutrophil levels in case reports and small patient series.[57,58] Prospective studies concerning the efficacy of G-CSF in neutropenic fever and sepsis of LGL leukemia patients have not been done. The effects of G-CSF have been studied in neutropenia related to Felty's syndrome, where it improved neutrophil counts.[59] The same authors performed a systematic review of literature, and G-CSF proved to be effective at elevating neutrophil counts, which was often associated with improvement of infectious complications.[59] In the French cohort, 14 patients had a poor or transient response in neutrophil levels when G-CSF was used as a monotherapy.[12]

Alemtuzumab is a recombinant humanized monoclonal antibody directed against the pan-lymphocyte antigen CD52. It has also been used in LGL leukemia with some success, and even some refractory patients have had durable responses in case reports.[60,61] Studies of larger patient cohorts have not been published.

LGL leukemia patients with splenomegaly have multilineage cytopenias with higher frequency compared to patients with normal spleen size, and splenectomy induces a hematological response so that patients achieve transfusion-independence.[62] First-line splenectomy was efficient in 4 of 10 patients studied in the French cohort.[12]

3.3. Evaluation of treatment response

The treatment response should be evaluated four months after the onset of the therapy.[51] In a hematologic complete response (CR), the blood counts are normalized: hemoglobin >120 g/L, neutrophil count >1.5×10^9/l, lymphocytes <4×10^9/l, neutrophils >1.5×10^9/l, and circulating LGLs are in the normal range.[51] A complete molecular response (CMR) is characterized by the disappearance of the T cell clone using a PCR method.[51] The term hematologic partial response (PR) is used when blood counts are improved, but do not fulfill the criteria for CR.[51] The term treatment failure is used when none of the response criteria are met during first four months of therapy.[51]

3.4. Prognosis of LGL leukemia

The prognosis of LGL leukemia is relatively good: most patients display a non-progressive chronic clinical course despite the presence of severe cytopenias, infections, and autoimmune disorders.[63] The survival rates vary between different studies. In the publication originally describing LGL leukemia, 9 of 25 patients died, primarily due to infections, during follow-up of over two years, and in the French cohort 25/229 patients died, related mostly to neutropenic sepsis.[7,12] Median survival was reported to be 161 months in one study.[64] The majority of patients required treatment at some time during the follow-up of ten years.[51]
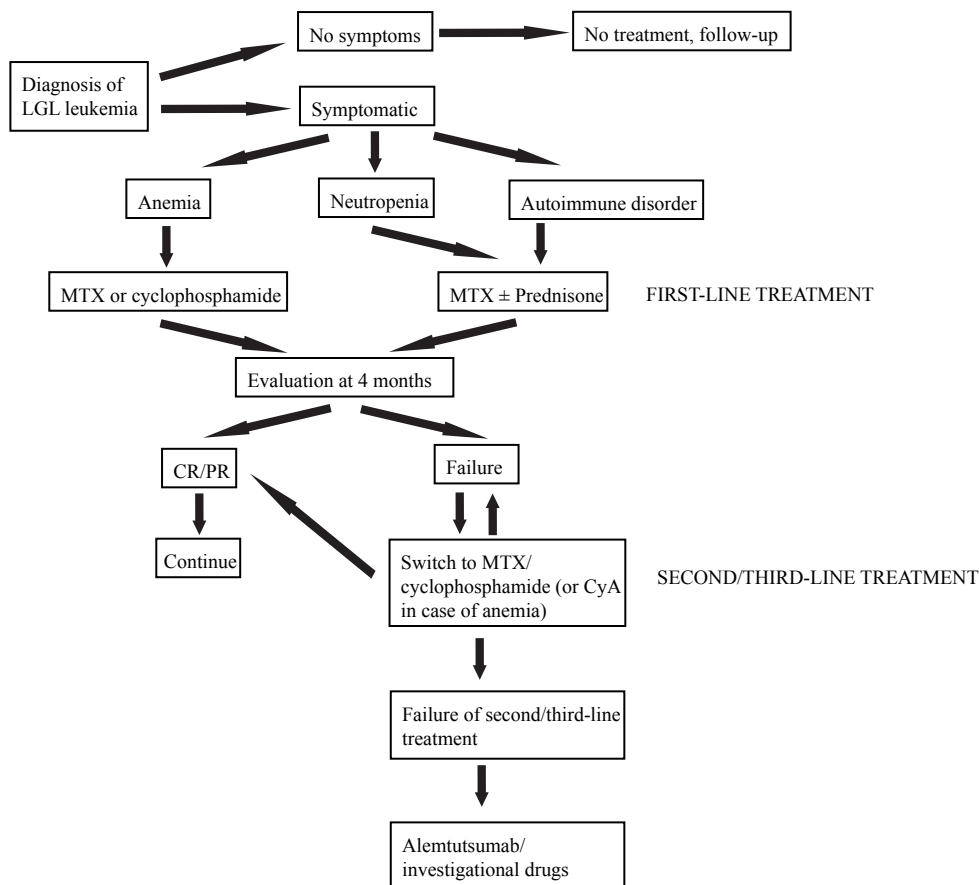

Figure 2. *Schematic representation of LGL leukemia treatment (modified from the review by Lamy and Loughran, Blood, 2011[51])*

4. Survival signaling in LGL leukemia

4.1. T cell receptor rearrangement and antigen specificity

The etiology of LGL leukemia is unknown. The phenotype of T-LGL, the effector-memory cytotoxic lymphocyte, suggests antigen stimulation behind leukemic expansion. The vast majority of T-LGL leukemia expansions express α/β-type TCR on their surface, and TCR-γ/δ LGL leukemia is rare.[16] The β chain is composed of rearranged variable (V), diversity (D), and joining (J) segments and a constant region, whereas the α gene consists of only V and J segments.[65] During T cell development, the TCR-α/β genes rearrange in a unique manner, and with junctional deletions and insertions, this recombination produces an incredible number of different combinations, thus creating the vast pool of T cells recognizing different antigens.[65] TCRα/β forms the TCR-CD3 receptor complex with different CD3 chains, in which TCRα/β is responsible for the antigen recognition and the CD3 chains transduce the signal inside cell, enhanced by coreceptors (CD4 or CD8). The TCRβ chain has three complementarity-determining regions (CDRs): CDR1 and CDR2 interact with antigen-presenting major histocompatibility (MHC) molecules stabilizing the antigen-TCR ligation while the hypervariable CDR3 region is responsible for the actual antigen recognition.[66,67] When a T cell encounters its specific antigen for the first time, contact between MHC/antigen complex of antigen-presenting cell and TCR of T cell causes changes in TCR and coreceptor conformation, resulting in binding of the cells to each other with high affinity. Simultaneously, different cytokines related to the immune response strengthen and guide the T cell activation and proliferation into effector T cells, which are capable of exerting different functions required for the clearance of the antigen, and do not need further co-stimulatory receptors in order to act if they encounter the antigen again.

A few studies have concentrated on elucidating the possible antigen target behind leukemic T-LGL expansion by using the analysis of TCRβ CDR3 repertoire. However, a specific antigen can be recognized by a large number of different TCR types and the variation in peptide-presenting HLA genes between individuals also affects the recognition, but in the context of the same HLA type, TCRβ CDR3 rearrangement seems to be typical for the antigen target.[68-70] Some T-LGL leukemia patients are reported to have identical TCRβ immunodominant clonotypes; sharing of similar major clonotypes or smaller expansions is even more common. In contrast, the physiological TCR repertoire in healthy individuals is diverse and common CDR3 rearrangements are not normally detected.[71]

The immunodominant clones in T-LGL leukemia can sometimes change through a phenomenon called clonal drift.[72] In a follow-up study with quantitative Vβ flow cytometry analysis of the TCR clonality, the patients represented four different patterns of TCR Vβ development during the course of the disease: of 71 analyzed T-LGL leukemia patients, 54% and 10% had either persistent large LGL expansion or had a slowly rising proportion of the same immunodominant clone during follow-up.[72] The rest had another type of TCR Vβ pattern: 20% of the patients displayed a rise in the major clone from the initially polyclonal-appearing background, and in 17% of the cases, the major clone disappeared while another clonal expansion simultaneously evolved.[72]

Despite the data suggesting restricted use of TCR repertoire in LGL leukemia, no common antigens causing clonal persistent lymphocyte expansions have been recognized. Considering infectious agents and LGL leukemia, the connection between

lymphoid malignancies and certain pathogens is well established. For example, Epstein-Barr virus (EBV) is often associated with rare aggressive NK-LGL leukemia and Hodgkin lymphomas.[73,74] Increased levels of human T lymphotrophic virus (HTLV) antibodies have been detected in T-LGL leukemia patients' sera (44% vs 0.12% in healthy controls), but only a minority of these patients (7.5%) were actually infected with HTLV-2 when analyzed with a HTLV-2-specific PCR method.[75] This could be explained by serologic crossreactivity to a cellular or retroviral antigen showing some amino acid sequence homology with the HTLV antigen. Similar seroreactivity to HTLV is seen in CLPD-NK patients sera, but no trace of HTLV nucleic acids were detected by PCR.[76] Rheumatoid arthritis, other autoimmune diseases, and autoimmune cytopenias are overrepresented in LGL leukemia patients, but no direct evidence of crossreactivity between pathogens and autoantigens exists.

4.2. NK cell receptors in CLPD-NK

NK cells are part of the innate immune system and are involved in immune surveillance against virus-infected and tumor cells. NK cells express various activating and inhibitory cell surface receptors, which control their ability to directly kill target cells and produce different cytokines. Normal NK cells express three types of receptors: killer cell lectin-like receptors (KLRC), killer cell immunoglobulin-like receptors (KIR), and natural cytotoxic receptors (NCR).[77] KIRs are divided into activating and inactivating receptors: Activating KIRs detect infectious ligands such as virus particles and also self-ligands from, for example, transformed malignant cells, whereas inhibiting KIRs ensure tolerance to self, but also give the ability to recognize the "missing self" on transformed cells.[78] NK cells express different co-receptors such as CD56 and CD16. The CD56^{dim}CD16+ NK cells display potent cytolytic function when activated, whereas the CD56^{bright}CD16$^{+/-}$ NK cells, which are thought to be the precursors of CD56^{dim} cells, are poorly cytolytic but secrete different cytokines.[79]

In healthy individuals, the NK cell receptor repertoire differs between individuals due to genetic background, but is consistent within one individual.[80,81] As a rule, every NK cell has at least one inhibiting NK cell receptor to ensure self-tolerance.[80,81] In a study of CLPD-NK, 11 of 18 patients expressed homogenously one KIR on the surface of NK cells, and further examination of *in vitro* cultured NK cell lines derived from 7 patients revealed the homogeneous expression of at least one activating KIR.[82] In another study, CLPD-NK cells also expressed activating KIR and the expression of inhibitory receptors, determined by RT-PCR, was lower in comparison to normal cells.[83] Leukemic NK-LGLs possess potent cytolytic activity in cytotoxicity assays, which could be related to the increased activating–to-inhibitory KIR ratio, and resulting inappropriate cytotoxicity or cytokine secretion can impact CLPD-NK pathogenesis.[83]

4.3. Impaired FAS/FASL-induced apoptosis

The encountering of its specific antigen by a T cell results in rapid proliferation of clonal effector cells during the first few days after the antigen recognition. These cells could cause harm if the response continues uncontrolled; therefore, activated T cells undergo activation-induced cell death (AICD), which leaves only a small memory cell population and is important for the maintenance of self-antigen tolerance.[84] When cytotoxic lymphocytes are activated, transcription of both FAS and FASL is

upregulated. Binding of FASL to FAS results in trimerization of FAS, which in turn enables binding of FAS-associated protein with death domain (FADD) to the intracellular part of the FAS trimer. The FAS trimer-FADD complex is called the death-inducing silencing complex (DISC).[84] The proapoptotic protein procaspase-8 binds to DISC and is cleaved to its active form caspase-8, which is capable of activating the downstream apoptotic pathway.[84]

One of the major molecular defects in LGL leukemia is resistance to FAS-mediated apoptosis, despite high levels of FAS and FASL (Table 2). [4,85] Impairing mutations in *FAS* are associated with lymphoproliferative disease and autoimmunity, but have not been found in LGL leukemia.[85,86] After PHA and interleukin-2 (IL-2) stimulation, LGL cells from almost all patients gained sensitivity to FAS-mediated apoptosis, suggesting that the actual FAS/FASL machinery in leukemic LGL is intact.[85]

4.4. Activation of STAT3

Aberrant and constitutive STAT3 activation is common in different malignancies and is caused by overactive tyrosine kinases arising from genetic or epigenetic alterations. The effects of constitutively active STAT3 depend on the cell type and co-activating proteins.

STAT3 is constitutively activated in LGL leukemia and seems to be connected to impaired sensitivity to FAS-mediated apoptosis: in electrophoretic mobility shift assays (EMSAs), STAT3 was localized in the nuclei of all (n=19) T-LGL leukemia cells, mostly as the sis-inducible element (SIE)-binding STAT3:STAT3-homodimers. The STAT3 phosphorylation levels were also higher in LGL leukemia samples compared to healthy controls and inhibition of STAT3, either by a JAK-inhibitor or direct STAT3-antisense-nucleotide method, restored FAS-mediated apoptosis and was associated with decreased expression of anti-apoptotic Myeloid cell leukemia-1 (MCL-1) protein (Table 2).[4]

4.5. PI3K/RAC/PAK/MAP2K-, PI3K/AKT-, and PI3K/NFκB-pathways

The Phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K)/Ras-related C3 botulinum toxin substrate (RAC)/P21 protein (Cdc42/Rac)-activated kinase (PAK)/ Mitogen-activated protein kinase kinase (MAP2K)-, PI3K/V-akt murine thymoma viral oncogene homolog (AKT)-, and PI3K/Nuclear factor kappa B (NFκB)-pathways are antiapoptotic signaling cascades that play critical roles in the transmission of signals from cell surface and intracytoplasmic receptors to regulate gene expression.[87] Components of these pathways are mutated or aberrantly expressed in different malignancies, and activating mutations in upstream regulators can also induce these pathways.

The cytotoxic activation of NK cells starts by the ligation and recognition of a target cell to the NK cell receptors. This triggers the activation cascade PI3K/RAC/PAK/MAP2K, which leads to the mobilization of lytic granules containing granzyme B and perforin toward the target cell.[77,88] The inhibition of this pathway leads to apoptosis of leukemic LGLs (Table 2).[89] The reason for constitutive phosphorylation and activation of the pathway is not known, but it could be related to the increased activating-to-inhibitory KIR ratio seen in leukemic NK-LGL.[83,89] The pathway is also activated in T-LGL leukemia.[90]

PI3K is also an upstream regulator of AKT, and the PI3K-AKT pathway is activated in T-LGL, depending on Src family kinases (Table 2).[90] Inhibition of the pathway leads to rapid FAS-mediated apoptosis of leukemic LGLs.[90] In a network model of survival signaling in LGL leukemia, PI3K-inhibition led to NFκB-inhibition and apoptosis, independent of AKT.[91] NFκB was also shown to inhibit apoptosis through MCL-1, independent of STAT3.[91] The PI3K activation could be related to increased cytokine production: leukemic T-LGLs constitutively produce the proinflammatory cytokines Chemokine (C-C motif) ligand 5 (CCL5), Macrophage inflammatory protein-1 (MIP-1), and IL-18, which all have been shown to induce the PI3K.[90,92]

4.6. IL-15 and PDGF signaling

IL-15, along with IL-2, stimulates proliferation of T cells, induces development of cytotoxic T cells, and maintains NK cell population.[93] The IL-15-receptor (IL-15R) and IL-2R contain the same β and γ subunits in addition to the specific IL-15Rα or IL-2Rα subunits, and they both induce the same downstream proteins such as JAK1, JAK3, PI3K, and Spleen tyrosine kinase (SYK).[93] However, IL-2 and IL-15 have distinct roles in T cell homeostasis: IL-2 is involved in AICD, elimination of autoreactive T cells, and promotion of the development of regulatory T cells, whereas IL-15 maintains the activated CD8+ memory cell population.[94-96] IL-15 transgenic mice first develop CD3+ and CD3[neg] LGL expansions, which transform into fatal leukemia.[97] Similarly, transduction of retroviral IL-15 vector to human CD8+ cells leads to clonal, cytokine-independent growth of CD8+ cells.[98]

Platelet-derived growth factors (PDGFs) are able to activate a number of downstream signaling cascades such as PI3K-AKT and JAK-STAT, and dysregulated PDGF signaling is associated with multiple malignancies.[99,100] The PDGF family consists of four gene products, which form five different isoforms and act via two receptor tyrosine kinases: PDGF receptors α and β.[101]

In a network model of survival signaling in LGL leukemia, the known dysregulations and survival signaling network of normal cytotoxic lymphocytes were integrated into one model, and the prediction was that the initial T cell activation, constitutive presence of IL-15, and activated PDGF signaling can sustain leukemic LGLs and produce all known signaling dysfunctions in LGL leukemia (Table 2).[91] IL-15 has previously been connected to LGL leukemia pathogenesis: both CD3+ and CD3[neg] LGL were stimulated by IL-15, and IL-15 enhanced their cytotoxicity.[102] In that study, IL-15 was probably secreted by monocytes, and it was bound to the membranes of the LGLs.[102] PDGF levels were elevated in the sera of LGL leukemia patients compared to the healthy controls, and inhibition of PDGF induced apoptosis of leukemic LGL.[91]

4.7. Altered sphingolipid-mediated signaling

Gene expression data from normal and normal phytohaemagglutinin (PHA)- and IL2-stimulated PB mononuclear cells (MNCs) were compared with leukemic LGLs. Interestingly, normal stimulated PB MNCs show upregulation of both prosurvival and proapoptotic pathways, probably related to the strict control of the T cell population with AICD seen in normal lymphocytes after activation and proliferation.[103] In contrast, in leukemic LGLs, genes and pathways related to leukocyte activation were upregulated simultaneously with antiapoptotic pathways, indicating the failure of

AICD.[103] The most clearly upregulated genes in a comparison of healthy stimulated PB MNCs and leukemic LGLs were tumor necrosis factor alpha-induced protein 3 (TNFAIP3)(NFκB inducible gene), antiapoptotic MCL-1, and different sphingolipid-metabolism related genes (Table 2).[103] The authors were especially interested in sphingosine-1-phosphate (S1P), which is generated when ceramidase converts ceramide to sphingosine, and sphingosine kinase subsequently converts it to S1P.[104] A ceramide–sphingosine–S1P rheostat has a role in multiple malignancies: ceramide and sphingosine are proapoptotic, whereas S1P promotes cell survival.[104] Antiapoptotic S1P promotes tumor growth and inflammation through both intracellular and receptor mediated mechanisms.[104] Inhibition of acid ceramidase by N-oleoylethanolamine (NOE, mechanism of inhibition unknown) and administration of FTY720 (binds S1P-receptors with high affinity and induces S1P downregulation) induced apoptosis of leukemic LGLs and had little effect on normal PB MNCs.[103]

Table 2. Summary of dysregulated pathways in LGL leukemia. Superscript numbers refer to the cited references.

Pathway	Description
FAS-FASL	Leukemic LGLs are resistant to FAS-mediated AICD, despite high levels of circulating FAS-FASL.[4,85]
JAK-STAT	STAT3 is constitutively active in LGL leukemia, and its inhibition leads to restored FAS-sensitivity, decreased expression of antiapoptotic MCL-1 and apoptosis of leukemic LGLs.[4]
IL-15, PDGF	In a network model of survival signaling in LGL leukemia IL-15 and PDGF are able to produce all known dysregulations in LGL leukemia.[91]
PI3K/RAC/PAK/MAP2K	The activation of PI3K in LGL leukemia can be related to the proinflammatory cytokines secreted by leukemic LGLs.[92] The pathway is related to the release of lytic granules toward the target cell of a cytotoxic lymphocyte. The inhibition of this pathway leads to apoptosis of leukemic LGLs.[89]
PI3K/AKT	Src activates PI3K-AKT pathway in LGL leukemia, and PI3K-AKT inhibition leads to FAS-mediated apoptosis of LGLs.[90]
PI3K/NFκB	PI3K inhibition leads to NFκB inhibition and apoptosis of leukemic LGLs and decreased MCL-1 expression.[91]
Ceramide-sphingosine-S1P rheostat	Sphingolipid imbalance is seen in LGL leukemia. Pro-apoptotic ceramide is decreased and anti-apoptotic S1P is increased.[103]

5. JAK-STAT signaling

5.1. Overview of JAK-STAT signaling

JAK-STAT signaling conveys signals of extracellular signaling peptides such as growth factors and cytokines from the cell surface into the cell and to the transcription of genes in the nucleus (Figure 3). The activation of the JAK-STAT pathway stimulates cell proliferation and migration, differentiation, and apoptosis. Historically speaking, the research into the JAK-STAT pathway started in 1957 with the discovery of interferon (IFN), which is one of the major stimulants of the pathway. By the end of the 1990s, the genomic and first crystal structures of STATs were described.[105] During the 2000s, the role of both inactivating and activating alterations in the pathway, seen in inherited diseases (germline mutations) and during malignant transformation (acquired somatic mutations), was studied and led to a deeper understanding of the pathway functions.

The JAK-STAT pathway is activated when a ligand binds to an extracellular receptor and induces the multimerization of receptor subunits. The cytoplasmic domains of the receptors are associated with JAKs. JAKs represent a family of non-receptor tyrosine kinases, and four members have been discovered: JAK1, JAK2, JAK3, and tyrosine kinase 2 (TYK2). Based on their crystal structure, they have four protein domains called kinase domain (called also Jak homology 1 [JH1] domain), pseudokinase domain (JH2), Src-like homology 2 domain (SH2) and FERM domain.[106] Phosphorylation of tyrosine residues in the C-terminal kinase domain is essential for the binding of the substrate. The second domain, the pseudokinase domain, regulates JAK activity: for example, the JH2 domain of JAK2 autophosphorylates two inhibitotory residues, and its function is crucial for the maintenance of low-level JAK2 activity in the absence of cytokine-induced activation.[107] JAKs are activated when cell surface receptors multimerize and cause JAK-induced phosphorylation of both the cytokine receptors and their major substrates, STATs (Figure 3). STATs reside in the cytoplasm as latent monomers. Seven STAT family members have been recognized: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6. They all have a SH2 domain and a conserved tyrosine residue near the C-terminus, which is phosphorylated by JAKs (Figure 4). Phosphorylated STAT monomers dimerize through reciprocal phosphotyrosine-SH2 interactions and enter the nucleus via various mechanisms depending on the STAT protein. In the nucleus, the STAT dimers bind to specific DNA elements with other transcription factors and change the transcription of their target genes (Figure 3).
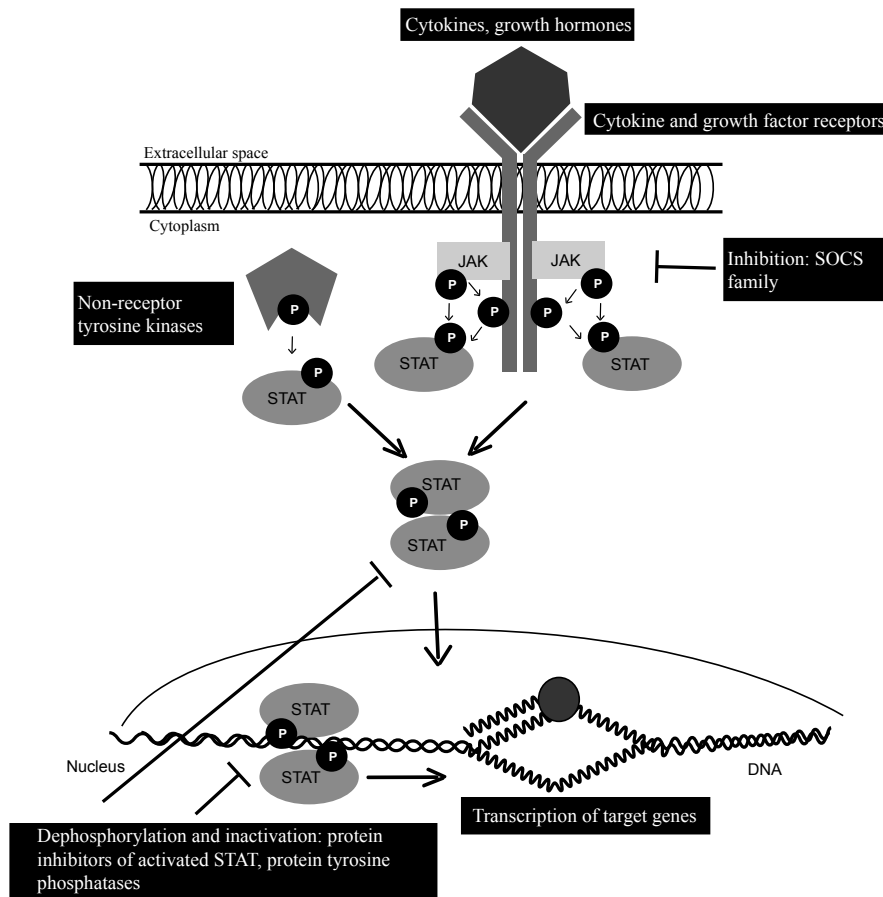
Figure 3. *Canonical representation of JAK-STAT signaling.*

5.2. Structure of STAT3, STAT5A and STAT5B

The *STAT*3, *STAT5A,* and *STAT5B* genes are located on the long arm of chromosome 17 in adjacent regions (17q11.2). The protein structure of STAT3 and STAT5A/B consists of coiled-coil, DNA-binding, linker, SH2, and transcriptional activation (TAD) domains (Figure 4). STAT3 exists in two major isoforms: full-length STAT3α is 770 amino acids long, while the shorter form STAT3β, arising from an alternative mRNA-splice site at exon 23, gives rise to a protein in which the last 55 C-terminal amino acid residues are replaced by 7 other residues. Thus, STAT3β lacks the TAD domain and seems to have specific functions of its own, as it can prevent the embryonic lethality of STAT3α-null mice and activate a set of STAT3 target genes.[108] The full-length STAT3α can exist in two isoforms: isoform 1 is one amino acid longer than isoform 2. STAT5A and STAT5B are separate genes. The coded proteins are 794 and 787 amino acids long, respectively, and are over 90% identical at the cDNA and protein level.[109] The data for proteolytically cleaved isoforms of STAT5A or STAT5B are controversial and their existence has not been proved with certainty.[110,111]

The dimer structure of STAT proteins is essential when they bind to the specific promoter sequences of the target genes. The crystal structure of a STAT3β homodimer bound to a short DNA fragment has been published (Figure 5, RCSB Protein Data Bank code 1BG1).[112] According to the publication, residues K591, R609, S611, and S613 form polar interactions with phospho-Y705 of the other STAT3 monomer and are conserved among STAT family proteins.[112] Phospho-Y705

and adjacent residues (702-709) are bound in a trans configuration to the opposite monomer, and the interaction is reciprocal (i.e., phosphotyrosine peptides of both monomers bind the monomers together).[112] The side chains of residues T708, F710, and C712 interact with the other monomer, and differences in the sequences of these side chains between different STATs may explain the differences in homo/heterodimerization tendency of the members of the STAT protein family.[112] Residues 643-644 and 647-648 also take part in the dimer interface, independent of the phosphorylation status, which may explain the loose dimer association occurring without phosphorylation, although the N-terminal domain also plays a role in unphosphorylated dimerization.[112] Another loop, which comes in close proximity to another STAT3 monomer, is situated nearer the C-terminal end, around amino acids 656-677, and cysteine substitutes that form sulfhydryl bonds between monomers at A662 and N664 create a constitutively active STAT3 molecule.[113]

The crystal structure of phosphorylated and DNA-bound STAT5A or STAT5B has not been reported yet. The structure of an unphosphorylated STAT5A dimer has been published (RCSB Protein Data Bank code 1Y1U).[114]
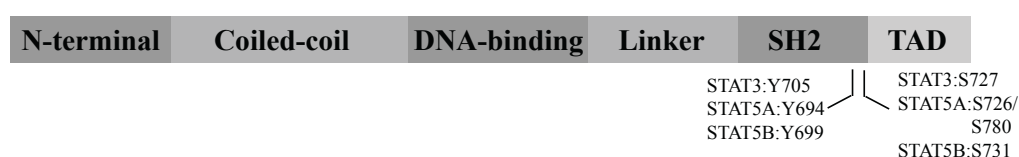
N-terminal	Coiled-coil	DNA-binding	Linker	SH2	TAD

STAT3:Y705
STAT5A:Y694
STAT5B:Y699

STAT3:S727
STAT5A:S726/ S780
STAT5B:S731

Figure 4. *Linear representation of common STAT protein structure. Tyrosine and serine residues that are phosphorylated during activation of STAT3 or STAT5A/B are marked in the structure.*

5.3. Activation and inactivation of STAT3

STAT3 was originally described as an IL-6 dependent transcription factor promoting acute-phase gene expression in IL-6 treated hepatocytes, but it can mediate signals from the entire IL-6 and IL-10 families with a notable variety, as well as G-CSF, leptin, IL-21, IL-27, protein tyrosine kinases such as Src and ALK, and receptor tyrosine kinases (epidermal growth factor [EGF], PDGF).[115,116] The IL-6 receptor complex consists of two molecules: IL-6 receptor (IL6R) in which IL-6 is bound and gp130 (called also IL6 signal transducer, IL6ST), which has a long cytoplasmic tail.[117] When IL-6 binds to IL-6R, it induces homodimerization of IL6ST: IL6ST does not have tyrosine kinase activity, but its cytoplasmic tail recruits JAKs, which can phosphorylate both IL6ST and STAT3.[118,119] The JAKs associated with STAT3 signaling are JAK1 and JAK2.[120] Activated STAT3 forms dimers through reciprocal phospho-tyrosine-SH2 domain binding (Figure 5).[121] Notably, unphosphorylated STAT3 can also exist in a homodimer form through N-terminal interaction, but the conformation of the dimer differs from the phosphorylated STAT3 dimer.[122] STAT3 can also be phosphorylated at S727: serine phosphorylation does not seem to influence the DNA binding of the STAT3 dimer, but it could affect the recruitment of other co-transcription factors.[123]

Access to the nucleus is limited by different nuclear pore complex proteins, which are located in the nuclear membrane. Small molecules can diffuse through, but the access of larger proteins such as STAT3 is restricted to those that possess a protein

sequence specific for nuclear import and export. Unphosphorylated STAT3 can also be transported to the nucleus. The transport mechanism of both the unphosphorylated and phosphorylated forms is related to the function of importins, the most important probably being the importin-β1-importin-α3 heterodimer on the nuclear membrane.[124] In the nucleus, STAT3 dimers bind to DNA-response elements of target genes, named IFN-γ-activated sequence (GAS) or SIE, and activate specific gene expression programs, which are often anti-apoptotic, pro-proliferative, and oncogenic.[125,126] Phosphorylated and activated STAT3 also activates transcription of itself, and increased amounts of unphosphorylated STAT3 protein can subsequently be chaperoned into the nucleus by NF-kB, where it binds to κB elements of gene promoters and can alter the transcription of a subset of genes such as CCL5 (known also as RANTES) and IL-6.[127]

STAT3 activation is reversed by different protein tyrosine phosphatases (PTPs), suppressors of cytokine signaling (SOCS), and protein activators of activated STATs (PIAS) (Figure 3).[128] There are eight members of SOCS protein family, and their expression is induced by different cytokines. As negative feedback regulators, they inhibit JAK-STAT signaling by directly binding to JAKs and cytokine receptors: SOCS3 seems to be the most important inhibitor of STAT3 function.[129] PIAS family proteins are located in the nucleus, and they inhibit and modify STAT3 signaling by blocking the DNA-binding of STAT3 and also by recruiting other transcription factors.[130] PTPs remove phosphate from phosphorylated tyrosine residue of STAT3.[131]

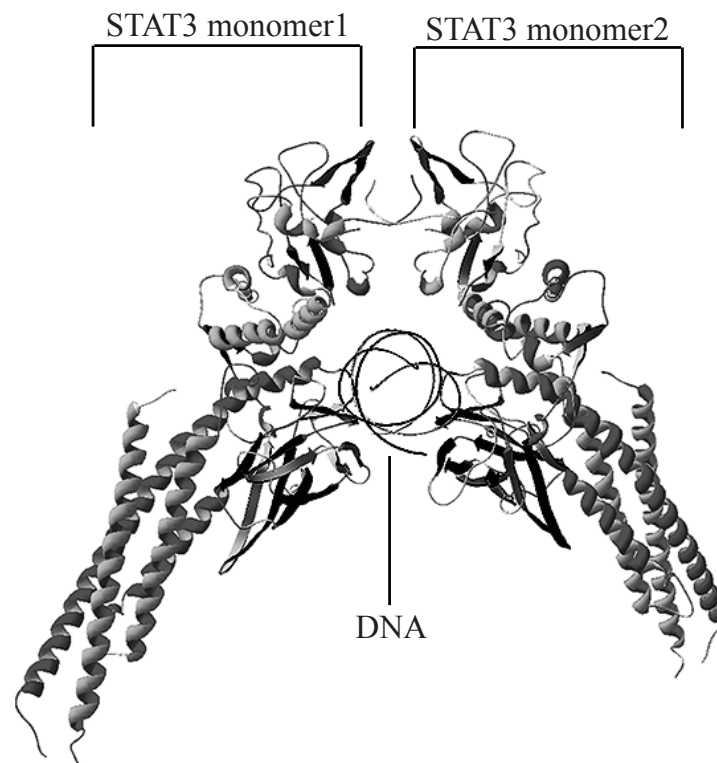**Figure 5.** *Three-dimensional model of dimerized STAT3, which has bound to the target DNA. The structure is predicted by Becker et al. Nature 1998[112], and the RCSB Protein Data Bank code is 1BG1. Figure created with Swiss-PdbViewer (Swiss Institute of Bioinformatics).*

5.4. Functions of STAT3 in different tissues

STAT3 is essential for development, and total Stat3 ablation leads to early embryonic lethality in mice.[132] Thus, the most widely used method to study the effects of *STAT3* depletion in different tissues is through targeted Cre-loxP recombination with tissue specific promoters in a murine model.[133] Targeted disruption of Stat3 function in mouse T-cells results in impaired IL-6-mediated survival and reduced proliferative response when stimulated with IL-2.[133] If Stat3 is absent in CD8+ lymphocytes, the levels of B cell leukemia/lymphoma 6 (Bcl6) and Socs3 are reduced and CMV-specific CD8+ T cells fail to transform into central memory T cells.[134] CD4+ T cell-specific *Stat3* deletion results in impaired IL17-producing effector T helper cell (Th17) response, whereas retroviral expression of constitutively active Stat3 promotes Th17 development.[135] Deletion of *Stat3* from neutrophils and macrophages promotes abolishment of the suppressive effects of IL-10 on inflammatory cytokines from neutrophils and promotes the production of inflammatory cytokines from macrophages in response to endotoxins: consequently, the Th1-type response is enhanced and elderly mice develop chronic enterocolitis.[133] B cell-specific *Stat3* deletion leads to reduced numbers of peripheral B cells and impaired plasma cell differentiation, resulting in a defective IgG response.[136,137] Targeted *Stat3*-deficiency in skin leads to a severely impaired hair cycle and wound healing.[133] In liver, Stat3 is essential for the induction of inflammatory genes.[138]

Heterozygous germline mutations in humans in DNA-binding and the SH2 domain of *STAT3* are seen hyper-IgE syndrome patients.[139,140] The syndrome is characterized by recurrent staphylococcal infections, eczema, cyst-forming pneumonias, and high serum-IgE levels. The current hypothesis is that mutated STAT3 in hyper-IgE syndrome works as a dominant-negative protein: PB MNCs carrying these mutations respond poorly to IL-6-stimulation, and co-expression of the mutated protein with wild-type STAT3 leads to impaired DNA-binding and transcriptional ability of STAT3.[139,140] The number of central memory CD4+ and CD8+ cells is reduced in hyper-IgE syndrome patients, related to the lower expression of BCL6 and SOCS3, and the development of a central memory T cell pool for varicella zoster and EBV is impaired, leading to poor control of latent infections.[141]

IL6-STAT3 activation has been observed in many autoimmune and chronic inflammatory diseases. T cell-expressed STAT3 is required for the Th17 immune response, and aberrant function of Th17 response is linked with STAT3 overexpression in RA, psoriasis, multiple sclerosis, and Crohn's disease.[142-145] Inflammatory cytokines induce an IL-6-STAT3 activation loop in RA, and STAT3 inhibition in a mouse model effectively reduced inflammation and destruction of synovial tissues.[146]

5.5. STAT3 in cancer

Tyrosine kinases such as JAKs are often overactive in cancer due to genetic or epigenetic alterations. Consequently, STAT3 activation is seen in many malignancies as a secondary event, one of the mechanisms being autocrine production of IL-6 or paracrine secretion of IL-6 by stromal or infiltrating inflammatory cells.[147] The effects of constitutively active STAT3 depend on the cell type and co-activating proteins. STAT3 has many effects that promote tumor growth: it downregulates transcription of the common tumor suppressor tumor protein p53 (TP53), activates transcription of downstream anti-apoptotic proteins including BCL-XL and MCL-1, and prevents cell cycle arrest through v-myc avian myelocytomatosis viral oncogene homolog

(MYC).[148] Aberrant STAT3 activation is often observed in head and neck cancer, multiple myeloma, breast cancer, and many hematological malignances.[149] STAT3 activation in AML shortens the time of disease-free survival, although it did not affect OS.[150] Interestingly, the oncogenic effects of STAT3 can be nongenomic and phosphorylation-independent: STAT3 is also located in the mitochondria, where it regulates cell metabolism and supports RAS-dependent malignant transformation.[151]

STAT3 has also oncogenic potential on its own: a spontaneously dimerizing and constitutively active STAT3-molecule, STAT3-C, was created by cysteine substitutes of A662 and N664 in the SH2 domain.[113] Mutated STAT3 is able to transform fibroblasts, and the transformed cells form tumors in nude mice.[113]

5.6. Activation and inactivation of STAT5A and STAT5B

Besides sharing the same chromosomal locus, *STAT5A* and *STAT5B* genes are also close functional relatives and share pleiotropic roles in mammary gland development, cytokine signaling, and immunoregulation in different tissues. STAT3 and STAT5A/B also exhibit the highest degree of homology to invertebrate STAT genes.[152] Of the two STAT5 proteins, STAT5A was first identified by its ability to mediate prolactin-induced upregulation of the expression of milk proteins. STAT5B, which was related to STAT5A with 96% sequence homology, was subsequently isolated. In this thesis, the term STAT5 refers to both STAT5A and STAT5B.

STAT5 proteins transduce signals mediated through common γ-chain cytokine receptor families (IL-2, IL-7, IL-9, IL-15, and possibly IL-21), and also through IL-3 (IL-3, IL-5, and GM-CSF) and single-chain (e.g. growth hormone, prolactin, thyroid peroxidase, and erythropoietin) cytokine receptor families.[116] STAT5A and STAT5B are also activated through receptor tyrosine kinases such as EGF and PDGF.[116] JAK3 is exclusively associated with common γ-chain receptors, whereas also other cytokine receptors and receptor tyrosine kinases can activate JAK1.[116] The effects of the single chain receptors are mediated by JAK2.[116]

Phosphorylated STAT5 dimerizes and translocates to the nucleus, probably through the Chromosome region maintenance (CRM1) nuclear transporter protein.[153] Similar to STAT3, STAT5 can also shuttle to the nucleus in monomeric form.[153] STAT5 binds to GAS motifs of the target genes, and in particular, STAT5A can also bind to DNA as a tetramer, probably enhancing the transcription in comparison with the STAT5A-dimer.[154] STAT5 regulates genes associated with apoptosis and cell proliferation: a recent study that utilized chromatin immunoprecipitation-sequencing (ChIP-Seq) showed high association between target genes of STAT5B, such as Dedicator of cytokinesis 8 (DOCK8), Forkhead box P3 (FOXP3), and IL2RA, and regulatory T cells (Tregs) and peripheral CD8+ cells and their activation, whereas STAT5A was involved in neural development and function.[155] Surprisingly, another ChIP-seq experiment showed that the number of shared binding sites of STAT3 and STAT5 in T cells is greater than the number of shared binding sites of STAT5 in T cells and other cells.[156] STAT5 regulates the CD8+ memory cell population by transmitting IL-2 and IL-15 signals.[157] The inactivation of STAT5 signaling is similar to that of the STAT3 pathway, involving the PTP, SOCS, and PIAS protein families.

5.7. Functions of STAT5 in different tissues

STAT5B is required for the development and normal function of lymphocytes. Germ-line deletion of either *Stat5a* or *Stat5b* gene in a mouse model does not lead to absolute embryonic lethality in the way that *Stat3* deletion does. The first studies done with double *Stat5a/b* deficiency used a gene targeting strategy leading to an N-terminally truncated and partially functional Stat5 proteins, and the results suggested that the effects of double deficiency were not that critical.[158] However, recent studies have shown that total *Stat5a/b* deletion in mice leads to high perinatal lethality, severe combined immunodeficiency, and a lack of lymphoid cells, further emphasizing the pleiotropic role of Stat5 proteins in immune system function.[159]

Stat5a$^{-/-}$ mice display impaired mammary gland development, whereas *Stat5b*$^{-/-}$ mice display impaired pituitary growth hormone production and consequently diminished body growth.[160,161] Considering the function of the immune system, BM-derived macrophages of *Stat5a*$^{-/-}$ mice exhibit defective GM-CSF–induced proliferation and gene expression in addition to other immunological defects, including a reduced number of T and NK cells *in vivo*.[162,163] The T-cell defect is associated with diminished IL-2–mediated signaling.[162,163] Stat5b$^{-/-}$ mice present with a diminished absolute number of NK cells and a maturation block in BM; the NK cell numbers are lower than in Stat5a$^{-/-}$ mice and NK-cell proliferation and cytolytic activity to IL-2 and IL-15 stimulation is poor.[164] Stat3 and Stat5 are important modulators of Th17 differentiation, and they bind to multiple common sites of the Il17 locus: in murine models, Stat3 enhances Th17 response, whereas Stat5b conversely represses the transcription of Il17.[135,165]

Impaired function of STAT5B caused by germline mutations has been described in humans in a few case reports: the patients are lymphopenic and have low numbers of CD4+CD25^{high} Tregs, NK cells, and γδ-T cells.[166-168] Clinically the patients present with growth retardation, interstitial pneumonia, severe eczema, and autoimmune disorders that are thought to be related to the dysfunctional T-regs.[168]

5.8. STAT5 in cancer

Constitutive STAT5 activation is an essential step in many malignancies such as CML and other myeloproliferative disorders (MPDs).[169,170] In CML, oncogenic BCR-ABL1 tyrosine kinase activates STAT5, and anti-apoptotic effects are conferred through transcription of antiapoptotic protein BCL-XL.[130,171] ABL oncogenes do not induce malignant transformation in STAT5A/B$^{null/null}$ cells.[172] Accordingly, STAT5 inhibition induces apoptosis of CML cells.[173] In other MPDs, STAT5 activation occurs through an activating JAK2 mutation, most commonly V617F, which has been observed in over 90% of patients with polycytemia vera and in 30-50% of cases with essential thrombocytosis or primary myelofibrosis.[170] The V617F mutation is located in the JH2 domain of JAK2, and it abrogates the inhibitory function of the JH2 domain.[107] Downstream STAT5 activation is an essential step in the oncogenic transformation through V617F-mutated JAK2.[174] Somatic *JAK*-mutations have been found in both adult and pediatric acute lymphoblastic leukemia (ALL) cases: the *JAK1/2/3*-mutations were seen in 20% of pediatric high-risk *BCR-ABL1*-negative cases and they caused constitutive STAT5 activation and cytokine-independent growth of Ba/F3-EpoR cell line.[175,176]

Two constitutively active STAT5 mutants (STAT5-C) have been described. The first of these has two amino acid substitutions, H299R in DNA-binding domain and S711F in TAD.[177] Cells transfected with STAT5-C$^{H299R+S711F}$ show increased STAT5

phosphorylation and transcriptional activity, and it sustains growth of IL-3-dependent cells in the absence of IL-3.[177] The other published STAT5-C was created by one missense mutation N642H in SH2 domain, and it has similar properties as STAT5-C$^{H299R+S711F}$ both biochemically and biologically.[178] Notably, phosphorylation and dimerization are essential for the aberrant function of the STAT5-C^{N642H}, and disruption of Y694 abolishes the acquired properties of STAT5-C^{N642H}.[178]

6. Cancer genetics and next-generation sequencing

6.1. Oncogenes and tumor-suppressors

Cancer results from the accumulation of genetic changes causing transformation of a normal cell.[179] These alterations can occur in different types of genes, which can be divided into three groups based on their function: oncogenes, tumor-suppressor genes, and stability genes. Several genetic changes are usually required for malignant transformation because mammalian cells have several mechanisms that protect cells from the effects of harmful mutations. The mutations of oncogenes typically create a constitutively active form of the coded protein through chromosomal translocations, gene amplifications, or even by a single amino acid change in a crucial spot in the protein, and a heterozygous mutation is often sufficient to alter the function.[179] On the other hand, usually both copies of the tumor-suppressor genes are damaged and inactivated by missense mutations in essential amino acids, deletions/insertions and larger chromosomal aberrations, or epigenetic changes.[179,180] The function of stability genes is related to chromosome and DNA maintenance and error repair, and mutations in these genes increase the mutation rate of a cell due to genomic instability or increased vulnerability to the effects of mutagens.[179] Lastly, genetic variants can be transmitted through the germline, in which case they are present in all the cells of an individual and contribute to inherited cancer susceptibility.[181]

The term epigenetics refers to chromatin-based events, which regulate processes related to DNA template. These include for example DNA methylation, histone modification, and non-coding RNA molecules. Epigenetic pathways also play a significant role in oncogenesis alongside with genetic changes.[182]

6.2. Next-generation sequencing technology

Automated capillary (known as Sanger or chain-termination method) sequencing had been the gold standard of sequencing for over two decades and its accomplishments include the sequencing of the first human genome.[183] Newer rapidly developing high-throughput sequencing methods are referred to as next-generation sequencing (NGS) or second-generation sequencing and their major benefit is the ability to produce a vast amount of data rapidly and inexpensively. The most popular methods are whole genome sequencing, whole exome sequencing (WES), RNA (transcriptome) sequencing, targeted sequencing, and ChIP-seq.[184] In this review, the Illumina NGS platform will be used as an example.

NGS technology consists of several steps which can be grouped as template preparation, sequencing and imaging, genome alignment and assembly, and data analysis.[184]

The template preparation method depends on the NGS application used. For example, in WES, the goal is to sequence the coding areas, consisting of 1% of the total genome. One method is to fractionate the DNA, ligate adaptors to the ends of the fragments, and subsequently capture fragments containing the exons using a kit consisting of custom oligonucleotides with magnetic beads.[185,186] In targeted sequencing, primers with adapter tails are used to produce enriched amplicons of the targeted area, with adapter sequences at both ends of every amplicon. Illumina technology uses solid-phase amplification of the DNA fragments: fragments are attached to a glass slide by the adapter sequences.[187] Adapter-specific forward and reverse primers are attached to the slide to produce 100–200 million spatially separate template clusters with bridging amplification strategy.[184,187]

Clusters consisting of identical templates are essential to the sequencing, because most imaging systems cannot detect single fluorescent signals.[184] In the sequencing and imaging step, the Illumina platform uses cyclic reversible termination, in which a fluorescent nucleotide with a terminating group is added to the primer-originating template, and imaging by lasers is performed to identify the nucleotide by fluorescence. The subsequent cleavage step removes the terminating group, enabling the start of the next incorporation cycle.[184]

The nucleotide sequence of a template is called a read (the length being approximately 25-250 bp in Illumina NGS platforms). The reads are stored as a text-based FASTQ file format, which contains both the sequence and quality scores (Phred score, in which score 20 corresponds to a 1% error rate) for every nucleotide.[188] After the sequencing step, every raw read is aligned with a known reference sequence.[189] Different tools are used in this process, one of which is called the Burrows-Wheeler aligner (BWA), a software package for mapping short sequences against a large reference genome, such as the human genome.[190] Sequencing data can be also visualized using different programs such as Integrative Genomics Viewer (IGV) but due to the vast amount of data, the next step (i.e., calling of variants from sequenced reads) is done with specific identification algorithms such as Varscan, which is developed for nomination of putative mutations, small insertions/deletions ("indels"), and copy number alterations.[191-193] In the case of cancer exome/genome analysis, comparison of the patient's tumor sample variants with the same patient's germline data is essential for the identification of somatic variants and it can be accomplished with Varscan2 or similar software.[191,194] Putative somatic mutations are annotated for functional consequences at gene and protein sequence levels, their conservation level in the evolution is estimated, and common variants seen in public databases such as 1000 Genomes Project and dbSNP are identified and then excluded from further analysis.[195]

6.3. Bias and error in NGS

Every NGS platform has its own error profile. The quantity and quality of the source DNA affects the error rate and has to be taken into account: for example, compared to fresh samples, tissue samples stored as formalin-fixed and paraffin-embedded (FFPE) blocks present a challenge for successful recovery of nucleic acids, which can be overcome with correct DNA isolation techniques.[196] Template amplification can introduce mutations to DNA templates, which are then considered as true variants of the source DNA or can affect the correct alignment of the read.[197] Parallel analysis of a reference DNA sample with a known sequence gives an estimate of the error profile of the actual sample sequenced and analyzed.

The amount of sequencing data can be described with the term coverage or depth, which means the average number of times a base pair is sequenced in the experiment. The coverage seems to associate with certain genomic locations (with high GC content) in a reproducible way, and this can cause bias, especially in experiments in which coverage levels of different regions are compared (RNA sequencing, ChIP-seq).[198] The detection of single nucleotide variants (SNVs) and structural variants in a low-coverage area is more difficult, and the minimum median coverage recommended for NGS experiments is x20-30 to ensure proper detection of variants.[199]

The data for the newest Illumina instruments have not been published, but earlier Illumina sequencers such as GA and HiSeq2000 seem to produce mostly substitution errors, especially in GC-rich areas, while false indels are rarer.[200] The error rate usually increases towards the end of the sequence (B-tail) and the quality of the data can be increased by trimming away the low-quality tails.[198,200] This is explained by the fact that the cyclic reversible termination process can become desynchronized if the sequencing reaction in a cluster proceeds in dyssynchrony between different copies of a template, and this error type is prone to pile up towards the end of the template when dyssynchrony increases.

AIMS OF THE STUDY

The overall aim of this PhD project was to uncover the pathogenesis of LGL leukemia and find new molecular markers, which can be used in the diagnostics and targeted therapy.

The project included four parts:

1) The analysis of somatic genomic changes of LGL leukemia expansion in T-LGL leukemia patients

2) Further characterization of *STAT3* mutations and their clinical impact in T-LGL leukemia and CLPD-NK

3) The pathogenesis of LGL leukemia in *STAT3* mutation-negative patients

4) The clonal hierarchy of expanded lymphocytes in LGL leukemia and the impact of *STAT3* mutations on treatment responses

PATIENTS AND METHODS

7. Patients and ethical permissions

T-LGL leukemia and CLPD-NK patients included in studies I-IV were diagnosed based on WHO 2008 guidelines; the diagnostic criteria are explained in detail in the review of the literature of this thesis.[9] The samples were collected at Helsinki University Central Hospital, Oulu University Hospital, Penn State Hershey Cancer Center (US), and Cleveland Clinic (US). The studies were conducted according to the Declaration of Helsinki and were approved by the Ethics Committees of Helsinki University Central Hospital, Penn State Hershey Cancer Center (US), and Cleveland Clinic (US). Written informed consents were obtained from patients and healthy controls in all studies.

Study I aimed to analyze the molecular background of LGL leukemia by WES of leukemic LGLs and paired CD4+ cells. The *STAT3* mutation of the index T-LGL leukemia patient was subsequently screened in a cohort, which consisted of 76 additional T-LGL patients with a large immunodominant clone. In addition, 8 healthy controls, 10 AML patients, 8 ALL patients, 4 CML patients with reactive LGL expansion during dasatinib therapy, 2 CMV-infected patients, and one graft-versus-host patient were analyzed as controls. DNA used in the *STAT3* mutation analysis was extracted either from whole blood or PB MNCs. Clinical and functional consequences of *STAT3* mutations were analyzed using patient samples and Hela-cell line.

Study II included a larger LGL leukemia cohort: 120 T-LGL leukemia (the patients already analyzed in study I were excluded from this study) and 50 CLPD-NK patients. Samples from 31 patients with idiopathic neutropenia not fulfilling all diagnostic criteria for LGL leukemia were analyzed as well. The *STAT3* mutation status was screened from the study cohort using capillary sequencing, and the source DNA was extracted from whole blood or PB MNCs. Due to the larger numbers of patients, clinical correlations were analyzed in greater detail in addition to the functional effects of the mutations.

In study III, the molecular background of LGL leukemia in *STAT3* mutation-negative patients was analyzed by WES, followed by screening of novel *STAT5B* mutations from 173 T-LGL leukemia and 38 CLPD-NK patients. Additional control samples included 12 patients with reactive LGL lymphocytosis and 2 T-ALL cell lines (Jurkat and MOLT-4). HEK-cells were used in the functional analysis of *STAT5B* mutations.

Study IV used deep targeted *STAT3* sequencing and deep T cell receptor beta chain (TCRB) sequencing to analyze the clonal hierarchy in both T-LGL leukemia and CLPD-NK. The original screening of *STAT3* mutations was conducted on whole blood and PB MNC DNA. In addition, sorted lymphocyte fractions were analyzed from 13 *STAT3* mutation-positive patients. The study cohort included both T-LGL leukemia (n=174) and CLPD-NK (n=39) patients. As a control, DNA was extracted from 8 CML patients with clonal LGL lymphocytosis during dasatinib treatment and from BM aspirates of 10 healthy controls.

Some of the patients included in studies III and IV were the same ones analyzed in studies I and II.

8. Mononuclear cell separation

Mononuclear cells were separated from peripheral blood using Ficoll-Paque gradient separation (GE Healthcare) according to the manufacturer's instructions. The resulting PB MNCs were either used fresh, stored as pellets (at -70C), or cryopreserved in fetal bovine serum containing 10% DMSO.

9. Flow cytometry analysis and selection/sorting of lymphocytes

9.1. Magnetic bead separation

In studies I and III, PB MNCs were labelled with magnetic CD4 or CD8 MicroBeads (Miltenyi Biotech) and separated with an AutoMACS Cell Sorter (Miltenyi Biotech). In study III, NK cells were isolated with negative selection, which depletes T cells, B cells, monocytes, platelets, dendritic cells, granulocytes, and erythrocytes (Dynabeads Untouched human NK cells kit, Invitrogen). The purities of the sorted fractions were analyzed using flow cytometry (FACSAria, Beckman-Coulter Immunotech).

9.2. Fluorescence-activated cell analysis and sorting

The TCR Vβ families were analyzed in studies I-IV using whole blood, fresh PB MNCs, or PB MNCs stored in liquid nitrogen. Samples were stained with a combination of monoclonal antibodies (mAb) against CD3, CD4, CD8 (BD Biosciences), and a panel of TCR Vβ mAbs corresponding to the 24 different families of TCR Vβ CDR3 region (Beckman-Coulter Immunotech)(Table 3). The kit recognizes about 70% of the normal human TCR Vβ repertoire: it consists of 8 tubes, each containing three different mAbs (PE-conjugate, FITC-conjugate, PE/FITC conjugate). Samples were analyzed with FACSAria/FACSAria II flow cytometers and FACSDiva software (Beckman-Coulter Immunotech) in studies I, III, and I. Study II utilized an FC500 flow cytometer with CXP Version 2.2 software (Beckman Coulter) and FCS Express Version 3.0 analysis software (De Novo Software).

The KIR phenotype was analyzed in CLPD-NK cases in study II with mAbs targeting KIR2DL1, KIR2DS1/2/4, KIR2DL2/3/4, KIR3DL1, and NKG2A/D, using a 2- or 4-color system (antibodies from BD Biosciences, Beckman-Coulter Immunotech, BioLegend).

The analysis and sorting of lymphocyte subpopulations by flow cytometry (FACSAria/FACSAria II) was done in studies I-IV with different mAb panels targeting lymphocyte surface antigens (antibodies from BD Biosciences and Beckman-Coulter Immunotech)(Table 3). In addition, T-LGL clones were sorted in study IV using appropriate mAbs from TCR Vβ panel of Table 3 (Beckman-Coulter Immunotech). The purities of the sorted fractions were checked with flow cytometry.

Table 3. Different flow cytometry antibody panels used in studies I-IV.

TCR Vbeta analysis	Lymphocyte subpopulations 1	Lymphocyte subpopulations 2	Lymphocyte subpopulations 3
CD3 APC	CD45 APC-Cy7	CD45 PerCP	CD45PerCP
CD4 PerCP	CD3 APC	CD3 FITC	CD3 FITC
CD8 PE-Cy7	CD8 PE-Cy7	CD4 APC	CD19 APC
TCR Vbeta PE	CD19 Pacific Blue	CD8 PE	CD16/56 PE
TCR Vbeta PE+FITC	CD16 PE		
TCR Vbeta FITC	CD56 PE-Texas Red		

10. DNA and RNA extraction

DNA was extracted from whole blood, PB MNCs, or sorted lymphocyte populations using Genomic DNA NucleoSpin Tissue kits or Tissue XS kits (when the number of cells was <100 000) (Macherey-Nagel). RNA was extracted with Total RNA Isolation: NucleoSpin RNA II kits (Macherey-Nagel) or miRNAeasy kits (QIAGEN). RNA and DNA concentrations were measured with Nanodrop (Thermo Fisher), Qubit 2.0 (Life Technologies), or Agilent 2100 (Agilent Technologies) bioanalyzers.

11. Sequencing methods

11.1. Capillary sequencing

Primers for capillary sequencing were designed using Primer-Blast search (http://blast.ncbi.nlm.nih.gov/). PCR products were either extracted from gels or purified from PCR reactions using the QIAquick Gel Extraction Kit (QIAGEN) in studies I and III, or the Montage Cleanup Kit (Millipore) in study II. Purified PCR products were sequenced with the BigDye v.1.1. Cycle Sequencing kit and an ABI PRISM 3730xl DNA Analyzer. Chromatograms were analyzed with ChromasPro (Technelysium), 4Peaks (Mek&Tosj), and Sequencer (Gene Codes Corporation).

All STAT3/STAT5 capillary sequencing primers are presented in Table 4.

Table 4. STAT3 and STAT5 primer sequences.

Primer use	Primer target	Sequence (5'-3')
STAT3 capillary sequencing (Studies I, II, IV)	STAT3 exons 18 and 19 F	ATCTCCACCCACCAGGGGGC
	STAT3 exons 18 and 19 R	AGGGAAGGGCTGGGATGGCA
	STAT3 exon 20 F	TCCCATCGGTCACCCCAACA
	STAT3 exon 20 R	GCCAGGCCACTGAACAGGGTG
	STAT3 exon 21 F	TCCCATTCCCAGGGATAACTGAGGA
	STAT3 exon 21 R	TCCTGCCGAGGCAGATGGCT
	STAT3 exon 22 F	AGAGCATCACACAAAGGGGACCA
	STAT3 exon 22 R	TCCTGCCGAGGCAGATGGCT
	STAT3 exon 23 F	GCAGGTAGGCGCCTCAGTCG
	STAT3 exon 23 R	TGCAGAGGGTGGACAACTGAACT
STAT5A/B capillary sequencing (Study III)	STAT5A exon14 F	TCTGTCCCTGCATGCCCCCA
	STAT5A exon14 R	GGGGCCCCCATCTCTCCTGG
	STAT5A exon15 16 F	GCCCTGACTCGGGGGTTCCT
	STAT5A exon15 16 R	GTGGCGGGGCAAGGGAACAA
	STAT5A exon 18 19 F	ACATGGGGCGTGGGCTTCCA
	STAT5A exon 18 19 R	GCCAGCCCTCCAGGAGTCCA
	STAT5A exon 20 F	AGCAGGCTGGAGGCTGTCCC
	STAT5A exon 20 R	AGGACTGCACAGGGGAGGCA
	STAT5B exon 14 15 F	AGGCGGGGATATTTGTATGCCTCT
	STAT5B exon 14 15 R	TGTTCATGTGTAACCATGCTGCCAT
	STAT5B exon 17 18 F	AGGTGGTTGTGTTCTGTCCT
	STAT5B exon 17 18 R	CAGTTCCTCCCCTGTGGAC
	STAT5B exon 19 F	CTGGTGGCCTGTGGGGCTTG
	STAT5B exon 19 R	TCTGTCTGTGGCCCCTCTGCT
STAT3 amplicon sequencing (Study IV)*	STAT3 exon 21 amplicon F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT<u>CCCAAAAATTAAATGCCAGGA</u>
	STAT3 exon 21 amplicon R	AGACGTGTGCTCTTCCGATCT<u>GGTTCCATGATCTTTCCTTCC</u>
STAT5 amplicon sequencing (Study III)	STAT5A exon 17 amplicon F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT<u>TCCTGCTGCTGGTGGATTAT</u>
	STAT5A exon 17 amplicon R	AGACGTGTGCTCTTCCGATCT<u>AGCCCAAGGCTTTGTCTATG</u>
	STAT5B exon 16 amplicon F	ACACTCTTTCCCTACACGACGCTCTTCCGATCT<u>TTGTTGGGGTTTTAAGATTCC</u>
	STAT5B exon 16 amplicon R	AGACGTGTGCTCTTCCGATCT<u>CAAATCAGAATGCGAACATTG</u>

*Target-specific sequences are underlined in amplicon primers. All amplicon primers contain Illumina adapter sequence tails.

11.2. Whole-exome sequencing

Paired samples from each patient were prepared and used for exome sequencing in studies I and III: CD8+ cells were used as LGL leukemia samples and CD4+ cells as controls. A total amount of 3 μg of each genomic DNA sample was fragmented with Covaris S-2 focused-ultrasonicator (Covaris) and processed with NEBNExt DNA Sample Prep Master Mix Set (New England Biolabs). Illumina Paired-End (PE) PCR primers were used for amplification of the library (Illumina). Exome capture was processed using the Agilent SureSelect Target Enrichment System for Illumina Paired-End Sequencing Library and SureSelect Human All Exon protocols (Agilent). Captured libraries were amplified in post-capture PCR with NimbleGen PE-POST1 and PE-POST2 primers (Roche). Cluster generation and sequencing of exome libraries were performed with Illumina GAII (Study I) or Illumina HiSeq2500 (Study III) instruments as 82 bp PE reads (Illumina).

After sequencing, the data were run through bioinformatics analysis tools. Raw reads were trimmed based on the quality scores, which were converted to Sanger phred scores of FASTQ files. Reads were then aligned against the current human genome build (hg19) with the Burrows-Wheeler Aligner.[190] Potential PCR duplicates and reads with poor mapping quality were eliminated. SAMtools was used to create pileup files (files that describe the base-pair information at every chromosomal position) to enable subsequent SNV/indel calling.

Somatic mutations specific for leukemic sample were called with the VarScan 2.2.3 or newer somatic mutation caller.[192,194] Missense, nonsense, frameshift, inframe coding indels, and splice site mutations with a VarScan somatic p-value below 0.01 were considered to be candidate mutations and their functional consequences were analyzed using software such as PolyPhen-2. GERP conservation scores were calculated with Annovar to assess the level of conservation of the mutated site in evolution.[195] Known public variants were excluded using different databases (1000 Genomes Project Consortium, dbSNP132). Candidate somatic mutations were checked visually in the Integrative Genomics Viewer (IGV). Capillary and/or amplicon sequencing primers were designed for every candidate mutation to validate it from the source DNA with a different method.

Different tools were used for the prediction of the effects of a mutation. The COSMIC database (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) is a catalogue of somatic mutations in cancer, and it was used to assess whether the same mutation has been reported previously in cancer.[201] PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) predicts the functional effects of human non-synonymous single nucleotide variants leading to an amino acid substitution.[202] PolyPhen-2 predictions are based both on the data of the structural conservation of amino acid sequence (functionally important residues are highly conserved in evolution) and available structural information of the mutated protein.[202]

The basic terminology of second-generation sequencing is explained in Figure 6.
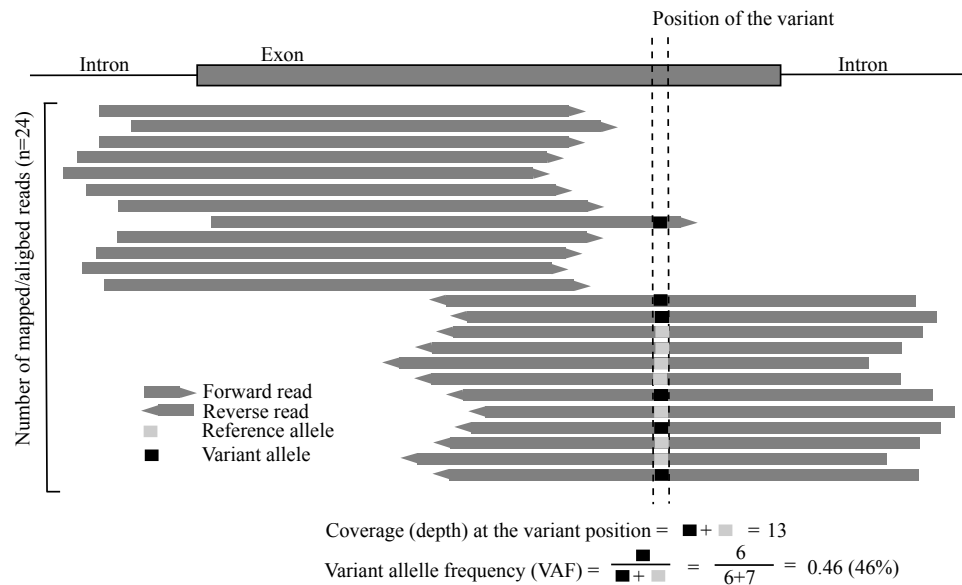
Coverage (depth) at the variant position = ■ + ▨ = 13

Variant allelle frequency (VAF) = $\dfrac{■}{■ + ▨}$ = $\dfrac{6}{6+7}$ = 0.46 (46%)

Figure 6. *Basic terminology of second-generation sequencing.*

11.3. RNA sequencing

RNA sequencing was used for gene expression analysis in study I: CD8+ RNA was extracted from a T-LGL leukemia patient and a healthy normal donor (anonymous Finnish Red Cross buffy coat) CD8+ RNA was used as a control. Ribosomal RNA was depleted with Ribo-Zero rRNA Removal Kit (Epicentre) and the sample was then reverse-transcribed to cDNA with SuperScript reverse transcriptase (SuperScript). The RNA sequencing library was prepared with Nextera technology (Epicentre) and amplified with primers containing Illumina adapters. The cluster generation and library sequencing (72 bp paired-end reads) were done with Illumina GAII.

RNA-sequencing reads from the index patient and normal control were aligned to the human reference genome hg19 using TopHat, which analyzes the mapping results to identify splice junctions between exons, and the number of mapped reads were counted with HTseq Python package. DEseq R/Bioconductor package was used for the analysis of differentially expressed genes, and the resulting gene expression data were analyzed with IPA software (Ingenuity). The expression levels of exome sequencing somatic variants were checked with IGV.

11.4. Targeted deep *STAT3* and *STAT5B* sequencing

Amplicon sequencing of *STAT3* (exon 21), *STAT5A* (exon 17) and *STAT5B* (exon 16) was used in studies III (STAT3) and IV (STAT5A, STAT5B). Locus-specific primers are presented in Table 4. The PCR reaction contained locus-specific PCR primers (with Illumina adapter sequence tails), Illumina TruSeq Universal Adapter primer, and Illumina TruSeq Adapter primer with a 6 bp index sequence (used for genetic barcoding of each sample). This initial amplification creates Illumina-compatible paired-end sequencing templates. Indexes are used to recognize sample-specific reads after the pooled sequencing run. Sequencing of PCR amplicons was performed using the Illumina MiSeq instrument with MiSeq Control Software (Illumina). Samples were sequenced as 151 bp or 251 paired-end reads using MiSeq amplicon workflow (Illumina).

Data were analyzed with Illumina MiSeq Reporter Software (Illumina) in study III. A specifically developed bioinformatics pipeline was used in study IV: first, after mapping the reads to *STAT3* gene, previously reported *STAT3* mutations were verified and all variants with a variant allele count over 5 and variant allele frequency (VAF) over 0.5% were taken into consideration. Second, a broader manner of approach was used, as all variants with variant allele count over 5 and VAF over 0.5% were called, and from these variants false positives were initially filtered out based on the estimated error rate from control sample in every run (a p-value vas calculated using binomial distribution). However, variants with VAF over 2% were called independent of the noise. In both scripts, a specific frequency ratio was calculated to filter out false positive variants, and all samples with a frequency ratio ≥0.80 were considered to be true mutations (Figure 7). The variants from both scripts with a borderline frequency ratio between 0.75-0.79 were verified manually with IGV.

$$\frac{\text{the ratio of variant calls/number of all the bases (at a position)}}{\text{the ratio of variant allele quality sum/quality sum of all the bases (at a position)}} \geq 0.80$$

Figure 7. Frequency ratio formula of amplicon sequencing analysis pipeline.

11.5. Targeted deep TCRB sequencing

In study IV, TCRB CDR3 regions were amplified and sequenced by Adaptive Biotechnologies Corp (Seattle, WA, US) using the ImmunoSEQ assay. The assay used analyzes the TCRB CDR3 repertoire and can detect a rearrangement of one cell in 40 000 T cells. Flow cytometry-sorted lymphocytes from three T-LGL leukemia patients were sequenced with 52 forward primers for the Vβ gene segment and 13 reverse primers for the Jβ segment in a multiplex-PCR reaction, to analyze the TCRB CDR3 VDJ rearrangement sequence. The resulting 60 bp amplicons were sequenced with Illumina HiSeq platform (Illumina). Raw sequences were aligned based on the V, D, and J gene definitions of IGMT database. The data were analyzed using ImmunoSEQ analyzer provided by Adaptive Biotechnologies. The Simpson's diversity index was calculated for every sample to assess the level of narrowing of T cell receptor repertoire (Figure 8).

$$D = 1 - \frac{\sum n(n-1)}{N(N-1)}$$

Figure 8. The calculation of Simpson's diversity index, where "n" represents the observed counts of a particular TCRB rearrangement seen in the sample, whereas "N" is the total number of all TCRB rearrangements (amplicon reads) in the sample.

11.6. Other sensitive PCR methods

In addition to amplicon sequencing, other sensitive sequencing methods were also used in the detection of *STAT3* mutations. In study II, the presence of the two most common *STAT3* mutations, D661Y and Y640F, was determined by amplification refractory mutation system (ARMS) assay.[203] In study IV, real-time allele-specific oligonucleotide quantitative PCR (ASO-qPCR) of *STAT3* mutations Y640F and D661V was performed to compare to amplicon sequencing of *STAT3:* A Y640F or D661V mutation-positive reference sample was used for preparing a log-linear

dilution series, and a standard curve was made based on the ASO-qPCR results. The data were normalized with qPCR of albumin reference gene, and the mutation load of the samples examined was quantified in relation to the standard curve.

12. Functional assays

12.1. STAT3 and STAT5B mutagenesis

STAT3 and STAT5B constructs were created for studies I and III, respectively.

In study I, the starting point was an open-reading frame (ORFeome) clone of the human *STAT3* gene coding sequence inserted in the pENTR201 vector (NM_003150, acquired from the Genome Biology Unit at the University of Helsinki). Vectors containing one common *STAT3* mutation—Y640F or D661V—were created by site-directed mutagenesis using 5'-phosphorylated primers presented in Table 5. After confirmation of the successful mutagenesis by sequencing, both wild type (wt) and mutated *STAT3* coding sequences were transferred to a V5-tagged pDEST-40 mammalian expression vector using Gateway® LR Clonase II enzyme mix (Life Technologies).

In study III, *STAT5B* mutations Y665F and N642H were introduced to an expression plasmid pCMV6-XL6 containing the wild type coding sequence of *STAT5B* (OriGene, SC115355) with the GENEART® Site-Directed Mutagenesis System (Invitrogen). The mutations were confirmed by capillary sequencing of the plasmid DNA. Mutagenesis primers are listed in Table 5.

Table 5. *STAT3 and STAT5B mutagenesis primers. The site of primer mismatch is underlined.*

Primer use	Primer target	Sequence (5'-3')
STAT3 mutagenesis (Study I)	STAT3 Y640F F	CTGAACAACATGTCATTTGCTG
	STAT3 Y640F R	CTGCTGCTTTGTGAATGGTTCCACGG
	STAT3 D661V F	CTGGTGTCTCCACTGGTCTATC
	STAT3 D661V R	GATATTGGTAGCAACCATGATCTT
STAT5B mutagenesis (Study III)	STAT5B Y665F F	GAGACTTGAATTTCCTTATCTACGTGTTTC
	STAT5B Y665F R	GAAACACGTAGATAAGGAAATTCAAGTCTC
	STAT5B N642H F	GAAAGAATGTTTTGGCATCTGATGCCTTTTAC
	STAT5B N642H R	AAAGGCATCAGATGCCAAAACATTCTTTC

12.2. Western blotting

Western blotting experiments were mainly conducted on whole cell-lysates or from subcellular-fractionated samples (cytosolic and nuclear lysates) from LGL leukemia patients: 20 million PB MNCs were first suspended in hypotonic buffer (10 mM hydroxyethyl piperazineethanesulfonic acid [HEPES, pH 7.9], 10 mM KCl, 1.5 mM MgCl2, 0.5 mM dithiothreitol [DTT]), and after swelling, the cells were sheared in a pre-cooled homogenizer (Douncer). After centrifuging, supernatants containing

the cytosolic fraction were collected, and pellets were centrifuged over a sucrose cushion to extract the nuclear fraction, which was subsequently lysed.

Protein concentrations of whole cell, nuclear, or cytosolic lysates were measured, and normalized aliquots of the different fractions were ran on an SDS-PAGE gel (8-15%) and subsequently transferred to a PVDF membrane (Bio-Rad). The membranes were blocked with 5% BSA or 5% fat-free milk and then incubated with primary antibodies. In study I, the following primary antibodies were used: rabbit anti-Erk1 (Santa Cruz sc-94) 1:1000, rabbit anti-phospho^{Y705} STAT3 (Cell signaling, #9145) 1:2000, or mouse anti-STAT3 (Cell Signaling, #9139) 1:1000. Primary antibodies of study III were rabbit anti-phospho^{Y694} STAT5B (Cell Signaling, #9351) 1:500, rabbit anti-STAT5 (Cell Signaling, #9358), rabbit anti-phospho^{Y705} STAT3 (Cell Signaling 9145) 1:2000, mouse anti-STAT3 (Cell Signaling, #9139) 1:1000, or mouse anti-GAPDH (Novus Biologicals 300-285) 1:5000. After ligation of appropriate secondary antibodies (containing infrared-readable dyes), protein bands were visualized with the LI-COR Odyssey (LI-COR).

12.3. Enzyme-linked immunosorbent assay (ELISA)

The total and phosphorylated STAT5 protein was analyzed by ELISA in study III using PhosphoTracer STAT5A/B pTyr694/699+Total ELISA Kit (Abcam). Mononuclear cells were lysed with kit reagents. Prior to analysis, protein concentrations were measured with the Qubit 2.0 fluorometer (Life Technologies) to ensure that the concentrations were similar between samples. The analysis was done in duplicate. PHERAstar FS reader (BMG Labtech) was used to measure fluorescence intensity.

12.4. Immunohistochemistry

Immunohistochemistry in study I was performed using a Leica BOND-MAX autostainer (Leica Microsystems, Wetzlar, Germany). Slides were prepared from FFPE tumor and healthy control bone marrow biopsy blocks. After baking, the slides were handled with a Bond Polymer Refine Detection kit (Leica Microsystems) including citrate buffer for antigen retrieval, and stained with monoclonal anti-phospo-STAT3^{Y705} antibody (9145, Cell Signaling Technology, Danvers, MA, US) 1:100 or anti-CD57 antibody (TB01, Dako, Glostrup, Germany) 1:100. The slides were visualized with the Zeiss Axio Imager AX10 microscope (Zeiss) and photographed with Nuance FX multispectral tissue imaging system (PerkinElmer).

12.5. STAT3 and STAT5B reporter assay

In study I, HEK293 cells stably expressing the GloResponse SIE-Luc2P luciferase reporter (Promega) were transfected with empty pDEST-40 or STAT3-containing pDEST-40 vectors using lipofectamine-assisted transfection. On the next day, the cells were mock-treated or stimulated with IL-6, and Dual-Glo luciferase detection reagent (Promega) was used to analyze firefly luciferase activity.

In study III, HeLa cells were transiently transfected using lipofectamine with pCMV6-XL6 vector containing STAT5B and another vector with Luc2P-luciferase reporter sequence under the control of STAT5 response element (pGL4.52 vector with Luc2P/STAT5RE/Hygro, Promega). On the following day, unstimulated firefly

luciferase activity was measured with One-Glo luciferase detection reagent (Promega).

13. Gene expression analysis

13.1. Illumina gene expression array

In study I, CD8+ cells from T-LGL leukemia patients with/without *STAT3* mutations were compared with healthy control CD8+ cells using Illumina Human HT-12 v4 Expression BeadChip array (Illumina). After hybridization, microarrays were read with a BeadArray Reader (Illumina) and analyzed with GeneSpring GX software (Agilent Technologies): samples were first transformed logarithmically by scaling to median baseline intensity of each chip, and then normalized.

In study III, the comparison was done between 3 *STAT5B*-mutated patients with LGL leukemia, 3 *STAT3*-mutated patients, 2 *STAT3*- and *STAT5A/B*-unmutated patients (unmutated by exome sequencing), and 4 healthy controls (2 CD8+ and 2 NK cell samples). An Illumina Human HT-12 v4 Expression BeadChip array (lllumina) was used for gene expression analysis. The chips were read with an iScan instrument (Illumina), followed by primary analysis with Genome Studio software (Illumina). Normalization and analysis of data were done with Chipster software (version 2.10): data were normalized and log2-transformed. Expression profiles were compared with pairwise Pearson correlation. Differentially expressed genes were detected with empirical Bayes test. Pathway analysis was performed with KEGG and hypergeometric test. The microarray data were uploaded into the open-access ArrayExpress database under accession number E-MTAB-1611.

14. Statistical methods

In study I, PASW Statistics software (version 18.0) was used to calculate clinical correlations of *STAT3* mutations, using Fisher's exact and unpaired 2-sided t-tests (p<0.05 considered statistically significant). In study II, with a larger patient cohort, overall survival and progression-free survival related to mutation status were also analyzed with the Kaplan-Meier and Cox methods. In study IV, GraphPad Prism software (version 5.0) was used to analyze clinical data: a chi-square test was used when categorical variants were compared, whereas either a 1 way ANOVA with Bonferroni post hoc test or an unpaired t test was used for parametric variables.

RESULTS

15. Exome sequencing of T-LGL leukemia patients (I, III)

The analysis of the molecular background of LGL leukemia started with exome sequencing of LGL leukemia patients. The aim was to discover novel somatic genetic variants related to the pathogenesis of the disease. The patients chosen fulfilled the diagnostic criteria of LGL leukemia by WHO 2008[9], were untreated, and had a large monoclonal expansion in CD8+ T cell fraction.

15.1. Exome sequencing results of the index T-LGL leukemia patient (I)

The index patient was a male and 71 years old at the time of the exome sequencing. Clinical characteristics of the patient are presented in Table 6. The diagnosis of LGL leukemia was based on persistent PB lymphocytosis, clonal TCR rearrangement, and chronic neutropenia; however, no severe infections were seen during follow-up. The Vbeta analysis of CD8+ fraction revealed one Vb16+ immunodominant expansion at all timepoints (Figure 9). Vbeta results of the CD4+ fraction were polyclonal (data not shown). The size of the Vb16+ clone was 94% of CD8+ cells at the time of the exome sequencing (Figure 9). During follow-up, the size of the clone diminished and at the same time, a decrease in lymphocyte counts and improvement of neutropenia was observed (Figure 9).

Table 6. *Clinical characteristics of the three exome sequenced patients. All laboratory values are presented as medians and range.*

Patient	Age at dg	Gender	Follow-up time	Hb, g/l[1]	Leuk, x10^9/L	Lymph, x10^9/L	Neut, x10^9/L	Plts, x10^9/L	Phenotype	Associated disorders
1 (Study 1)	70	Male	60 months	123 (87-160)	9.0 (4.8-23.2)	3.7 (1.7-14.0)	1.6 (0.6-7.6)	311 (164-594)	CD8+ TCRα/β+	Neutropenia
2 (Study III)	71	Female	40 months	126 (102-133)	11.8 (8.5-18.2)	8.3 (5.8-11.7)	4.2 (3.0-5.5)	249 (156-287)	CD8+ CD56+ TCRα/β+	RA-like symptoms
3 (Study III)	49	Female	36 months	126 (120-132)	16.9 (13.2-19.7)	13.4 (10.2-15.4)	2.5 (1.15-3.44)	279 (260-301)	CD8+ CD56+ TCRα/β+	None

[1]All laboratory values are presented as medians and range during follow-up time.

Abbreviations: dg, diagnosis; Hb, hemoglobin; Leuk, leukocytes; Lymph, lymphocytes; Plts, platelets; RA, rheumatoid arthritis
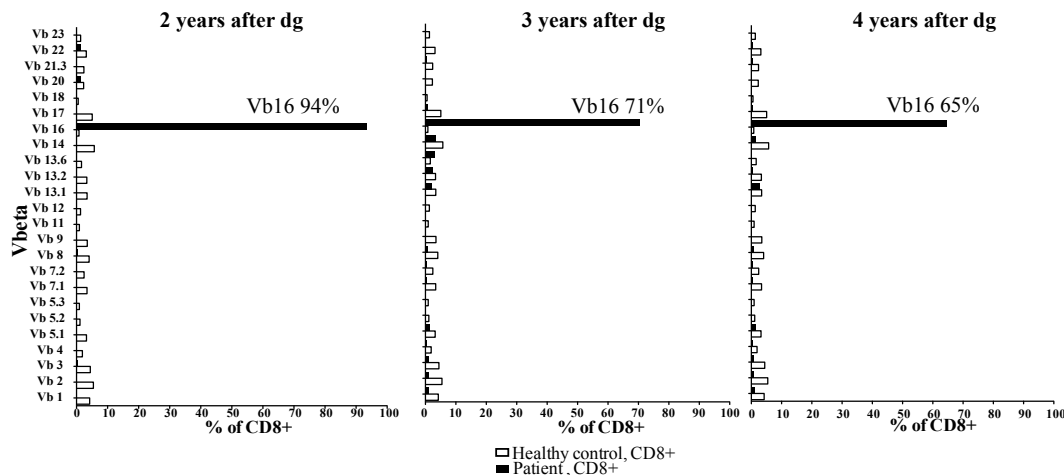

Figure 9. *TCR Vbeta results (CD8+ lymphocytes) of the patient 1 in study I (Table 6).*

Exome sequencing of patient 1 was done from the magnetic bead-sorted CD8+ fraction and CD4+ cells were used as a control. The exome sequencing produced 60.2 million (CD8+) and 61.1 million (CD4+) reads, which were aligned to the reference genome. The early version of the variant calling pipeline did not calculate the median coverage, but 59% and 62% of the exons in CD8+ and CD4+ samples, respectively, were sequenced with depth more than x20.

Exome sequencing pipeline and somatic mutation calling are described in the methods section (Chapter 11). Briefly, Varscan was used to call somatic mutations specific for CD8+ fraction. Candidate mutations (somatic $p<0.01$) were taken into consideration, and their functional effects were inspected using PolyPhen-2 and other annotation software. Known population variants were excluded using the data from the 1000 Genomes Project Consortium and dbSNP132. Plausible variants with connection to the immune system or cancer were taken into further consideration and were checked visually using the IGV. The expression of the variants was verified from RNA sequencing data. Capillary and/or amplicon sequencing primers were designed for every candidate mutation to validate it from the source DNA with a different method.

After comparison of CD8+ and CD4+ samples, seven variants were chosen for capillary sequencing validation, and 4/7 of these variants were confirmed specific for CD8+ sample (Table 7). The mutated genes included calpain 13 (*CAPN13*), homeobox C9 (*HOXC9*), macrophage scavenger receptor 1 (*MSR1*), and signal transducer and activator of transcription 3 (*STAT3*). Of these, *STAT3* mutation D661V was the only one expressed in the RNA sequencing data.

***Table 7.** Exome sequencing somatic variants of patient 1 (Table 6), CD8+ fraction*

Gene	Chr	Posit	Mutations type	Protein	Ref. base	Var. base	CD8 ref (n)	CD8 var (n)	CD4 ref (n)	CD4 var (n)	Somatic p [1]	RNA-seq. [2]
CAPN13	2	30966280	Nonsense	R472*	G	A	92	65	168	2	2.94×10^{-22}	NE
HOXC9	12	54396263	Frameshift	NA	T	delT	25	9	38	0	0.00062	NE
MSR1	8	50692579	Missense	R26C	G	A	11	8	28	0	0.00024	0/1
STAT3	17	118219358	Missense	D661V	T	A	39	28	63	0	2.66×10^{-10}	119/295

Abbreviations: chr, chromosome; del, deletion; NA, not accessible; NE, not expressed; Ref, reference; Var, variant.
[1] Somatic p-value for somatic/loss-of-heterozygosity events: significance of tumor read count vs. normal read count
[2] Number of variant alleles/total sequencing depth in RNA sequencing data.

CAPN13 is a member of the calpain protein family. Calpains are intracellular Ca^{2+}-dependent proteases that cleave their substrates at specific inter-domain sites, affecting several signaling cascades connected to cell growth and survival.[204] The mutation of patient 1 in CAPN13 introduced a stop codon at amino acid R472. The mutation leads to translation of a truncated protein lacking the EF-hand domain, thus possibly impairing the calcium-binding ability of the mutated CAPN13. However, the exon with the mutation was not expressed in RNA sequencing.

HOXC9 and other homeobox genes regulate transcription, are involved in development, and their aberrant expression is associated with cancer.[205] The frameshift mutation of patient 1 occurred at the beginning of the DNA-binding homeodomain of *HOXC9* (after amino acid K195), and could impair the transcriptional activity of HOXC9. The mutated spot was not expressed in the RNA sequencing data.

MSR1, known also as SR-A, is important in the innate immune system as it is expressed on the surface of macrophages, where it recognizes pathogens and modified self-proteins such as lipids.[206] The COSMIC database reports the same missense mutation, R26C, in a carcinoma of the large intestine, and sporadic mutations of adjacent amino acids have been reported in breast and endometrium carcinomas. PolyPhen-2 predicted the R26C mutation to be probably damaging, and amino acid R26 is highly conserved in evolution. The mutation was not expressed in the RNA sequencing.

The function of the normal STAT3 protein is reviewed in the first section of this thesis (Chapter 5). The D661V mutation was in the SH2 domain, which is important for the dimerization of STAT3. Based on the PolyPhen-2 report, the mutation was probably damaging and situated in a highly conserved area. Amino acid D661 is in a SH2 domain surface loop, which comes to close proximity of the other STAT3 monomer during dimerization.[112]

15.2. *STAT3* sequencing of the index T-LGL leukemia patient (I)

STAT3 D661V mutation of the patient 1 (Table 6) was the most interesting of four somatic mutations detected (Table 7). To analyze the location of the mutation, different lymphocyte fractions were sorted with flow cytometry from a PB MNC sample. The gating strategy is shown in Figure 10. The mutation was confirmed specific for CD8+ fraction by capillary sequencing (Figure 10). Vb16+ clone consisted of 94% of the CD8+ fraction, and both exome sequencing (Table 7) and capillary sequencing suggested that mutation was heterozygous in the leukemic Vb16+ population.
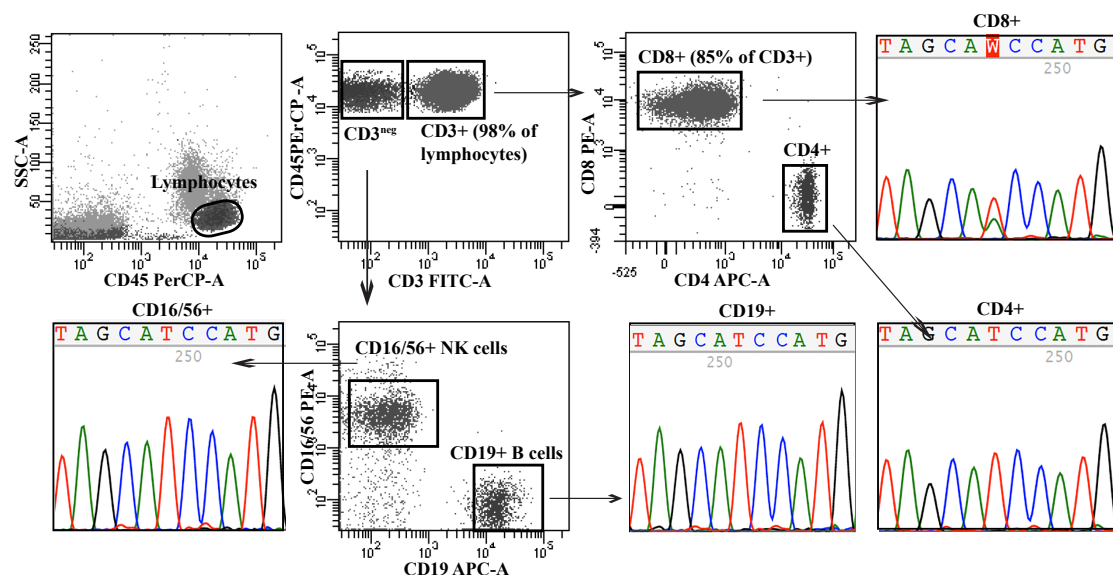

Figure 10. *Lymphocyte fractions and STAT3 capillary sequencing results of patient 1 in study I (Table 6). CD4+/CD8+ and CD16/56+/CD19+ cells were sorted from different tubes. 85% of CD3+ lymphocytes were CD8+, and the STAT3 mutation D661V was seen only in CD8+ fraction.*

15.3. Exome sequencing results of *STAT3* mutation-negative T-LGL leukemia patients (III)

Two *STAT3*-mutation negative patients were chosen for exome sequencing in study III. The clinical characteristics of patients 2 and 3 are presented in Table 6. The Vbeta analysis of patient 2 (Table 6) was done once, and 91% of the CD8+ cells were major Vb22+ expansion (Figure 11A). Patient 3 (Table 6) had also large clone: Vb17+ cells consisted of 94% of the CD8+ fraction at two separate timepoints (Figure 11B-C).
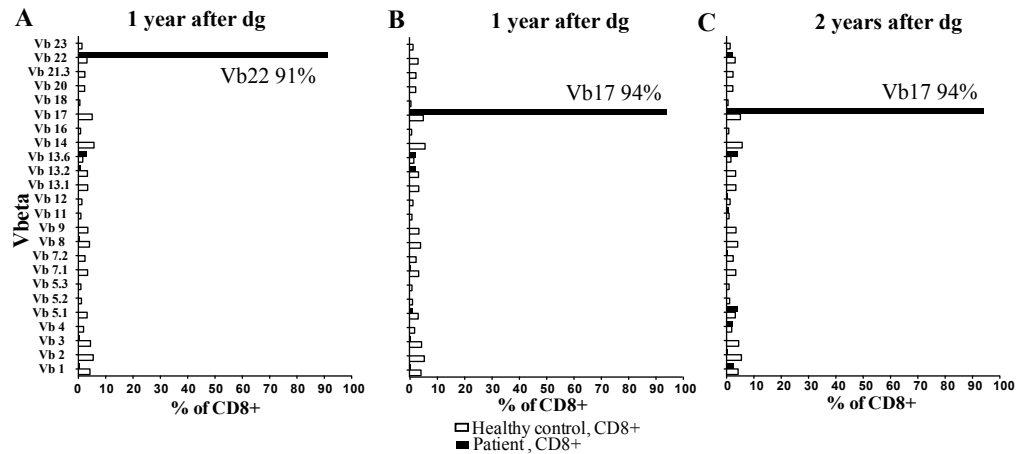

Figure 11. TCR Vbeta results of patients 2 (A) and 3 (B-C) (Table 6) in study III. (A) CD8+ Vbeta results of patient 2 one year after the diagnosis. CD8+ Vbeta expression of patient 3 was analyzed one (B) and two (C) years after the diagnosis.

Similar to patient 1 (Table 6), a CD8+ fraction was used as a tumor and CD4+ cells as a control in exome sequencing. In the case of patient 2 (Table 6), sequencing yielded 62 million and 64 million aligned reads for CD8+ and CD4+ samples, respectively. The median coverage was not analyzed, but 54% (CD8+) and 55% (CD4+) of the target exons were sequenced with more than x20 coverage. In the case of patient 3, the number of aligned reads was 68 million (CD8+) and 53 million (CD4+), and more than x20 coverage was achieved in 55% and 52% of target bases, respectively.

Somatic variants with $p<0.01$ were validated with capillary sequencing. Of the four variants of patient 2 (Table 6), two were confirmed to be specific for CD8+ fraction by capillary sequencing, and the mutations were located in bone marrow morphogenetic receptor 2 (*BMPR2*) and *STAT5B* (Table 8). The capillary sequencing of patient 3 (Table 6) validated 3 of 4 candidate mutations in following genes: early B cell factor (*EBF3*), sphingomyelin synthase 2 (*SGMS2*), and STAT5B (Table 9). Thus, both patients carried the same Y665F mutation in the *STAT5B* gene, and in both cases this mutation was expressed in the RNA sequencing data (Tables 8 and 9).

Table 8. *Exome sequencing somatic variants of patient 2 (Table 6), CD8+ sample.*

Gene	Chr	Posit	Mutation type	Protein	Ref. base	Var. base	CD8 ref (n)	CD8 var (n)	CD4 ref (n)	CD4 var (n)	Somatic p [1]	RNA-seq. [2]
BMPR2	2	203379664	Missense	E195K	G	A	39	42	113	1	7.10×10^{-19}	13/25
STAT5B	17	40359659	Missense	Y665F	T	A	43	30	99	0	9.34×10^{-14}	73/171

Abbreviations: chr, chromosome; del, deletion; NA, not accessible; NE, not expressed Ref, reference; Var, variant
[1]*Somatic p-value for somatic/loss-of-heterozygosity events: significance of tumor read count vs. normal read count*
[2]*Number of variant alleles/total sequencing depth in RNA sequencing data.*

Table 9. *Exome sequencing somatic variants of patient 3 (Table 6), CD8+ sample.*

Gene	Chr	Posit	Mutation type	Protein	Ref. base	Var. base	CD8 ref (n)	CD8 var (n)	CD4 ref (n)	CD4 var (n)	Somatic p [1]	RNA-seq. [2]
EBF3	10	131761204	Splice-5	NA	A	G	73	41	93	0	4.42×10^{-13}	NE
SGMS2	4	108820744	Missense	R157C	C	T	123	89	212	0	1.20×10^{-32}	0/1
STAT5B	17	40359659	Missense	Y665F	T	A	38	41	63	0	5.98×10^{-14}	149/343

Abbreviations: see Table 8.

Patient 2 (Table 6) had a CD8-specific mutation in the *BMPR2* gene, which belongs to the receptor family of serine/threonine kinases. *BMPR2* loss-of-function mutations have been detected in pulmonary hypertension patients.[207] The mutated amino acid E195 was conserved in the *BMPR2* gene of vertebrates and was estimated to be probably damaging by PolyPhen-2. The location of the mutation was near the kinase domain and the ATP-binding site of BMPR2. Two sporadic cases of endometrial and large intestine cancer were reported by the COSMIC database to carry the same mutation. The mutation was expressed in the RNA sequencing.

STAT5B, which was mutated in both patients 2 and 3, is reviewed in Chapter 5 of this thesis. Similar to the *STAT3* D661V mutation of patient 1, *STAT5B* Y665F was also located in the SH2 domain. PolyPhen-2 reported the mutation as benign, and the area was conserved in mammals. COSMIC did not report mutations in Y665, although missense mutations in nearby residues have been detected in kidney and lung carcinoma. The mutation was expressed in the RNA sequencing.

The remaining mutations of patient 3 (Table 6) included a splice-5 site mutation in *EBF3*. EBF3 is a member of EBF family of transcription factors, which mediate B cell, bone, and neuronal differentiation and development, and act as tumor-suppressors: silencing mutations of *EBF3* are frequently observed in glioblastoma.[207] A splice-site mutation could potentially be a loss-of-function variant, as it affects the splicing of the pre-messenger RNA and subsequent translation of the functional EBF3 protein.

SGMS2 catalyzes the production of sphingomyelin, which is a major component of cell membrane. SGMS2 activity affects the cell surface composition and formation of lipid rafts, which harbor cell surface receptors.[208] Considering the connection to cancer, sphingomyelins were downregulated in the glioma cell line and activation of SGMS caused an increase in sphingomyelin levels with simultaneous apoptosis of glioma cells.[209] Amino acid R157 is located in a loop between two transmembrane-regions, and no mutations are reported in this residue. PolyPhen-2 predicted the mutation to be probably damaging and located in a conserved area.

16. Results of *STAT3* and *STAT5B* mutation screening in LGL leukemia patients (I-IV)

Amplicon sequencing and/or capillary sequencing were used in the screening of *STAT3* and *STAT5B* mutations in exons 21 and 16, respectively, and other SH2 domain exons of *STAT3* and *STAT5B* genes were sequenced with capillary sequencing. In the following section, the screening results of recurrent *STAT3* and *STAT5B* mutations are presented.

The primers were also designed for the capillary sequencing validation of the other exome sequencing variants than STAT3/STAT5B mutations (Tables 7-9) of patients 1-3. These mutations proved to be non-recurrent in the patient cohorts.

16.1. *STAT3* screening by capillary sequencing in LGL leukemia (I, II, IV)

The most promising CD8-specific variant of patient 1 (Table 7), STAT3 D661V, was screened first from the cohort of 8 Finnish T-LGL leukemia patients with monoclonal disease (patients included in study I). Among these patients, 3 of 8 harbored a *STAT3* mutation in exon 21 of *STAT3* gene. Recurrent *STAT3* mutations led to the initiation of study I, in which the exons 18-23 of SH2 domain of *STAT3* were sequenced by capillary sequencing from a cohort of 77 T-LGL leukemia patients, including patient 1 (Table 6). The samples were either whole blood or PB MNC DNA. To ensure that the sensitivity of capillary sequencing would be sufficient to detect mutations, the whole blood patient samples chosen for screening had a monoclonal expansion in Vbeta analysis (data available from 41/77 patients, median Vbeta size 83% of CD8+ cells, range 32-99%). The remaining samples were PB MNC DNA, and the assumption was that the sensitivity of capillary sequencing would be sufficient to detect mutations.

The frequency of *STAT3* mutations in the study I cohort was 40% (31 of 77 patients sequenced) (Figure 12). The mutations were focused in two amino acids coded by exon 21, Y640 and D661, which accounted for 90% (28/31) of all mutations. None of the mutations was either reported by two datasets describing genetic variants in population (1000 genomes and dbSNP database) or detected in CML patients with reactive LGL lymphocytosis (n=6), AML patients (n=9), ALL patients (n=6), or in two T-ALL cells lines (Jurkat, MOLT-4).
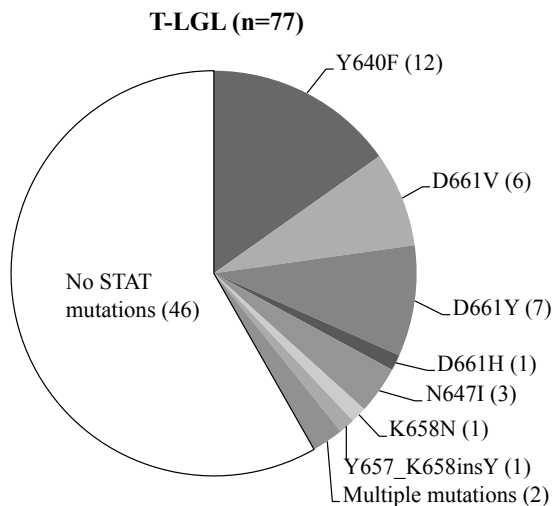

Figure 12. STAT3 mutations in the T-LGL leukemia cohort of study I. Numbers after the mutations indicate the amount of patients in each group. Two patients had two STAT3 mutations (Y640F+D661Y, D661V+N647I).

Study I included only T-LGL leukemia patients, and the aim of study II was to analyze the frequency of *STAT3* mutations in a larger cohort containing both patients with T-LGL leukemia (n=120) and CLPD-NK (n=50). The frequency of *STAT3* mutations, determined by a combination of capillary sequencing (exons 18-23) and ARMS-PCR assay (Y640F and D661Y variants), was 28% (33/120) and 30% (15/50) in T-LGL and CLPD-NK, respectively (Figure 13). One T-LGL leukemia patient harbored two different *STAT3* mutations (Y640F+D661Y) (Figure 13). Similar to study I, mutations in Y640 and D661 were the most common and consisted of 80% of the variants detected: in ARMS-PCR assay designed for these mutants additional 7 *STAT3*-mutated cases were identified. The *STAT3* mutations in study I were seen in exon 21, and two additional variants, S614R and G618R, which were coded by adjacent exon 20, were detected in study II. The SH2 domains of STAT family members (STAT1, STAT2, STAT4, STAT5A, STAT5B, STAT6) were screened in addition to other JAK-STAT pathway hotspots, which are frequently mutated in cancer (IL6ST exon 6, JAK2 V617F, JAK3 exons 13-17), using samples from 40 *STAT3* mutation-negative LGL leukemia patients, but no mutations were seen in these genes. No *STAT3* mutations were detected in a cohort of 31 patients with idiopathic neutropenia failing to meet complete diagnostic criteria of LGL leukemia.

The results of study II showed that *STAT3* mutations occurred in T-LGL leukemia and CLPD-NK with equal frequency. Study II included also two patients with a rarer TCR-γδ LGL leukemia, one of which had a *STAT3* mutation.
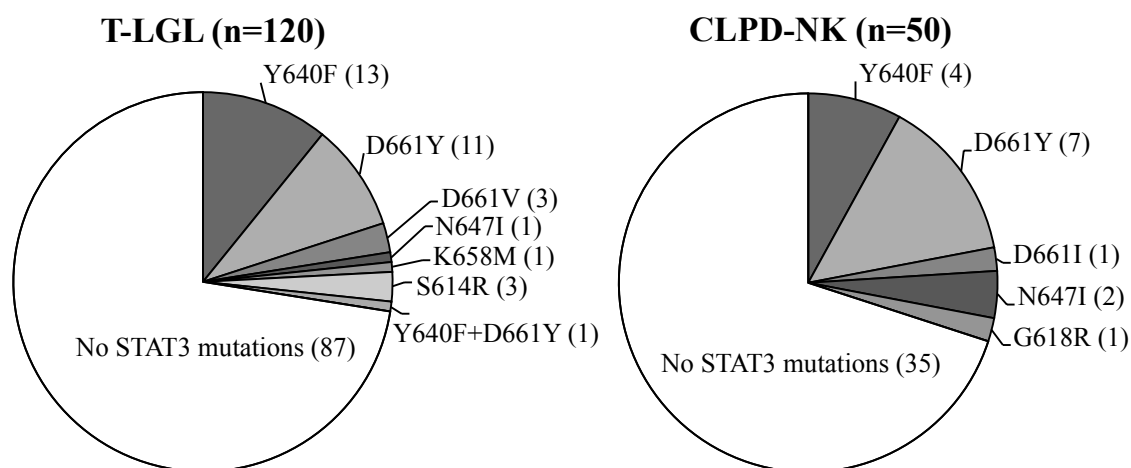

Figure 13. *STAT3 mutations in T-LGL leukemia and CLPD-NK patients in study II. Numbers after the mutations indicate the number of patients in each group.*

16.2. *STAT3* screening by amplicon sequencing in LGL leukemia (IV)

LGL leukemia is a heterogenous disease. The size of the mutated clone in T-LGL leukemia can vary and some patients have multiple CD8+ clones in Vbeta analysis.[72] In CLPD-NK, the analysis of the existence and size of LGL clones is challenging, as good clonal markers do not exist. Deep amplicon sequencing is a novel method with high sensitivity and ability to quantitatively analyze the sizes of the mutated clones. It detects all variants in the target area, whereas ARMS-PCR and ASO-qPCR are restricted to the use of mutation-specific primers.

The study IV patient cohort was the largest among the studies of this thesis, and it included 174 T-LGL leukemia and 39 CLPD-NK patients, who all met the criteria for

LGL leukemia diagnosis. However, the T-LGL leukemia cohort included also cases with smaller CD8+ clones, which were excluded from the monoclonal T-LGL population of study I. Amplicon sequencing samples were either PB MNC DNA or whole blood DNA.

The frequency of *STAT3* mutations in exon 21 by amplicon sequencing was 43% (75/174) and 18% (7/39) in T-LGL leukemia and CLPD-NK patients, respectively. In total, 82 of 213 (38%) patients carried a *STAT3* mutation by amplicon sequencing (Figure 14), whereas capillary sequencing was able to recognize mutations only in 23% (49/213) of the cases. Notably, among the patients of study I with large immunodominant clones, the frequency of mutations was 63% (45/72) by amplicon sequencing.

STAT3 screening revealed that a surprisingly high number, 22% (18/82), of mutated patients had multiple mutations in the *STAT3* exon 21 (Figure 14, Table 10). The mutations were either in different clones (n=14) or in the same clone (n=4), in which case 2 of 4 patients harbored two single SNVs in one codon resulting in one amino acid change (Table 10). In agreement with previous publications describing T-LGL leukemia patients with smaller expansions in addition to the large immunodominant clone[72], 8 of 14 patients with multiple mutated clones had one dominant SNV with bigger variant allele frequency (VAF), accompanied by one or multiple low-frequency SNVs (Table 10).
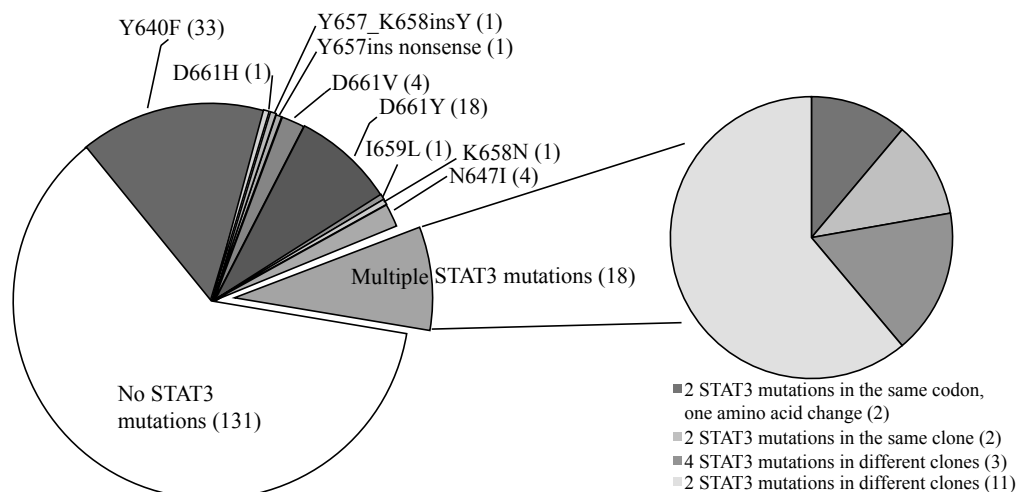

Figure 14. *STAT3 mutations detected by amplicon sequencing in the study IV cohort. Different patterns of multiple mutations are presented in the small pie. Numbers in parenthesis indicate the amount of patients in each group.*

Table 10. *Different examples of patients with multiple STAT3 mutations in the same or different lymphocyte clones (study IV).*

Patient	VAF1	Variant 1	VAF2	Variant 2	VAF3	Variant 3	VAF4	Variant 4
Example A	18%	2 SNVs make D661Y						
Example B	36%, 36%	N647I, D661V						
Example C	13%	D661Y	3%	Y640F	3%	N647I	1%	I659L
Example D	3%	Y640F	2%	D661Y	1%	I659L	1%	D661V
Example E	39%	Y640F	2%	I659L				
Example F	2%	K658R	2%	Y640F				

Abbreviations: SNV, single nucleotide variant; VAF, variant allele frequency.

16.3. Amplicon sequencing sensitivity and reliability (IV)

In a comparison of amplicon sequencing and ASO-qPCR data of two common *STAT3* mutants, Y640F and D661V, the results were congruent in 8 of 9 cases: one patient was weakly positive for Y640F mutation by ASO-qPCR, but even sorted lymphocyte subpopulations appeared unmutated by amplicon sequencing, indicating that either the mutated clone was under the detection limit of amplicon sequencing, or the Y640F mutation in ASO-qPCR was false positive.

A dilution series of a sample with *STAT3* D661V (VAF 37%) and Y640F (VAF 1.4%) was sequenced with an amplicon platform (Table 11). The analysis pipeline does not filter low quality reads prior to alignment, but uses quality values (QV, i.e. Phred scores) to exlude low quality calls. Low average QVs (under 20) are considered as errors. The D661V mutation was under the detection limit with the dilution of 1:100, at which point the number of error calls with low QV probably interfered with the detection of the variant. The Y640F mutation with lower VAF was detected with the dilution of 1:2. Based on the results, the variants with VAF over 0.5% and frequency ratio over 0.80 (borderline 0.75-0.80 checked manually) were considered true positives in amplicon analyses (Table 11).

The screening of STAT3 mutations in the study IV cohort was also a test of amplicon sequencing sensitivity. The sequencing depth and VAFs are shown in Table 12. Amplicon sequencing had superior sensitivity when compared to capillary sequencing: the median VAF in samples, which were positive by amplicon but negative by capillary sequencing, was 3%, whereas in all amplicon-positive samples it was 17% (Table 12). Considering the possibility of false negative results, the median number of mapped reads was lower in *STAT3* mutation-negative samples (14196 vs. 6931), and it is theoretically possible that some of these samples are false negatives and harbor small mutated clones (Table 12). Finally, the DNA source material could also affect the detection threshold of mutation: the median VAF in PB MNC samples was naturally higher than in whole blood samples (24% vs 10%, Table 12). However, the frequency of *STAT3* mutations was similar in PB MNC (39%) and whole blood (40%) DNA samples.

Table 11. *Allele frequencies and quality values from STAT3 amplicon sequencing analysis of dilution series made from a T-LGL sample with STAT3 D661V (chr17:40474419 T>A) and Y640F (chr17:40474482 T>A) mutations (study IV). Frequency ratio under 0.8 and QVs under 20 are marked in gray.*

Mutation	Dilution	Freq ratio[1]	Total depth[2]	T	T, QV[3]	A	A, QV[3]	G	G, QV[3]	C	C, QV[3]	N	N, QV[3]
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]The calculation of frequency ratio is explained in the Methods section of this manuscript.

[2]Number of reads mapped to *STAT3* exon 21.

[3]The average QV (Phred score) of bases: a non-negative 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 a sequencing error.

Table 12. Amplicon sequencing coverage and VAF of STAT3 mutations of study IV (VAF of the largest variant was used in the analysis in case of multiple mutations).

	All mapped reads, median range)[1]	Total coverage at the variant allele location[2]	VAF, median (range)
Positive by capillary sequencing (n=49)	14472 (840-565170)	4060 (264-94038)	27% (5-51)
Positive only by amplicon sequencing (n=32)	12616 (1255-309580)	5281 (395-21839)	3% (0.6-25)
All amplicon sequencing positive samples (n=82)	14196 (840-565170)	4103 (264-94038)	17% (0.6-51)
All amplicon sequencing-negative samples (n=131)	6931 (1270-804232)	NA	NA
Whole blood DNA samples, amplicon positive (n=35)	7524 (983-29824)	3638 (488-20873)	10% (0.7-49%)
PB MNC DNA samples, amplicon positive (n=36)	20427 (840-565170)	3948 (264-11815)	24% (0.6-48%)

[1,2]*The concepts of mapped reads and coverage are explained in the Methods section, Figure 6.*
Abbreviations: NA, not applicable; VAF, variant allele frequency.

16.4. *STAT5B* screening by capillary and amplicon sequencing in LGL leukemia patient cohort (III)

The exome sequencing of two *STAT3* mutation-negative patients (patients 2 and 3, Table 6) identified a novel *STAT5B* Y665F mutation, and similar to *STAT3* mutations, the discovery led to the sequencing of additional LGL leukemia patients to see if *STAT5B* mutations were recurrent. The screening of *STAT5B* SH2 domain mutations utilized the combination of amplicon (exon 16) and capillary (exons 14-15 and 17-18) sequencing methods, and the patient cohort of study III consisted of 171 T-LGL leukemia and 38 CLPD-NK patients.

Amplicon sequencing reported two additional *STAT5B*-mutated LGL leukemia patients, both of whom carried the N642H mutation in exon 16 of *STAT5B* gene. Thus, the total frequency of *STAT5B* mutations was 2% (4/211). No mutations were detected either in other exons of the *STAT5B* SH2 domain, or in corresponding areas of the *STAT5A* gene screened (exon 17 by amplicon sequencing).

The two patients with N642H mutations had abnormal clinical pictures and phenotypes: the disease presented with an untypically aggressive, fatal, and treatment-resistant form in both cases. Clinical characteristics of the patients with N642H mutations are presented in Table 13, and similar data for the two patients harboring Y665F are in Table 6.

Table 13. Clinical characteristics of the two patients with STAT5B N642H mutation

Patient	Age at dg	Gender	Hb, g/l[1]	Leuk, x10^9/L	Lymph, x10^9/L	Neut, x10^9/L	Plts, x10^9/L	Phenotype	Associated disorders
4 (Study III)	74	Male	122	90	85.5	0.2	280	CD3+ CD56+	Splenomegaly, lymphadenopathy, neutropenia
5 (Study III)	75	Male	168	164.7	131.8	3.3	167	CD3[neg] CD56+	Neutropenia, hemolytic anemia, splenomegaly

[1]*Laboratory values presented are taken at diagnosis.*
Abbreviations: dg, diagnosis; Hb, hemoglobin; Leuk, leukocytes; Lymph, lymphocytes; Neut, neutrophils; Plts, platelets.

Lymphocyte fractions were sorted from all four patients with *STAT5B* mutation to assess the exact location of mutation. The analysis of mutation load in lymphocyte subpopulations was done with capillary sequencing (patients 2 and 3, Table 6) or by amplicon sequencing (patients 4 and 5, Table 13). The results are presented in Figures 15–18.

In the capillary sequencing of sorted fractions from patient 2 (Table 6), *STAT5B* Y665F mutation was specific for CD8+CD56+ leukemic cells, whereas CD3^{neg}CD16+CD56+ NK cells were unmutated (Figure 15). The CD4+ fraction was unmutated in the exome sequencing analysis (Table 8). The amount of DNA extracted from CD3^{neg}CD19+ B cells was too low for capillary sequencing.
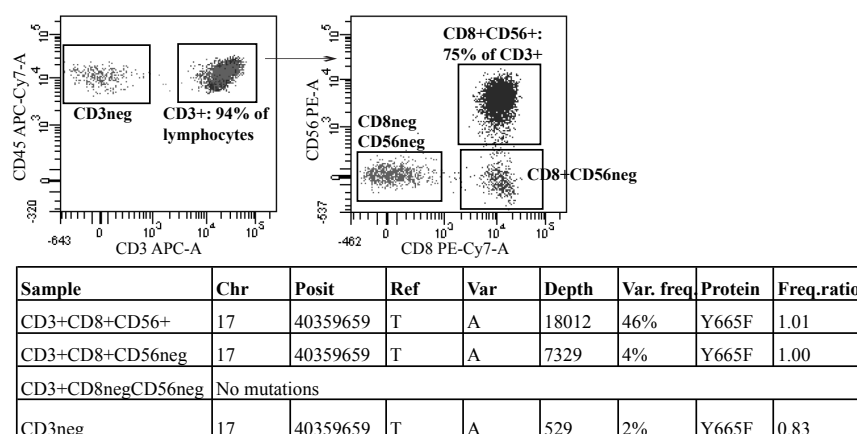

Sample	Chr	Posit	Ref	Var	Depth	Var. freq	Protein	Freq.ratio
CD3+CD8+CD56+	17	40359659	T	A	18012	46%	Y665F	1.01
CD3+CD8+CD56neg	17	40359659	T	A	7329	4%	Y665F	1.00
CD3+CD8negCD56neg	No mutations							
CD3neg	17	40359659	T	A	529	2%	Y665F	0.83

Figure 15. *STAT5B amplicon sequencing results of patient 2 (Table 6), who had CD8+CD56+ T-LGL leukemia (study III). Abbreviations: Chr, chromosome; Freq, frequency; Posit, position; Ref, reference; Var, variant.*

A similar sorting and sequencing strategy was utilized for the PB MNC sample of patient 3 (Table 6). The majority of CD8+ lymphocytes expressed CD56 (Figure 16). Leukemic CD8+CD56+ cells were positive for *STAT5B* Y665F (VAF 48%), and in addition, the mutation was seen in CD3^{neg} cells with VAF of 18% (Figure 16). However, 30% of CD3^{neg} cells expressed aberrantly CD8 and CD56, and thus Y665F-mutated cells in CD3^{neg} fraction could be part of the leukemic clone lacking surface CD3 expression. CD4+ fraction was unmutated (exome sequencing results, Table 9), as were B cells (data not shown).

Sample	Chr	Posit	Ref	Var	Depth	Var. freq	Protein	Freq.ratio
CD3+CD8+CD56+	17	40359659	T	A	15507	48%	Y665F	1.02
CD3+CD8+CD56neg	17	40359659	T	A	10217	2%	Y665F	0.96
CD3+CD8+CD56neg	17	40359729	T	G	10216	1%	N642H	0.93
CD3+CD8negCD56neg	17	40359659	T	A	7660	1%	Y665F	0.89
CD3neg	17	40359659	T	A	3022	18%	Y665F	1.02
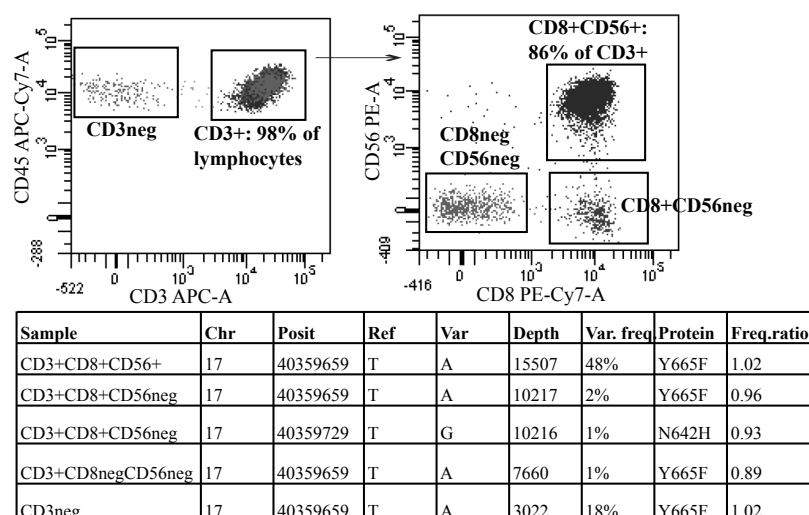
Figure 16. *STAT5B amplicon sequencing results of patient 3 (Table 6), who had CD8+CD56+ T-LGL leukemia (study III). Abbreviations: see Figure 15.*

Patients 4 and 5 (Table 13) carried the *STAT5B* N642H mutation, and the mutation burden in different lymphocyte subpopulations was analyzed using amplicon sequencing.

Patient 4 (Table 13) had a rare, aggressive CD56+ T-LGL leukemia. N642H was seen with extremely high VAF (92%) in the screening of PB MNC DNA. Concordantly, the proportion of CD3+CD56+ leukemic cells of lymphocytes was 97% and the VAF of N642H mutation in leukemic CD3+CD56+ fraction was 99%, suggesting either a homozygous mutation or a loss-of-heterozygosity event (Figure 17). Normal NK cells and B cells were too low in number for sorting. Small CD3+CD16/56^{neg}CD19^{neg} T cell population harbored N642H with low frequency, but this could be related to the purity of the sorted fraction (Figure 17).

Sample	Chr	Posit	Ref	Var	Depth	Var. freq	Protein	Freq.ratio
CD3+CD16/56+	17	40359729	T	G	3677	99%	N642H	1.00
CD3+CD16/CD56neg	17	40359729	T	G	641	6%	N642H	1.01
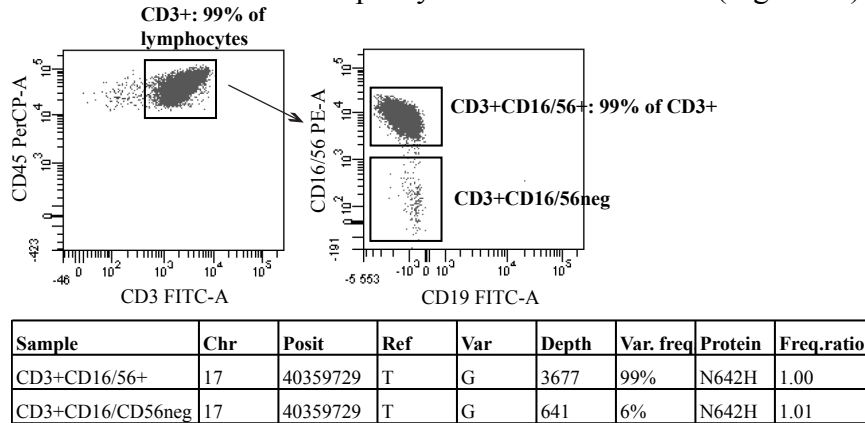
Figure 17. Sorting and STAT5B amplicon sequencing results of patient 4 in study III (Table 13).

Patient 5 (Table 13) had LGL leukemia with an aggressive clinical presentation. In addition to the leukemic CD3^{neg}CD16/56+ NK cell expansion, the patient also had an abnormal CD3+CD16/56+ T cell population, and both were mutated with high VAF (Figure 18). Due to the low number of cells, other lymphocyte fractions were not sequenced for *STAT5B* mutations.

Sample	Chr	Posit	Ref	Var	Depth	Var. freq	Protein	Freq.ratio
CD3negCD16/56+	17	40359729	T	G	2212	48%	N642H	1.00
CD3+CD16/CD56+	17	40359729	T	G	4897	56%	N642H	1.01
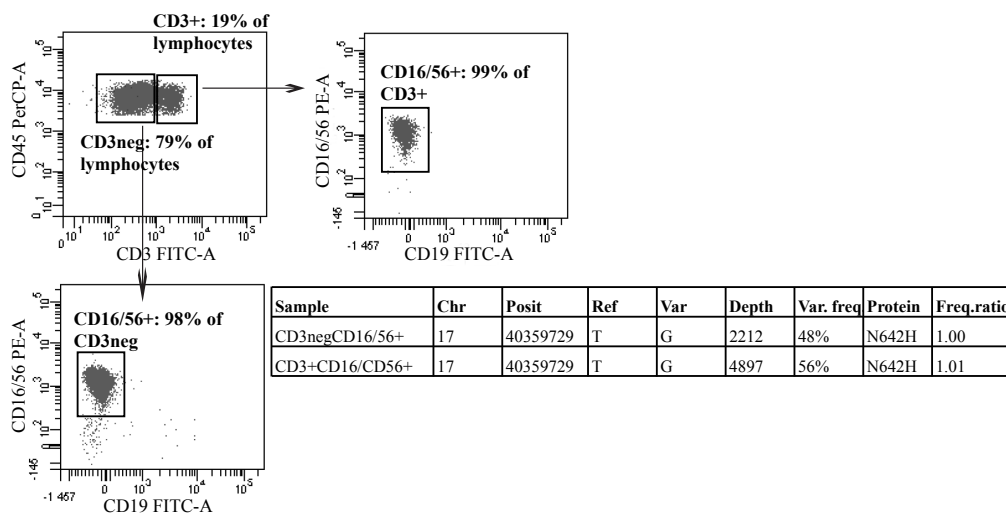
Figure 18. Sorting and STAT5B amplicon sequencing results of patient 5 in study III (Table 13).

To summarize the results of flow cytometry-assisted sorting and amplicon sequencing of *STAT5B*-mutated patients, all had leukemic LGL-populations of either T or NK lineage expressing the CD56 antigen. In the case of patients 2 and 3 (Table 6), CD4+ fractions were unmutated but patient 3 also harbored *STAT5B* Y665F in NK cells (Figure 16). The last two patients (Table 13) had more aggressive disease, and the numbers of CD4+ or CD19+ cells were too low for sorting. In addition to the mutated leukemic population, patient 4 carried *STAT5B* N642H with low VAF in CD3+CD16/56[neg] cells (Figure 17), whereas the mutation was detected in both CD3[neg] and CD3+ cells of patient 5 with high VAF (Figure 18).

16.5. Location of the LGL leukemia mutations in STAT3 and STAT5B protein structure (I, II, III, IV)

In total, 16 different *STAT3* mutations were detected in studies I, II, and IV. *STAT3* mutations and their frequencies in patient cohorts are summarized in Table 14. The localization of mutations in the STAT3 protein structure is presented in Figures 19 and 20.

Table 14. *Different STAT3 SH2 domain mutations in studies I, II, and IV. Mutations seen in patients with multiple mutations in study IV are marked with an "x."*

Study	Study I	Study II		Study IV			
Patients in the study	T-LGL (n=77)	T-LGL (n=120)	CLPD-NK (n=50)	T-LGL (n=174)		CLPD-NK (n=39)	
				1 mutation	≥2 mutations	1 mutation	≥2 mutations
S614R		3 (3%)					
G618R			1 (2%)				
Y640F	13 (17%)	14 (12%)	4 (8%)	32	X	1	X
Q643H					X		
N647I	3 (4%)	1 (1%)	2 (4%)	2	X	2	
K658M		1 (1%)					
K658N	1 (1%)			1			
K658R					X		
I659L					X	1	
D661H	1 (1%)			1			
D661I			1 (2%)				
D661V	7 (9%)	3 (3%)		4	X		
D661Y	7 (9%)	12 (10%)	7 (14%)	16	X	2	X
P678S					X		
Y657_K658insY	1 (1%)			1			
Y657_ins nonsense				1			
STAT3-mutated patients, total	31/77 (43%)[1]	34/120 (28%)	15/50 (30%)	58	17	6	1
				75/174 (43%)		7/39 (18%)	

[1]Two patients in study I had two *STAT3* mutations (Y640F+D661Y, D661V+N647I)
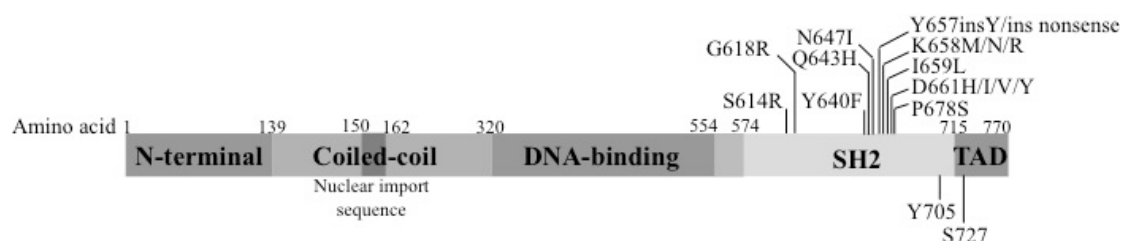
Figure 19. *STAT3 mutations of studies I, II, and IV are placed in linear presentation of STAT3 protein domains. Transcriptional activity-regulating, phosphorylating Y705 and S727 residues are marked below.*
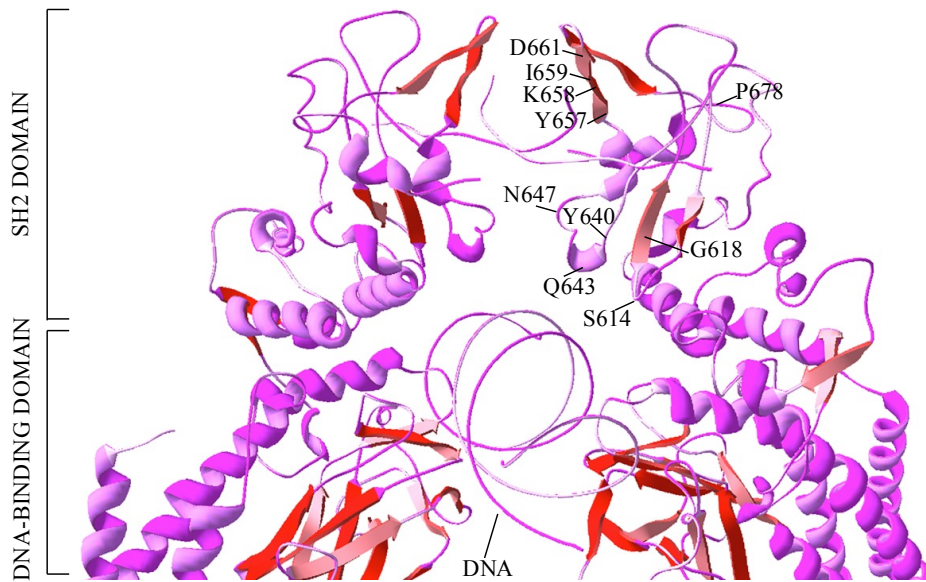
Figure 20. *3D model of the SH2 and DNA-binding domains of dimerized STAT3 attached to target DNA. Mutated residues seen in studies I, II and IV are marked in one monomer. All recurrently mutated amino acids are localized in protein surface loops, facing the other STAT3 monomer. The structure is predicted in Becker et al. Nature 1998[112], and RCSB Protein Data Bank code is 1BG1.*

The only *STAT5B* mutations seen in LGL leukemia so far are Y665F and N642H (2/211, 1% and 2/211, 1% in study III, respectively). Their localization in the STAT5B protein structure is shown in Figures 21 and 22.
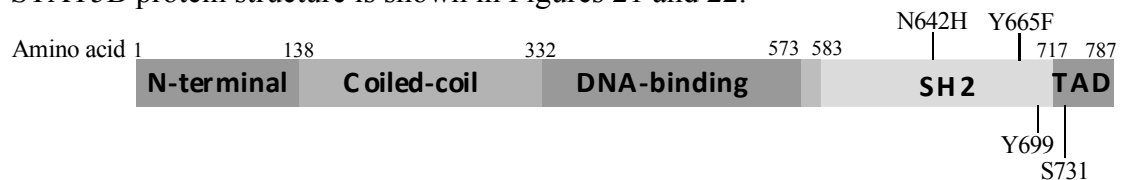
Figure 21. *Linear model of STAT5B protein with mutations N642H and Y665F detected in study III.*
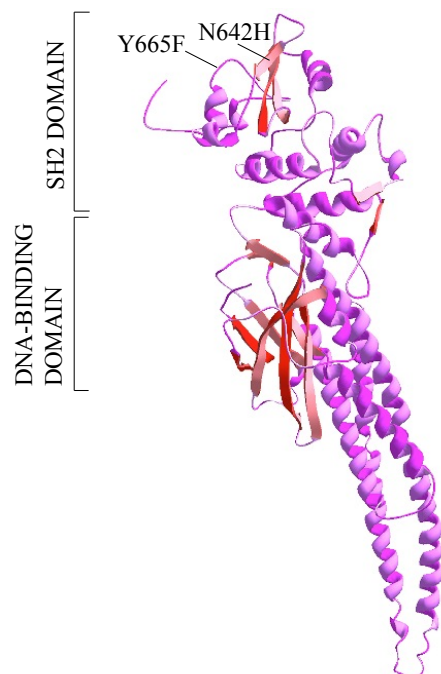
Figure 22. *STAT5B mutations N642H and Y665F projected on a 3D model of STAT5A monomer protein structure, isolated from the unphosphorylated STAT5A trimer complex (RCSB Protein Data Bank 1Y1Y, published by Neculai et al. J Biol Chem 2005[114]). STAT5A and STAT5B share over 90% sequence homology. Similar to STAT3 mutations, N642H and Y665F are located on the surface loops of the SH2 domain (study III).*

17. Functional studies with mutated STAT3 (I, II)

17.1. Western blotting results of T-LGL leukemia patients with *STAT3* mutation (I, II)

STAT3 protein is tyrosine-phosphorylated in its active dimer form. The effects of *STAT3* mutations on the tyrosine phosphorylation status were examined using western blotting with pSTAT3 and total-STAT3 antibodies. In addition to the PB MNC lysate, the analysis was performed from cytosolic and nuclear fractions to decipher the cellular localization of mutated STAT3 protein. The two patients with *STAT3* mutations D661V and Y640F expressed pSTAT3 in nuclear fractions (Figure 23).

In study II, pSTAT3 levels were analyzed from PB MNC lysates, and patients with mutated and wild type STAT3 protein alike expressed pSTAT3 (data not shown).
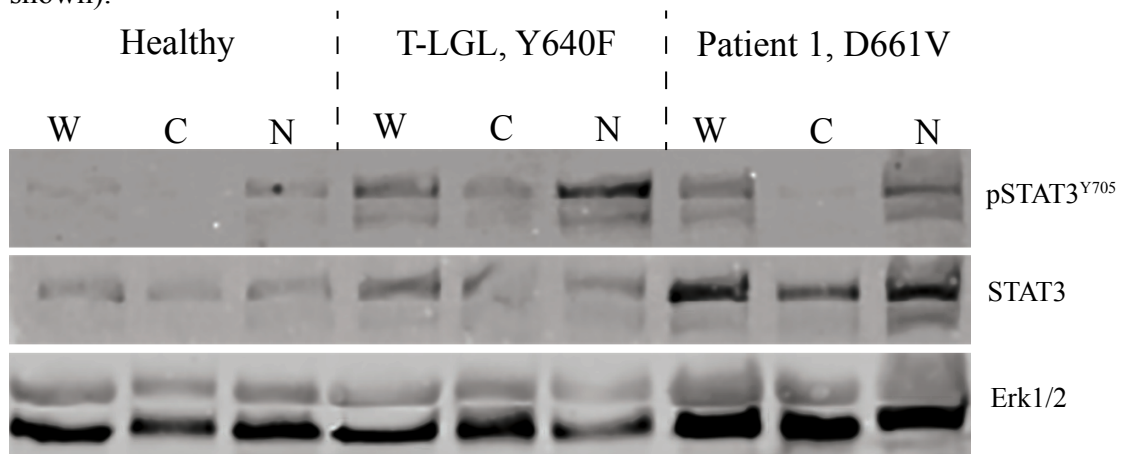
Figure 23. STAT3 western blotting results of a healthy control and two T-LGL leukemia patients with STAT3 mutations (study I). Patient 1 is presented in Table 6. Abbreviations: W, PB MNC whole cell-lysate; C, cytosolic fraction (of PB MNC); N, nuclear fraction (of PB MNC).

17.2. Phospho-STAT3 immunohistochemistry from LGL leukemia patient bone marrow samples (I)

BM biopsies stained with anti-pSTAT3^{Y705} and anti-CD57 antibodies were examined from 3 *STAT3*-mutated patients and 1 healthy control in study I. All LGL leukemia patients displayed positive nuclear staining of BM-infiltrating lymphocytes with pSTAT3^{Y705} antibody, and the lymphocytes were also CD57+ (Figure 24).

Similar anti-pSTAT3 staining was performed using slides obtained from two *STAT3* mutation-negative patients, both of whom demonstrated nuclear localization of pSTAT3^{Y705}, although to a lesser extent compared to the *STAT3*-mutated patients (Figure 25).
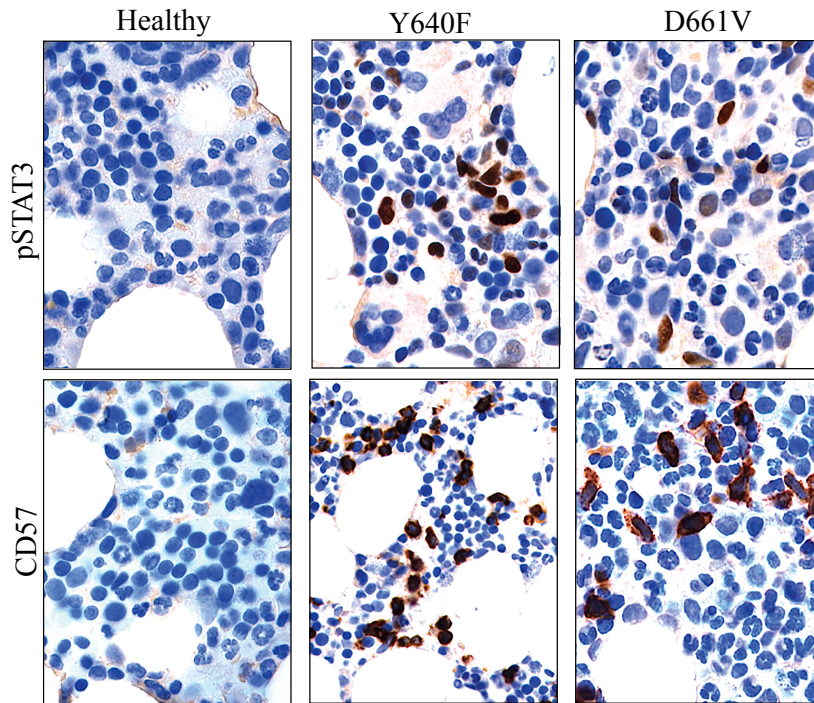
Figure 24. *IHC results from two STAT3-mutated T-LGL leukemia patients and a healthy control (study I). Anti-pSTAT3^{Y705} (upper panel) and anti-CD57 (lower panel) are marked with brown staining (magnification x63)*
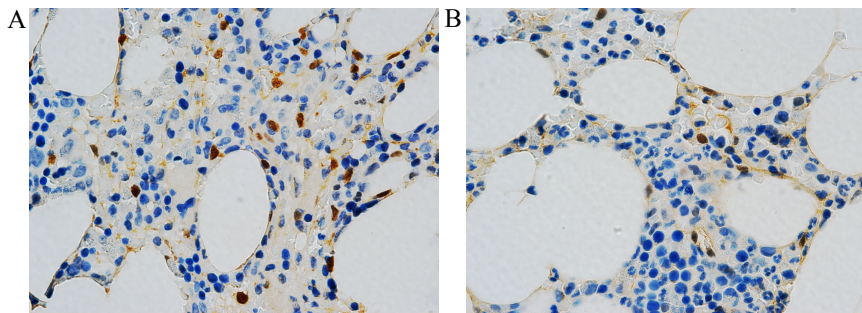
Figure 25. *Anti-pSTAT3^{Y705} IHC results from two T-LGL leukemia patients (A and B present two different patients) with wild type STAT3 (magnification x63)*

17.3. Transcriptional activity of mutated STAT3 using reporter assay (I)

STAT3 is a transcription factor, and one way to measure the effects of SH2 domain mutations is to analyze the transcriptional activity of mutated versus wild type STAT3. HEK2 cells were simultaneously transfected with constructs containing either wild type or Y640F/D661V mutated *STAT3*, and a luciferase construct driven by STAT3-responsive SIE. Luciferase activity was measured in triplicates after mock or IL-6 stimulation. Both mutated constructs, especially Y640F, showed increased transcriptional activity at baseline and after IL-6 stimulation (Figure 26).
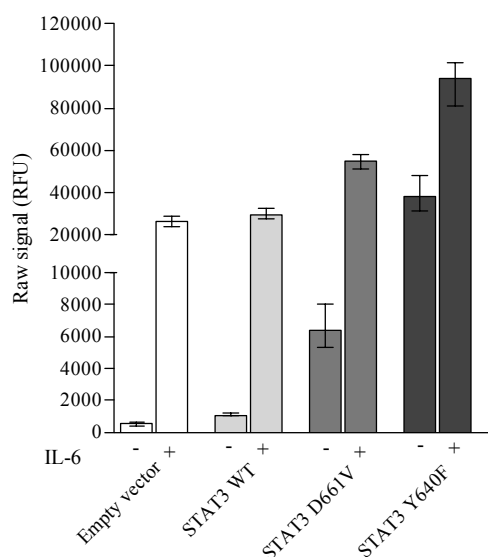
Figure 26. *STAT3 reporter assay results. Analysis was done in triplicates (study I). Columns present mean intensities and* I *bars mark range.*

18. Functional studies with mutated STAT5B (III)

18.1. Western blotting results of T-LGL leukemia patients with *STAT5B* mutation (III)

Similar to STAT3, activation of STAT5B is also tyrosine phosphorylation-dependent. The tyrosine phosphorylation levels of mutated STAT5B were analyzed with Western blotting, using antibodies for pSTAT5^{Y694}, total STAT5, pSTAT3^{Y705} and total STAT3. The analysis was done from PB MNC nuclear lysates of STAT5B Y665F-mutated T-LGL leukemia patients (n=2), STAT3/STAT5B-unmutated T-LGL leukemia patients (n=3) and healthy controls (n=5) (Figure 27). The STAT5B-mutated patients expressed pSTAT5 in nuclear fraction, but it was not seen in other samples (Figure 27). GAPDH was used as a control of protein load, but no nuclear-specific control was included in the assay.
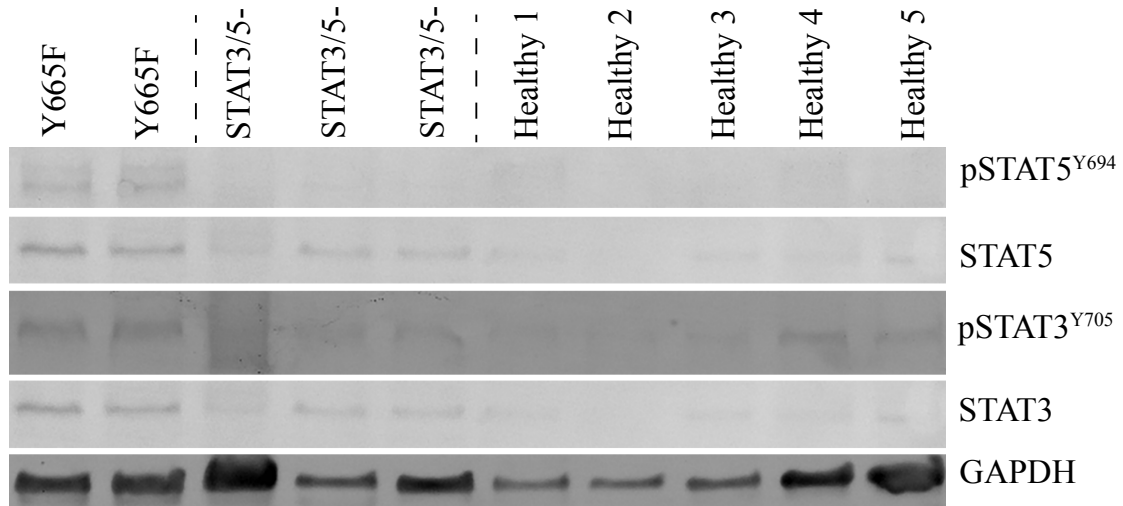
Figure 27. *Western blotting results from PB MNC nuclear lysates of T-LGL leukemia patients and healthy controls (study III).*

18.2. Analysis of STAT5 phosphorylation using enzyme-linked immunosorbent assay (III)

Phosphorylated and total STAT5 protein was analyzed with enzyme-linked immunosorbent assay (ELISA) in addition to Western blotting in study III. Protein lysates from PB MNC fraction were used in the analysis. The pSTAT5 antibodies detect pSTAT5A^{Y699} and pSTAT5B^{Y964}. The two T-LGL leukemia patients with *STAT5B* Y665F mutation showed increased amount of pSTAT5 when compared to STAT3-mutated (n=2) or STAT3/STAT5B wild type (n=4) LGL leukemia patients and healthy controls (n=4) (Figure 28). Thus, the results were in accordance with the pSTAT5 western blotting analysis.
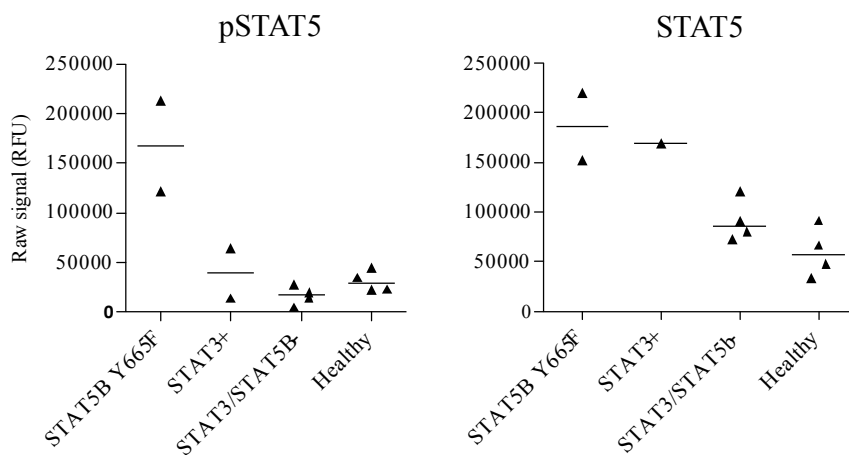
Figure 28. *Results of pSTAT5 and STAT5 ELISA (study III). Horizontal lines define median intensities.*

60

18.3. Transcriptional activity of mutated STAT5B using reporter assay (III)

The transcriptional activity of the STAT5B proteins was analyzed in HeLa cells. A wild-type STAT5B construct, or a construct containing either the N642H or Y665F mutation, were transfected into HeLa cells together with a luciferase reporter gene under the control of STAT5 response element. Both mutated vectors, especially N642H, activated the transcription of luciferase without cytokine stimulation (Figure 29).
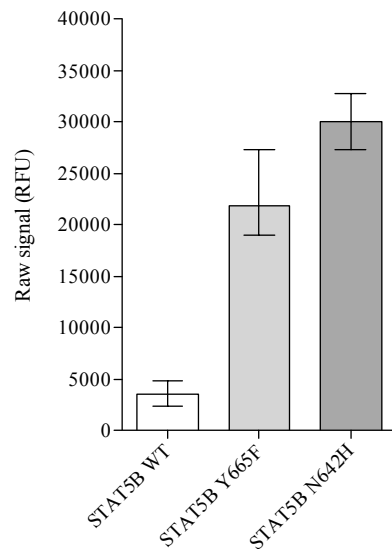
Figure 29. *Results of STAT5B reporter assay with wild-type, Y665F and N642H constructs (study III). Analysis was done in triplicate. Columns present means intensities and I bars mark range.*

19. Gene expression analysis in LGL leukemia patients (I, II, III)

Gene expression analysis was performed in studies I (RNA sequencing and microarray), II (microarray), and III (RNA sequencing and microarray). RNA sequencing data in study III was used only to analyze the expression of genes with somatic mutations.

In study I, the gene expression in CD8+ cells of *STAT3*-mutated (n=8) and unmutated (n=10) T-LGL leukemia patients and healthy controls (n=5) were compared using an Illumina Human HT-12v4 microarray. The expression levels of selected STAT3-responsive genes and other JAK-STAT-related genes are shown in Figure 30. The expression of IL-6, an upstream regulator of STAT3, showed a wide variation within the group of *STAT3*-mutated patients (Figure 30). The expression of JAK2, which is both STAT3 upstream regulator and STAT3 target gene[126], did not differ significantly between mutated and unmutated patients, although its expression was higher in *STAT3*-mutated patients versus healthy controls (Figure 30). SOCS3 expression levels were similar in all groups (Figure 30). In the comparison of the RNA sequencing data of CD8+ T cells from T-LGL leukemia index patient (carrying the D661V mutation) and healthy control CD8+ T cell RNA several STAT3-responsive genes were upregulated, including FASL and JAK2 (data not shown).[126]
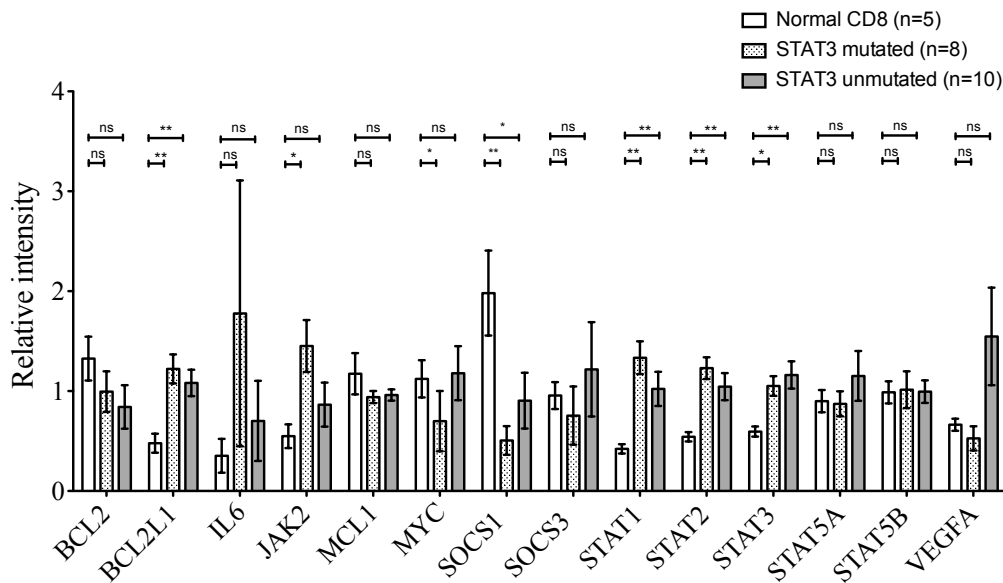
Figure 30. *Expression array results of selected STAT3 responsive genes and other JAK-STAT related genes of study I. Mean relative intensity is shown (logarithmic scale) and I bars present standard error of mean. Statistical significance of differences between expression levels are marked as follows: *, p<0.05; **, p<0.01; ***, p<0.001; ns, not significant.*

The gene expression comparison using Affymetrix microarray in study II consisted of both *STAT3*-mutated (n=2) and unmutated (n=1) T-LGL leukemia patients and healthy controls (n=3). T-LGL expansions from LGL leukemia patients were sorted based on the Vbeta expression on CD8+ cells, and CD8+CD57+ T cells from healthy individuals were used as controls. Similar to study I, several STAT3-downstream genes[125] were upregulated in both *STAT3*-mutated and –unmutated patients (data not shown).

In study III, novel *STAT5B* mutations added a new group to the gene expression analysis: CD8+ RNA was extracted from *STAT5B* Y665F mutated (n=2) and N642H mutated (n=1) patients, 3 patients harboring a *STAT3* mutation, and 2 patients with wild type *STAT3* and *STAT5B* (confirmed by exome sequencing, results are not presented in this thesis). CD8+ (n=2) and NK cells (n=2) were separated from healthy individuals (anonymous Finnish Red Cross buffy coat samples) and were used both as controls and biological replicates in the assay. LGL leukemia patients carrying the *STAT5B* Y665F mutation grouped together in the clustering analysis, and belonged to the same branch of the dendrogram with other LGL leukemia patients (Figure 31). Healthy controls were one group, and patient 5 with the *STAT5B* N642H mutation (Table 13) was an outlier as was expected based on the aggressive clinical presentation (Figure 31). Related to the data of the distance dendrogram, the gene expression profiles of LGL leukemia patients were very similar, independent of the mutation status, and no significantly differentially expressed genes were found in a comparison of *STAT3+* vs. *STAT5B+* vs. wild-type *STAT3/STAT5B* LGL leukemia patients. The analysis of STAT5B mutated patients versus healthy CD8+ controls yielded 4894 genes with diverging expression (p<0.05). In subsequent comparison, the expression levels of the most over- and underexpressed of these genes were remarkably similar between all LGL leukemia patients (patient 5, Table 13, was excluded from the analysis) (Figure 32).
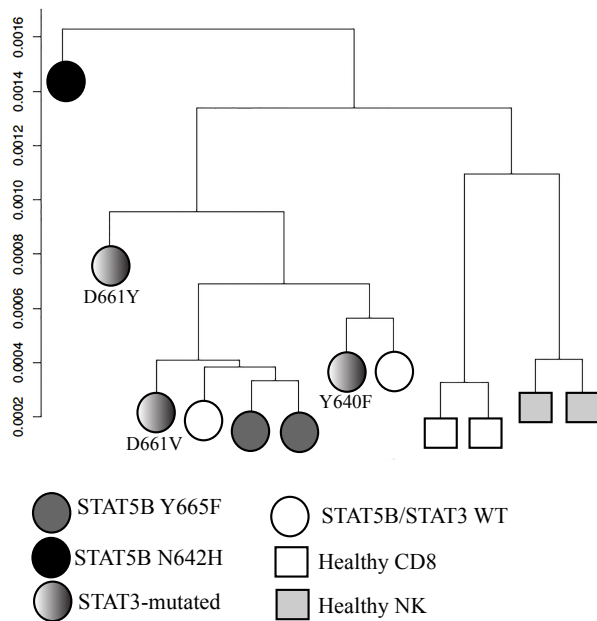
Figure 31. *Distance dendrogram of the gene expression profiles between LGL leukemia patients and healthy controls (study III). Each balloon represents one LGL leukemia patient or one healthy control. The type of the STAT3 mutation is marked under the balloon.*
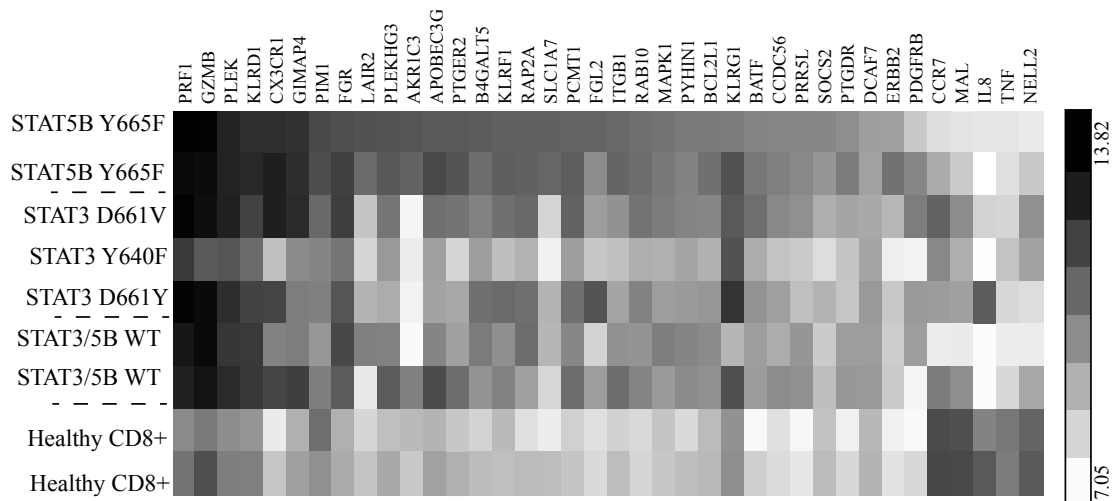

Figure 32. *Gene expression levels of selected differentially expressed genes in LGL leukemia patients and healthy controls in study III (log2 scale).*

20. Clonal hierarchy in LGL leukemia (IV)

In study I, the *STAT3* D661V mutation of patient 1 (Table 6) was located only in the CD8+Vb16+ immunodominant clone and other lymphocyte subpopulations were unmutated (Figure 10), whereas in the case of *STAT5B* mutated patients, the mutations were also harbored by other lymphocyte fractions in addition to the mutated major clone (Figures 15-18). In study IV, amplicon sequencing revealed that a substantial proportion of the LGL leukemia patients carried multiple *STAT3* mutations, existing either in the same or different clones (Figure 14, Table 14). Consequently, the exact location of *STAT3* mutations and clonal hierarchy of lymphocytes was analyzed in study IV using flow cytometry-assisted sorting and amplicon sequencing of the *STAT3* exon 21. Samples were available from 12 T-LGL leukemia and 1 CLPD-NK patients. Healthy control BM samples (n=10) and reactive LGL expansions of dasatinib-treated CML patients (n=8) were also analyzed, but they all had wild type exon 21 of *STAT3*.

20.1. *STAT3* mutations in lymphocyte subpopulations (IV)

In the case of T-LGL leukemia patients, Vbeta analysis data were the basis of the sorting of different lymphocyte fractions. CD8+ lymphocyte fraction was divided into Vbeta-positive (LGL expansion) and −negative cells, and in addition CD4+ lymphocytes, CD3^{neg}CDCD16/56+ NK cells and CD3^{neg}CD19+ B cells were sorted and analyzed.

Samples from T-LGL leukemia patients with one major immunodominant expansion in Vbeta analysis and only one *STAT3* mutation in the original screening revealed that in 4 of 5 cases analyzed, the STAT3 Y640F mutation was located solely in the immunodominant expansion (Figure 33). In one patient, three additional STAT3 mutations with low VAF were detected in the seemingly polyclonal CD8+Vb21.3^{neg} fraction (Figure 33C). In all five cases, CD4+, NK, and B cells were unmutated.
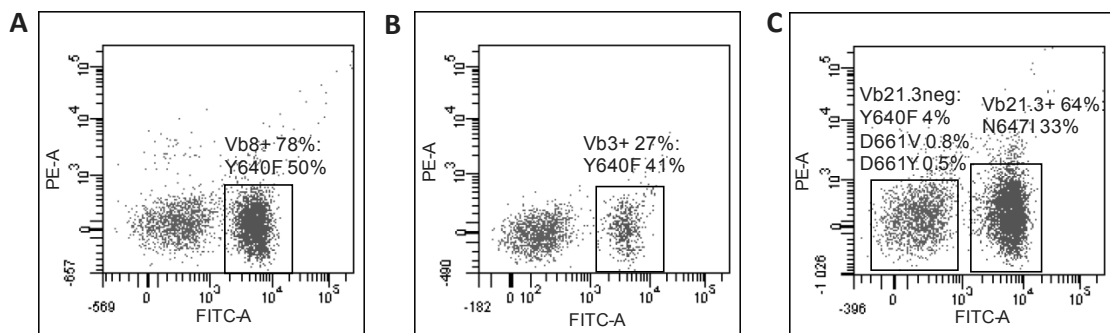

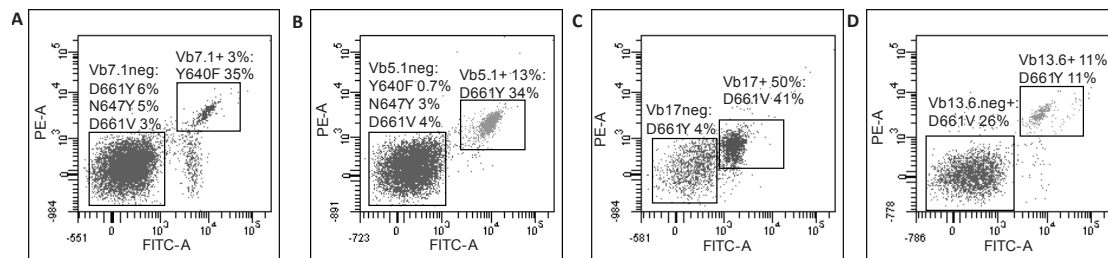***Figure 33.*** *Three examples of Vbeta results (proportion of LGL expansion/CD8+ cells) and the location and VAF of STAT3 mutations in sorted fractions in study IV. Patients in A and B had monoclonal STAT3-mutated expansion, whereas the patient shown in C had three minor STAT3-mutated clones in the Vb21.3^{neg} fraction in addition to the mutated Vb21.3+ clone.*
Axes: A: Vb13.1-PE, Vb8-FITC, Vb13.6-PE+FITC. B: Vb5.3-PE, Vb3-FITC, Vb7.1-PE+FITC. C: Vb23-PE, Vb21.3-FITC, Vb1-PE+FITC.

Two patients with multiple *STAT3* mutations in the original screening were similarly sorted and analyzed. As an extreme example, the first of these patients had a total number of four different *STAT3* mutations, two of which were situated in separate Vb-positive expansions and the other two in polyclonal CD8+ fraction (Figure 34A-B). The second patient harbored two *STAT3* mutations in two CD8+ expansions (Figure 34C-D). Other lymphocyte populations did not carry *STAT3* mutations.

Figure 34. *Vbeta results (proportion of LGL expansion/CD8+ cells) and the location and VAF of STAT3 mutations in sorted fractions in study IV. The first patient had four different STAT3 mutations (A-B), whereas the second patient harbored only two STAT3 mutations in two separate Vbeta expansions (C-D). The percentage of Vbeta expansion is shown, and the numbers after mutations stand for VAF.*
Axes:A: Vb5.3-PE, Vb3-FITC, Vb7.1-PE+FITC. B: Vb18-PE, Vb20-FITC, Vb5.1-PE+FITC. C: Vb9-PE, Vb16-FITC, Vb17-PE+FITC. D: Vb13.1-PE, Vb8-FITC, Vb13.6-PE+FITC.

20.2. TCRB sequencing of *STAT3* mutated T-LGL leukemia patients (IV)

Second generation sequencing enables the analysis of TCRB CDR3 rearrangement repertoires at a new level, compared to capillary sequencing where only major clones can be detected and sequenced, or flow cytometry analysis, which does not recognize all TCRB rearrangements and does not give data on the CDR3 amino acid sequence. In many flow cytometry-sorted cases, *STAT3* mutations were located in apparently polyclonal parts of the CD8+ lymphocyte population. The exact amino acid sequence of the TCRB CDR3 region could also clarify the quality (nature) of antigens behind leukemic expansions. Thus, three T-LGL patients with multiple *STAT3* mutations were chosen for detailed analysis of the TCRB CDR3 repertoire. The patients were not HLA-matched. The results of two patients are presented here as an example.

The first patient harbored two *STAT3* mutations: Y640F in the immunodominant Vb3+ expansion and I659L in Vb3[neg] CD8+ cells (Figure 35A). The patient was anemic and was treated with MTX: both clones persisted during and after MTX treatment (Figure 35B). In the TCRB CDR3 analysis, the Vb3+ population consisted mostly of a major rearrangement (74%), and the Vb3[neg] fraction was similarly skewed with a single rearrangement of 74% of the reads (Figure 35C-D). Simpson's diversity indexes for Vb3+ and Vb3[neg] fractions were 0.45 and 0.45, respectively, emphasizing the level of narrowing of the TCRB CDR3 repertoire in the CD8+ lymphocytes.
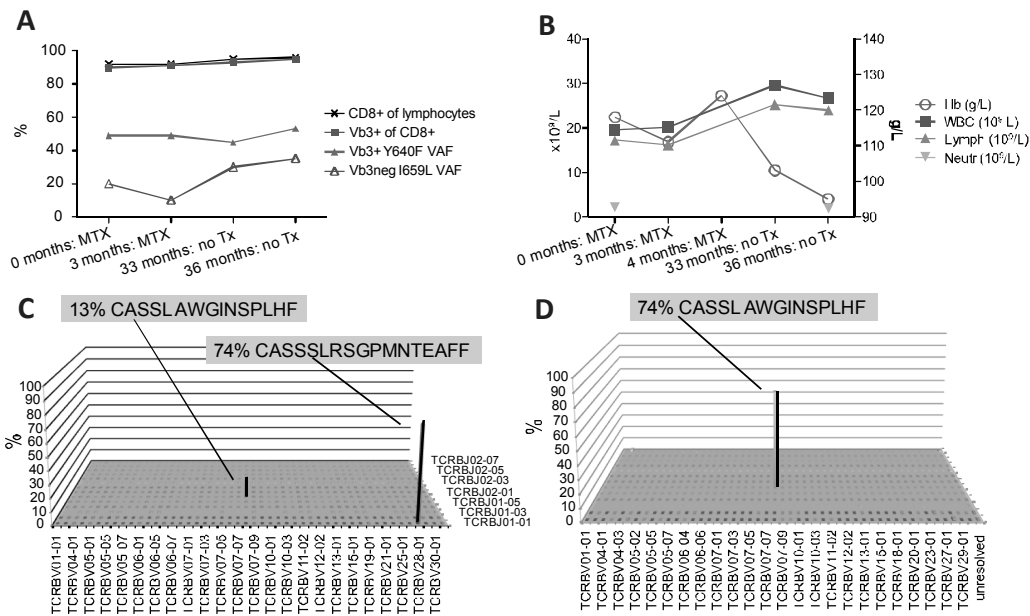
Figure 35. *Vbeta analysis (A), STAT3 amplicon sequencing results (A) and laboratory values (C) of a T-LGL leukemia patient with two STAT3 mutations in study IV. TCRB CDR3 deep sequencing was done after 33 months follow-up: the results of sorted Vb3+ expansion (C) and Vb3^{neg} cells (D) are shown, with CDR3 amino acid sequences of the largest rearrangements indicated.*

The *STAT3* mutation status of the second patient is presented in Figure 33C: in addition to the N647I-mutated Vb21.3+ immunodominant clone, the patient carried three *STAT3* mutations in Vb21.3^{neg} cells. The Vb21.3+ fraction harbored a major TCRB rearrangement of 63%, and surprisingly also another rearrangement with an unresolved V-gene (26%) (Figure 36A). The sorted Vb21.3+ population looked almost 100% pure in flow cytometry analysis, indicating that both expansions were recognized by the same Vbeta mAb despite the different CDR3 amino acid sequence. The VAF of *STAT3* N647I was 33%, corresponding well to the percentage of the major CDR3 rearrangement, whereas the smaller expansion was probably unmutated (Figure 36). The Vb21.3^{neg} cells contained only three small expansions of 4%, 3%, and 2%, which could represent small *STAT3* mutated clones seen in amplicon sequencing (Figure 36B). Simpson's diversity index for the Vb21.3+ fraction was 0.53, related to the narrowing of the TCRB repertoire. The Vb21.3^{neg} sample did not contain large expansions, and the diversity index was 0.97, suggesting a relatively normal TCRB CDR3 repertoire.
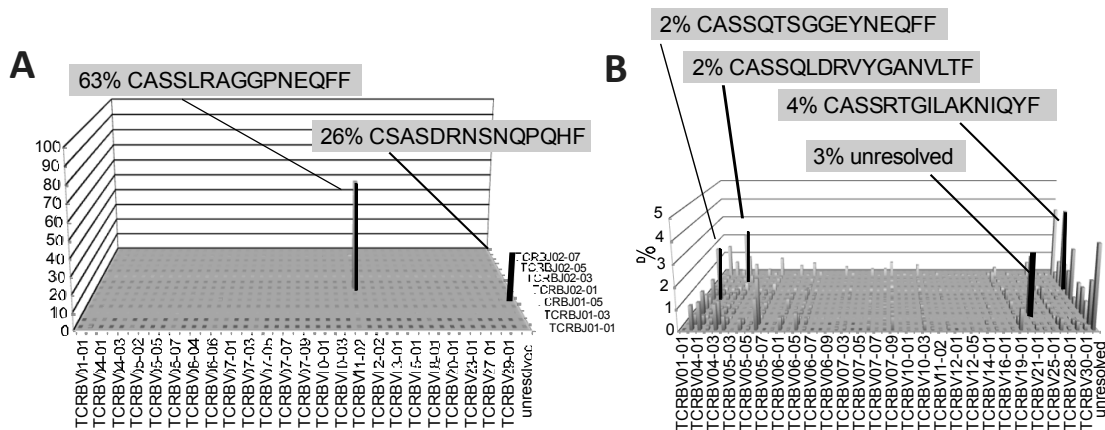
Figure 36. *TCRB sequencing results and CDR3 amino acid sequences of a T-LGL leukemia patient with four STAT3 mutations, all in different clones in study IV (shown in Figure 33C). Sorted Vb21.3+ fraction consisted mostly of two different TCRB rearrangements (A), both recognized by the same Vbeta mAb although the amino acid sequences were dissimilar. The Vb21.3^{neg} fraction appeared more polyclonal and the four largest rearrangements are marked (B).*

The 20 most frequent TCRB CDR3 amino acid sequences of each patient were compared to the TCRB CDR3 repertoire of other patients, but all sequences were private and did not exist in the other T-LGL leukemia patients analyzed. No similarity was seen between amino acid sequences of multiple rearrangements within one patient. However, the CDR3 rearrangements of the immunodominant clone in two patients were seen with low frequency in healthy controls (only 100% matches taken into consideration, data not shown). The patient presented in Figure 35 had the most skewed TCRB repertoire, and in the comparison of the five largest rearrangements, none were shared with healthy individuals (n=165).

20.3. *STAT3* mutations during follow-up and treatment of LGL leukemia patients (IV)

Sorted samples from different timepoints were analyzed from *STAT3*-mutated LGL leukemia patients treated with MTX (n=2) and cyclophosphamide (n=3). The median observation period was 33 months.

Both patients treated with MTX had T-LGL leukemia. The results of a patient with *STAT3* mutations Y640F and I659L are presented in Figure 35A-B: the size of Vb3+ Y640F-mutated clone was unchanged during therapy, whereas the proportion of I659L-mutated cells in Vb3^{neg} fraction was slightly lower during MTX therapy. The patient achieved partial response as hemoglobin values improved during treatment (Figure 35B). The treatment response of the second patient was similar, with improvement in hemoglobin levels, but the Y640F-mutated Vb17+ clone was unchanged (Figure 37).
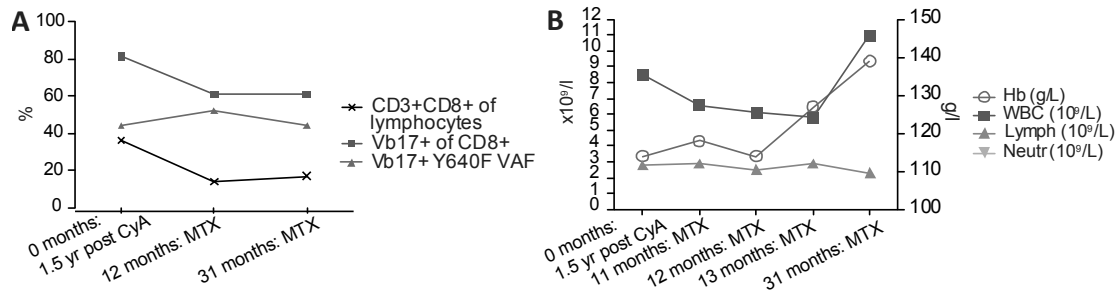
Figure 37. *Results of flow cytometry and STAT3 amplicon sequencing of a T-LGL leukemia patient (A), and laboratory values of the same patient during follow-up (B) in study IV.*

Among the patients (n=3) treated with cyclophosphamide, 2 patients showed better treatment responses compared to MTX-treated patients. The first patient was diagnosed with T-LGL leukemia, but the major CD8+Vb20+ clone was unmutated, whereas an aberrant CD3[neg]CD16/56[neg] fraction harbored a D661Y mutation (Figure 38A). The patient went to complete remission, during which both the Vb20+ clone and D661Y-mutated cells were suppressed and normal NK cell population was restored (Figure 38A-B). The second patient was diagnosed with CLPD-NK, but similarly to the first patient, D661I-mutated cells had the abnormal phenotype of CD3[neg]CD16/56[neg]. The mutated cells were eradicated during cyclophosphamide therapy, with the rise of a normal NK cell fraction (Figure 38C-D). After cessation of cyclophosphamide administration, the mutated cells returned, but no follow-up data from later timepoints exist (Figure 38C).
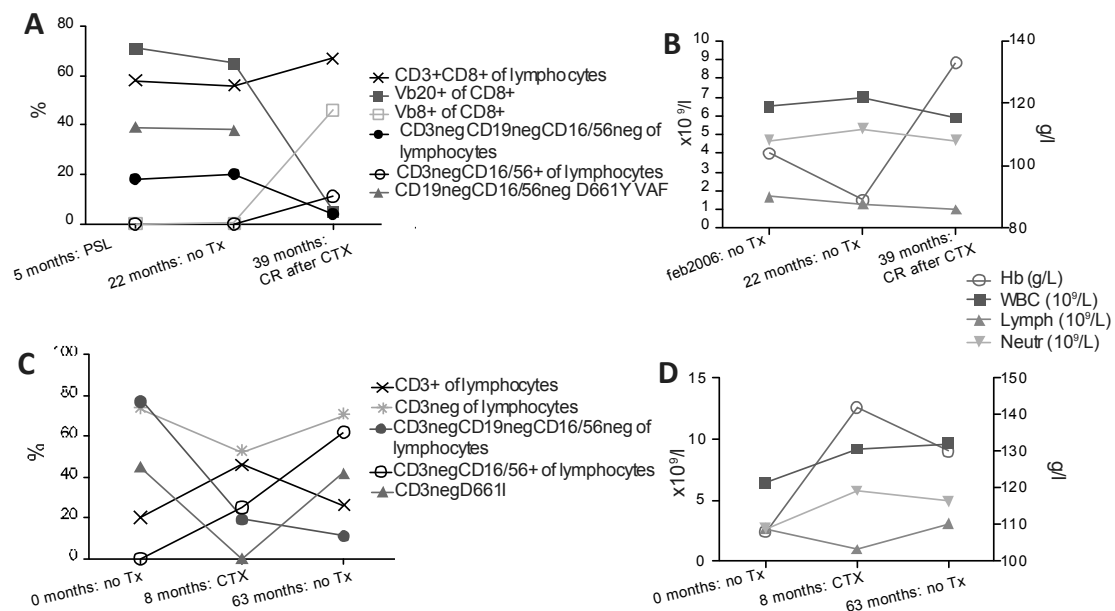

Figure 38. *Results of flow cytometry analysis, STAT3 amplicon sequencing, and blood counts at different timepoints from two cyclophosphamide-treated patients in study IV. (A-B) The first patient was diagnosed with T-LGL leukemia, but a D661Y-mutation was in CD3[neg] cells; this disappeared during complete remission. (C-D) The second patient had CLPD-NK and complete remission was connected to the eradication of mutated cells and recovery from anemia.*

The third cyclophosphamide-treated T-LGL leukemia patient had two *STAT3* mutations in two separate clones (Figure 39). The patient failed therapy, after which only the major Vb13.6+ clone with the D661Y mutation was seen in amplicon sequencing analysis, whereas the minor clone harboring the D661V mutation disappeared (Figure 39).
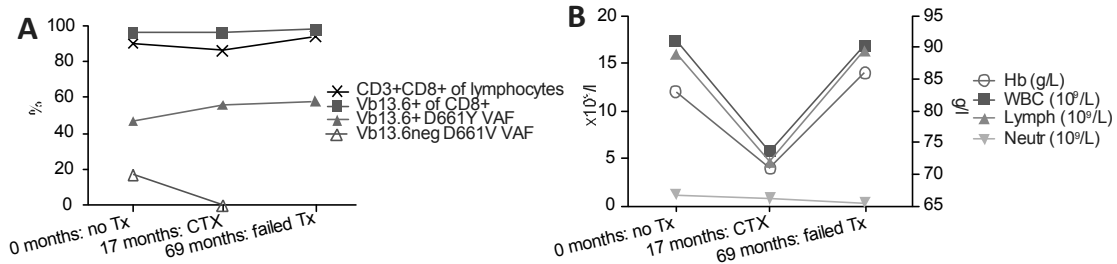

Figure 39. *Results of a T-LGL leukemia patient who failed cyclophosphamide therapy (study IV). The minor D661V-mutated clone disappeared, while cyclophosphamide did not have any effect on the major Vb13.6+ clone (A). (B) Laboratory values at different time points.*

21. Clinical impact of *STAT3* and *STATB* mutations in LGL leukemia (I, II, III, IV)

The clinical impact of *STAT3* mutations was assessed in studies I, II and IV.

In study I, the comparison was made between *STAT3* mutated and unmutated T-LGL leukemia patients and also between the three most common mutations (Y640F, D661Y, and D661V). *STAT3* mutation-carrying patients were more often neutropenic (24/31, 77% vs. 23/46, 50%, p=0.02) and had concomitant RA (26% vs. 6%, p=0.02). The patients chosen for study I had all monoclonal T-LGL leukemia: a slight difference was seen in the size of the major clone between mutated and unmutated cases (median clone sizes 89% [range 62-99%] and 77% [range 34-96%], p=0.03). No significant differences were observed between carriers of different *STAT3* mutations, but this could be due to the low number of patients in each group.

The analysis in study II was concordant with study I: RA was more frequent in *STAT3*-mutated cases (18% vs. 7%, p=0.024) and autoimmune hemolytic anemia was also more common in mutated patients (14% vs. 4%, p=0.039). The size of the immunodominant clone was over 25% of the CD8+ cells in all cases harboring *STAT3* mutations, whereas among nonmutated cases, 23% of the patients had clones smaller than 25%. The pattern of KIR expression was analyzed from 25 CLPD-NK patients, 7 of which carried *STAT3* mutations: almost all analyzed patients (23/25) had abnormal KIR expression, including all mutated cases. *STAT3*-mutated patients needed more lines of therapy during the observation period (median number of therapies 2.4 vs. 1.2, p=0.03). The presence of mutations or the type of LGL leukemia (T-LGL or CLPD-NK) did not affect overall survival.

A large proportion of patients in the study IV cohort harbored multiple mutations in the *STAT3* gene. Thus, T-LGL leukemia patients were divided into three groups: unmutated patients (n=99), patients with one *STAT3* mutation (n=58), and patients harboring ≥2 mutations (n=17). Only 1 of 7 *STAT3*-mutated CLPD-NK patients had

≥2 mutations, and the comparison of CLPD-NK patients revealed no statistically significant differences between mutated or nonmutated patients. The most interesting difference in the T-LGL leukemia cohort was the prevalence of RA: it was seen in 41% of patients with ≥2 mutations vs. 22% of patients with one mutation vs. 6% of unmutated patients (p=0.0001). The median size of the immunodominant clone was the highest in patients with one mutation (83%), whereas in patients with multiple mutations it was 65%, probably related to the existence of multiple clones (p<0.0001). Neutrophil count at diagnosis was the lowest in patients with multiple mutations, but did not reach statistical significance. The T-LGL leukemia patients carrying Y640F (n=33) or D661Y (n=18) mutation were also compared: patients harboring D661Y had lower hemoglobin values than Y640F-mutated patients, but higher leukocyte and lymphocyte counts (median hemoglobin 104 vs. 122 g/l, p=0.0449, leukocytes 12.4 vs. 5.1 x106/l, p=0.0440, lymphocytes 7.5 vs. 3.2 x106/l, p=0.0028, respectively).

The number of patients harboring *STAT5B* mutations in study III was low (4/211, 2%) and the calculation of statistics was not possible. However, some notions were made based on the mutation status and clinical information. Leukemic LGLs of all *STAT5B*-mutated patients expressed CD56 antigen on their surface, one case being classified as CD3[neg] NK-cell type LGL leukemia (with concomitant aberrant CD3+CD56+ population) whereas 3 cases were rare CD8+CD56+ T-LGL leukemias (Tables 6 and 13). The two patients (Table 13) carrying a *STAT5B* N642H mutation were the only cases in the cohort of 211 patients with untypical, aggressive LGL-leukemia, suggesting a correlation between *STAT5B* N642H and an aggressive clinical picture.

DISCUSSION

Since the discovery and characterization of LGL leukemia nearly thirty years ago, researchers have speculated whether the disease is caused by dysregulated T cell homeostasis resulting in extremely prolonged immune response, or by malignant transformation leading to chronic lymphoproliferative disease.[1,2] Considering the former hypothesis, several studies have shown that multiple cell survival and apoptosis-related signaling pathways function aberrantly in LGL leukemia, but the actual antigen causing the initial immune activation has not been discovered. Despite the detailed mapping of several molecular abnormalities in leukemic LGLs, these discoveries have not yet led to the development of targeted therapies.[100] The latter theory of oncogenic transformation behind LGL leukemia contains an idea of possible somatic mutations, which lead to the expansion of LGL clone. In many malignancies, emerging second-generation sequencing and bioinformatics tools have deepened the understanding of molecular events causing transformation and consequently led to the discovery of novel targeted therapies.

The goal of the first study was to explore the genetic landscape in LGL leukemia using second-generation sequencing methods. The application of exome sequencing proved to be successful and led to the major discovery of the thesis: 40% of T-LGL leukemia patients harbored somatic mutations in the *STAT3* gene, and in study II, similar *STAT3* mutations were seen also in 30% of CLPD-NK patients. The original results of *STAT3* mutations in study I have been reproducible: the frequency rate of mutations is up to 70% in T-LGL leukemia and 30-40% in CLPD-NK in different publications.[210,211] Sporadic *STAT3*-mutated cases have been reported among diffuse large B cell lymphoma (DLBCL) and T cell lymphoma (especially in anaplastic CD30+ T cell lymphoma) patients, but in solid tumors, they are practically absent.[211-214]

The inspection of the STAT3 protein 3D model suggested that these mutations could change the dimerization properties of STAT3. All recurrent mutations were located in the specific surface loops of STAT3 SH2 domain. Mutated residues Y657, K658, I659, and D661 belong to the BG loop (amino acids 656-677) near the C-terminal, and interestingly, substitution of A662 and N664 of BG loop with cysteine results in sulfhydryl bonds between STAT3 monomers and creates a constitutively active STAT3 dimer (STAT3-C).[113] Frequently mutated amino acid Y640 is located in another surface loop annotated as DB, and S614 and G618 seen in study II are situated in a third loop (βC) next to DB.[112] All these loops are in close contact with the other STAT3 monomer during dimerization of STAT3, and the BG loop is situated next to the phosphotyrosine-binding pocket.[112] Notably, the mutations seen in LGL leukemia are different from the germline mutations of hyper-IgE syndrome patients, which are thought to function in a dominant-negative manner.[139,140] STAT3 SH2 domain mutations caused constitutive phosphorylation of STAT3 and increased its transcriptional activity in the experiments of study I and II. Some DLBCL mutations were located in the coiled-coil domain of STAT3 near nuclear import sequence, and functional data of one of these (M206K) showed increased phosphorylation and transcriptional potential, consistent with the effects of SH2 domain mutations.[213] The most common mutation, Y640F, was able to cause leukemic transformation in a mouse bone marrow model.[212] Taken together, structural data and results of functional experiments led to the hypothesis that STAT3 SH2

domain mutations could stabilize the reciprocal phosphotyrosine-SH2 interaction and strengthen the dimer structure of STAT3. STAT3-C is activated independent of phosphorylation status, but it has not been confirmed whether the mutations of LGL leukemia patients require phosphorylation of Y705 in order to convey their effects.

Study III focused on the analysis of somatic mutations in *STAT3* mutation-negative LGL leukemia patients by means of exome sequencing. This resulted in a new candidate gene, as two T-LGL leukemia patients harbored a novel gain-of-function *STAT5B* mutation Y665F in the leukemic CD8+ clone. In the following deep sequencing-based screening of *STAT5B* exon 16, two additional patients carried another activating *STAT5B* mutation N642H. Residues N642 and Y665 are situated on the surface of the SH2 domain, but their location in relation to the other STAT5B monomer in the dimerized and phosphorylated structure cannot be assessed in detail, as a molecular model does not exist. The four LGL leukemia patients (one NK cell, three T-LGL) with STAT5B mutations expressed CD56, which is a typical NK cell antigen. CD56-expressing CD8+ T cells possess more robust cytotoxic activity, which can also be MHC-unrestricted, similar to NK cells.[215] Mutated *STAT5B* could cause transcription of genes that program cells to an NK-like CD56+ phenotype. The two patients carrying a *STAT5B* N642H mutation were the only cases with untypical and fatal LGL-leukemia, indicating a connection between the mutation and aggressive phenotype. This is supported by the recent discovery of similar *STAT5B* mutations in T-ALL: in a cohort of 64 adult T-ALL patients, 7/64 (11%) carried activating *STAT5B* mutations (the N642H was harbored by 6 cases).[216]

STAT3 and STAT5B share interesting structural and functional properties: both are pleiotropic transcription factors with a crucial role in lymphocyte biology.[116] Furthermore, although their upstream activators are different and their target genes depend on the cell type, ChIP-seq based analysis of T lymphocytes showed that over 90% of the STAT5 target sequences are co-occupied with STAT3.[156] However, in some cases their action seems to reciprocal, such as in the case of Th17 differentiation through binding of STAT3 or STAT5B to Il17 gene locus.[165] The combined prevalence of *STAT3* and *STAT5B* mutations does not cover all LGL leukemia cases. The gene expression profiles of LGL leukemia patients in the microarray analysis of studies I, II, and III resembled each other, independent of *STAT3/STAT5B* mutation status, indicating that in patients with the wild type *STAT3/STAT5B,* JAK-STAT axis could be activated by alternative means. Furthermore, previous studies have shown that STAT3 is universally activated in LGL leukemia.[4] Two recent studies using exome sequencing have suggested potential candidate genes both upstream and downstream of STAT3: in the first paper, one of the mutations was located in a conserved miRNA binding site of IL6R and it could potentially alter translation of *IL6R* gene.[217] Mutations in a related gene, *IL6ST*, have been described in inflammatory hepatocellular adenoma, where *STAT3* mutations are also seen with low frequency.[218,219] In another study, the most interesting finding was a missense mutation in PTPRT, which dephosphorylates STAT3: a defect in dephosphorylation of Y705 could lead to constitutive activation of STAT3.[220] However, all the variants in these publications were seen in individual LGL leukemia patients and only *STAT3* and *STA5B* mutations have been recurrent thus far.

The diagnosis of T-LGL leukemia requires confirmation of clonal rearrangement of TCR, but the number of clones and their size varies: some patients have

monoclonal LGL leukemia, while others harbor smaller subclones in addition to the major expansion.[72] The clonal pattern can even change during follow-up, a phenomenon designated as clonal drift.[72] The data for multiple clones are in concordance with the results of study IV, in which 20% of *STAT3* mutated patients harbored multiple *STAT3* mutations, in most cases existing in different clones when analyzed by deep sequencing. The mutations were specific for CD8+ and NK cells, whereas CD4+ and CD19+ lymphocytes were unmutated in all samples. LGL leukemia patients do not commonly share immunodominant TCRB CDR3 sequences with each other, although in some cases the major clonotypes are seen in other patients.[71,221] If multiple *STAT3* mutated clones of one patient shared a similar TCRB CDR3 rearrangement, the expansions could have originally developed in response to the same antigen. Consequently, in study IV, the repertoire of the TCRB CDR3 region was analyzed using a deep sequencing method, but the clones sorted and sequenced did not share common CDR3 rearrangements with other clonal expansions within the same or different patients. Although the results did not suggest a common antigen behind clonal expansion, the possibility of one was not excluded.

One intriguing feature of T-LGL leukemia is the high incidence of concomitant autoimmune syndromes. The initial expansion and self-reactivity of cytotoxic LGLs could be related to crossreactivity between self-antigens and infectious agents. Several mechanisms causing activation of autoreactive lymphocytes have been described: in addition to crossreactivity (molecular mimicry), a tissue-destroying inflammatory reaction can lead to enhanced presentation of autoantigens by antigen-presenting cells, followed by activation of autoreactive lymphocytes.[222] Inflammation can arouse widespread immune activation, during which self-reactive lymphocyte clones also expand.[222] The connection between lymphoid malignancies and certain lymphotropic viruses has been established. As an example, rare aggressive NK-LGL leukemia is often associated with EBV.[223] HTLV-1 and -2 are retroviruses that have been connected with chemorefractory T cell lymphomas and autoimmunity.[224] Interestingly, autocrine IL-2 signaling is often reversed in HTLV-infected transformed cells, and IL-2-independent malignant lymphocytes start to express both STAT3 and STAT5 constitutively.[224] No reports of possible activating JAK/STAT axis mutations in these cells exist. HTLV antibodies have been detected in the sera of LGL leukemia patients with a prevalence of 44% (23/53), but only a few patients were actually infected with HTLV-1/2, suggesting that positive titer results from contact with an HTLV-resembling epitope in most cases.[75]

Normal control mechanisms of immunologic tolerance malfunction in LGL leukemia. The rearrangement of TCR determines the antigen target of a lymphocyte, and the positive selection of central tolerance allows only CD4+CD8+ thymocytes expressing TCR (which can interact with self-peptide-MHC complexes with moderate affinity) to differentiate into CD4+ or CD8+ lymphocytes. Central tolerance cannot eliminate all autoreactive lymphocytes, and two important mechanisms, anergy and activation-induced cell death (AICD), regulate autoreactivity in the periphery together with Tregs. Anergy is related to T cell surface receptors, which convey inhibitory signals and these include, for example, Cytotoxic T lymphocyte antigen 4 (CTLA-4) and Programmed cell death-1 (PD-1).[225] No definitive data of the role of these proteins in LGL leukemia pathogenesis exist, although one study reported that T-LGL leukemic cells were CTLA-4-negative.[226] AICD, on the other hand, has a pronounced role in LGL leukemia pathogenesis: leukemic LGLs are resistant to FAS-FASL-

signaling, which is one of the main AICD-triggering pathways.[85] Another T cell apoptosis-related pathway is mediated by the BCL-2 family proteins: in a previous publication describing STAT3 activation in LGL leukemia, the JAK2 inhibitor AG-490 had no effect on BCL-2 expression, but reduced MCL-1 levels were observed together with STAT3-inhibition and restored FAS-sensitivity.[4] Thus, mutated *STAT3* could possibly cause FAS-resistance and expansion of autoreactive cells, which are able to escape AICD.

This idea is supported by the fact that *STAT3* mutations have recently been observed in different autoreactive conditions. RA is the most common autoimmune disorder in LGL leukemia patients.[7,12,24] In studies I, II and IV, the prevalence of RA was highest in *STAT3*-mutated patients, especially in those patients with multiple mutations. Felty's syndrome is typically a late complication of RA, and *STAT3* mutations have been reported in patients with Felty's syndrome and LGL expansion, suggesting that Felty's syndrome and LGL leukemia with RA are truly overlapping conditions.[227,228] LGL leukemia can also appear in the context of certain bone marrow defects such as PRCA, aplastic anemia (AA), and MDS, and it has been suspected that hematopoietic precursors could be the target of clonal self-reactive LGLs.[229-231] *STAT3* mutations are relatively common in these syndromes, and they were detected in AA and MDS patients both with and without LGL leukemia. In patients with detectable LGL expansion, the frequency rates were 55% and 38% for AA and MDS, respectively, and the mutations were located in T cells when sorted bone marrow samples were analyzed. In patients without apparent LGL clones, the mutations were also seen in 7% (AA) and 3% (MDS) of the cases.[232] In an Asian LGL leukemia cohort, the prevalence of *STAT3* mutations was higher in LGL leukemia patients with concomitant PRCA (71% vs 5%), and *STAT3*-mutated patients more often had neutropenia and had lower treatment-free survival.[233]

Why do *STAT3,* or *STAT5B,* mutations occur with such a high rate in LGL leukemia? One possible reason is the amplification and overexpression of the IL-6-STAT3 axis caused by chronic immune activation.[149,234] This could result in an increased number of somatic mutations in this specific pathway, because transcription of a gene correlates positively with the rate of mutagenesis.[235] A hypothesis stating that LGL leukemia initiates from polyclonal immune activation and that mutations occur secondarily to activated lymphocytes is further supported by the fact that a marked proportion of study IV patients harbored multiple *STAT3* mutations, existing in multiple clones in most cases analyzed. Considering mutations in *STAT5B,* IL-15 and PDGF were able to sustain a leukemic population in a network model of LGL survival signaling.[91] These both were connected to STAT3 activation in the same model, but they are also upstream regulators of STAT5B.[91] IL-15 possesses interesting leukemogenic effects: If LGLs are cultured with IL-15 and then transplanted into mice, the mice develop malignant lymphoproliferation possibly due to genomic instability caused by prolonged IL-15 exposure. A similar mechanism could be related to the development of *STAT5B* mutations: IL-15 secretion is increased during different virus infections and autoimmune disorders, and it enhances NK cell function and stimulates the proliferation of memory T cells.[236,237]

It is likely that *STAT3*-mutated clones accumulate during the course of LGL leukemia: the clonal landscape of LGL leukemia is often dynamic, and even major immunodominant clones can disappear with simultaneous appearance of other

expansions.[72] A simplified model of clonal evolution in acute leukemias suggests that a sequential acquisition of mutations results in the development and dominance of certain subclones, but in reality this process is more complex and branched. In LGL leukemia, the clonal hierarchy is probably different, and in the case of multiple clones with or without *STAT3* mutation, they could arise from originally polyclonal immune response to an antigen. Some LGL leukemia patients even harbor both NK and T cell expansions, which was the case in one patient analyzed in study IV.[14]

The driving force of dynamic expansions could arise from the genetic background. Transcriptional activity of all different *STAT3* mutations has not been analyzed, but possibly their potential to activate STAT3 is not similar. In most cases analyzed with *STAT3* deep sequencing of sorted lymphocyte fractions, the patients with multiple mutations had one major *STAT3*-mutated clone and additional smaller mutated expansions, some of which were too small or otherwise undetectable by flow cytometry Vbeta analysis. This could be related to the activating effects of different *STAT3* mutations, but also to other factors causing changes in proliferation potency or apoptosis rate of expanded cells. First, these could be additional lymphocyte survival-affecting somatic mutations such as those variants that were seen in patient exomes sequenced in studies I and III. Second, genome-wide association studies have revealed predisposing host factors in LGL leukemia-related autoimmune disorders, which could affect the proliferation potency also in LGL leukemia. For example, single nucleotide polymorphisms (SNP) near IL6ST and IL2RA loci have been reported to associate with higher risk to develop RA.[238] The HLA-DR4 haplotype is connected to both Felty's syndrome and LGL leukemia, and a polymorphism in MICA, which is an MHC class I-related gene, is associated with higher risk of developing LGL leukemia. [32] Thirdly, immune system dysregulation could play a role: the activated IL6-STAT3 loop and suppression of SOCS3 expression are seen in LGL leukemia patients independent of the STAT3 mutation status.[4,234] Lastly, changes in the antigenic background, developing T cell anergy or exhaustion, and immunosuppressive therapy could activate or deactivate multiple clones with variable intensity.[72]

STAT3 and *STAT5B* mutations discovered in LGL leukemia are clinically relevant. STAT3 mutations can be used as molecular markers at diagnosis, along with other already established factors, and they could prove useful during follow-up of treatment responses. The prevalence of *STAT5B* mutations is lower, but they could be specific markers to the rare CD56+ T-LGL leukemia. In study IV, the screening of *STAT3* mutations in lymphocyte subpopulations showed that in the two cases analyzed, methotrexate did not eradicate the mutated clones and the partial responses seen were probably due to the MTX-caused immunosuppression of cytotoxic lymphocytes. Cyclophosphamide induced better responses, which were connected to the disappearance of mutated clones during treatment. Large prospective trials concerning the treatment of LGL leukemia have not been done, but the ORR for methotrexate is approximately 50% and relapses are common, whereas for cyclophosphamide ORR is 70% and relapses or failures are more uncommon.[12,53,55]

STAT3 activation is common in malignancies and autoimmune disorders, and consequently it presents a potential and attractive therapeutic target in multiple previously uncured disorders. LGL leukemia with activating *STAT3* mutations would be interesting platform to study the effects of different STAT3 inhibitors. Small

molecule SH2 domain inhibitors have shown promise in preclinical trials, their mechanism being the disruption of the SH2-phosphotyrosine interaction. Tested compounds such as BP-1-102 and OPB-3112 have both shown efficacy in tumor-specific growth inhibition.[239,240] DNA binding domain-targeting decoy oligonucleotides block the nuclear transportation of STAT3, but their specificity and penetrance to different tissues could be a problem.[241] *STAT3* mutations could affect the binding of small molecule SH2-inhibitors, but the small-molecule STAT3 inhibitor STA-21 suppressed the growth of two *STAT3*-mutated cell lines.[212] Similarly, STA-21 induced apoptosis of LGL leukemia patient-derived CD8+ T cells independent of the *STAT3*-mutation status, but treatment with JAK1/2-inhibitor ruxolitinib did not have remarkable effect on leukemic LGLs.[242] OPB-3112 is currently in phase I/II clinical trials, but no results have been published to date. Although the prevalence of *STAT5B* mutations was lower in LGL leukemia, STAT5 could be an appropriate therapeutic target, especially in aggressive CD56+ T/NK-LGL if N642H mutations are truly the ultimate drivers of this rare disease. Small molecule SH2-domain inhibitors are being developed in preclinical trials for treatment of STAT5-related cancer such as CML.[243,244] One interesting compound is pimozide, which is an antipsychotic drug, and already in clinical use. It induces apoptosis of *BCR-ABL1*-positive CML cells and FMS-like tyrosine kinase-3 (FLT3)-mutated AML-cells through STAT5 inhibition, but the mechanism is unknown.

The discovery of activating *STAT3* and *STAT5B* mutations provides a novel molecular basis for LGL leukemia and has provided new information about the clonal landscape and clonal evolution during the course of the disease. These observations support the theory of a common pathogenesis for T-LGL leukemia and CLPD-NK. Furthermore, the discovery of *STAT3* mutations in patients with concomitant autoimmune disorder or bone marrow failure and LGL leukemia may indicate that acquired genomic changes also underlie a proportion of autoimmune diseases, and these disorders can result from chronic antigen exposure in combination with immunogenetic factors such as *STAT* mutations. Finally, *STAT3* mutations present a novel therapeutic target and STAT3 inhibition may improve the treatment results of LGL leukemia.
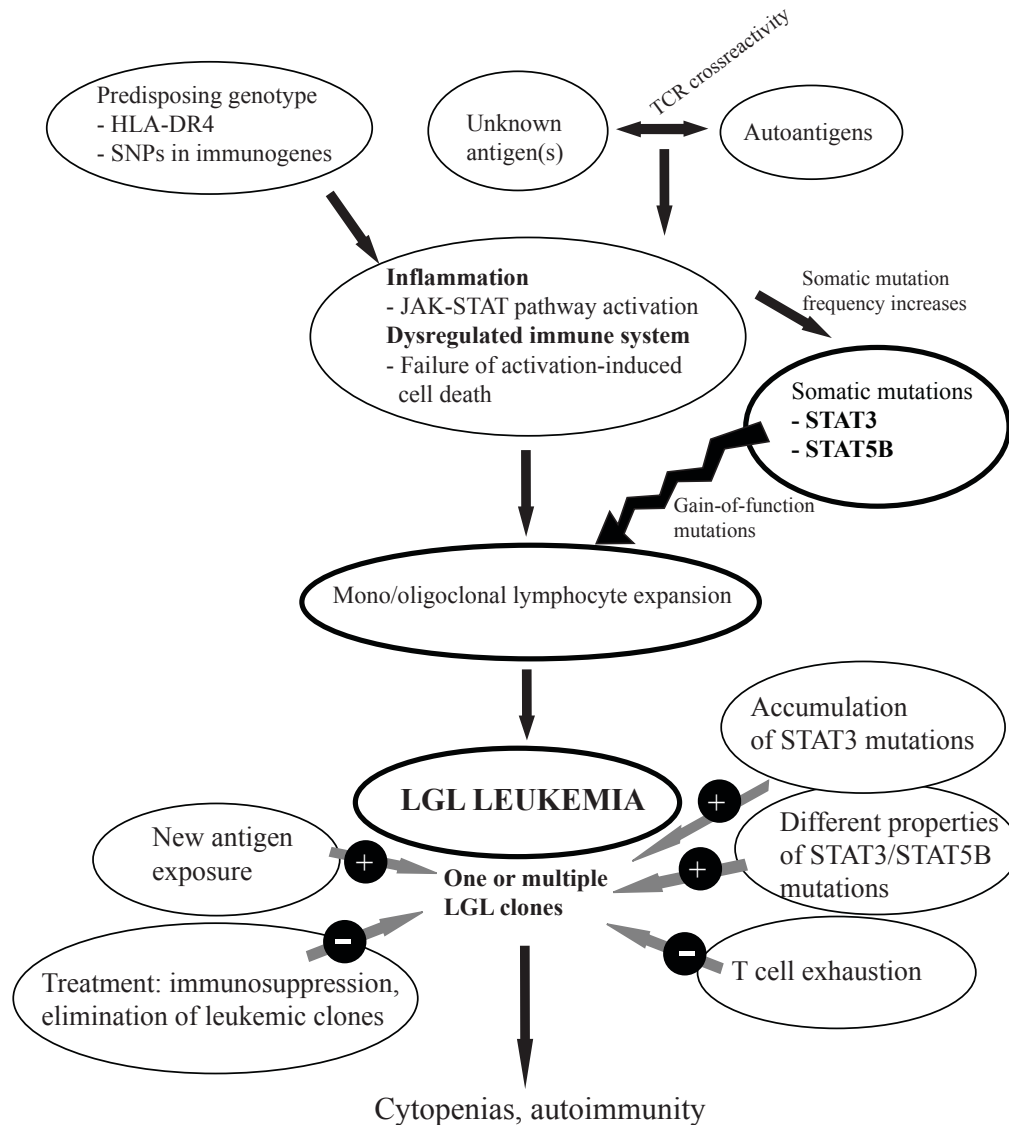
Figure 40. *Hypothesis of LGL leukemia pathogenesis. An exposure to an unknown antigen (such as microbial antigen) may lead to the initial oligoclonal expansion of cytotoxic lymphocytes, some of which could have a T cell receptor cross reacting with autoantigens. Predisposing immunogenic factors lead to the failure in restraining the immune response. Persistent autoantigen stimulation and inflammation induces the activation of JAK-STAT pathway, increasing the rate of mutagenesis in STAT genes. Gain-of-function STAT3 and STAT5B mutations give growth advantage to the mutated lymphocyte clones, and facilitate the development of LGL leukemia. A substantial proportion of LGL leukemia patients harbor multiple STAT3-mutated lymphocyte clones. The events shaping the clonal landscape include both stimulatory and inhibitory factors: STAT3 and STAT5B mutations have different functional properties, and their potential to drive leukemic LGLs is different. A new antigen exposure can give rise to new lymphocyte clones. Inhibitory factors suppressing leukemic LGLs include immunosuppressive treatment and T cell exhaustion. Depending on the antigen target of leukemic LGLs, symptoms include cytopenias and autoimmune manifestations, which are caused by the cytotoxicity of leukemic LGLs.*

CONCLUSIONS

This thesis provides new information about the molecular pathogenesis of LGL leukemia. Exome sequencing of a T-LGL leukemia patient led to the discovery of a *STAT3* SH2 domain mutation D661V, and subsequent screening revealed a prevalence of *STAT3* mutations of 40% in T-LGL leukemia and 30% in CLPD-NK patients. The mutations were activating both *in vivo* and i*n vitro.* The patients who carried *STAT3* mutations were more often neutropenic and had rheumatoid arthritis, suggesting that mutated cells could have more potent cytotoxicity. The dysregulation of anti-apoptotic and pro-survival pathways and gene expression patterns were similar in different LGL leukemia patients, independent of the mutation status, which supports the notion of universal STAT3 activation also in *STAT3*-unmutated LGL leukemia patients and suggests that other mutations in JAK-STAT3-axis may also exist and drive leukemic LGLs.

The screening of *STAT3* mutations by deep amplicon sequencing revealed multiple *STAT3* mutations in 20% of the mutated patients, and they existed mostly in different lymphocyte clones. The finding supports the model of a dynamic clonal landscape in LGL leukemia in which T cell exhaustion, different functional properties of *STAT3* mutations, other immunogenetic factors, therapies given, and dysregulated immune system all affect the clonal LGL population, giving rise to new clones while others are suppressed. The analysis of the TCRB CDR3 region by deep sequencing did not suggest a common antigen target between multiple clones of a same patient or in a comparison of the TCRB CDR3 repertoire between the T-LGL leukemia patients analyzed. However, this does not exclude the existence of a common antigen behind LGL leukemia expansion. *STAT3* mutations are seen in patients suffering from rheumatoid arthritis and different bone marrow failure syndromes, with or without concomitant LGL leukemia. One theory is that after a pathogen exposure, crossreactivity of antigen epitopes leads to the expansion of self-targeting lymphocytes, which acquire somatic mutation resulting in development of LGL leukemia with concomitant autoimmunity or bone marrow failure. Mutations in *STAT3* were also clinically relevant: complete remission during immunosuppressive therapy was connected to the eradication of leukemic *STAT3*-mutated clone.

The genetic landscape of *STAT3*-unmutated patients is still mostly undiscovered, but a small proportion of these patients carried similar activating mutations in *STAT5B* gene. One of these mutations, N642H, was connected to rare atypically aggressive CD56+ T-LGL leukemia, and could present a driver for the disease.

To conclude, activating S*TAT3* and *STAT5B* mutations are novel, highly specific molecular markers for LGL leukemia. The mutations can be used as diagnostic tools alongside the other established clinical and hematological parameters, and they also prove the clonality in the case of CLPD-NK. STAT3-targeted therapies could be used in future to treat LGL leukemia. Further studies are warranted to understand the antigen targets initiating LGL leukemia expansion, the lymphocyte population where *STAT3/STAT5B* mutations occur, the reason for the frequent loss of self-tolerance in LGL leukemia, and also the transcriptional targets of mutated *STAT3/STAT5B* and the effects they have on the signaling pathways of leukemic LGL.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Docent Satu Mustjoki and Professor Kimmo Porkka. Satu, you are the most supportive, warm, and brilliant teacher and group leader. Without your effort, comprehensive understanding of immunology and leukemia biology, and collaborative skills, the LGL leukemia project would not have been as successful as it has been. I am grateful for the increasing responsibilities you gave me during these years and for understanding the difficulties of combining clinical duties with a PhD project. Kimmo, your vast knowledge in the field of cancer biology and ability to connect basic research with clinical hematology have been essential for this thesis.

The official reviewers of this thesis, Professor Sirpa Leppä and Docent Marko Pesu, are acknowledged for the excellent comments and constructive criticism they provided to improve the manuscript. Professor Leppä and Professor Markku Heikinheimo were members of my thesis committee, and I would like to thank them for their feedback and suggestions.

Working in the enthusiastic atmosphere of the HRUH has been one of the most important aspects on the path to my PhD, and I wish to thank all the former and present group members. My fellow PhD students, Sari Hernesniemi, Anna Kreutzman, Emma Andersson, Mette Ilander, Paavo Pietarinen, Mika Kontro, Heikki Kuusanmäki, Perttu Koskenvesa, Paula Savola, and Mohamed El Missiry, have been great friends at the office, in the lab, and during social activities. Anna and Sari, I would like to thank you for your help and patience with me at the beginning of my research career, your friendship, and for sharing the best shopping tips when travelling to conferences. Emma, I am deeply grateful to you for being a great friend, making our office a better place with your optimistic energy, and for sharing your biochemical expertise and LGL leukemia projects with me. Mette, your positive attitude has cheered me up many times, and I greatly appreciate your skills and help with flow cytometry. Paavo, I would like to thank you for all the enlightening conversations we have had; I have greatly enjoyed your friendship and company. I wish to thank our postdocs, Can Hekim and Tiina Kelkka, for your help, especially on the immunological issues, and for your friendship and support during the final stages of my PhD. I would like to acknowledge senior researcher Jukka Vakkila for his support and for sharing his immunological knowledge. Minna Pajuportti and Saara Vaalas have helped me a lot with the arrangement of samples, and I thank you for sharing your knowledge about clinical studies. Hanna Lähteenmäki and Tiina Kasanen, thank you for always finding time to help me with laboratory and flow cytometry experiments. Eeva Mäkinen is acknowledged for always being kind and

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;102:169-175.

2.   Lamy T, Loughran TP, Jr. Current concepts: large granular lymphocyte leukemia. Blood Rev. 1999;13:230-240.

3.   Sokol L, Loughran TP, Jr. Large granular lymphocyte leukemia. Oncologist. 2006;11:263-273.

4.   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;107:351-362.

5.   Timonen T, Ranki A, Saksela E, Hayry P. Human natural cell-mediated cytotoxicity against fetal fibroblasts. III. Morphological and functional characterization of the effector cells. Cellular Immunology. 1979;48:121-132.

6.   McKenna RW, Parkin J, Kersey JH, Gajl-Peczalska KJ, Peterson L, Brunning RD. Chronic lymphoproliferative disorder with unusual clinical, morphologic, ultrastructural and membrane surface marker characteristics. Am J Med. 1977;62:588-596.

7.   Loughran TP, Jr. Clonal diseases of large granular lymphocytes. Blood. 1993;82:1-14.

8.   Bennett JM, Juliusson G, Mecucci C. Morphologic, immunologic, and cytogenetic classification of the chronic (mature) B and T lymphoid leukemias: Fourth meeting of the MIC Cooperative study group. Cancer Res. 1990;50:2212.

9.   Swerdlow SH, Campo E, Harris NL, et al. WHO classification of tumours of haematopoietic and lymphoid tissues. 4th ed Lyon, France: International Agency for Research on Cancer Press. 2008.

10.   O'Malley DP. T-cell large granular leukemia and related proliferations. Am J Clin Pathol. 2007;127:850-859.

11.   Mohan SR, Maciejewski JP. Diagnosis and therapy of neutropenia in large granular lymphocyte leukemia. Curr Opin Hematol. 2009;16:27-34.

12.   Bareau B, Rey J, Hamidou M, et al. Analysis of a French cohort of patients with large granular lymphocyte leukemia: a report on 229 cases. Haematologica. 2010;95:1534-1541.

13.   Timonen T, Ortaldo JR, Herberman RB. Characteristics of human large granular lymphocytes and relationship to natural killer and K cells. J Exp Med. 1981;153:569-582.

14.   Sandberg Y, Dezentje VO, Szuhai K, et al. Clonal T- and natural killer-cell large granular lymphocyte proliferations in a single patient established by array-based comparative genomic hybridization analysis. Leukemia. 2006;20:2212-2214.

15.   Morice WG, Kurtin PJ, Leibson PJ, Tefferi A, Hanson CA. Demonstration of aberrant T-cell and natural killer-cell antigen expression in all cases of granular lymphocytic leukaemia. Br J Haematol. 2003;120:1026-1036.

16.   Bourgault-Rouxel AS, Loughran TP, Jr., Zambello R, et al. Clinical spectrum of gammadelta+ T cell LGL leukemia: analysis of 20 cases. Leuk Res. 2008;32:45-48.

17.   Yang J, Epling-Burnette PK, Painter JS, et al. Antigen activation and impaired Fas-induced death-inducing signaling complex formation in T-large-granular lymphocyte leukemia. Blood. 2008;111:1610-1616.

18.   Mitsui T, Maekawa I, Yamane A, et al. Characteristic expansion of CD45RA CD27 CD28 CCR7 lymphocytes with stable natural killer (NK) receptor expression in NK- and T-cell type lymphoproliferative disease of granular lymphocytes. Br J Haematol. 2004;126:55-62.

19.   Gentile TC, Uner AH, Hutchison RE, et al. CD3+, CD56+ aggressive variant of large granular lymphocyte leukemia. Blood. 1994;84:2315-2321.

20.   van Dongen JJ, Langerak AW, Bruggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia. 2003;17:2257-2317.

21.   Langerak AW, Groenen PJ, Bruggemann M, et al. EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. Leukemia. 2012;26:2159-2171.

22.   Langerak AW, van Den Beemd R, Wolvers-Tettero IL, et al. Molecular and flow cytometric analysis of the Vbeta repertoire for clonality assessment in mature TCRalphabeta T-cell proliferations. Blood. 2001;98:165-173.

23.	Fischer L, Hummel M, Burmeister T, Schwartz S, Thiel E. Skewed expression of natural-killer (NK)-associated antigens on lymphoproliferations of large granular lymphocytes (LGL). Hematol Oncol. 2006;24:78-85.

24.	Semenzato G, Zambello R, Starkebaum G, Oshimi K, Loughran TP, Jr. The lymphoproliferative disease of granular lymphocytes: updated criteria for diagnosis. Blood. 1997;89:256-260.

25.	Liu JH, Wei S, Lamy T, et al. Chronic neutropenia mediated by fas ligand. Blood. 2000;95:3219-3222.

26.	Morice WG, Kurtin PJ, Tefferi A, Hanson CA. Distinct bone marrow findings in T-cell granular lymphocytic leukemia revealed by paraffin section immunoperoxidase stains for CD8, TIA-1, and granzyme B. Blood. 2002;99:268-274.

27.	Mailloux AW, Zhang L, Moscinski L, et al. Fibrosis and subsequent cytopenias are associated with basic fibroblast growth factor-deficient pluripotent mesenchymal stromal cells in large granular lymphocyte leukemia. J Immunol. 2013;191:3578-3593.

28.	Howard MT, Bejanyan N, Maciejewski JP, Hsi ED. T/NK large granular lymphocyte leukemia and coexisting monoclonal B-cell lymphocytosis-like proliferations. An unrecognized and frequent association. Am J Clin Pathol. 2010;133:936-941.

29.	Viny AD, Lichtin A, Pohlman B, Loughran T, Maciejewski J. Chronic B-cell dyscrasias are an important clinical feature of T-LGL leukemia. Leuk Lymphoma. 2008;49:932-938.

30.	Rawstron AC, Bennett FL, O'Connor SJ, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. N Engl J Med. 2008;359:575-583.

31.	Bockorny B, Dasanu CA. Autoimmune manifestations in large granular lymphocyte leukemia. Clin Lymphoma Myeloma Leuk. 2012;12:400-405.

32.	Liu X, Loughran TP, Jr. The spectrum of large granular lymphocyte leukemia and Felty's syndrome. Curr Opin Hematol. 2011;18:254-259.

33.	O'Keefe CL, Plasilova M, Wlodarski M, et al. Molecular analysis of TCR clonotypes in LGL: a clonal model for polyclonal responses. J Immunol. 2004;172:1960-1969.

34.	Bowman SJ, Hall MA, Panayi GS, Lanchbury JS. T cell receptor alpha-chain and beta-chain junctional region homology in clonal CD3+, CD8+ T lymphocyte expansions in Felty's syndrome. Arthritis Rheum. 1997;40:615-623.

35.	Kwong YL, Au WY, Leung AY, Tse EW. T-cell large granular lymphocyte leukemia: an Asian perspective. Ann Hematol. 2010;89:331-339.

36.	Friedman J, Schattner A, Shvidel L, Berrebi A. Characterization of T-cell large granular lymphocyte leukemia associated with Sjogren's syndrome-an important but under-recognized association. Semin Arthritis Rheum. 2006;35:306-311.

37.	Schwab R, Szabo P, Manavalan JS, et al. Expanded CD4+ and CD8+ T cell clones in elderly humans. J Immunol. 1997;158:4493-4499.

38.	Posnett DN, Sinha R, Kabak S, Russo C. Clonal populations of T cells in normal elderly humans: the T cell equivalent to "benign monoclonal gammapathy". J Exp Med. 1994;179:609-618.

39.	Smith PR, Cavenagh JD, Milne T, et al. Benign monoclonal expansion of CD8+ lymphocytes in HIV infection. J Clin Pathol. 2000;53:177-181.

40.	Gillespie GM, Wills MR, Appay V, et al. Functional heterogeneity and high frequencies of cytomegalovirus-specific CD8(+) T lymphocytes in healthy seropositive donors. J Virol. 2000;74:8140-8150.

41.	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;116:772-782.

42.	Mustjoki S, Ekblom M, Arstila TP, et al. Clonal expansion of T/NK-cells during tyrosine kinase inhibitor dasatinib therapy. Leukemia. 2009;23:1398-1405.

43.	Herling M, Khoury JD, Washington LT, Duvic M, Keating MJ, Jones D. A systematic approach to diagnosis of mature T-cell leukemias reveals heterogeneity among WHO categories. Blood. 2004;104:328-335.

44.	Cooke CB, Krenacs L, Stetler-Stevenson M, et al. Hepatosplenic T-cell lymphoma: a distinct clinicopathologic entity of cytotoxic gamma delta T-cell origin. Blood. 1996;88:4265-4274.

45.	Kwong YL. Natural killer-cell malignancies: diagnosis and treatment. Leukemia. 2005;19:2186-2194.

46.	Suzuki R, Suzumiya J, Nakamura S, et al. Aggressive natural killer-cell leukemia revisited: large granular lymphocyte leukemia of cytotoxic NK cells. Leukemia. 2004;18:763-770.

47.	Gottschalk S, Rooney CM, Heslop HE. Post-transplant lymphoproliferative disorders. Annu Rev Med. 2005;56:29-44.

48.   Paya CV, Fung JJ, Nalesnik MA, et al. Epstein-Barr virus-induced posttransplant lymphoproliferative disorders. ASTS/ASTP EBV-PTLD Task Force and The Mayo Clinic Organized International Consensus Development Meeting. Transplantation. 1999;68:1517-1525.

49.   Sneller MC, Straus SE, Jaffe ES, et al. A novel lymphoproliferative/autoimmune syndrome resembling murine lpr/gld disease. J Clin Invest. 1992;90:334-341.

50.   Rao VK, Straus SE. Causes and consequences of the autoimmune lymphoproliferative syndrome. Hematology. 2006;11:15-23.

51.   Lamy T, Loughran TP, Jr. How I treat LGL leukemia. Blood. 2011;117:2764-2774.

52.   Kremer JM, Lee JK. The safety and efficacy of the use of methotrexate in long-term therapy for rheumatoid arthritis. Arthritis Rheum. 1986;29:822-831.

53.   Loughran TP, Jr., Kidd PG, Starkebaum G. Treatment of large granular lymphocyte leukemia with oral low-dose methotrexate. Blood. 1994;84:2164-2170.

54.   Fujishima N, Sawada K, Hirokawa M, et al. Long-term responses and outcomes following immunosuppressive therapy in large granular lymphocyte leukemia-associated pure red cell aplasia: a Nationwide Cohort Study in Japan for the PRCA Collaborative Study Group. Haematologica. 2008;93:1555-1559.

55.   Moignet A, Hasanali Z, Zambello R, et al. Cyclophosphamide as a first-line therapy in LGL leukemia. Leukemia. 2013.

56.   Haas JF, Kittelmann B, Mehnert WH, et al. Risk of leukaemia in ovarian tumour and breast cancer patients following treatment by cyclophosphamide. Br J Cancer. 1987;55:213-218.

57.   Weide R, Heymanns J, Koppler H, et al. Successful treatment of neutropenia in T-LGL leukemia (T gamma-lymphocytosis) with granulocyte colony-stimulating factor. Ann Hematol. 1994;69:117-119.

58.   Osuji N, Matutes E, Tjonnfjord G, et al. T-cell large granular lymphocyte leukemia: A report on the treatment of 29 patients and a review of the literature. Cancer. 2006;107:570-578.

59.   Hellmich B, Schnabel A, Gross WL. Treatment of severe neutropenia due to Felty's syndrome or systemic lupus erythematosus with granulocyte colony-stimulating factor. Semin Arthritis Rheum. 1999;29:82-99.

60.   Monjanel H, Hourioux C, Arbion F, et al. Rapid and durable molecular response of refractory T-cell large granular lymphocyte leukemia after alemtuzumab treatment. Leuk Res. 2010;34:e197-199.

61.   Rosenblum MD, LaBelle JL, Chang CC, Margolis DA, Schauer DW, Vesole DH. Efficacy of alemtuzumab treatment for refractory T-cell large granular lymphocytic leukemia. Blood. 2004;103:1969-1971.

62.   Subbiah V, Viny AD, Rosenblatt S, Pohlman B, Lichtin A, Maciejewski JP. Outcomes of splenectomy in T-cell large granular lymphocyte leukemia with splenomegaly and cytopenia. Exp Hematol. 2008;36:1078-1083.

63.   Pandolfi F, Loughran TP, Jr., Starkebaum G, et al. Clinical course and prognosis of the lymphoproliferative disease of granular lymphocytes. A multicenter study. Cancer. 1990;65:341-348.

64.   Dhodapkar MV, Li CY, Lust JA, Tefferi A, Phyliky RL. Clinical spectrum of clonal proliferations of T-large granular lymphocytes: a T-cell clonopathy of undetermined significance? Blood. 1994;84:1620-1627.

65.   Bassing CH, Swat W, Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. Cell. 2002;109 Suppl:S45-55.

66.   Garcia KC, Teyton L, Wilson IA. Structural basis of T cell recognition. Annu Rev Immunol. 1999;17:369-397.

67.   Borg NA, Ely LK, Beddoe T, et al. The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. Nat Immunol. 2005;6:171-180.

68.   Peggs K, Verfuerth S, Pizzey A, Ainsworth J, Moss P, Mackinnon S. Characterization of human cytomegalovirus peptide-specific CD8(+) T-cell repertoire diversity following in vitro restimulation by antigen-pulsed dendritic cells. Blood. 2002;99:213-223.

69.   Bourcier KD, Lim DG, Ding YH, Smith KJ, Wucherpfennig K, Hafler DA. Conserved CDR3 regions in T-cell receptor (TCR) CD8(+) T cells that recognize the Tax11-19/HLA-A*0201 complex in a subject infected with human T-cell leukemia virus type 1: relationship of T-cell fine specificity and major histocompatibility complex/peptide/TCR crystal structure. J Virol. 2001;75:9836-9843.

70.   Matsumoto I, Tsubota K, Satake Y, et al. Common T cell receptor clonotype in lacrimal glands and labial salivary glands from patients with Sjogren's syndrome. J Clin Invest. 1996;97:1969-1977.

71.   Wlodarski MW, O'Keefe C, Howe EC, et al. Pathologic clonal cytotoxic T-cell responses: nonrandom nature of the T-cell-receptor restriction in large granular lymphocyte leukemia. Blood. 2005;106:2769-2780.

72.   Clemente MJ, Wlodarski MW, Makishima H, et al. Clonal drift demonstrates unexpected dynamics of the T-cell repertoire in T-large granular lymphocyte leukemia. Blood. 2011;118:4384-4393.

73.   Kimura H, Ito Y, Kawabe S, et al. EBV-associated T/NK-cell lymphoproliferative diseases in nonimmunocompromised hosts: prospective analysis of 108 cases. Blood. 2012;119:673-686.

74.   Saha A, Robertson ES. Epstein-Barr virus-associated B-cell lymphomas: pathogenesis and clinical outcomes. Clin Cancer Res. 2011;17:3056-3063.

75.   Thomas A, Perzova R, Abbott L, et al. LGL leukemia and HTLV. AIDS Res Hum Retroviruses. 2010;26:33-40.

76.   Loughran TP, Jr., Hadlock KG, Yang Q, et al. Seroreactivity to an envelope protein of human T-cell leukemia/lymphoma virus in patients with CD3- (natural killer) lymphoproliferative disease of granular lymphocytes. Blood. 1997;90:1977-1981.

77.   Djeu JY, Jiang K, Wei S. A view to a kill: signals triggering cytotoxicity. Clin Cancer Res. 2002;8:636-640.

78.   Vivier E, Tomasello E, Baratin M, Walzer T, Ugolini S. Functions of natural killer cells. Nat Immunol. 2008;9:503-510.

79.   Moretta L, Ferlazzo G, Bottino C, et al. Effector and regulatory events during natural killer-dendritic cell interactions. Immunol Rev. 2006;214:219-228.

80.   Shilling HG, Young N, Guethlein LA, et al. Genetic control of human NK cell repertoire. J Immunol. 2002;169:239-247.

81.   Valiante NM, Uhrberg M, Shilling HG, et al. Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. Immunity. 1997;7:739-751.

82.   Zambello R, Falco M, Della Chiesa M, et al. Expression and function of KIR and natural cytotoxicity receptors in NK-type lymphoproliferative diseases of granular lymphocytes. Blood. 2003;102:1797-1805.

83.   Epling-Burnette PK, Painter JS, Chaurasia P, et al. Dysregulated NK receptor expression in patients with lymphoproliferative disease of granular lymphocytes. Blood. 2004;103:3431-3439.

84.   Zhang J, Xu X, Liu Y. Activation-induced cell death in T cells and autoimmunity. Cell Mol Immunol. 2004;1:186-192.

85.   Lamy T, Liu JH, Landowski TH, Dalton WS, Loughran TP, Jr. Dysregulation of CD95/CD95 ligand-apoptotic pathway in CD3(+) large granular lymphocyte leukemia. Blood. 1998;92:4771-4777.

86.   Rieux-Laucat F, Le Deist F, Hivroz C, et al. Mutations in Fas associated with human lymphoproliferative syndrome and autoimmunity. Science. 1995;268:1347-1349.

87.   McCubrey JA, Steelman LS, Abrams SL, et al. Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. Adv Enzyme Regul. 2006;46:249-279.

88.   Jiang K, Zhong B, Gilvary DL, et al. Pivotal role of phosphoinositide-3 kinase in regulation of cytotoxicity in natural killer cells. Nat Immunol. 2000;1:419-425.

89.   Epling-Burnette PK, Bai F, Wei S, et al. ERK couples chronic survival of NK cells to constitutively activated Ras in lymphoproliferative disease of granular lymphocytes (LDGL). Oncogene. 2004;23:9220-9229.

90.   Schade AE, Powers JJ, Wlodarski MW, Maciejewski JP. Phosphatidylinositol-3-phosphate kinase pathway activation protects leukemic large granular lymphocytes from undergoing homeostatic apoptosis. Blood. 2006;107:4834-4840.

91.   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;105:16308-16313.

92.   Kothapalli R, Nyland SB, Kusmartseva I, Bailey RD, McKeown TM, Loughran TP, Jr. Constitutive production of proinflammatory cytokines RANTES, MIP-1beta and IL-18 characterizes LGL leukemia. Int J Oncol. 2005;26:529-535.

93.   Waldmann TA. The biology of interleukin-2 and interleukin-15: implications for cancer therapy and vaccine design. Nat Rev Immunol. 2006;6:595-601.

94.   Fontenot JD, Rasmussen JP, Gavin MA, Rudensky AY. A function for interleukin 2 in Foxp3-expressing regulatory T cells. Nat Immunol. 2005;6:1142-1151.

95.   Ku CC, Murakami M, Sakamoto A, Kappler J, Marrack P. Control of homeostasis of CD8+ memory T cells by opposing cytokines. Science. 2000;288:675-678.

96.   Malek TR. The main function of IL-2 is to promote the development of T regulatory cells. J Leukoc Biol. 2003;74:961-965.

97.   Fehniger TA, Suzuki K, Ponnappan A, et al. Fatal leukemia in interleukin 15 transgenic mice follows early expansions in natural killer and memory phenotype CD8+ T cells. J Exp Med. 2001;193:219-231.

98.   Hsu C, Jones SA, Cohen CJ, et al. Cytokine-independent growth and clonal expansion of a primary human CD8+ T-cell clone following retroviral transduction with the IL-15 gene. Blood. 2007;109:5168-5177.

99.   Yu J, Ustach C, Kim HR. Platelet-derived growth factor signaling and human cancer. J Biochem Mol Biol. 2003;36:49-59.

100.   Leblanc F, Zhang D, Liu X, Loughran TP. Large granular lymphocyte leukemia: from dysregulated pathways to therapeutic targets. Future Oncol. 2012;8:787-801.

101.   Fredriksson L, Li H, Eriksson U. The PDGF family: four gene products form five dimeric isoforms. Cytokine Growth Factor Rev. 2004;15:197-204.

102.   Zambello R, Facco M, Trentin L, et al. Interleukin-15 triggers the proliferation and cytotoxicity of granular lymphocytes in patients with lymphoproliferative disease of granular lymphocytes. Blood. 1997;89:201-211.

103.   Shah MV, Zhang R, Irby R, et al. Molecular profiling of LGL leukemia reveals role of sphingolipid signaling in survival of cytotoxic lymphocytes. Blood. 2008;112:770-781.

104.   Pyne NJ, Pyne S. Sphingosine 1-phosphate and cancer. Nat Rev Cancer. 2010;10:489-503.

105.   Stark GR, Darnell JE, Jr. The JAK-STAT pathway at twenty. Immunity. 2012;36:503-514.

106.   Laurence A, Pesu M, Silvennoinen O, O'Shea J. JAK Kinases in Health and Disease: An Update. Open Rheumatol J. 2012;6:232-244.

107.   Ungureanu D, Wu J, Pekkala T, et al. The pseudokinase domain of JAK2 is a dual-specificity protein kinase that negatively regulates cytokine signaling. Nat Struct Mol Biol. 2011;18:971-976.

108.   Maritano D, Sugrue ML, Tininini S, et al. The STAT3 isoforms alpha and beta have unique and specific functions. Nat Immunol. 2004;5:401-409.

109.   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;285:311-318.

110.   Ramos HL, O'Shea JJ, Watford WT. STAT5 isoforms: controversies and clarifications. Biochem J. 2007;404:e1-2.

111.   Schuster B, Hendry L, Byers H, Lynham SF, Ward MA, John S. Purification and identification of the STAT5 protease in myeloid cells. Biochem J. 2007;404:81-87.

112.   Becker S, Groner B, Muller CW. Three-dimensional structure of the Stat3beta homodimer bound to DNA. Nature. 1998;394:145-151.

113.   Bromberg JF, Wrzeszczynska MH, Devgan G, et al. Stat3 as an oncogene. Cell. 1999;98:295-303.

114.   Neculai D, Neculai AM, Verrier S, et al. Structure of the unphosphorylated STAT5a dimer. J Biol Chem. 2005;280:40782-40787.

115.   Sasse J, Hemmann U, Schwartz C, et al. Mutational analysis of acute-phase response factor/Stat3 activation and dimerization. Mol Cell Biol. 1997;17:4677-4686.

116.   Schindler C, Levy DE, Decker T. JAK-STAT signaling: from interferons to cytokines. J Biol Chem. 2007;282:20059-20063.

117.   Hibi M, Murakami M, Saito M, Hirano T, Taga T, Kishimoto T. Molecular cloning and expression of an IL-6 signal transducer, gp130. Cell. 1990;63:1149-1157.

118.   Lutticken C, Wegenka UM, Yuan J, et al. Association of transcription factor APRF and protein kinase Jak1 with the interleukin-6 signal transducer gp130. Science. 1994;263:89-92.

119.   Mihara M, Hashizume M, Yoshida H, Suzuki M, Shiina M. IL-6/IL-6 receptor system and its role in physiological and pathological conditions. Clin Sci (Lond). 2012;122:143-159.

120.   Kisseleva T, Bhattacharya S, Braunstein J, Schindler CW. Signaling through the JAK/STAT pathway, recent advances and future challenges. Gene. 2002;285:1-24.

121.   Shuai K, Horvath CM, Huang LH, Qureshi SA, Cowburn D, Darnell JE, Jr. Interferon activation of the transcription factor Stat91 involves dimerization through SH2-phosphotyrosyl peptide interactions. Cell. 1994;76:821-828.

122.   Braunstein J, Brutsaert S, Olson R, Schindler C. STATs dimerize in the absence of phosphorylation. J Biol Chem. 2003;278:34133-34140.

123.   Wen Z, Darnell JE, Jr. Mapping of Stat3 serine phosphorylation to a single residue (727) and evidence that serine phosphorylation has no influence on DNA binding of Stat1 and Stat3. Nucleic Acids Res. 1997;25:2062-2067.

124.   Cimica V, Chen HC, Iyer JK, Reich NC. Dynamics of the STAT3 transcription factor: nuclear import dependent on Ran and importin-beta1. PloS one. 2011;6:e20188.

125.   Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. Nat Rev Cancer. 2009;9:798-809.

126.   Snyder M, Huang XY, Zhang JJ. Identification of novel direct Stat3 target genes for control of growth and differentiation. J Biol Chem. 2008;283:3791-3798.

127.   Yang J, Liao X, Agarwal MK, Barnes L, Auron PE, Stark GR. Unphosphorylated STAT3 accumulates in response to IL-6 and activates transcription by binding to NFkappaB. Genes Dev. 2007;21:1396-1408.

128.   Rebe C, Vegran F, Berger H, Ghiringhelli F. STAT3 activation: A key factor in tumor immunoescape. JAKSTAT. 2013;2:e23010.

129.   Alexander WS, Hilton DJ. The role of suppressors of cytokine signaling (SOCS) proteins in regulation of the immune response. Annu Rev Immunol. 2004;22:503-529.

130.   Shuai K, Liu B. Regulation of gene-activation pathways by PIAS proteins in the immune system. Nat Rev Immunol. 2005;5:593-605.

131.   Zhang X, Guo A, Yu J, et al. Identification of STAT3 as a substrate of receptor protein tyrosine phosphatase T. Proc Natl Acad Sci U S A. 2007;104:4060-4064.

132.   Takeda K, Noguchi K, Shi W, et al. Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. Proc Natl Acad Sci U S A. 1997;94:3801-3804.

133.   Akira S. Roles of STAT3 defined by tissue-specific gene targeting. Oncogene. 2000;19:2607-2611.

134.   Cui W, Liu Y, Weinstein JS, Craft J, Kaech SM. An interleukin-21-interleukin-10-STAT3 pathway is critical for functional maturation of memory CD8+ T cells. Immunity. 2011;35:792-805.

135.   Yang XO, Panopoulos AD, Nurieva R, et al. STAT3 regulates cytokine-mediated generation of inflammatory helper T cells. J Biol Chem. 2007;282:9358-9363.

136.   Fornek JL, Tygrett LT, Waldschmidt TJ, Poli V, Rickert RC, Kansas GS. Critical role for Stat3 in T-dependent terminal differentiation of IgG B cells. Blood. 2006;107:1085-1091.

137.   Chou WC, Levy DE, Lee CK. STAT3 positively regulates an early step in B-cell development. Blood. 2006;108:3005-3011.

138.   Alonzi T, Maritano D, Gorgoni B, Rizzuto G, Libert C, Poli V. Essential role of STAT3 in the control of the acute-phase response as revealed by inducible gene inactivation [correction of activation] in the liver. Mol Cell Biol. 2001;21:1621-1632.

139.   Holland SM, DeLeo FR, Elloumi HZ, et al. STAT3 mutations in the hyper-IgE syndrome. N Engl J Med. 2007;357:1608-1619.

140.   Minegishi Y, Saito M, Tsuchiya S, et al. Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. Nature. 2007;448:1058-1062.

141.   Siegel AM, Heimall J, Freeman AF, et al. A critical role for STAT3 transcription factor signaling in the development and maintenance of human T cell memory. Immunity. 2011;35:806-818.

142.   Harris TJ, Grosso JF, Yen HR, et al. Cutting edge: An in vivo requirement for STAT3 signaling in TH17 development and TH17-dependent autoimmunity. J Immunol. 2007;179:4313-4317.

143.   Gaston JS. Cytokines in arthritis--the 'big numbers' move centre stage. Rheumatology (Oxford). 2008;47:8-12.

144.   Fitch E, Harper E, Skorcheva I, Kurtz SE, Blauvelt A. Pathophysiology of psoriasis: recent advances on IL-23 and Th17 cytokines. Curr Rheumatol Rep. 2007;9:461-467.

145.   Kebir H, Kreymborg K, Ifergan I, et al. Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. Nat Med. 2007;13:1173-1175.

146.   Mori T, Miyamoto T, Yoshida H, et al. IL-1beta and TNFalpha-initiated IL-6-STAT3 pathway is critical in mediating inflammatory cytokines and RANKL expression in inflammatory arthritis. Int Immunol. 2011;23:701-712.

147.   Cavarretta IT, Neuwirt H, Untergasser G, et al. The antiapoptotic effect of IL-6 autocrine loop in a cellular model of advanced prostate cancer is mediated by Mcl-1. Oncogene. 2007;26:2822-2832.

148.   Yu H, Jove R. The STATs of cancer--new molecular targets come of age. Nat Rev Cancer. 2004;4:97-105.

149.   Hodge DR, Hurt EM, Farrar WL. The role of IL-6 and STAT3 in inflammation and cancer. Eur J Cancer. 2005;41:2502-2512.

150.   Benekli M, Xia Z, Donohue KA, et al. Constitutive activity of signal transducer and activator of transcription 3 protein in acute myeloid leukemia blasts is associated with short disease-free survival. Blood. 2002;99:252-257.

151.   Gough DJ, Corlett A, Schlessinger K, Wegrzyn J, Larner AC, Levy DE. Mitochondrial STAT3 supports Ras-dependent oncogenic transformation. Science. 2009;324:1713-1716.

152.   Miyoshi K, Cui Y, Riedlinger G, et al. Structure of the mouse Stat 3/5 locus: evolution from Drosophila to zebrafish to mouse. Genomics. 2001;71:150-155.

153.   Zeng R, Aoki Y, Yoshida M, Arai K, Watanabe S. Stat5B shuttles between cytoplasm and nucleus in a cytokine-dependent and -independent manner. J Immunol. 2002;168:4567-4575.

154.   Soldaini E, John S, Moro S, Bollenbacher J, Schindler U, Leonard WJ. DNA binding site selection of dimeric and tetrameric Stat5 proteins reveals a large repertoire of divergent tetrameric Stat5a binding sites. Mol Cell Biol. 2000;20:389-401.

155.   Kanai T, Seki S, Jenks JA, et al. Identification of STAT5A and STAT5B Target Genes in Human T Cells. PloS one. 2014;9:e86790.

156.   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.

157.   Kelly J, Spolski R, Imada K, Bollenbacher J, Lee S, Leonard WJ. A role for Stat5 in CD8+ T cell homeostasis. J Immunol. 2003;170:210-217.

158.   Teglund S, McKay C, Schuetz E, et al. Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. Cell. 1998;93:841-850.

159.   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;103:1000-1005.

160.   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;11:179-186.

161.   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;94:7239-7244.

162.   Feldman GM, Rosenthal LA, Liu X, et al. STAT5A-deficient mice demonstrate a defect in granulocyte-macrophage colony-stimulating factor-induced proliferation and gene expression. Blood. 1997;90:1768-1776.

163.   Nakajima H, Liu XW, Wynshaw-Boris A, et al. An indirect effect of Stat5a in IL-2-induced proliferation: a critical role for Stat5a in IL-2-mediated IL-2 receptor alpha chain induction. Immunity. 1997;7:691-701.

164.   Imada K, Bloom ET, Nakajima H, et al. Stat5b is essential for natural killer cell-mediated proliferation and cytolytic activity. J Exp Med. 1998;188:2067-2074.

165.   Yang XP, Ghoreschi K, Steward-Tharp SM, et al. Opposing regulation of the locus encoding IL-17 through direct, reciprocal actions of STAT3 and STAT5. Nat Immunol. 2011;12:247-254.

166.   Hwa V, Camacho-Hubner 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;68:218-224.

167.   Nadeau K, Hwa V, Rosenfeld RG. STAT5b deficiency: an unsuspected cause of growth failure, immunodeficiency, and severe pulmonary disease. J Pediatr. 2011;158:701-708.

168.   Kanai T, Jenks J, Nadeau KC. The STAT5b Pathway Defect and Autoimmunity. Front Immunol. 2012;3:234.

169.   Shuai K, Halpern J, ten Hoeve J, Rao X, Sawyers CL. Constitutive activation of STAT5 by the BCR-ABL oncogene in chronic myelogenous leukemia. Oncogene. 1996;13:247-254.

170.   Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell. 2005;7:387-397.

171.   Gesbert F, Griffin JD. Bcr/Abl activates transcription of the Bcl-X gene through STAT5. Blood. 2000;96:2269-2276.

172.   Hoelbl A, Kovacic B, Kerenyi MA, et al. Clarifying the role of Stat5 in lymphoid development and Abelson-induced transformation. Blood. 2006;107:4898-4906.

173.   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;117:3421-3429.

174.   Funakoshi-Tago M, Tago K, Abe M, Sonoda Y, Kasahara T. STAT5 activation is critical for the transformation mediated by myeloproliferative disorder-associated JAK2 V617F mutant. J Biol Chem. 2010;285:5296-5307.

175.   Mullighan CG, Zhang J, Harvey RC, et al. JAK mutations in high-risk childhood acute lymphoblastic leukemia. Proc Natl Acad Sci U S A. 2009;106:9414-9418.

176.   Jeong EG, Kim MS, Nam HK, et al. Somatic mutations of JAK1 and JAK3 in acute leukemias and solid cancers. Clin Cancer Res. 2008;14:3716-3721.

177.   Onishi M, Nosaka T, Misawa K, et al. Identification and characterization of a constitutively active STAT5 mutant that promotes cell proliferation. Mol Cell Biol. 1998;18:3871-3879.

178.   Ariyoshi K, Nosaka T, Yamada K, et al. Constitutive activation of STAT5 by a point mutation in the SH2 domain. J Biol Chem. 2000;275:24407-24413.

179.   Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 2004;10:789-799.

180.   Knudson AG. Cancer genetics. Am J Med Genet. 2002;111:96-102.

181.   Foulkes WD. Inherited susceptibility to common cancers. N Engl J Med. 2008;359:2143-2153.

182.   Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150:12-27.

183.   Consortium IHGS. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931-945.

184.   Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11:31-46.

185.   Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. Nat Methods. 2007;4:903-905.

186.   Hodges E, Xuan Z, Balija V, et al. Genome-wide in situ exon capture for selective resequencing. Nat Genet. 2007;39:1522-1527.

187.   Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res. 2006;34:e22.

188.   Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010;38:1767-1771.

189.   Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. Nat Biotechnol. 2009;27:455-457.

190.   Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-1760.

191.   Allen VJ, Methven L, Gosney MA. Use of nutritional complete supplements in older adults with dementia: systematic review and meta-analysis of clinical outcomes. Clin Nutr. 2013;32:950-957.

192.   Koboldt DC, Chen K, Wylie T, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25:2283-2285.

193.   Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24-26.

194.   Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568-576.

195.   Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

196.   Schweiger MR, Kerick M, Timmermann B, et al. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PloS One. 2009;4:e5548.

197.   Taub MA, Corrada Bravo H, Irizarry RA. Overcoming bias and systematic errors in next generation sequencing data. Genome Med. 2010;2:87.

198.   Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008;36:e105.

199.   Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011;12:443-451.

200.   Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biology. 2011;12:R112.

201.   Forbes SA, Bhamra G, Bamford S, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet. 2008;Chapter 10:Unit 10 11.

202.   Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248-249.

203.   Ye S, Dhillon S, Ke X, Collins AR, Day IN. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res. 2001;29:E88-88.

204.   Sato K, Kawashima S. Calpain function in the modulation of signal transduction molecules. Biol Chem. 2001;382:743-751.

205.   Abate-Shen C. Deregulated homeobox gene expression in cancer: cause or consequence? Nat Rev Cancer. 2002;2:777-785.

206.   Greaves DR, Gordon S. The macrophage scavenger receptor at 30 years of age: current knowledge and future challenges. J Lipid Res. 2009;50 Suppl:S282-286.

207.   Liao D. Emerging roles of the EBF family of transcription factors in tumor suppression. Mol Cancer Res. 2009;7:1893-1901.

208.   Simons K, Toomre D. Lipid rafts and signal transduction. Nat Rev Mol Cell Biol. 2000;1:31-39.

209.   Barcelo-Coblijn G, Martin ML, de Almeida RF, et al. Sphingomyelin and sphingomyelin synthase (SMS) in the malignant transformation of glioma cells and in 2-hydroxyoleic acid therapy. Proc Natl Acad Sci U S A. 2011;108:19569-19574.

210.   Fasan A, Kern W, Grossmann V, Haferlach C, Haferlach T, Schnittger S. STAT3 mutations are highly specific for large granular lymphocytic leukemia. Leukemia. 2013;27:1598-1600.

211.   Ohgami RS, Ma L, Merker JD, Martinez B, Zehnder JL, Arber DA. STAT3 mutations are frequent in CD30+ T-cell lymphomas and T-cell large granular lymphocytic leukemia. Leukemia. 2013;27:2244-2247.

212.   Couronne L, Scourzic L, Pilati C, et al. STAT3 mutations identified in human hematologic neoplasms induce myeloid malignancies in a mouse bone marrow transplantation model. Haematologica. 2013;98:1748-1752.

213.   Hu G, Witzig TE, Gupta M. A Novel Missense (M206K) STAT3 Mutation in Diffuse Large B Cell Lymphoma Deregulates STAT3 Signaling. PloS One. 2013;8:e67851.

214.   Kim MS, Lee SH, Yoo NJ, Lee SH. STAT3 exon 21 mutation is rare in common human cancers. Acta Oncol. 2013;52:1221-1222.

215.   Ohkawa T, Seki S, Dobashi H, et al. Systematic characterization of human CD8+ T cells with natural killer cell markers in comparison with natural killer cells and normal CD8+ T cells. Immunology. 2001;103:281-290.

216.   Kontro M, Kuusanmaki H, Eldfors S, et al. Novel activating STAT5B mutations as putative drivers of T-cell acute lymphoblastic leukemia. Leukemia. 2014.

217.   Bedoya-Reina OC, Ratan A, Burhans R, et al. Galaxy tools to study genome diversity. Gigascience. 2013;2:17.

218.   Pilati C, Amessou M, Bihl MP, et al. Somatic mutations activating STAT3 in human inflammatory hepatocellular adenomas. J Exp Med. 2011;208:1359-1366.

219.   Rebouissou S, Amessou M, Couchy G, et al. Frequent in-frame somatic deletions activate gp130 in inflammatory hepatocellular tumours. Nature. 2009;457:200-204.

220.   Andersson EI, Rajala HL, Eldfors S, et al. Novel somatic mutations in large granular lymphocytic leukemia affecting the STAT-pathway and T-cell activation. Blood Cancer J. 2013;3:e168.

221.   Clemente MJ, Przychodzen B, Jerez A, et al. Deep sequencing of the T cell receptor repertoire in CD8+ T-large granular lymphocyte leukemia identifies signature landscapes. Blood. 2013;122:4077-4085.

222.   Wucherpfennig KW. Mechanisms for the induction of autoimmunity by infectious agents. J Clin Invest. 2001;108:1097-1104.

223.   Hart DN, Baker BW, Inglis MJ, et al. Epstein-Barr viral DNA in acute large granular lymphocyte (natural killer) leukemic cells. Blood. 1992;79:2116-2123.

224.   Poiesz BJ, Poiesz MJ, Choi D. The human T-cell lymphoma/leukemia viruses. Cancer Invest. 2003;21:253-277.

225.   Fife BT, Bluestone JA. Control of peripheral T-cell tolerance and autoimmunity via the CTLA-4 and PD-1 pathways. Immunol Rev. 2008;224:166-182.

226.   Pistillo MP, Tazzari PL, Palmisano GL, et al. CTLA-4 is not restricted to the lymphoid cell lineage and can function as a target molecule for apoptosis induction of leukemic cells. Blood. 2003;101:202-209.

227.   Schrenk KG, Krokowski M, Feller AC, et al. Clonal T-LGL population mimicking leukemia in Felty's syndrome--part of a continuous spectrum of T-LGL proliferations? Ann Hematol. 2013;92:985-987.

228.   Loughran TP, Jr., Starkebaum G, Kidd P, Neiman P. Clonal proliferation of large granular lymphocytes in rheumatoid arthritis. Arthritis Rheum. 1988;31:31-36.

229.   Go RS, Li CY, Tefferi A, Phyliky RL. Acquired pure red cell aplasia associated with lymphoproliferative disease of granular T lymphocytes. Blood. 2001;98:483-485.

230.   Wlodarski MW, Gondek LP, Nearman ZP, et al. Molecular strategies for detection and quantitation of clonal cytotoxic T-cell responses in aplastic anemia and myelodysplastic syndrome. Blood. 2006;108:2632-2641.

231.   Risitano AM, Maciejewski JP, Green S, Plasilova M, Zeng W, Young NS. In-vivo dominant immune responses in aplastic anaemia: molecular tracking of putatively pathogenetic T-cell clones by TCR beta-CDR3 sequencing. Lancet. 2004;364:355-364.

232.   Jerez A, Clemente MJ, Makishima H, et al. STAT3-mutations indicate the presence of subclinical T cell clones in a subset of aplastic anemia and myelodysplastic syndrome patients. Blood. 2013;122:2453-2459.

233.   Qiu ZY, Fan L, Wang L, et al. STAT3 mutations are frequent in T-cell large granular lymphocytic leukemia with pure red cell aplasia. J Hematol Oncol. 2013;6:82.
234.   Teramo A, Gattazzo C, Passeri F, et al. Intrinsic and extrinsic mechanisms contribute to maintain the JAK/STAT pathway aberrantly activated in T-type large granular lymphocyte leukemia. Blood. 2013;121:3843-3854, S3841.
235.   Park C, Qian W, Zhang J. Genomic evidence for elevated mutation rates in highly expressed genes. EMBO Rep. 2012;13:1123-1129.
236.   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;22:645-655.
237.   Fehniger TA, Caligiuri MA. Interleukin 15: biology and relevance to human disease. Blood. 2001;97:14-32.
238.   Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010;42:508-514.
239.   Zhang X, Yue P, Page BD, et al. Orally bioavailable small-molecule inhibitor of transcription factor Stat3 regresses human breast and lung cancer xenografts. Proc Natl Acad Sci U S A. 2012;109:9623-9628.
240.   Hayakawa F, Sugimoto K, Harada Y, et al. A novel STAT inhibitor, OPB-31121, has a significant antitumor effect on leukemia with STAT-addictive oncokinases. Blood Cancer J. 2013;3:e166.
241.   Souissi I, Najjar I, Ah-Koon L, et al. A STAT3-decoy oligonucleotide induces cell death in a human colorectal carcinoma cell line by blocking nuclear transfer of STAT3 and STAT3-bound NF-kappaB. BMC Cell Biol. 2011;12:14.
242.   Clemente MJ, Bilori B, Thota S, et al. Inhibition of the JAK/STAT signaling pathway as a novel therapy in CD8+ T-large granular lymphocytic leukemia. Blood (2013 ASH Annual Meeting and Exposition). 2013;122:3902.
243.   Page BD, Fletcher S, Yue P, et al. Identification of a non-phosphorylated, cell permeable, small molecule ligand for the Stat3 SH2 domain. Bioorg Med Chem Lett. 2011;21:5605-5609.
244.   Miklossy G, Hilliard TS, Turkson J. Therapeutic modulators of STAT signalling for human diseases. Nat Rev Drug Discov. 2013;12:611-629.