UNIVERSITÀ DEL PIEMONTE ORIENTALE

School of Medicine

Department of Health Sciences

Master's degree in medical biotechnologies

Prognostic impact of the mutational status of the immunoglobulin light chain genes in chronic lymphocytic leukemia

Tutor: **Chiar.mo Prof. Gianluca GAIDANO**

Firmato digitalmente da Gianluca Gaidano
Data: 26.06.2024 10:11:58 CEST
Organizzazione: UNIVERSITA' DEGLI STUDI DEL PIEMONTE ORIENTALE/01943490027

Candidate: **Bashar AL DEEBAN**

Matricula Number: **20053530**

Academic Year 2023/2024

TABLE OF CONTENTS

TABLES

FIGURES

SUMMARY

The B-cell receptor (BCR) plays a pivotal role in chronic lymphocytic leukemia (CLL) pathogenesis. Between the two components of the BCR variable region, the Immunoglobulin heavy chain (IGHV) has been extensively studied and its mutational status not only serves as a robust prognostication tool but also represents a predictive biomarker for therapeutic choices. In contrast, the clinical relevance of the light chain gene recombination and mutational status in CLL is still largely unexplored. This study aimed to elucidate the prognostic role of light chain genes and their mutational status. A real-life training-validation approach was used in this study. The light chain repertoire was analyzed by Sanger and by NGS sequencing methods. In a training cohort of 573 CLL patients (median follow-up 11.6 years), A total of 530 productive rearrangements were identified, in which the most frequently rearranged kappa gene was IGKV4-1, and among lambda rearrangements, the most frequent one was IGLV3-21, out of which 59.4% harbored the R110 somatic point mutation. A recursive partitioning approach identifies 99.0% of homology as the best cut-off that maximizes the log-rank statistics for TTFT[1] in 414 Binet A CLL. Unmutated (UM) light chain patients (homology ≥99.0%) associated with shorter TTFT, with a 10-year probability of 32.4% compared to 73.2% for mutated (homology <99.0%) light chain patients. A multivariate analysis adjusted for the (IPS-E)[2] variables revealed that both UM light chains and UM-IGHV maintained an independent association with shorter TTFT. The prognostic role of light chain mutational status in terms of TTFT using the established homology cut-off of 99.0% was validated in an independent cohort that comprised 299 Rai 0 CLL patients. In addition, by combining the two cohorts (673 early stage CLL), UM light chains independently predicted shorter TTFT in a multivariate analysis when adjusted for the IPS-E variables. In conclusion, this study represents the largest real-world cohort of unselected CLL analyzed for the immunoglobulin light chain gene repertoire. The mutational status of light chain genes refines the clinical impact of IGHV mutational status and independently predicts shorter TTFT in early stage CLL.

[1] Time to First Treatment
[2] International Prognostic Score for Early-stage CLL

1. INTRODUCTION

1.1. Chronic Lymphocytic Leukemia

Chronic lymphocytic leukemia (CLL) or small lymphocytic lymphoma (SLL) is a slow-progressing cancer marked by an overproduction of mature but malfunctioning B lymphocytes. It is identified as a monoclonal lymphoproliferative disorder, involving the proliferation and accumulation of mature-looking but immunologically ineffective B-cell lymphocytes, often seen as smudge cells on a peripheral blood smear. The primary areas affected by the disease include the peripheral blood, spleen, lymph nodes, and bone marrow. Pathologically and immunophenotypically, CLL and SLL are identical, both originating from B-cell lymphocytes. The distinction between the two lies in their presentation: CLL typically refers to the leukemic phase where abnormal cells are found in the blood, while SLL refers to the lymphoma phase where these cells are located in the lymph nodes. SLL is generally used to describe the lymphoproliferative process confined to the lymph nodes [1].

Chronic lymphocytic leukemia (CLL) primarily affects older adults. Due to its often-slow progression and the long survival of many patients, CLL has a high prevalence and is the most common leukemia in adults in Western countries. The disease arises from the overgrowth of a single CD5+ B lymphocyte that co-expresses low levels of surface membrane immunoglobulin (smIg) of a single IG light (L) chain type, as well as CD79B, CD20, and CD23. The clinical outcomes of this clonal overgrowth are highly variable: some patients may die within 2-3 years of diagnosis, while others may live for several decades. This variability is influenced by intrinsic factors related to the leukemic B cell, such as genetic and epigenetic alterations in coding and noncoding genes, as well as extrinsic factors, including signals from the tissue microenvironment [2].

1.2. Pathogenesis

The most frequently observed genetic abnormalities in mature CLL cells include del13q, trisomy 12, del11q, and del17p, in that sequence [3-5]. Deletions in 13q and trisomy 12 are often detected at diagnosis, indicating these are early events in the disease process. In contrast, deletions in 11q and 17p usually emerge later, suggesting their role in disease

progression and clonal evolution. The critical genetic factors in del13q are the microRNAs miR-15a and miR-16b [6]. Loss of miR15a/16b leads to overproduction of the anti-apoptotic protein Bcl-2[7], and heightened cell-cycle progression[8], both oncogene-like effects. Moreover, mice with a defective miR15/16 spontaneously develop a CLL-like disease [9], and deletion of the chromosomal region corresponding to 13q in mice leads to murine CLL [8, 10].

The mutational status of IGHV genes plays a pivotal role in the biological and clinical profile of CLL. Mutated CLL M-CLL shows a similarity of the IGHV genes less than 98% when compared to the corresponding germline IGHV gene counterpart, and this is due to the somatic hypermutation (SHM) process that takes place in the germinal center (GC) in the lymphatic node, which generally has an indolent disease course[11]. On the other hand, the unmutated CLL U-CLL harbours IGHV gene that is similar to the germline equal or more than 98%, which might suggest the devoid of SHM process in the GC and displays a more aggressive disease[12, 13]. It was emphasized that the BCR is substantial in the molecular pathogenesis of CLL, whether due to its biased usage of IGHV genes (and hypothetically the light chain genes too) and for its role in survival and growth signals in malignant B cells[14, 15]. These milestones not only transparentized the blurring vision of CLL pathogenesis, but also served as a well-established platform to develop new drugs (BTKi) added to CLL therapeutic armamentarium [14].

More disease-driving aberrations are found in IGHV-unmutated CLL (U-CLL) than in IGHV-mutated CLL (M-CLL). Additionally, certain mutations occur more frequently in U-CLL (*NOTCH1*, *XPO1*, and *POT1*) and others in M-CLL (del(13q), MyD88, and CHD2), or in CLL subsets with specific genomic aberrations (e.g., MyD88 with del13q; *SF3B1* with del11q; *NOTCH1*, *BIRC3*, *FBXW7*, and *BCOR* with tri12. Moreover, sets of mutations associate with the use of certain stereotyped IGHV-IGHD-IGHJ gene rearrangements [16, 17]. Selected mutations occur in specific regions of individual genes (e.g., PEST domain of *NOTCH1* and HEAT domain of *SF3B1*). In some instances, a relationship exists between a specific abnormality and aggressive clinical course and/or shortened survival [18]; this might be caused by development of Richter transformation [18].

1.3. Clinical staging systems

Two staging systems are currently applied to CLL patients to define disease burden and treatment indication: Rai [19] and Binet [20] staging systems. Clinical staging using the Binet and Rai classification systems provides a simple and inexpensive approach to assess prognosis in CLL [21]. Rai staging system was the first used in the clinical management of CLL. The system includes lymphadenopathy, organomegaly and cytopenia (anemia and thrombocytopenia) to establish five prognostic groups that can be used to predict median survival of newly diagnosed CLL patients.

The modified Rai classification defines low-risk disease as occurring in patients who have lymphocytosis with leukemic cells in the blood and/or bone marrow (formerly considered Rai stage 0) [22]. Patients with peripheral blood lymphocytosis, enlarged lymph nodes in any site, and splenomegaly and/or hepatomegaly (palpable lymph nodes or not) are defined as having intermediate-risk disease (formerly considered Rai stage I or II). High-risk disease that includes patients with disease-related anemia (hemoglobin [Hb] level < 1 g/dL) are formerly defined as stage III or thrombocytopenia (platelet count of <100 × 109/L) are formerly defined as stage IV [22]. Conversely, in the Binet staging system, published in 1981, patients are classified according to the number of affected superficial tissue regions (cervical, axillary, and inguinal lymph nodes), spleen and liver as well as the presence or not of anemia and thrombocytopenia.

In 2016, a new prognostic model called the CLL International Prognostic Index (CLL-IPI) was released. Five independent prognostic biomarkers were identified and specific points were attributed to them: TP53 status (no abnormalities vs del17p, TP53 mutations, or both), IGHV mutational status (mutated vs unmutated), serum beta-2 microglobulin (β2M) concentration (<3.5 mg/L vs > 3.5 mg/L), clinical stage (Binet A or Rai 0 vs. Binet B-C or Rai I-IV), age (<65 years vs > 65 years) [23].

1.4. IPS-E (international prognostic score – early stage)

Asymptomatic early-stage CLL is a challenging condition. Patients experience different outcomes varying from a stable or a slowly progressive disease not requiring treatment for decades to a disease that requires treatment soon after diagnosis because of the development of cytopenia or lymphadenopathy. The International Prognostic Scoring

system for early stage CLL (IPS-E) is a robust model that predicts the likelihood of disease progression at the time of diagnosis, and hence the need for therapy. IGHV mutation status is one of the calculated variables in IPS-E. since it never changes during the course of the disease, the model suggests that it should be evaluated at diagnosis to estimate time to first treatment and assist the physician into making the correct decision [24].

1.5. B cell receptor structure and activation

The structures of the BCR complex are composed of membrane forms of antibodies bound to two transmembrane proteins, Igα (CD79A) and Igβ (CD79B) [25, 26]. These structures are expected to guide the creation of highly effective therapies for eradicating cancers and managing and preventing autoimmune diseases. Previous research revealed that the membrane antibody of the BCR has the characteristic Y shape, featuring two fragment antigen-binding (Fab) arms connected via hinges to a single fragment crystallizable (Fc) leg, which is linked to a transmembrane domain. Igα and Igβ escort the membrane antibody to the B cell surface and are crucial for initiating biochemical signaling within the B cell upon antigen binding [27, 28]. The key interactions anchoring the membrane antibody to Igα and Igβ are embedded within the cell membrane [29], making the BCR difficult to study structurally through crystallography.

The B cell receptor (BCR) is a complex system that regulates most B cell functions. When conditions are suitable, antigen binding induces B cell proliferation and differentiation into cells that produce antibodies. However, this activation depends on additional signals; without them, BCR signaling results in cell death. Many of these supplementary signals come from CD4+ helper T cells. To gain T cell assistance, the BCR acts as an endocytic receptor, capturing antigens and transporting them to intracellular vesicles for processing and presentation on major histocompatibility complex class II (MHC II) molecules, which antigen-specific helper T cells recognize. Additionally, T cells encourage class-switching from the default IgM and IgD antibodies to IgG, IgA, or IgE. These different antibody classes create distinct BCRs on B cell surfaces, refining the response to antigens [30]. For instance, IgG BCRs are more effective at stimulating antibody secretion than IgM BCRs.

Another crucial aspect of the B cell receptor (BCR) is its ability to detect the affinity of antigen binding. Instead of acting as a simple on-off switch, the BCR adjusts responses to favor B cells with a higher affinity for the antigen over those with lower affinity. This mechanism underlies antibody affinity maturation, driven by the selection of B cells that have developed somatic mutations in the antigen-binding sites of their antibodies [31]. Even in the absence of an antigen, the BCR provides essential survival signals to resting B cells[32]. Therefore, a comprehensive understanding of B cell activation necessitates a detailed structural knowledge of the BCR.

The exact mechanism by which antigen binding signals are transmitted through the B cell receptor (BCR) to the cell's interior has long been mysterious. Initially, it was believed that antigens, often multivalent like virus particles, merely caused BCRs to aggregate on the B cell surface. This aggregation was thought to trigger cross-phosphorylation of the immunoreceptor tyrosine-based activation motifs (ITAMs) on the Igα-Igβ intracellular domains by associated tyrosine kinases, without requiring conformational changes to be communicated through the BCR's flexible hinge region [33]. However, phosphorylation was also observed in BCRs not directly engaged by antigens, suggesting that resting BCR oligomers might be separated by antigen binding, revealing their ITAM phosphorylation sites [34]. Additionally, because the BCR often needs to extract antigens from other cell surfaces, mechanical forces on the BCR could induce conformational changes or clustering [35]. This process is likely regulated by the submembrane cytoskeleton, which helps maintain the BCR in its resting state [36].

1.6. B-cell receptor (BCR) genes recombination and response

Before depicting BCR signaling proteins as biomarkers in CLL, it is important to review normal BCR formation and the implications of antigen encountering. B cells create a vast platform of antigen specific BCRs through gene recombination of immunoglobulin (IG) genes. This process begins early in B cell development in the bone marrow with the random rearrangement of the variable (V), diversity (D), and joining (J) genes of the IG heavy (IGH) chain repertoire. A pro-B cell selects and joins -by ligation- one V, D, and J gene from the present options, resulting in a functional IGH chain if correctly assembled which forms a pre-BCR by pairing with a surrogate light chain, a receptor capable of

signaling for B cell survival and maturation. The growing B-cell then goes in positive selection, and next the VJ recombination of the IG kappa or lambda (IGK/IGL) light chain locus takes place. The V(D)J recombination process also introduces additional diversity through enzymatic editing of the junctional regions, known as complementarity determining region 3 (CDR3) [37]. Hereafter, the IGH and IGK/IGL chains get linked and are expressed on the cell surface in a very small cluster with the supporting proteins CD79α/β. Still in the bone marrow, a further negative selection to the BCR complex is held, in which strong antigen-binding capacity leads to clonal deletion or receptor-editing, avoiding self-reactivity. Liberation to peripheral blood occurs to the mature B-cell after surviving the strict negative selection, which substantially raises the probability of antigen-encountering.

After the mature B-cell binds to its specific antigen, a cascade of phosphorylation reactions occurs involving a series of tyrosine kinases downstream the BCR. Starting with Lyn that phosphorylates the cytoplasmic ends of CD79α/β. Syk, a key player in BCR signaling, is then incorporated, and is activated through phosphorylation. Subsequently, Syk activates BTK with assistance from BLNK, and this complex phosphorylates PLCγ2 in which the latter becomes activated and generates second messengers PKC and Ca2+, leading to the activation of ERK, JNK, p38, NF-κB, and NFAT. Occurring in parallel, Syk activates PI3K leading to AKT pathway triggering, and Lyn (another key player) phosphorylates CD19 [38]. These implications provoke intracellular B-cell survival program preparing for the next steps. It is worth mentioning here that based on the antigenicity of the enemy and the affinity of the involved BCR, a process called Somatic Hypermutation (SHM) takes place in the germinal center of the lymph-node's follicle to further diversify the BCR by introducing point mutations in the V genes of IGH and IGK/IGL chains which optimizes BCR antigenic grasp [37]. The final negative selection mechanism occurs before exiting the lymph node, in which immunogenic tolerance is checked leading to B-cell apoptosis or anergy if its BCR engages an epitope with improper co-stimulatory signals or with weak binding strength [39].

1.7. Risk stratification of CLL in terms of SHM

The importance of BCR signaling in CLL was first highlighted by immunogenetic studies, which scoped the unique features of clonotypic BCR and revealed their association with CLL leukemogenesis and prognosis. In the late 1990s, the somatic hypermutation status on IGHV gene was unprecedentedly correlated to overall survival (OS) by two constitutional studies which classified CLL to IGHV unmutated (U-CLL; ≥98% germline homology) and IGHV mutated (M-CLL; <98% germline homology) [12, 13]. In terms of treatment initiation, it was revealed that U-CLL patients are in need of significant shorter time to therapy than M-CLL patients and this could be attributed to the occurrence of dismal gene mutations in U-CLL or possibly due to the difference in BCR downstream activation capabilities between the two mutational subsets [40, 41]. This might not be more than a reflection of two different ontogenies; for example, the B-cell maturation stage could be elicited indirectly by the mutational status of BCR in which it discriminates between early-stage and late-stage memory B-cell leukemic transformation [42]; it might also be a reflection of B-cell response, as the B-cell responds differently when in germinal center or extra-follicle [43]. These differences create unique epigenetic, transcriptomic, and metabolic profiles for each mutational subset resulting in higher potential of U-CLL cases for experiencing clonal expansion and evolution, which in turn increases the probability of progression in this mutational subset.

Interestingly, BCR rearrangement in CLL narrows the spectrum of V(D)J gene usage, where the most frequent IGHV genes encountered are IGHV1-69, IGHV3-7/3-23, and IGHV4-34. Nonetheless, BCR stereotypy is conspicuously seen in unrelated CLL patients who harbor quasi-identical BCRs [44]. CLL clones among stereotyped patients have immunogenetic characteristics in common, i.e. IGHV/IGHD/IGHJ gene usage and heavy chain CDR3 sequences [44-47], and remarkably, major subsets possess HCDR3 sequences restricted to particular AA residues [44, 46]. In current practice, about 40% of CLL patients could be assigned to specific subsets using their HCDR3 sequences with major subsets #1, #2, #4, and #8 collectively comprising the largest fraction (6-7%) of CLL cases [44]. This frequency could be increased by implementing larger cohorts according to recent studies, which facilitates classifying the vast majority of CLL patients into stereotypic clusters [48]. Noticeably, stereotypic CLL cases tend to have consistent clinical and biological profiles,

as seen for example in patients with subset #2 who have aggressive manifestations regardless of the mutational status of IGHV. Collectively, the BCR stereotypy serves as an additional prognostic biomarker to stratify CLL patients [45, 47, 49].

2. OBJECTIVES OF THE STUDY

Until recently, IGH chain recombination data were the sole source for BCR stereotypy assignment. Intriguingly enough, subset #2 BCR stereotype is frequently associated with an IGLV3-21 lambda light chain (with highest frequency of the R110 SHM) and IGHV3-21, which suggest associations between other unique IGHV-based subsets and specific light chain rearrangements[50, 51]. Although this might be interesting, the prognostic significance of IGK/IGL light chain repertoire in CLL has been poorly studied rather reported. This study aimed to:

1) Investigate the light chain gene repertoire and how it might affect the prognosis of CLL patients.
2) Assess the interplay between the light chain repertoire and the heavy chain repertoire and study the combined effect on CLL prognosis.

3. MATERIALS AND METHODS

2.1. Study population

The training cohort included 573 CLL patients referring at our institution from which tumour genomic DNA was extracted from peripheral blood mononuclear cells. Samples were collected at the time of diagnosis and all cases were provided with the main clinical and biological data of the disease. The validation cohort included 343 Rai 0 CLL patients deriving from a previously published study aiming at the development of a scoring system for predicting early treatment requirement in Rai 0 patients [21]. Patients informed consents were provided in accordance with the local institutional review board requirements and the Declaration of Helsinki. The study was approved by the local ethical committee (study number CE 120/19).

2.2. Separation of mononuclear cells from peripheral blood

The separation of mononuclear cells was performed by density gradient centrifugation. Peripheral Blood (PB) samples were diluted 1:2 with physiological solution (NaCl 0.9%) and then centrifuged at the following program (1800 revolutions per minute (rpm) for 25 minutes) in a gradient differentiation Sigma Diagnostic Histopaque-1077 Cell Separation Medium (Sigma-Aldrich, St. Louis, MO, USA) solution that is used to separate blood cells into specific cell populations based on their densities. It separates white blood cells into granulocytes and mononuclear cells (monocytes and lymphocytes).

2.3. Genomic DNA extraction from peripheral blood mononuclear cells

DNA extraction was performed using the technique of "Salting out" which provides a rapid extraction of DNA by cell lysis, precipitation of proteins with salts and recovery of DNA with ethanol. The cells were suspended in a lysis buffer, consisting of (10 mM Tris-HCl, 10 mM NaCl and 2 mM EDTA), to which a surfactant with denaturing function, sodium dodecyl sulphate (SDS) was added to a final concentration of 0.5%, and Pronase E, at a final concentration of 400 µg/ml. The cell lysate was incubated for 16-20 hours at 37° C in a thermostatic stirrer. After the incubation period and after verifying the complete cell lysis, the proteins were eliminated by NaCl precipitation at a final concentration equal

to 1.6 M and centrifugation at 3200 rpm for 20 minutes. The DNA was recovered in a jelly-like form of DNA which was seen after the addition of 100% ethanol. The DNA was extracted with Pasteur glass loops and subjected to three washes with 75% ethanol. Once the ethanol has evaporated, each sample was dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). This reagent has an alkaline pH and allows a correct dissolution and conservation of DNA, since the buffering activity counteracts the acidification of the solution and the degradation of the DNA, while the calcium chelating activity inhibits the proliferation of microorganisms.

2.4. Quantification of DNA

DNA concentration was determined by the NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo Scientific, Massachusetts, USA). Nanodrop is a UV-visible spectrophotometer that uses small amounts of starting material exploiting a technology based on the surface tension that small volumes of DNA (1-2 µl) exert when they are placed between two neighbouring surfaces. The instrument allows to analyse samples with a concentration range between 2.0 and 27,500 ng/µL allowing the elimination of the pre-measurement dilution phase. The Blank solution was TE buffer. A drop of the sample placed on the podium is in direct contact with two optical fibres that allow a quick analysis in the order of a few seconds. The sample is read at a wavelength of 230 nm, 260 nm and 280 nm to evaluate, in addition to DNA concentration, the presence of contaminants such as ethanol or proteins. A ratio is calculated between the different readings. A ratio value between 1.9 and 2.1 is indicative of a good-quality DNA. The DNA concentration is calculated by the Lambert-Beer law, an empirical relation that correlates the quantity of light absorbed by a medium to its chemical nature and concentration.

2.5. Light chain gene rearrangements identification

2.5.1. Training cohort

In the training cohort, light chain gene rearrangements were amplified with a polymerase chain reaction (PCR) method using GoTaq G2 hot start polymerase kit (Promega) and the reaction mixture was prepared as stated in **Table 1**.

PCR Reaction			
Reagent	**Initial concentration**	**Final concentration in mix**	**Volume in mix**
GoTaq Flexi Buffer (Promega)	5X	1X	6 µl
Mgcl2	25 mM	1.5 mM	1.8 µl
dNTPs	10 mM	100 µM	0.3 µl
GoTaq (Promega)	5 U/µL	0.25 U/µL	0.05 µl
Forward Primer	5 µM	0.16 µM	1 µl
Reverse Primer	5 µM	0.16 µM	1 µl
DNA	50ng/µl	100 ng/µl	2 µl
H2O			17.85 µl
PCR cycling conditions			
Stage	**Temperature**	**Time**	**Cycles**
Stage 1.	95°C	3 min	
Stage 2. Step 1.	95°C	30 sec	Step 2. Temperature is lowered by 0.5 for each cycle
Step 2.	65°C	40 sec	
Step 3.	72°C	40 sec	(x 12 cycles)
Stage 3. Step 1.	95°C	30 sec	
Step 2.	59°C	40 sec	(x 35 cycles)
Step 3.	72°C	40 sec	
Stage 4.	72°C	10 min	

Table 1 PCR reaction and conditions

Leader primers were used for the amplification of the entire sequence of the rearranged light chain gene, thus enabling the true level of somatic hypermutation (SHM) to be determined. The sequences of leader primers are detailed in **Table 2** for kappa light chains and in **Table 3** for lambda light chains.

Primer Mix	Name	Sequence 5'→ 3'
Kappa light chain primers		
LVK1mix	LVK1a	CCTGCTCAGCTCCTGGGRCTCCTGC
	LVK1b	CCCGCTCAGCTCCTGGGGCTYCTGC
LVK2	LVK2	CTCCTGGGGCTGCTAATGCTCTGG
LVK3	LVK3	CTCTTCCTCCTGCTACTCTGGCTC
LVK4	LVK4	ATGGTGTTGCAGACCCAGGTCTT
LVK5-6 mix	LVK5	TTCCTCCTCCTTTGGATCTCTG
	LVK6	TTCTGCTSCTCTGGGTTCCAG
FR1VK1	FR1VK1	GACATCCRGWTGACCCAGTCTCCWTC
FR1VK2mix	FR1VK2a	CAGWCTCCACTCTCCCTGCCCGTC
	FR1VK2b	CAGACTCCACTCTCTCTGTCCGTC
	FR1VK2c	CAGACTCCACTCTCCTCACCTGTC

FR1VK3mix	FR1VK3a	TCTCCAGSCACCCTGTCTTTGTCTC
	FR1VK3b	TCTCCAGCCACCCTGTCTGTGTCTC
FR1VK4	FR1VK4	AGACTCCCTGGCTGTGTCTCTGGGC
FR1VK5-6mix	FR1VK5	CAGTCTCCAGCATTCATGTCAGCGA
	FR1K6a	TTTCAGTCTGTGACTCCAAAGGAG
	FR1K6b	TTCCTCTCTGTGACTCCAGGGGAG
JK outer mix	JK1-4 out	GTTTGATYTCCASYTTGGTCCC
	JK5 out	GTTTAATCTCCAGTCGTGTCCC

Table 2 Leader primers and Framework region (FR) 1 primers for kappa light chain genes

Primer Mix	Name	Sequence 5'→ 3'
Lambda light chain primers		
LVL1	LVL1	CTGGTCYCCTCTCYTCCTCAC
LVL2mix	LVL2a	CTCCTCAGCCTCCTCACTCAGG
	LVL2b	CTCCTCACYCTCCTCACTCAGG
LVL3amix	LVL3	CTCCTKCTCCCCCTCCTCACT
	LVL3e	CTCCTGCTCCCACTCCTCAAC
	LVL3h	GCCTGGACCGTTCTCCTCCTC
LVL3bmix	LVL3j	GCCTGGACCGCTCTCCTTCTG
	LVL3l	TCTGGCTCACTCTCCTCACTC
	LVL3r	CCTCGGCGTCCTTGCTTACTG
LVL4mix	LVL4a	CTCTTCCCTCTCCTCCTCCAC
	LVL4b	CCTCACCCTCCTCCTCCACTG
	LVL4c	CCTACTGCCCTTCATTTTCTC
LVL5-6mix	LVL5	TGYTCCTCTCTCACTGCACAG
	LVL6	CCTCCTCGCTCACTGCACAG
LVL7-10mix	LVL7	ATGGCCTGGACTCCTCTCTTT
	LVL8	ATGGCCTGGATGATGCTTCTC
	LVL9	TGCTCCTCACCCTCCTCAGTC
	LVL10	CTGGGCTCTGCTCCTCCTGAC
FR1VL1	FR1VL1	GGTCCTGGGCCCAGTCTGTG
FR1VL2kupp	FR1VL2kupp	CAGTCTGCCCTGACTCAGCCT
FR1VL3mix	FR1VL3a	CTCAGCCACCCTCAGTGTCCGT
	FR1VL3b	CTCAGCCACCCTCGGTGTCAGT
	FR1VL3c	CTCAGGACCCTGCTGTGTCTGT
FR1VL4-6mix	FR1VL4kupp	TTTCTTCTGAGCTGACTCAGGAC
	FR1VL5-9	CAGSCTGTGCTGACTCAGCC
	FR1VL6kupp	GAGTCTCCGGGGAAGACGGTA
FR1VL7-10mix	FR1VL7kupp	GTGGTGACTCAGGAGCCCTCAC
	FR1VL7-8	CTGTGGTGACYCAGGAGCC
	FR1VL9kupp	GCTGACTCAGCCACCTTCTGCA
	FR1VL10	CAGGCAGGGCWGACTCAGC
JL outer mix	JLa out	CTAGGACGGTSASCTTGGTCCC
	JL7 out	CGAGGACGGTCAGCTGGGTGCC
IGLV3-21 R110 primers		
	IGLV3-21_F	GCTCTGTGACCTCCTATGTGC

	JC1 intron	GGCCCCATGGAGAAAGGTAA
	JC2/3 intron	CAAGCAAGGGTCTGAACAGG
	JC7 intron	AGGGGCAAAGATTCCAGACA

Table 3 Leader primers and Framework region (FR) 1 primers for lambda light chain genes including IGLV3-21^{R110} primers.

In cases of leader primers' inability to identify the rearranged gene, framework region 1 (FR1) primers were used as alternatives (forward primers) with the same reverse primers (JK, JL outer mix) mentioned above, and their sequences are listed in **Table 2**. PCR products were then visualized by 2% agarose gel electrophoresis containing a DNA interlayer, in TBE 1X buffer (1M Tris, 1M boric acid and 20 mM EDTA, pH 8.0). The bands were visualized with UV using the ChemiDoc MP imaging system instrument (Bio-Rad Laboratories S.R.L. Milan, Italy). The templates were then purified using MultiScreen PCRμ96 Filter Plate (Merck-Millipore, Damstadt, Germany), using a vacuum pump (Merck-Millipore), according to the company's protocol.

The purified DNA was subjected to bi-directional Sanger sequencing. The sequence reaction was performed using the BrilliantDye Terminator v1.1 (Nimagen, Gelderland, Netherlands). The Sanger reaction mixture can be seen detailed in **Table 4**. The Sanger reaction protocol was set on the Thermal Cycler Veriti 96 well (Applied Biosystem, Forest City, USA) as shown in **Table 4**.

Sanger Sequencing Reaction			
Reagent	**Initial concentration**	**Final concentration in mix**	**Volume in mix**
BrilliantDye Buffer v1.1 (Nimagen)	5X	0.4X	1.6 µL
BrilliantDye v1.1 RR Premix	2.5X	0.1X	0.8 µL
Forward Primer	5 µM	0.175 µM	0.7 µL
DNA	PCR product		1 µL
H2O			15.9 µL
Sanger cycling conditions			
Stage	**Temperature**	**Time**	**Cycles**
Stage 1.	98°C	3 min	
Stage 2.	96°C 60°C	10 sec 4 min	x 27 cycles

Table 4 Sanger sequencing reaction

Sequence reaction products were purified with the Montage SEQ 96 Sequencing reaction cleanup kit (Merck-Millipore, Burlington, USA). A volume of 5 µl of sequenced DNA was added to 10µl of formamide, and together they were denatured and subjected to capillary electrophoresis with the automated ABI PRISM 3130XL Genetic Analyzer (Applied Biosystem, Forest City, USA) and SeqStudio Genetic Analyzer for Human Identification (Applied Biosystem, Forest City, USA). The migration of DNA fragments was detected by fluorescence emission, and the data were analysed and displayed as a nucleotide sequence in a chromatogram as in **Figure 1**.

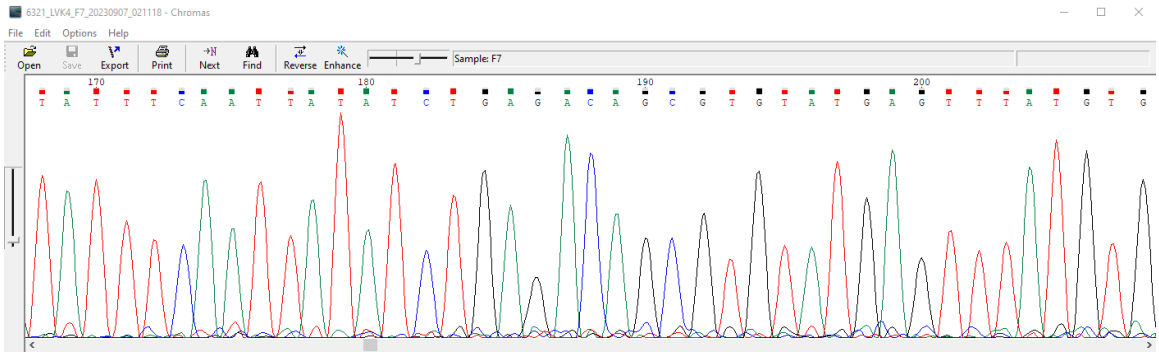

Figure 1 Sequence chromatogram illustrated by Chromas

2.5.2.Validation cohort

In the validation cohort, the light chain rearrangements were analysed by NGS. Kappa and lambda rearranged genes were amplified using a pool of dedicated primers in a first PCR reaction (**Table 5**), followed by a second PCR reaction assigned for indexing the amplicons in a step to add the universal adapter and the patient's specific 5' end and 3' end barcodes (**Table 6**).

IGK Primers	Sequence with Illumina partial adapter
VK1	GTTCTACAGTCCGACGATCGGACATCCRGWTGACCCAGTCTCCWTC
VK2	GTTCTACAGTCCGACGATCGCAGWCTCCACTCTCCCTGYCCGTCA
VK3	GTTCTACAGTCCGACGATCGTTGTGWTGACRCAGTCTCCAGSCACC
VK4	GTTCTACAGTCCGACGATCGAGACTCCCTGGCTGTGTCTCTGGGC
VK5	GTTCTACAGTCCGACGATCGCAGTCTCCAGCATTCATGTCAGCGA
VK6	GTTCTACAGTCCGACGATCGTTTCTSTCTGTGACTCCARRGGAGAA
VK7	GTTCTACAGTCCGACGATCGTCTCCATGSCACCCTGTCTKTGTCTC

JK1	TTGGCACCCGAGAATTCCACTGACGTTTGATCTCCACCTTGGTCCC
JK2	TTGGCACCCGAGAATTCCACTGACGTTTGATATCCACTTTGGTCCC
JK3	TTGGCACCCGAGAATTCCACTGACGTTTAATCTCCAGTCGTGTCCC
IGL Primers	**Sequence with Illumina partial adapter**
VL1	GTTCTACAGTCCGACGATCGGGTCCTGGGCCCAGTCTGTGCTG
VL2	GTTCTACAGTCCGACGATCGGGTCCTGGGCCCAGTCTGCCCTG
VL3	GTTCTACAGTCCGACGATCGGCTCTGWGCCCTCCTATGAGCTG
VL4	GTTCTACAGTCCGACGATCGTCTCTCTCSCAGCYTGTGCTG
VL6	GTTCTACAGTCCGACGATCGGTTCTTGGGCCAATTTTATGCTG
VL7	GTTCTACAGTCCGACGATCGGGTCCAATTCYCAGGCTGTGGTG
VL8	GTTCTACAGTCCGACGATCGGAGTGGATTCTCAGACTGTGGTG
JL1	TTGGCACCCGAGAATTCCACTGGCCACTTACCTAGGACGGTGAC
JL2	TTGGCACCCGAGAATTCCACTGGAAGAGACTCACCTAGGACGGTC
JL3	TTGGCACCCGAGAATTCCACTGGGAGACTYACCGAGGACGGTC
JL4	TTGGCACCCGAGAATTCCACTGGGACGGTGACCTTGGTCCCAGT
JL5	TTGGCACCCGAGAATTCCACTGGACGGTCAGCTTGGTSCCTCC
JL6	TTGGCACCCGAGAATTCCACTGGACGGTCACCTTGGTGCCACT

Table 5 Kappa and lambda primers used for the first PCR reaction

Index S	**Sequence**
Index_S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S503	AATGATACGGCGACCACCGAGATCTACACTATCCTTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S504	AATGATACGGCGACCACCGAGATCTACACAGAGTATCTACACGTTCAGAGTTCTACAGTCCGA
Index_S505	AATGATACGGCGACCACCGAGATCTACACGTAAGGTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S506	AATGATACGGCGACCACCGAGATCTACACACTGCATCTACACGTTCAGAGTTCTACAGTCCGA
Index_S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S508	AATGATACGGCGACCACCGAGATCTACACCTAAGCTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S510	AATGATACGGCGACCACCGAGATCTACACCGTCTATCTACACGTTCAGAGTTCTACAGTCCGA
Index_S513	AATGATACGGCGACCACCGAGATCTACACTCGACTTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S515	AATGATACGGCGACCACCGAGATCTACACTTCTAGTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S516	AATGATACGGCGACCACCGAGATCTACACCCTAGATCTACACGTTCAGAGTTCTACAGTCCGA
Index_S517	AATGATACGGCGACCACCGAGATCTACACGCGTAATCTACACGTTCAGAGTTCTACAGTCCGA
Index_S518	AATGATACGGCGACCACCGAGATCTACACCTATTATCTACACGTTCAGAGTTCTACAGTCCGA
Index_S520	AATGATACGGCGACCACCGAGATCTACACAAGGCTTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S521	AATGATACGGCGACCACCGAGATCTACACGAGCCTTCTACACGTTCAGAGTTCTACAGTCCGA
Index_S522	AATGATACGGCGACCACCGAGATCTACACTTATGCTCTACACGTTCAGAGTTCTACAGTCCGA
Index A	**Sequence**
Index_A1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

Index_A5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A13	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A14	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A15	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A16	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A18	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A19	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A20	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A21	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A22	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A23	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A25	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
Index_A27	CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

Table 6 Index S and Index A used for the second PCR reaction

The amplicon pool was later purified with AMPure XP magnetic beads (Beckman Coulter Inc, Brea, CA, USA) and the resulting amplicons were quantified using Quant-iTM PicoGreen dsDNA Assay kit (ThermoFisher Scientific, Eugene, OR, USA) and checked for an average length of ~470 base pairs (bp) using Agilent 2100 Bioanalyzer (Agilent Technologies, St. Clara, CA, USA). They were later sequenced in an Illumina MiSeq sequencing system using MiSeq Reagent Nano Kit v2 (paired-end, 2 × 250[bp], 500-cycles) (Illumina Inc., San Diego, CA, USA). Repertoire analysis was performed using Vidjil (http://www.vidjil.org) [52], and International ImMunoGeneTics information (IMGT) [53] resulting in consensus sequences for each clonotype. The dominant light chain clone was considered as the CLL clone if it had number of reads that is significantly higher than other clones found in each patient. A Z-test was performed, and a Bonferroni-corrected p-value was calculated to confirm significance. Only productive dominant clones were included in the clinical analysis. Double productive cases were identified by

possessing two dominant clones with the previous criteria. In this case, the clone with the highest V-gene homology percentage was included in the clinical analysis.

2.6. Statistical analysis

The primary endpoint of this study was time to first treatment (TTFT), defined as the time between CLL diagnosis and the date of first line therapy (event) as a result of progression to a symptomatic disease according to the International Workshop on Chronic Lymphocytic Leukemia (iwCLL) guidelines and death/last follow up (censoring). A maximally selected rank statistic (Maxstat) R package was used to create a cut-off for light chain mutational status based on the Log-rank statistics for TTFT. Survival analysis was performed using the Kaplan-Meier method and compared between strata using the Log-rank test. A false discovery rate approach was used to account for multiple testing, and adjusted p-values were calculated using Bonferroni correction. Cox regression test was used in the univariate and the multivariate analysis. Pearson's correlation was used to compare the prevalence of rearrangements in different molecular and clinical subgroups of patients. The tests were performed using the Statistical Package for the Social Sciences (SPSS) software v.24.0 (Chicago, IL, USA) and R Studio Version 1.2.1335 2009-2019, Inc. Statistical significance was defined as p-value <0.05.

4. RESULTS

3.1. Patient characteristics

A total of 573 patients were analysed in the training cohort and 530 harboured at least one productive rearrangement. Forty-two (7.9%) patients presented double productive rearrangements. Patient characteristics were consistent with a real-world cohort of unselected newly diagnosed CLL (**Table 7**), as the median age at diagnosis was 70 years, 58.5% were male, the median lymphocyte count was 9900/µl, 36.8% had unmutated (UM)-IGHV, 46.5% had del13q, and 11.8% harboured TP53 disruption (i.e. TP53 mutation and/or 17p deletion).

Characteristics	Values
Median age at diagnosis (range)	70 (36 - 92)
Sex	
Male	310 (58.5%)
Female	219 (41.3%)
Median lymphocyte count	9900/µl
Median haemoglobin level	13.7 g/dl
Median platelet count	$200x10^3$/µl
Light chain	
Kappa	305 (57.6%)
Lambda	167 (31.5%)
Undefined	58 (10.9%)
IGHV	
Mutated	327 (63.1%)
Unmutated	191 (36.8%)
Binet staging system	
Binet A	414 (81.5%)
Binet B	53 (10.4%)
Binet C	41 (8.1%)
***TP53* disrupted**	62 (11.8%)
del17p	43 (8.7%)
TP53	50 (9.6%)
Trisomy 12	92 (18.7%)
NOTCH1	35 (11.4%)
del13q	229 (46.5%)
del11q	37 (7.5%)
MYD88	11 (3.6%)
SF3B1	21 (6.8%)

BIRC3	5 (1.6%)
Stereotyped cases	
Subset 1	11 (13.8%)
Subset 2	12 (15.0%)
Subset 4	6 (7.5%)
Subset 8	6 (7.5%)

Table 7 Patient characteristics in the training cohort

According to the Binet staging system, 414 patients (81.5%) were classified as Binet A, 53 patients (10.4%) as Binet B and 41 patients (8.1%) as Binet C. CLL stereotyped subsets were identified in 80/530 patients (15.1%), namely 11/80 patients (13.8%) belonged to subset #1, 12/80 (15%) to subset #2, 6/80 (7.5%) to subset #4 and 6/80 (7.5%) to subset #8. **Table 8** includes all CLL subsets identified in this study.

CLL Subset	No. (%)	CLL Subset	No. (%)
Subset#2	12 (15)	**Subset#66**	1 (1.3)
Subset#1	11 (13)	**Subset#7E**	1 (1.3)
Subset#8	6 (7.5)	**Subset#6**	1 (1.3)
Subset#4	6 (7.5)	**Subset#216**	1 (1.3)
Subset#59	3 (3.8)	**Subset#77**	1 (1.3)
Subset#68	2 (2.5)	**Subset#218**	1 (1.3)
Subset#204	3 (2.5)	**Subset#95**	1 (1.3)
Subset#10	4 (2.5)	**Subset#223**	1 (1.3)
Subset#16	5 (2.5)	**Subset#11**	1 (1.3)
Subset#7H	6 (2.5)	**Subset#246**	1 (1.3)
Subset#215	1 (1.3)	**Subset#5**	1 (1.3)
Subset#247	1 (1.3)	**Subset#248**	1 (1.3)
Subset#221	1 (1.3)	**Subset#23**	1 (1.3)
Subset#31	1 (1.3)	**Subset#250**	1 (1.3)
Subset#251	1 (1.3)	**Subset#206**	1 (1.3)
Subset#41	1 (1.3)	**Subset#7B**	1 (1.3)
Subset#217	1 (1.3)	**Subset#209**	1 (1.3)
Subset#9	1 (1.3)	**Subset#29**	1 (1.3)
Subset#226	1 (1.3)	**Subset#210**	1 (1.3)
Subset#63	1 (1.3)	**Subset#201**	1 (1.3)
Subset#249	1 (1.3)	**Subset#203**	1 (1.3)

Table 8 The frequency of CLL subsets in the training cohort

After a median follow-up of 11.6 years, the median overall survival for CLL patients was 11.7 years and the median time to first treatment for patients who were

classified as Binet A was 15.7 years. The 10-year overall survival (OS) for entire cohort was 61.0%. The well-known CLL prognostic biomarkers were associated with poor outcomes. The 10-year overall survival (OS) for UM-IGHV patients was 48.1% compared to 67.9% for mutated (M)-IGHV patients (p<0.0001) (Figure 2A). Similarly, the 10-year OS for TP53 disrupted cases was 36.7% compared to 64.2% for wild type TP53 patients (p<0.0001) (Figure 2B).

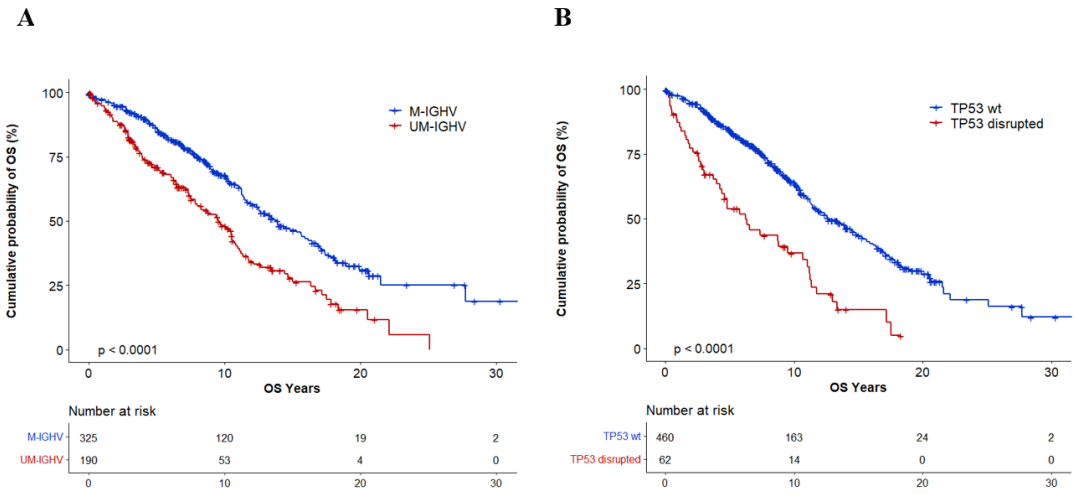

Figure 2 Kaplan-Meier curve in terms of overall survival of the well-known prognostic factors of CLL. **(A)** IGHV mutational status. The blue line represents patients with mutated heavy chains (M-IGHV), while the red line includes patients with unmutated heavy chains (UM-IGHV). **(B)** *TP53* disrupted (*TP53* mutation or del17p). Patients with disruptive aberrations of TP53 are indicated in red, while patients which harbour a wild type (wt)-*TP53* are indicated in blue. The log-rank statistics p values are indicated adjacent to the curves. The number of patients at risk in each group is indicated in the risk table under each graph.

3.2. Light chain rearrangements

According to flow-cytometry analysis, kappa light chains were expressed in 306 (57.7%) patients and lambda light chains in 166 (31.3%), while in 58 (10.9%) cases the expression of light chains was too low on the surface of the B cell that impaired determination of the CLL light chain restriction.

Overall, a total of 51 different productive IGKV (N=25) and IGLV (N=26) genes were identified. Leader primers were used to identify rearrangements in 404 patients (76.2%), while, in cases without productive rearrangements with the leader primers, the

FR1 primers were used (N=126, 23.8%). For kappa light chains, the most frequently rearranged gene was IGKV4-1 in 84 patients, that comprised 20.5% of the total kappa rearrangements, followed by IGKV3(D)-20 in 60 patients (14.8%) and IGKV1-39 in 42 patients (10.4%) (**Figure 3A**). For lambda light chains, the most frequently rearranged gene was IGLV3-21 in 32 patients, that comprised 19% of the total lambda rearrangements (N=168), followed by IGLV2-14 in 27 patients (16.1%), and IGLV3-25 in 12 patients (7.1%). (Figure 1B). Among the 32 patients who expressed the lambda rearrangement IGLV3-21, nineteen of them (59.4%) harboured the IGLV3-21*04^{R110} configuration (**Figure 3B**).

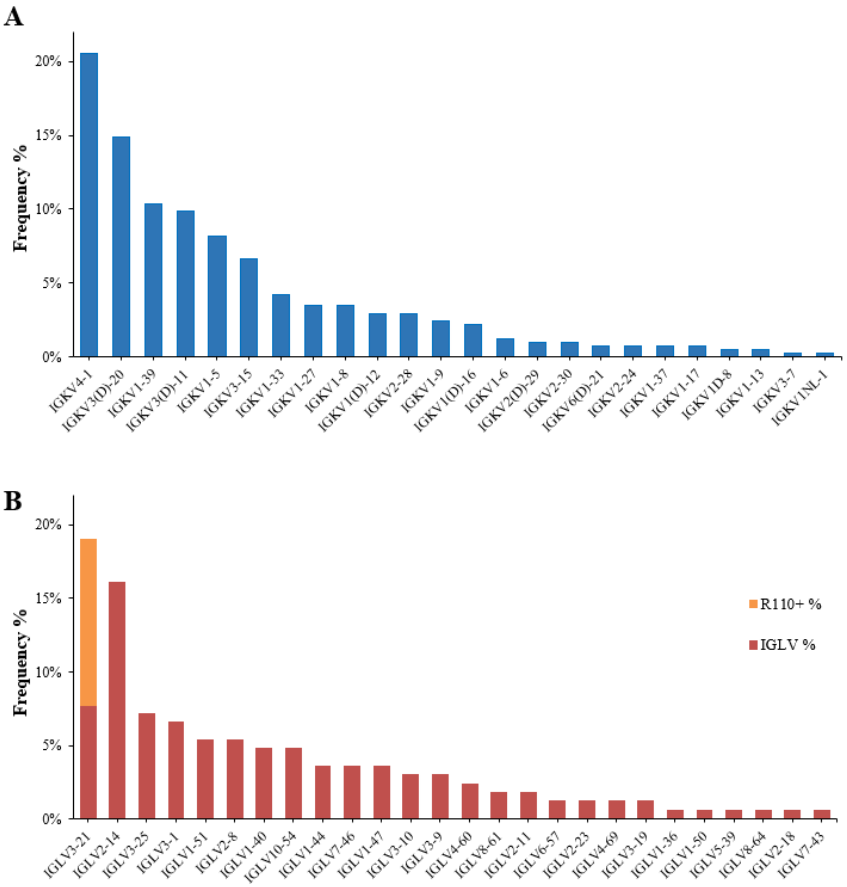

Figure 3 Frequency of rearranged light chain genes found in the training cohort and their association to CLL molecular features. **(A)** Bar chart depicting the frequency of kappa light chain rearrangements **(B)** Bar chart depicting the frequency of lambda light chain rearrangements. IGLV3-21^{R110} rearrangements are colour coded in orange.

Interestingly, 57/58 cases whose light chain expression was undefined through flow-cytometry analysis, displayed kappa rearrangements through sequencing. However, they did not present similarities with regards to rearrangements or any other characteristics. One unique case presented with double rearrangements that included the two most frequent genes, IGKV4-1, and IGLV3-21.

By evaluating the association of light chain rearrangements with baseline CLL molecular features, among lambda rearrangements, both IGLV3-21 and IGLV3-21^{R110} genes significantly associated with the usage of IGHV3-21 genes, with IGHV borderline mutational status (defined as a variable region homology to the germline sequence of 97-97.99%), with subset#2, and with *SF3B1* mutations (all p<0.001). Among kappa rearrangements, IGKV4-1 gene showed neither positive nor negative association with any of the CLL molecular features, whereas IGKV1-39 gene significantly associated with *NOTCH1* mutations (p<0.0001), trisomy 12 (p=0.0002), UM-IGHV (p=0.0001), subset #1 (p<0.0001), and subset #8 (p<0.0001). Additionally, IGKV3-11 gene associated with IGHV4-34 genes (p=0.0086) and subset #4 (p<0.0001) (**Figure 4**).

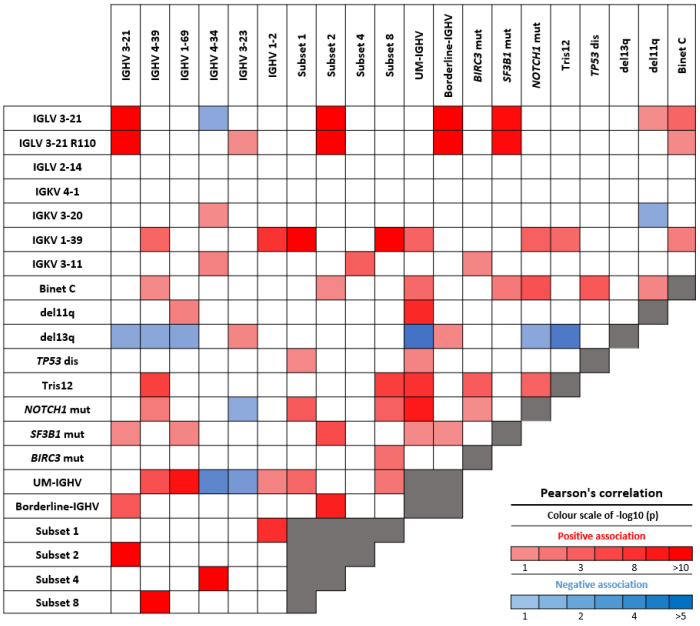

Figure 4 Correlation map comparing the most recurrent light chain rearrangements to the molecular features of CLL. The red color scale of -$\log_{10}$(p) points to a co-occurrence between two variables. The blue color scale of -$\log_{10}$(p) points to exclusivity between two variables. The intensity of the color corresponds to the strength of the correlation.

3.3. Association of specific light chain rearrangements with TTFT in Binet A

Light chain rearrangements with a frequency of at least 4% were investigated for their association to TTFT in Binet A CLL patients. IGKV1-39 (N=27), IGKV1-33 (N=13), and IGLV3-21 (N=14) were found to be significantly associated with shorter TTFT in Binet A (p=0.003, p=0.007, and p=0.004, respectively). Interestingly, only patients with IGKV4-1 rearrangement (N=47) exhibited longer TTFT when compared to all other rearrangement (p=0.02) (**Figure 5**). On the other hand, IGKV-3(D)-20, IGKV3(D)-11, IGKV1-5, IGKV3-15, IGLV2-14, IGLV3-25, IGLV3-1, IGLV1-51, and IGLV2-8 rearrangements did not show any association with TTFT.

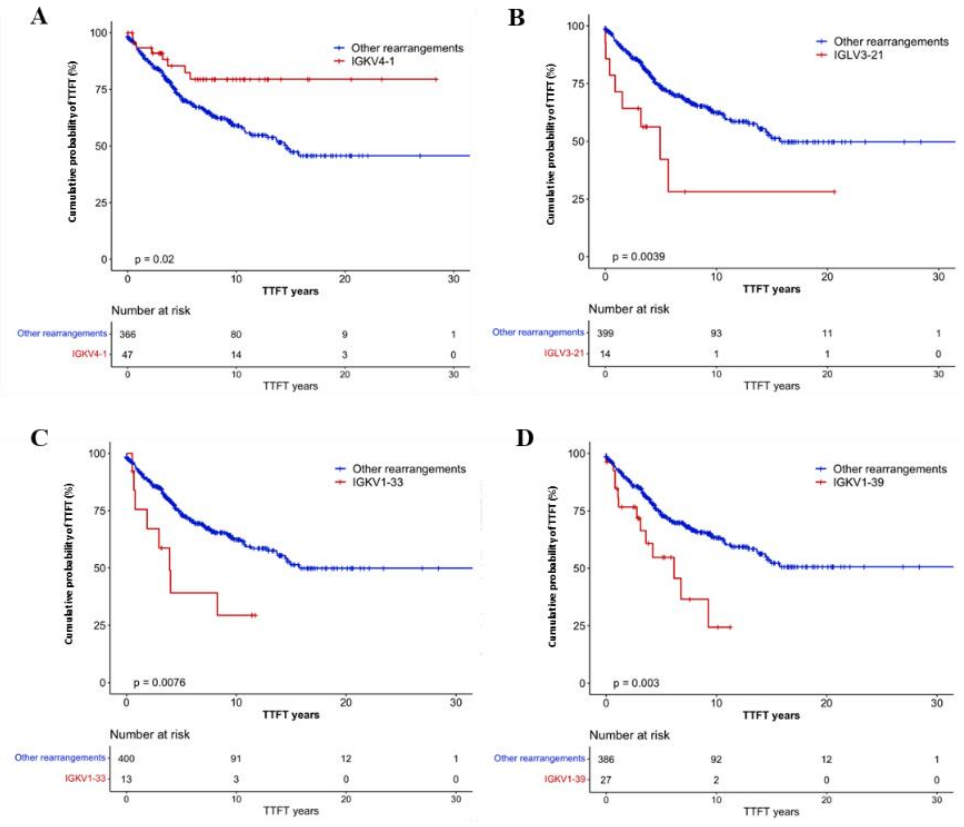

Figure 5 Light chains associated with TTFT in Binet A CLL. Kaplan-Meier curves of TTFT **(A)** in patients who express IGKV4-1 (in red) vs other rearrangements (in blue) (p=0.02); **(B)** in patients who express IGLV3-21 (in red) vs other rearrangements (in blue) (p=0.0039); **(C)** in patients who express IGKV1-33 (in red) vs other rearrangements (in blue) (p=0.0076); **(D)** in patients who express IGKV1-39 (in red) vs other rearrangements (in blue) (p=0.0003).

3.4. Light chain mutational status predicts TTFT in early stage CLL

Since no validated cut-off of homology has been established for analysing the mutational status of the light chain genes, a recursive partitioning approach identifies cut-off values ranging from 99.0% to 99.5% as the best cut-offs maximizing the log-rank statistics for TTFT in 414 Binet A CLL (**Figure 6**).

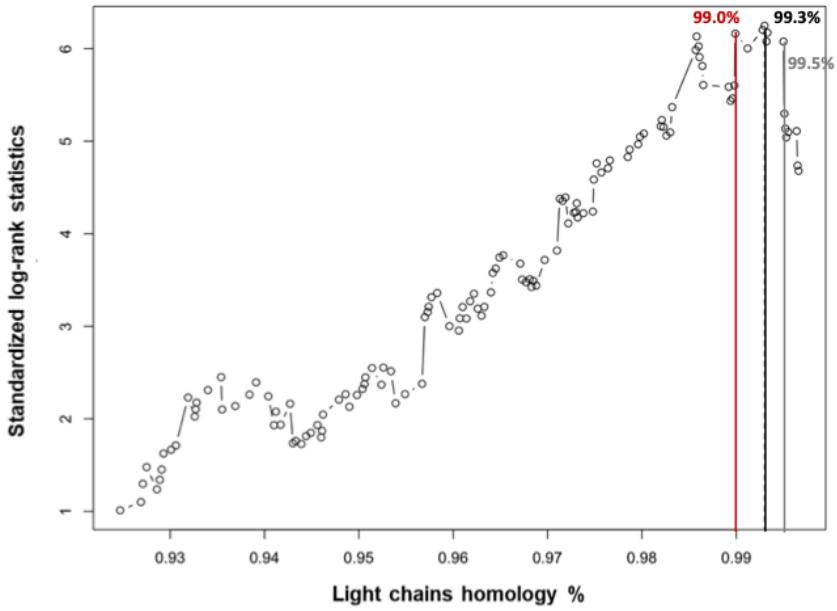

Figure 6 The standardized log rank statistics of the Maxstat test for the identification of the best cut-off of light chains gene mutational status that best predict TTFT. The optimal cut-off in prediction TTFT for light chains is represented by a homology of 99.3% **(black line)**. The closed integer value that showed very similar prediction is 99.0% **(red line).**

By choosing a cut-off value of 99% for clinical purposes, patients whose light chain gene sequences with a percentage that is equal to or higher than 99% of homology to the germline sequences were considered unmutated UM), while patients with less than 99% of homology were considered mutated (M). Using this new optimal cut-off, unmutated light chain patients presented a worse TTFT with a 10-year TTFT of 32.4% compared to 73.2% for mutated patients ($p<0.0001$) (**Figure 7**).

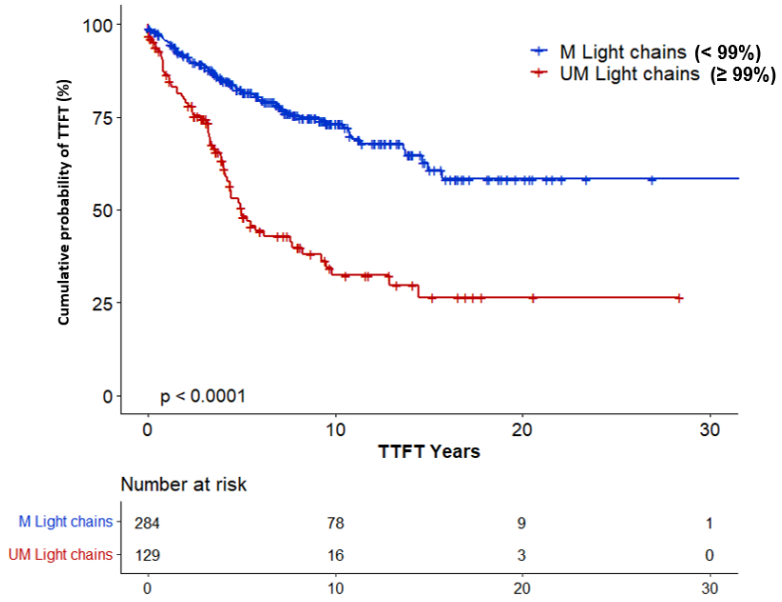

Figure 7 Kaplan-Meier curve of the light chain mutation status in terms of time to first treatment in Binet A patients using the optimal cut-off of homology. UM light chains patients (≥99%) are represented by the red line and M light chains patients (<99%) are represented by the blue line**.**

By considering separately kappa and lambda patients similar results were obtained. M-IGKV patients had a 10-year TTFT of 73.3% compared to 39.4% for UM-IGKV patients (p<0.001) (**Figure 8A**) and M-IGLV had a 10-year TTFT of 70.4% compared to 7.9% for UM-IGLV patients (p<0.001) (**Figure 8B**).

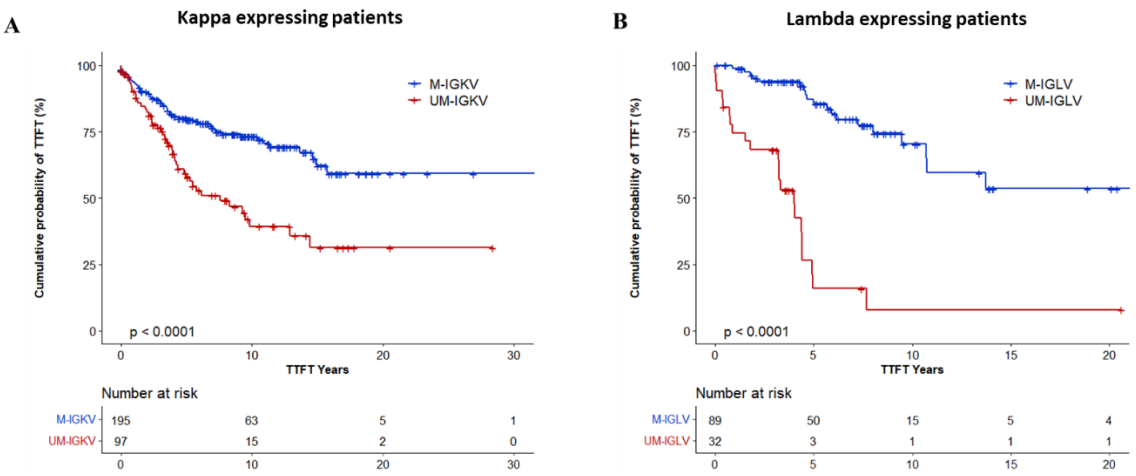

Figure 8 Kaplan-Meier curves in terms of time to first treatment for light chain mutation status, studied separately. **(A)** Kaplan-Meier curve of time to first treatment in kappa light chain Binet A patients using the optimal cut-off of homology. UM-IGKV patients (≥99%) are represented by the red line and M-IGKV patients (<99%) are represented by the blue line. **(B)** Kaplan-Meier curve of time to first treatment in lambda light chain Binet A patients using the optimal cut-off of homology. UM-IGLV patients (≥99%) are represented by the red line and M-IGLV patients (<99%) are represented by the blue line.

Parallel assessment of the light chain and its corresponding heavy chain was performed using the 99% cut-off set in this study for light chains and the standard 98% cut-off of homology for the heavy chains. This analysis showed that mutated heavy chain rearrangements associated with mutated light chain rearrangements and *vice versa* ($p<0.0001$). More precisely, 83.42% of M-IGHV patients also harboured M light chains, while 16.6% harboured UM light chains and 73.3% of UM-IGHV patients also harboured UM light chains, while 26.7% harboured M light chains (**Figure 9A**). Superimposable results were obtained by considering separately kappa and lambda expressing patients. (**Figure 9B-C**).

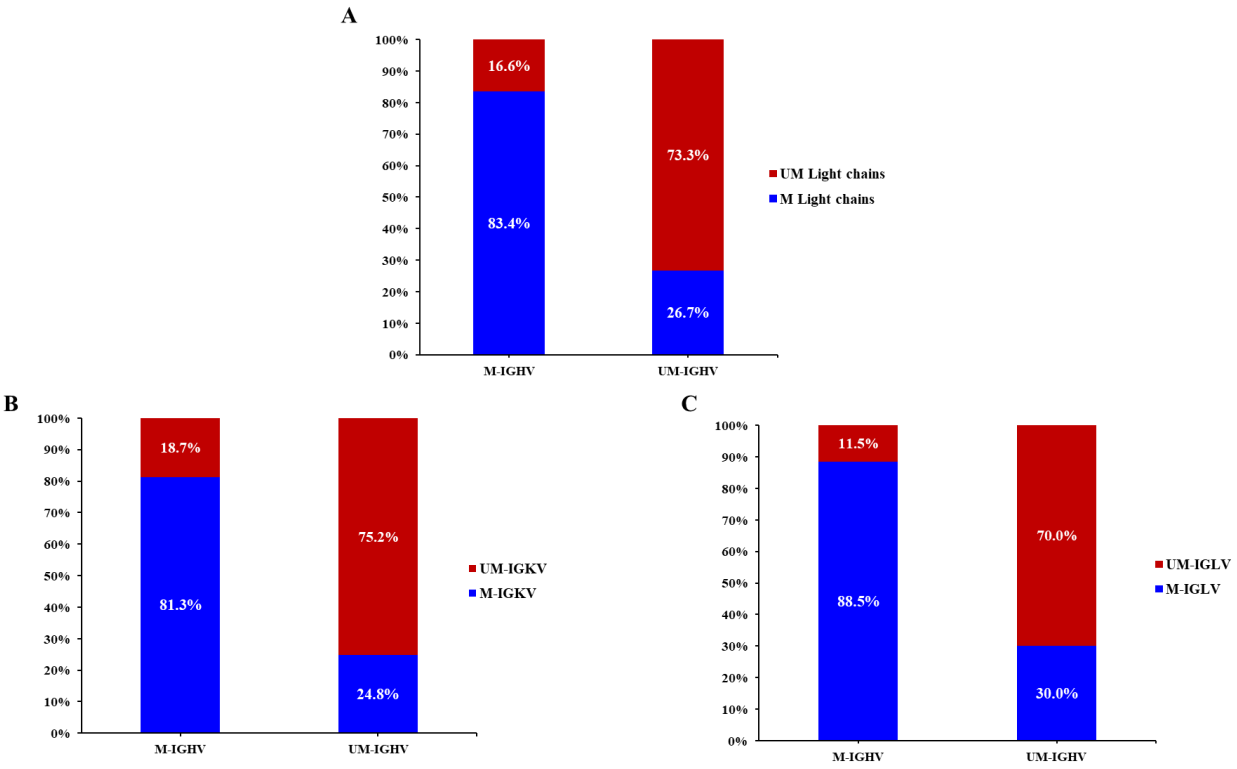

Figure 9 Parallel assessment of the IGHV mutation status (using the standard cut-off of 98%) and their corresponding light chains mutation status (using the optimal cut-offs of 99%). **(A)** A stacked column chart showing the % of patients with mutated and unmutated IGHV according to the total light chain rearrangements. **(B)** A stacked column chart showing the % of patients with mutated and unmutated IGHV according to kappa light chain rearrangements. **(C)** A stacked column chart showing the % of patients with mutated and unmutated IGHV according to lambda light chain rearrangements.

Therefore, TTFT was evaluated according to the combined heavy and light chains mutation status. CLL patients expressing both UM light chains and UM heavy chains (in red) associated with shorter TTFT (10 years TTFT of 13.6%) compared to patients with both M light chains and M heavy chains (in blue) (10 years TTFT of 76.1%, $p<0.001$). Discordant cases showed an intermediate outcome (10 years TTFT of 54.3%) that significantly differs from both mutated and both unmutated patients ($p=0.01$ and $p<0.001$, respectively) (**Figure 10**).

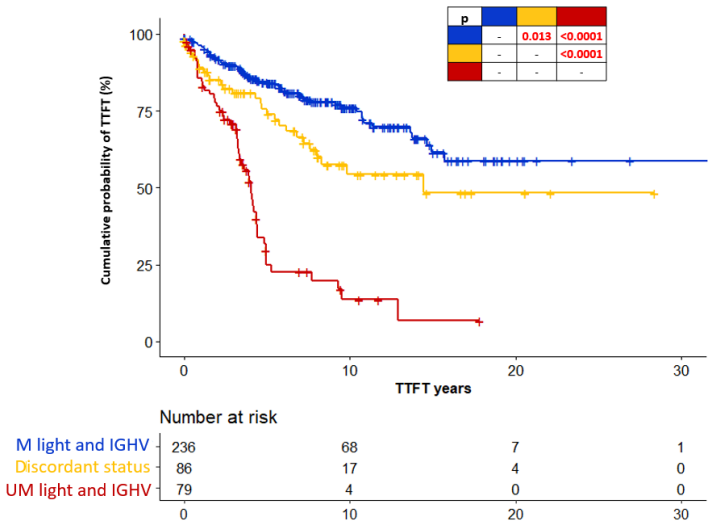

Figure 10 Kaplan-Meier curve of TTFT according to the combined heavy and light chain mutation status. M light chains and M heavy chains are represented by the blue line, UM light chains and UM heavy chains are represented by the red line and discordant cases (i.e. M light and UM heavy of UM light and M heavy) are represented by the yellow line. The pairwise log-rank statistics p values are indicated in the tables adjacent to the curve.

Subsequently, a multivariate analysis adjusted for the International Prognostic Score for Early-stage CLL (IPS-E) variables [24], namely UM-IGHV, palpable lymph nodes and lymphocyte count >15x10^9/L, was performed. Notably, UM light chains (both kappa and lambda) (HR=2.24, 95% CI 1.49-3.36, p<0.001) and UM-IGHV (HR=2.02, 95% CI 1.33-3.05, p<0.001) maintained an independent association with a shorter TTFT (**Figure 11**).

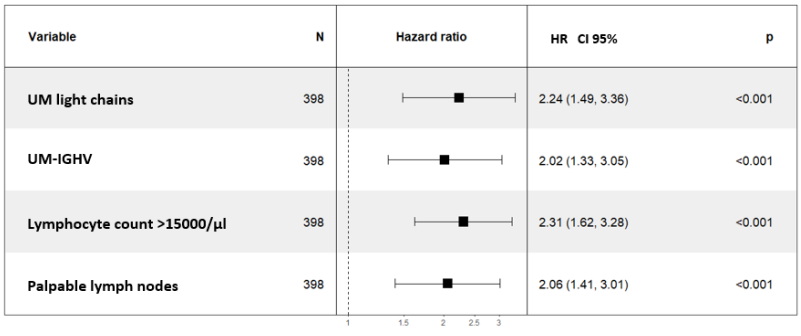

Figure 11 Forest plot multivariate analysis in terms of time to first treatment of light chain mutation status and the IPS-E variables.

3.5. The prognostic role of light chain mutational status validated in an independent cohort

The validation cohort included 343 Rai 0 patients (patient characteristics are reported in **Table 9**) and 299 patients showed productive results after NGS analysis. The median number of reads was 5316 for each patient and the dominant light chain gene clone identified for each patient accounted for a median total number of reads of 80.9%.

Characteristics	Values
Median WBC	20300/μl
Median B2M	1.7mg/l
Light chain	
Kappa	146 (48.8%)
Lambda	76 (25.4%)
Undefined	77 (25.7%)
IGHV	
Mutated	213 (71.7%)
Unmutated	84 (28.3%)
TP53 disrupted	37 (12.3%)
del17p	22 (7.4%)
TP53	29 (9.7%)
Trisomy 12	50 (16.7%)
NOTCH1	50 (16.7%)
del13q	169 (56.5%)
del11q	29 (9.7%)

Table 9 Patient characteristics in the validation cohort

The most frequent kappa rearrangement was IGKV4-1 identified in 35 patients, that comprised 17.9% of the total kappa rearrangements (N=196). It was followed by IGKV3(D)-20 expressed by 25 patients (12.8%) and IGKV1-5 by 16 patients (8.2%) (**Figure 12A**). For lambda light chains, the most frequent rearrangement was IGLV3-21 in 13 patients that comprised 12.6% of the total lambda rearrangements (N=103), followed by IGLV1-47 in 12 patients (11.7%), and IGLV2-14 in 10 patients (9.7%). (**Figure 12B**).

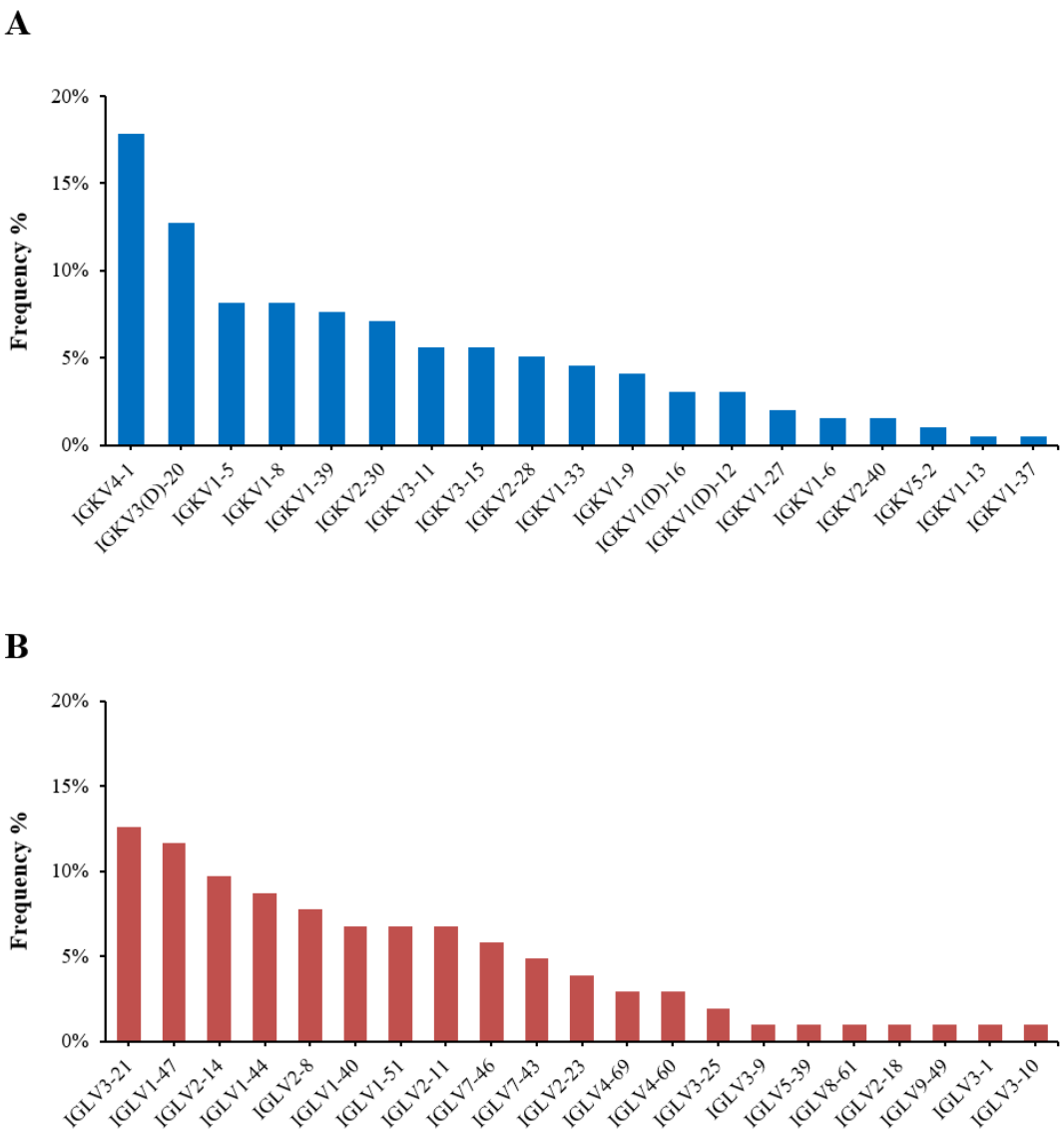

Figure 12 The frequency of light chain gene rearrangements in the validation cohort of Rai 0 patients and their clinical association to TTFT. **(A)** Bar chart depicting the frequency of kappa light chain rearrangements. **(B)** Bar chart depicting the frequency of lambda light chain rearrangements.

The prognostic role in terms of TTFT using the established homology cut-off of 99% was validated. UM light chain patients presented a 10-year TTFT of 25.2% compared to 63.1% for M light chain patients (p<0.0001) (**Figure 13**).

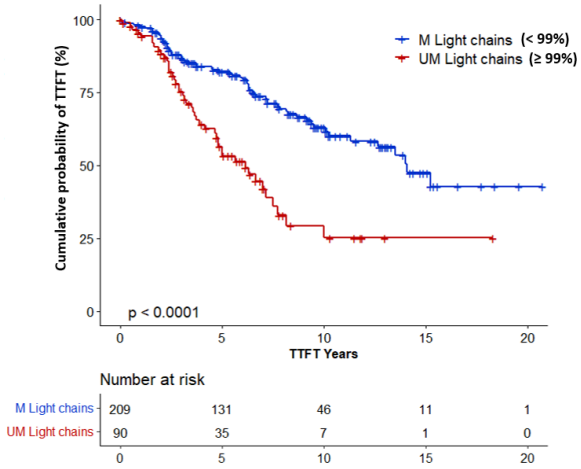

Figure 13 Kaplan-Meier curve in terms of TTFT regarding the total light chain genes mutation status using the optimal cut-off of homology. UM light chains patients (≥99%) are represented by the red line and M light chains patients (<99%) are indicated in blue.

Superimposable results were obtained by considering separately kappa and lambda expressing patients (**Figure 14A,** and **B**).

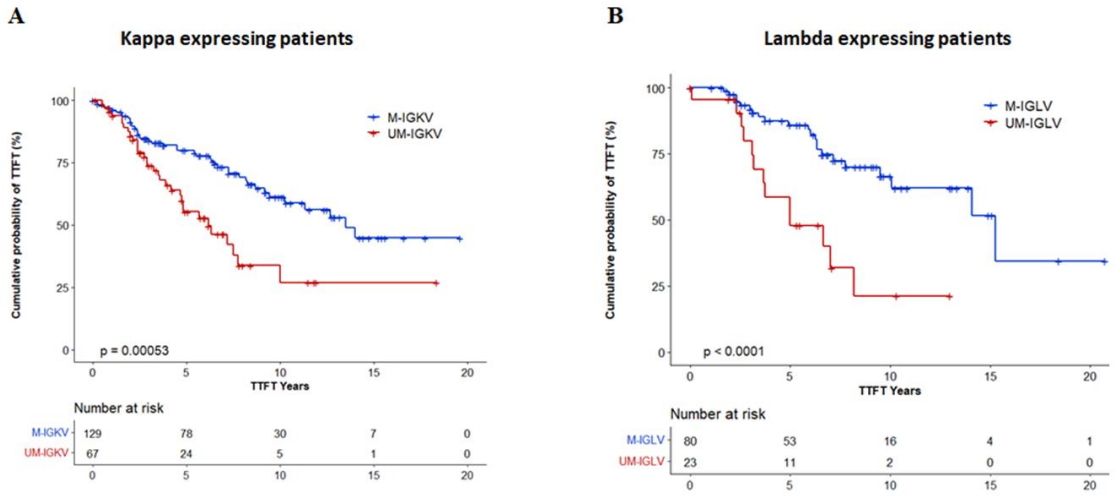

Figure 14 Light chains mutation status association with TTFT in Rai 0 CLL patients in the validation cohort. **(A)** Kaplan-Meier curve of time to first treatment for kappa expressing patients, (using the cut-off of 99%), where UM-IGKV patients are indicated in red and M-IGKV patients are represented in blue. **(B)** Kaplan-Meier curve of time to first treatment for lambda expressing patients, (using the cut-off of 99%), where UM-IGLV patients are indicated in red and M-IGLV patients are represented in blue.

3.6. Light chain mutational status independently predicted TTFT in early-stage CLL

By combining the training and the validation cohort, a total of 829 patients were obtained. The prognostic role of unmutated light chains in terms of TTFT was therefore adjusted in multivariate analysis in 673 Binet A or Rai 0 patients. Unmutated light chains (combining both kappa and lambda) independently predict shorter TTFT (HR=1.69, 95% CI 1.18-2.41, p=0.004) when adjusted for the IPS-E variables (**Figure 15**).

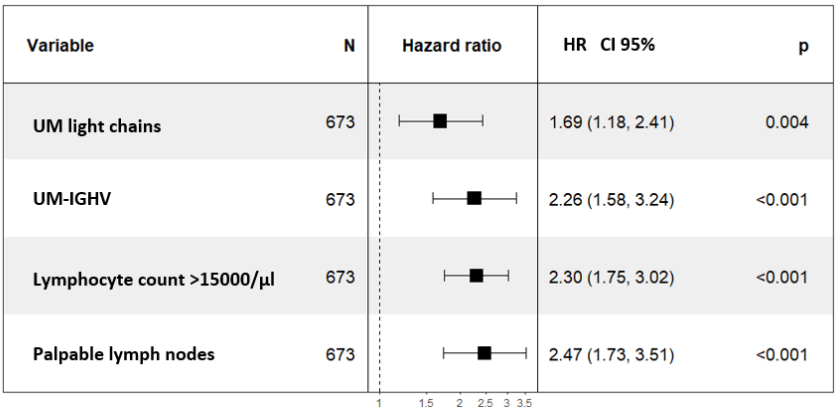

Figure 15 Forest plot multivariate analysis in terms of TTFT in early stage CLL including both cohorts (Binet A patients from the training cohort and Rai 0 from the validation cohort) depicting the light chain mutation status and the IPS-E variables.

5. DISCUSSION

In the present study, the immunoglobulin light chain gene rearrangement has been evaluated in the largest real-world cohort of unselected CLL. The somatic hypermutation of the light chain gene sorted out as an independent predictor of shorter TTFT and identified a fraction of early stage CLL patients destined to early treatment requirement, independently of the IPS-E variables.

Currently, few studies have focused on the role of light chain gene rearrangement in CLL. The first study that analysed the prevalence of light chain gene in CLL was conducted in 2005 by Stamatopoulos *et al* [54]. In the present study, the majority of CLL patients expressed kappa light chain (70%) and the rest expressed lambda chains (30%), in keeping with CLL biology as well as with other studies on healthy B cells [55]. Also, the prevalence of light chain rearrangements in the present study are similar to the ones identified by Stamatopoulos et al., as the most frequent lambda rearrangements were IGLV3-21, IGLV2-14, IGLV3-25. Similarly, in kappa expressing patients, the IGKV4-1, IGKV3(D)-20 and IGKV1-39 were the most frequently rearranged genes in both studies. In normal B-cell, no evidence for pairing of specific IGHV or IGKV/IGLV rearrangements was observed [56, 57]. In CLL, the relative skewed repertoire of IGHV rearrangements is driven by antigen stimulation from the microenvironment [58]. Consistently, in the present study some light chain rearrangements preferentially associated with unique IGHV rearrangements (i.e. IGLV3-21 with IGHV3-21 and IGKV1-39 with IGHV4-39 and IGHV1-2) suggesting that also for light chain the repertoire is strongly influenced by external specific antigen stimulation and that the combination of specific light and heavy chain rearrangements may be crucial in CLL pathogenesis.

Most of CLL cases are asymptomatic and are managed with the watch & wait strategy [22]. The identification of molecular markers in asymptomatic patients may be useful to plan the most appropriate follow-up and also to identify high-risk patients that may benefit from early interventional clinical trials. Currently, for light chain mutational status, no validated cut-off for somatic hypermutation has been established. Similar to IGHV, in the present study the closer the percentage of homology of the rearranged light chain to the germline counterpart the worse the TTFT of patients is. By using the Maxstat statistics, we propose a cut-off of 99% to predict TTFT in early stage Binet A CLL patients for kappa

and for lambda expressing patients. Interestingly, by evaluating the TTFT according to the heavy and the light chain immunoglobulin rearrangements, the TTFT prediction does not solely rely on the mutational status of the heavy chain but also on the mutational status of the light chain. Consistently, the prognostic value of light chain mutational status maintained an independent prognostic value when adjusted for the IPS-E variables (UM-IGHV, palpable lymph nodes and lymphocyte count $>15x10^9/L$) [24] indicating that light chain mutational status refines IGHV mutational status in the TTFT prediction in early stage CLL. Importantly, these findings were independently validated in a previously reported and comprehensively annotated cohort of 343 Rai 0 CLL patients [21].

In conclusion, this study represents the largest real-world cohort of unselected CLL analysed for the immunoglobulin light chain gene repertoire and further underscores the pivotal role of the BCR in CLL. The somatic hypermutation status of light chain genes independently predicts shorter TTFT. The step forward will be to assess the role of the immunoglobulin light chain gene in CLL patients treated with new biological agents, particularly within the context of fixed-duration therapy with anti-CD20 antibodies plus BCL2 inhibitors, where IGHV still retain prognostic value.

6. BIBLIOGRAPHY

1 Mukkamalla SKR, T.A., Malipeddi D, et al. *Chronic Lymphocytic Leukemia*. 2024.

2 Chiorazzi, N., S.S. Chen, and K.R. Rai, *Chronic Lymphocytic Leukemia.* Cold Spring Harb Perspect Med, 2021. **11**(2).

3 Döhner, H., et al., *Genomic aberrations and survival in chronic lymphocytic leukemia.* New England Journal of Medicine, 2000. **343**(26): p. 1910-1916.

4 Haferlach, C., et al., *Comprehensive genetic characterization of CLL: a study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgVH status and immunophenotyping.* Leukemia, 2007. **21**(12): p. 2442-2451.

5 Stilgenbauer, S., et al., *Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial.* Blood, The Journal of the American Society of Hematology, 2014. **123**(21): p. 3247-3254.

6 Calin, G.A., et al., *Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.* Proceedings of the national academy of sciences, 2002. **99**(24): p. 15524-15529.

7 Cimmino, A., et al., *miR-15 and miR-16 induce apoptosis by targeting BCL2.* Proceedings of the National Academy of Sciences, 2005. **102**(39): p. 13944-13949.

8 Klein, U., et al., *The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia.* Cancer cell, 2010. **17**(1): p. 28-40.

9 Hayakawa, K., et al., *Loss of a chromosomal region with synteny to human 13q14 occurs in mouse chronic lymphocytic leukemia that originates from early-generated B-1 B cells.* Leukemia, 2016. **30**(7): p. 1510-1519.

10 Kasar, S., et al., *Alterations in the mir-15a/16-1 loci impairs its processing and augments B-1 expansion in de novo mouse model of chronic lymphocytic leukemia (CLL).* PLoS One, 2016. **11**(3): p. e0149331.

11 Crombie, J. and M.S. Davids, *IGHV mutational status testing in chronic lymphocytic leukemia.* American journal of hematology, 2017. **92**(12): p. 1393-1397.

12 Hamblin, T.J., et al., *Unmutated Ig VH genes are associated with a more aggressive form of chronic lymphocytic leukemia.* Blood, The Journal of the American Society of Hematology, 1999. **94**(6): p. 1848-1854.

13 Damle, R.N., et al., *Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic Indicators in Chronic Lymphocytic Leukemia: Presented in part at the 40th Annual Meeting of The American Society of Hematology, held in Miami Beach, FL, December 4-8, 1998.* Blood, The Journal of the American Society of Hematology, 1999. **94**(6): p. 1840-1847.

14 Ten Hacken, E. and J.A. Burger, *Microenvironment interactions and B-cell receptor signaling in Chronic Lymphocytic Leukemia: Implications for disease pathogenesis and treatment.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2016. **1863**(3): p. 401-413.

15 Fabbri, G. and R. Dalla-Favera, *The molecular pathogenesis of chronic lymphocytic leukaemia.* Nature Reviews Cancer, 2016. **16**(3): p. 145-162.

16 Rossi, D., et al., *Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome.* Clinical Cancer Research, 2009. **15**(13): p. 4415-4422.

17 Sutton, L.-A., et al., *Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors.* haematologica, 2016. **101**(8): p. 959.

18 Rossi, D., et al., *Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia.* Blood, The Journal of the American Society of Hematology, 2012. **119**(2): p. 521-529.
19 Rai, K.R., et al., *Clinical staging of chronic lymphocytic leukemia.* Blood, 1975. **46**(2): p. 219-34.
20 Binet, J.L., et al., *A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis.* Cancer, 1981. **48**(1): p. 198-206.
21 Cohen, J.A., et al., *A laboratory-based scoring system predicts early treatment in Rai 0 chronic lymphocytic leukemia.* Haematologica, 2020. **105**(6): p. 1613-1620.
22 Hallek, M., et al., *iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL.* Blood, 2018. **131**(25): p. 2745-2760.
23 Molica, S., et al., *The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: Independent validation in a prospective cohort of early stage patients.* Am J Hematol, 2016. **91**(11): p. 1090-1095.
24 Condoluci, A., et al., *International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia.* Blood, 2020. **135**(21): p. 1859-1869.
25 Su, Q., et al., *Cryo-EM structure of the human IgM B cell receptor.* Science, 2022. **377**(6608): p. 875-880.
26 Ma, X., et al., *Cryo-EM structures of two human B cell receptor isotypes.* Science, 2022. **377**(6608): p. 880-885.
27 Hombach, J., et al., *Molecular components of the B-cell antigen receptor complex of the IgM class.* Nature, 1990. **343**(6260): p. 760-2.
28 Campbell, K.S. and J.C. Cambier, *B lymphocyte antigen receptors (mIg) are non-covalently associated with a disulfide linked, inducibly phosphorylated glycoprotein complex.* Embo j, 1990. **9**(2): p. 441-8.
29 Shaw, A.C., et al., *Mutations of immunoglobulin transmembrane and cytoplasmic domains: effects on intracellular signaling and antigen presentation.* Cell, 1990. **63**(2): p. 381-92.
30 Venkitaraman, A.R., et al., *The B-cell antigen receptor of the five immunoglobulin classes.* Nature, 1991. **352**(6338): p. 777-81.
31 Victora, G.D. and M.C. Nussenzweig, *Germinal Centers.* Annu Rev Immunol, 2022. **40**: p. 413-442.
32 Lam, K.P., R. Kühn, and K. Rajewsky, *In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death.* Cell, 1997. **90**(6): p. 1073-83.
33 Reth, M. and J. Wienands, *Initiation and processing of signals from the B cell antigen receptor.* Annu Rev Immunol, 1997. **15**: p. 453-79.
34 Schamel, W.W. and M. Reth, *Monomeric and oligomeric complexes of the B cell antigen receptor.* Immunity, 2000. **13**(1): p. 5-14.
35 Tolar, P., et al., *The constant region of the membrane immunoglobulin mediates B cell-receptor clustering and signaling in response to membrane antigens.* Immunity, 2009. **30**(1): p. 44-55.
36 Treanor, B., et al., *The membrane skeleton controls diffusion dynamics and signaling through the B cell receptor.* Immunity, 2010. **32**(2): p. 187-99.
37 González, D., et al., *Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma.* Blood, The Journal of the American Society of Hematology, 2007. **110**(9): p. 3112-3121.
38 Tanaka, S. and Y. Baba, *B cell receptor signaling.* B Cells in immunity and tolerance, 2020: p. 23-36.

39 Hsueh, R.C. and R.H. Scheuermann, *Tyrosine kinase activation in the decision between growth, differentiation, and death responses initiated from the B cell antigen receptor.* 2000.

40 Mansouri, L., et al., *Different prognostic impact of recurrent gene mutations in chronic lymphocytic leukemia depending on IGHV gene somatic hypermutation status: a study by ERIC in HARMONY.* Leukemia, 2023. **37**(2): p. 339-347.

41 Knisbacher, B.A., et al., *Molecular map of chronic lymphocytic leukemia and its impact on outcome.* Nature genetics, 2022. **54**(11): p. 1664-1674.

42 Oakes, C.C., et al., *DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia.* Nature genetics, 2016. **48**(3): p. 253-264.

43 Darwiche, W., et al., *Chronic lymphocytic leukemia B-cell normal cellular counterpart: clues from a functional perspective.* Frontiers in immunology, 2018. **9**: p. 683.

44 Agathangelidis, A., et al., *Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies.* Blood, The Journal of the American Society of Hematology, 2012. **119**(19): p. 4467-4475.

45 Ghia, P., et al., *Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: the lesson of the IGHV3-21 gene.* Blood, 2005. **105**(4): p. 1678-1685.

46 Murray, F., et al., *Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis.* Blood, The Journal of the American Society of Hematology, 2008. **111**(3): p. 1524-1533.

47 Bomben, R., et al., *Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study.* British journal of haematology, 2009. **144**(4): p. 492-506.

48 Agathangelidis, A., T. Chatzikonstantinou, and K. Stamatopoulos, *B cell receptor immunoglobulin stereotypy in chronic lymphocytic leukemia: Key to understanding disease biology and stratifying patients.* Semin Hematol, 2023.

49 Jaramillo, S., et al., *Prognostic impact of prevalent chronic lymphocytic leukemia stereotyped subsets: analysis within prospective clinical trials of the German CLL Study Group.* Haematologica, 2020. **105**(11): p. 2598.

50 Tobin, G., et al., *Somatically mutated Ig VH3-21 genes characterize a new subset of chronic lymphocytic leukemia.* Blood, The Journal of the American Society of Hematology, 2002. **99**(6): p. 2262-2264.

51 Maity, P.C., et al., *IGLV3-21* 01 is an inherited risk factor for CLL through the acquisition of a single-point mutation enabling autonomous BCR signaling.* Proceedings of the National Academy of Sciences, 2020. **117**(8): p. 4320-4327.

52 Duez, M., et al., *Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing.* PLoS One, 2016. **11**(11): p. e0166126.

53 Manso, T., et al., *IMGT® databases, related tools and web resources through three main axes of research and development.* Nucleic Acids Res, 2022. **50**(D1): p. D1262-d1272.

54 Stamatopoulos, K., et al., *Immunoglobulin light chain repertoire in chronic lymphocytic leukemia.* Blood, 2005. **106**(10): p. 3575-83.

55 Creyssel, R., et al., *The frequency distribution of heavy chain classes and light chain types of 1,000 monoclonal immunoglobulins.* Biomedicine, 1975. **22**(1): p. 41-8.

56 Brezinschek, H.P., et al., *Pairing of variable heavy and variable kappa chains in individual naive and memory B cells.* J Immunol, 1998. **160**(10): p. 4762-7.

57 de Wildt, R.M., et al., *Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire.* J Mol Biol, 1999. **285**(3): p. 895-901.

58 Kostareli, E., et al., *Immunoglobulin gene repertoire in chronic lymphocytic leukemia: insight into antigen selection and microenvironmental interactions.* Mediterr J Hematol Infect Dis, 2012. **4**(1): p. e2012052.