



HAL
open science

Développement de nouvelles molécules pour caractériser la leucémie lymphocytaire chronique.

Xuan Lan

► **To cite this version:**

Xuan Lan. Développement de nouvelles molécules pour caractériser la leucémie lymphocytaire chronique.. Médecine humaine et pathologie. Normandie Université, 2022. Français. NNT : 2022NORMR085 . tel-04033754

HAL Id: tel-04033754

<https://theses.hal.science/tel-04033754>

Submitted on 17 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité ASPECTS MOLECULAIRES ET CELLULAIRES DE LA BIOLOGIE

Préparée au sein de l'Université de Rouen Normandie

Development of new molecular tools to characterise Chronic lymphocytic leukemia

Présentée et soutenue par

XUAN LAN

Thèse soutenue le 05/12/2022
devant le jury composé de

MME MARIE-HÉLÈNE DELFAU LARUE	PROFESSEUR DES UNIV - PRATICIEN HOSP., UNIVERSITE PARIS-EST CRETEIL	Rapporteur du jury
MME BRIGITTE GUBLER	PROFESSEUR DES UNIV - PRATICIEN HOSP., UNIVERSITE AMIENS PICARDIE JULES VERNE	Rapporteur du jury
M. DAVID RIZZO	MAITRE DE CONF UNIV. - PRATICIEN HOSP., UNIVERSITE LIMOGES	Président du jury
M. FABRICE JARDIN	PROFESSEUR DES UNIV - PRATICIEN HOSP., Université de Rouen Normandie	Directeur de thèse
M. HERVÉ TILLY	PROFESSEUR DES UNIV - PRATICIEN HOSP., Université de Rouen Normandie	Co-directeur de thèse

Thèse dirigée par FABRICE JARDIN (CANCER AND BRAIN GENOMICS) et HERVÉ TILLY (CANCER AND BRAIN GENOMICS)

TABLE OF MATERIAL

INDEX OF FIGURES AND TABLES	2
RESUME (French).....	7
ABSTRACT (English).....	8
INTRODUCTION.....	10
I. General Information of CLL	10
A. The development of CLL.....	10
B. The risk classification of CLL.....	11
C. The therapeutic strategies of CLL.....	16
II. Basic biomarkers in CLL	20
A. Gene mutation	20
B. Chromosome aberration.....	21
C. The clinical value of micro-RNA in CLL	22
III. The importance of <i>IGHV</i> in CLL	24
A. The normal development of B-cell.....	25
B. The development of BCR in CLL progress and its clinical value	26
C. The routine method to test the mutational status of <i>IGHV</i>	32
D. The research on the light chain in CLL	35
IV. The function of <i>tp53</i> aberration in CLL	38
A. The role of <i>tp53</i> in cancer	38
B. Alternative splicing.....	38
C. The research on the isoform of <i>tp53</i> in cancer	39
D. The routine method of testing the expression of p53 isoforms	42
V. The object of this thesis	45
MATERIALS AND METHODS.....	46
I. Patients and samples.....	47
II. Preparation of RNA.....	47
III. 5'RACE.....	47
IV. RT-MLPA.....	53
V. QMPSF	56
VI. Statistical analysis.....	56
RESULTS.....	58

I. Relationship between the mutational status of <i>IGHV</i> and <i>IGLV</i> and the clinical value of <i>IGLV</i>	59
A. Problem.....	59
B. Results	60
1) The stable of 5'RACE in testing the mutational status of <i>IGHV</i>	60
2) The mutational relationship between <i>IGHV</i> and <i>IGLV</i>	74
3) The value of <i>IGLV</i>	75
C. Conclusion and discussion	83
II. The expression of <i>tp53</i> at mRNA level in CLL and the influence of other important biomarkers.....	85
A. The critical biomarkers in CLL	85
1) Problem.....	85
2) Results.....	86
B. The function of <i>tp53</i> in CLL	92
1) Problem.....	92
2) Results.....	92
C. The influence of isoforms of <i>tp53</i> in CLL	94
1) Problem.....	94
2) Results.....	94
3) Conclusion and discussion	102
CONCLUSION AND GENERAL DISCUSSION	105
I. Conclusion	106
II. General discussion	106
A. NGS in test <i>IGHV</i> in CLL	106
B. Using a more straightforward method to test the expression of <i>tp53</i> mRNA in CLL	109
ANNEXE	111
REFERENCES	150

INDEX OF FIGURES AND TABLES

FIGURES

Figure 1: Diagnosis and differentiation of CLL.	10
Figure 2: The risk factors of the different risk staging systems.	13
Figure 3: Analysis of overall survival classified by the CLL-IPI.	15
Figure 4: For fit patients, the long-term benefit of first-line FCR regimens (fludarabine, cyclophosphamide, and rituximab).	17
Figure 5: The different pathways of new target drugs.	19
Figure 6: The frequency of chromosome aberrations in CLL.	21
Figure 7: the relationship among miR-15/16, 13q14-deleted region, and the expression of <i>BCL2</i>	24
Figure 8: The rearrangement of VDJ in the heavy chain and VJ in the light chain.	25
Figure 9: The progress of CLL evolution, from antigen (Ag) stimulation to molecular abnormalities.	26
Figure 10: The process of maturation of B cell.	27
Figure 11: Characteristics of UM-CLL and M-CLL patient subsets.	28
Figure 12: The information of major and other subsets in CLL.	30
Figure 13: The sequence comparison from subsets #6 and subset #2 in CLL.	31
Figure 14: The sequence comparison from subsets #8 and #4 in CLL.	32
Figure 15: Some problematic cases indicated by ERIC.	33
Figure 16: The process of Biomed-2.	34
Figure 17: Ca ⁺ signaling by CLL subset #2 IgM BCRs depends on a specific mutation from the germline sequence.	36
Figure 18: The usage of heavy chains in IGLV3-21 with or without point mutation.	37
Figure 19: The mechanisms of the alternative transcription and alternative splicing.	39
Figure 20: Human <i>tp53</i> Isoforms.	40
Figure 21: The promoters of p53 and each sequence of isoform.	42
Figure 22: The specific antibodies of p53 isoforms in humans.	43
Figure 23: Procedure of 5' RACE.	51

Figure 24: The principle of the RT-MLPA.....	54
Figure 25: The results of two methods.	64
Figure 26: Situation of SHM and CSR in each productive rearrangements.	68
Figure 27: Relationship between <i>IGHV</i> and <i>IGLV</i> regions and the mutational statuses of different <i>IGHVs</i>	69
Figure 28: The relationship of the mutational status of heavy and light chains.....	75
Figure 29: Sequence of all <i>IGLV3-21</i> alleles, including the <i>IGLV3-21</i> , <i>IGLV3-21^{R110}</i> , and subset 2 <i>IGLV3-21^{R110}</i> alleles.....	76
Figure 30: The information on <i>IGLV3-21</i>	76
Figure 31: The identification of <i>IGHV</i> and <i>IGLV</i> in different groups.	78
Figure 32: Identifying <i>IGHV</i> in M-CLL non- <i>IGLV3-21^{R110}</i> , <i>IGLV3-21^{R110}</i> , and U-CLL non- <i>IGLV3-21^{R110}</i>	79
Figure 33: The survival analysis of different groups.	80
Figure 34: The different characteristics of the two methods.....	84
Figure 35: Common genetic markers in CLL.....	86
Figure 36: The characteristic of each important genetic molecule in CLL.	87
Figure 37: The relationship between expression of <i>tp53</i> and different prognostic markers.	88
Figure 38: The survival analysis of different prognostic markers.....	91
Figure 39: The structure of <i>tp53</i>	94
Figure 40: The stable of RT-MLPA used in CLL.	95
Figure 41: TP53 splicing isoforms of SU-DHL-16 cell line characterized by RT-MLPA and RT-PCR.....	96
Figure 42: The expression of each exon junction in 114 CLL patients.	97
Figure 43: The results of RT-MLPA in a CLL patient with splicing site mutation.	99
Figure 44: The markers expressed a lower level of exon 4-5.....	99
Figure 45: The survival analysis of different groups classified by the expression of exon4-5.	100
Figure 46: The markers expressed a lower level of exon9-9b and exon9bg-10.....	100
Figure 47: The survival analysis of different groups classified by the expression of exon 9-9b.	101

Figure 48: The survival analysis of different groups classified by all the important biomarkers in CLL..... 102

TABLES

Table 1: Different immunotyping to diagnosis by the expression of surface markers on lymphoid cells.....	11
Table 2: Classic risk stratification systems for CLL.....	12
Table 3: First and second-line therapy strategies of CLL in healthy and unhealthy patients	18
Table 4: Clinical predictive biomarkers of CLL.....	20
Table 5: Summaries the clinical and molecular characteristics of the most common stereotyped subsets.....	29
Table 6: Suggestions for reporting the <i>IGHV</i> mutational status for various types of sequences	35
Table 7: p53 isoforms: their cellular effects and appearance in various cancer types.	41
Table 8: RT Primers.....	47
Table 9: The temperature of the first amplification.....	48
Table 10: Nested I2 Primers.....	49
Table 11: The temperature of the second amplification.....	49
Table 12: Primer of the search pattern.....	52
Table 13: The probes of aim exon of <i>tp53</i>	55
Table 14: 5' RACE and Biomed-2 Results.....	60
Table 15: The results of patients tested by 5'RACE at different times.....	65
Table 16: The top 5 usage of IgκV in CLL patients.....	70
Table 17: The top 5 usage of IgλV in CLL patients.....	70
Table 18: List of light chains that share identical CDR3 in 80 CLL patients.....	71
Table 19: Compare two methods.....	73
Table 20: The relationship of the mutational status of heavy and light chains.....	74
Table 21: The sequence of all the IGLV3-21 CLL patients.....	77
Table 22: Clinical information of CLL patients.....	81
Table 23: The clinical information of 114 CLL patients.....	89
Table 24: Multiple linear regression model of the expression of exon junction.....	97

ABBREVIATIONS

5'RACE	5' Rapid Amplification of cDNA ends
BCR	B-cell Receptor
CDR	Complementarity Determining Region
CLL	Chronic Leukemia Lymphoma
CNV	Copy Number Variation
ECOG	Eastern Cooperative Oncology Group
FFPE	Formalin-Fixed Paraffin-Embedded
FISH	Fluorescent In-Situ Hybridization
<i>IGHV</i>	Immunoglobulin Heavy Chain Variable region gene
<i>IGLV</i>	Immunoglobulin Light Chain Variable region gene
LDH	Lactate Dehydrogenase
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
QMPSF	Quantitative Multiplex PCR of Short Fluorescent Fragments
RT-MLPA	Reverse-Transcriptase Multiplex Ligation-dependent Probe Amplification
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
UMI	Unique Molecular Identifier

RESUME (French)

La leucémie lymphocytaire chronique (LLC) est très répandue dans le monde occidental et se concentre sur les hommes âgés. La maladie présente une hétérogénéité importante, ce qui rend particulièrement important de trouver des biomarqueurs et des tests stables pour guider le pronostic et les options de traitement de la maladie. Ce projet de recherche doctorale vise à améliorer l'étude des biomarqueurs dans la LLC par le séquençage de nouvelle génération (NGS). D'une part, nous avons développé une nouvelle approche pour détecter le statut mutationnel de l'*IGHV*, un biomarqueur pronostique essentiel dans la LLC. Par rapport à la méthode de routine actuelle, le 5'RACE est plus sensible pour traiter les réarrangements multiples de l'*IGH*. De plus, grâce à la supériorité du NGS, nous pouvons obtenir simultanément des informations sur la chaîne légère de l'immunoglobuline (IGL). Par conséquent, en combinaison avec les connaissances sur les chaînes légères, nous avons constaté que la plupart du temps, la chaîne lourde présentait un statut d'identité mutationnelle avec la chaîne légère. Cette découverte confirme que la chaîne légère peut également guider la stratégie diagnostique et thérapeutique de la LLC, en particulier dans les cas limites. Nous avons également confirmé l'existence d'un groupe spécial présentant des mutations ponctuelles en position 110 de l'*IGLV3-21*, dont le pronostic est aussi mauvais que celui des patients non mutés, même si la plupart de leurs *IGHV* sont mutés.

D'autre part, la fonction de *tp53*, un autre biomarqueur important de la LLC, n'est pas très claire. Pour étudier la relation entre l'épissage alternatif de *tp53* et l'aberration de *tp53*, nous avons détecté l'expression de chaque exon-jonction de *tp53* pour obtenir la situation de l'épissage alternatif par transcription inverse, amplification de sonde multiplexée par ligature (RT-MLPA) combinée au séquençage à haut débit. Nous avons également comparé les résultats de NGS et FISH au niveau de l'ADN de l'aberration *tp53* avec les résultats de RT-MLPA au niveau de l'ARNm. Un patient présentant une mutation du site d'épissage testé par NGS a également montré un épissage alternatif anormal entre les exons 4 et 5 détecté par RT-MLPA. Le plus intéressant est que certaines jonctions d'exon spéciales ont montré des niveaux significativement différents chez les patients de bon et de mauvais pronostic. Et lorsque nous combinons d'autres biomarqueurs pronostiques de la LLC pour établir un système de notation, le niveau d'expression de ces jonctions d'exon spéciales concorde également avec le pronostic. Ces jonctions d'exon appartiennent spécifiquement à un type d'isoformes de *p53*, ce qui nous rappelle le rôle critique des isoformes de *p53* dans la LLC.

C'est la première fois que l'on utilise le NGS pour détecter la chaîne lourde et légère de l'immunoglobuline et l'épissage alternatif de *tp53* au niveau du transcriptome. À l'ère du NGS, nos résultats soulignent l'efficacité et la précision des méthodes permettant de découvrir davantage d'informations sur la LLC. Notre étude a mis en évidence l'importance des isoformes d'IGL et de *p53* dans la LLC et leur valeur pronostique et thérapeutique en clinique.

Mots clés : Leucémie lymphocytaire chronique, immunoglobuline, récepteur des cellules B, 5'RACE, RT-MLPA, *tp53*.

ABSTRACT (English)

Chronic lymphocytic leukemia (CLL) is highly prevalent in the Western world and is concentrated in older men. The disease exhibits significant heterogeneity, making it particularly important to find stable biomarkers and assays to guide the prognosis and treatment options for the disease. This Ph.D. research project aims to improve the study of biomarkers in CLL by next-generation sequencing (NGS). On the one hand, we developed a novel approach to detecting the mutational status of the *IGHV*, an essential prognostic biomarker in CLL. Compared with the current routine method, 5'RACE is more sensitive to handling multiple rearrangements of IGH. Moreover, due to the superiority of NGS, we can obtain information on the immunoglobulin light chain (IGL) simultaneously. Therefore, combined with the knowledge of light chains, we found that mostly, the heavy chain showed mutational identity status to the light chain. This finding support that the light chain can also guide the diagnostic and therapeutic strategy of CLL, especially in borderline cases. We also confirmed a special group with point mutations in position 110 of IGLV3-21, which showed a similarly poor prognosis to *IGHV*-unmutated patients, even though most of their *IGHV* are mutated.

On the other hand, the function of *tp53*, another important biomarker of CLL, isn't very clear. To investigate the relationship between alternative splicing of *tp53* and *tp53* aberration, we detected the expression of each exon-junctions of *tp53* to get the situation of alternative splicing by reverse transcription multiplex ligation-dependent probe amplification (RT-MLPA) combined with high-throughput sequencing. We also compared the results of NGS and FISH at the DNA level of *tp53* aberration with the results of RT-MLPA at the mRNA level. One patient with splicing site mutation tested by NGS also showed abnormal alternative splicing between exons 4 to 5 detected by RT-MLPA. The most exciting thing is that some special exon-junctions showed significantly different levels in favor and poor prognosis patients. And when we combine other prognostic CLL biomarkers to establish a scoring system, the expression level of these special exon-junctions is also concordant with the prognosis. These exon junctions specifically belong to one kind of p53 isoforms, which reminds us of the critical role of p53 isoforms in CLL.

It is the first time using NGS to detect the immunoglobulin heavy and light chain and the alternative splicing of *tp53* at the transcriptome level. In the era of NGS, our results emphasized the efficiency and accuracy of methods to discover more information about CLL. Our study highlighted the importance of IGL and p53 isoforms in CLL and their prognostic and therapeutic value in the clinic.

Keywords: Chronic lymphocytic leukemia, immunoglobulin, B-cell receptor, 5'RACE, RT-MLPA, *tp53*.

INTRODUCTION

INTRODUCTION

I. General Information of CLL

A. The development of CLL

The general approach to the workup of lymphocytosis.

Chronic lymphocytic leukemia (CLL) is the most popular adult leukemia in the western world, mainly found in older males¹ (the ratio of males to females is around two, and the median age at diagnosis is around 72 years old). It is diagnosed based on the clonal expansion of mature cells marketed by CD5+ and CD23+² detected in the blood, bone marrow, lymph nodes, and spleen, and $> 5 \times 10^9/L$ of lymphocytes in the peripheral blood persisting for more than three months (**Figure 1**). The expression of surface immunoglobulins³, CD19, CD20, and CD79b was significantly lower in CLL than in normal B cells⁴ (**Table 1**). Each clone of CLL cells expresses κ or/and λ immunoglobulin light chains⁴. And CLL shows significant disease heterogeneity, as some patients are found to only have increased lymphocyte counts by annual physical examination and survive for a long time without appropriate treatment. In contrast, some patients experience rapid disease progression and respond poorly to treatment regimens. Therefore, for the last two decades, researchers have been working to better identify relevant therapeutic and prognostic predictors to understand this disease's development and provide precise treatments for patients.

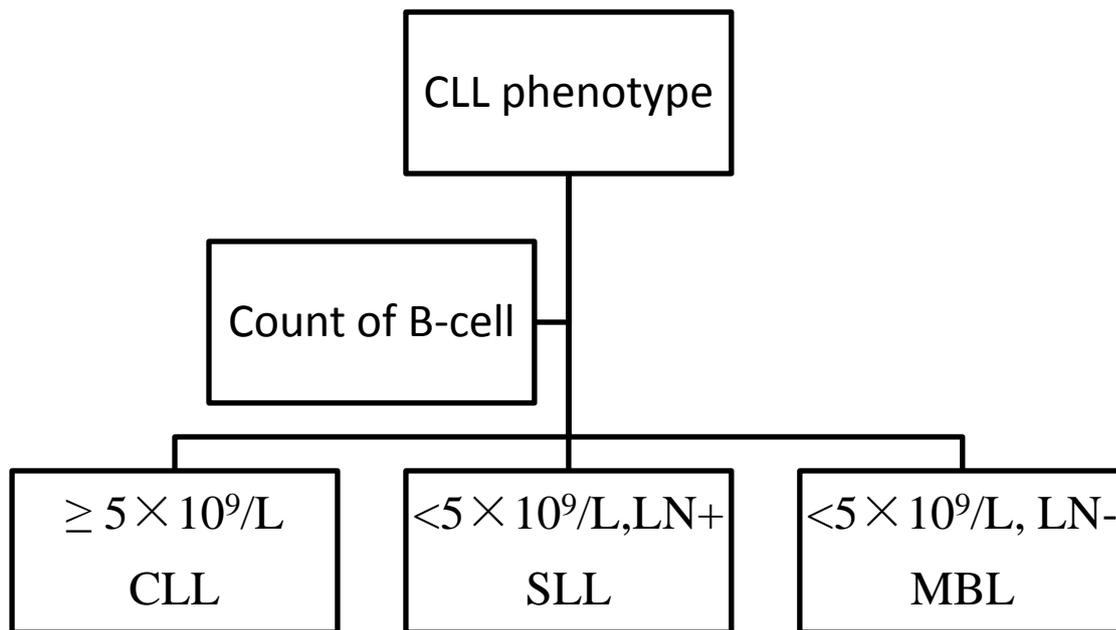


Figure 1: Diagnosis and differentiation of CLL.

After the appearance of lymphocytosis, the relevant and complete tissue physical examination and blood tests are performed. These tests aim to determine the monoclonal nature of the cells by flow cytometry and detecting the chronic lymphocytic leukemia phenotype, which is then classified according to the B-cell count as MBL, CLL, and SLL. MBL: monoclonal B lymphocytosis; SLL: Small lymphocytic lymphoma.

B lymphocytes with absolute B cell count below $5 \times 10^9/L$ and without cytopenias, lymphadenopathy, hepatomegaly, or splenomegaly are called monoclonal B lymphocytosis (MBL). Recent research has shown that the expression levels of CD38, ZAP70, - amongst others in MBL, which progresses to CLL at a rate of 1-2% per year⁵, are lower than in CLL. In comparison, the levels of corresponding prognostic biomarkers are significantly higher, making MBL a prodromal state of CLL³. It is also associated with an increased risk of infection and second malignancies⁶.

Table 1: Different immunotyping to diagnosis by the expression of surface markers on lymphoid cells.

	slg	CD5	CD23	CD79b	FMC7	CD20	CD22	CD103	CD200	CD25	ROR1
Normal	High	No	No	High	High	High	High	No	No	Low	No
CLL	Low	High	High	Low	Low	Low	Low	No	Very High	Low	High
MCL	High	High	No	High	Very High	Very High	High	No	Low	No	High
FL	Very High	No	No	High	Low	High	Low	No	No	No	Low
HCL	Very High	No	No	Low	High	High	Very High	High	No	Very High	High
MZL	High	No	No	High	High	High	High	No	No	High	Low

FL: Follicular lymphoma¹

B. The risk classification of CLL

As previously mentioned, one of the significant characteristics of CLL is its highly heterogeneous nature, which makes risk stratification for the disease especially important. The two backbone of CLL clinical staging systems are RAI⁷, and Binet⁸, which have played a considerable role in the treatment choice and prognostic evaluation in the past four decades (**Table 2**). In these two staging systems, patients can be reliably classified into different risk groups by physical examination and basic clinical laboratory tests alone, without manipulating imaging and molecular biology, which can equally apply in poorly equipped medical facilities. Moreover, in the past two decades, with the development of medical research and technology, we have found that CLL with highly heterogeneous requires us to search for relevant biomarkers. These biomarkers were combined with two clinical staging systems to enhance the understanding of CLL's genetic and molecular biology, determine the risk of disease progression, and complement the information of classical staging systems⁹.

Table 2: Classic risk stratification systems for CLL.

Rai⁷			
Risk status	Stage	Characteristics	Median survival
Low risk	0	Lymphocytosis	>150 months
Intermediate risk	I	Stage 0 + lymphadenopathies	Eight years
	II	Stage 0 +splenomegaly and /or hepatomegaly associated with or without lymphadenopathies	Six years
High risk	III	Anemia, with serum hemoglobin level <11g/dL	19 months
	IV	Thrombocytopenia, platelet count <100,000/mm ²	19months
Binet⁸			
Low risk	A	< 3 involved nodal areas enlarge	>15 years
Intermediate risk	B	≥ 3 involved nodal areas enlarge	7-8years
High risk	C	Hemoglobin < 10g/dL and/or platelets < 10×10 ⁴ /μL	5-6 years

The most critical biomarkers include the mutational status of the *IGHV* and *tp53* aberrations. More and more studies have demonstrated that the mutational status of *IGHV* can predict the prognosis of CLL patients stable and independently. Patients with mutated *IGHV* showed a slow disease progression and a better prognosis. In contrast, patients with unmutated *IGHV* had a more rapid disease progression and suggested the necessity of early treatment¹⁰. However, due to the 98% threshold of the *IGHV*, a group of patients expressed 97-99% identity, which we called “borderline” cases, can’t always predict the correct outcome. The treatment and prognostic evaluation of this group of patients should pay attention to and require more research.

Another critical biomarker is *tp53* aberrations, which include the mutations of the *tp53* gene and deletion of chromosome 17 short arm. *tp53* aberrations also can predict rapid disease progression, poor prognosis, and an inadequate response to chemoimmunotherapy (CIT)¹¹. The mechanisms involved need to be further explored. The expression of p53 isoforms has been gaining attention recently. The method for stable detection of the expression levels of different isoforms remains to be explored.

Other relevant prognostic factors include the expression levels of molecules CD38, ZAP70^{12,13}, gene mutations such as *NOTCH1* and *SF3B1*^{9,14}, and changes in serum concentrations of β2-microglobulin¹⁵. In an international prospective randomized clinical trial study, 3472 patients were recruited to develop a regular prognostic index for CLL. Five independent prognostic factors were identified, including *tp53* aberrations, mutational status of *IGHV*, the concentration

of serum β 2-microglobulin, clinical stage, and age. After grading these five independent factors, patients were classified into four risk groups with significantly different survival rates¹⁶. The reliability of this classification method was further validated in subsequent studies^{17,18}. It, therefore, provides an excellent theoretical basis for further individualized and precise treatment of patients.

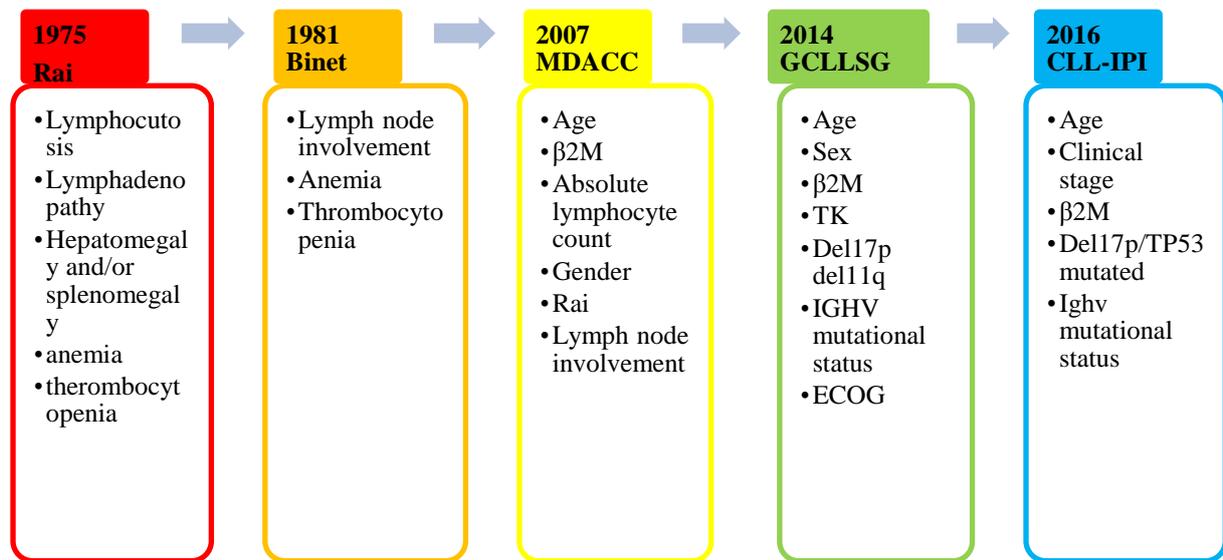


Figure 2: The risk factors of the different risk staging systems.

Each risk system is described in the figure, including the Rai stage, Binet stage, MDACC nomogram, GCLLSG, and CLL-IPI.

As we can see from the previous summary, the RAI and Binet staging system are the most widely used classical risk stratification methods in clinical practice^{7,8}. Both systems are based on routine and inexpensive physical examination findings, including peripheral blood, lymph nodes, liver, and spleen, to classify patients into different prognostic groups and provide a reliable basis for other predictive models. However, these two classifications, developed 40 years ago, also have some drawbacks. First, they cannot determine whether early-stage patients will develop aggressive disease progression. They do not incorporate promising biomarkers discovered recently, such as molecules found by immunofluorescence assays or NGS technologies. Hence, both staging systems have a minimal prediction for treatment option selection and response to therapy. Several centers have continuously updated their predictive models for CLL through their strengths (**Figure 2**). The most classic include MD Anderson Cancer Center in 2007¹⁹, 2010²⁰, and 2012²¹ through three further optimizations, adding mutational status of *IGHV*, del17p, Binet staging, β 2M, sex, and age in an integrated model to estimate overall survival.

After two years, GCLLSG published a new prognostic scoring system that integrated sex, age, Eastern Cooperative Oncology Group (ECOG) status, del17, mutational status of *IGHV*, s- β 2M, and serum thymidine kinase (s-TK) into four risk categories²². This novel classification system remedied our previous analysis of the two cornerstone stratification systems by emphasizing the role of biomarkers in CLL. The classification system was subsequently validated by other

experimental groups²³. Because early management of s-TK levels was not widespread, investigators proposed a revised version that did not include such tests. And based on seven other independent prognostic factors, the revised version still proved to be a stable predictor for patients with early CLL at the time of the first treatment²⁴.

A widely used risk scoring system in clinical management was developed in 2016 by combining clinical parameters and cellular biological molecules, namely the CLL International Prognostic Index (CLL-IPI)¹⁶ (**Figure 3**). It includes the mutational status of *IGHV*, *tp53*, level of serum β 2-microglobulin, clinical stage, and age. This system classifies patients into four risk groups by these five independent prognostic factors. The CLL-IPI has considered cellular biological molecular abnormalities compared to the previous risk system. Several studies have confirmed this system to provide stable predictive value for overall survival (OS) time, time to first treatment (TTFT), free survival time, and response to CIT¹⁷ in newly diagnosed patients^{25,26}. Because detection of *IGHV* mutational status is not available universally²⁷, researchers have replaced *IGHV*-unmutated class with a doubling of lymphocyte count within six months because of their relationship¹⁸. Compared to *IGHV* detection, the LDT doubling time is easier and cheaper to measure, and CLL-LIPI plays a significant role in the absence of *IGHV* mutation status results. It should be mentioned that this single-center study enrolled only 218 patients, and CLL-LIPI could predict TTFT but not time to next treatment (TTNT). In addition, CLL-IPI is more sensitive in predicting overall outcome²⁸.

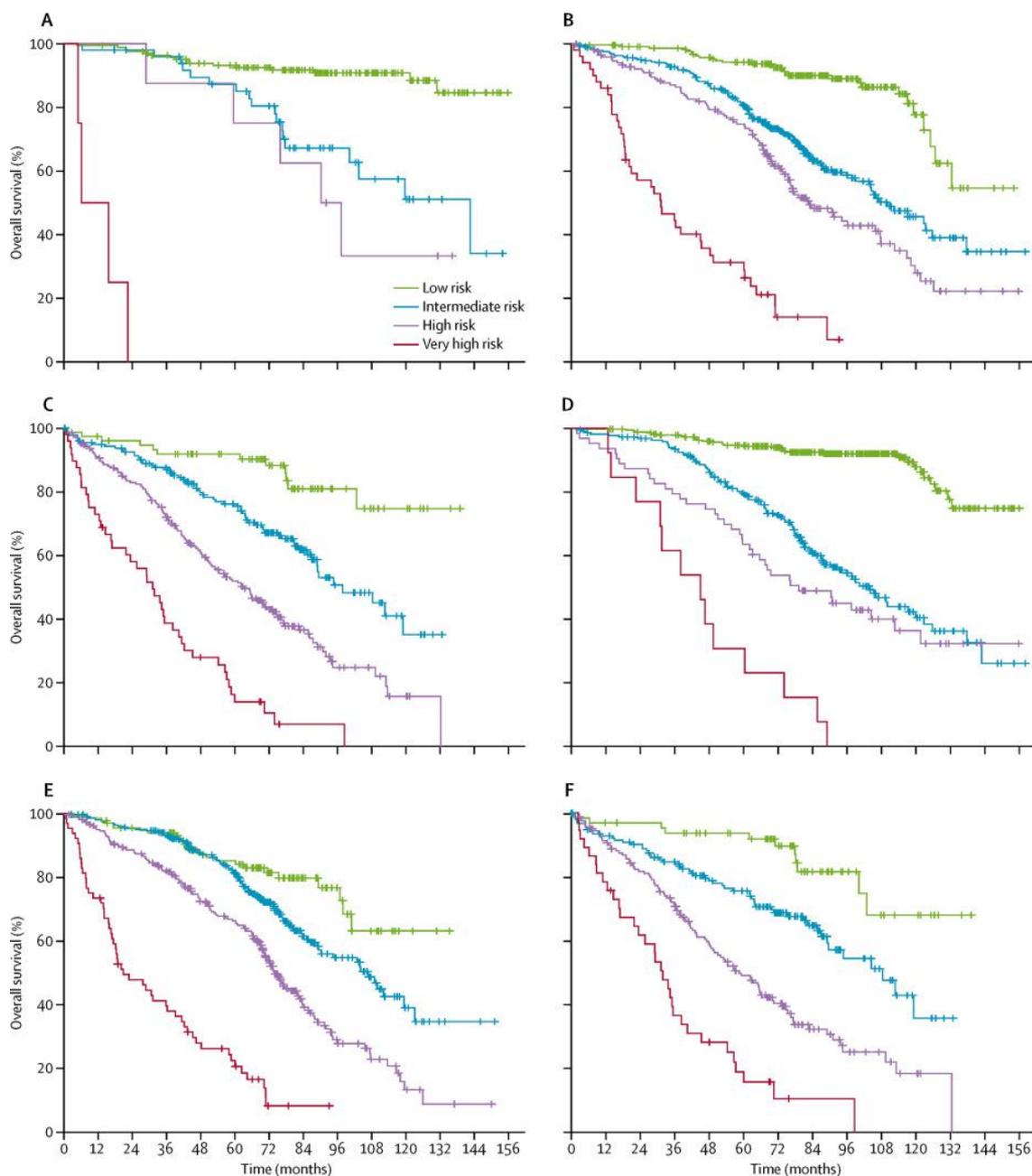


Figure 3: Analysis of overall survival classified by the CLL-IPI.

(A) Rai stage 0. (B) Rai stage I–II. (C) Rai stage III–IV. (D) Binet stage A. (E) Binet stage B. (F) Binet stage C. Adapted from the International CLL IPI working group (2016)¹⁶.

With the development of NGS technology in recent years, more somatic mutations and molecular pathways have been detected in CLL. Besides these biomarkers' ability to help us stratify risk for patient prognosis, the analysis of genetic molecules in CLL has improved our understanding of the therapeutic targets^{29,30}. It has been reported clinically significant improvements in specific patients with novel target inhibitors^{31,32}. Moreover, cellular immunotherapy, including chimeric antigen receptor T cells (CAR-T) and allogeneic stem cell transplantation (allo-SCT), have also been used in high-risk patients. However, the emergence of novel drugs targeting different pathways, including the Bruton tyrosine kinase (BTK)

inhibitors ibrutinib, acalabrutinib, and zanubutinib, the phosphatidylinositol 3-kinase (PI3K) inhibitor idelalisib, and the *BCL-2* inhibitor venetoclax, has revolutionized the treatment approach and prognostic evaluative system³³. In the era of novel targeted drugs, researchers found that these prognostic factors, including unmutated *IGHV*, *tp53* aberration, and del11q, have no predictive value in relapsed/refractory CLL patients who receive targeted therapy³⁴, which emphasized the need to improve the methods to understand the relationship between target drugs and CLL. It has also been shown that the clinical impact of high-risk features in the era of new drugs is not the same as in the era of CIT, with only del17p/*tp53* mutations and CD49d expression remaining indicators of poorer prognosis³⁵. *Tp53* aberrations play an important role in predicting PFS of patients treated with ibrutinib ± rituximab³⁶ or obinutuzumab + venetoclax³⁷. Therefore, the predictive value of traditional biomarkers needs to be further investigated, and the relevant assays for biomarkers should be more easy, rapid, and individualized.

C. The therapeutic strategies of CLL

It is crucial to decide the necessary and choose timing of treatment for CLL patients, which still determines through the staging system of RAI and Binet³. Early asymptomatic patients include RAI stage 0 and Binet stage A patients, who can engage in a watch-and-wait management program until disease progression. At the same time, relevant studies have shown that early-stage patients can't get benefits from early intervention, while applying chemotherapeutic drugs may cause specific side effects to patients³⁸. However, for patients with advanced stages, including RAI stage 3-4 and Binet stage C patients, as well as patients with clinical symptoms, including B symptoms (unexplained weight loss, 10% weight loss within six months, fever of 38 degrees or more, night sweats lasting more than one month without evidence of infection), and doubling of lymphocyte count within a short period, early therapeutic interventions will provide benefits to patients.

The choice of therapeutic strategy has also improved with the growing understanding of the disease. Immunotherapy combined with FCR is the first-line treatment of physically fit CLL patients^{11,39}. Early improvements in survival time in younger patients were mainly achieved through the combination of purines plus cyclophosphamide⁴⁰. The introduction of CD20 monoclonal antibodies has increased overall survival in CLL patients¹¹. At the same time, treatment regimens with conventional chemotherapeutic agents in combination with CD20 monoclonal antibodies have also shown significant outcomes in older patients⁴¹, prolonging PFS and increasing overall and complete response. In the CLL8 trial, patients with *IGHV* unmutated, 13q deletion, 11q deletion, or trisomy 12 also benefited from this treatment combination and were further able to achieve durable remission. Thus FCR chemotherapy regimens have become the first-line treatment option for most CLL patients (**Figure 4**).

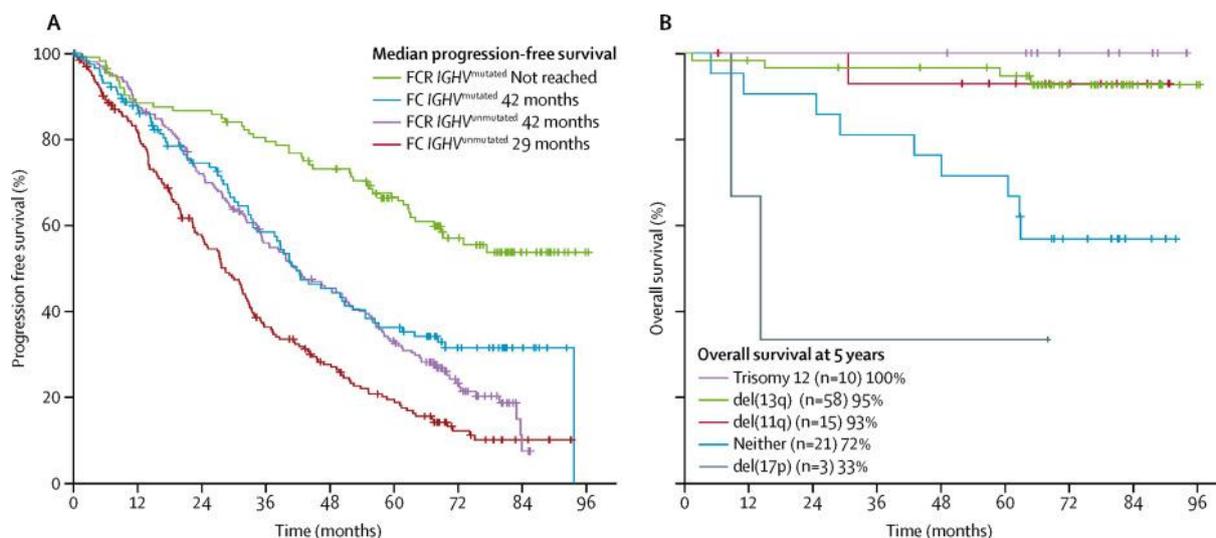


Figure 4: For fit patients, the long-term benefit of first-line FCR regimens (fludarabine, cyclophosphamide, and rituximab).

(A) Analyse the Progression-free survival of patients with FCR according to *IGHV* status in phase 3 study, including 817 patients for frontline therapy with a median observation time of 5.9 years⁴². (B) The overall survival of patients with FCR with mutated *IGHV* status in phase 3 study, including 817 patients for frontline therapy with a median observation time of 5.9 years⁴².

Tp53 aberration often predicts an aggressive disease process and poor response to FCR regimens^{17,43} (**Table 3**). In the last decade, with the increasing research on the molecular level of CLL, new targeted drugs have gradually appeared to be efficacious and approved for the first-line treatment of CLL. As a BTK inhibitor, Ibrutinib binds to the C481 residue of the BTK kinase structural domain and irreversibly inhibits BTK. Large-scale clinical trials have shown good benefits of ibrutinib in multiple CLL subgroups, including relapsed or refractory cases, del17p⁴⁴, and elderly patients, with an overall response of approximately 60-90% and showing durable remission. In *tp53* aberrations CLL patients, the best therapeutic results were achieved by ibrutinib combined with idelalisib^{44,45}. Despite this progress, *tp53* aberrations in leukemic clones still seem to harm prognosis compared to wild-type patients^{44,45}. With a quickly developing treatment field, making an excellent therapeutic strategy and choosing the best initial treatment time for patients has become a task requiring experience and extensive clinical trials.

Table 3: First and second-line therapy strategies of CLL in healthy and unhealthy patients

First line treatment		
	Without <i>TP53</i> aberration	<i>TP53</i> aberration
Physically fit	Fludarabine plus cyclophosphamide plus rituximab (age ≤65 years); or bendamustine plus rituximab (age >65 years)	Ibrutinib or idelalisib plus rituximab or venetoclax (if ibrutinib therapy is not suitable because of comorbidities or comedication)
Physically unfit	Chlorambucil plus obinutuzumab; or chlorambucil plus ofatumumab; or chlorambucil plus rituximab; or ibrutinib monotherapy	Ibrutinib or idelalisib plus rituximab or venetoclax (if ibrutinib is not suitable because of comorbidities or comedication)
Second line treatment		
	Standard therapy	Alternative therapies
Refractory or progression within three years		
Physically fit	Ibrutinib possibly followed by an allogeneic hemopoietic stem-cell transplant	Idelalisib plus rituximab; venetoclax; lenalidomide
Physically unfit	Change therapy	Ibrutinib; idelalisib plus rituximab; venetoclax; lenalidomide; CD20 antibody alone
Progression after three years		
All fitness levels	Repeat front-line therapy	Change to another chemoimmunotherapy; ibrutinib; idelalisib plus rituximab

Factors that should be considered include body composition, age, comorbidities, and genetic status of *tp53*. With the development of large international clinical trials, researchers found that the rate of complete remission (CR) with monotherapy of ibrutinib was not very high. Less than 5% of patients achieved CR and 15% of therapeutic non-responders, and 20% of relapsed/refractory patients experienced disease progression after monotherapy. Ibrutinib also requires sustained dosing until disease progression or the occurrence of intolerable drug toxicity causes a financial burden to patients. These limitations have emphasized the complicated molecular pathway in CLL and the need to develop a suitable combination of therapeutic strategies. Researchers have developed combination regimens for ibrutinib based on different signaling pathways. Anti-CD20 monoclonal antibodies^{31,45}, FCR or bendamustine-rituximab (BR) regimens, CAR-T cell therapies, and novel targeted inhibitors, such as *BCL-2* inhibitors (venetoclax) and PI3K inhibitors (idelalisib and duvelisib), have been studied. In various

clinical trials, the combination regimen of ibrutinib with venetoclax has shown better efficacy in both first-accept-treatment and relapsed/refractory CLL patients, improving CR rates and reducing minimal residual disease (MRD)⁴⁶ rates.

Another class of targeted agents, the third-generation *BCL-2* antagonist venetoclax, is a potent BH3 mimetic that selectively binds and neutralizes *BCL-2* and has had a favorable response in the treatment of CLL. High-risk cases with unfavorable prognostic factors, including *TP53* aberrations^{47,48}, unmutated *IGHV*, and therapy with previous 3rd-line therapy⁴⁹, or fludarabine tolerance, especially in relapse/refractory (R/R) CLL patients, appear unaffected by treatment. Venetoclax has recently been approved for treating patients with relapsed *tp53* aberrations. Also, in Europe, clinical trials of its combination with CD20 monoclonal antibody have shown some favorable results⁴⁸. It will be approved for the treatment of relapsed CLL⁵⁰ (**Figure 5**).

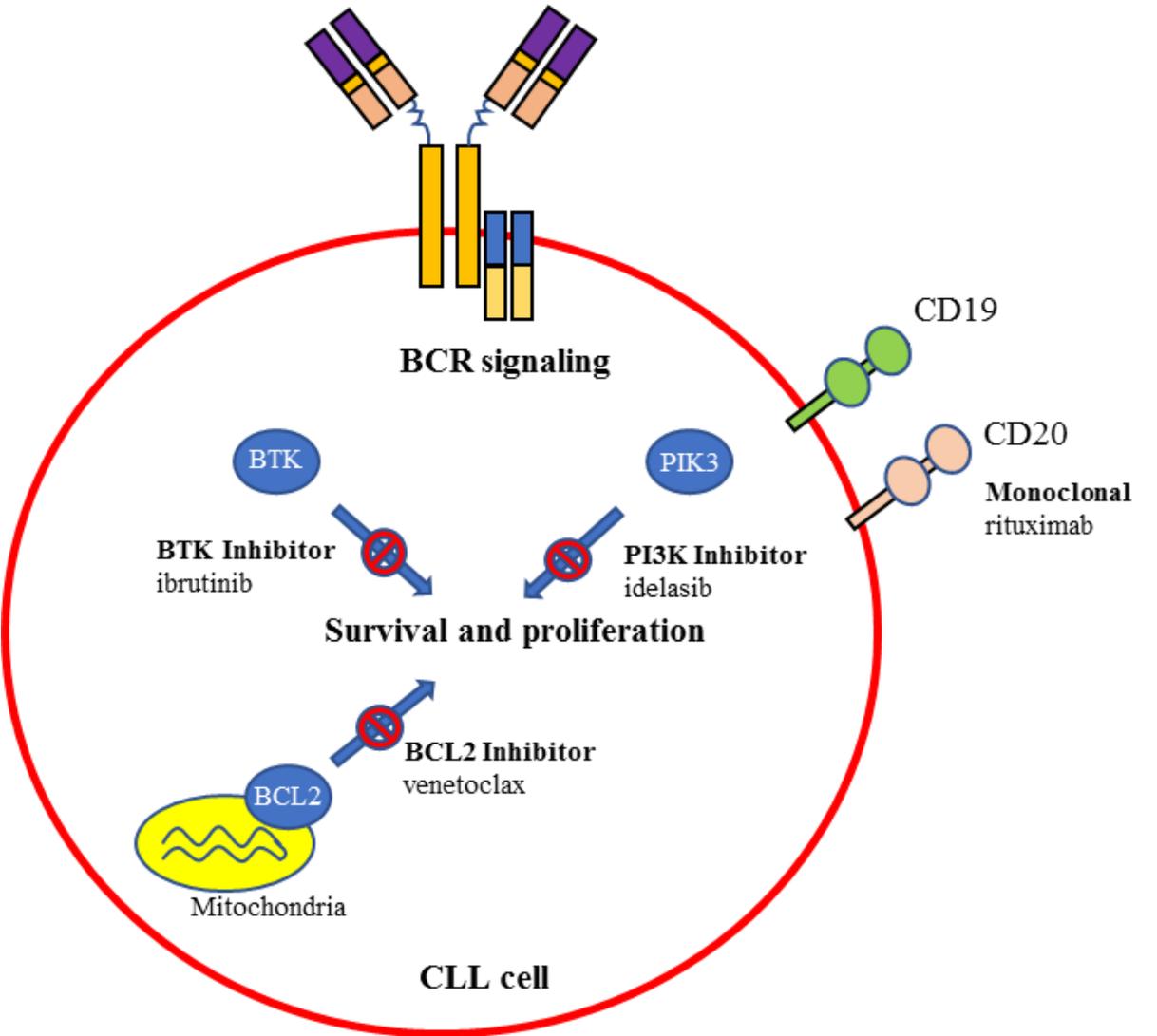


Figure 5: The different pathways of new target drugs.

II. Basic biomarkers in CLL

The progression of CLL is not fully understood at this stage regarding which factors contribute to the ongoing disease progression from MBL to CLL and later deterioration to Richter syndrome. Considering CLL-related cytogenetic abnormalities have been revealed in the MBL, we also have to reconsider the pathogenic mechanisms of various classical molecular biological abnormalities and whether we can detect the disease progression to control the disease progression from an early stage. At this time of the whole genome sequencing (WGS), more and more new genes are being revealed, including *MYD88*¹⁴, *NOTCH1*¹⁴, and *SF3B1*⁵¹, but these genes are less mutated overall CLL cases, and there are no significant associations between them. It is worth noting that there is often a close correlation between gene mutations and chromosomal abnormalities, such as between *SF3B1* and 11q⁵¹, *NOTCH1*, and trisomy 12⁵². As evidenced in several studies, gene mutations that recurrently appear in CLL cases seem to play an essential role in disease progression, finally leading to the transformation to malignant lymphoma^{53,54}. Therefore, exploring these genes and their interrelationships will reveal the molecular biological basis of CLL progression. These molecules also emphasize using different techniques and experimental approaches for studying the molecular mechanisms in CLL (**Table 4**).

Table 4: Clinical predictive biomarkers of CLL.

Different kinds of biomarkers	Clinical outcomes
<i>TP53</i> aberrations (del 17p and/or <i>TP53</i> mutation)	Poor response to CIT
Mutated <i>NOTCH1</i>	Poor response to anti-CD20 therapy
Mutational status of <i>IGHV</i>	Predict a long-term remission in younger, fit patients with M- <i>IGHV</i>
Complex karyotypes	Poor response to CIT
Micro-RNAs	MiR-15/16: associates with <i>BCL-2</i> ; MiR-34a: associates with chemotherapy-refractory disease; MiR-155: predicts therapy response
CD49d	Inhibits cell trafficking in the setting of novel BCR target therapy

A. Gene mutation

Thanks to extensive sample cohort studies, somatic mutations that recurrently appear in CLL are receiving increasing attention. The signaling pathways involved in these mutated genes and

their roles have also been investigated. These genes function in pathways including DNA damage, mRNA transcription, chromosome modification, WNT signaling, NOTCH signaling, and inflammatory pathways. Although there is variability in data from different study cohorts, two large international collections examined and characterized over 500 CLL samples from around the world for whole gene or whole exon sequencing. Since one of the studies analyzed pre-treatment CLL samples requiring initial treatment and the other one analyzed early CLL patients and MBL patients, the percentage of frequently mutated genes has increased. As a result, there was a slight difference in the rate of frequently mutated genes, including *NOTCH1* (6% and 12.6%, respectively), *ATM* (15% and 11%, respectively), *SF3B1* (21% and 8.6%, respectively), and *BIRC3* (4% and 8.8%, respectively)^{9,55}. In addition to these most frequently mutated genes, these extensive sample cohort studies also provided data to support new CLL candidate genes, which include MYD88, XPO1, KLHL6, and IRAK1¹⁴. Meanwhile, mutations in the noncoding region of some genes have also been found to affect patient survival. One study reported that patients with mutations in the noncoding region at the 3' end of *NOTCH1* showed a poorer overall survival and suggested the need for early therapeutic intervention⁵⁶.

Another advantage of NGS is the ability to analyze the heterogeneity between individual cells within CLL tumors. Reports show that some somatic mutations, such as MYD88, are often expressed in all CLL tumor cells, suggesting that these genes are altered early in the development of leukemia. In contrast, other genes, such as *NOTCH1* and *SF3B1*, are usually present only in a subset of CLL tumor cells and therefore represent subclonal populations of cells, suggesting that their mutational events occur later in the development of CLL. And notably, it was reported in subsequent studies that usually mutated genes of subclonal type are more aggressive, especially when multiple mutations occur^{9,57}. Moreover, the increased proportion of *tp53* aberrations after receiving chemotherapy also suggests that CLL tumor cells undergo clonal metastasis while indicating the importance of such genes in studies related to treatment regimens and drug resistance⁹.

B. Chromosome aberration

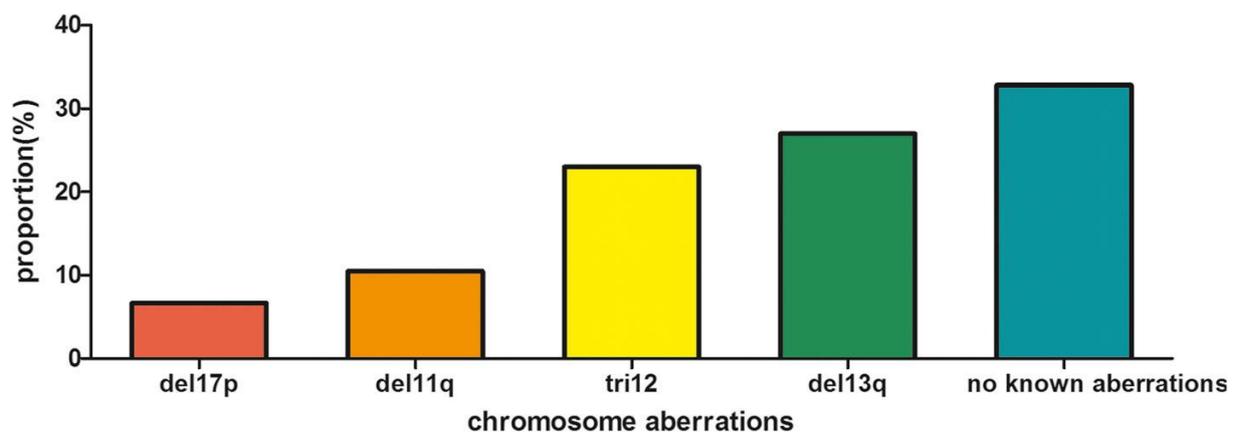


Figure 6: The frequency of chromosome aberrations in CLL.

In addition to mutations in hotspot genes, chromosomal abnormalities including deletion of the long arm of chromosome 13 (del13q), trisomy of chromosome 12 (tri12), deletion of the long

arm of chromosome 11 (del11q), and deletion of the short arm of chromosome 17 (del17p) have been extensively studied in CLL patients (**Figure 6**).

The most common chromosomal abnormality in CLL patients is **the deletion of 13q (del13q)**, occurring in 50-60% of CLL patients^{9,43}. Interestingly, del13q is primarily found in patients with mutated *IGHV*, suggesting that most patients show a favorable prognosis. However, it is essential to note that more than 60% of patients with del13q need to receive early therapeutic intervention.

Following del13q, the second widespread chromosomal abnormalities are **trisomies 12 (tri12)**, which account for approximately 10% of CLL patients and are assessed as intermediate risk cytogenetic aberrations⁴³. Tri12 is heterogeneous in terms of clinical presentation and biological characteristics. When the proportion of cells containing tri12 increases, patients show a poorer prognosis⁵⁸, while tri12 can co-presence with other trisomic chromosomal abnormalities in some patients⁵⁹, TTFT is prolonged, and the risk of recurrence is reduced⁶⁰.

The deletion of 11q (del11q) represents a poorer prognosis, although clinical studies show shortened TFS with no effect or prolonged OS in del11q patients⁶¹. However, most studies have reported that the lower the frequency of del11q, the more favorable the prognostic outcome of patients⁶². For patients after receiving allogeneic HSCT, del11q can more consistently predict shorter PFS and OS⁶³.

The deletion of 17p (del17p) is mainly detected in 5-8% of chemotherapy-naïve CLL patients and usually suggests a poorer prognosis⁶⁴. Del17p could affect the therapeutic efficacy of CD20 monoclonal antibodies, which can lead to resistance or rapid relapse in CLL patients⁶⁵. Due to properties such as cell subcloning, the less ratio of del17p cells, the longer TTFT the patient will have⁶⁶. The results of another study indicate that, in addition to del17p, mutation of *tp53* and complex karyotypes (CK) can also affect TTFT, suggesting that these biological factors have a limited role in early stage development in CLL and mainly act on related molecular activities after disease relapse⁶⁷.

Complex karyotypes (CK) are associated with a poor disease prognosis, defined as the existence of at least three cytogenetic abnormalities⁶⁸. However, it has been found that not all complex karyotypes have an equal impact on disease⁶⁹. Analyzed by a large clinical trial, patients were classified into subgroups based on the number of abnormal genetic carriage. Disease prognosis was consistently predicted between subgroups, independent of *IGHV* and *TP53* mutation status independently⁷⁰.

Besides the chromosomal abnormalities commonly seen in CLL patients described above, other chromosomal abnormalities can also impact the prognosis of CLL, such as del6q, del9p, and del10q with full or partial triplication of chromosomes⁷¹. However, because of the low percentage of these chromosomal abnormalities, large-scale clinical trials are needed later to verify the biological significance of the molecules involved.

C. The clinical value of micro-RNA in CLL

Besides directly affecting the survival of patients, chromosomal abnormalities can contribute to the aggression and progression of CLL by affecting the expression of other small molecules. Some findings showed that del13q affects the expression of mir-15 and mir-16⁷², negatively regulating the expression of *BCL-2* by interacting with the 3'UTR (**Figure 7**). Thus the loss of mir-15/16 is the leading cause of *BCL-2* and *ZAP70* overexpression in CLL⁷³. The current study showed that mir-15/16 deleted or downregulated in approximately 60% of CLL patients and exhibits dysfunction in a minority of cases. And these two small molecules also target *MCL1*⁷⁴ and affect the expression of anti-apoptotic proteins of the *BCL-2* family. Thus, the reduced level of these micro RNAs results in increased expression of the relative target genes that encode proteins which would protect tumor cells from apoptosis⁷⁴ and enhance BCR signaling¹².

Moreover, the most favorable mouse model equally supports our speculation. New Zealand black mice, who exhibit low expression levels of mir-16-1⁷⁵, appear the characters of monoclonal lymphoproliferative expansion by increasing CD5⁺ B220^{dull} B cells. In contrast, the complete emergence of CLL occurs in the mouse's later life. This animal model not only emphasizes the critical role of the deleted region of 13q14 for disease progression but also suggests a strong correlation between this deleted region and the early development of the disease, the progression from MBL to CLL. Also, considering the prolonged pathogenesis of the mouse model, it suggests that aberrant gene expression, in addition to disrupting cell growth, also acts synergistically with antigenic stimulation to induce clonal expansion of tumor cells⁷⁶. In a subsequent study, this conclusion was supported by the human CLL metastasis model established in NOD/Sci-scid-IL2rg^{-/-} mice⁷⁷. At the same time, autologous CD4⁺ T cells activated in vivo by alloantigens have a relevant role in promoting CLL cell survival and expansion⁷⁸.

Moreover, researchers have focused on micro-RNAs expressed in subpopulations with unique clinical or biological characteristics⁷⁹. For example, miR-223 and miR-125a were associated with *BCL-2* genes and STAT3 pathway⁸⁰. Loss or reduced expression of miR-29a/b/c⁸¹, miR-34a⁸², miR-150⁸³, and miR-3676 can reduce the target of 3' UTR of *TCL1A*⁸⁴, which enhances the expression of *TCL1A* while promoting the development of CLL in transgenic mice⁸⁵. Conversely, increased expression of mir-155 is associated with enhanced BCR signaling, B-cell proliferation, and lymphoma development⁸⁶.

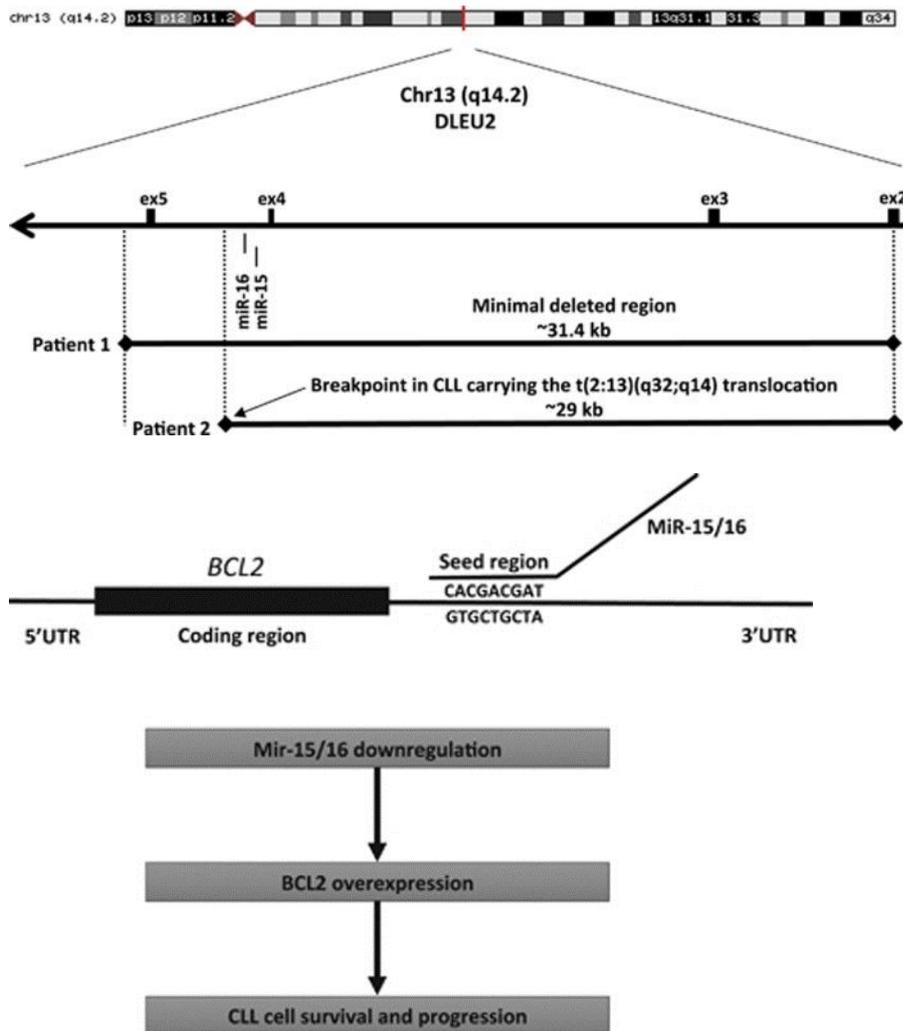


Figure 7: the relationship among miR-15/16, 13q14-deleted region, and the expression of *BCL2*.

III. The importance of *IGHV* in CLL

To investigate the evolutionary process of CLL, researchers have observed that in MBL, which represents the early stage of CLL, patients carry minimal genetic abnormalities, and some patients do not have molecular biological alterations. Therefore, for 1-2% of MBL patients who eventually transform into CLL, besides found out the key drivers, the stimulation in the microenvironment may also support the continuous expansion and deterioration of subclonal cells through various signaling pathways, the most critical of which is the antigenic stimulation of B-cell receptors. Over the past 30 years, the continued progress of studies emphasized the importance of the mutational status of *IGHV* on disease prognosis. The most exciting thing is that more than 30% of CLL patients expressed similar sequences in the CDR3 of *IGHV*, according to which we classified patients into different subsets, each with its molecular biology and prognostic survival characteristics^{30,87}. These findings reaffirm that the mutational status of *IGHV* and the sequences of antigenic epitopes in the BCR play a vital role in tumor development. Other experiments have confirmed that core receptors such as CD40/Toll-like receptors (TLRs) play an active role in acquiring full activation of CLL cells⁸⁸. Therefore, a deeper understanding of the BCRs will allow exploring the key induction events for MBL development, further

understanding the molecular mechanisms of different subgroups of CLL, and identifying promising therapeutic strategies.

A. The normal development of B-cell

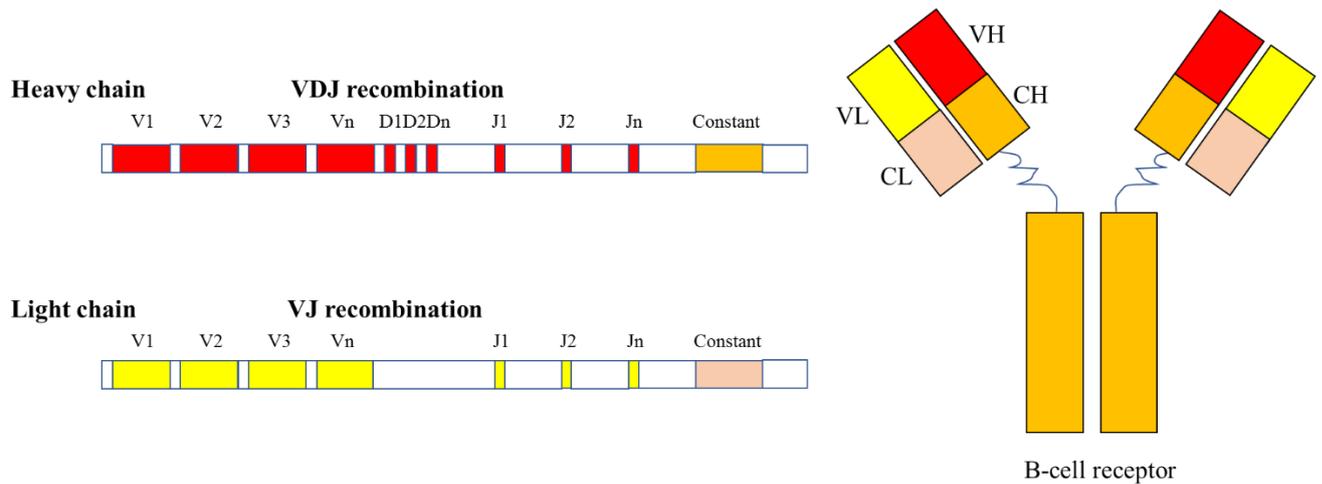


Figure 8: The rearrangement of VDJ in the heavy chain and VJ in the light chain.

In the pro- and pre-B cell stages, the rearrangement of VDJ regions of heavy chain and VJ regions of light chain can generate the mature BCR. Many VDJ (VJ) regions allow for generating a large variability of BCRs. The full BCR is made up of two variable heavy (V_H) and light (V_L) chains, respectively, which are responsible for binding the antigen, as well as two constant heavy (C_H) and light (C_L) chains involved in effector functions.

Maturation of B lymphocytes

The maturation of B lymphocytes can be divided into several stages. Firstly, pro-B cells are generated in primary lymphoid tissues, including the liver of fetal, bone marrow, and adult bone marrow⁸⁹, after which the immature lymphocytes enter through peripheral tissues to stimulate further maturation in second lymphoid tissue, including lymph nodes and spleen. The most critical aspect of B lymphocyte maturation is the completion of rearrangement of immunoglobulin heavy and light chains⁹⁰. First, during the progenitor B-cell and precursor B-cell phases, the V-D-J segment of the heavy chain is rearranged to produce a highly specific antigen-binding HCDR3 region⁹¹. At this time, the immature BCR express on the surface of the precursor B-cells. After B cells enter the differentiation phase, the V-J segment of the immunoglobulin light chain is rearranged to form a complete BCR IgM heavy chain expression. Due to the numerous VDJ segments of the heavy chains and VJ segments of light chains, the immune system can generate approximately 3×10^{11} BCRs after the rearrangement⁹² (Figure 8).

After finishing the rearrangement, B lymphocytes undergo negative selection, eliminating lymphocytes expressing self-reactive BCRs, and non-self-reactive cells that are not eliminated continue to develop and evolve into mature B cells with IgD. IgM/IgD+ mature B lymphocytes, through lymphatic circulation, enter lymph node germinal centers(GC), where the most critical molecular events, including somatic hypermutation (SHM) in the *IGHV* and *IGLV* and class switch recombination (CSR), happened. While the process of CSR induces DNA recombination

within chromosomes, replacing immunoglobulin constant regions and generating different antigens in response to varying infections⁹².

B. The development of BCR in CLL progress and its clinical value

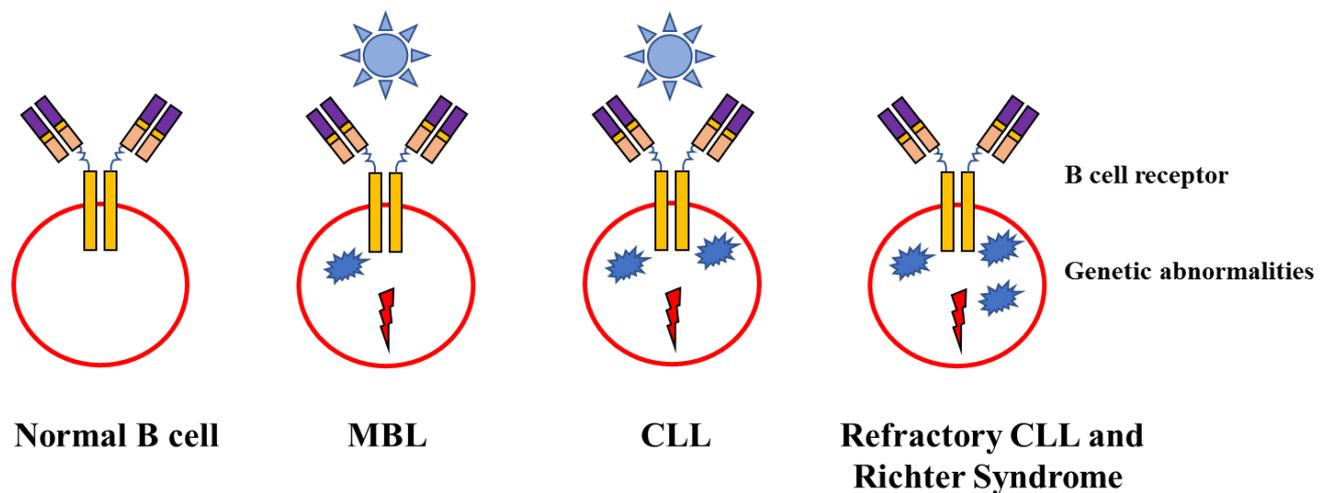


Figure 9: The progress of CLL evolution, from antigen (Ag) stimulation to molecular abnormalities.

MBL is considered to be a preclinical stage of CLL. In clinical practice, MBL is more common than CLL, with only 1-2% of MBL progressing to CLL. In CLL, only a tiny percentage of patients transform to advanced disease, evolving to Richter's syndrome or undergoing prolymphocytic transformation (**Figure 9**). It has been suggested that CLL patients may carry an intrinsic abnormality in the earliest stages of the disease. This abnormal B-cell encounter with an appropriate external stimulus triggers the clonal development of MBL and may enhance cellular stimulation.

Another possibility is that external stimuli may precede the appearance of genetic abnormalities. Over time, the pressure of the stimulus leading to enhanced proliferation favors the acquisition of molecular abnormalities, and some MBLs may develop into apparent CLL. More aggressive CLL is acquired if and when the acquired genetic abnormality (e.g., relapse can replace the necessity of external stimulation by shocking the same final pathway (either in the form of refractory disease or Richter syndrome or in the form of proto form of lymphocytic transformation).

The diagnostic feature of CLL is the accumulation of clonal CD5⁺ mature B cells. As we concluded in part on B cell maturation, the essential processes for B cell maturation and normal differentiation are SHM and CSR, so any instability affecting these two processes will increase the possibility of malignant transformation of the cells⁹³. With the results of several studies demonstrating that CLL is a BCR-associated immune cell malignant transforming disease⁹⁴⁻⁹⁶, the study of SHM in the *IGHV* has helped to track tumor transformation stages. Because of the independence of *IGHV* from clinical staging and other biological markers and the maintenance of a stable state over time has made it possible to achieve patient classification, which has been continuously demonstrated in the last 20 years⁹⁷.

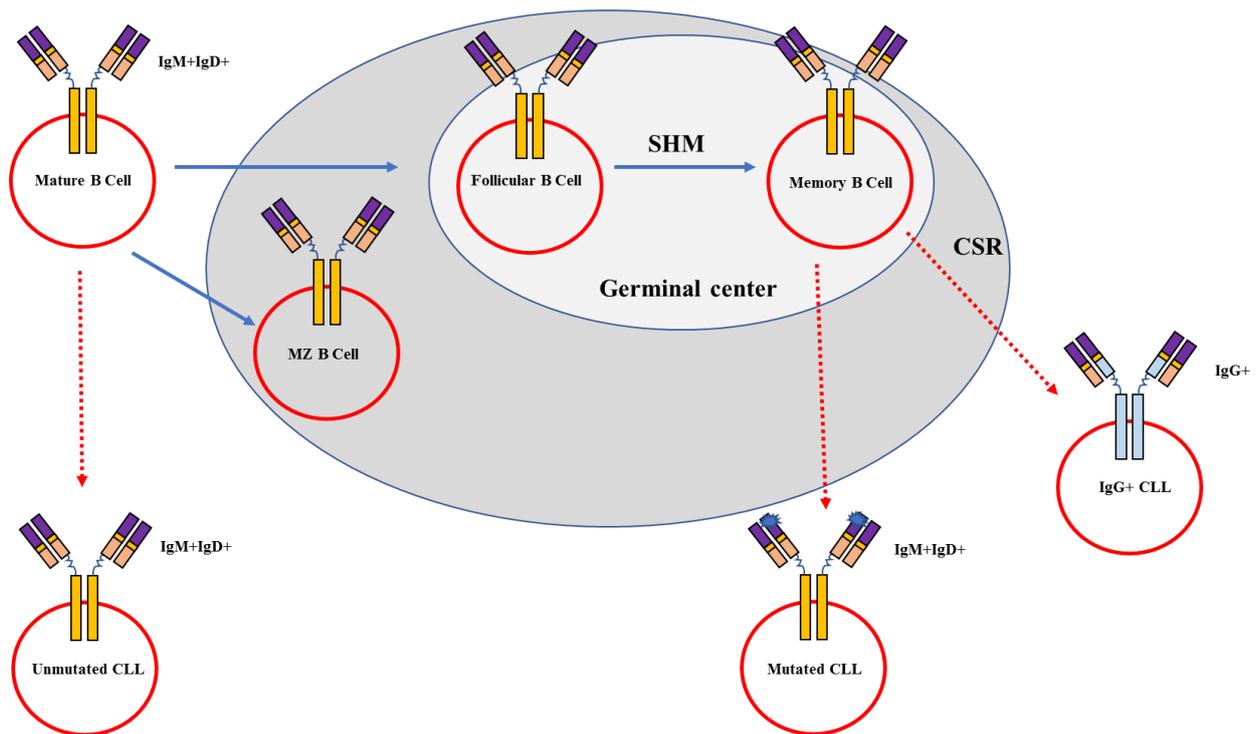
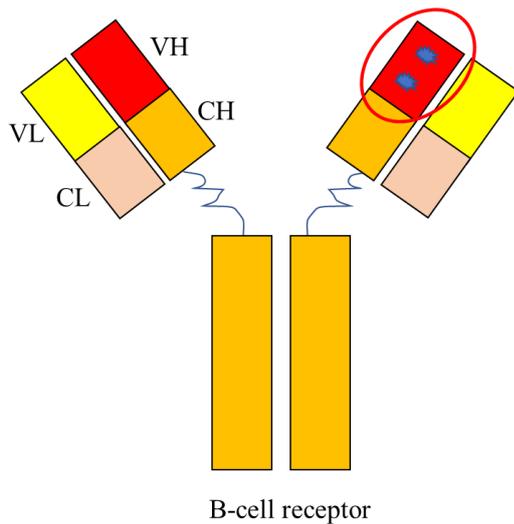


Figure 10: The process of maturation of B cell.

In the bone marrow, B cells undergo a series of steps to produce mature B cells that express IgM and IgD isotype receptors on their surface. B cells then travel with the peripheral circulation to secondary lymphoid organs to continue maturation. After antigenic stimulation, BCRs increase their diversity through somatic hypermutation (SHM) and class switch recombination (CSR). CLL can arise from different stages of B cell abnormalities. U-CLL is derived from naive B cells expressing unmutated *IGHV*, while M-CLL is derived from memory B cells undergoing SHM. A smaller, IgG-expressing subpopulation can arise from memory B cells that have undergone CSR⁹⁸.

Somatic mutational status of *IGHV*

Based on a comparison with the germline gene, the cut-off value for SHM was defined as 2%. Patients were classified as mutated CLL (M-CLL) when the cellular mutation was $\geq 2\%$ and as unmutated CLL (UM-CLL) when the percentage of somatic mutations was lower than 2% (**Figure 11**). Many studies have shown that M-CLL patients have a slower disease progression and are more sensitive to conventional chemotherapeutic agents than UM-CLL patients^{99,100}. In addition to the clinical differences between the two groups of CLL patients, it has been found that the two types of cells in different groups belong to different origins and are only morphologically similar. UM-CLL cells were derived from pre-GC naive B cells with low differentiation maturity (**Figure 10**). It highly expressed IgM and ZAP70 on the cell surface, and increased tyrosine phosphorylation levels further promote CD38 expression, which supports expansion and avoids apoptosis of tumor cells. In contrast, M-CLL cells showed anaphylactic responsiveness with reduced expression levels of ZAP70 and CD38 and reduced activation of Ca^{2+} signaling.



	UM-CLL	M-CLL
Frequency	40%	60%
Homology with germline IGHV	$\geq 98\%$	$< 98\%$
BCR characteristic	High responsiveness	Low, anergic
Clinical outcome	Poor prognosis	Indolent

Figure 11: Characteristics of UM-CLL and M-CLL patient subsets.

CLL patients can be classified into two groups, UM-CLL and M-CLL, according to the different mutational statuses of *IGHV*. These two groups showed significantly different phenotypes and clinical outcomes.

However, since a simple numerical threshold can't provide a comprehensive overview of all biological events, different somatic mutation percentage thresholds have also been proposed in recent years, including 95%¹⁰¹ and 97%¹⁰². These cut-offs are valuable in predicting treatment responsiveness, disease progression, overall survival (OS), and time to first treatment (TTFT) in CLL. But the most common cut-off is 98%. However, there is a group of CLL patients whose mutation percentage is near the threshold, i.e., 97-99%, called borderline cases. The predictive assessment appears contradictory to the mutational status of these patients. So some problems are necessary to pay attention to. Firstly, many factors can influence the judgment of BCR characteristics, such as primer design for polymerase conjugate reactions (PCR), choice of sequencing method (Sanger sequencing or NGS), and database selection for data analysis. Secondly, some high-frequency immunoglobulin heavy chain types in CLL patients, such as VH1-69, VH3-07, VH3-23 and VH4-34, showed a bias towards the mutational status, such as VH1-69 and VH4-39 are more likely to be unmutated status, while VH4-34 and some VH3 family genes, such as VH3-07 and VH3-23 are more likely to express mutated. Some IGHV showed a bias to use specific light chains and predicted a different disease's course, for example IGHV4-34/IGKV2-30 expressed in younger patients and followed an indolent disease, on contrary, IGHV3-21/IGLV3-21 experienced and aggressive disease, regardless of IGHV mutations. These findings emphasizes the crucial role of BCR in CLL. This also makes it particularly important for finding standard assays and data analysis systems for immunoglobulins, large-scale assessments, and other valuable molecular studies such as light chains.

Stereotype of IG

Another characteristic of BCRs is nearly 30%¹⁰³ CLL patients expressed highly similar sequences of HCDR3. More notably, 60% of all CLLs are UM-CLLs (**Figure 12**).

The most common subset is **#2**¹⁰⁴, which accounts for 2.8% of all CLL patients (**Table 5**). Even though 70% of *IGHV3-21* exhibit a mutated status, #2 has a significantly poorer survival than other M-CLL¹⁰⁵. In addition to expressing highly similar or even identical VH CDR3 sequences (**Figure 13**), this group of patients is characterized by the fixation of their combined light chain gene to *IGLV3-21*¹⁰⁶. Compared to other types of *IGHV3-21/IGLV*, *IGHV3-21/IGLV3-21* patients have a significantly worse prognosis and are often accompanied by mutations in *SF3B1*, high expression of *MYC* and *CXCR4* expression at ground level.

Another collaborative group with a significantly poor prognosis is subset **#1**, which is involved in the regulation of cell proliferation and BCR signaling, as well as in genetic changes related to apoptosis and NF-KB signaling pathways. Most cases express *IGHV* unmutated phenotypes.

Other subsets that exhibit homogeneous clinical and biological features are the stereotyped #4 and #8. The two categories show some sequence similarity but belong to different cell types (**Figure 14**). Subset **#4** CLL cells are characterized by features such as an allergic functional phenotype¹⁰⁷, identified by mutated *IGHV4-34* and *IGKV2-30*, have long VH CDR3, are primarily diagnosed in young patients, show an unstable clinical course, and a good prognosis. At the same time, researchers indicate that IgG-CLL mainly expresses *IGHV4* family immunoglobulin heavy chains, similar to the edited autoantibodies¹⁰⁸, expressed by CLL cells of subset #4.

Subset **#8** expresses *IGHV4-39/IGKV1(D)-39*, which is the opposite prognosis of subset #4, is a poor prognostic biomarker with a high risk of Richter transformation and is highly associated with several different genetic aberrations, such as trisomy 12 and mutated *NOTCH1*. Some parts of subset #8 can also express γ -constant, while the mechanism of IgG conversion is not fully understood.

Table 5: Summaries the clinical and molecular characteristics of the most common stereotyped subsets.

Stereotyped subset	Subset #1	Subset #2	Subset #4	Subset #8
Frequency	~2.4% ^{103,109}	~2.8% ^{103,109}	~1% ^{103,109}	~0.5 % ¹⁰³
<i>IGHV/IGVL</i> gene identity	<i>IGHV1-69</i> genes/ <i>IGHD6-19</i> <i>/IGHJ4/IGKV1[D]-39</i> ^{103,110}	<i>IGHV3-21/</i> <i>IGHJ6/</i> <i>IGLV3-21</i> ^{103,104}	γ -switched <i>IGHV4-34/</i> <i>IGHJ6/</i> <i>IGKV2-30</i> ^{103,110}	γ -switched <i>IGHV4-39/</i> <i>IGHD6-13/</i> <i>IGHJ5/</i> <i>IGKV1[D]-39</i> ^{103,110}
<i>IGHV</i> mutational status	Unmutated ^{103,110}	Mutated (60%) Unmutated (40%) ^{103,110,111}	Mutated ^{103,108}	Unmutated ^{103,110}
BCR signaling properties	functional BCR signaling ¹¹²	functional BCR signaling ¹¹²	anergic BCRs ¹⁰⁷	functional BCR signaling ¹¹²

Stereotyped subset	Subset #1	Subset #2	Subset #4	Subset #8
Genetic lesions in treatment-naïve CLL	del11q, del17p, <i>NOTCH1</i> mutation, <i>NFKB1E</i> mutations ¹¹³⁻¹¹⁶	del13q, <i>SF3B1</i> mutations ^{30,114}	del13q ^{30,114,117}	trisomy 12, <i>NOTCH1</i> mutations ^{30,114}
Clinical course/Risk of transformation	Aggressive (median TTFT of 1.6 yrs) ⁸⁷	Aggressive (median TTFT of 1.9 yrs) ⁸⁷	Indolent (median TTFT of 11 yrs) ⁸⁷	Aggressive (median TTFT of 1.5 yrs), ⁸⁷ increased risk of Richter's transformation ¹¹⁸

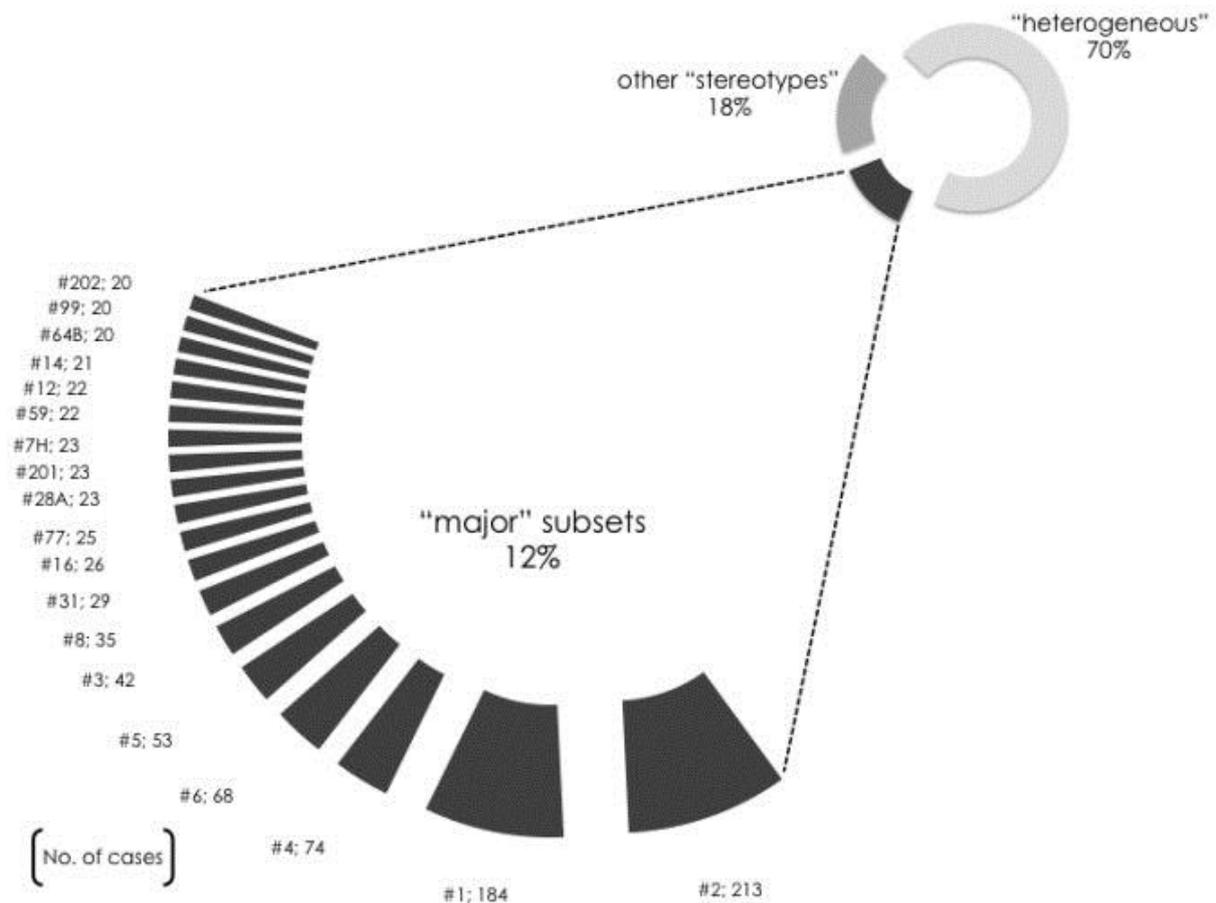


Figure 12: The information of major and other subsets in CLL.

30% of CLL patients expressed VH CDR3 stereotypes, of which 12% are a major subset, suggesting a significant fraction of CLL cases represent stereotypes¹⁰³.

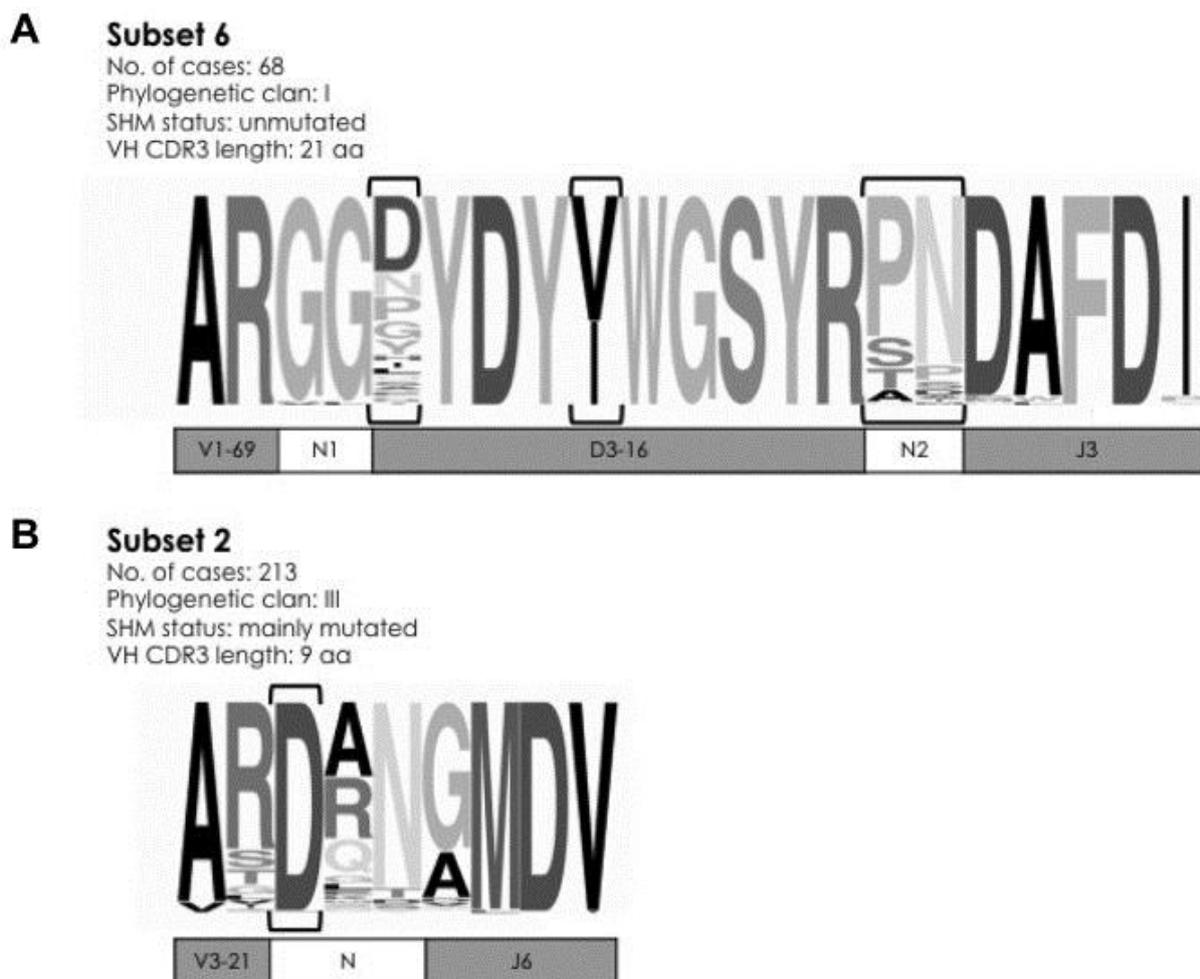


Figure 13: The sequence comparison from subsets #6 and subset #2 in CLL.

(A) There are 68 UM-CLL *IGHV1-69/IGHD3-16/IGHJ3* rearrangements classified into subset# 6, except 4 VH CDR3 positions (show in parentheses), are characterized by variability, and 17 other positions, even if not fully conserved.

(B) Subset # 2 is the most popular subset in CLL. It character a 9-aa long VH CDR3 and an acidic residue (aspartate D) at position 107 (indicated in parentheses)¹⁰³.

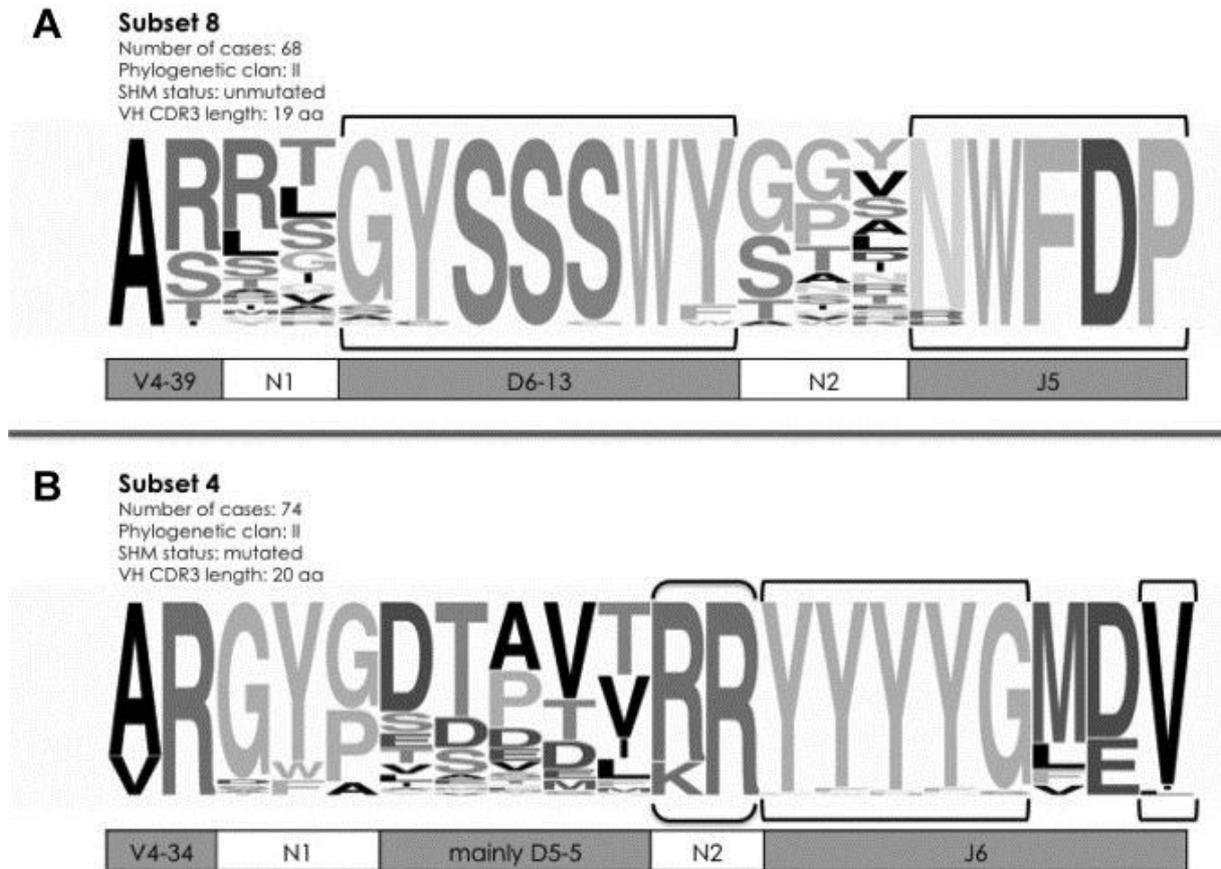


Figure 14: The sequence comparison from subsets #8 and #4 in CLL.

(A) Subset # 8 consists of the unmutated IGHD6-13 and IGHJ5 regions. In contrast, the N-diversity areas at the junction (N1 and N2) are diverse.

(B) The pattern of subset # 4 consists of junctional N2 amino acids [KR]R located at 112.4 (top of CDR3 loop) and 112.3 patterns, YYYG encoded by IGHJ6¹⁰³.

C. The routine method to test the mutational status of *IGHV*

Firstly, we amplified the sample by PCR and detected the genetic clonotype by capillary electrophoresis. Secondly, we can obtain the sequence of the corresponding cloned immunoglobulin heavy chain using Sanger sequencing. Two databases are used for comparison, IMGT¹²¹, and IgBlast¹²². Both websites automatically give the identity of immunoglobulin heavy chain V, D, and J used segments and the mutation percentage of V segment. Researchers will classify CLL patients as M-CLL or UM-CLL according to a consensus 2% threshold.

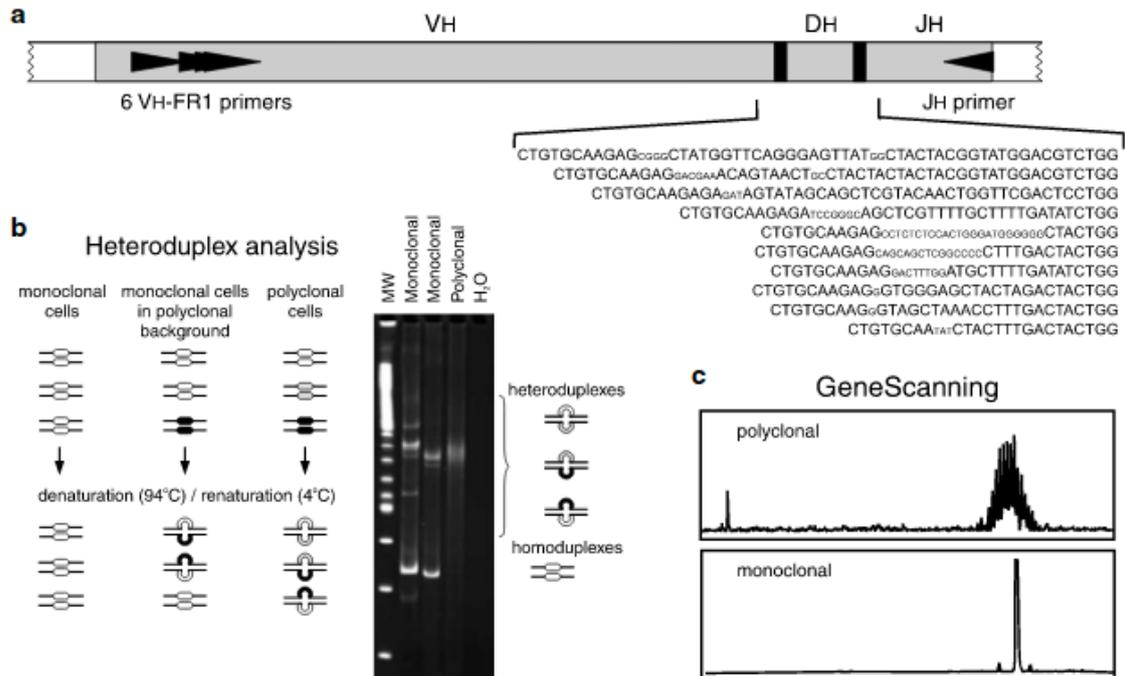


Figure 16: The process of Biomed-2.

This conventional assay is sufficient to provide stable and accurate results for patients expressing a single clonal rearrangement. Moreover, Biomed-2 is available in most hospitals and laboratories, and the procedure has been highly standardized. However, at the same time, in the face of the increasing research on CLL, this conventional method also exhibits several shortcomings. Firstly, it is less sensitive for samples expressing two or more rearrangements, especially multiple rearrangements with different mutational status¹¹⁹. According to the guideline of ERIC, the diagnosis of multiple rearrangements patients are always inconclusive (**Table 6**). Secondly, because of its low throughput, Sanger sequencing requires high-quality samples, and some results obtained for formalin-fixed paraffin-embedded (FFPE) samples are inaccurate. Thirdly, Biomed-2 sequencing starts from the V segment, which is often not long enough to cover all CDR3 sequences, so there is a possibility of missing some critical information. Last but not least, this method cannot be operated simultaneously with other assays and does not provide much data to support the study of light chains, immunoglobulin subclonal, and internal structure. Therefore, developing and standardizing new methods for CLL immunoglobulin assays is particularly important.

Table 6: Suggestions for reporting the *IGHV* mutational status for various types of sequences

Category	Suggestion for report
Single rearrangement	
productive	Definitive conclusion
unproductive	Inconclusive
Double rearrangements	
<i>productive+unproductive</i>	
similar mutational status	Definitive conclusion
unmutated productive+mutated unproductive	Inconclusive
mutated productive+unmutated unproductive	Mutated
<i>double productive rearrangements</i>	
similar mutational status	Definitive conclusion
discordant mutational status	Inconclusive

D. The research on the light chain in CLL

Studies on immunoglobulins have mainly focused on rearranging the heavy chain VDJ segment, while studies related to the light chain κ and λ have been limited. During the last 30 years, the mutational status of *IGHV*, as an independent predictive factor, was used to evaluate the prognosis and therapeutic strategy of CLL patients. However, with the ongoing studies of CLL, several studies have begun to discuss and uncover the biological and clinical value of light chains in CLL. Studies on light chains have mainly focused on their repertoire, function, and mutational relationship with their heavy chains. A large-scale study, which included 279 CLL patients, showed that 21 functional genes were identified in 179 IGK rearrangements, in which the five most frequently used V region of light chain genes were IGKV3-20, IGKV1-39/1D-39, IGKV4-1, IGKV1-5, and IGKV2-30¹²³.

In comparison, 20 functional IGLV genes were identified in 97 IGL rearrangements, and the top five expressed light chain genes were IGLV3-21, IGLV2-8, IGLV2-14, IGLV1-40, and IGLV3-1. It is interesting to note that 77% of 179 IGKV and 71% of 79 IGLV gene rearrangements used proximal cluster (closest to the IGLJ-IGLC pairs).

Moreover, a complete analysis of the CDR3 region showed that for most IGKV genes, similar recombination patterns of CLL cells were observed in the sequences of normal and autoreactive cells. A “CLL-biased” recombination pattern of IGKJ1-2 with IGKJ3-5 was found in some IGKVs. For example, the IGKV3-11 is preferent combined with the IGKJ3-5. The length of CDR3 regions of light chains was considered to have a relationship with multiple antibodies or

self-reaction reactions. KCDR3 was 6 to 11 amino acids in length (median nine amino acids); LCDR3 was 8 to 13 amino acids in length (median 11 amino acids).

A study on the classification of Ig from CLL patients revealed that not only did the heavy chain sequences show high similarity, but also their paired light chains showed a highly restricted bias, thus opening up the understanding of light chains. Among these researches, one of the most well-known light chains is IGLV3-21, expressed explicitly in subset #2 CLL, which predicted patients to exhibit a poor prognosis. In recent studies, BCR-BCR interactions were found in IGLV3-21 cells. Along with point mutation at the shear site between IGLJ3 and the constant variable region due to SHM, resulting in the conversion of glycine to arginine at amino acid position 110, the researchers found that this class of mutations, called IGLV3-21^{R110}, exhibits unique biological and clinical features in CLL¹²⁴ (**Figure 17**). 62% of IGLV3-21^{R110} patients do not express stereotyped *IGHV*, and all subset #2 CLL carry IGLV3-21^{R110}, which means that subset #2 may be a subgroup of IGLV3-21^{R110} and emphasizes the importance of this point mutation in CLL.

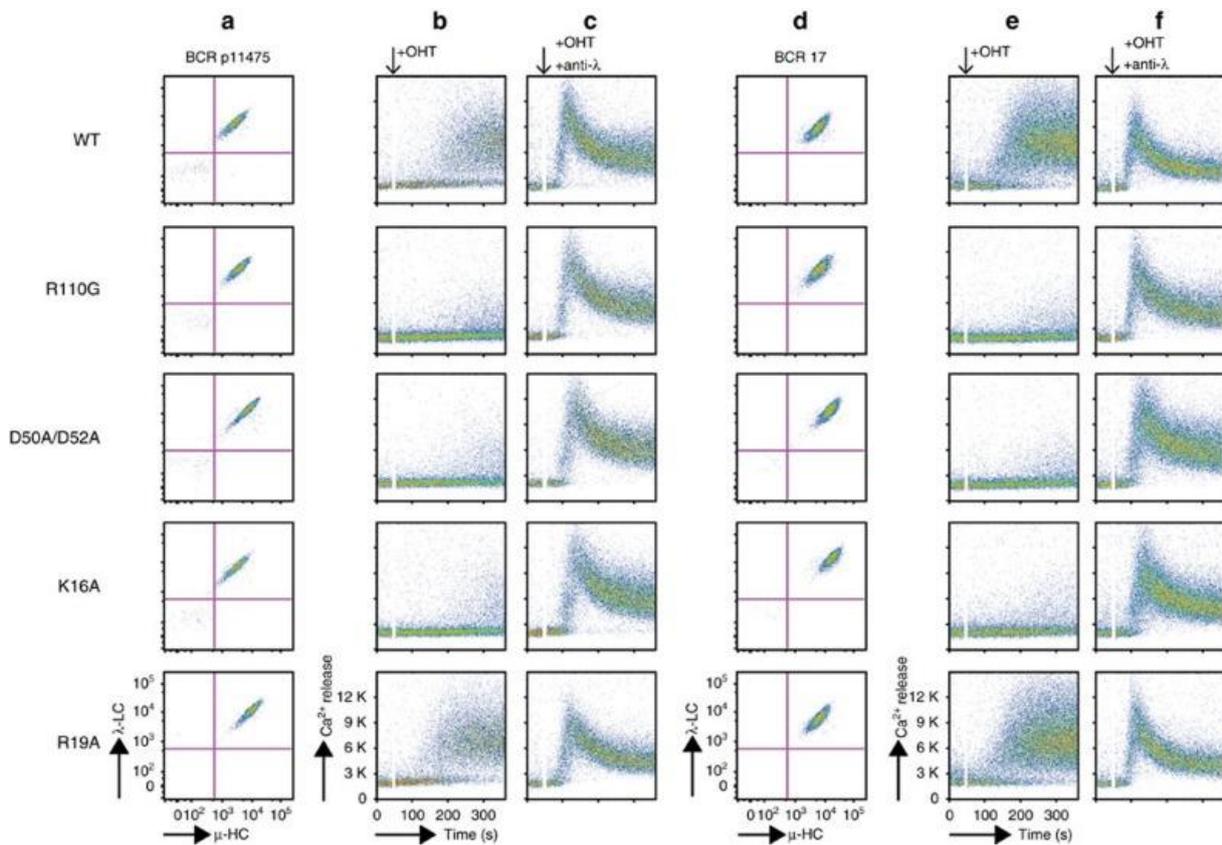


Figure 17: Ca⁺ signaling by CLL subset #2 IgM BCRs depends on a specific mutation from the germline sequence.

The somatic mutation on the position of Arg110LGly (R100G), Asp50LAla/Asp52LAla (D50A/D52A), and Lys16LAla (K16A) can influence the Ca⁺ signal¹²⁴.

To uncover the function of IGLV3-21^{R110}, researchers have compared the survival, and phenotypic clusters between normal B cell, M-CLL, UM-CLL, and IGLV3-21R110 analyzed

the characteristic of IGLV3-21^{R110} patients, including identity to germline *IGHV*, VJ segments usage. Most light chains expressed the same mutational status as their heavy chain. But IGLV3-21^{R110} always showed composite mutational status compared with their *IGHV*. It showed similar phenotype and clinical outcomes of U-CLL, independent of the mutational status of *IGHV*, even though most of its concordant heavy chain identify between 97-99%. It has been reported that the IGLV3-21^{R110} and lysine 16 (K16) in one BCR and aspartate (D) 50 and 52 in the YSDS motif of a neighbor BCR, which correspond to trigger autonomous BCR signaling of aggressive clinical course. And in the three reported IGLV3-21 alleles, only IGLV3-21*01 was associated with K16, and YSDS motif, compared with IGLV3-21 just used IGLV3-21*02 and IGLV3-21*03¹²⁵ (Figure 18).

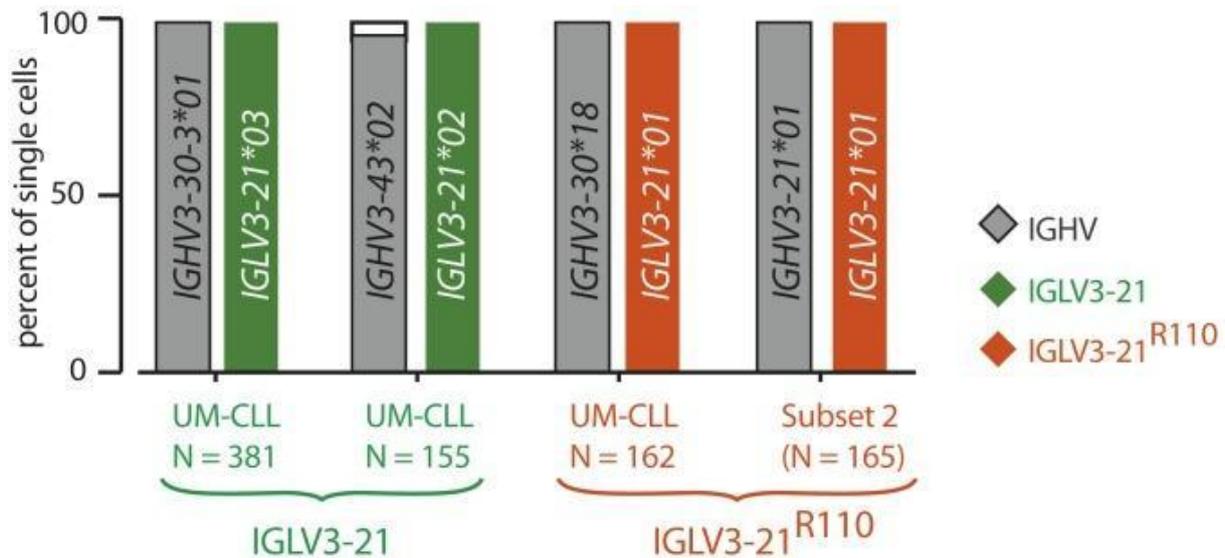


Figure 18: The usage of heavy chains in IGLV3-21 with or without point mutation

The unique usage of *IGHV* and IGLV was shown in the bar graph. IGLV3-21 without mutation was more likely combined with other *IGHV* or *IGHV3-21*02* and *03*. On the contrary, IGLV3-21 with point mutation was more likely to be combined with *IGHV3-21*01*, especially in subset # 2¹²⁵.

According to the results, researchers identified a subtype group 2L, which should expand subset #2 CLL patients into IGLV3-21^{R110} CLL. Subtype 2L represented a poor prognosis and similar clinical phenotype to UM-CLL, which comprises a large group of CLL patients, especially borderline cases. The appearance of subtype 2L can answer the question of diagnosis in CLL patients with 97-99% identity to germline *IGHV*.

IV. The function of *tp53* aberration in CLL

A. The role of *tp53* in cancer

The tumor suppressor protein p53 is regarded as the guardian of the genome¹²⁶ and the coordinator of the underlying processes of cancer¹²⁷. Deletions of the short arm of chromosome 17 or *tp53* mutations, collectively known as *tp53* aberrations, are the most common genetic damage in solid and hematologic cancers. They lead to a significantly increased risk of cancer development by many different pathways. P53 is a DNA-binding protein in which various stress signals can act, including genotoxic damage, nutritional deficiencies, hypoxia, and sustained inflammatory stress (e.g., I-interferon, interleukin 6)¹²⁸. As a tumor suppressor protein, it promotes tumor cell clearance by inducing cell death, fundamentally hindering the proliferation of cells with DNA damaged¹²⁹ and helping to inhibit the growth of tumor cells. At the same time, p53 induces cell cycle arrest¹³⁰, which helps to resolve damage or helps cells to adapt to stress¹³¹.

DNA-damaged cells lacking p53 protein due to *tp53* aberration can tolerate genomic instability and increase the risk of activating oncogenic signaling and the likelihood of cancer development because p53 cannot bind to the target DNA¹³². As p53 can combine and regulate many vital genes, it has been used as a predictive factor of different cancer types, evaluated the developmental course and prognosis of cancer, as well as help to make therapeutic strategies for routine chemotherapy and novel target drugs. However, the mechanism of p53 is not understood very well. So the study of p53, whether in DNA, RNA or protein level, is increasingly important in disease research and clinical practice, especially in the development and progression of tumors.

B. Alternative splicing

Alternative splicing is a complex regulatory process. Exons are usually taken out of the genes, after which they are arranged, joining adjacent exons. The differential selection of alternative splicing leads to some exons or part exons being skipped and producing mRNAs in different alignments. Alternative splicing is a necessary process because of the procedure of various mRNA. One gene can generate different protein isoforms, which promotes protein diversity in different tissues and plays a crucial role in biological differentiation and genome evolution. It has been found that more than 95% of human genes depend on alternative splicing for development, tissue specificity, and cell signaling. With the different isoforms, an isoforms network exists, whether in isoforms of the same genes or among different genes. The regulation of alternative splicing is a complex process that includes regulating cis-acting elements and trans-acting factors. While additional molecular features, such as chromosome structure and RNA structure, act synergistically to regulate alternative splicing and biological protein diversity. Disruption of these processes may disrupt normal cellular function and disease development.

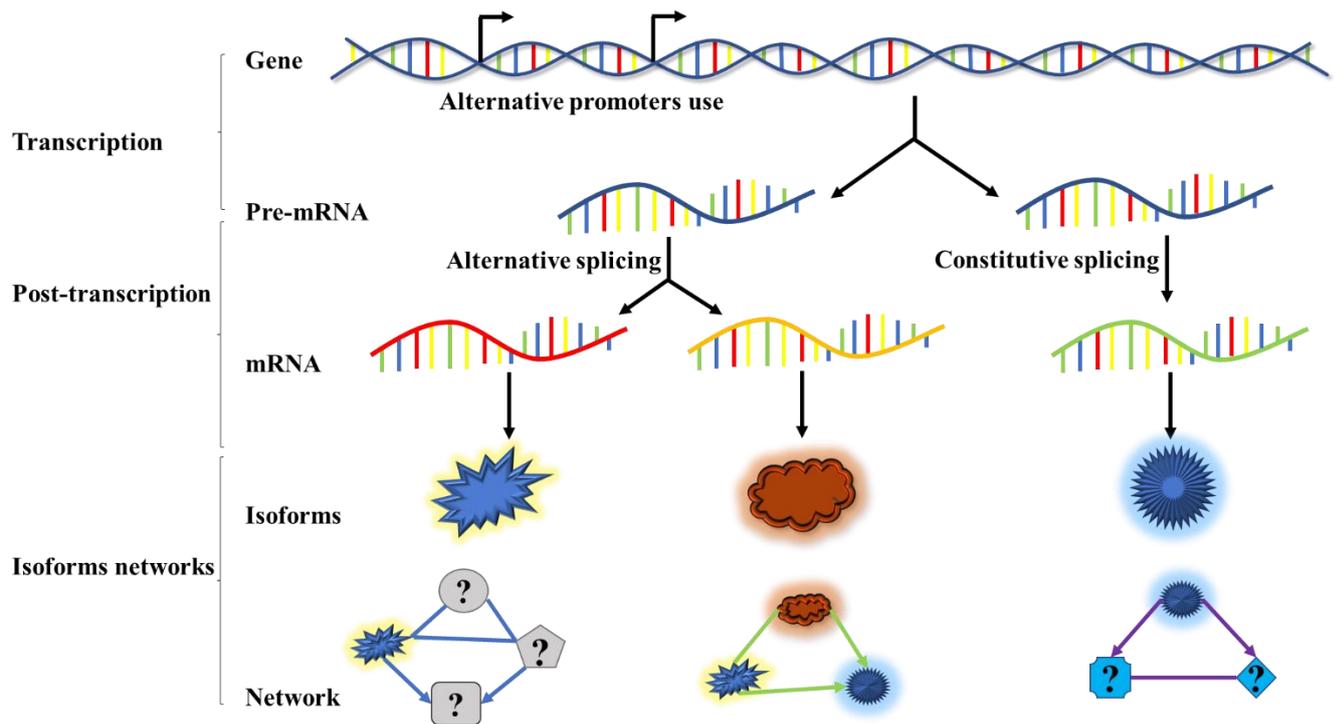


Figure 19: The mechanisms of the alternative transcription and alternative splicing.

In recent years, more and more studies have reported the critical role of alternative splicing in cancer. Abnormal alternative splicing can generate tumor-specific protein or break the balance among isoforms (Figure 19). Besides the regulate factors, genetic mutations or chromosomal aberration can also influence protein expression. Therefore, abnormal alternative splicing could be essential and promising research whose prognostic and therapeutic values are not understood clearly. Consequently, it is necessary to combine the mutational information of genes on DNA, the chromosomal aberration, and the situation of alternative splicing on RNA to figure out the relationship between chromosomal aberration, DNA mutations, and mRNA expression.

C. The research on the isoform of *tp53* in cancer

In CLL, the mutation rate of *tp53* is around 10%, which increases in patients who have relapsed after accepting chemotherapy. Recently, with the development of NGS technology, there has been an increasing number of studies on low-load *tp53* mutations with a variant allele frequency (VAF) of less than 10%. There has been attention to the fact that for patients with a VAF of 0.1%-10%¹³³, although their mutation burden is low, studies have found that most low mutation-load *tp53* has a poor prognostic impact on the patients. Therefore, despite the current generation of VAF threshold determination with only a 10% *tp53* mutation rate, some other factors must influence the function of *tp53*, including protein and transcriptome level analyses. Hence, it is crucial to make a combination of different kinds of abnormality, like *tp53* mutation, deletion of chromosome 17p, and expression level of p53 isoforms in transcription and translation level, to reveal its role in the development and progression of CLL. To accomplish

this, we need to understand the principles and potential mechanisms of p53 isoforms and molecular interactions.

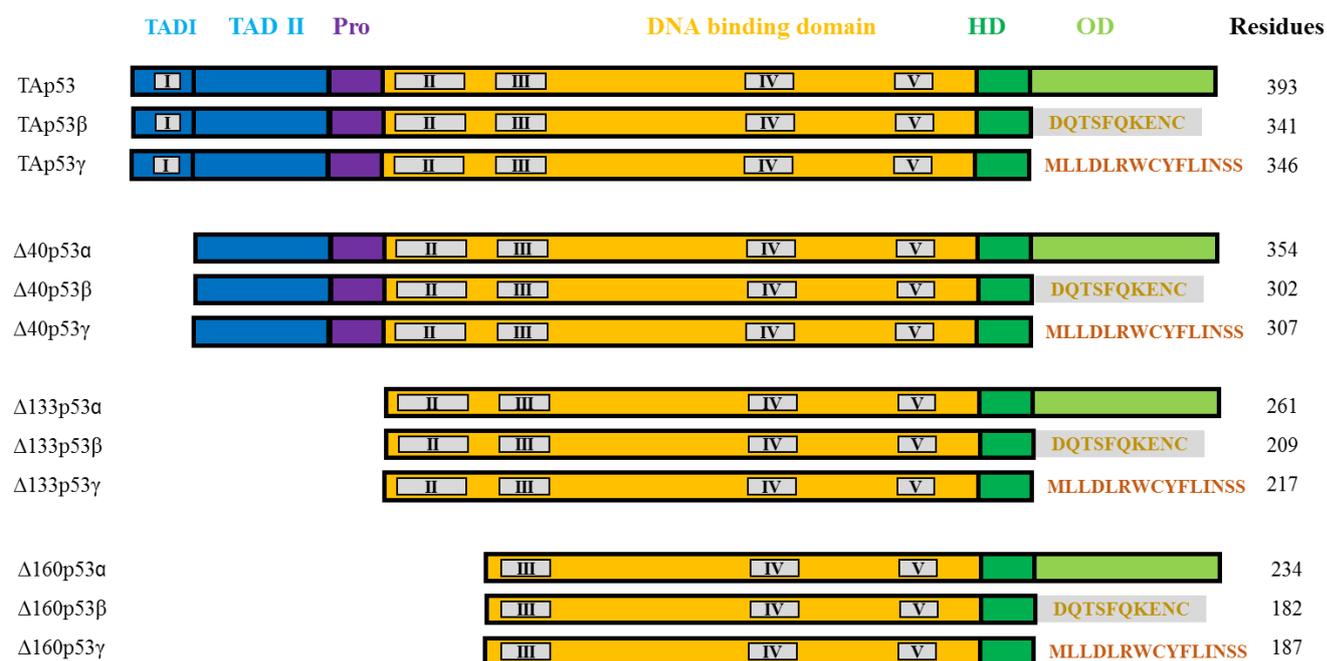


Figure 20: Human *tp53* Isoforms.

Because of two promoters and alternative splicing, the *tp53* gene can transcript into 12 kinds of mRNA. As shown in the figure 20, at the N-terminal, they can start at codon 1 or 40, expressing the full-length p53 and Δ40p53, respectively. The other two transcriptions were produced by promoter 2. Their protein begins at positions 133 and 160, respectively, named Δ133p53 and Δ160p53. At the c-terminal, the alternative splicing site was located on exons 9, named α, β, and γ.

The *tp53* gene is distributed on 11 exons on chromosome 17p13. So far, researchers have found that *tp53* can generate 12 different kinds of p53 protein isoforms¹³⁴, where isoform expression is regulated at the transcriptional level by two promoters and alternative splicing among 11 exons (Figure 20). The mRNA fragment transcript by promoter 1 (P1) translates either full length p53 (TAp53) or Δ40p53 starting from codon 40. While mRNA transcript by promoter 2 (P2) would translate from methionine at codons 133 and 160, synthesizing Δ133p53 and Δ160p53 isoforms. In addition to different splicing sites at the N-terminal, alternative splicing of C-terminal exon 9 generates three other isoforms, α, β and γ. Among them, α retains all exons, while both exons 9β and 9γ have stop codons that result in post-translational C-terminal truncations of different lengths.

The function of p53 is dependent on the six structural domains. Starting at N-terminal, there are two trans-activating structural domains I and II (TAD I and TAD II) with amino acid codes 1-40 and 41-67, respectively, which can independently activate the transcription of target genes. The proline-rich structural domain (PRD) with amino acid codes 68-98 mediates the interaction between proteins in signal transduction¹³⁵ that are required to trigger apoptosis and growth inhibition¹³⁶. The DNA binding domain (DBD), the most critical part of p53, can specifically recognize and bind the double-stranded response element on target genes¹³⁷. The oligomeric

structural domain (OD) contributes to the deformation of bound DNA, and its interaction with the DBD facilitates binding to DNA. It is noteworthy that all isoforms of p53 contain OD. The C-terminal structural domain (CTD) controls the structure and function of the complete protein.

Table 7: p53 isoforms: their cellular effects and appearance in various cancer types.

Description of p53 Isoforms ¹³⁸		Biological effects	Associated cancers
TAp53 β	TAp53 β lacks 52 amino acids from the C-terminus	Promotes senescence	Colorectal cancer AML Renal cell carcinoma Breast cancer Squamous cell carcinoma
TAp53 γ	TAp53 lacks 47 amino acids from the C-terminus	Cytotoxic predict a favorable overall survival and a low cancer reoccurrence	Breast cancer AML Uterine carcinoma
Δ 40p53	N-terminal truncated p53 starts from Met 40	Induces cell death tumor suppression	Mucinous ovarian cancer Ovarian cancer Li-Fraumeni syndrome
Δ 133p53 α	N-terminal truncated p53 starts from Met 133	Pro-survival factor Inhibits senescence May rescue embryogenic cells from radiation-induced apoptosis	Colorectal cancer Ovarian cancer Primary breast cancer
Δ 133p53 β	N-terminal truncated p53 starts from Met 133 and also lacked 52 amino acids from the C-terminus	Promotes epithelial-mesenchymal transition in breast cancer cells Fosters cancer stem cell potential	Enhances cancer cell stemness in breast cancer Expressed in melanoma cell lines Increased breast cancer pervasiveness
Δ 160p53	Truncated at N-terminus, starts with Met 160	Induce mutant-like phenotypes Involved in oncogenic mutant p53 gain of functions	-

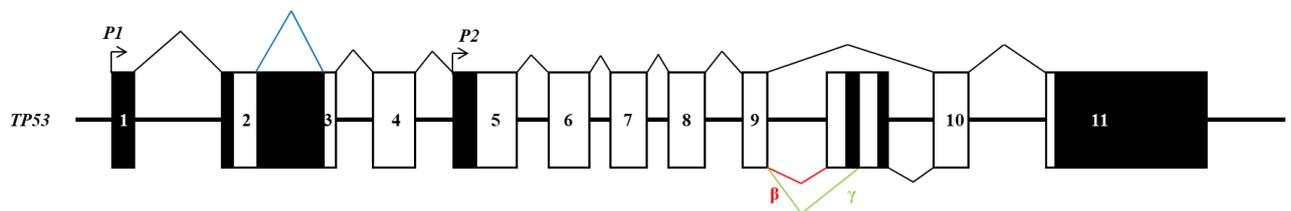
P53 isoforms have clear clinical relevance¹³⁹, as p53 isoforms have shown differential expression in different tumor studies¹³⁴, and some isoforms are aberrantly expressed in specific tumor tissues. For example, in breast cancer, the *tp53* gene mutation rate was only 25% in all patients, but its 60% reduction in P53 β and γ expression was accompanied by a 40% increase in Δ 133p53 expression¹⁴⁰. Similar results were observed in acute myeloid leukemia (AML), where the expression of both p53 β and γ isoforms were significantly altered, but only 10% of patients showed *tp53* mutations¹⁴¹. Thus, all these results emphasize the vital role of p53 isoforms in cancer disease and the need for a clinical database of p53 isoforms between normal and tumor samples from different tissues.

The different isoforms of p53 can homodimerize or oligodimerize to perform their function. As summarized in the **Table 7**, $\Delta 40p53$ is translated from codon 40. It retains most of its functional regions compared to the full-length p53 protein, lacking only TAD I but maintaining the function of the TAD II region. This isoform binds to TAp53 in vivo and thus regulates its activity¹⁴². Due to its different trans-activating structural domains can combine with various target genes and induce a different response than full-length p53 protein. $\Delta 40p53$ has a higher propensity to form oligomers than TAp53, so a small number of $\Delta 40p53$ can significantly impact the p53 pathway¹⁴³, and studies have shown that $\Delta 40p53$ balances the regulation of tissue regeneration and tumor suppression¹⁴⁴.

$\Delta 133p53$ truncates a region that included all TAD (I and II region) and part of DBD, so the $\Delta 133p53$ isoform showed a more dramatic effect on the p53 pathway. Studies have shown that this isoform can support cell survival by inducing the expression of specific genes to inhibit apoptosis¹⁴⁵ and regulate immunity and angiogenesis¹⁴⁶. It is also concordance that in many malignance diseases, $\Delta 133p53$ is always expressed more highly than in normal tissues. $\Delta 160p53$ is the minimal isoform of p53. Due to its deletion of a significant part of the DBD region, it may produce effects similar to *tp53* mutations. Interestingly for frequent *tp53* point mutations, such as in R175H, R248Q, $\Delta 160p53$ expression levels are significantly elevated¹⁴⁷.

D. The routine method of testing the expression of p53 isoforms

The researchers usually detect the p53 isoforms at mRNA and protein levels.



Promoter	P53 mRNA variants	Sequence
P1	TAp53 (α, β, γ)	TAp53α: E1-E2-E3/8-E9-E10-E11
		TAp53β: E1-E2-E3/8-E9-E9β-E10-E11
	TAp53γ: E1-E2-E3/8-E9-E9γ-E10-E11	
	$\Delta 40p53$ (α, β, γ)	$\Delta 40p53$ α: E1-E2-I2-E3/8-E9-E10-E11
		$\Delta 40p53$ β: E1-E2-I2-E3/8-E9-E9β-E10-E11
$\Delta 40p53$ γ: E1-E2-I2-E3/8-E9-E9γ-E10-E11		
P2	$\Delta 133p53$ (α, β, γ)	$\Delta 133p53$ α: I4-E5/8-E9-E10-E11
		$\Delta 133p53$ β: I4-E5/8-E9-E9β-E10-E11
		$\Delta 133p53$ γ: I4-E5/8-E9-E9γ-E10-E11

Figure 21: The promoters of p53 and each sequence of isoform.

On the transcriptional level, the *tp53* can generate at least nine different mature mRNA variants by two other promoters. Researchers used reverse transcription polymerase chain reaction (RT-PCR) to detect the expression of p53 at the RNA level (**Figure 21**). Real-time quantitation PCR (RT-qPCR) is also used if the studies need precisely quantify the expression level. For quantifying the p53 isoforms, specific regions corresponding to each isoform were selected. The primers forward and reverse were near the particular areas. After amplification, the probe

belonging to the specific region will quantify the expression level of different isoforms. There is also a common region expressed in all isoforms to quantify total p53 mRNA expression. Many reports have investigated that different human tissues express different p53 isoforms at transcriptional levels. However, cancers can develop when the balance between different isoforms is broken. For example, many cancers have shown a higher expression level of $\Delta 133p53$ than normal tissues, whereas β and γ isoforms seem to disappear in human cancer cells¹⁴⁸. Considering different isoforms showed different functional regions, it is necessary to build up a database of the mRNA expression level of normal and malignant tumors. And so, because of the limited number of antibodies of all p53 isoforms, the only method to quantify the expression level of p53 isoforms is RT-qPCR. Therefore it's crucial to establish the standard protocol to uncover the characteristic of p53 isoforms at the mRNA level.

Although the limited antibodies with a different efficiency are used in p53 isoforms, it is crucial to detect the expression of p53 isoforms at protein level for analysing the function of each isoform and understanding the network among different proteins (**Figure 22**). The standard method to test the expression of proteins is Western Blotting by several specific antibodies to distinguish functional regions of p53.

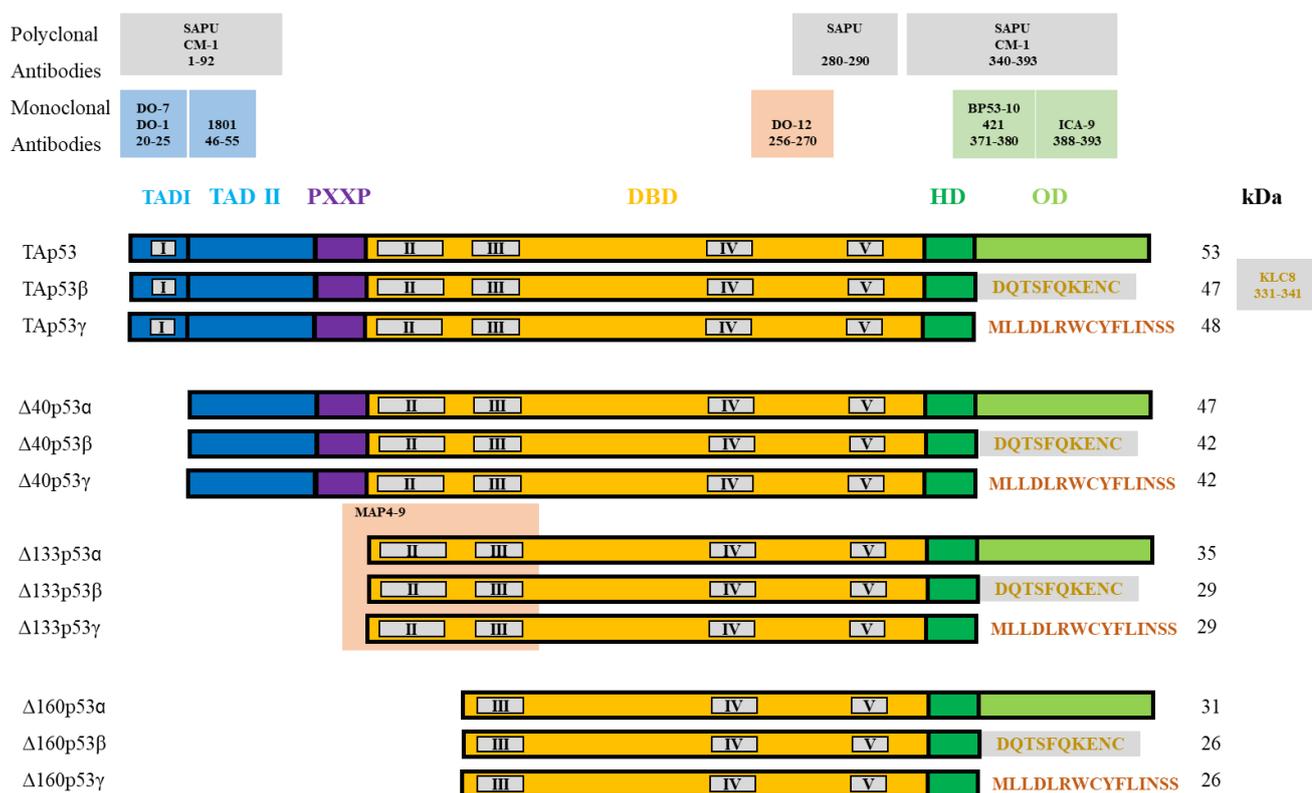


Figure 22: The specific antibodies of p53 isoforms in humans.

According to the specific antibodies that combine with distinctive epitopes, we can compare the different expression levels of isoforms. Some antibodies can also distinguish different kinds of p53 isoforms shown in the picture. The MAP4-9 can detect the expression of $\Delta 133p53$ isoforms, including $\Delta 133p53\alpha$, $\Delta 133p53\beta$, and $\Delta 133p53\gamma$. The KLC8 can combine with the

epitopes unique to β p53 isoforms, including TAp53 β , Δ 40p53 β , Δ 133p53 β , and Δ 160p53 β . Besides monoclonal antibodies, there are also polyclonal antibodies that are used in research. They can combine N-terminal and C-terminal simultaneously, making it possible to compare the different expression levels between isoform expressed N/C-terminal and truncated N/C-terminal. However, the method of detection of p53 isoforms still needs to be improved, especially in developing specific antibodies that can recognize single isoforms.

V. The object of this thesis

This study aims to improve the practical application of NGS in CLL continuously. Researchers can co-operate with multiple experiments on one platform and improve the cytological information detection rate in CLL patients. By placing the detection of numerous vital genes within the same testing environment and combining it with clinical data, we can assess the survival prognosis and treatment of patients and identify biomarkers with the potential to play a vital role in staging the disease and evaluating prognosis.

Therefore, we selected to detect the mutational status of heavy and light chains and the expression of *tp53* gene alternative splicing at the transcriptional level. Firstly, we want to compare the conventional detection methods and get more information about immunoglobulin. Because most studies now focus on the heavy chain's mutational status, the light chain's value has been underestimated. The information on light chains obtained simultaneously by 5'RACE can also help us further analyze the pathogenesis and prognostic value of light chains in CLL. Secondly, *tp53* aberration is another important biomarker in CLL, which play a crucial role in relapse patients and evaluating immunochemotherapy strategy. The current methods to diagnose the *tp53* aberration are FISH, which detects chromosomal deletion at the *tp53* gene-related position on the short arm of chromosome 17, and sequencing exon 4-9 of *tp53* to calculate mutation VAF value. However, all these methods tested the abnormality at the DNA level, and more and more studies have emphasized the value of alternative splicing of p53 transcription. So, in this study, we try to investigate the expression level of p53 mRNA, the clinical significance of different isoforms of p53 mRNA, and analysis the influence of other genes on the transcription of p53, including *tp53* aberration, unmutated *IGHV*, *NOTCH1*mut, *SF3B1*mut, and abnormal chromosome. It is the first study of alternative splicing of *tp53* gene expression at the transcriptional level in CLL. Also, it's the first time RT-MLPA combined with high-throughput sequencing to analyze the bio-information of mRNA p53 in CLL patients.

MATERIALS AND METHODS

I. Patients and samples

A series of 81 patients diagnosed with B-cell CLL with *IGHV* mutational status results from Biomed-2 in Centre Henri Becquerel were included between July 1, 2002, and July 1, 2020. To compare the sensitivity and ability of the two methods (Biomed-2 and 5'RACE), we chose cases with different intervals of *IGHV* identity. There were 14 unmutated cases (99-100%), 21 borderline cases (97-99%), 34 mutated cases (<97%), and 12 double rearrangement cases which Biomed-2 diagnosed.

To analyze the alternative splicing of *tp53* in CLL, we retrospectively collected data from 114 patients diagnosed with CLL in Centre Henri Becquerel between January 1, 2002, and July 1, 2021. The CLL diagnosis was determined by at least two cytologists with ten years of experience, according to the criteria of the recently updated WHO diagnosis guideline². Inclusion criteria were the complete clinical data and follow-up, available tumor RNA at the time of diagnosis. Mutations of *tp53* were identified through NGS sequencing (covering exons 2 to 11), allowing the detection of low-frequency variants down to 1% VAF. *IGHV* mutational status was available for 94 patients. We also get the evaluation of single nucleotide polymorphisms (SNPs) identified through the NGS by the same sequencing platform.

II. Preparation of RNA

Mononuclear cells were isolated by density gradient centrifugation (lymphoprep, 15 min at 2000 g). According to the instruction manual, RNA was isolated from peripheral blood mononuclear cells (PBMCs) using the RNA NOW kit (Biogentex, Seabrook, TX, USA) and stored at -80°C in water. RNA concentrations were determined using a nanodrop spectrophotometer.

III. 5'RACE

It is a molecular biology technique that makes it easy to determine the 5' sequence of an RNA transcript expressed by a cell. We start by producing cDNAs from RNA, after which we add a poly-A tail via a terminal transferase (TT) to allow the initiation of PCR amplification.

Stage 1: A mix containing, for each sample, 4 μL of buffer 5 \times , 2 μL DTT at 100mM, 1 μL Primers RT 0.1 μM (Table 8), and 11 μL RNA. And then, the mixture was incubated for 2 minutes at 80°C , 5 minutes at 37°C , and finally cooled at four $^{\circ}\text{C}$. Added 1 μL of transcriptase inverse (RT MMLV), incubated 1 hour at 37°C , 2 minutes at 98°C , and cooled at four $^{\circ}\text{C}$.

Table 8: RT Primers.

Primers RT	5'---3'
Primer_RT_IgM	AAGTGATGGAGTCGGGAAGG
Primer_RT_IgG	CGCCTGAGTTCCACGACA

Primer_RT_IgA	GGAAGAAGCCCTGGACCAG
Primer_RT_IgE	AGGTGGCATTGGAGGGAAT
Primer_RT_IgD	AGGGCTGTTATCCTTTGGGT
Primer_RT_IgK	CACACAACAGAGGCAGTTCC
Primer_RT_IgL123	TCAGAGGAGGGCGGGAAC
Primer_RT_IgL7	CCGGGTAGAAGTCACTTACGA

Stage 2: The resulting product from stage 1 is purified using bills Ampure. For each sample, we added 20 μL H₂O and 40 μL bills Ampure, mixed, and incubated at room temperature for 5 minutes. We used a plate magnet and magnetic beads to wash the samples. After adding 80% alcohol to wash the magnetic beads twice, we entirely remove the alcohol and wait until after the magnetic beads are dry (not more than 5 minutes). Add 12 ml of water to wash away the cDNA contained on the magnetic beads. We used the magnetic plate again to suck out 10 ml of the liquid. For the terminal transferase, we added 5 μL TdT Buffer 10X, 5 μL CoCl₂ at 2.5 mM, 0.5 μL dATP at 10mM, 0.5 μL terminal transferase (20 units/ μL), and 29 μL water for each sample. The mixture was incubated for 30 minutes at 37 °C, 10 minutes at 70 °C, and cooled at four °C.

Stage 3: The last step of 5'RACE is PCR (polymerase Chain Reaction). PCR allows the in vitro exponential enzymatic synthesis of multiple copies of a DNA sequence by a cyclic process (denaturation, hybridization, extension) based on temperature variations. The objective is to obtain a considerable quantity of the DNA sequence studied to detect it.

1. First Amplification

For each sample, we mixed 25 μL Q5 High-Fidelity 2X Master Mix, 1 μL I1_PolyT(20) at 2.5 μM , and 23 μL cDNA with A-tailed. Then the mixture was placed in a thermocycler. After the first cycle (**Table 9**) finished, 1 μL I2_Primer_Nested_Pool at 2.5 μM was added (**Table 10**). The program restarted on the second cycle.

Table 9: The temperature of the first amplification.

Number of cycles	Denaturation	hybridization	extension
First cycle	3 min at 98°C	5 min at 52°C	3 min at 72°C
Infinite			
Second cycle	30 sec at 98°C	5 min at 52 °C	3 min at 72°C
17 cycles	30 sec at 98°C	30 sec at 50 °C	1 min at 72 °C

Last cycle	30 sec at 98°C	30 sec at 50°C	2 min at 72°C
------------	----------------	----------------	---------------

Table 10: Nested I2 Primers.

	Primers Nested
I2_Primer_Nested_IgM	AAAAGGGTTGGGGCGGAT
I2_Primer_Nested_IgG	GGAAGACCGATGGGCCCTTG
I2_Primer_Nested_IgA	GAGGCTCAGCGGGAAGAC
I2_Primer_Nested_IgE	AAGGGGAAGACGGATGGGCT
I2_Primer_Nested_IgD	ATGGGGAACACATCCGGAGC
I2_Primer_Nested_IgK	CAGATGGCGGGAAGATGAAG
I2_Primer_Nested_IgL1	AGTGACAGTGGGGTTGGCCT
I2_Primer_Nested_IgL237	ACAGAGTGACCGAGGGGCA

2. Second Amplification

For each sample, 25 μ L Q5 High-Fidelity 2X Master Mix, 18 μ L H₂O, 1 μ L P5_I2 at 5 μ M, and 1 μ L P7_Index_Hyb11-Truseq at five μ M (Bar code) have been prepared for each 5 μ L PCR1. The mixture was lanced on the thermocycler (**Table 11**).

Table 11: The temperature of the second amplification.

Number of cycles	Denaturation	Hybridization	Extension
First cycle	3 min at 98°C	30 sec at 52°C	1 min at 72°C
18 cycles	30 sec at 98°C	30 sec at 52°C	1 min at 72°C
Last cycle	30 sec at 98°C	30 sec at 52°C	2 min at 72°C

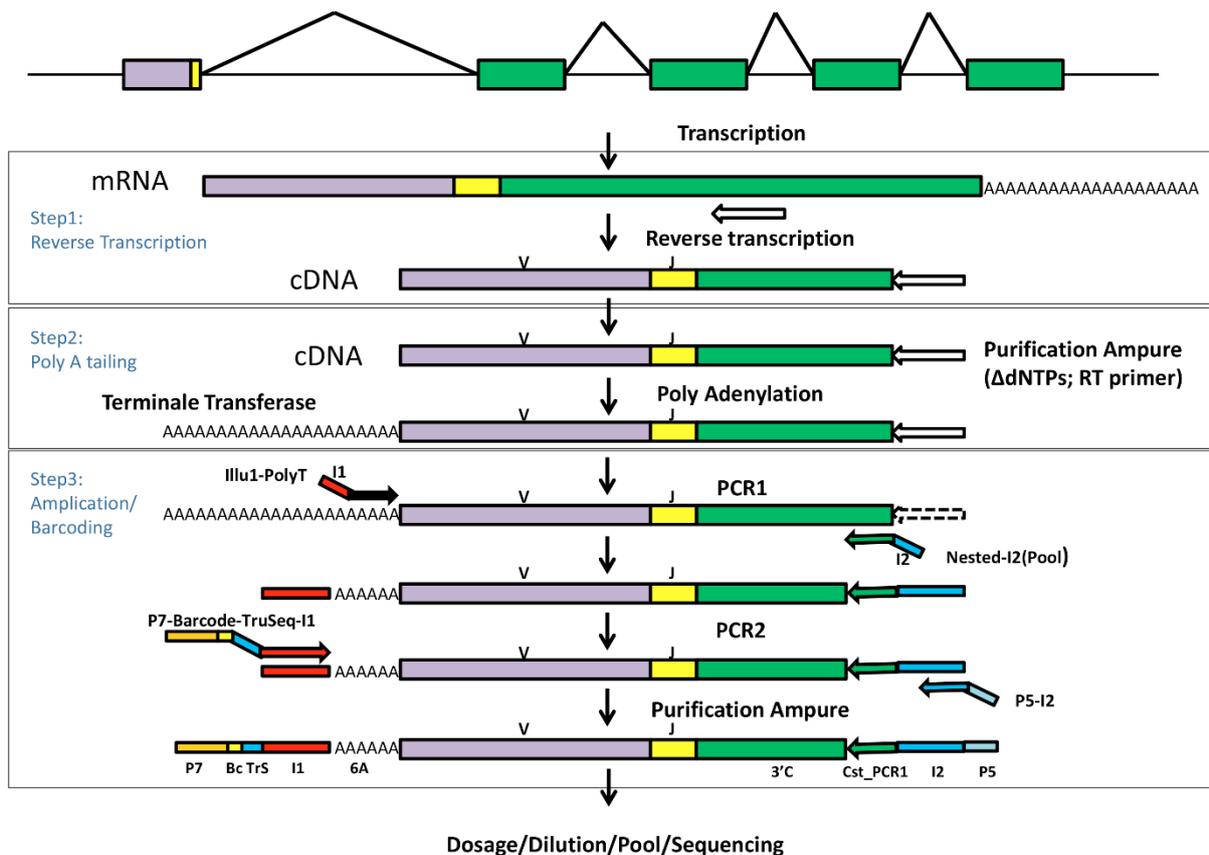
To 50 μ L of PCR product, 50 μ L of Ampure XP beads were added (mixed and incubated for 5 minutes at room temperature), then the ethanol washing steps were repeated (see above).

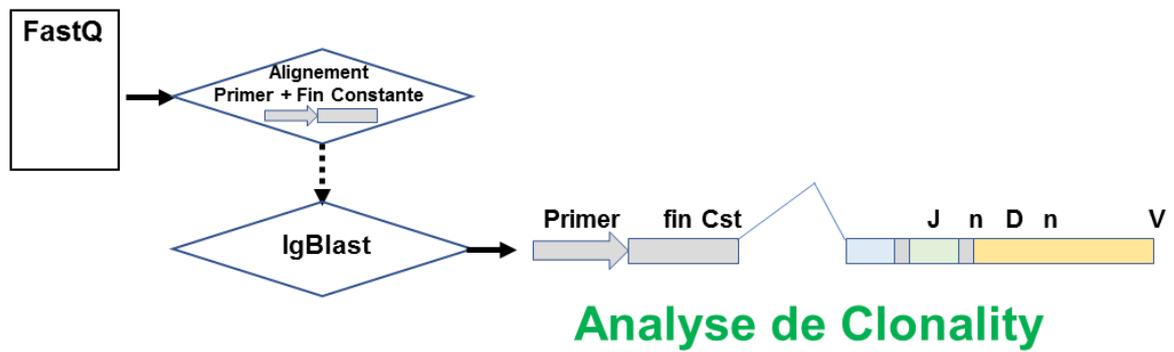
The sample was removed from the magnet, and 40 μ L of TE was added (mixed and incubated for 2 minutes). Finally, the plate was put back on the magnet (1 minute) to transfer 38 μ L of the supernatant to a new plate. The samples were dosed to know the quantity of cDNA (pg / μ L) and to determine the dilution factor using a Qubit type fluorimeter and a Qubit HS kit.

Next Generation Sequencing (NGS) (Figure 23)

1. Sample preparation

It is done by dilution, denaturation, and mixing of the purified PCR products. These products were diluted to 4nM with EBT (convert $\text{pg}/\mu\text{L}$ into nmol/l), and the pool samples (groups) volume by volume to constitute a library. For the denaturation of the library, a mix containing 5 μL of a fresh solution of NaOH (probe) at 0.2 N and 5 μL of the library at 4 μM was prepared (vortex, centrifuge, incubate for 5 minutes at room temperature). To this mix was added 990 μL of cold HT1. We obtained 1ml of library denaturation at 20 pM.





Result summary: CLL-1	Productive IGH rearranged sequence (no stop codon and in-frame junction) CLL subset #2 ?		
V-GENE and allele	Homsap IGHV3-21*01 F	score = 1377	identity = 97.57% (281/288 nt)
J-GENE and allele	Homsap IGHJ6*02 F	score = 215	identity = 90.38% (47/52 nt)
D-GENE and allele by IMGT/JunctionAnalysis	No results	-	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.11]	[8.8.9]	CARDQNDMDVW
JUNCTION length (in nt) and decryption	33 nt = (11)+2{3}-15(17)	(3'V)3'(N)5'(5'J)	

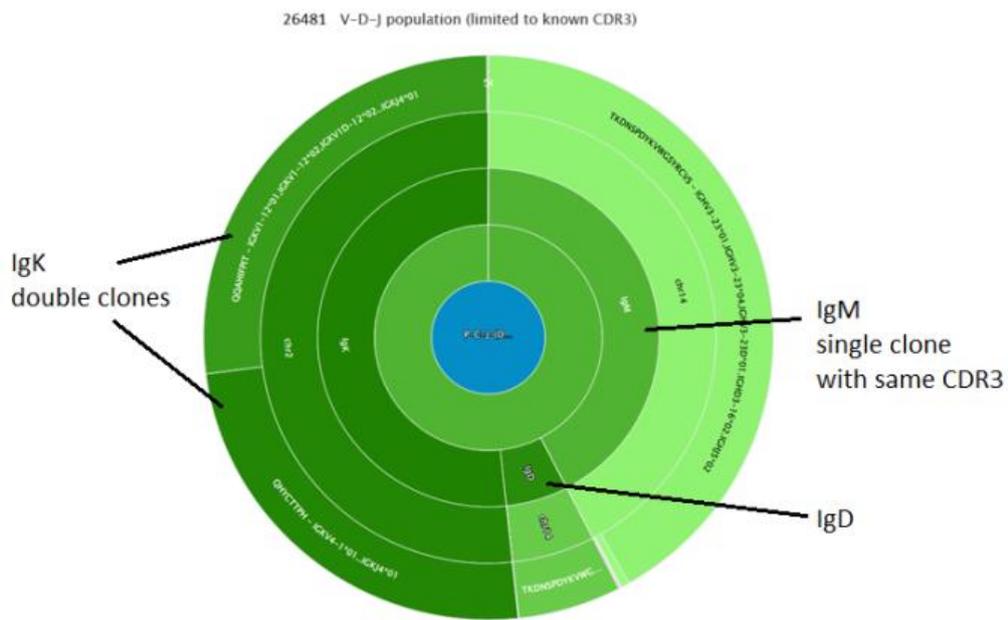


Figure 23: Procedure of 5' RACE.

It included three steps before sequencing: step 1 is reverse transcription, producing a single strand cDNA, step 2 consists of adding the poly-A tail, and step 3 is a two-round PCR amplification adding P7 and P5 adaptors as well as individual barcodes.

2. Sequence

The sequencing cartridge, supplied in the NGS kit, and reagents were loaded with our reagents: 600 µl of sequencing primers at 0.5 µM in wells 18 and 19, 600 µl of the library in the "Load Sample" well. The flow cell was previously rinsed, dried, cleaned, and placed on the Miseq plate. The flow cell is a slide comprising channels lined with primers sequences (complementary to a portion of the adapters), on which the strands of DNA that one wishes to sequence will attach. PR2 replaced the washing bottle, the cartridge was inserted in the refrigerated compartment, and the run started.

Illumina method: according to the Illumina solexa method, the DNA fragments to be sequenced hybridize and are amplified directly on the flow cell. We then obtain clusters (a large number of copies of the fragment to be sequenced). Subsequently, the sequencer proceeds to sequence these fragments via reversible fluorescent nucleotides (nucleotides A, T, C, and G carry a fluorochrome and a blocker). The sequencing occurs in a succession of cycles (addition of a complementary labeled nucleotide-fluorescence signal-photo). The blocker and the fluorescent group are removed to move on to the following process. The sequencer converts the fluorescence signal obtained in the photos by the incorporated nucleotide, forming a sequence.

Table 12: **Primer of the search pattern.**

Primer	Search pattern
IgM	GAATTCTCACAGGAGACGAGGGGAAAAGGGTTGGGGCGGATGCACTCC
IgG1234	GGAAGACCGATGGGCCCTTGGTGGAGG
IgA	GAGGCTCAGCGGAAGACCTTGGGGCTGGTCGGGGATG
IgE	AAGGGGAAGACGGATGGGCTCTGTGTGGAGG
IgD	ATGGGGAACACATCCGGAGCCTTGGTGGGTG
IgK	CAGATGGCGGGAAGATGAAGACAGATGGTGCAGCCACAGTTC
IgL1	AGTGACAGTGGGGTTGGCCTTGGGCTGAC
IgL237	ACAGAGTGACCGAGGGGGCAGCCTTGGGCTGAC

Bioinformatic analysis

Each FASTQ was analyzed using a dedicated bioinformatics pipeline composed of four steps: identifying and trimming primer sequences, correcting sequencing artifacts, identifying VDJ recombination, and finally, mutation rate estimation.

1. Sequence identification

Primers were designed to capture immunoglobulin (Ig) heavy and light chains. An exact linear search for primer patterns was first performed to assign each read to its chain. An Aho-Corrasick¹⁴⁹ data structure is computed from the designs presented in **Table 12** to ensure an efficient linear search of the primer sequence in each read. Only sequences having an expected primer were retained for further analysis.

2. Sequencing artifact correction

An occurrence table is built from all the trimmed sequences. To prevent reads with sequencing errors from being considered real mutations, reads with a low occurrence are compared to those with a high occurrence to identify whether a low base quality score could explain observed substitutions. In practice, this comparison is performed exclusively for sequences of the same length and within the limit of two substitutions at most. If the observed mismatches have a score strictly less than Q30, then the wrong read is discarded.

3. Identification of VDJ recombination

Reads were compared against known V, D, and J germline sequences using IgBlast v1.15.0¹⁵⁰. Germline V, D, and J sequences of Ig were downloaded from the international ImMunoGeneTics information system® (IMGT/V-QUEST¹⁵¹). All Ig V, D, and J sequences were combined into separate FASTA files and used as search databases. Rearrangement summary reports are then used to annotate the sequences and estimate the mutation rate.

4. Identification of mutation status

Mutation status was determined by percent homology with germline. A clone with more than 98% homology was considered unmutated and vice versa. Some cDNAs can carry one or more STOP codons (TGA, TAG, TAA), also called nonsense codons, or be out-of-frame (out of reading frame). Not translated into immunoglobulin, this kind of transcribed DNA would be identified as unproductive rearrangements.

IV. RT-MLPA

RT-MLPA high-throughput sequencing

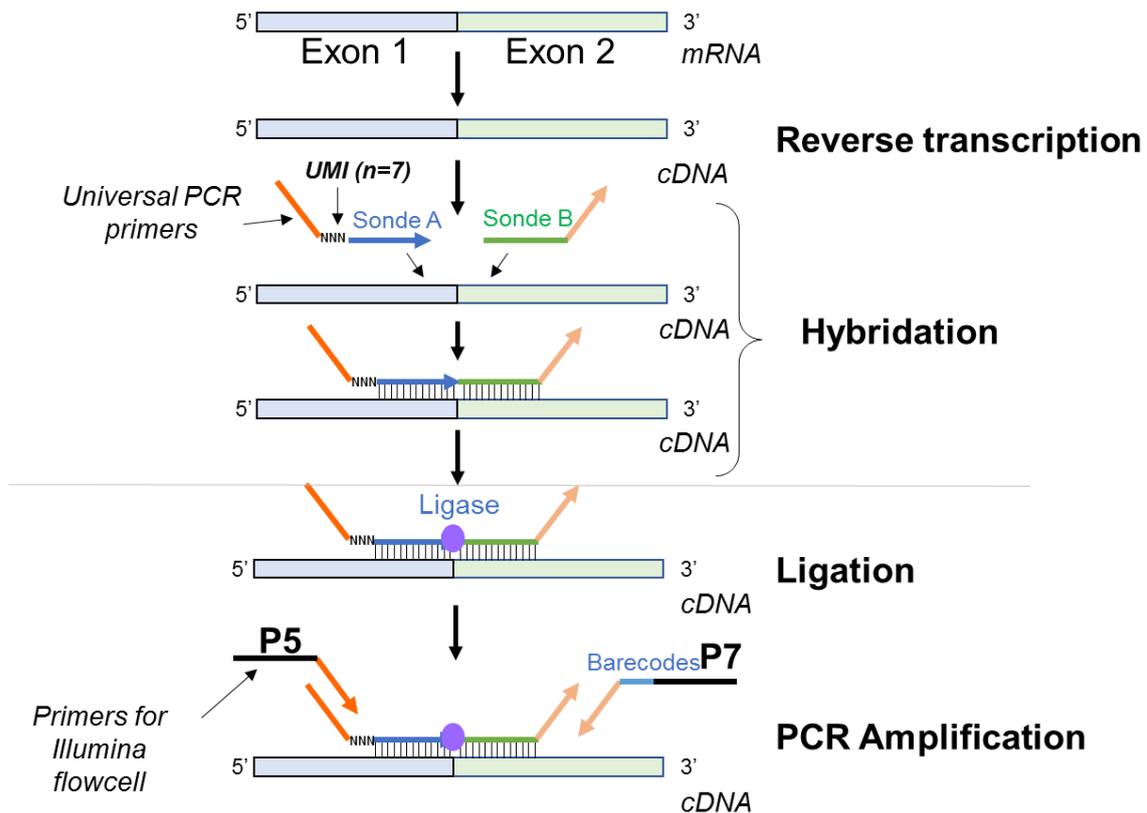


Figure 24: The principle of the RT-MLPA.

The probe pair for each *tp53* exon junction comprised left hybridizing sequencing (LHS) and right hybridizing sequencing (RHS). LHS included 5' universal adaptor sequences, 7N random sequences (UMI), and left probe oligonucleotide (LPO) from 5' to 3'. RHS had right probe oligonucleotide (RPO) and 3' universal adaptor sequences from 5' to 3'. After the probes are hybridized and ligated, we use libraries by P5 and barcode P7 adaptor primers to classify samples.

The probes of MLPA were designed according to the procedure named “Designing synthetic MLPA probes” from MRC-Holland except for PCR primer sequences¹⁵². Probe pair for each *tp53* exon junction, including Left and Right Probe Oligonucleotide (as presented in figure 24). To prepare the sequencing library, we added forward and reverse primers binding sites as 5' adaptor sequences (5'GTGCCAGCAAGATCCAATCTAGA3') on LPO and 3' adaptor sequences (5'TCCAACCCTTAGGGAACCC3') on RPO, named Left and Right Hybridizing Sequence as sequencing primers. To eliminate PCR bias, we added 7 N random sequences at the 5' end of LHS as Unique Molecular Identifier (UMI). As illustrated in **Figure 24**, LHS and RHS are directly adjacent and exactly reverse-complimentary to cDNA from target RNA.

The RNA obtained from samples was reverse transcribed into cDNA. Each sample added a 3.75 μ L mix RT, which included 1.25 μ L Tampon 5X, 0.5 μ L DTT 100mM, 1 μ L dNTPs 10mM, and one hexamers 100 pmoles/ μ L, reverse transcription by 80°C 2 minutes, followed 37°C for 5 minutes and added 0.5 μ L RT 100u in 37°C 15minutes and 98°C 2 minutes.

After we got the cDNA, we hybridized it with LPOs and RPOs for each exon (**Table 13**). For each sample from the first step, add 1.5 μ L MLPA Buffer and 1.5 μ L Mix Oligos RT MLPA for 95 $^{\circ}$ C 2 minutes and 60 $^{\circ}$ C for an hour. Following that, we used Mix Ligase 65, including 25 μ L H₂O, 3 μ L Ligation buffer A, 3 μ L Ligation buffer B, and 1 μ L ligase 65, and directly added the mixture on the Thermo at 54 $^{\circ}$ C. After 54 $^{\circ}$ C 15 minutes and 98 5 $^{\circ}$ C minutes, we can do the PCR to amplify the sequence.

We take five products of ligation of each sample and add 12.5 μ L mix Q5, 5.5 μ L H₂O, and 2 μ L mix oligos P5 I2 and P5 BC I1. The program of PCR starts at 94 $^{\circ}$ C for 6 minutes, followed by 35 cycles at 94 $^{\circ}$ C 30 seconds, 58 $^{\circ}$ C 30 seconds, and 72 $^{\circ}$ C 30 seconds, and ends at 72 $^{\circ}$ C 4 minutes.

The ligated probes were purified by Ampure magnetic beads and used as PCR templates to prepare libraries using P5 and barcode P7 adaptor primers. Different samples constituted libraries were pooled into a single sequencing assay and then classified according to the barcode contained in the P7 adaptor primers.

Sequencing data were aligned to the hybridizing sequences of all gene probes and analyzed to calculate the number of sequence reads that contained both LHS and RHS. Relative expression levels of each exon junction were compared with the ratio of corresponding reads among the total reads of *tp53*.

Table 13: The probes of aim exon of *tp53*.

Probes	Seq
<i>TP53</i> Exon1 LPO	GACACGCTTCCCTGGATTGG
<i>TP53</i> Exon2 RPO	CAGCCAGACTGCCTTCCGG
<i>TP53</i> Exon2 LPO	TCAGGAAACATTTTCAGACCTATGGAAACT
<i>TP53</i> Exon3 RPO	ACTTCCTGAAAACAACGTTCTG
<i>TP53</i> Exon3 LPO	ACTTCCTGAAAACAACGTTCTG
<i>TP53</i> Exon4 RPO	TCCCCCTTGCCGTCCCAAGC
<i>TP53</i> Exon4 LPO	AGCCAAGTCTGTGACTTGCACG
<i>TP53</i> Exon5 RPO	TACTCCCCTGCCCTCAACAAGAT
<i>TP53</i> Exon5b RPO	ATGTTTTGCCAACTGGCCAAGA
<i>TP53</i> Exon5 LPO	AGCGCTGCTCAGATAGCGATG
<i>TP53</i> Exon6 RPO	GTCTGGCCCCCTCCTCAGCAT
<i>TP53</i> Exon6 LPO	TGCCCTATGAGCCGCCTGAG

<i>TP53</i> Exon7 RPO	GTTGGCTCTGACTGTACCACCATCC
<i>TP53</i> Exon7 LPO	CCATCATCACACTGGAAGACTCCAG
<i>TP53</i> Exon8 RPO	TGGTAATCTACTGGGACGGAACAGCT
<i>TP53</i> Exon8 LPO	CCCCAGGGAGCACTAAGCGAG
<i>TP53</i> Exon9 RPO	CACTGCCCAACAACACCAGCT
<i>TP53</i> Exon9 LPO	ACTGGATGGAGAATATTTACCCCTCAG
<i>TP53</i> Exon9b RPO	GACCAGACCAGCTTTCAAAAAGAAAATTGT
<i>TP53</i> Exon9g RPO	ATGCTACTTGACTTACGATGGTGTTACTTC
<i>TP53</i> Exon9bg LPO	CTGATAAACTCGTCGTAAGTTGAAAATATT
<i>TP53</i> Exon10 RPO	ATCCGTGGGCGTGAGCGCTT
<i>TP53</i> Exon10 LPO	GGGAGCAGGGCTCACTCCAG
<i>TP53</i> Exon11 RPO	CCACCTGAAGTCCAAAAAGGGTC
<i>TP53</i> Exon4 codon72 mut RPO	GCGTGGCCCCTGCACCAGCAG
<i>TP53</i> Exon4 codon72 LPO	ATGCCAGAGGCTGCTCCCC
<i>TP53</i> Exon4 codon72 wt RPO	CCGTGGCCCCTGCACCAGCAG

V. QMPSF

Based on the simultaneous amplification of short genomic fragments using dye-labeled primers under quantitative conditions, QMPSF was proved a sensitive method for detecting genomic deletions or duplications^{153,154}. Seven genes located in the commonly deleted regions of chromosomes 6, 11, 13, 17 and the more frequently gained regions in trisomy 12 were selected as target genes, namely *SEC63* at 6q21, *ATM* at 11q22, *DLEU2* at 13q14, and *TP53* at 17p13 and 17p13.1, *POU6F1* at 12q13 and *MDM2* at 12q15.

VI. Statistical analysis

The statistical analysis was performed using SPSS v17.0¹⁵⁵ and GraphPad Prism 5.0 software¹⁵⁶. OS was measured from the time of diagnosis to the date of death or last living follow-up. TTFT was calculated from the time of diagnosis to the date of the first treatment or last living follow-up. All survival data were compared using the Kaplan–Meier and log-rank tests. The correlation between heavy and light chains' mutations were established using Spearman's correlation test.

Comparisons between variables were assessed by the Chi-square test (and Yates' correction), Fisher's exact test for qualitative variables, and t-tests or non-parametric Mann–Whitney test for quantitative data. To describe the linear association between the expression of each exon junction, a multiple linear regression model was developed by the stepwise method. Correlation

between the mutation ratio of p53R72P obtained by NGS and RT-MLPA high-throughput sequencing was established using Spearman's correlation test. All tests were 2-sided, and statistical significance was defined as a P value < 0.05 .

RESULTS

I. Relationship between the mutational status of *IGHV* and *IGLV* and the clinical value of *IGLV*

A. Problem

B-cell CLL, the most common leukemia¹⁵⁷, is a heterogeneous disorder with a remarkably variable natural history and prognosis with leukemic B cells ranging from functionally allergic to highly proliferating. Some CLL patients survive for years without therapy, while others pass away quickly due to progressive disease. Even though different biologic markers are available to help clinicians predict prognosis and make the best treatment decision, the biology and pathogenesis of CLL remain poorly understood.

One of the best prognostic and therapeutic markers in CLL identified more than three decades ago⁹⁷ is the mutational status of *IGHV*¹⁵⁸. Survival analysis has shown that patients with mutated BCR (M), where the *IGHV* region has undergone SHM and expressed less than 98% identity than the germline sequence, have more prolonged survival than unmutated cases (UM) significantly. Furthermore, multivariate analyses have confirmed that this mutational status is a solid independent poor prognostic marker in this pathology¹⁵⁹.

In the clinic, *IGHV* mutational status is commonly evaluated by PCR-based immunoglobulin clonality testing addressed through the alignment of the sequence detected in CLL cells with dedicated immunoglobulin gene databases using a 98% homology threshold¹²⁰. However, as the presence of a single mutation (representing the deviation from germline) can significantly impact the biological results, this threshold has been challenged. The observation of so-called 'borderline' CLL cases with 97–98.9% *IGHV* identity has raised the question of whether they should be considered M- or UM-CLL¹⁶⁰. To address this question, the European Research Initiative on CLL (ERIC) has recently published guidelines emphasizing that caution should be used when employing any threshold to evaluate biological results¹⁶¹. The ERIC initiative has also identified three additional categories of problematic cases¹¹⁹: (1) an unproductive *IGH* VDJ rearrangement (which does not code a functional BCR) is exclusively detected; (2) two productive rearrangements with discordant mutational status are identified; and (3) the CDR3 region, which codes the whole D segment and part of the V and J segments, lacks critical residues. The guidelines also emphasize the importance of analyzing the entire variable region of the *IGH* gene to obtain reliable results.

In addition to the mutational status of the heavy chain, other features of the BCR can also participate in classifying CLL patients into different prognostic subsets^{161–163}. For example, the light chain repertoire is biased, and some *IGKV/IGLV* genes are used more often in CLL than in normal B cells^{31,123}. Other studies have also demonstrated that independent of *IGHV* mutational status, the *IGLV3-21* light chain is associated with a poor prognosis^{125,164}. However, this characteristic is rarely evaluated during routine diagnosis due to the difficulty of analyzing the rearrangements of the BCR light chains at the molecular level.

Today, amplifying the *IGH* VDJ region by PCR starting from gDNA extracted from a blood sample followed by Sanger sequencing is regarded as the standard method for evaluating the

IGHV mutational status¹²⁰. This analysis requires six independent PCR amplifications, which is time-consuming and needs a high level of expertise, particularly for interpreting problematic cases. Other methodologies have recently been proposed to address these difficulties associated with the democratization of NGS platforms in routine diagnosis laboratories¹⁶⁵.

In this study, we evaluated the advantages of one of the NGS methodologies based on a simplified multiplex 5' RACE BCR analysis method adapted to a routine diagnosis workflow, which allows the simultaneous characterization of the immunoglobulin heavy and light chains repertoire in CLL.

B. Results

1) The stable of 5'RACE in testing the mutational status of *IGHV*

a. The concordance and discordance of two different methods

Comparison of the characterization of the mutational status by conventional methods and by 5' RACE

The results of 81 patients assessed by 5' RACE and conventional multiplex PCR following the Biomed-2 protocol are summarized in **Table 14** and **Figure 25**. The evaluation of the *IGHV* mutational status returned concordant results in **76 patients (93.8%) with 27 unmutated, 47 mutated, and two patients with double (one mutated and one unmutated) rearrangements**. Five patients obtained discordant results from the two methods. Specifically, three unproductive, 1 unmutated, and 1 double *IGHV* rearrangement were reported based on Biomed-2 results, which were reported as 3 productive mutated, 1 mutated, and 1 triple *IGHV* rearrangement by 5' RACE, respectively. When we compared the sequences of these five patients, different conclusions were obtained for four patients (#64, 76, 78, and 80) based on the two methods, mainly due to the poor quality of the Sanger sequencing, even though clones were identified on gel electrophoresis.

In contrast, 5' RACE unambiguously demonstrated that all 4 cases expressed productive sequences and solved technical problems due to the co-expression of three productive rearrangements, 2 mutated and one unmutated gene. For patient #42, the two methods yielded the same sequence and CDR3 results, but the mutation identification rate differed (98.3% for Biomed-2 and 97.5% for 5' RACE). This result indicates that caution should be taken when reporting conclusions regarding the mutational status of *IGHV* in borderline cases with 97-99% homology. And we also observed different kinds of heavy and light chains were biased.

Table 14: 5' RACE and Biomed-2 Results.

	IGH (Biomed-2)				IGH (5' RACE)				IGL (5' RACE)							
	Mut	V	D	J	Mut	Ig	V	D	J	Mut	Ig	V	J			
	U/M	Id%			U/M	Id%				U/M	Id%					
1	U	99	<i>IGHV1</i>	<i>IGHD6</i>	<i>IGHJ4</i>	U	100	M/D	<i>IGHV1-2</i>	<i>IGHD6-19</i>	<i>IGHJ4</i>	U	100	K	<i>IGKV1-39</i>	<i>IGKJ2</i>

2	U	99.5	IGHV1-3	IGHD6-19	IGHJ4	U	100	M/D	IGHV1-3	IGHD6-19	IGHJ4	U	100	K	IGKV1-39	IGKJ3
3	U	99.6	IGHV3-21	IGHD1-26	IGHJ4	U	100	M/D	IGHV3-21	IGHD1-26	IGHJ4	U	100	K	IGKV6-21	IGKJ1
4	U	99.6	IGHV3-21	IGHD3-9	IGHJ5	U	100	M/D	IGHV3-21	IGHD3-9	IGHJ5	U	100	K	IGKV1-33	IGKJ5
5	U	100	IGHV3-23	IGHD3-3	IGHJ4	U	100	M/D	IGHV3-23	IGHD3-3	IGHJ4	U	99.6	K	IGKV4-1	IGKJ5
6	U	99.7	IGHV3-23	IGHD3-3	IGHJ6	U	99.5	M/D	IGHV3-23	IGHD3-3	IGHJ6	U	100	K	IGKV2-28	IGKJ4
7	U	100	IGHV3-23	IGHD3	IGHJ4	U	100	M/D	IGHV3-23	IGHD3	IGHJ4	U	99.5	K	IGKV1-5	IGKJ1
8	U	99.5	IGHV3-23	IGHD4-17	IGHJ4	U	99.7	M	IGHV3-23	IGHD4-17	IGHJ4	U	100	K	IGKV3-15	IGKJ2
9	U	98.6	IGHV3-33	IGHD3-16	IGHJ6	U	100	M/D	IGHV3-33	IGHD3-10	IGHJ6	U	100	K	IGKV2-28	IGKJ3
10	U	100	IGHV5-51	IGHD3-3	IGHJ4	U	100	M/D	IGHV5-51	IGHD3-3	IGHJ4	U	100	K	IGKV3-15	IGKJ2
11	U	99.2	IGHV1-58	IGHD3-3	IGHJ4	U	100	M/D	IGHV1-58	IGHD3-3	IGHJ4	U	100	K	IGKV2-28	IGKJ1
	U	99.6	IGHV3-71*	NA	IGHJ5	U	100	M/D	IGHV3-71*	IGHD3-9	IGHJ4					
12	U	99.2	IGHV5-51	IGHD1-26	IGHJ4	U	100	M/D	IGHV5-51	IGHD1-26	IGHJ4	U	100	K	IGKV1-39	IGKJ3
	U	99.2	IGHV3-30*	IGHD2-2	IGHJ4	U	100	G	IGHV3-30*	IGHD2-2	IGHJ4					
13	U	100	IGHV4-31	IGHD5-18	IGHJ4	U	100	M/D	IGHV4-31	IGHD5-18	IGHJ4	U	100	K	IGKV3-11	IGKJ4
	U	100	IGHV1-2	IGHD6-19	IGHJ4	U	100	M/D	IGHV1-2	IGHD6-19	IGHJ4					
14	U	98.6	IGHV4-39	IGHD3-22	IGHJ3	U	100	M/D	IGHV4-39	IGHD3-22	IGHJ3	U	99.5	K	IGKV3-11	IGKJ3
						U	100	M/D	IGHV3-23	IGHD4-11	IGHJ6					
15	U	100	IGHV4-39	IGHD3-3	IGHJ4	U	100	M/D	IGHV4-39	IGHD3-3	IGHJ4	U	100	K	IGKV1-33	IGKJ4
	U	100	IGHV3-33	IGHD3-3	IGHJ6	U	100	M/D	IGHV3-33	IGHD3-3	IGHJ6	U	100	K	IGKV1-12	IGKJ1
16	U	99	IGHV3-15	IGHD3-16	IGHJ4	U	99.3	M/D	IGHV3-15	IGHD3-16	IGHJ5	U	100	L	IGLV3-21	IGLJ3
17	U	98.6	IGHV3-15	IGHD2-21	IGHJ4	U	98.6	M/D	IGHV3-15	IGHD2-21	IGHJ4	U	99.1	L	IGLV3-21	IGLJ1
18	U	98.3	IGHV3-23	IGHD6-25	IGHJ4	U	98.1	M/D	IGHV3-23	IGHD6-25	IGHJ4	U	99.6	L	IGLV3-21	IGLJ3
19	U	98.6	IGHV3	IGHD6	IGHJ6	U	98.6	M/D	IGHV3-33	IGHD3-3	IGHJ6	U	99.1	L	IGLV3-1	IGLJ2
20	U	98.6	IGHV3-48	IGHD6-13	IGHJ4	U	98.1	M/D	IGHV3-48	IGHD6-13	IGHJ4	U	99.1	L	IGLV3-21	IGLJ3
21	U	100	IGHV4-34	IGHD3-3	IGHJ6	U	100	M/D	IGHV4-34	IGHD3-3	IGHJ6	U	100	L	IGLV1-44	IGLJ1
22	U	99	IGHV4-59	IGHD3-22	IGHJ4	U	99.1	M/D	IGHV4-59	IGHD2-8	IGHJ4	U	100	L	IGLV3-21	IGLJ1
23	U	98.6	IGHV3-48	IGHD4-23	IGHJ1	U	98	M/D	IGHV3-48	IGHD1-26	IGHJ4	U	99.1	L	IGLV3-21	IGLJ3
						U	99.5	M/D	IGHV5-51*	IGHD1-1	IGHJ4					
24	U	98.3	IGHV3-7	IGHD3-10	IGHJ4	U	98.6	M/D	IGHV3-7	IGHD3-10	IGHJ4	M	95.3	K	IGKV1-39	IGKJ4
25	U	100	IGHV1-69	IGHD3-3	IGHJ6	U	100	M/D	IGHV1-69	IGHD3-3	IGHJ6	U	100	K	IGKV3-20	IGKJ5
	M	96.9	IGHV3-48	IGHD6-6	IGHJ3	M	96.7	M/D	IGHV3-48	IGHD6-6	IGHJ3	M	96.5	K	IGKV1-8	IGKJ5
26	M	90	IGHV3-23	IGHD3-10	IGHJ4	M	92.4	M/D	IGHV3-23	IGHD3-10	IGHJ4	M	97.1	K	IGKV1-8	IGKJ4
27	M	88.5	IGHV3-23	IGHD6-19	IGHJ4	M	89.6	M/D	IGHV3-23	IGHD6-19	IGHJ4	M	94.7	K	IGKV3-15	IGKJ2

28	M	92	IGHV3-23	IGHD6-6	IGHJ6	M	90.2	M/D	IGHV3-23	IGHD6-6	IGHJ6	M	95.9	K	IGKV1-8	IGKJ3
29	M	88.3	IGHV3-23	IGHD3-3	IGHJ4	M	89.7	M/D	IGHV3-23	IGHD3-3	IGHJ4	M	93	K	IGKV1-5	IGKJ1
30	M	90	IGHV3-23	IGHD1-26	IGHJ5	M	90.5	M/D	IGHV3-23	IGHD1-26	IGHJ5	M	95.8	K	IGKV1-5	IGKJ1
31	M	94.4	IGHV3-23	IGHD1-26	IGHJ4	M	94.7	M/D	IGHV3-23	IGHD1-26	IGHJ4	M	96.4	K	IGKV1-8	IGKJ1
32	M	94.4	IGHV3-23	IGHD3-16	IGHJ5	M	94.9	M/D	IGHV3-23	IGHD3-16	IGHJ5	M	90.1	K	IGKV4-1	IGKJ4
	M					M						M	90.5	K	IGKV1-12	IGKJ4
33	M	95.9	IGHV4-34	IGHD3-22	IGHJ4	M	96	M/D	IGHV4-34	IGHD3-22	IGHJ4	M	96.4	K	IGKV3-20	IGKJ1
34	M	96.8	IGHV4-34	IGHD4-17	IGHJ6	M	97.5	M/D	IGHV4-34	IGHD4-17	IGHJ6	M	95	K	IGKV1-5	IGKJ2
35	M	92.3	IGHV4-34	IGHD2-15	IGHJ4	M	91.3	M/D	IGHV4-34	IGHD2-15	IGHJ4	M	93.2	K	IGKV3-20	IGKJ2
36	M	93.8	IGHV5-51	NA	IGHJ3	M	93.5	M/D	IGHV5-51	IGHD5-18	IGHJ6	M	96.3	K	IGKV3-15	IGKJ2
	M	75	IGHV1-3*	IGHD3-22	IGHJ5	M	90.1	M/D	IGHV1-3*	IGHD3-22	IGHJ5					
37	M	95.3	IGHV6-1	IGHD3-10	IGHJ6	M	93.7	D	IGHV6-1	IGHD3-10	IGHJ6	M	95.3	K	IGKV6-21	IGKJ2
38	M	97.6	IGHV3-21	IGHD3-22	IGHJ6	M	97.8	M/D	IGHV3-21	NA	IGHJ6	M	97	L	IGLV3-21	IGLJ3
39	M	87.3	IGHV3-23	IGHD6-13	IGHJ5	M	88.2	M/D	IGHV3-23	IGHD6-13	IGHJ5	M	90.9	L	IGLV2-23	IGLJ3
40	M	97.2	IGHV3-23	IGHD6-13	IGHJ4	M	96.5	M	IGHV3-23	IGHD6-13	IGHJ4	M	91.9	L	IGLV2-11	IGLJ1
41	M	92.3	IGHV3-23	IGHD6-13	IGHJ4	M	92.5	M/D	IGHV3-23	IGHD6-13	IGHJ4	M	95.7	L	IGLV2-14	IGLJ3
42	U	98.3	IGHV3-48	IGHD3-3	IGHJ6	M	97.5	M/D	IGHV3-48	IGHD3-10	IGHJ6	M	97.9	L	IGLV3-21	IGLJ1
43	M	89.1	IGHV4-34	IGHD1-1	IGHJ4	M	91	M/D	IGHV4-34	IGHD1-1	IGHJ4	M	96.5	L	IGLV1-47	IGLJ2
44	M	91.5	IGHV4-34	IGHD4-11	IGHJ5	M	91.5	M/D	IGHV4-34	IGHD4-11	IGHJ5	M	96.9	L	IGLV2-11	IGLJ1
45	M	90.2	IGHV4-39	IGHD2-21	IGHJ4	M	97.6	M/D	IGHV4-39	IGHD2-21	IGHJ4	M	97.4	L	IGLV3-1	IGLJ2
46	M	94.8	IGHV4-61	IGHD1-14	IGHJ6	M	96.1	M/D	IGHV4-61	IGHD1-14	IGHJ6	M	96.1	L	IGLV3-25	IGLJ2
47	M	93.8	IGHV3-23	IGHD3-10	IGHJ3	M	96.9	M/D	IGHV3-23	IGHD3-10	IGHJ6	Polyclone light chain is short				
48	M	92.7	IGHV3-23	IGHD6-19	IGHJ4	M	91.8	M/D	IGHV3-23	IGHD6-19	IGHJ4	U	98.7	K	IGKV2-30	IGKJ3
49	M	92	IGHV3-72	IGHD2-8	IGHJ6	M	94.8	M/D	IGHV3-72	IGHD2-8	IGHJ6	U	98.8	K	IGKV1-16	IGKJ4
50	M	97.2	IGHV4-39	IGHD3-10	IGHJ4	M	93.6	M/D	IGHV4-39	IGHD3-10	IGHJ4	U	98.8	K	IGKV4-1	IGKJ2
51	M	97.2	IGHV1-24	IGHD3-10	IGHJ4	M	95.7	M/D	IGHV1-24	IGHD3-10	IGHJ4	U	98.7	L	IGLV3-21	IGLJ1
52	M	96.9	IGHV3-15	IGHD4-17	IGHJ4	M	97.5	M/D	IGHV3-15	IGHD4-17	IGHJ4	U	99.3	L	IGLV3-21	IGLJ3
53	M	95.6	IGHV3-21	IGHD3-9	IGHJ5	M	97	M/D	IGHV3-21	IGHD3-9	IGHJ6	U	99.4	L	IGLV3-21	IGLJ3
54	M	96.7	IGHV1-58	IGHD2-2	IGHJ5	M	96.1	M/D	IGHV3-21	NA	IGHJ6	U	98.9	L	IGLV3-21	IGLJ3
55	M	96.3	IGHV3-21	IGHD1-14	IGHJ6	M	96.9	M/D	IGHV3-21	IGHD1-14	IGHJ6	U	99.6	L	IGLV3-21	IGLJ3
56	M	97.2	IGHV3-21	NA	IGHJ6	M	97.4	M/D	IGHV3-21	NA	IGHJ6	U	99.6	L	IGLV3-21	IGLJ3
57	M	97.6	IGHV3-23	IGHD6-6	IGHJ4	M	97.5	M/D	IGHV3-23	IGHD6-6	IGHJ4	U	98.3	L	IGLV3-21	IGLJ3
58	M	97.2	IGHV3-23	IGHD5-24	IGHJ4	M	96.3	M/D	IGHV3-23	IGHD5-24	IGHJ4	U	98.3	L	IGLV3-21	IGLJ3

59	M	97.2	<i>IGHV3-48</i>	IGHD2-15	IGHJ4	M	97.1	M/D	<i>IGHV3-48</i>	IGHD2-15	IGHJ4	U	99.6	L	IGLV3-21	IGLJ3
60	M	96.8	<i>IGHV3-53</i>	IGHD1-26	IGHJ4	M	96.8	M/D	<i>IGHV3-53</i>	IGHD1-26	IGHJ4	U	99.1	L	IGLV3-21	IGLJ3
61	M	97.9	<i>IGHV3-21</i>	NA	IGHJ6	M	94.7	M/D	<i>IGHV3-69</i>	NA	IGHJ6	U	99.4	L	IGLV3-21	IGLJ3
62	M	97.6	<i>IGHV3-74</i>	IGHD5-18	IGHJ4	M	97.6	M/D	<i>IGHV3-74</i>	IGHD5-18	IGHJ4	U	99.6	L	IGLV3-21	IGLJ1
63	M	88	<i>IGHV3-23</i>	NA	IGHJ4	M	97.2	M/D	<i>IGHV3-23</i>	IGHD1-26	IGHJ4	U	98.7	L	IGLV3-21	IGLJ1
	M	97.5	<i>IGHV2-10*</i>	IGHD6-13	IGHJ5	U	99.5	G/M	<i>IGHV2-10*</i>	IGHD6-13	IGHJ5					
64	M	97.6	<i>IGHV3-13*</i>	IGHD2-2	IGHJ4	M	97.9	M/D	<i>IGHV3-48</i>	IGHD4-17	IGHJ4	U	99.6	L	IGLV3-21	IGLJ3
						M	97.3	G/M	<i>IGHV3-13*</i>	IGHD2-2	IGHJ4					
65	M	96.8	<i>IGHV3-74</i>	IGHD3-10	IGHJ3	M	95.4	M/D	<i>IGHV3-74</i>	IGHD3-10	IGHJ3	U	98.7	L	IGLV3-21	IGLJ1
						U	99.3	M/D	<i>IGHV6-1*</i>	IGHD6-6	IGHJ4					
66	M	97.5	<i>IGHV4-61</i>	IGHD4-11	IGHJ3	M	97.3	M/D	<i>IGHV4-59</i>	IGHD4-11	IGHJ3	U	98.7	L	IGLV3-21	IGLJ1
67	U	100	<i>IGHV4-34</i>	IGHD6-6	IGHJ6	U	100	G	<i>IGHV4-34</i>	IGHD6-6	IGHJ6	U	100	K	IGKV3-20	IGKJ1
68	U	99.6	<i>IGHV3-30</i>	IGHD1-20	IGHJ6	U	100	M/D/G	<i>IGHV3-30</i>	IGHD5-12	IGHJ6	U	98.9	L	IGLV3-21	IGLJ2/3
69	U	98.6	<i>IGHV4-39</i>	IGHD6-13	IGHJ5	U	98.9	G	<i>IGHV4-39</i>	IGHD6-13	IGHJ5	U	100	L	IGLV3-19	IGLJ2
	M	91	<i>IGHV1-69</i>	IGHD2-15	IGHJ4	M	91.2	G	<i>IGHV1-69</i>	IGHD2-15	IGHJ4					
70	M	96.7	<i>IGHV4-34</i>	IGHD3-22	IGHJ6	M	97.3	G	<i>IGHV4-34</i>	IGHD3-22	IGHJ6	M	97.7	K	IGKV2-30	IGKJ2
71	M	94.7	<i>IGHV4-34</i>	NA	IGHJ2	M	93.6	M/D	<i>IGHV4-34</i>	IGHD1-26	IGHJ4	M	94.6	K	IGKV3-20	IGKJ1
						M	94.5	G	<i>IGHV1-8</i>	IGHD6-25	IGHJ5	M	97.7	K	IGKV1-9	IGKJ2
72	M	95.9	<i>IGHV1-2</i>	IGHD3-22	IGHJ4	M	96.6	G	<i>IGHV1-2</i>	IGHD3-22	IGHJ4	M	91.8	L	IGLV1-40	IGLJ2/3
73	M	95.9	<i>IGHV4-34</i>	IGHD3-9	IGHJ3	M	95.5	G	<i>IGHV4-34</i>	IGHD3-9	IGHJ3	M	94.8	L	IGLV3-1	IGLJ2
74	M	90.7	<i>IGHV4-34</i>	IGHD6-13	IGHJ5	M	89.6	G	<i>IGHV4-34</i>	IGHD6-13	IGHJ5	M	94.4	L	IGLV1-40	IGLJ3
75	M	96.5	<i>IGHV4-34</i>	IgHD3-16	IGHJ5	M	95.9	G	<i>IGHV4-34</i>	IgHD3-16	IGHJ5	M	95.2	L	IGLV6-57	IGLJ3
76	M	89.8	<i>IGHV4-34*</i>	NA	NA	M	90	G	<i>IGHV4-59</i>	IGHD6-19	IGHJ4	M	96.5	L	IGLV3-1	IGLJ2
77	M	94.7	<i>IGHV4-34</i>	IGHD6-6	IGHJ6	M	94.3	M/D	<i>IGHV4-34</i>	IGHD6-6	IGHJ6	M	94.8	L	IGLV3-1	IGLJ1
	M	64.9	<i>IGHV1-3</i>	IGHD2-15	IGHJ6	M	83.5	M/G	<i>IGHV1-3</i>	IGHD2-15	IGHJ6					
78	M	69.5	<i>IGHV4-61*</i>	IGHD5-18	IGHJ4	M	89.2	G	<i>IGHV4-4</i>	IGHD3-10	IGHJ4	M	94.4	L	IGLV1-44	IGLJ3
	M	95	<i>IGHV3-33*</i>	IGHD2-2	IGHJ5	M	93.7	M	<i>IGHV3-23</i>	IGHD2-21	IGHJ4					
79	U	100	<i>IGHV1-69</i>	IGHD3-9	IGHJ6	U	100	M/D	<i>IGHV1-69</i>	IGHD3-9	IGHJ6	U	100	K	IGKV1D-8	IGKJ1
	U	100				U	100	M/D	<i>IGHV1-3</i>	IGHD4/5	IGHJ3	U	100	K	IGKV2-28	IGKJ4
	U	100	<i>IGHV3-11</i>	IGHD2-2	IGHJ6	U	100	G	<i>IGHV3-11</i>	IGHD2-2	IGHJ6	U	99.4	L	IGLV2-14	IGLJ2/3
80	M	97.2	<i>IGHV1-69</i>	IGHD3-10	IGHJ6	U	100	M/D	<i>IGHV1-69</i>	IGHD3-10	IGHJ6	U	100	L	IGLV3-27	IGLJ3
	U	99.7	<i>IGHV3-33</i>	IGHD5-24	IGHJ6	M	93.1	M/D	<i>IGHV2-5</i>	IGHD1	IGHJ4	M	96.4	K	IGKV3-11	IGKJ5
	U					U	100	M/D	<i>IGHV3-33</i>	IGHD5-24	IGHJ6	U	100	L	IGLV1-47	IGLJ1

81	M	90	<i>IGHV3-23</i>	<i>IGHD6-19</i>	<i>IGHJ4</i>	M	89.4	M/D	<i>IGHV3-23</i>	<i>IGHD6-19</i>	<i>IGHJ4</i>	U	99.8	K	<i>IGKV2-28</i>	<i>IGKJ2</i>
												M	92.3	L	<i>IGLV3-16</i>	<i>IGLJ3</i>

U: unmutated. M: mutated. * Unproductive rearrangement.

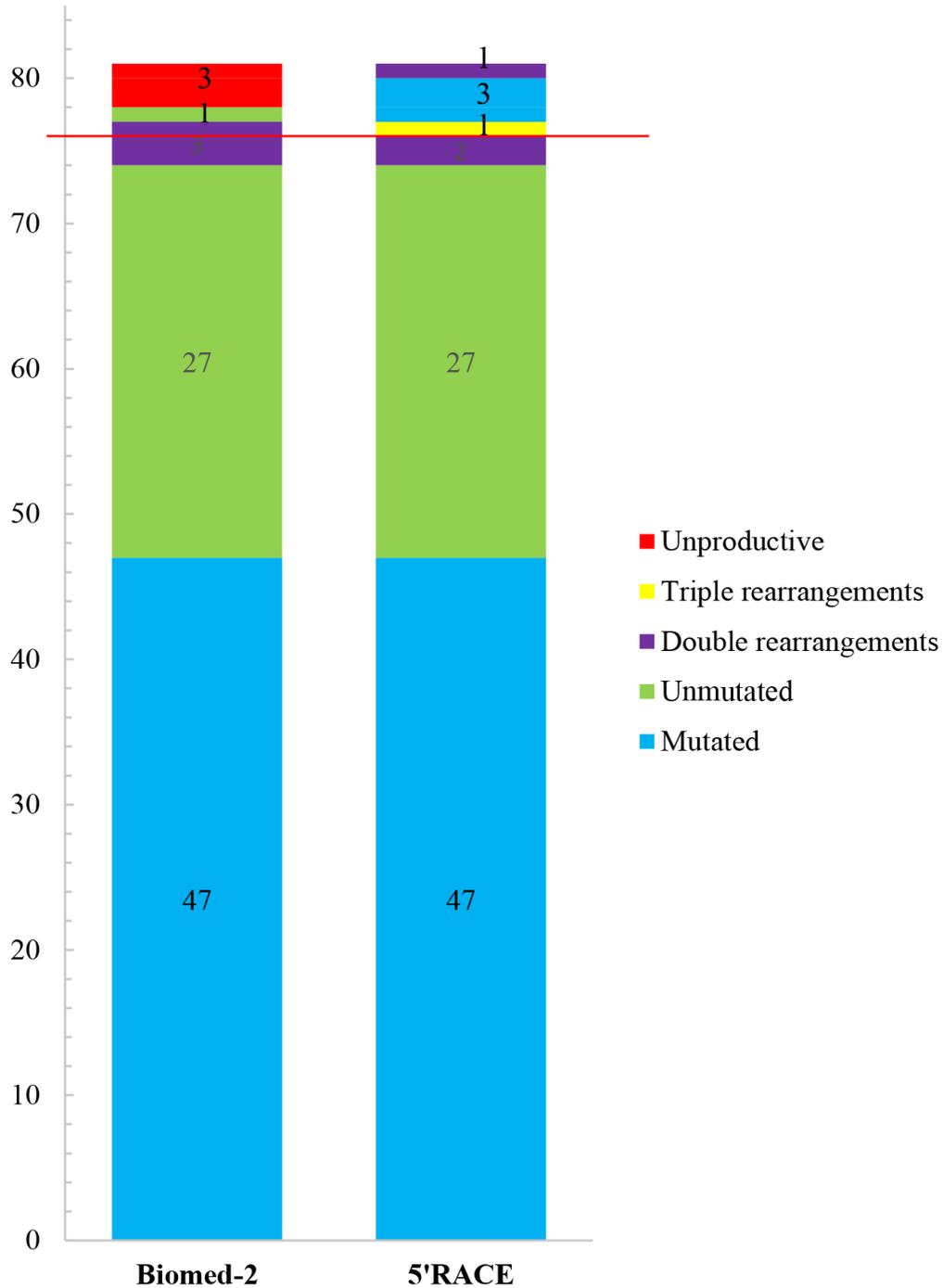


Figure 25: The results of two methods.

Results of Biomed-2 and 5' RACE in 81 patients. Under the red line, patients received the same results from the two methods. The rate of concordance is 93.8%. Green indicates *IGHV*-unmutated patients, blue indicates *IGHV*-mutated patients, orange indicates *IGHV*-unproductive patients, and yellow and purple indicate multiple productive rearrangements patients.

We also chose 14 additional patients with blood samples collected at different times after CLL diagnosis and tested the sample using 5' RACE (**Table 15**). The interval time between sample collections in one patient is from 0.64 to 10.89 years, and the median time is 4.59 years. Our methods detect the same productive immunoglobulin heavy and light chain and mutation status for the samples collected at different time points. Moreover, six patients had the same *IGHV* and *IGLV* mutation percentages. This result shows the stability and reproducibility of 5' RACE, representing a reliable method for judging the treatment plan and patient prognosis.

Table 15: The results of patients tested by 5'RACE at different times.

	Date of sample	Interval time/year	No. patient	IGH					IGL/K			
				Mut %	Kind of IgH	V	D	J	Mut %	Kind of IgL	V	J
1	2009/10/2	10.89	11252	99.37	M/D	<i>IGHV3-15</i>	<i>IGHD3-16</i>	<i>IGHJ5</i>	100	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
	2020/6/26		37139	99.34	M/D	<i>IGHV3-15</i>	<i>IGHD3-16</i>	<i>IGHJ5</i>	100	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
2	2011/1/1	7.09	14034	99.15	M/D	<i>IGHV4-59</i>	<i>IGHD2-8</i>	<i>IGHJ4</i>	100	L	<i>IGLV3-21</i>	<i>IGLJ1</i>
	2017/12/27		27541	99.15	M/D	<i>IGHV4-59</i>	<i>IGHD2-8</i>	<i>IGHJ4</i>	100	L	<i>IGLV3-21</i>	<i>IGLJ1</i>
3	2015/8/26	2.12	21731	98.56	M/D	<i>IGHV3-15</i>	<i>IGHD2-21</i>	<i>IGHJ4</i>	99.14	L	<i>IGLV3-21</i>	<i>IGLJ1</i>
	2017/9/27	1.62	26784	98.56	M/D	<i>IGHV3-15</i>	<i>IGHD2-21</i>	<i>IGHJ4</i>	99.14	L	<i>IGLV3-21</i>	<i>IGLJ1</i>
	2019/5/3	3.74	31556	98.56	M/D	<i>IGHV3-15</i>	<i>IGHD2-21</i>	<i>IGHJ4</i>	99.14	L	<i>IGLV3-21</i>	<i>IGLJ1</i>
4	2015/7/9	3.03	21479	100	M/D	<i>IGHV3-33</i>	<i>IGHD3-10</i>	<i>IGHJ6</i>	100	K	<i>IGKV2-28</i>	<i>IGKJ3</i>
	2018/7/5	2.23	29093	100	M/D	<i>IGHV3-33</i>	<i>IGHD3-10</i>	<i>IGHJ6</i>	100	K	<i>IGKV2-28</i>	<i>IGKJ3</i>
	2020/9/14	5.26	39073	100	M/D	<i>IGHV3-33</i>	<i>IGHD3-10</i>	<i>IGHJ6</i>	100	K	<i>IGKV2-28</i>	<i>IGKJ3</i>
5	2014/2/28	3.32	18699	98.58	M/D	<i>IGHV3-33</i>	<i>IGHD3-3</i>	<i>IGHJ6</i>	99.13	L	<i>IGLV3-1</i>	<i>IGLJ2</i>
	2017/6/6		25912	98.58	M/D	<i>IGHV3-33</i>	<i>IGHD3-3</i>	<i>IGHJ6</i>	99.13	L	<i>IGLV3-1</i>	<i>IGLJ2</i>
6	2015/5/20	3.60	21184	97.47	M/D	<i>IGHV3-15</i>	<i>IGHD4-17</i>	<i>IGHJ4</i>	99.31	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
	2018/12/5		30312	97.18	M/D	<i>IGHV3-15</i>	<i>IGHD4-17</i>	<i>IGHJ4</i>	99.57	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
7	2018/4/25	1.54	28534	96.85	M/D	<i>IGHV3-53</i>	<i>IGHD1-26</i>	<i>IGHJ4</i>	99.13	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
	2019/10/30		32986	96.64	M/D	<i>IGHV3-53</i>	<i>IGHD1-26</i>	<i>IGHJ4</i>	99.31	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
8	2019/7/23	0.64	32212	94.71	M/D	<i>IGHV3-21</i>	NA	<i>IGHJ6</i>	99.44	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
	2020/3/11		34099	94.67	M/D	<i>IGHV3-21</i>	NA	<i>IGHJ6</i>	99.57	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
9	2018/3/16	2.29	28186	97.06	M/D	<i>IGHV3-48</i>	<i>IGHD2-15</i>	<i>IGHJ4</i>	99.57	L	<i>IGLV3-21</i>	<i>IGLJ3</i>
	2020/6/16		37700	97.06	M/D	<i>IGHV3-48</i>	<i>IGHD2-15</i>	<i>IGHJ4</i>	99.57	L	<i>IGLV3-21</i>	<i>IGLJ3</i>

10	2016/10/28	2.01	24345	97.27	G	<i>IGHV3-13*</i>	IGHD2-2	IGHJ4	99.57	L	IGLV3-21	IGLJ3
				97.91	M/D	<i>IGHV3-48</i>	IGHD4-17	IGHJ4				
	2018/10/20		29984	97.61	G	<i>IGHV3-13*</i>	IGHD2-2	IGHJ4	99.31	L	IGLV3-21	IGLJ3
11	2015/9/30	2.80	21946	97.58	M/D	<i>IGHV3-74</i>	IGHD5-18	IGHJ4	99.56	L	IGLV3-21	IGLJ1
				2018/7/5		29092	97.10	M/D				
12	2009/9/17	9.18	11199	95.43	M/D	<i>IGHV3-74</i>	IGHD3-10	IGHJ3	98.67	L	IGLV3-21	IGLJ1
				99.34	M/D	<i>IGHV6-1*</i>	IGHD6-6	IGHJ4				
	2018/10/4		29783	96.73	M/D	<i>IGHV3-74</i>	IGHD3-10	IGHJ3	98.72	L	IGLV3-21	IGLJ1
13	2018/4/11	1.38	28421	97.55	M/D	<i>IGHV3-48</i>	IGHD3-10	IGHJ6	97.85	L	IGLV3-21	IGLJ1
				2019/8/21		32412	97.55	M/D				
14	2011/9/1	7.52	14469	97.46	M/D	<i>IGHV4-34</i>	IGHD4-17	IGHJ6	94.98	K	IGKV1-5	IGKJ2
				2019/1/28		30742	96.18	M/D				

b. The distribution of *IGHV* and *IGLV* in CLL

Characterization of *IGHV* using 5' RACE

Based on 5'RACE, the isotype of the immunoglobulin heavy chain and sequences of the light chain was available as determined by RT-PCR for mu, delta, gamma kappa, and gamma constant region transcripts and NGS.

In the cohort of 81 patients, the male-to-female ratio was 2.86 (60 males: 21 females), and the median age was 64 (range 41–93). Here, 5' RACE identified specific clonal *IGH* and *IGL* rearrangements in 81/81 and 80/81 patients (one patient showed polyclone light chain), respectively, confirming the method's efficiency.

At least one productive *IGH* rearrangement was identified in all cases. Considering the number and mutational status of arrangements, 64 cases presented a single productive rearrangement (79.01%), 15 cases harbored two rearrangements (two productive sequences in 8 cases and one productive and one unproductive sequence in 7 cases), and 2 cases presented three different rearrangements (all productive). The most frequent variable segment was *IGHV3* (51 cases [63.0%] with a bias toward *IGHV3-23* in 23 cases [28.4%]) followed by *IGHV4* (24 cases [29.6%] with a bias toward *IGHV4-34* in 13 cases [16%]) (Figure 26).

Regarding the mutational status, 44 of the 64 cases that presented a single productive *IGH* rearrangement were mutated, and 20 were unmutated. For the 15 cases that presented two rearrangements, 8 harbored two productive rearrangements, including 3 concordant mutated

cases, 3 concordant unmutated cases, and 2 discordant cases with one mutated and one unmutated sequence. The 7 remaining cases showed one productive and one unproductive rearrangement, including 2 with one productive mutated and one unproductive unmutated arrangement (whose diagnosis should pay attention to), 2 with one productive mutated and one unproductive mutated sequence (considered mutated cases), and 3 with one productive unmutated and one unproductive unmutated sequence (considered unmutated cases). For the last two cases, we detected three productive rearrangements. Specifically, one case showed three unmutated sequences, and the other showed one mutated and two unmutated sequences (considered mutated).

More importantly, 5' RACE also provides information about the BCR isotype. Most CLL cases in our series expressed only C μ and C δ heavy chains (65/81; 80.2%). Fifty-six of these cases expressed a single productive rearrangement (with 18 unmutated and 38 mutated sequences), 4 expressed two productive rearrangements (3 concordant unmutated and 1 discordant with mutated and unmutated sequences), and 4 expressed one productive and one unproductive rearrangement (2 concordant unmutated, 1 concordant mutated and 1 discordant productive mutated and unproductive unmutated). The last IgM case expressed three productive rearrangements with 2 mutated and 1 unmutated sequence.

Eight other cases in our series expressed only a C γ heavy chain. Seven showed a single productive rearrangement (6 mutated, 1 unmutated), and 1 showed two productive IgG rearrangements. We observed a bias toward using the VH4 family in this population (6/8 cases), with 5 patients expressing IGHV4-34 sequences.

Unexpectedly, the 8 remaining cases simultaneously expressed the *IGH* chain with different isotypes, suggesting that the class switch recombination process was ongoing when the tumor cell developed. One case expressed three unmutated productive rearrangements (two IgM/D and one IgG). Three cases expressed one productive IgM and one unproductive IgG rearrangement, 3 cases expressed two productive but different IgM and IgG sequences, and the last case showed a very unusual pattern, expressing three different unmutated heavy chains sharing the same CDR3 (IgM, D, and G). No case in our series expressed IgA or IgE.

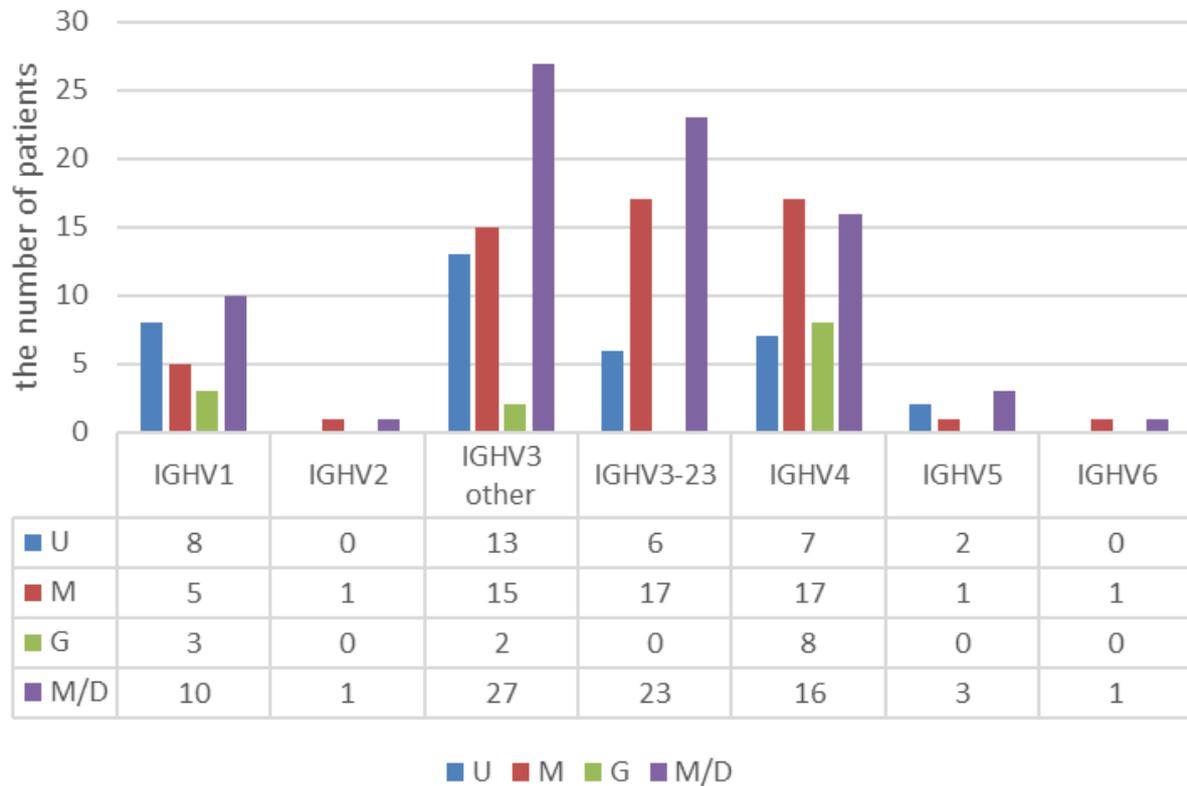


Figure 26: Situation of SHM and CSR in each productive rearrangements.

81 CLL patients have detected 93 productive rearrangements by 5'RACE, including 71 single productive rearrangement cases, 8 double productive rearrangements cases and 2 triple productive rearrangements, for one cases (#68), the same rearrangement present IgM/D/G which counts as G and M/D group.

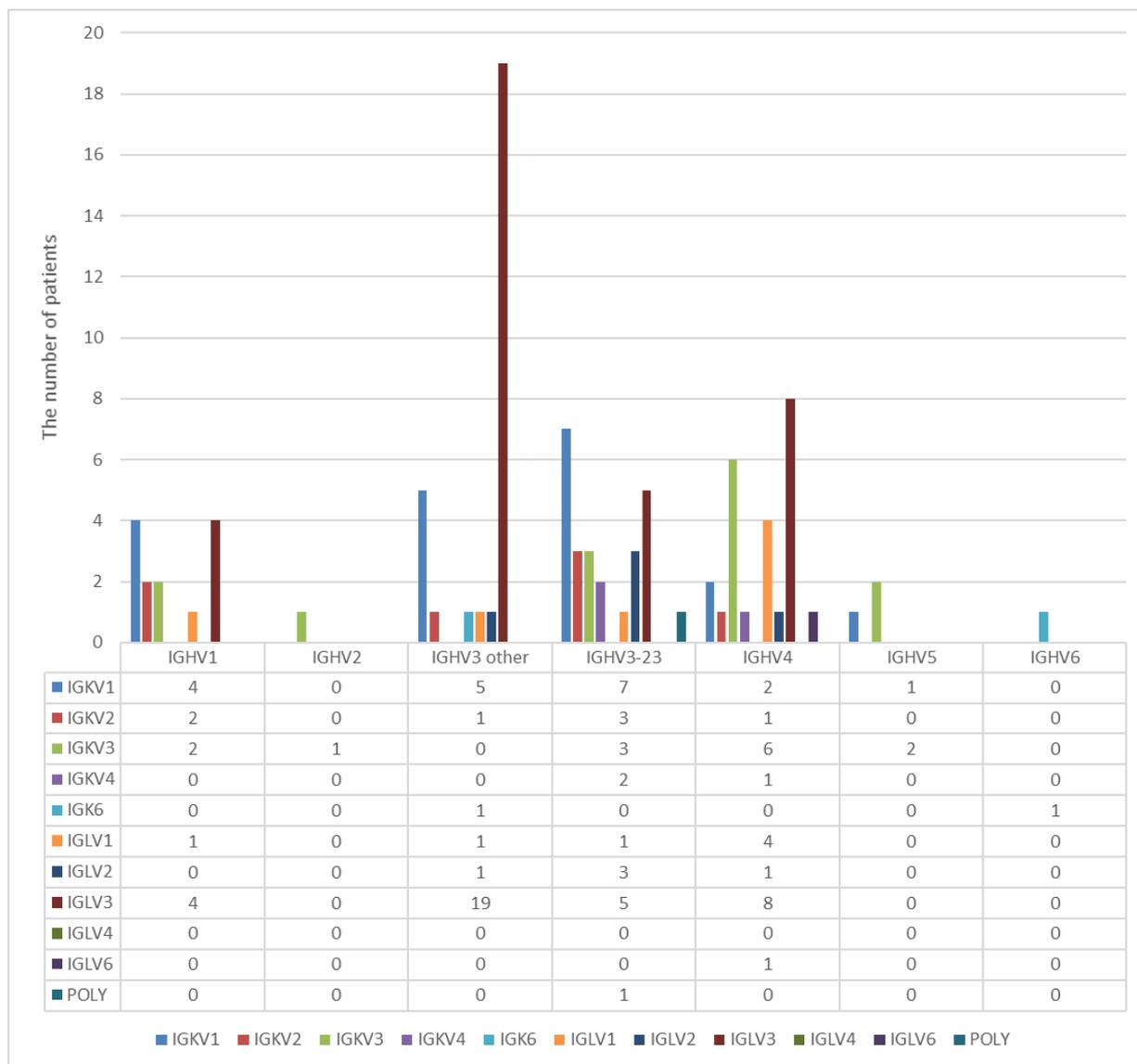


Figure 27: Relationship between *IGHV* and *IGLV* regions and the mutational statuses of different *IGHVs*.

35 Igκ-positive cases expressed 39 productive rearrangements, 42 Igλ-positive cases expressed 42 productive rearrangements. 3 multiple cases expressed 8 productive rearrangements. One case showed poly clonal light chains. There are also five cases who expressed double heavy chain only with single light chain. In order to calculate the relationship of heavy and light chain, we counts these 5 cases twice.

Characterization of the *IGLV* gene using 5' RACE

Another advantage of 5' RACE is that it provides information about the sequence of the immunoglobulin light chain. In our series, we characterized a monoclonal light chain in 80 cases, whereas only 1 case appeared to express polyclonal light chains. Different heavy chains showed a bias to combined with different light chain (Figure 27). And in the Table 16 and Table 17, we have summarized the top 5 rearrangements in κ and λ chain in different mutational

status. As the data showed, mutated and unmutated light chain showed a bias in using the different rearrangements.

Table 16: The top 5 usage of IgκV in CLL patients.

Rank	IGKV mutated	No.	%among mutated IGKV (n=19)	Igkv unmutated	No.	%among unmutated IGKV (n=24)
1	1-8	4	21.1	2-28	5	20.8
2	1-5	3	15.8	1-39	3	12.5
3	3-20	3	15.8	3-20	2	8.3
4	3-15	2	10.5	3-15	2	8.3
5	1-39	1	5.3	3-11	2	8.3

Table 17: The top 5 usage of IgλV in CLL patients.

Rank	IGLV mutated	No.	% among mutated IGLV (n=17)	IGLV unmutated	No.	% among unmutated IGLV (n=29)
1	3-1	4	23.5	3-21	23	79.3
2	1-40	2	11.8	3-1	1	3.4
3	2-11	2	11.8	1-44	1	3.4
4	3-21	2	11.8	1-47	1	3.4
5	1-44	1	5.9	2-14	1	3.4

Thirty-five patients expressed productive Igκ rearrangements, 42 patients expressed productive Igλ rearrangements, and 3 patients showed a mixed Igκ and Igλ phenotype.

Thirty-one of the 35 Igκ-positive cases expressed a single productive rearrangement (19 unmutated and 12 mutated), and 4 expressed two productive rearrangements (concordant unmutated in one case, concordant mutated in two cases, and discordant unmutated and mutated in the last case). All 42 Igλ-positive cases expressed a single productive rearrangement (26 unmutated and 16 mutated). We also observed a significant bias with 23 of the 26 unmutated Igλ-positive cases expressing IGLV3-21 light chains (**Table 17**). The last 3 cases expressed both Igκ and Igλ sequences, 1 harboring two productive rearrangements of one mutated Igλ and

one unmutated Igκ, 1 with three unmutated productive rearrangements of one Igλ and two Igκ, and 1 showing three productive rearrangements with two unmutated Igλ and one mutated Igκ.

Complete analysis of the LCDR3 region was possible in 80 patients. The CDR3 of the Igκ light chains ranged from 8 to 11 amino acids (median nine amino acids); the CDR3 of the Igλ light chains ranged from 8 to 13 amino acids (median 11 amino acids). Interestingly, we observed that some cases in our series shared the same light chains with very similar CDR3 sequences, especially for those cases that expressed *IGLV3-21* rearrangements (**Table 18**).

Table 18: List of light chains that share identical CDR3 in 80 CLL patients.

Light chain					Heavy chain				
ARN	IGLV	IGLJ	HOM (%)	CDR3	<i>IGHV</i>	IGHD	IGHJ	HOM (%)	CDR3
25	IGKV1-8	IGKJ5	96.5	QQYYDYPST	<i>IGHV3-48</i>	IGHD6-6	IGHJ3	96.7	ARGRAYTSSPYDAFDV
31	IGKV1-8	IGKJ1	96.4	-H--S--W-	<i>IGHV3-23</i>	IGHD1-26	IGHJ4	94.7	AKSLSGSYGQIDY
26	IGKV1-8	IGKJ4	97.1	---G--L-	<i>IGHV3-23</i>	IGHD3-10	IGHJ4	92.4	AKVHSDTEYGSLSY
18	IGKV1-8	IGKJ3	95.9	----F-R-	<i>IGHV3-23</i>	IGHD6-6	IGHJ6	90.2	AKGGSSGSYYGMDV
79	IGKV1D-8	IGKJ1	100	----SF-W-	<i>IGHV1-69</i>	IGHD3-9	IGHJ6	100	ARGNTYYDILTYDGVVYGMVDV
12	IGKV1-39	IGKJ3	100	QQSYSTPPFT	<i>IGHV5-51</i>	IGHD1-26	IGHJ4	100	ARHPWYSGSYSFDY
1	IGKV1-39	IGKJ2	100	-----Y-	<i>IGHV1-2</i>	IGHD6-19	IGHJ4	100	ARLQWLVIHFYDY
2	IGKV1-39	IGKJ3	100	-----	<i>IGHV1-3</i>	IGHD6-19	IGHJ4	100	--E---LR----
24	IGKV1-39	IGKJ4	95.4	---H-I-L-	<i>IGHV3-7</i>	IGHD3-10	IGHJ4	98.6	ARSYSGSYSLRGRFDY
9	IGKV2-28	IGKJ3	100	MQUALQTPGFT	<i>IGHV3-30</i>	IGHD3-10	IGHJ6	100	ARDRTSLGDPPLWGYYYYGMDV
6	IGKV2-28	IGKJ4	100	-----L-	<i>IGHV3-23</i>	IGHD3-3	IGHJ6	99.5	AKVTLEWLGYGMDV
11	IGKV2-28	IGKJ1	100	-----	<i>IGHV1-58</i>	IGHD3-3	IGHJ4	100	AASYDFWSGYSH
81	IGKV2-28	IGKJ2	99.1	-----Y-	<i>IGHV3-23</i>	IGHD6-19	IGHJ4	87.4	ASGGPYGVAGNPFYDY
79	IGKV2-28	IGKJ4	100	-----Q-	<i>IGHV1-3</i>	IGHD4-23	IGHJ3	100	ARDDSGGYAFDI
10	IGKV3-15	IGKJ2	100	QQYNNWPPSYT	<i>IGHV5-51</i>	IGHD3-3	IGHJ4	100	ARRNYDFWSGYTPAPFDY
8	IGKV3-15	IGKJ2	100	-----L--	<i>IGHV3-23</i>	IGHD4-17	IGHJ4	99.5	ANGPRGDYVFPFDY
27	IGKV3-15	IGKJ2	94.7	-----F-	<i>IGHV3-23</i>	IGHD2-8	IGHJ4	89.6	AKASVLQWVSSTYYFDY
36	IGKV3-15	IGKJ2	96.3	---DT--R--	<i>IGHV5-51</i>	IGHD5-18	IGHJ6	93.5	VKGGYSYGYHYMDV
76	IGLV3-1	IGLJ2	96.5	QAWDTSTAAYV	<i>IGHV4-59</i>	IGHD6-19	IGHJ4	90.0	ARDLRSSGWSPYFDY
19	IGLV3-1	IGLJ2	99.1	---S--VV	<i>IGHV3-33</i>	IGHD3-3	IGHJ6	98.6	AKDAALTVSCMDV

45	IGLV3-1	IGLJ2	97.4	---S-- M-	IGHV4-39	IGHD2-21	IGHJ4	97.6	GRRYGDSTNAIFDY
73	IGLV3-1	IGLJ2	94.8	---S-- --	IGHV4-34	IGHD3-9	IGHJ3	95.5	ARKQGEDMALRYFDWSYAFDI
77	IGLV3-1	IGLJ1	94.8	---S---Y-	IGHV4-34	IGHD6-6	IGHJ6	94.3	ARVAGRPYGMVDV
55	IGLV3-21	IGLJ3	99.6	QVWDSSSDHPWV	IGHV3-21	IGHD1-14	IGHJ6	96.9	ATDRNGMDV
56	IGLV3-21	IGLJ3	98.3	----G-----	IGHV3-21	NA	IGHJ6	97.4	TR--T----
53	IGLV3-21	IGLJ3	99.4	-----	IGHV3-21	IGHD3-9	IGHJ6	97.0	-L-----
61	IGLV3-21	IGLJ3	99.6	-----	IGHV3-21	NA	IGHJ6	94.7	-A-----
54	IGLV3-21	IGLJ3	99.1	-----	IGHV3-21	NA	IGHJ6	96.1	-R-A-V---
38	IGLV3-21	IGLJ3	97.0	--E---Q--	IGHV3-21	NA	IGHJ6	97.8	-R-ESV--
52	IGLV3-21	IGLJ3	99.3	-----	IGHV3-15	IGHD4-17	IGHJ4	97.5	TADGDYRGTLRN
16	IGLV3-21	IGLJ3	100	-----	IGHV3-15	IGHD3-16	IGHJ5	99.4	TTDSFS
17	IGLV3-21	IGLJ1	99.1	-----L -	IGHV3-15	IGHD2-21	IGHJ4	98.6	TTDLHCGGDCNFYD
51	IGLV3-21	IGLJ1	98.7	-----LY-	IGHV1-24	IGHD3-10	IGHJ4	95.7	ATGGGSGSYTFDY
18	IGLV3-21	IGLJ3	99.6	-----L-M	IGHV3-23	IGHD6-25	IGHJ4	98.1	AKDLGDSSGYFDY
63	IGLV3-21	IGLJ1	98.7	-----Q-Y-	IGHV3-23	IGHD1-26	IGHJ4	97.2	--A-ELQ ----
58	IGLV3-21	IGLJ3	98.3	-----	IGHV3-23	IGHD5-24	IGHJ4	96.3	---P-SDGFKKN
57	IGLV3-21	IGLJ3	98.3	----G-----	IGHV3-23	IGHD6-6	IGHJ4	97.4	--GD-T---H---
68	IGLV3-21	IGLJ2	98.9	-----VRG -	IGHV3-30	IGHD5-12	IGHJ6	100	ARDKGYDPSLGMDV
20	IGLV3-21	IGLJ3	99.1	-----	IGHV3-48	IGHD6-13	IGHJ4	98.2	ARVSSSYFDY
23	IGLV3-21	IGLJ3	99.1	-----	IGHV3-48	IGHD1-26	IGHJ4	98.0	ATDGVGAPH
59	IGLV3-21	IGLJ3	99.6	-----	IGHV3-48	IGHD2-15	IGHJ4	97.1	ASDPHCSGGSCYSYGY
64	IGLV3-21	IGLJ3	99.6	-----	IGHV3-48	IGHD4-17	IGHJ4	97.7	ARAVTTGGGC
42	IGLV3-21	IGLJ1	97.9	-----D---Y-	IGHV3-48	IGHD3-10	IGHJ6	97.6	ARDFGGETVQPPWNV
60	IGLV3-21	IGLJ3	99.1	-----	IGHV3-53	IGHD1-26	IGHJ4	96.9	ARGQGANDY
65	IGLV3-21	IGLJ1	98.7	-----G- -F	IGHV3-74	IGHD3-10	IGHJ3	95.4	ARDDSGGNSFDI
62	IGLV3-21	IGLJ1	99.6	-----L-Y-	IGHV3-74	IGHD5-18	IGHJ4	97.6	AGDLVDTAGGVFDY
22	IGLV3-21	IGLJ1	99.7	-----L Y-	IGHV4-59	IGHD2-8	IGHJ4	99.2	ARAPGDTGGYRIDY
66	IGLV3-21	IGLJ1	98.7	-----L-Y-	IGHV4-59	IGHD4-11	IGHJ3	97.4	AVWTVTDPFDI

c. The advantage and disadvantages of the two methods

Compared with the two methods, although the routine test is relatively cheap and shows the data more directly because of the gel electrophoresis, there are also some disadvantages (**Table 19**). The Sanger sequencing doesn't have insight into subclonal architecture and intracloal diversity and can't combine with other assays in the same program. Even for the target segment,

the PCR-amplified transcripts are necessarily shorter, which leads high risk of an inaccurate diagnosis of immunoglobulin and calculation of percent identity of mutation. Last but not least, many manuals manipulate the routine method, which can increase labor costs and the possibility of making mistakes. So this method is just suitable for a small number of cases. However, the 5'RACE can solve these shortcomings. Because of the NGS workflow, 5'RACE makes it possible to operate more than one assay simultaneously. Though it needs a relatively long time to sequence the amplification, it adapts to a large number of cases. Automated operation saves more time and human resources and reduces the possibility of making errors. And a large number of test cases can reduce sequencing financial and timing costs.

What's more, the results of 5'RACE provide more details insight into subclonal architecture and intraclonal diversity. With more and more research on IGH in CLL, the need to detect more rearrangement is necessary¹⁶⁶. 5'RACE can circumvent the requirement of laborious techniques in cases with biallelic rearrangement. But data interpretation of 5'RACE is more complicated and not yet standardized, so we need more tests and research.

Table 19: Compare two methods.

	Sanger sequencing	5'RACE
Pros	Relatively cheap	Cooperated with other NGS-assays
	Suitable for a small scale cases	Suitable for large-scale cases
	Data investigate directly	More details of subclonal architecture and intraclonal diversity
		Avoid laborious (cloning) techniques in cases with multiple rearrangements
		Automatic processing and analysis can reduce the operation and errors
Cons	No date of subclonal architecture and intraclonal diversity	Relatively costly ☹
	No combination with other assays in one workflow is possible	Unavailable dedicated bioinformatics tools ☹
	PCR-amplified transcripts are necessarily shorter	Data is more complicated, and the analysis system is not yet standardized

	high risk of misamplification and inaccurate calculation of percent identity	
	There is a financial and human cost	

2) The mutational relationship between *IGHV* and *IGLV*

The relationship between the somatic mutational status of the heavy and light chains

We observed a strong relationship between the mutational status of the heavy chain and the light chain in this cohort (**Table 20 and Figure 28**). In most cases, the mutational status of *IGHV* rearrangements was associated with the mutational status of *IGK/IGL* rearrangements, even though multiple light chain rearrangements were noted in some cases. For 3 cases co-expressing $Ig\kappa$ and $Ig\lambda$ light chains, the mutational status of the heavy and light chains were concordant in 2 cases. One expressed three unmutated rearrangements of heavy and light chains, and the other expressed one mutated and two unmutated rearrangements of heavy and light chains. Finally, the last case expressed a mutated heavy chain and two productive light chains with different mutational statuses.

Table 20: The relationship of the mutational status of heavy and light chains.

	<i>IGHV</i>	<i>IgLV</i>	
26	U	U	
1	U	M	1 borderline case
30	M	M	
19	M	U	<div style="display: flex; align-items: center; justify-content: space-between;"> <div style="text-align: center;"> <p>16 <i>IgLV3-21</i></p> <p>3 disagreement</p> </div> <div style="text-align: center;"> <p>15 borderline cases</p> <p>5 subset 2</p> </div> </div>
1	M	M+U	
1	M	POLY	
3	U+M	U+M	

Of the 35 $Ig\kappa$ -positive cases, the mutational statuses of the heavy and light chains were concordant in 30 cases, with 16 double-unmutated and 14 double-mutated cases. One case expressed two productive heavy chains (one mutated and one unmutated) and two productive

light chains (one mutated and one unmutated). Only 4 cases showed a discordant profile with a mutated heavy chain and an unmutated light chain.

Of the 42 Igl λ -positive cases, the mutational statuses of the heavy and light chains were concordant in 25 cases, with 9 double-unmutated and 16 double-mutated cases. One patient expressed double mutated and unmutated heavy chain with a unmutated light chain. All 16 other cases expressed a mutated heavy chain and an unmutated light chain. Strikingly, all 16 cases expressed an IGLV3-21 light chain with similar CDR3 sequences (QVWDSSSDHPWV).

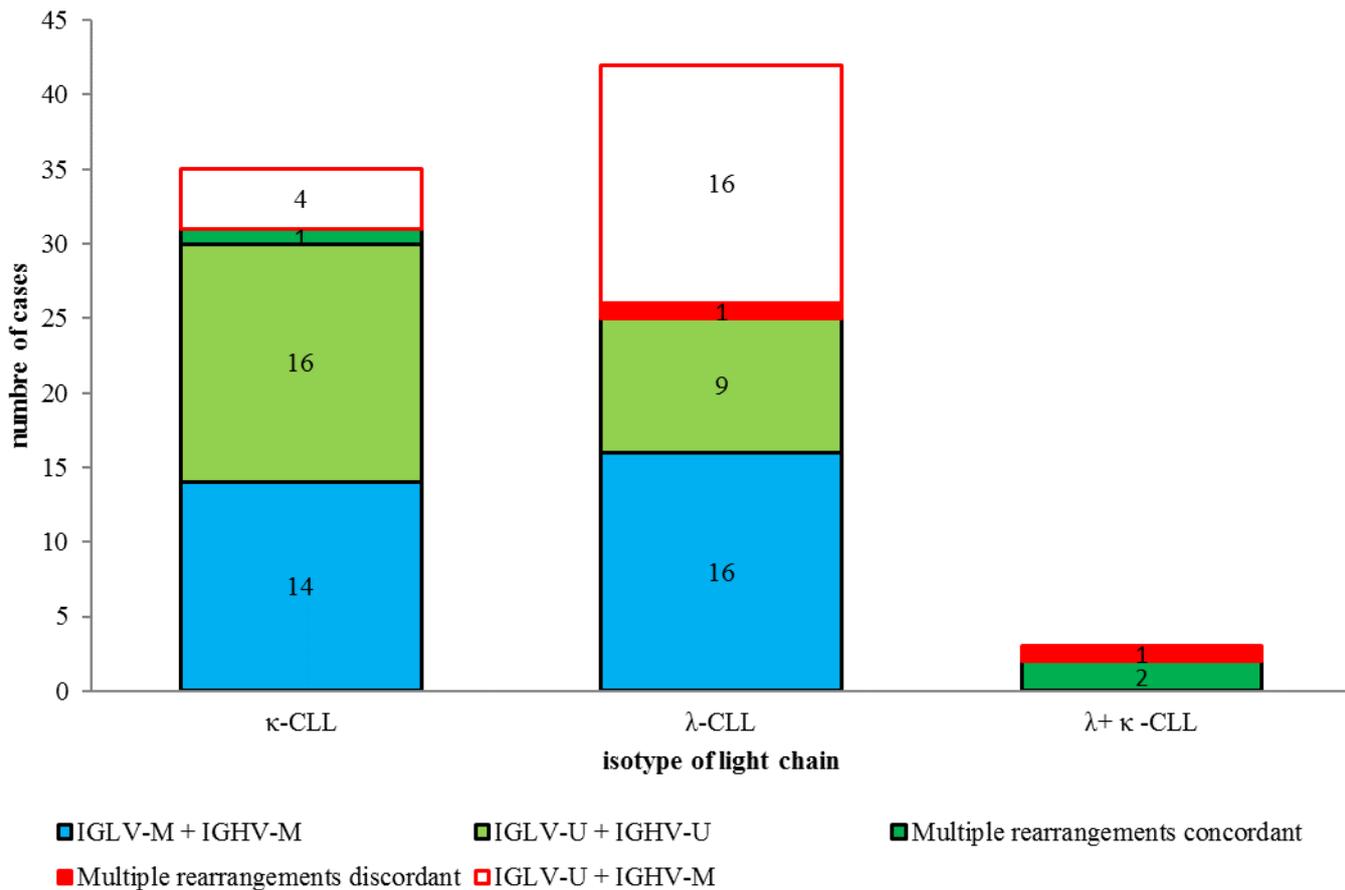


Figure 28: The relationship of the mutational status of heavy and light chains.

In different isotypes, the relationship of the mutational status of heavy and light chains. Green represents the concordant unmutated status of heavy and light chains. Blue represents the concordant mutated status of heavy and light chains. Dark green represents concordant mutational statuses of multiple rearrangements of heavy and light chains. Red means the discordant mutational status of heavy and light chains (the blank red box represents a single rearrangement, and the solid red box represents multiple rearrangements).

3) The value of IGLV

a. Find a new kind of light chain

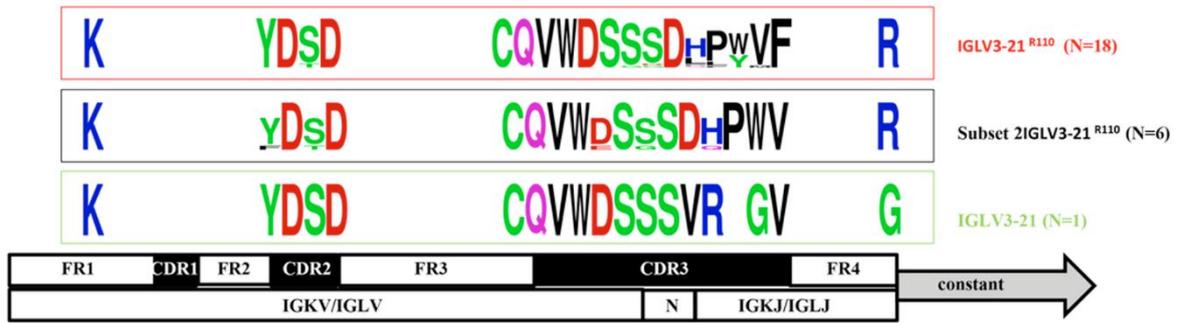


Figure 29: Sequence of all IGLV3-21 alleles, including the IGLV3-21, IGLV3-21^{R110}, and subset 2 IGLV3-21^{R110} alleles.

Then, we assessed all light chains and found that 31% (25/80) were IGLV3-21. The light chain in 25 IGLV3-21-CLL patients always shared a similar LCDR3, and most of them (24/25) had a point mutation, termed R110, between the variable and constant junction, named IGLV3-21^{R110} (**Figure 29**). All IGLV3-21^{R110} cases carried germline K16 residue, and 20/24 cases expressed the germline YDSD motif required for homotypic BCR-BCR interaction. The remaining 4 cases carried a motif changed by one residue, which has similar properties in all cases (three YDTD and one FDS^D).

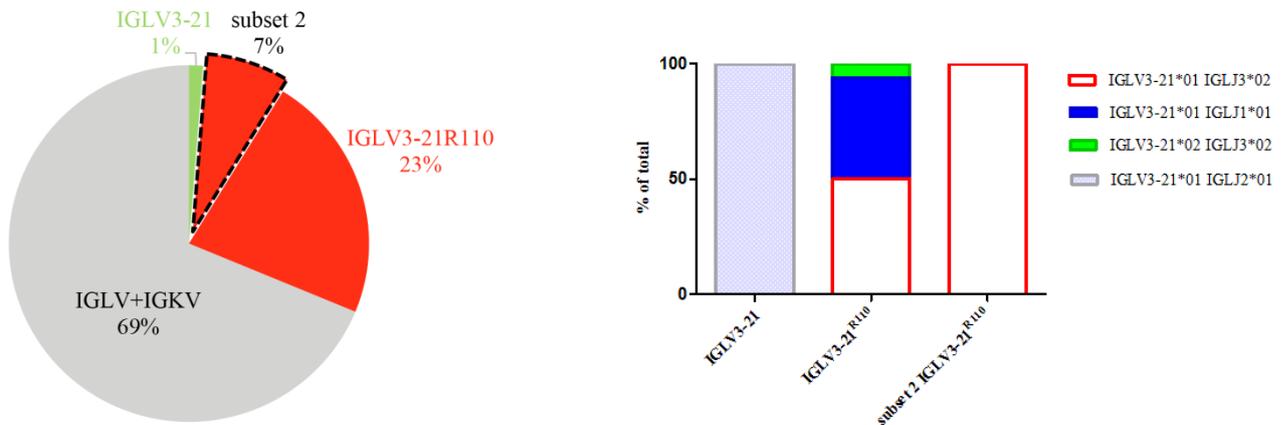


Figure 30: The information on IGLV3-21.

The pie chart showed the mutational relationship of IG heavy and light chains based on different types of light chains. And the bar chart showed the distribution of the J segment in all IGLV3-21 rearrangements.

In these 24 IGLV3-21^{R110}-CLL patients, six belonged to subset#2 (**Figure 30** left). The light chain rearrangements differed among subset 2 IGLV3-21^{R110}-CLL, IGLV3-21^{R110}-CLL, and IGLV3-21-CLL patients (**Figure 30** right). All six subset 2 IGLV3-21^{R110}-CLLs expressed IGLV3-21*01 with IGLJ3*02, whereas only half of the IGLV3-21^{R110}-CLL patients (9/18) showed identical rearrangements. The other half of IGLV3-21^{R110}-CLL patients (8/18)

showed IGLV3-21*01 with IGLJ1*01. Only one patient with IGLV3-21 without the R110 mutation expressed IGLV3-21*01 with IGLJ2*01.

Table 21: The sequence of all the IGLV3-21 CLL patients.

Num	IgH		Mut %	M/U	IgL		Mut %	M/U
	IGHV	CDR3			IgLV	CDR3		
22451	IGHV3-21	A R D L S P G D I L T G Y S R W F D P	100	U	IgKV1-33	Q Q Y D N L L P T	100	U
33901	IGHV3-21	A R D S G S Y R S Y Y F D Y	100	U	IgKV6-21	H Q S S S L P W T	100	U
28959	IGHV3-21	T R D R T G M D V	97.386	M	IgLV3-21	Q V W D S G S D H P W V	98.276	U
25640	IGHV3-21	A L D R N G M D V	97.006	M	IgLV3-21	Q V W D S S S D H P W V	99.441	U
34099	IGHV3-21	A A D R N G M D V	97.92	M	IgLV3-21	Q V W D S S S D H P W V	99.565	U
30176	IGHV3-21	A R D A N V M D V	96.078	M	IgLV3-21	Q V W D S S S D H P W V	99.057	U
22743	IGHV3-21	A T D R N G M D V	96.903	M	IgLV3-21	Q V W D S S S D H P W V	99.565	U
36281	IGHV3-21	A R D E S V M D V	97.758	M	IgLV3-21	Q V W E S S S D Q P W V	96.957	M
22850	IGHV1-24	A T G G G S G S Y T F D Y	97.22	M	IgLV3-21	Q V W D S S S D L Y V	98.696	U
21731	IGHV3-15	T T D L H C G G D C N F D Y	98.558	U	IgLV3-21	Q V W D S S S D H L V	99.142	U
21184	IGHV3-15	T A D G D Y R G T L R N	97.468	M	IgLV3-21	Q V W D S S S D H P W V	99.31	U
11252	IGHV3-15	T T D S F S	99	U	IgLV3-21	Q V W D S S S D H P W V	100	U
25104	IGHV3-23	A K A L E L Q G Y F D Y	97.183	M	IgLV3-21	Q V W D S S S D Q P Y V	98.712	U
15784	IGHV3-23	A K D P G S D G F K K N	97.22	M	IgLV3-21	Q V W D S S S D H P W V	98.261	U
30629	IGHV3-23	A K G D G T S S G H F D Y	97.484	M	IgLV3-21	Q V W D S G S D H P W V	98.324	U
34234	IGHV3-23	A K D L G D S S G Y F D Y	98.104	U	IgLV3-21	Q V W D S S S D L P W M	99.558	U
30838	IGHV3-48	A R V G S S S Y F D Y	98.148	U	IgLV3-21	Q V W D S S S D H P W V	99.13	U
28186	IGHV3-48	A S D P H C S G G S C Y S G Y	97.059	M	IgLV3-21	Q V W D S S S D H P W V	99.565	U
24345	IGHV3-48	A R A V T T G G G C	97.059	M	IgLV3-21	Q V W D S S S D H P W V	99.565	U
28421	IGHV3-48	A R D F G G E T V Q P P W N V	97.706	M	IgLV3-21	Q V W D S S D D H P Y V	97.854	M
28652	IGHV3-48	A T D G V G A P H	98.013	U	IgLV3-21	Q V W D S S S D H P W V	99.13	U

28534	IGHV3-48	A R G Q G A N D Y	96.847	M	IgLV3-21	Q V W D S S S D H P W V	99.13	U
21946	IGHV3-53	A G D L V D T A G G V F D Y	97.576	M	IgLV3-21	Q V W D S S S D L P Y V	99.558	U
11199	IGHV3-74	A R D D S G G N S F D I	96.79	M	IgLV3-21	Q V W D S S G D P F	98.673	U
22500	IGHV4-59	A V W T V T D P F D I	97.345	M	IgLV3-21	Q V W D S S S D L P Y V	98.696	U
14034	IGHV4-59	A R A P G D T G G Y R I D Y	98.95	U	IgLV3-21	Q V W D S S S D L Y V	99.648	U
27953	IGHV3-7	A R S Y G S G S Y S L R G R F D Y	98.26	U	IgKV1-39	Q Q S H S I P L T	95.349	M
18699	IGHV3-30	A K D A A L T V S C M D V	98.61	U	IgLV3-1	Q A W D S S T V V	99.134	U
37336	IGHV3-30	A K E S Y S S S W P S Y Y F D Y	97.22	M	IgLV2-11	C S Y A G T Y T Y L	91.903	M
14469	IGHV3-30	A R G S E E V D Y D D M E Y Y Y Y Y M D V	96.75	M	IgKV1-5	Q Q Y D S H P Y T	94.982	M

b. The relationship between *IGHV* and *IGLV3-21*

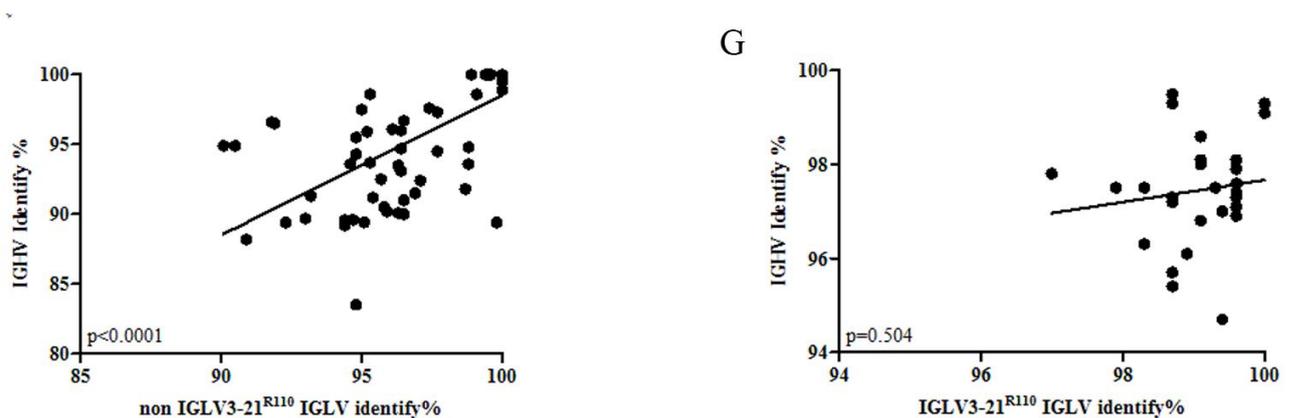


Figure 31: The identification of *IGHV* and *IGLV* in different groups.

Left scatter plots of the identification of *IGHV* and *IGLV* in non-*IGLV3-21*^{R110} CLL (N=56). And right scatter plots for identifying *IGHV* and *IGLV* in *IGLV3-21*^{R110} CLL (N=24).

Compared with other light chains, *IGLV3-21*^{R110}-CLL was more likely to exhibit a discordant mutational status of heavy and light chains (16/20 cases) (**Table 21**). We want to validate the value of light and if, in the same patients, there always exists a light chain expressing the same mutational status as the heavy chain. If the relationship works, in borderline cases, we can test the mutational status of the light chain to aid the diagnosis and predict the prognosis of CLL patients. As we can see in the **Figure 31**, the identification of the heavy chain and light chain showed a linear correlation in the non-*IGLV3-21*^{R110} group ($p < 0.0001$), which means that in most CLL patients, the mutational status of the light chain is concordance with a heavy chain.

Still, it isn't suitable for the IGLV3-21^{R110} group (p=0.504), which advises the critical role of IGLV3-21^{R110} in the development and progress of CLL.

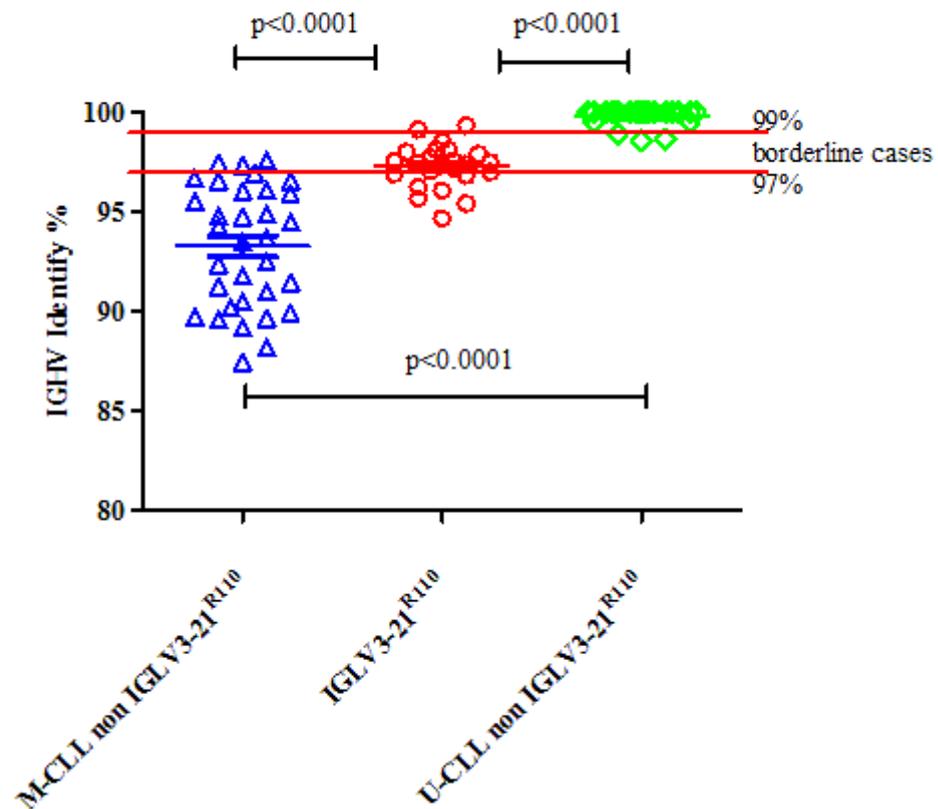


Figure 32: Identifying *IGHV* in M-CLL non-IGLV3-21^{R110}, IGLV3-21^{R110}, and U-CLL non-IGLV3-21^{R110}.

So that we divided patients into three groups based on the mutation status of the heavy chain and the point mutation of IGLV3-21: M-CLL non-IGLV3-21^{R110}, IGLV3-21^{R110} and U-CLL non-IGLV3-21^{R110}, the interesting thing is that when we compared the identification of the heavy chain in different groups, we found that IGLV3-21^{R110} was more likely to exhibit approximately 97-99% homology, representing borderline cases (Figure 32). The results indicated the importance of the light chain in CLL, especially in borderline cases.

c. The clinical value of the IGLV3-21^{R110}

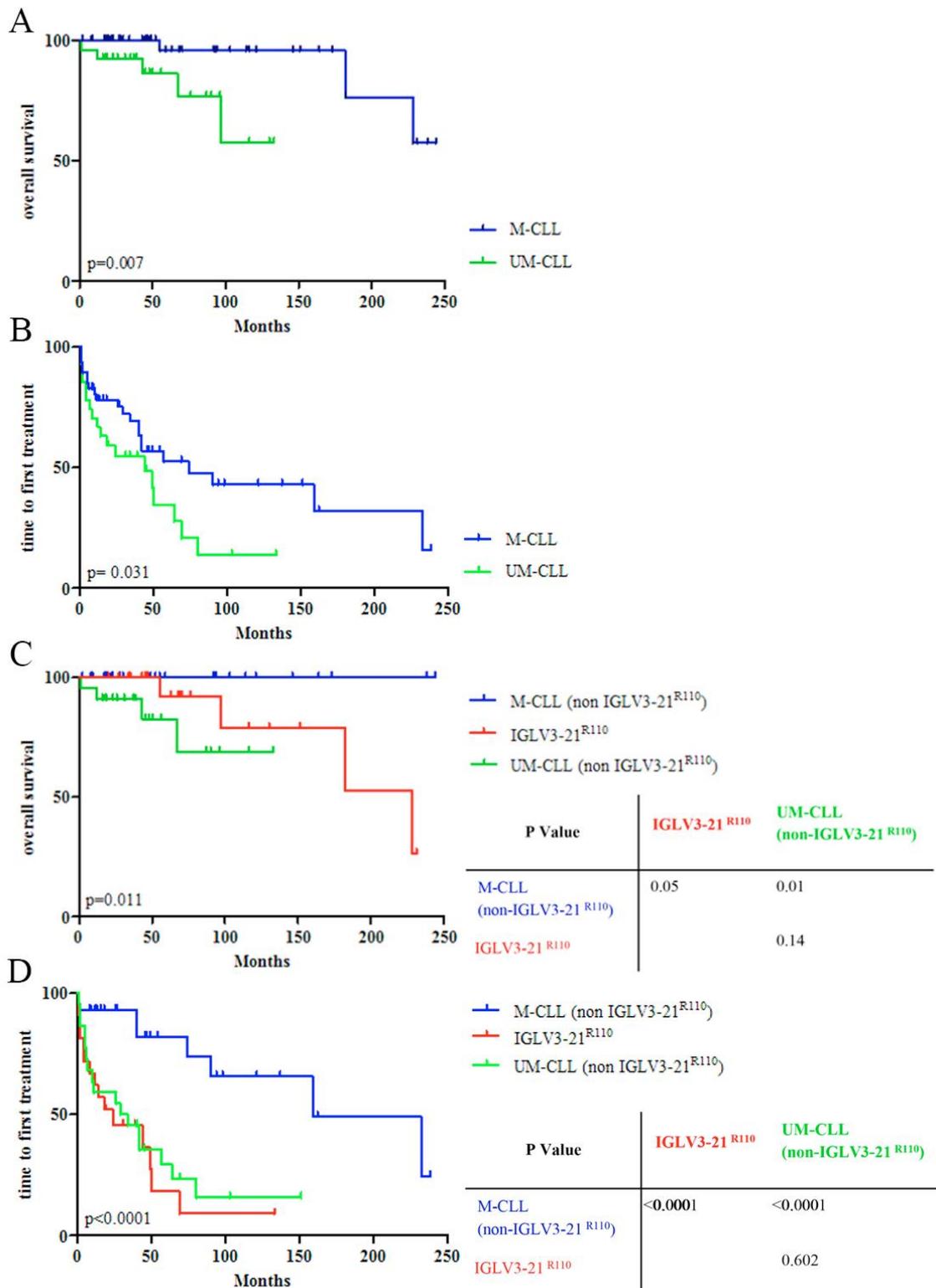


Figure 33: The survival analysis of different groups.

(A) Overall survival analysis of M-CLL and U-CLL ($p < 0.05$). (B) TTFT analysis of M-CLL and U-CLL ($p = 0.031$). (C) Overall survival analysis of M-CLL non-IGLV3-21^{R110}, IGLV3-21^{R110}, and U-CLL non-IGLV3-21^{R110} ($p < 0.05$). (D) TTFT analysis of M-CLL non-IGLV3-21^{R110}, IGLV3-21^{R110}, and U-CLL non-IGLV3-21^{R110} ($p < 0.0001$).

Table 22: Clinical information of CLL patients.

	IgG-CLL N=13	IgM-CLL N=67	p	IgLV3- 21 ^{R110} -CLL N=24	M-CLL (non- IGLV3-21 ^{R110}) N=33	U-CLL (non- IGLV3-21 ^{R110}) N=23	p
Age at diagnosis (years) mean ±SD	73.4±10.8	69.8±11.0	0.28	68.5±11.8	71.0±10.7	71.4±10.7	0.62
Gender (F: M)	4 : 9	16 : 51	0.86	3: 21	13: 20	4: 19	0.042
Binet-no. (%)			0.90				0.013
A	6 (46.1%)	38 (56.7%)		9 (37.5%)	24 (72.7%)	11 (47.8%)	
B	3 (23.1%)	18 (26.9%)		7 (29.2%)	4 (12.1%)	10 (43.5%)	
C	2 (15.4%)	8 (11.9%)		6 (25.0%)	2 (6.1%)	2 (8.7%)	
unknown	2 (15.4%)	3 (4.5%)		2 (8.3%)	3 (9.1%)	0 (0.0)	
FISH aberrations, no. (%)							
Del 11q12 (<i>ATM</i>)	2 (15.4%)	6 (9.0%)	0.61	2 (8.3%)	1 (3.0%)	5 (21.7%)	0.17
Del 13q14 (<i>D-LEU2</i>)	5 (38.5%)	31 (46.3%)	0.83	13 (54.2%)	13 (39.4%)	10 (43.5%)	0.39
Trisomy 12 (<i>MDM2</i>)	2 (15.4%)	6 (9.0%)	0.61	1 (4.2%)	2 (6.1%)	5 (21.7%)	0.19
Del 17q13 (<i>TP53</i>)	0 (0.0)	2 (3.0%)	1.00	0 (0.0)	0 (0.0)	2 (8.7%)	0.18
Del 17q13 (<i>TP53</i> exon 1)	0 (0.0)	1 (1.5%)	1.00	0 (0.0)	0 (0.0%)	1 (4.3%)	0.61
Trisomy 12 (<i>POU6F1</i>)	2 (15.4%)	4 (6.0%)	0.17	1 (4.2%)	2 (6.1%)	3 (13.0%)	0.66
Del 6q21 (<i>SEC63</i>)	2 (15.4%)	2 (3.0%)	0.08	1 (4.2%)	2 (6.1%)	1 (4.3%)	0.68
unknown	4 (30.8%)	7 (10.4%)		4 (16.7%)	6 (18.2%)	1 (4.3%)	
<i>TP53</i> gene, no. (%)	1 (7.7)	4 (6.0)	0.53	0 (0.0)	2 (6.1%)	3 (13.0%)	0.22
Karyotype no. (%)							

Del11	2 (15.4%)	4 (6.0%)	0.28	2 (8.3%)	1 (3.0%)	3 (13.0%)	0.86
Tri12	3 (23.1%)	8 (12.0%)	0.61	0 (0.0)	4 (12.1%)	7 (30.4%)	0.011
Del13	2 (15.4%)	10 (15.0%)	1.00	6 (25.0%)	4 (12.1%)	2 (8.7%)	0.35
Del6	2 (15.4%)	2 (3.0%)	0.14	1 (4.2%)	1 (3.0%)	2 (8.7%)	1.0
unknown	2 (15.4%)	15 (22.4%)		3 (12.5%)	12 (36.4%)	2 (8.7%)	
SHM IGH (M: U)	9: 4	42: 25	0.98	18: 6	33:0	0:23	
Same SHM IgH and IgL no. (%)	13 (100)	46 (68.7)	0.04	8 (33.3%)	29 (87.9%)	22 (95.7%)	<0.0001

Laboratory parameters at diagnosis

Platelets (150-450 Giga/L) mean \pm SD	276.0 \pm 49.8	218.7 \pm 49.8	0.00	215.0 \pm 56.2	212.8 \pm 67.7	252.2 \pm 97.3	0.31
Median (min, max)	271.0 (219.0, 330.0)	225.0 (80.0, 438.0)	2	212.0 (80.0, 320.0)	233.0 (85.0, 330.0)	250.0 (120.0, 438.0)	
LDH (240-480 UI/L) mean \pm SD	415.4 \pm 67.0	367.3 \pm 167.2	0.01	335.9 \pm 90.1	396.2 \pm 233.6	388.0 \pm 106.0	0.18
Median (min, max)	387.0 (349.0, 523.0)	353.0 (125.0, 1288.0)	3	325.0 (177.0, 562.0)	366.0 (125.0, 1288.0)	376.0 (233.0, 670.0)	
sIgM (0.4-2.3 g/L) mean \pm SD	0.56 \pm 0.27	0.52 \pm 0.32	0.71	0.38 \pm 0.30	0.60 \pm 0.31	0.61 \pm 0.35	0.025
Median (min, max)	0.5 (0.3, 1.0)	0.45 (0.1, 1.4)		0.3 (0.1, 0.9)	0.60 (0.1, 1.2)	0.60 (0.2, 1.4)	
Beta-2 microglobulin (0.8-2.4 mg/L) mean \pm SD	2.1 \pm 0.64	2.61 \pm 1.13	0.53	2.63 \pm 1.07	2.03 \pm 0.43	3.16 \pm 1.43	0.92
Median (min, max)	1.9 (1.3, 3.0)	2.3 (1.4, 6.4)		2.4 (1.4, 5.0)	2.1 (1.3, 3.0)	2.9 (1.5, 6.4)	

We have summarized the clinical data of all the patients (**Table 22**). A significant difference in survival was noted between UM-CLL and M-CLL, previously reported in many studies (**Figure 33 A and B**). As noted above, we divided patients into three groups (M-CLL non-IGLV3-21^{R110}, IGLV3-21^{R110}-CLL, and U-CLL non-IGLV3-21^{R110}). We found that IGLV3-21^{R110}-CLL and

U-CLL non-IGLV3-21^{R110} were more likely to occur in males than females ($p=0.042$), and these patients were more often classified as Binet stage B or C than A ($p=0.013$). Next, we analyzed the predictive value of IGLV3-21^{R110} ($n=24$). Expectedly, M-CLL non IGLV3-21^{R110} ($n=33$) showed a longer OS and TTFT than UM-CLL non IGLV3-21^{R110} ($n=23$) (**Figures 33 C and D**, $p=0.001$ and $p<0.0001$, respectively). Then, we compared the prognosis of IGLV3-21^{R110}-positive cases. Regardless of the mutational status of *IGHV*, IGLV3-21^{R110}-positive cases presented a similarly poor prognosis compared with UM-CLL patients ($p=0.14$ for OS and $p=0.602$ for TTFT), even though most of them (18/24) presented mutated *IGHV* and showed a shorter survival time than M-CLL non-IGLV3-21^{R110} ($p=0.05$ for OS and $p<0.0001$ for TTFT). IGLV3-21^{R110}-CLL showed a short TTFT and OS, similar to UM-CLL cases, whereas M-CLL non-IGLV3-21^{R110} presented a favorable prognosis. IGLV3-21^{R110} defines a CLL subgroup representing a clinically aggressive group and is exclusively defined by the mutational status of the light chain regardless of *the IGHV*.

C. Conclusion and discussion

CLL, a most popular leukemia disease, generally occurs in older patients. Only 3.8% of patients with CLL were younger than 50 in our study. More male than female patients (2.86 (60 males: 21 females)) were affected in this study. Over the past two decades, more and more molecular and clinical studies have reported that immunoglobulin BCR sequences are the ideal biologic marker for CLL. This marker is present from the birth of every CLL clone and remains stable over time, *IGHV* mutations never change over time, and thus represent the fingerprint of the disease¹⁶⁷, and is an independent prognostic factor of CLL, which was also identified in this study^{94,97}. However, progress has also raised new questions, such as assessing the prognosis of borderline cases and cases expressing multiple rearrangements, the value of the light chain, and the difference between IgG-CLL and IgM/D-CLL. It highlights the need to extend the research on the immunoglobulin gene. Not only on the mutational status of the *IGHV* but also a comprehensive understanding of the biological and clinical value of classification and characteristic of stereotype and more information and value of light chains¹⁶¹.

The current routine method of detection sequences of the *IGHV* transcript requires analysis of clones with gel electrophoresis and six subsequent rounds of PCR amplification by the primers of leader region^{120,168}. Sanger sequencing was used to obtain each clone's information and calculate the mutational status identity from the germline gene websites. Although this method is relatively cheap and more directly presents the data based on gel electrophoresis, there are also some shortcomings. Sanger sequencing does not provide insight into subclonal architecture and intraclonal diversity and cannot be combined with other assays in the same program. Even for the target segment, the PCR-amplified transcripts are shorter, which leads to a high risk of an inaccurate diagnosis of immunoglobulin and calculations of the percent identity of the mutation. Finally, numerous manual manipulations were performed in the routine method, which caused a waste of time and human resources. For example, after detecting the clonal rearrangements by gel electrophoresis, the amplified products must then be extracted from the gel, purified, and sequenced, increasing labor costs and the possibility of making mistakes. This

method can also be complicated when several rearrangements are present. Therefore, this method is suitable for only a small number of cases.

Our method overcomes these shortcomings. Because of the NGS workflow, 5' RACE allows it to expand the field of using RNA-sequencing assays, directly leading to clinical application development in CLL. With concordance levels greater than 90% concerning the mutation status and the characterization of the V, D, and J segments compared with routine methods, 5' RACE appears more stable and sensitive in handling multiple rearrangements and poor-quality RNA samples. The 5' RACE can provide more data on subclonal and intraclonal diversity of immunoglobulin in CLL. Given the increasing research on *IGH* in CLL, the need to detect more rearrangement is necessary¹⁶⁶ (**Figure 34**). In addition, this kind of RNA-sequencing approach can also extend to the light chains and the details of the T-cell receptor at the same time. Although most studies investigating the significant value of predicting survival have focused exclusively on the heavy chain^{10,97}, some researchers have found that the light chains can also influence the function of BCR and showed a predictive value in the clinical analysis¹⁶⁹. With the development of CLL studies, there are an increasing number of problematic cases¹¹⁹, and it is not easy to evaluate the treatment programs and prognosis of patients by simply determining the mutation status of the heavy chain. Through 5' RACE, we obtained more details of the Ig gene, allowing us to get more BCR pathway information and closely understand immunoglobulin's value in CLL. When using NGS-based methods to determine *IGHV* gene SHM status, it is necessary to consider that this biomarker is useful for prognosis and aids in guiding therapeutic decisions. All processes must be highly stable and reproducible, including lab workflow, library preparation, and bioinformatics data analysis system.

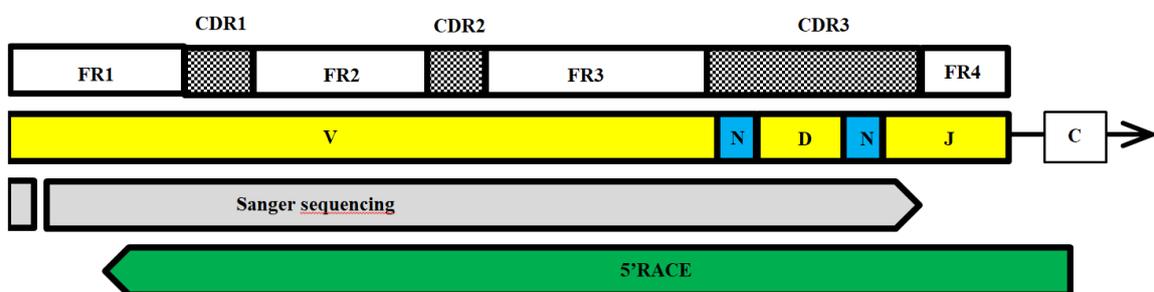


Figure 34: The different characteristics of the two methods.

Through 5' RACE, we found that most cases expressed identical mutational statuses of the heavy and light chains, except IGLV3-21. Interestingly, 92% (23/25) of IGLV3-21 light chains showed unmutated status considering a strict 98% cut-off, but their corresponding heavy chains frequently expressed mutated with an identity between 97-99% (16/25). Then, we analyzed the characteristics of IGLV3-21-positive cases, which were the bias to particular heavy chains. According to our study, the *IGHV3* family represented 60% (15/25) of all *IGHVs*/IGLV3-21, especially *IGHV3*-21, *IGHV3*-48, and *IGHV3*-23. This finding confirms that the heavy and light chains in CLL patients are globally conserved and biased^{31,103,123,162}. Some reports found a different RNA expression characteristic between IGLV3-21 and others^{170,171}. In our method,

we found that 94% (24/25) of IGLV3-21, which showed a similar CDR3 (QVWDSSSDHPWV), harbored a point mutation in the position between FR4 and the constant region termed R110. Combined with clinical data, we found that IGLV3-21^{R110}-CLL presented a high-risk stage and worse prognosis than M-CLL non-IGLV3-21^{R110}, even though 75% of IGLV3-21^{R110}-CLL (18/24) cases harbored heavy chain mutations. Researchers have also reported that the allele IGLV3-21*01 is more prevalent in IGLV3-21^{R110} patients, including more than half of *IGHV* others/IGLV3-21^{R110} and all of subset#2, which reminds us that IGLV3-21*01 may be a risk factor for IGLV3-21^{R110} positive CLL. As reported, the residues K16 and R110 of one BCR and the residues D50 and D52, made of YDSD motif on the other BCR, can promote the interaction between BCR¹²⁴. Recently, studies have also shown that IGLV3-21^{R110}-CLL, although their corresponding *IGHV* showed mutated status, exhibits a similar clinical outcome to UM-CLL^{22,24}. These findings suggest that the immunoglobulin light chain plays an essential role in Ig genes' clonal selection and somatic hypermutation (SHM). There is a greater need for improved methods for detecting immunoglobulins.

There are also some limitations with 5' RACE that we need to mention. First, expanding high-throughput RNA-sequencing methods in clinical and laboratory environments need a standard NGS working platform and automatic data analysis application and specialists that bring immune biological knowledge, technical experience with NGS, and immune informatics expertise¹⁶⁵. These requirements suggest that this method cannot be applied on a large scale. Second, the data interpretation of NGS is more complicated, and a standardized methodology does not currently exist. Thus, more studies of NGS in BCR, especially in the light chain, are necessary.

In conclusion, the results suggest that 5' RACE could be used as an alternative to routine methods for analyzing immunoglobulin rearrangements in research and clinical assessment of chronic lymphoid leukemia.

II. The expression of *tp53* at mRNA level in CLL and the influence of other important biomarkers

A. The critical biomarkers in CLL

1) Problem

Del17p, in about 5–8% of cases, usually showed a mutation of *tp53* gene, these two genetic abnormalities were usually called *tp53* aberration. No matter whether the del17p or mutated *tp53* can cause the functional inactive of the *tp53* pathway¹⁷². Until now, the mechanism of *tp53* in CLL is not understood very well. As we have discussed before, many biomarkers showed the function in progress of CLL. The del13q in about 55% of CLL patients causes low expression of miRNAs (miR-15a and miR-16-1), which can also influence the expression of *BCL-2*¹⁷³. Del11q is observed in about 10% of CLL patients, and it can influence the expression of the *ATM* gene, which code the DNA damage kinase. The other crucial somatic mutation gene included *NOTCH1*, *XPO1*, *KLHL6*, *MYD88*, and *SF3B1*^{14,174} (**Figure 35**).

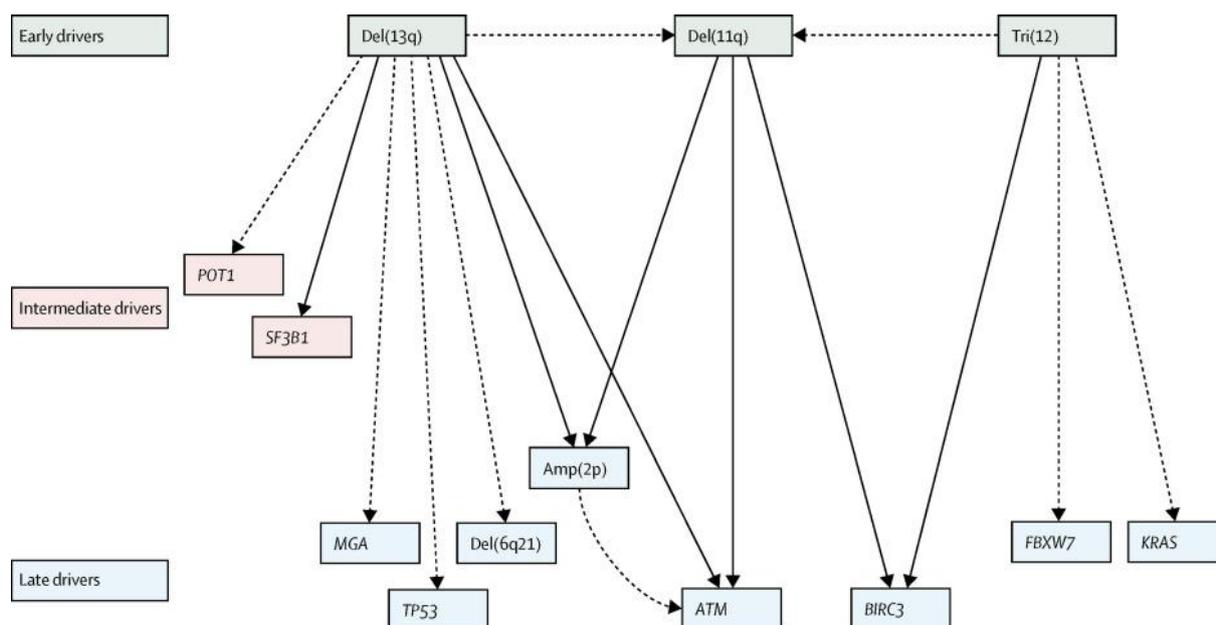


Figure 35: Common genetic markers in CLL

With understanding the function of each molecule in CLL, a pathway of genetic abnormality in CLL patients was shown in the figure¹. The gain or loss of the chromosome can initially drive the disease. An additional gene mutation can promote the disease's progress. At the late stage of the disease, specific gene mutation and aberration of chromosomes during illness would make the condition more aggressive and resistant to chemotherapy⁹.

The aim of this project is to take advantage of NGS to improve the method of detection the influence of *tp53* at the transcriptional level and figure out the relationship between the expression of *tp53* and other genetic abnormalities in CLL.

2) Results

a. The relationship of each biomarker

According to the results from clinic, we summarized the relevant genetic biomarkers and the survival data of the patients. All the mutational information of 114 patients have shown in the **Figure 36**. The red areas represented the poorer prognostic factors, including *tp53*mut, deletion of *tp53* at 17p13 and 17p13.1, unmutated *IGHV*, deletion of chromosome 11 and *ATM* at 11q22, deletion of chromosome 11 and *DLEU2* at 13q14, trisomy 12 and *POU6F1* at 12q13 and *MDM2* at 12q15, deletion of chromosome 6 and *SEC63* at 6q21. The white box represented the favor wildtype. The grey box means the results unavailable. We also studied the relationship of these genetic factors by correlation analysis in our 114 CLL patients.

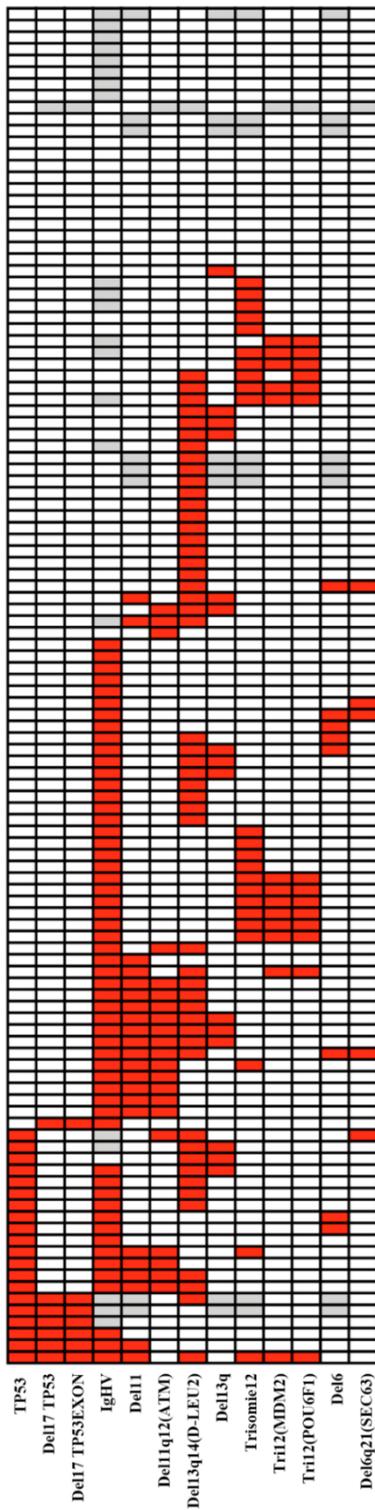


Figure 36: The characteristic of each important genetic molecule in CLL.

The summary of different genetic factors in the 114 CLL patients. Red presented unmutated *IGHV* and mutated other genes, which are all poor prognostic factors. White presents mutated *IGHV* and wildtype of other genes, and grey presents not evaluated items.

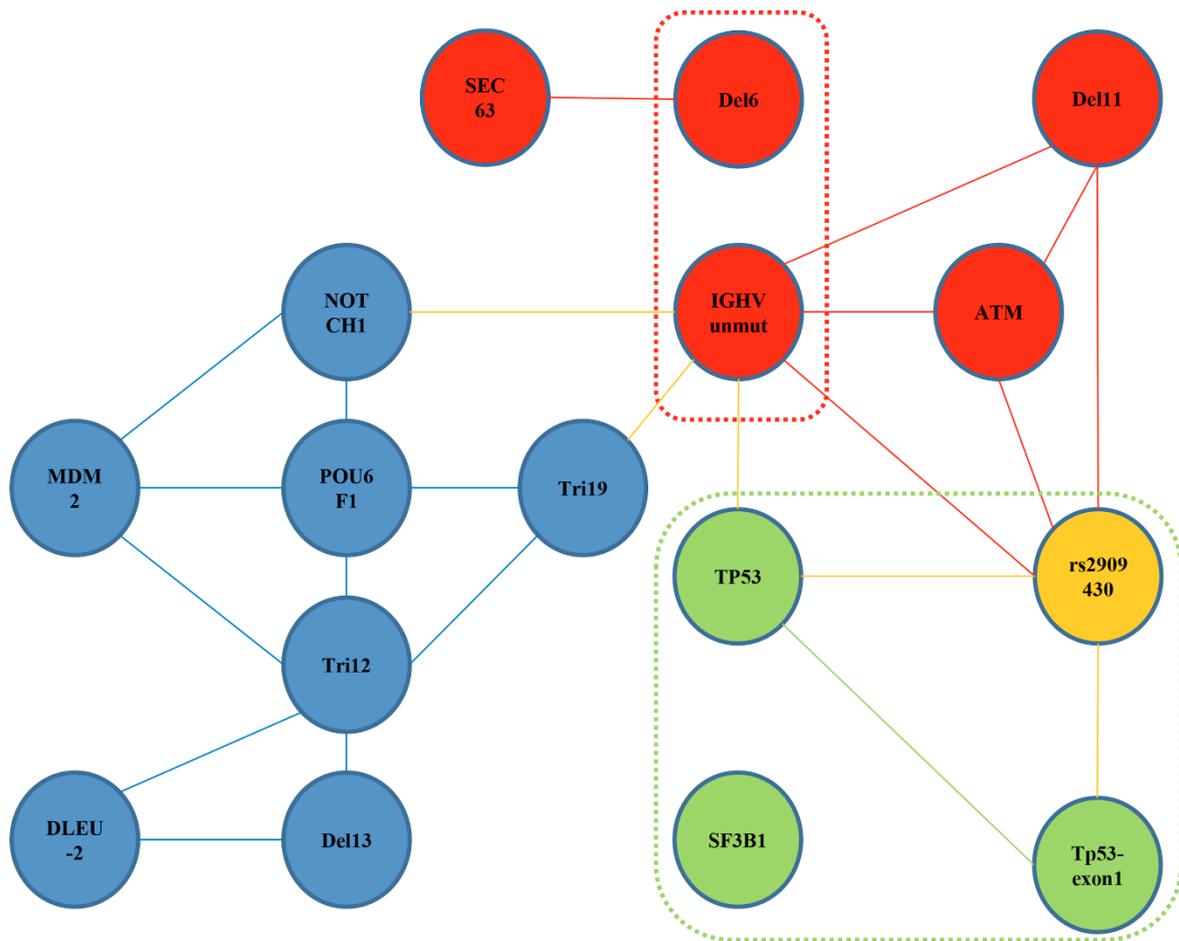


Figure 37: The relationship between expression of *tp53* and different prognostic markers.

In correlation of different genes, the line connected two genes showed positive relation ($p < 0.05$).

As we can see from the picture of relationship (**Figure 37**), it showed a large network of each factors in CLL, what remind us the interaction of each biomarker. For the next step to analysis the expression of *tp53* at the mRNA level, we would use this network as a guideline to classify different groups. For example, *tp53*mut showed a high correlation to the deletion of *tp53* at the del17p13 and SNP rs2909430, all these three will be considered as one factor in score system.

Table 23: The clinical information of 114 CLL patients.

Variable	<i>TP53</i> MUT (n=18)	<i>TP53</i> WT (n=96)	P-value
Gender			0.58
Male	12	63	
Female	6	33	
Age at diagnosis (years) Median (min; max)	72.2 (56-100)	71.3 (32-94)	0.75
Binet stage			0.87
A	11	60	
B	3	19	
C	4	17	
Laboratory parameters at diagnosis			
Platelets (150-450 Giga/L)	227.7±88.8 (n=17)	201.1±77.0 (n=88)	0.25
Haemoglobin (13.0-17.0 g/dL)	12.1±2.5(n=18)	13.2±2.1 (n=88)	0.04
White blood cells (4.0-11.0 Giga/L)	55.8±109.4 (n=18)	47.7±117.8 (n=89)	0.21
LDH (240-480 UI/l)	391.2±145/9 (n=17)	343.0±118.9 (n=86)	0.12
Protein (66-87 g/L)	68.2±6.3 (n=17)	67.5±4.3 (n=86)	0.59
Albumin (40-49 g/L)	41.1±5.0 (n=17)	41.5±3.8 (n=86)	0.97
Beta-2 microglobulin (0.8-2.4 mg/L)	3.4±2.4 (n=18)	2.9±1.6 (n=95)	0.16
Karyotype	n=16	n=90	
Deletion11	4(25.0%)	17(18.9%)	0.82
Trisomy12	1(6.3%)	21(23/3%)	0.22
Deletion13	3(18.8%)	11(12.2%)	0.76
Deletion6	2(12.5%)	6(6.7%)	0.76
Trisomy19	0(0.0%)	3(3.3%)	1.00
QMPSF	n=18	n=95	
<i>ATM</i> (del11q12)	4(22.2%)	17(17.9%)	0.92
<i>D-LEU2</i> (del13q14)	10(55.5%)	35(36.8%)	0.14
<i>MDM2</i> (trisomy12)	2(11.1%)	11(11.6%)	1.00

<i>TP53</i> (del17p13)	6(33.3%)	1(1.1%)	<0.001
<i>TP53</i> exon1 (del17p13)	6(33.3%)	1(1.1%)	<0.001
<i>POU6F1</i> (trisomy12)	1(5.6%)	12(12.6%)	0.73
<i>SEC63</i> (del6q12)	1(5.6%)	4(4.2%)	0.59
<i>IGHV</i> Unmut (98% cut-off)	13(92.9%, n=14)	44(55.0%, n=80)	0.006
NGS	n=18	n=96	
RS1042522	17(94.4%)	84(87.5%)	0.66
RS2909430	8(44.4%)	10(10.4%)	0.001
RS1625895	8(44.4%)	10(10.4%)	0.001
RS1800372	3(16.6%)	4(4.2%)	0.14
<i>SF3B1</i>	2(11.1%)	6(6.3%)	0.81
<i>NOTCH1</i>	2(11.1%)	18(18.8%)	0.66

According to the results of NGS, we have divided patients into *tp53* mutated and wildtype groups to confirm the characteristic of *tp53* mutation (**Table 23**). We found that unmutated *IGHV*, deletions of the short arm of chromosome 17 (del17p13) correlated with *tp53* and *tp53* exon1 were highly correlated with the mutated *tp53* ($p=0.005$, $p<0.0001$ and $p<0.0001$, respectively). Regarding the results of NGS, the most frequent single nucleotide polymorphisms (SNP) were rs1042522 (101, 88.6%), rs162895 (18, 15.8%) and rs2909430 (18, 15.8%). The last two SNP were systematically present in the same patients and showed similar mutational frequency and a close correlation with *tp53* mutation ($p=0.001$). Patients with mutated *tp53* were more likely to have a lower hemoglobin level (mean of 12.1 g/dL with mutated *tp53*) than a mean of 13.2 g/dL with wild type *tp53*, $p=0.04$).

b. The predictive value of each biomarker

Not only the relationship of these genetic factors, we also confirmed their clinical value in our series. Consistently, the mutational status of *IGHV* can consistently predict the overall survival and the time from diagnosis to first treatment in CLL patients. In contrast, the *tp53* gene mutation does not well reflect the characteristics of its prognostic biomarkers, which may be related to the small sample size or the lack of calculation of deletion of chromosome 17. Therefore, we can find that the detection of *tp53* mutations does not consistently predict patient survival. In contrast, the mutational status of other biomarkers such as *NOTCH1* and *SF3B1* can consistently predict the overall survival of patients but not the time to first treatment. Finally, we also identified deletion of chromosome 6, affecting poor prognosis in CLL.

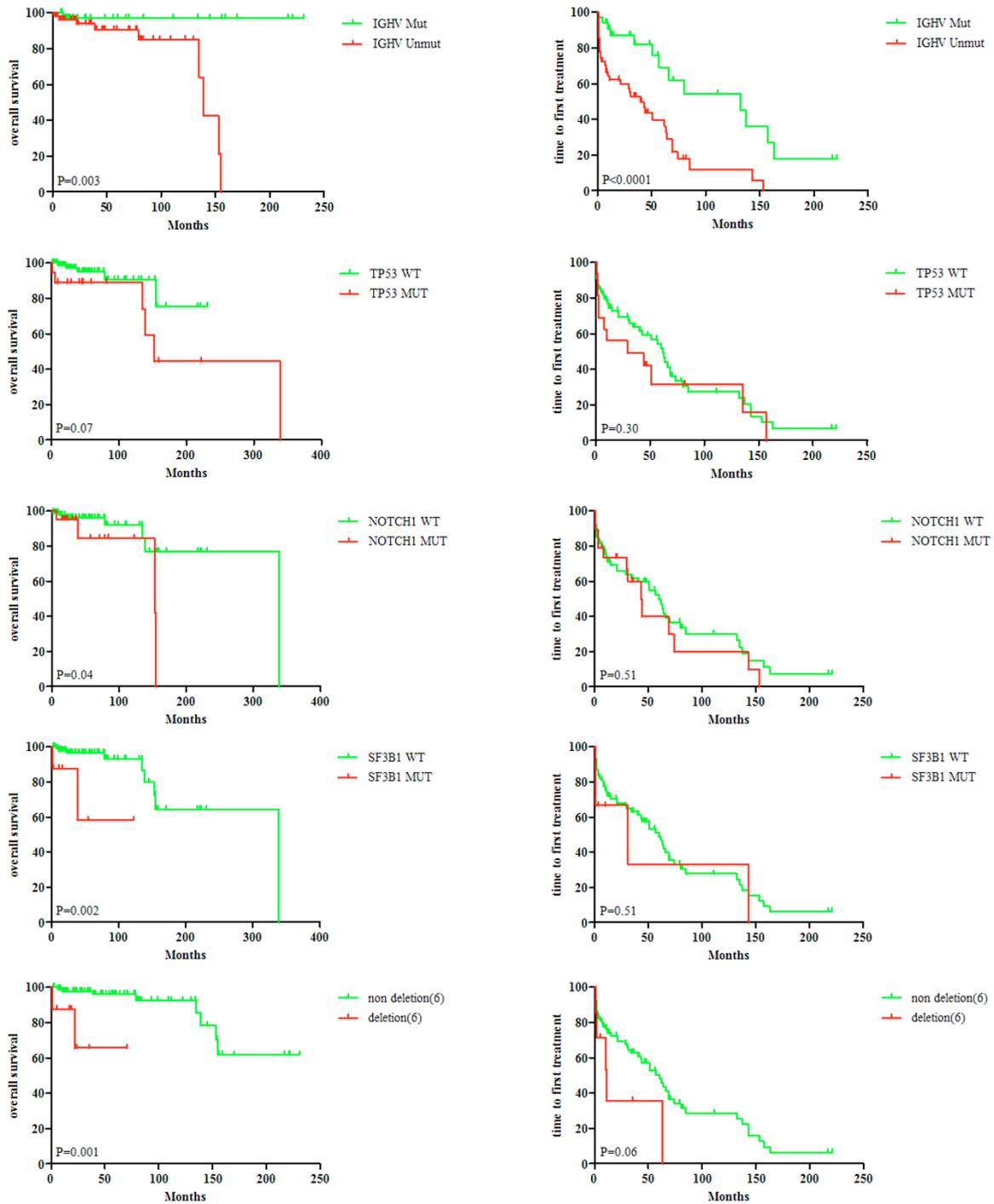


Figure 38: The survival analysis of different prognostic markers.

The curve of the OS of *IGHV* mutated (green) with unmutated (red) ($p=0.003$) and the curve of the TTFT of *IGHV* mutated (green) with unmutated (red) ($p<0.0001$).

The curve of the OS of *tp53* wildtype (green) with mutated (red) ($p=0.07$) and the curve of the TTFT of *tp53* wildtype (green) with mutated (red) ($p=0.30$).

The curve of the OS of *NOTCH1* wildtype (green) with mutated (red) ($p=0.04$) and the curve of the TTFT of *NOTCH1* wildtype (green) with mutated (red) ($p=0.51$).

The curve of the OS of *SF3B1* wildtype (green) with mutated (red) ($p=0.002$) and the curve of the TTFT of *SF3B1* wildtype (green) with mutated (red) ($p=0.51$).

The curve of the OS of chromosome 6 wide type (green) with deletion 6 (red) ($p=0.001$) and the curve of the TTFT of chromosome 6 wide type (green) with deletion 6 (red) ($p=0.06$).

In general, through our survival analysis of each biomarker in 114 CLL patients (**Figure 38**), we can see that, as expected, unmutated *IGHV*, *NOTCH1*mut, *SF3B1*mut, and deletion 6 predicted poor OS in CLL patients ($p=0.003$, 0.04, 0.002, 0.001, respectively). Unmutated *IGHV* also showed a shorter TTFT ($p<0.0001$). Multivariate Cox proportional hazards regression analysis indicated that unmutated *IGHV* (RR=11.5, $p=0.03$), del(6) (RR=10.0, $p=0.021$) and *NOTCH1* mutation (RR=9.1, $p=0.017$) were associated with poor survival and unmutated *IGHV* (RR=2.7, $p=0.002$) and Binet stage (RR=1.5, $p=0.021$) predicted for treatment to be initiated.

However, mutation of *tp53* can't be a stable predict factors in our CLL patients, which means that besides mutation of *tp53*, there are also other mechanisms which can regulate the activation of *tp53* family.

B. The function of *tp53* in CLL

1) Problem

As many researches have confirmed, *tp53* aberration is a crucial factor in CLL. The routine methods to diagnose *tp53* aberration are targeted NGS of exon 4 to 9 of *tp53* and FISH to detect the deletion of chromosome 17. However, the threshold of VAF in targeted NGS has been controversial because many researchers have found that a low tumor mutational load, VAF around 0.1-10%, still predicts an unfavorable prognosis in CLL patients. That may explain why we can't confirm the clinical value of *tp53* mutation in our series. So it is crucial to understand the role of *tp53* in CLL and improve the method to detect the expression change of *tp53* at mRNA level.

2) Results

a. The mutation of *tp53* in CLL

In our 114 CLL patients, there are 18 patients showed mutated *tp53* diagnosed by target NGS. We have described the mutation sites of *tp53* on the gene map. Interestingly, all the mutations were concentrated in the DNA binding region, which was related to the N-terminal of p53 isoforms. In order to analysis the expression of *tp53* at mRNA level, we introduced the RT-MLPA to detect each exon junction. On the one hand, it can show the specific regions of each isoforms. On the other hand, it can also reflect the total expression of p53 mRNA and the relationship among each exon junctions.

Figure 39: The structure of *tp53*.

Eleven exons make up six domains, and the different colors represent different domains: as we have shown in the figure, according to the amino acid residues, the position from 1 to 67 make up the transactivation domains, including TAD I (red) and TAD II (light purple)). From residues, 68-98 are the domains of the proline-rich region (PRD, yellow). The core domain of *tp53* is the DNA-binding domain (DBD, green), made up of residues 94–292. On the C-terminal, there are three functional domains, including the hinge domain (HD, dark purple), the oligomerization domain (OD, orange), and the carboxy-terminal regulatory domain (CTD, brown). Other variants of the C-terminal are exon9b (light blue) and 9g (dark blue). All the mutated sites were directed at the genetic picture, all gathered on the DNA-binding domain.

C. The influence of isoforms of *tp53* in CLL

1) Problem

As we mentioned, there is no large-scale study on the variability splicing of *tp53*, resulting in differential expression at the transcriptome level in CLL. The commonly used methods for the detection of *tp53* gene mutation are targeted sequencing and whole exon sequencing, in which the definition of the VAF value threshold is still debated. So the impact of the *tp53* gene on the disease solely from the mutation ratio can no longer explain the poor prognosis of patients, we need to develop new methods to reveal the important role of variability splicing of the *tp53* gene in CLL from the transcriptional level.

2) Results

a. The total information of the *tp53* mRNA expression in CLL

The structure of *tp53* is presented in **Figure 39**. Theoretically, the structure can generate 12 kinds of proteins because of alternative splicing. For the N-terminus, there are two promoters named P1 and P2. The mRNA transcribed by P1 will translate two types of proteins, the full-length p53 (TAp53) and the $\Delta 40$ p53 isoform, which started at codon 40. Promotor P2 can transcribe the isoforms $\Delta 133$ p53 and $\Delta 160$ p53, started at codon133 or 160, respectively. Three variants in the C-terminus are also present due to alternative splicing of the exon9, named α , β , and γ . Exon9 β and γ have stop codons, leading to translational termination. Isoform α also translates exon10 and exon11. There are six domains of *tp53* which are directed by different colours.

We detected the expression of each exon junction by RT-MLPA to find more information in the mRNA and to determine the percent of mutation on codon 72 to confirm the stability of this method.

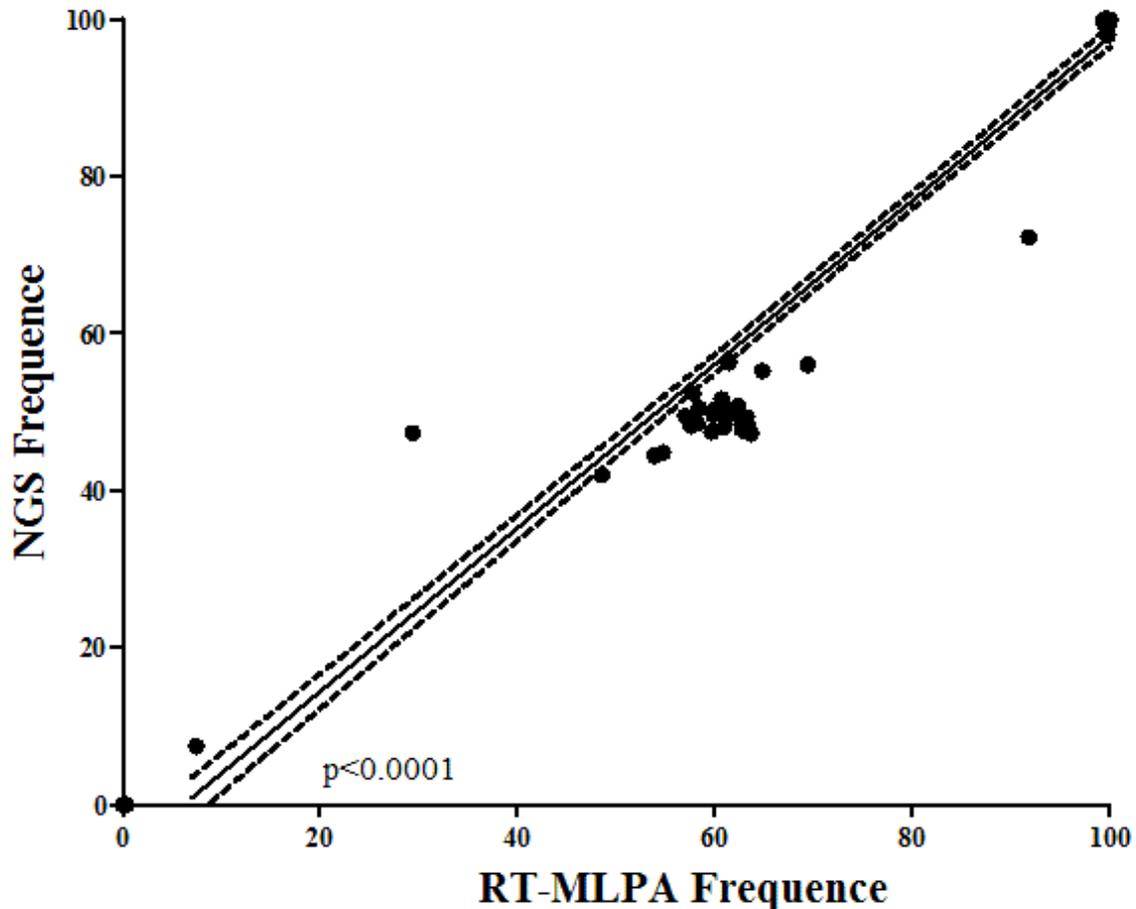


Figure 40: The stable of RT-MLPA used in CLL.

Comparing the mutational frequency of codon 72 on the exon 4 detected by NGS and RT-MLPA showed positive correlations ($p < 0.0001$).

More than 3000 reads of *tp53* in each patient among the 114 included were detected. Firstly, we evaluated the frequency of codon 72 mutation detected by NGS versus RT-MLPA, illustrated in figure 39, showing a positive correlation ($p < 0.0001$). It means that RT-MLPA can also catch the mutation frequency in the coding site and reflects the situation of mRNA expression stably (**Figure 40**).

To detect the accuracy and efficiency of this method, we also established a positive control cell line derived from human origin diffuse large cell B lymphoma (DLBCL), named SU-DHL-16, which showed a splicing site mutation between exon 9 to 10 (NM_000546: c.994-1G>A) with an allelic frequency of 100% (**Figure 41 A**). Point mutation on the acceptor consensus site at the end of the intron 9 causes a junction between exon 9 to 11 and skipping exon 10, detected by RT-MLPA (**Figure 41 D**). We also used RT-PCT and Sanger sequencing methods to verify mRNA expression in the SU-DHL-16 sample. Although we observed the exon junction of 9 to 11 and exon 10 skipping, it was not the main isoform. The sequencing result indicated another segment with the retention of a part of intron 9 between exon 9 to 10, using an alternative

splicing site acceptor in intron 9 (**Figure 41 C**). However, because of the absence of a specific probe, RT-MLPA cannot detect this kind of isoform.

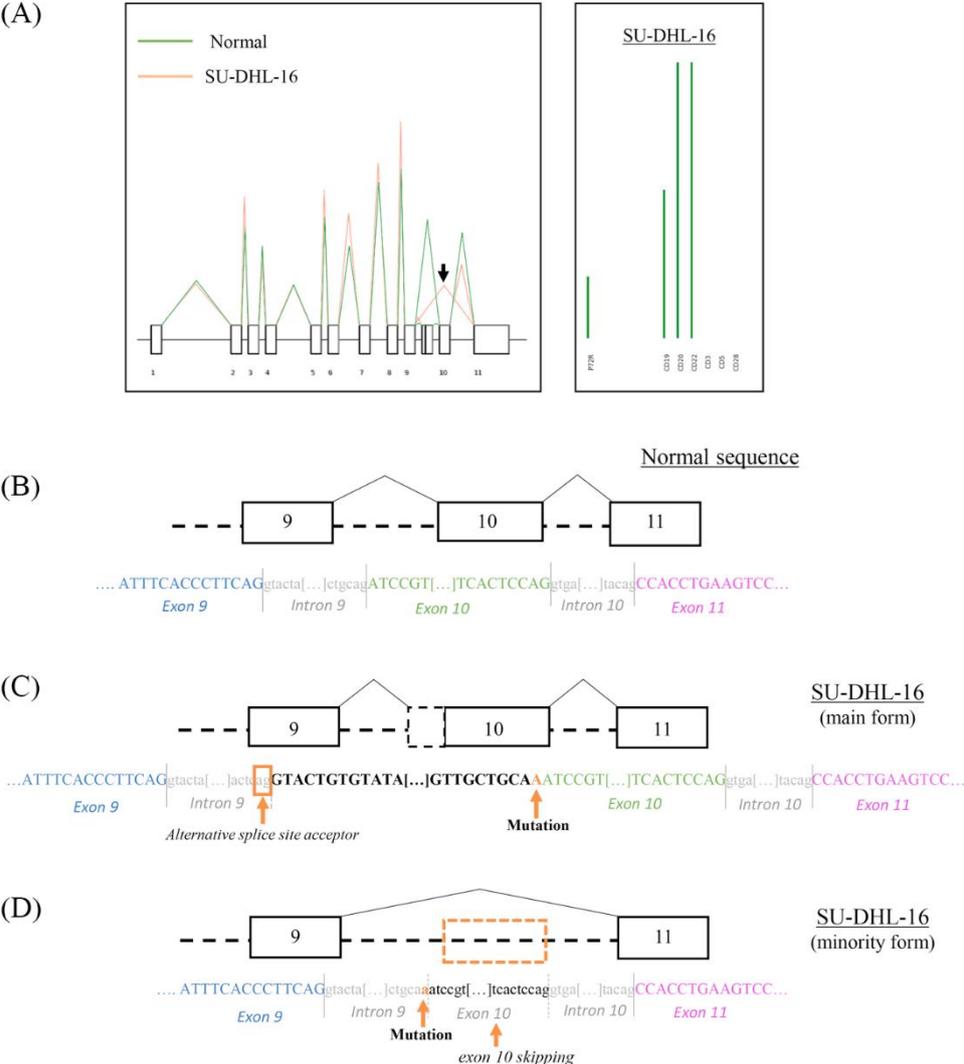


Figure 41: TP53 splicing isoforms of SU-DHL-16 cell line characterized by RT-MLPA and RT-PCR.

(A) In the left side, the green line showed the alternative splicing in normal, and orange line showed the howed an exon 10 skipping and the expression of a junction between exon 9 and 11. In the right side, by using the biomarkers of B-cell and T-cell, we can ensure the purify of the DLBCL cell lines. (B,C,D) *Tp53* splicing isoforms characterization by RT-PCR and Sanger sequencing. The capital letters correspond to the sequence transcript on mRNA, and the small caps correspond to the deduced DNA sequence. (A) The sequencing of the normal control cell has shown the expression of exons 9 to 10 to 11 in normal order. (B) The main isoform detected in the SU-DHL-16 cell line which expression part of intron 9 because of an alternative splicing site acceptor in intron 9. (C) The minor isoform found in the cell line RT-MLPA also detected, expressed junction of exon 9 to 11 and skipping exon 10.

The expression percent of all *tp53* exon junctions of 114 CLL patients was shown below (**Figure 42**). Using a multiple linear regression model that was developed by a stepwise method, we presents the linear association between the expression of exon9-10 (p53 α), exon9-9b (p53 β), exon9bg-10 (p53 β and γ), and a set of other expression exon junction, respectively (**Table 24**).

The expression of exon9-9b and exon9bg-10 had the most impact according to the value of standardized coefficients.

Table 24: Multiple linear regression model of the expression of exon junction.

Model	B	t	R ²	F	p	Durbin-Watson
Exon9-10 (p53α)	Constant	7.01	0.724	56.72	<0.0001	1.932
	Exon9bg-10	-3.26				
	Exon2-3	4.86				
	Exon7-8	2.81				
	Exon8-9	-2.76				
	Exon9-9b	-2.26				
Exon9-9b (p53β)	Constant	2.93	0.841	593.03	<0.0001	2.057
	Exon9bg-10	24.35				
Exon9bg-10 (p53β and γ)	Constant	5.45	0.881	271.23	<0.0001	2.171
	Exon9-9b	14.16				
	Exon6-7	-5.23				
	Exon9-10	-3.19				

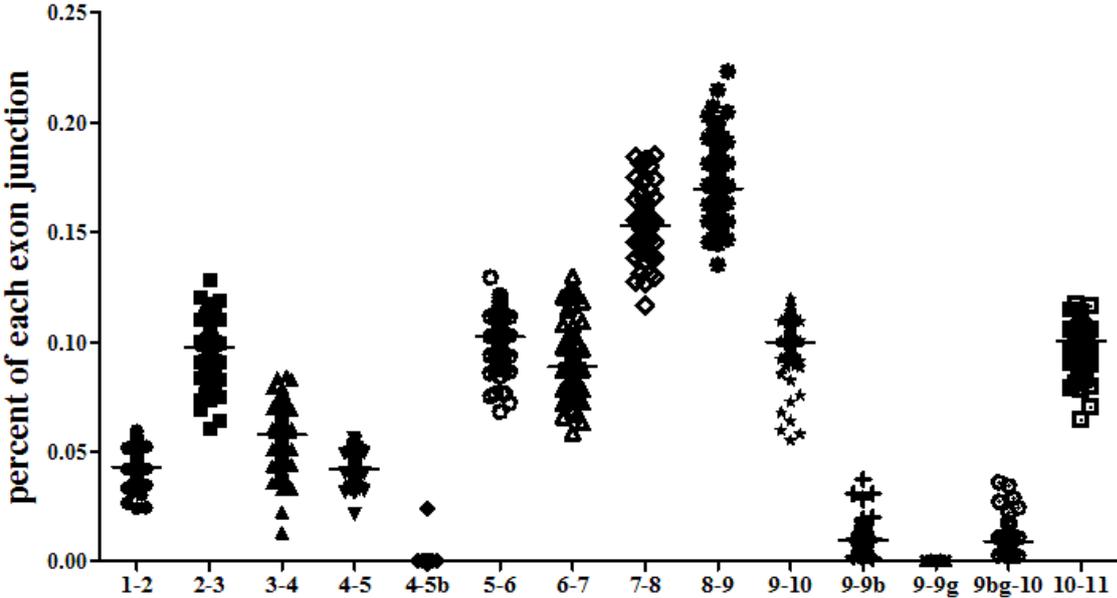


Figure 42: The expression of each exon junction in 114 CLL patients.

b. Special patient change

Moreover, RT-MLPA can also reflect the frequency of the splice site mutation. In our series, we found a patient that received 6 cycles of chemotherapy and showed a splice site mutation between exon 4 and 5 (c.376-2A>G) with a mutational frequency of 45% tested by NGS. Interestingly, when we tested this sample by RT-MLPA, this splice site mutation caused 47.4% of exon 4 LHS ligated with normal exon 5 RHS and 52.6% ligated with exon5 truncated at Δ 133 site (exon5b) (**Figure 43**), which is concordance with the data of NGS. After 2-year treatment of ibrutinib (a kind of Bruton's tyrosine kinase (BTK) inhibitors, which can irreversibly inhibit BTK), the expression of exon4-5b decreased to 5.9%, while the mutational frequency of splice site detected by NGS was of 1.4%. Therefore, besides the fact that RT-MLPA can reflect the mutational frequency, it can also notice the change of mRNA expression caused by splice site mutation and treatment strategy.

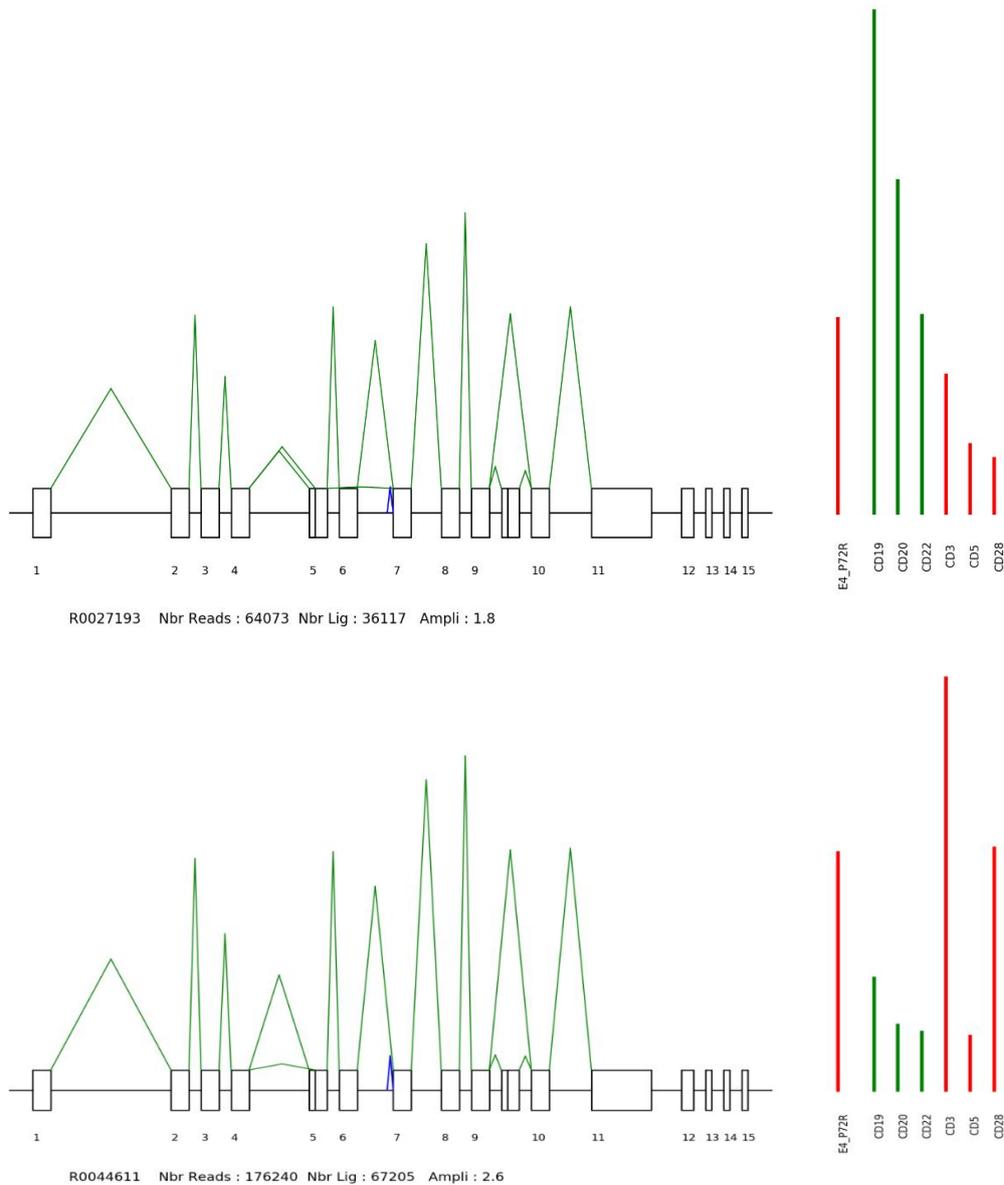


Figure 43: The results of RT-MLPA in a CLL patient with splicing site mutation.

The RT-MLPA results of a patient presented splice site mutation between exon 4 to 5 (c.376-2A>G). The top picture showed the expression of exon4-5 and exon4-5b when the patient had just received six cycles of FCR. The picture below presented a change in expression after the 2-year treatment of BTK inhibitor ibrutinib, which concord with the result of NGS. The picture on the right has showed the expression level of biomarkers of T-cell and B-cell, which made a control to compared the expression level of *TP53* and B-cell.

c. In the different expressions of exon junction groups
i. Exon 4-5 special for TA/ Δ 40p53

Although several studies investigate the clinical influence of *tp53*, the mRNA expression of *tp53* is controversial. Therefore, firstly, we analyzed the relationship between different genes and illustrated the significant positive correlations ($p < 0.05$) between them with lines. Secondly, we highlighted that the expression of some exon junctions was statistically different between various mutational statuses of genes.

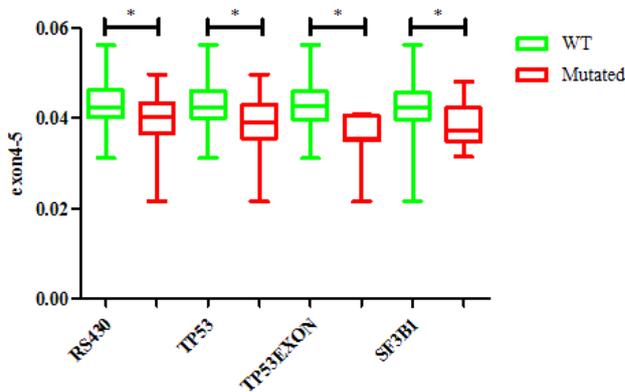


Figure 44: The markers expressed a lower level of exon 4-5.

For example, *tp53* mutated cases expressed a lower level of exon 4-5, the only part in FL and Δ 40p53 isoforms. The same situation was identified in del17p13 (*tp53*exon1), rs2909430 and *SF3B1*mut patients (Figure 44). Positives correlations were also detected between *tp53*mut, del17p13 (*tp53*exon1) and rs2909430 (Figure 37, represented in green). Consequently, we divided patients into three groups, as presented in table classification 1. The three groups not only expressed different levels of exon4-5 but also showed a difference in the OS, as illustrated in Figure 45.

Classification 1					
Biomarker	Score	Biomarker	Score	Risk group	Group B
<i>SF3B1</i>		<i>TP53</i> net		Group A	0
Wildtype	0	Wildtype	0	Group B	1
Mutated	1	<i>TP53</i> mut/del17p13/rs2909430	1	Group C	2

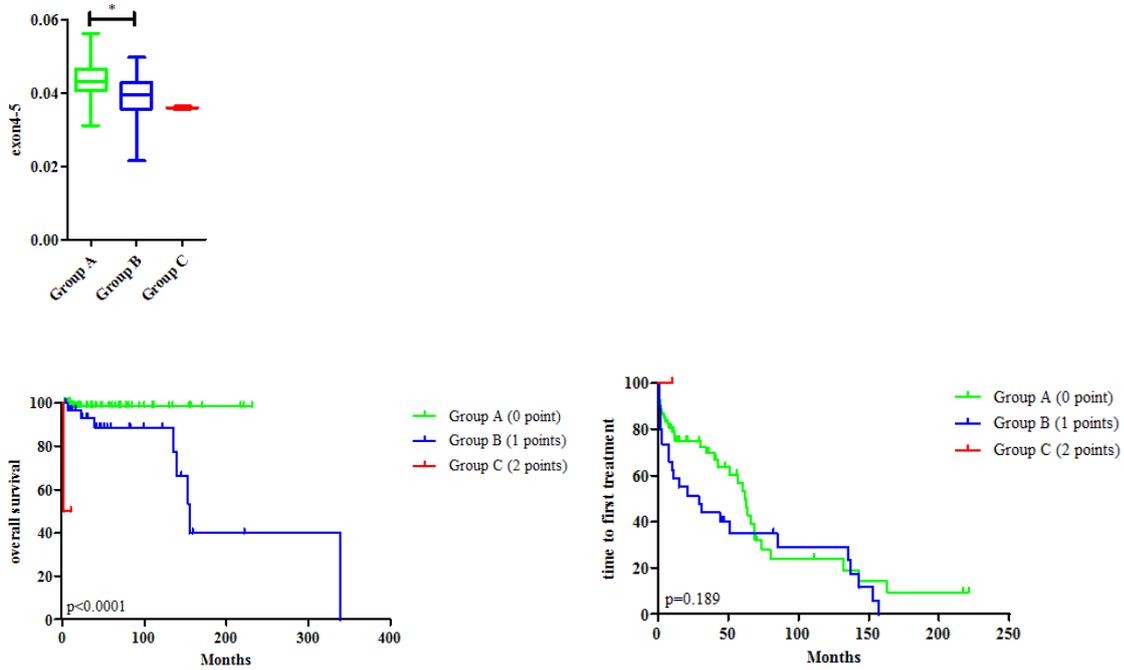


Figure 45: The survival analysis of different groups classified by the expression of exon4-5.

According to classification 1, the genetic groups showed significant differences in the expression of exon4-5 and OS ($p < 0.0001$).

ii. Exon9-9b and exon9bg-10 special for p53 β

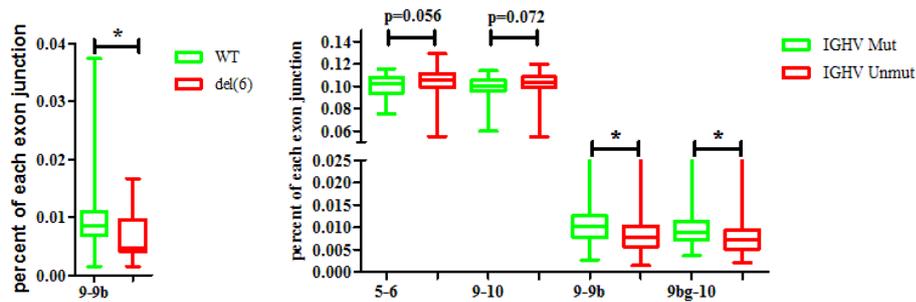


Figure 46: The markers expressed a lower level of exon9-9b and exon9bg-10.

Additionally, we also compared the expression of *tp53* in other genetic groups. Unmutated *IGHV* cases expressed a lower level of exon9-9b and exon9bg-10, which may represent the isoform of p53 β and p53 γ . A similar situation also appeared in the deletion 6 group (Figure 46). According to Figure 37 red box, we made the score system of classification 2. We divided patients into three groups with increasingly different levels of expression; the higher points they scored, the higher level of exon 9-10 and the lower level of exon 9-9b and 9bg-10 it expressed

(Figure 47). This trend was maintained in the OS and TTFT analysis ($p < 0.0001$ and 0.001 , respectively).

Classification 2					
Biomarker	Score	Biomarker	Score	Risk group	Group B
Deletion 6		<i>IGHV</i>		Group A	0
Wildtype	0	Mutated	0	Group B	1
Deletion	1	Unmutated	1	Group C	2

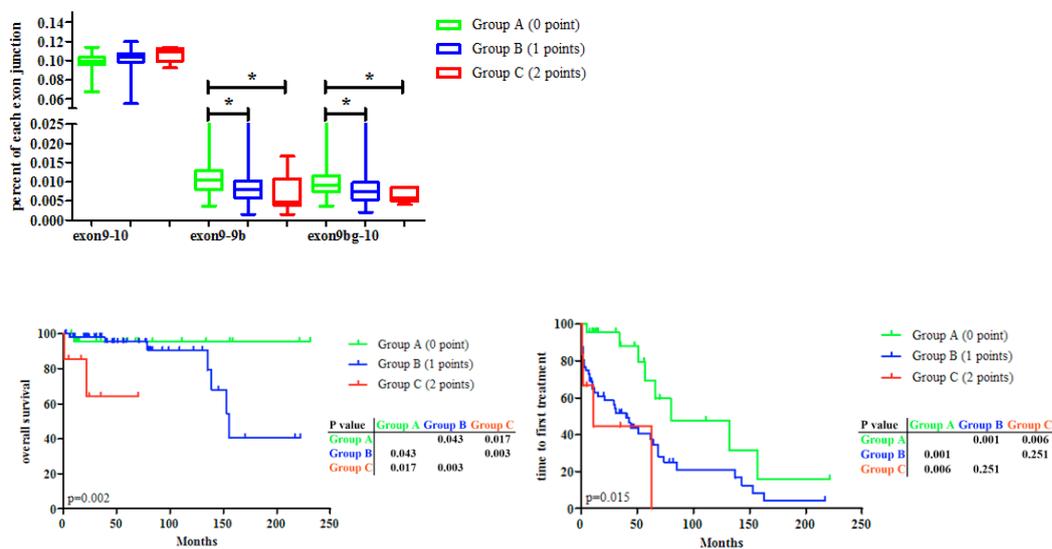


Figure 47: The survival analysis of different groups classified by the expression of exon 9-9b.

According to classification 2, the different groups showed different expressions of exon9-9b and exon9bg-10 and OS, TTFT ($p < 0.0001$ and 0.001 , respectively).

iii. All crucial biomarkers in CLL

As the results shown, patients divided into high-risk groups are more likely to expressed higher level of exon 4-5, which is special for isoform TA/ $\Delta 40p53$, and lower level of exon 9-9b, which is special for isoform $p53\beta$. Considering the influence of all the clinical biomarkers in our series, we build the classification 3, and divided patients into three groups. We found that higher risk groups expressed a lower level of exon9-9b and exon9bg-10 and a higher level of exon10-11. The same trend could also be observed in the analysis of OS and TTFT ($p = 0.0003$ and 0.004 , respectively) (Figure 48).

Classification 3					
Biomarker	Score	Biomarker	Score	Risk group	Score
Deletion 6		<i>IGHV</i>		Group A	0
Wildtype	0	Mutated	0	Group B	1-2
Deletion	1	Unmutated	1	Group C	3-4
<i>NOTCH1</i>		<i>Tp53</i> network			
Mutated	0	Wildtype	0		
Unmutated	1	<i>TP53</i> mut/del17p13/rs2909430	1		
<i>SF3B1</i>					
Mutated	0				
Unmutated	1				

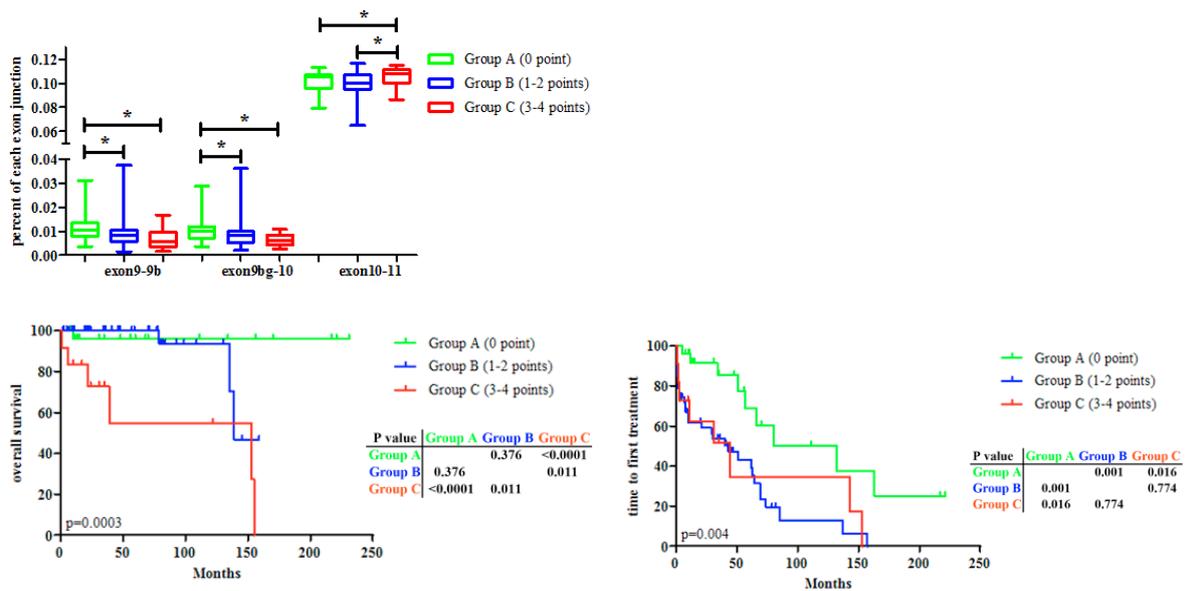


Figure 48: The survival analysis of different groups classified by all the important biomarkers in CLL.

According to classification 3 (including all genes that showed prognostic value), the different groups showed different expressions of exon9-9b, exon9bg-10, and exon10-11, and OS, TTFT (p=0.0003, 0.004, respectively). * Present the significant analysis (p<0.05)

3) Conclusion and discussion

With the discovery of *tp53* in CLL, we found that most mutations of *tp53* were detected in the DNA binding domain. This kind of mutation has been considered the leading cause of p53

inactivation¹⁷⁵. Although only 10% of patients showed a genetic mutation in CLL, *tp53* is still a significant predictor of treatment response and prognosis. Among the clinical data of our CLL patients, mutational status of *IGHV*, *tp53* at 17p13 and 17p13.1 genes detected by QMPSF, SNP rs29009430 and rs1625895 tested by NGS, and low level of haemoglobin were shown a close relationship with mutated *tp53*. As we know, the frequency of *tp53* mutation is consistently higher in chemo-refractory patients than in treatment-naïve CLL patients. It may decrease after the treatment of target inhibitors. Compared the results of one patient after receiving routine first-line treatment followed by BTK inhibitor, we found that RT-MLPA high-throughput sequence can also serve as a surveillance tool to monitor the change of the mutational frequency in *tp53*.

Traditionally, point mutations of the *tp53* gene have been identified as a fundamental cause of the occurrence of cancer¹⁷⁶. However, various truncated forms of p53 can also be responsible for its inactivation. Especially the N-terminally truncated deletion part of the DNA binding domain. Some researchers have suggested that p53 isoforms involve in carcinogenesis¹⁴⁸. It has been reported in many different kinds of tumours that, compared to normal control samples, tumorigenic tissue expressed various levels of the p53 isoforms¹⁷⁷.

Moreover, several researchers have reported that even though some cancers showed a low mutation rate of *tp53*, their expression of p53 isoforms has significantly changed. For example, breast cancer with a 25% mutation rate of *tp53* expressed different levels of p53 isoforms. They decreased more than half of the expression of p53 β , and p53 γ isoforms increased half of the expression of the $\Delta 133p53$ ¹⁷⁸. A similar situation was also observed in acute myeloid leukemia, which only shared a 10% mutation rate and showed a large different level of the p53 β and p53 γ isoforms¹⁷⁹. These studies have expanded the perspective to discover the mechanisms of the p53 isoforms and improve cancer treatment.

In the report, RT-MLPA was designed to detect transcript expression levels of mRNA. Because this is an RNA-sequencing technology, the procedure started with combining probes with the RNA and ligating the cDNA. PCR followed it amplified with the universal PCR primer. Because each probe of *tp53* needs only about 30 bps, the length of the cDNA template will not be too big. The other advantage of our method was the combination with second-generation sequencing to detect target RNA. The high-throughput platform can analyse the specific sequences in left and right hybridizing sequencing, allowing as many probes in one assay as possible.

Furthermore, the Illumina-compatible sequencing primers allowed a more flexible pooling strategy when the RT-MLPA sequence needs extremely little data. The disjunction between cytogenetic and molecular testing leads to the challenge of estimating these factors into a prognosis model to guide the clinical decision. We combined the results of the RT-MLPA high-throughput sequence with other molecular factors to find more information on mRNA and the relationship among multiple genes with *tp53* in CLL. It is the first time to use the RT-MLPA high-throughput sequence in CLL patients to discover the transcriptional level of mRNA.

The different expressions of exon junction between various prognostic factors are always related to the exon4-5, exon5-6, exon9-10, exon9-9b, and exon9bg-10, which may represent the distribution of TAp53 and $\Delta 40p53$, $\Delta 133p53$ and $\Delta 160p53$, p53 α , p53 β , and p53 γ isoforms, respectively. And when we classify patients according to the factors expressing a similar level of exon junction, different groups can predict the treatment and prognosis of CLL. The low expression level of exon4-5, exon9-9b, and exon9bg-10 and high expression level of exon5-6, exon9-10, and exon10-11 indict shorter overall survival and a highly necessary to be treated.

In summary, RT-MLPA high-throughput technology is a stable and economical technology to analyse the expression levels of RNA in CLL patients, which can be used in both research and clinical. It's the first time we have studied the expression of *tp53* by testing the expression level of all exon junctions. Our finding is in concordance with the research on other cancers. The short N-terminal truncated isoforms, including $\Delta 133$ and $\Delta 160$ p53, are more likely to aggressive the disease, and the expression of p53 β and p53 γ isoforms may release the disease. There are also some limits. Firstly, there are not many studies of mRNA expression in CLL. The standard guideline is needed in the future. And our study, which included 114 patients, was also too small. More multi-centre cooperation is required.

CONCLUSION AND GENERAL DISCUSSION

I. Conclusion

In this study, we aimed to extend the use of NGS technology in CLL and explore valuable biomarkers with the help of this high-throughput sequencing tool, which can be used to evaluate treatment options and predict prognosis. Among them, *IGHV* is the crucial prognostic indicator in CLL. The conventional detection tool Bio-med2 has shown some drawbacks in operation. It is necessary for detection throughput and subsequent automated analysis and combination with other experiments, so we used the 5'RACE technique with the help of second-generation sequencing technology to detect immunoglobulin heavy chain mutations. At the same time, information about the immunoglobulin light chain can also be obtained. Also, since the 5'RACE technique starts from the constant region sequence, we can get the complete sequence information of the CDR3 region, which is crucial in evaluating the function of immunoglobulins. We also found that most CLL patients' heavy and light chains express the same mutation status. This finding gives a solution for the borderline cases group previously defined by the international organization ERIC. When the percentage of immunoglobulin heavy chain mutations in patients is between 97-99%, we cannot directly determine the prognosis of patients. We can use the information obtained from the light chain to guide the evaluation of the patient's prognosis by the light chain mutation. A particular group of patients expressing the IGLV3-21 sequence with a point mutation at amino acid sequence 110, resulting in a change from G to R, is called IGLV3-21^{R110}. When analyzed with the clinical data, we found that although most of the *IGHV* in this group were mutated, the patients still showed a similar immunoglobulin heavy chain non-mutation status. The OS and TTFT were significantly shorter, even though most of the patients in this group exhibited a mutated immunoglobulin heavy chain status. The percentage of *IGHV* mutations in this group of patients was mainly distributed in borderline cases, thus further emphasizing the role of immunoglobulin light chains in the diagnosis and prognosis of CLL.

II. General discussion

A. NGS in test *IGHV* in CLL

1) The importance of *IGHV* in CLL

The role of the mutational status of *IGHV* in CLL has been strengthened over the past 30 years^{10,97}. It is a stable predictor of patient therapeutic outcome and survival prognosis independent of other prognostic factors, making *IGHV* mutation status a critical consideration in most clinical staging at this phase.

The mutational status of the *IGHV* is defined by the percentage of similarity between the detected heavy chain sequence in the patient and the identified germline gene. The threshold value for mutation status is 98%¹⁸⁰. When the identity of the *IGHV* is greater than or equal to 98%, the patient is defined as mutated, called M-CLL. On the contrary, when the patient's identity is less than 98%, it is defined as non-mutation, called UM-CLL. Some scholars have studied the use of 97%¹⁰² or even 95% to describe the immunoglobulin heavy chain mutation status.

However, no mathematical threshold can define a biological phenomenon because even a point mutation at a relatively important site can also lead to a variation in the disease process. Therefore, as the study of immunoglobulin heavy chains continues to progress and many clinical problems cannot be explained simply by the 98% threshold, scholars have raised another issue that needs to be addressed: the consideration of borderline cases.

Borderline cases¹¹⁹, whose identity to the germline genes is between 97-99%, can not judge the treatment outcome and the survival prognosis simply by the numerical threshold, suggesting that there may be other biological molecules at work. Therefore, the International Chronic Lymphocytic Leukemia Organization appeals to clinicians to exercise great caution when using the 98% threshold to evaluate patients with immunoglobulin heavy chain mutation rates from 1-3%.

With continued research on *IGHV*, researchers have identified that 30% of individuals with CLL have highly similar CDR3 sequences in their immunoglobulin heavy chain sequences. They always deviate from the same light chains that make up the immunoglobulin, a group of cases known as subsets¹⁶³. Instead, they belong exclusively to a different group exhibiting specific clinical features and prognoses.

In developing the body's immune system, high-frequency mutations in somatic cells increase the diversity of immunoglobulins and thus respond to different in vitro invasions. However, nearly 30% of CLL patients express identical or similar heavy chain CDR3 sequences while matching the corresponding light chain with bias. Thus, the role of immunoglobulins in CLL goes well beyond the mutant state. The study of HCDR3 sequences and the critical value of immunoglobulin light chains still needs to be continuously investigated and discovered.

2) The value of IGLV in CLL

Late studies of immunoglobulin light chains in CLL¹²³ reported the immunoglobulin light chain repertoire in 276 and 207 patients. By comparing with normal groups and associated tumor cells, the authors defined λ and κ light chain, commonly used functional genes. When identifying the light chain homologous complementary determining region 3, specific light chains bind to fixed heavy chains, and a subset consisting of a given *IGLV-J* with homologous *IGHV-D-J* provides a theoretical basis for the role of immunoglobulins in the development of CLL.

Subsequent large-scale studies have also revealed non-random pairing between immunoglobulin heavy chains and specific light chains in CLL. Most notably, immunoglobulin heavy chain *IGLV3-21*, whose third complementary determining region expresses two types of amino acid sequences, motif-1 (DANGMDV) and motif-2 (DPSFTSSSWTLFDY). In contrast, HCDR3 with motif-1 usually pairs with *IGLV3-21*, and HCDR3 with motif-2 pairs with *IGKV3-20*. In contrast, in 2021, some researchers found that *IGLV3-21*-expressing CLL cells may acquire a point mutation at position R110 that triggers autonomous BCR signaling, which defines a subgroup of CLL with specific biological characteristics and prognosis and is independent of *IGHV* mutation status and epigenetic subtype.

The lack of research on immunoglobulin light chains is mainly because the results of the current assay do not give the sequence information of heavy and light chains at the same time very conveniently. The Bio-med 2 method¹²⁰, which includes PCR, gel electrophoresis, and Sanger sequencing after PCR amplification, is routinely used in clinical practice and requires a lot of labor input. At the same time, the later analysis of sequencing results is also manual. The study of sequencing results is also manual, compared to the gene sequence results on the website. Another drawback is that, with the increasing research on CLL, more and more critical biomarkers play an important role in the development and prognosis of the disease, and BIO-MED2, as a single-item study, does not allow the simultaneous detection of multiple items on the same research platform. Therefore, finding a more efficient and convenient assay that can be used in combination with other items is important.

At the present stage, thanks to the continuous development of second-generation sequencing technology, more and more technologies are emerging for the detection of immunoglobulins. And we use the second-generation sequencing platform in our laboratory to detect immunoglobulin sequences in CLL patients using 5'RACE technology, and its accuracy is up to 94% compared with the traditional gold standard. We also analyzed the cases with inconsistent results. We found that 5'RACE technology can more sensitively handle poor RNA quality samples and detect multiple and non-functional rearrangements more accurately than BIO-MED2, the second-generation sequencing technology.

The advantage of 5'RACE is that its preparation work is relatively simple, reducing manual operation steps. At the same time, using high-throughput sequencing means more information can be obtained regarding subcloning and molecular internal structure. The subsequent sequencing analysis work is all done automatically by the corresponding bioinformatics software. The final results are output as a combination of data and graphical means, which facilitates statistical analysis and visualization of data, and is simple and easy to understand. It saves labor costs and increases the probability of human error.

Another advantage of 5'RACE is that the sequence information of both immunoglobulin heavy chain and light chain can be obtained for a single sample. In most of the cases we analyzed, the mutational status of the light chain was always the same as that of its corresponding heavy chain. It reminds us that for the CLL borderline cases, where the percentage of *IGHV* mutations is 1-3%, the therapeutic and prognostic analyses are inconsistent with their heavy chain mutation status. And we found that in most cases, the mutation status of the heavy chain is consistent with that of the light chain. Clinicians can also use the light chain information obtained by 5'RACE to complement the heavy chain mutation status to guide the selection of treatment options. We also identified a group of specific mutations in the light chain.

We also confirm a specific subgroup of CLL cases expressing IGLV3-21 with a single point mutation at amino acid sequence 110 of the light chain, which we named IGLV3-21^{R110}. The mutational rate of their corresponding *IGHV* was between 97-99%. When they were divided into three groups, namely *IGHV*-M non-IGLV3-21^{R110} CLL, *IGHV*-U non-IGLV3-21^{R110} CLL, and IGLV3-21^{R110} CLL groups, we found that this particular group consistently exhibited

similar clinical features as the unmutated patients, i.e., shorter survival and the need to receive treatment as soon as possible, even though 75% of the IGLV3-21^{R110} patients expressed mutant immunoglobulin heavy chains.

Therefore, in our study, we not only highlighted the impact of efficient methods on *IGHV* but also demonstrated that immunoglobulin light chains play an important role in CLL. Therefore, efficient sequencing methods exploring immunoglobulin heavy and light chains will provide more information and data to support our study in future research and clinical treatment.

B. Using a more straightforward method to test the expression of *tp53* mRNA in CLL

Tp53 tumor suppressor gene is frequently mutated in cancer, and its mutation usually leads to enhanced chromosomal instability, including amplification of oncogenes and profound deletion of oncogenes¹⁷⁵. *Tp53* mutations can predict treatment outcomes and survival in many cancers. Numerous studies have shown that mutations in *tp53* are involved in various signaling pathways, like cell cycle¹²⁹, apoptosis, and tumor microenvironment regulation.

The primary outcome of *tp53* mutations is to destroy the function of p53, depriving the cell of its tumor suppressive ability. Most *tp53* undergoes missense mutations, i.e., a single base change within the DNA binding domain of the protein. Mutated *tp53* can provide tumor cells with the ability to cope with challenging conditions during tumor development, including DNA damage, nutritional deficiencies, and anti-tumor immune responses.

Specific *tp53* mutations have been reported to disrupt key cellular pathways while promoting tumor cell proliferation and invasion and increasing their resistance to chemotherapy. Although *tp53* mutant phenotypes contributing to tumor growth have been described, the mechanisms determining *tp53* promotion of cancer maintenance and progression remain incompletely understood.

In recent years, the study of p53 isoforms has gained much attention^{141,177} because of the deletion of some of its N-terminal and C-terminal amino acid sequences, its ability to produce 12 different types of proteins through variable splicing, and the absence of its N-terminal deletion resulting in the lack of the DNA-binding region of some isoforms, making its function different from that of normal wild-type p53. Therefore, at the transcriptome level, the level of mRNA obtained by alternative splicing will be able to indirectly respond to the alteration of the molecules involved in the transcription of the *tp53* gene to p53.

We analyzed the expression levels between different exon-binding regions in CLL patients with the help of RT-MLPA technology to detect whether variable splicing between different exons occurs in patients and its impact on the disease. By targeted sequencing, we used probes to detect the status of mutations in the *tp53* gene. We combined data from other common biomarkers in CLL to analyze the p53 mRNA expression levels between different gene mutations. Our results showed that both mutations in the *tp53* gene itself and deletions in the *tp53* gene-related part of chromosome 17 and mutations in other genes, such as *NOTCH1*,

SF3B1, Del6, and *IGHV* mutation states, the differences in p53 expression levels were always present in exons 4-5, 5-6, 9-9b, 9b-10 and 10-11. And these exons are the missing parts in the middle of the N-terminal and C-terminal isoforms. A molecular prognostic scoring system can be established by analyzing the relationship between biomarkers. We found that the expression levels of different *tp53* exon-junction appear to be hierarchically different between molecule groups. Therefore, we can assume that differences in the expression of p53 mRNA levels also affect the disease progression and prognosis of CLL.

Meanwhile, we found that one of the patients, who developed *tp53* shear site mutation, resulted in mRNA levels expressing the 4-5b binding site, while the normal exon 4-5 expression level was reduced by half. After the patient was treated with BTK inhibitors for one year, *tp53* splice site mutations decreased to 5%. And his RT-MLPA assay results likewise showed that exon 4-5 expression levels largely returned to normal, while exon 4-5b levels were significantly reduced. Therefore, using RT-MLPA to detect inter-exon expression levels can reliably reflect the changes of relevant biological markers in patients after receiving treatment. It can help clinicians evaluate the treatment effect on patients.

Compared with other real-time quantitative PCR methods, we have studied the relationship between each exon-junction by high-throughput sequencing. It can more comprehensively reflect the expression of the *tp53* gene at the transcriptional level mRNA, not only to detect abnormal exon variable splicing but also to analyze its mutational status on the *tp53* transcriptional level in combination with other biological markers. The relationship between mRNA expression and *tp53* transcript levels can be established more accurately with the prognostic assessment system of patients.

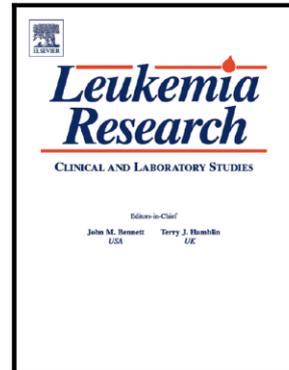
However, our study was mainly focused on the mRNA level and was not validated at the subsequent protein level. At the same time, we did not do a more in-depth analysis of the effects of *tp53* expression differences due to variable splicing on the signaling pathways involved, so more research is needed in the future to investigate the mechanisms behind this and the effects of other signaling pathways downstream.

ANNEXE

Journal Pre-proof

5' Rapid amplification of cDNA ends (5'RACE):
A simpler method to analyze immunoglobulin
genes and discover the value of the light chain in
chronic lymphocytic leukemia

Xuan Lan, Philippe Ruminy, Elodie Bohers,
Vinciane Marchand, Mathieu Viennot, Pierre-Julien
Viailly, Pascaline Etancelin, Hervé Tilly, Sorina
Mihailescu, Florian Bouclet, Stéphane Leprêtre,
Fabrice Jardin



PII: S0145-2126(22)00178-3

DOI: <https://doi.org/10.1016/j.leukres.2022.106952>

Reference: LR106952

To appear in: *Leukemia Research*

Received date: 1 August 2022

Revised date: 4 September 2022

Accepted date: 15 September 2022

Please cite this article as: Xuan Lan, Philippe Ruminy, Elodie Bohers, Vinciane Marchand, Mathieu Viennot, Pierre-Julien Viailly, Pascaline Etancelin, Hervé Tilly, Sorina Mihailescu, Florian Bouclet, Stéphane Leprêtre and Fabrice Jardin, 5' Rapid amplification of cDNA ends (5'RACE): A simpler method to analyze immunoglobulin genes and discover the value of the light chain in chronic lymphocytic leukemia, *Leukemia Research*, (2022)
doi:<https://doi.org/10.1016/j.leukres.2022.106952>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier.

5' Rapid amplification of cDNA ends (5'RACE): A simpler method to analyze immunoglobulin genes and discover the value of the light chain in chronic lymphocytic leukemia

Xuan Lan^{1,2}, Philippe Ruminy¹, Elodie Bohers¹, Vinciane Marchand¹, Mathieu Viennot¹, Pierre-Julien Viailly¹, Pascaline Etancelin², Hervé Tilly^{1,3}, Sorina Mihailescu⁴, Florian Bouclet³, Stéphane Leprêtre³ and Fabrice Jardin^{1,3}

1 INSERM U1245, Team Genetic and biomarkers in lymphomas and solid tumors, Centre Henri Becquerel, University of Normandie; Rouen, France

2. Department of molecular biology, Centre Henri Becquerel, Rouen, France

3. Department of clinical Hematology, Centre Henri Becquerel, Rouen, France

4. Clinical Research Unit, Centre Henri Becquerel, Rouen France

Corresponding author: Fabrice Jardin fabrice.jardin@chb.unicancer.fr

Permanent address: Centre Henri Becquerel Rue d'Amiens 76038 Rouen

Abstract

The mutational status of the variable region of the immunoglobulin heavy chain gene (*IGHV*) is a very important biomarker for chronic lymphocytic leukemia (CLL) patients. However, the routine detection of *IGHV* mutational status is time-consuming and costly. Therefore, we performed 5' Rapid amplification of cDNA ends (5' RACE) in 81 CLL patients who previously underwent detection using Biomed-2. The agreement rate of these two methods was 93.8%. Regarding the discordant cases, 5' RACE was more sensitive to identify unproductive and multiple rearrangements. Furthermore, 5' RACE can also be used to simultaneously sequence light chains. In most CLL cases, the mutational statuses of heavy and light chains are concordant, except in IGLV3-21. Most IGLV3-21 (24/25) rearrangement shared a similar LCDR3 (QVWDSSSDHPWV) and harbored a single point mutation, namely, IGLV3-21^{R110}. Compared to mutated-CLL non IGLV3-21^{R110}, IGLV3-21^{R110}-CLL exhibited a shorter overall survival (OS) and time to first treatment (TTFT) ($p=0.05$, $p<0.0001$, respectively) even though 75% (18/24) of these patients expressed mutated heavy chains. Altogether, IGLV3-21^{R110} defines a CLL subgroup with specific biological features and an unfavorable prognosis independent of the *IGHV* mutational status and emphasizes the important value of the light chain. This study is the first to use 5' RACE to detect the mutational status of IGH in CLL. Here, 5' RACE was a reliable and effective method to test the mutational status of heavy and light chains. In addition, 5' RACE can be combined with other assays in the NGS workflow to obtain more detailed insight into subclonal architecture and intraclonal diversity.

Keywords: Chronic lymphocytic leukemia (CLL), 5'Rapid amplification of cDNA ends (5'RACE), immunoglobulin heavy chain (IGH), immunoglobulin light chain (IGL), IGLV3-21^{R110}

Introduction

B-cell chronic lymphocytic leukemia (CLL), the most common leukemia in the western world^{1,2}, is a heterogeneous disorder with a remarkably variable natural history and prognosis with leukemic B cells ranging from functionally allergic to highly proliferating. Some CLL patients survive for years without any treatment, whereas others pass away rapidly due to progressive disease. Even though different biologic markers are available to help clinicians predict prognosis and make the best treatment decision, the biology and pathogenesis of CLL remain poorly understood.

At present, one of the best prognostic and therapeutic markers in CLL identified more than 20 years ago^{3,4} is the mutational status of *IGHV*⁵. Survival analysis has shown that patients with mutated BCR (M), where the *IGHV* region has undergone somatic hypermutation (SHM) and shows less than 98% identity when compared to the germline sequence, have a significantly longer survival than unmutated cases (UM)^{3,4,6,7}. Furthermore, multivariate analyses have confirmed that this mutational status is a strong independent poor prognostic marker in this pathology^{8,9}.

In routine, *IGHV* mutational status is commonly evaluated by PCR-based immunoglobulin clonality testing addressed through the alignment of the sequence detected in CLL cells with dedicated immunoglobulin gene databases using a 98% homology threshold¹⁰. However, as the presence of a single mutation (representing the deviation from germline) can have a major impact on the results, this threshold has been challenged, and the observation of so-called 'borderline' CLL cases with 97–98.9% *IGHV* identity has raised the question of whether they should be considered M or UM CLL^{11–13}. To address this question, the European Research Initiative on CLL (ERIC) has recently published guidelines emphasizing that caution should be used when employing any threshold to evaluate biological results¹⁴. The ERIC initiative has also identified three additional categories of problematic cases¹⁵: (1) an unproductive *IGH VDJ* rearrangement (which does not code a functional BCR) is exclusively detected; (2) two productive rearrangements with discordant mutational status are identified; and (3) the CDR3 region, which codes the whole D segment and part of the V and J segments, lacks critical residues. The guidelines also emphasize the importance of the analysis of the entire variable region of the *IGH* gene to obtain reliable results.

In addition to the mutational status of the heavy chain, other features of the BCR can also participate in the classification of CLL patients into different prognostic subsets^{14,16–19}. The light chain repertoire is, for example, biased, and some *IGKV/IGLV* genes are used more often in CLL than in normal B cells^{20,21}. Other studies have also demonstrated that, independent of *IGHV* mutational status, the use of the *IGLV3-21* light chain is associated with a poor prognosis^{22–24}. However, this characteristic is rarely evaluated during routine diagnosis due to the difficulty of analyzing the rearrangements of the BCR light chains at the molecular level.

Today, amplification of the *IGH VDJ* region by polymerase chain reaction (PCR) starting from genomic DNA extracted from a blood sample followed by Sanger sequencing is regarded as the standard method for the evaluation of the *IGHV* mutational status¹⁰. This analysis requiring six independent PCR amplifications is time-consuming and needs a high level of expertise, particularly for the interpretation of problematic cases. To address these difficulties, other methodologies have recently been proposed that are associated with the democratization of next-generation sequencing (NGS) platforms in routine diagnosis laboratories^{25,26}.

In this study, we evaluated the advantages of one of NGS methodologies based on a simplified multiplex 5' RACE BCR analysis method adapted to a routine diagnosis workflow, which allows the simultaneous characterization of the immunoglobulin heavy and light chains repertoire in CLL.

Materials and Methods

Blood sample preparation

A series of 81 patients who were routinely diagnosed B-cell CLL with *IGHV* mutational status results from Biomed-2 in Henri Becquerel Centre between July 1, 2002 and July 1, 2020 were included. To compare the sensitivity and ability of the two different methods, we chose cases with different intervals of *IGHV* identification. There were 14 unmutated cases (99-100%), 21 borderline cases (97-99%), 34 mutated cases (<97%) and 12 double rearrangement cases. This study was performed according to the Declaration of Helsinki and local laws, and the protocol was approved by the Institutional Review Board of Henri Becquerel Centre. Biological material was obtained after receiving patient consent. The diagnosis of CLL was made according to the established iwCLL/National Cancer Institute (NCI) criteria²⁷. RNA was isolated from peripheral blood mononuclear cells (PBMCs) using the RNA NOW kit (Biogenex, Seabrook, TX, USA) according to the instruction manual and stored at -80 °C in water. RNA concentrations were determined using a Nanodrop spectrophotometer.

5' RACE (rapid amplification of cDNA ends)

The first step of 5' RACE consists of cDNA synthesis (as presented in Figure 1) followed by an AMPure XP magnetic beads purification. The primers of Ig reverse transcription were listed in Appendix 1. The second step involves the addition of a poly A tail. The cDNA is amplified by two rounds of Polymerase chain reaction (PCR) extension, performed using a thermal cycler. The first PCR amplified the total first-strand cDNA with I1 and I2 (Appendix 2), and the second PCR was performed to add P5 and P7 adaptators. AMPure XP was added directly to the second PCR for product purification. The samples concentrations were assessed to determine the quantity of cDNA (pg/μL) and the dilution factor using a Qubit HS kit (ThermoFisher, Waltham, MA, USA). Using P5 and barcode P7 adaptor primers, different sample libraries could be pooled into a single sequencing assay and classified according to the barcode contained in the P7 adaptor primers.

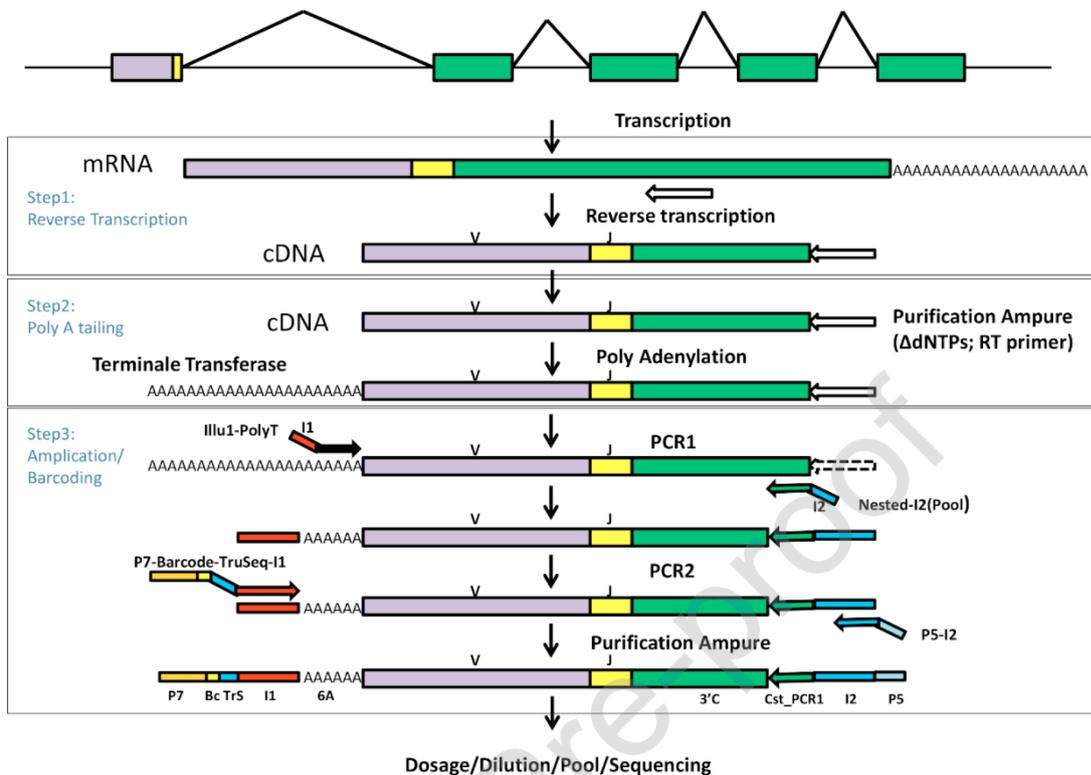


Figure 1 Procedure of 5' RACE. It included three steps before sequencing: step 1 is reverse transcription, producing a single strand cDNA, step 2 consists in the addition of the poly A tail, and step 3 is a two rounds PCR amplification adding P7 and P5 adaptors as well as individual barcodes.

Next generation sequencing (NGS)

For preparation, PCR products were diluted to 4 nM with EBT (converting pg/μL into nmol/L), and the samples (groups) were mixed by volume to constitute a library. The concentration of library is 8 pM, and we analysis the sequence according to the Illumina sequence manufacturer's instructions. The library concentration is 15491.9±7549.1 (4920-38900) pg/μl, the number of raw reads per sample is 217057.5±1.1 (3571-779822), the number of P-C-* reads is 86408.6±79576.0 (6613-722655), the percent of P-C-* in raw reads is 36.9±12.8 (2.1-56.27) %, the mean length of P-C-* is 234.5±56.0 (112.1-339.9). All the data was showed mean±SD (Min-Max). In the 81 patients, only one patient showed polyclone light chain.

Bioinformatic analysis

Each FASTQ was analyzed using a dedicated bioinformatics pipeline composed of four steps: the identification and trimming of primer sequences, the correction of sequencing artifacts, the identification of VDJ recombination and finally mutation rate estimation.

1. Sequence identification

Primers were designed to capture immunoglobulin (Ig) heavy and light chains. We implement Aho-Corrasick string-searching algorithm²⁸ to find primer sequences and assign each read to its chain. Found primers were trimmed and only raw sequences having an expected primer were retained for further analysis. List of primers are presented in Appendix 3.

2. Sequencing artifact correction

An occurrence table is built from all the trimmed sequences. To prevent reads with sequencing errors from being considered real mutations, reads with a low occurrence are compared to those with a high occurrence to identify whether observed substitutions could be explained by a low base quality score. In practice, this comparison is performed exclusively for sequences of the same length and within the limit of two substitutions at most. If the observed mismatches have a score strictly less than Q30, then the wrong read is discarded.

3. Identification of VDJ recombination

Reads were compared against known V, D and J germline sequences using IgBlast v1.15.0²⁹. Germline V, D and J sequences of Ig were downloaded from the international ImMunoGeneTics information system® (IMGT/V-QUEST³⁰). All Ig V, D and J sequences were combined into separate FASTA files and used as search databases. Rearrangement summary reports are then used to annotate the sequences and estimate the mutation rate.

4. Identification of mutation status

Mutation status is determined by percent homology with germline. A clone with greater than 98% homology is considered to be unmutated and vice versa. Some cDNAs can carry one or more STOP codons (TGA, TAG, TAA), which are also called nonsense codons, or be out-of-frame (out of reading frame). This kind of transcribed DNA not translated into immunoglobulin, will be identified as unproductive rearrangements.

Statistical analysis

The statistical analysis was performed using SPSS v17³¹ and GraphPad Prism 5.0³² software. Comparisons between variables were assessed by Chi-square test (and Yates' correction) or Fisher's exact test for qualitative variables and t tests or non-parametric Mann-Whitney test for quantitative data. Correlation between mutation ratio of heavy and light chain obtained by 5' RACE was established using Spearman's correlation test in non-IGLV3-21^{R110} group and Pearson test in IGLV3-21^{R110} group. Overall survival (OS) was measured from the time of diagnosis to the date of death from any cause. Time-to-first-treatment (TTFT) was calculated from the time of diagnosis to the date of the first treatment initiation. Survival probabilities were calculated using the Kaplan-Meier method. Log-rank tests were performed to evaluate the impact of mutational status on survival. A two-tailed p-value below 0.05 was considered statistically significant.

Results

1. Comparison of the characterization of the mutational status by conventional methods and by 5' RACE

The results of 81 patients assessed by 5' RACE and by conventional multiplex PCR following the Biomed-2 protocol are summarized in Table 1 and Figure 2A. Evaluation of the *IGHV* mutational status returned concordant results in 76 patients (93.8%) with 27 unmutated, 47 mutated and 2 patients with double (one mutated and one unmutated) rearrangements. Five patients obtained discordant results from the two methods. Specifically, 3 unproductive, 1 unmutated and 1 double *IGHV* rearrangements were reported based on Biomed-2 results, which were reported as 3 productive mutated, 1 mutated and 1 triple *IGHV* rearrangement by 5' RACE, respectively. When we compared the sequences of these 5 patients, different conclusions were obtained for four patients (#64, 76, 78 and 80) based on the two methods mainly due to the poor quality of the Sanger sequencing, even though clones were identified on gel electrophoresis. In contrast, 5' RACE unambiguously demonstrated that all 4 cases expressed productive sequences and solved technical problems due to the co-expression of three productive rearrangements, 2 mutated and one unmutated genes. For patient #42, the two methods yielded the

same sequence and CDR3 results, but the mutation identification rate differed (98.3% for Biomed-2 and 97.5% for 5' RACE) (Figure 2B).

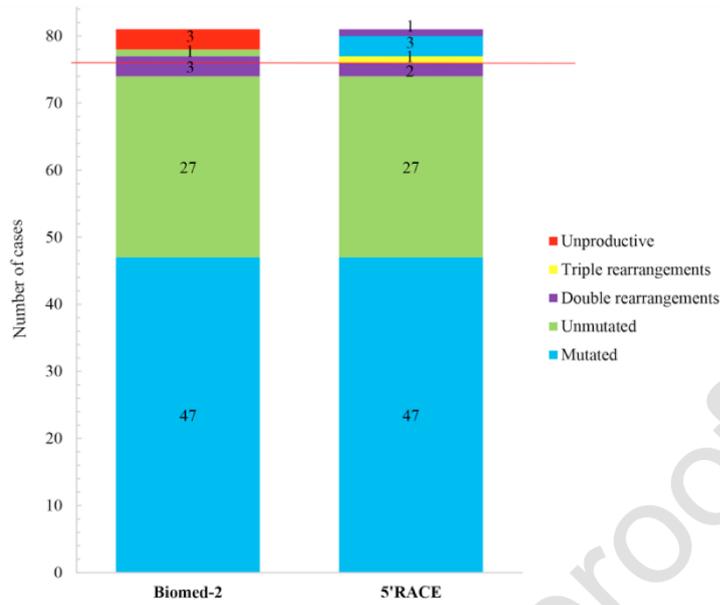
Table 1. 5' RACE and Biomed-2 Results

IGH (Biomed-2)					IGH (5' RACE)					IGL (5' RACE)						
Mut					Mut					Mut						
U/M	Id%	V	D	J	U/M	Id%	Ig	V	D	J	U/M	Id%	Ig	V	J	
1	U	99	IGHV1	IGHD6	IGHJ4	U	100	M/D	IGHV1-2	IGHD6-19	IGHJ4	U	100	K	IGKV1-39	IGKJ2
2	U	99.5	IGHV1-3	IGHD6-19	IGHJ4	U	100	M/D	IGHV1-3	IGHD6-19	IGHJ4	U	100	K	IGKV1-39	IGKJ3
3	U	99.6	IGHV3-21	IGHD1-26	IGHJ4	U	100	M/D	IGHV3-21	IGHD1-26	IGHJ4	U	100	K	IGKV6-21	IGKJ1
4	U	99.6	IGHV3-21	IGHD3-9	IGHJ5	U	100	M/D	IGHV3-21	IGHD3-9	IGHJ5	U	100	K	IGKV1-33	IGKJ5
5	U	100	IGHV3-23	IGHD3-3	IGHJ4	U	100	M/D	IGHV3-23	IGHD3-3	IGHJ4	U	99.6	K	IGKV4-1	IGKJ5
6	U	99.7	IGHV3-23	IGHD3-3	IGHJ6	U	99.5	M/D	IGHV3-23	IGHD3-3	IGHJ6	U	100	K	IGKV2-28	IGKJ4
7	U	100	IGHV3-23	IGHD3	IGHJ4	U	100	M/D	IGHV3-23	IGHD3	IGHJ4	U	99.5	K	IGKV1-5	IGKJ1
8	U	99.5	IGHV3-23	IGHD4-17	IGHJ4	U	99.7	M	IGHV1-58	IGHD4-17	IGHJ4	U	100	K	IGKV3-15	IGKJ2
9	U	98.6	IGHV3-33	IGHD3-16	IGHJ6	U	100	M/D	IGHV3-33	IGHD3-10	IGHJ6	U	100	K	IGKV2-28	IGKJ3
10	U	100	IGHV5-51	IGHD3-3	IGHJ4	U	100	M/D	IGHV5-51	IGHD3-3	IGHJ4	U	100	K	IGKV3-15	IGKJ2
11	U	99.2	IGHV1-58	IGHD3-3	IGHJ4	U	100	M/D	IGHV1-58	IGHD3-3	IGHJ4	U	100	K	IGKV2-28	IGKJ1
	U	99.6	IGHV3-71*	NA	IGHJ5	U	100	M/D	IGHV3-71*	IGHD3-9	IGHJ4	U	100	K	IGKV2-28	IGKJ1
12	U	99.2	IGHV5-51	IGHD1-26	IGHJ4	U	100	M/D	IGHV5-51	IGHD1-26	IGHJ4	U	100	K	IGKV1-39	IGKJ3
	U	99.2	IGHV3-30*	IGHD2-2	IGHJ4	U	100	G	IGHV3-30*	IGHD2-2	IGHJ4	U	100	K	IGKV1-39	IGKJ3
13	U	100	IGHV4-31	IGHD5-18	IGHJ4	U	100	M/D	IGHV4-31	IGHD5-18	IGHJ4	U	100	K	IGKV3-11	IGKJ4
	U	100	IGHV1-2	IGHD6-19	IGHJ4	U	100	M/D	IGHV1-2	IGHD6-19	IGHJ4	U	100	K	IGKV3-11	IGKJ4
14	U	98.6	IGHV4-39	IGHD3-22	IGHJ3	U	100	M/D	IGHV4-39	IGHD3-22	IGHJ3	U	99.5	K	IGKV3-11	IGKJ3
	U	100	IGHV3-23	IGHD4-11	IGHJ6	U	100	M/D	IGHV3-23	IGHD4-11	IGHJ6	U	100	K	IGKV3-11	IGKJ3
15	U	100	IGHV4-39	IGHD3-3	IGHJ4	U	100	M/D	IGHV4-39	IGHD3-3	IGHJ4	U	100	K	IGKV1-33	IGKJ4
	U	100	IGHV3-33	IGHD3-3	IGHJ6	U	100	M/D	IGHV3-33	IGHD3-3	IGHJ6	U	100	K	IGKV1-12	IGKJ1
16	U	99	IGHV3-15	IGHD3-16	IGHJ4	U	99.3	M/D	IGHV3-15	IGHD3-16	IGHJ5	U	100	L	IGLV3-21	IGLJ3
17	U	98.6	IGHV3-15	IGHD2-21	IGHJ4	U	98.6	M/D	IGHV3-15	IGHD2-21	IGHJ4	U	99.1	L	IGLV3-21	IGLJ1
18	U	98.3	IGHV3-23	IGHD6-25	IGHJ4	U	98.1	M/D	IGHV3-23	IGHD6-25	IGHJ4	U	99.6	L	IGLV3-21	IGLJ3
19	U	98.6	IGHV3	IGHD6	IGHJ6	U	98.6	M/D	IGHV3-33	IGHD3-3	IGHJ6	U	99.1	L	IGLV3-1	IGLJ2
20	U	98.6	IGHV3-48	IGHD6-13	IGHJ4	U	98.1	M/D	IGHV3-48	IGHD6-13	IGHJ4	U	99.1	L	IGLV3-21	IGLJ3
21	U	100	IGHV4-34	IGHD3-3	IGHJ6	U	100	M/D	IGHV4-34	IGHD3-3	IGHJ6	U	100	L	IGLV1-44	IGLJ1
22	U	99	IGHV4-59	IGHD3-22	IGHJ4	U	99.1	M/D	IGHV4-59	IGHD2-8	IGHJ4	U	100	L	IGLV3-21	IGLJ1
23	U	98.6	IGHV3-48	IGHD4-23	IGHJ1	U	98	M/D	IGHV3-48	IGHD1-26	IGHJ4	U	99.1	L	IGLV3-21	IGLJ3
	U	99.5	IGHV5-51*	IGHD1-1	IGHJ4	U	99.5	M/D	IGHV5-51*	IGHD1-1	IGHJ4	U	99.1	L	IGLV3-21	IGLJ3
24	U	98.3	IGHV3-7	IGHD3-10	IGHJ4	U	98.6	M/D	IGHV3-7	IGHD3-10	IGHJ4	M	95.3	K	IGKV1-39	IGKJ4
25	U	100	IGHV1-69	IGHD3-3	IGHJ6	U	100	M/D	IGHV1-69	IGHD3-3	IGHJ6	U	100	K	IGKV3-20	IGKJ5
	M	96.9	IGHV3-48	IGHD6-6	IGHJ3	M	96.7	M/D	IGHV3-48	IGHD6-6	IGHJ3	M	96.5	K	IGKV1-8	IGKJ5
26	M	90	IGHV3-23	IGHD3-10	IGHJ4	M	92.4	M/D	IGHV3-23	IGHD3-10	IGHJ4	M	97.1	K	IGKV1-8	IGKJ4
27	M	88.5	IGHV3-23	IGHD6-19	IGHJ4	M	89.6	M/D	IGHV3-23	IGHD6-19	IGHJ4	M	94.7	K	IGKV3-15	IGKJ2
28	M	92	IGHV3-23	IGHD6-6	IGHJ6	M	90.2	M/D	IGHV3-23	IGHD6-6	IGHJ6	M	95.9	K	IGKV1-8	IGKJ3
29	M	88.3	IGHV3-23	IGHD3-3	IGHJ4	M	89.7	M/D	IGHV3-23	IGHD3-3	IGHJ4	M	93	K	IGKV1-5	IGKJ1
30	M	90	IGHV3-23	IGHD1-26	IGHJ5	M	90.5	M/D	IGHV3-23	IGHD1-26	IGHJ5	M	95.8	K	IGKV1-5	IGKJ1
31	M	94.4	IGHV3-23	IGHD1-26	IGHJ4	M	94.7	M/D	IGHV3-23	IGHD1-26	IGHJ4	M	96.4	K	IGKV1-8	IGKJ1
32	M	94.4	IGHV3-23	IGHD3-16	IGHJ5	M	94.9	M/D	IGHV3-23	IGHD3-16	IGHJ5	M	90.1	K	IGKV4-1	IGKJ4
	M	90.5	IGHV1-3*	IGHD3-22	IGHJ5	M	90.5	M/D	IGHV1-3*	IGHD3-22	IGHJ5	M	90.5	K	IGKV1-12	IGKJ4
33	M	95.9	IGHV4-34	IGHD3-22	IGHJ4	M	96	M/D	IGHV4-34	IGHD3-22	IGHJ4	M	96.4	K	IGKV3-20	IGKJ1
34	M	96.8	IGHV4-34	IGHD4-17	IGHJ6	M	97.5	M/D	IGHV4-34	IGHD4-17	IGHJ6	M	95	K	IGKV1-5	IGKJ2
35	M	92.3	IGHV4-34	IGHD2-15	IGHJ4	M	91.3	M/D	IGHV4-34	IGHD2-15	IGHJ4	M	93.2	K	IGKV3-20	IGKJ2
36	M	93.8	IGHV5-51	NA	IGHJ3	M	93.5	M/D	IGHV5-51	IGHD5-18	IGHJ6	M	96.3	K	IGKV3-15	IGKJ2
	M	75	IGHV1-3*	IGHD3-22	IGHJ5	M	90.1	M/D	IGHV1-3*	IGHD3-22	IGHJ5	M	96.3	K	IGKV3-15	IGKJ2
37	M	95.3	IGHV6-1	IGHD3-10	IGHJ6	M	93.7	D	IGHV6-1	IGHD3-10	IGHJ6	M	95.3	K	IGKV6-21	IGKJ2
38	M	97.6	IGHV3-21	IGHD3-22	IGHJ6	M	97.8	M/D	IGHV3-21	NA	IGHJ6	M	97	L	IGLV3-21	IGLJ3
39	M	87.3	IGHV3-23	IGHD6-13	IGHJ5	M	88.2	M/D	IGHV3-23	IGHD6-13	IGHJ5	M	90.9	L	IGLV2-23	IGLJ3
40	M	97.2	IGHV3-23	IGHD6-13	IGHJ4	M	96.5	M	IGHV3-23	IGHD6-13	IGHJ4	M	91.9	L	IGLV2-11	IGLJ1
41	M	92.3	IGHV3-23	IGHD6-13	IGHJ4	M	92.5	M/D	IGHV3-23	IGHD6-13	IGHJ4	M	95.7	L	IGLV2-14	IGLJ3
42	U	98.3	IGHV3-48	IGHD3-3	IGHJ6	M	97.5	M/D	IGHV3-48	IGHD3-10	IGHJ6	M	97.9	L	IGLV3-21	IGLJ1
43	M	89.1	IGHV4-34	IGHD1-1	IGHJ4	M	91	M/D	IGHV4-34	IGHD1-1	IGHJ4	M	96.5	L	IGLV1-47	IGLJ2
44	M	91.5	IGHV4-34	IGHD4-11	IGHJ5	M	91.5	M/D	IGHV4-34	IGHD4-11	IGHJ5	M	96.9	L	IGLV2-11	IGLJ1
45	M	90.2	IGHV4-39	IGHD2-21	IGHJ4	M	97.6	M/D	IGHV4-39	IGHD2-21	IGHJ4	M	97.4	L	IGLV3-1	IGLJ2
46	M	94.8	IGHV4-61	IGHD1-14	IGHJ6	M	96.1	M/D	IGHV4-61	IGHD1-14	IGHJ6	M	96.1	L	IGLV3-25	IGLJ2
47	M	93.8	IGHV3-23	IGHD3-10	IGHJ3	M	96.9	M/D	IGHV3-23	IGHD3-10	IGHJ6	Polyclone light chain is short				
48	M	92.7	IGHV3-23	IGHD6-19	IGHJ4	M	91.8	M/D	IGHV3-23	IGHD6-19	IGHJ4	U	98.7	K	IGKV2-30	IGKJ3
49	M	92	IGHV3-72	IGHD2-8	IGHJ6	M	94.8	M/D	IGHV3-72	IGHD2-8	IGHJ6	U	98.8	K	IGKV1-16	IGKJ4
50	M	97.2	IGHV4-39	IGHD3-10	IGHJ4	M	93.6	M/D	IGHV4-39	IGHD3-10	IGHJ4	U	98.8	K	IGKV4-1	IGKJ2
51	M	97.2	IGHV1-24	IGHD3-10	IGHJ4	M	95.7	M/D	IGHV1-24	IGHD3-10	IGHJ4	U	98.7	L	IGLV3-21	IGLJ1
52	M	96.9	IGHV3-15	IGHD4-17	IGHJ4	M	97.5	M/D	IGHV3-15	IGHD4-17	IGHJ4	U	99.3	L	IGLV3-21	IGLJ3
53	M	95.6	IGHV3-21	IGHD3-9	IGHJ5	M	97	M/D	IGHV3-21	IGHD3-9	IGHJ6	U	99.4	L	IGLV3-21	IGLJ3
54	M	96.7	IGHV1-58	IGHD2-2	IGHJ5	M	96.1	M/D	IGHV3-21	NA	IGHJ6	U	98.9	L	IGLV3-21	IGLJ3
55	M	96.3	IGHV3-21	IGHD1-14	IGHJ6	M	96.9	M/D	IGHV3-21	IGHD1-14	IGHJ6	U	99.6	L	IGLV3-21	IGLJ3

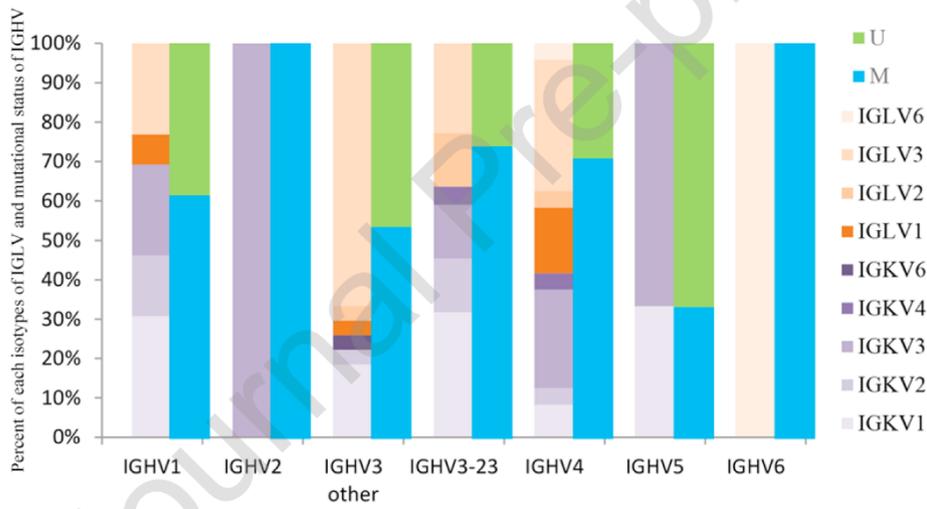
56	M	97.2	IGHV3-21	NA	IGHJ6	M	97.4	M/D	IGHV3-21	NA	IGHJ6	U	99.6	L	IGLV3-21	IGLJ3
57	M	97.6	IGHV3-23	IGHD6-6	IGHJ4	M	97.5	M/D	IGHV3-23	IGHD6-6	IGHJ4	U	98.3	L	IGLV3-21	IGLJ3
58	M	97.2	IGHV3-23	IGHD5-24	IGHJ4	M	96.3	M/D	IGHV3-23	IGHD5-24	IGHJ4	U	98.3	L	IGLV3-21	IGLJ3
59	M	97.2	IGHV3-48	IGHD2-15	IGHJ4	M	97.1	M/D	IGHV3-48	IGHD2-15	IGHJ4	U	99.6	L	IGLV3-21	IGLJ3
60	M	96.8	IGHV3-53	IGHD1-26	IGHJ4	M	96.8	M/D	IGHV3-53	IGHD1-26	IGHJ4	U	99.1	L	IGLV3-21	IGLJ3
61	M	97.9	IGHV3-21	NA	IGHJ6	M	94.7	M/D	IGHV3-69	NA	IGHJ6	U	99.4	L	IGLV3-21	IGLJ3
62	M	97.6	IGHV3-74	IGHD5-18	IGHJ4	M	97.6	M/D	IGHV3-74	IGHD5-18	IGHJ4	U	99.6	L	IGLV3-21	IGLJ1
63	M	88	IGHV3-23	NA	IGHJ4	M	97.2	M/D	IGHV3-23	IGHD1-26	IGHJ4	U	98.7	L	IGLV3-21	IGLJ1
	M	97.5	IGHV2-10*	IGHD6-13	IGHJ5	U	99.5	G/M	IGHV2-10*	IGHD6-13	IGHJ5					
64	M	97.6	IGHV3-13*	IGHD2-2	IGHJ4	M	97.9	M/D	IGHV3-48	IGHD4-17	IGHJ4	U	99.6	L	IGLV3-21	IGLJ3
						M	97.3	G/M	IGHV3-13*	IGHD2-2	IGHJ4					
65	M	96.8	IGHV3-74	IGHD3-10	IGHJ3	M	95.4	M/D	IGHV3-74	IGHD3-10	IGHJ3	U	98.7	L	IGLV3-21	IGLJ1
	U					U	99.3	M/D	IGHV6-1*	IGHD6-6	IGHJ4					
66	M	97.5	IGHV4-61	IGHD4-11	IGHJ3	M	97.3	M/D	IGHV4-59	IGHD4-11	IGHJ3	U	98.7	L	IGLV3-21	IGLJ1
67	U	100	IGHV4-34	IGHD6-6	IGHJ6	U	100	G	IGHV4-34	IGHD6-6	IGHJ6	U	100	K	IGKV3-20	IGKJ1
68	U	99.6	IGHV3-30	IGHD1-20	IGHJ6	U	100	M/D/G	IGHV3-30	IGHD5-12	IGHJ6	U	98.9	L	IGLV3-21	IGLJ2/3
69	U	98.6	IGHV4-39	IGHD6-13	IGHJ5	U	98.9	G	IGHV4-39	IGHD6-13	IGHJ5	U	100	L	IGLV3-19	IGLJ2
	M	91	IGHV1-69	IGHD2-15	IGHJ4	M	91.2	G	IGHV1-69	IGHD2-15	IGHJ4					
70	M	96.7	IGHV4-34	IGHD3-22	IGHJ6	M	97.3	G	IGHV4-34	IGHD3-22	IGHJ6	M	97.7	K	IGKV2-30	IGKJ2
71	M	94.7	IGHV4-34	NA	IGHJ2	M	93.6	M/D	IGHV4-34	IGHD1-26	IGHJ4	M	94.6	K	IGKV3-20	IGKJ1
						M	94.5	G	IGHV1-8	IGHD6-25	IGHJ5	M	97.7	K	IGKV1-9	IGKJ2
72	M	95.9	IGHV1-2	IGHD3-22	IGHJ4	M	96.6	G	IGHV1-2	IGHD3-22	IGHJ4	M	91.8	L	IGLV1-40	IGLJ2/3
73	M	95.9	IGHV4-34	IGHD3-9	IGHJ3	M	95.5	G	IGHV4-34	IGHD3-9	IGHJ3	M	94.8	L	IGLV3-1	IGLJ2
74	M	90.7	IGHV4-34	IGHD6-13	IGHJ5	M	89.6	G	IGHV4-34	IGHD6-13	IGHJ5	M	94.4	L	IGLV1-40	IGLJ3
75	M	96.5	IGHV4-34	IgHD3-16	IGHJ5	M	95.9	G	IGHV4-34	IgHD3-16	IGHJ5	M	95.2	L	IGLV6-57	IGLJ3
76	M	89.8	IGHV4-34*	NA	NA	M	90	G	IGHV4-59	IGHD6-19	IGHJ4	M	96.5	L	IGLV3-1	IGLJ2
77	M	94.7	IGHV4-34	IGHD6-6	IGHJ6	M	94.3	M/D	IGHV4-34	IGHD6-6	IGHJ6	M	94.8	L	IGLV3-1	IGLJ1
	M	64.9	IGHV1-3	IGHD2-15	IGHJ6	M	83.5	M/G	IGHV1-3	IGHD2-15	IGHJ6					
78	M	69.5	IGHV4-61*	IGHD5-18	IGHJ4	M	89.2	G	IGHV4-4	IGHD3-10	IGHJ4	M	94.4	L	IGLV1-44	IGLJ3
	M	95	IGHV3-33*	IGHD2-2	IGHJ5	M	93.7	M	IGHV3-23	IGHD2-21	IGHJ4					
79	U	100	IGHV1-69	IGHD3-9	IGHJ6	U	100	M/D	IGHV1-69	IGHD3-9	IGHJ6	U	100	K	IGKVID-8	IGKJ1
	U	100	IGHV3-11	IGHD2-2	IGHJ6	U	100	M/D	IGHV1-3	IGHD4/5	IGHJ3	U	100	K	IGKV2-28	IGKJ4
	U	100				U	100	G	IGHV3-11	IGHD2-2	IGHJ6	U	99.4	L	IGLV2-14	IGLJ2/3
80	M	97.2	IGHV1-69	IGHD3-10	IGHJ6	U	100	M/D	IGHV1-69	IGHD3-10	IGHJ6	U	100	L	IGLV3-27	IGLJ3
	U	99.7	IGHV3-33	IGHD5-24	IGHJ6	M	93.1	M/D	IGHV2-5	IGHD1	IGHJ4	M	96.4	K	IGKV3-11	IGKJ5
	U					U	100	M/D	IGHV3-33	IGHD5-24	IGHJ6	U	100	L	IGLV1-47	IGLJ1
81	M	90	IGHV3-23	IGHD6-19	IGHJ4	M	89.4	M/D	IGHV3-23	IGHD6-19	IGHJ4	U	99.8	K	IGKV2-28	IGKJ2
						M						M	92.3	L	IGLV3-16	IGLJ3

U: unmutated, M: mutated. * Unproductive rearrangement.

A



B



C

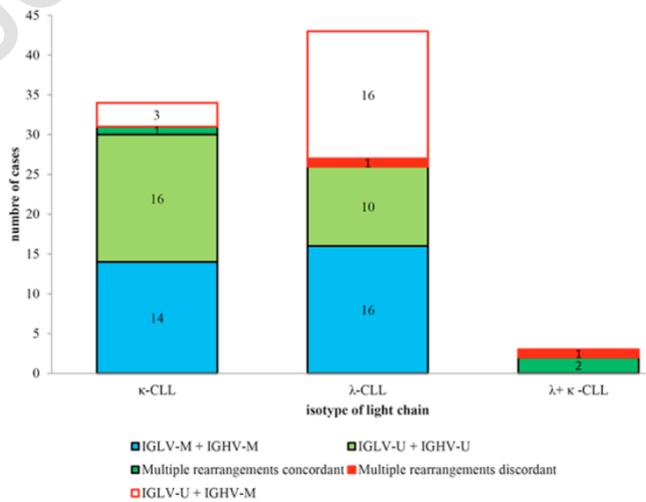


Figure 2 the results of two methods and complete information regarding *IGHV* and *IGLV*.

(A) Results of Biomed-2 and 5' RACE in 81 patients. Below the red horizontal line, patients' results were concordant using the two methods; the rate of concordance is 93.8%. Green indicates *IGHV*-unmutated patients, blue indicates *IGHV*-mutated patients, red indicates *IGHV*-unproductive patients, and purple and yellow indicate double and triple, respectively, productive rearrangements patients. (B) Relationship between *IGHV* and *IGLV* regions and the mutational statuses of different *IGHV*. (C) In different isotypes, relationship of mutational status of heavy and light chains. Green represents concordant unmutated status of heavy and light chains. Blue represents concordant mutated status of heavy and light chain. Dark green represents concordant mutational statuses of multiple rearrangements of heavy and light chains. Red represents discordant mutational status of heavy and light chains (Red blank box represent single rearrangement and red solid box represent multiple rearrangements).

We also chose 14 additional patients who had blood samples collected at different times after CLL diagnosis and tested the sample using 5' RACE (Appendix 4). The interval time between sample collections in one patient is from 0.64 to 10.89 years, and the median time is 4.59 years. For the samples collected at different time points, our methods detect the same productive immunoglobulin heavy and light chain and mutation status. Moreover, 6 patients had the same *IGHV* and *IGLV* mutation percentages.

2. Characterization of *IGHV* using 5' RACE

Based on 5'RACE, the isotype of the immunoglobulin heavy chain and sequences of light chain were available as determined by RT-PCR for mu, delta, gamma kappa and gamma constant region transcripts and NGS.

In the cohort of 81 patients, the male-to-female ratio was 2.86 (60 males: 21 females), and the median age was 64 (range 41–93). Here, 5' RACE identified specific clonal *IGH* and *IGL* rearrangements in 81/81 and 80/81 patients (one patient showed polyclone light chain), respectively, confirming the efficiency of the method.

At least one productive *IGH* rearrangement was identified in all cases. Considering the number and productive status of arrangements, 64 cases presented a single productive rearrangement (79.01%), 15 cases harbored two rearrangements (two productive sequences in 8 cases and one productive and one unproductive sequence in 7 cases), and 2 cases presented three different rearrangements (all productive). The most frequent variable segment was *IGHV3* (50 cases [61.7%] with a bias toward *IGHV3-23* in 23 cases [28.4%]) followed by *IGHV4* (24 cases [29.6%] with a bias toward *IGHV4-34* in 13 cases [16%]).

Regarding the mutational status, 44 of the 64 cases who presented a single productive *IGH* rearrangement were mutated, and 20 were unmutated. For the 15 cases who presented two rearrangements, 8 harbored two productive rearrangements, including 3 concordant mutated cases, 3 concordant unmutated cases, and 2 discordant cases with one mutated and one unmutated sequence. The 7 remaining cases showed one productive and one unproductive rearrangement, including 2 with one productive mutated and one unproductive unmutated sequences, 2 with one productive mutated and one unproductive mutated sequence (considered mutated), and 3 with one productive unmutated and one unproductive unmutated sequence (unmutated cases). For the last two cases, we detected three productive rearrangements. Specifically, one case showed three unmutated sequences, and the other showed one mutated and two unmutated sequences (considered mutated).

More importantly, 5' RACE also provides information about the BCR isotype. A majority of CLL cases in our series expressed only C μ and C δ heavy chains (65/81; 80.2%). Fifty-six of these cases expressed a single productive rearrangement (with 18 unmutated and 38 mutated sequences), 4 expressed two productive rearrangements (3 concordant unmutated and 1 discordant with mutated and unmutated sequences) and 4 expressed one productive and one unproductive rearrangement (2 concordant unmutated, 1 concordant mutated and 1 discordant productive mutated and unproductive unmutated). The last IgM case expressed three productive rearrangements with 2 mutated and 1 unmutated sequences.

Eight other cases in our series expressed only a C γ heavy chain. Seven showed a single productive rearrangement (6 mutated, 1 unmutated), and 1 showed two productive IgG rearrangements. Interestingly, we observed a bias toward the usage of the VH4 family in this population (6/8 cases) with 5 patients expressing IgG4-34 sequences.

Unexpectedly, the 8 remaining cases simultaneously expressed the *IGH* chain with different isotypes, suggesting that the class switch recombination process was ongoing in the tumor. One case expressed three unmutated productive rearrangements (two IgM/D and one IgG). Three cases expressed one productive IgM and one unproductive IgG rearrangement, 3 cases expressed two productive but different IgM and IgG sequences, and the last case showed a very unusual pattern, expressing three different unmutated heavy chains sharing the same CDR3 (IgM, D and G). No case in our series expressed IgA or IgE.

3. Characterization of the *IGLV* gene using 5' RACE

Another advantage of 5' RACE is that it provides information about the sequence of the immunoglobulin light chain. In our series, we characterized a monoclonal light chain in 80 cases, whereas only 1 case appeared to express polyclonal light chains. Thirty-four patients expressed productive Igκ rearrangements, 43 patients expressed productive Igλ rearrangements, and 3 patients showed a mixed Igκ and Igλ phenotype.

Thirty of the 34 Igκ-positive cases expressed a single productive rearrangement (18 unmutated and 12 mutated), and 4 cases expressed two productive rearrangements (concordant unmutated in one case, concordant mutated in two cases and discordant unmutated and mutated in the last case). All 43 Igλ-positive cases expressed a single productive rearrangement (26 unmutated and 17 mutated). We also observed a very significant bias with 23 of the 26 unmutated Igλ-positive cases expressing IGLV3-21 light chains. The last 3 cases expressed both Igκ and Igλ sequences, 1 harboring two productive rearrangements of one mutated Igλ and one unmutated Igκ, 1 with three unmutated productive rearrangements of one Igλ and two Igκ and 1 showing three productive rearrangements with two unmutated Igλ and one mutated Igκ.

Complete analysis of the LCDR3 region was possible in 80 patients. The length of the CDR3 of the Igκ light chains ranged from 8 to 11 amino acids (median 9 amino acids); the length of the CDR3 of the Igλ light chains ranged from 8 to 13 amino acids (median 11 amino acids). Interestingly, we observed that some cases in our series shared the same light chains with very similar CDR3 sequences, especially for those cases that expressed *IGLV3-21* rearrangements (Table 2).

Table 2. List of light chains that share identical CDR3 in 80 CLL patients

ARN	Light chain			CDR3	Heavy chain				CDR3
	IGLV	IGLJ	HOM (%)		IGHV	IGHD	IGHJ	HOM (%)	
25	IGKV1-8	IGKJ5	96.5	QQYYDYFST	IGHV3-48	IGHD6-6	IGHJ3	96.7	ARGRAYTSSPYDAFDV
31	IGKV1-8	IGKJ1	96.4	-H--S--W-	IGHV3-23	IGHD1-26	IGHJ4	94.7	AKSLGSGYGQIDY
26	IGKV1-8	IGKJ4	97.1	----G--L-	IGHV3-23	IGHD3-10	IGHJ4	92.4	AKVHSDTEYGSLSY
18	IGKV1-8	IGKJ3	95.9	----F-R-	IGHV3-23	IGHD6-6	IGHJ6	90.2	AKGSSSSGSYYGMDV
79	IGKV1D-8	IGKJ1	100	----SF-W-	IGHV1-69	IGHD3-9	IGHJ6	100	ARGNTYYDILTGYYDGVYYGMDV
12	IGKV1-39	IGKJ3	100	QQSYSTPPFT	IGHV5-51	IGHD1-26	IGHJ4	100	ARHPWYSGYSYFDY
1	IGKV1-39	IGKJ2	100	-----Y-	IGHV1-2	IGHD6-19	IGHJ4	100	ARLQWLVIHFYDY
2	IGKV1-39	IGKJ3	100	-----	IGHV1-3	IGHD6-19	IGHJ4	100	---E---LR----
24	IGKV1-39	IGKJ4	95.4	---H-I- L-	IGHV3-7	IGHD3-10	IGHJ4	98.6	ARSYSGSGSYSLRGRFDY
9	IGKV2-28	IGKJ3	100	MQALQTPGFT	IGHV3-30	IGHD3-10	IGHJ6	100	ARDRTSLGDPPLWGYYGMDV
6	IGKV2-28	IGKJ4	100	-----L -	IGHV3-23	IGHD3-3	IGHJ6	99.5	AKVTLEWLGYGMDV
11	IGKV2-28	IGKJ1	100	----- -	IGHV1-58	IGHD3-3	IGHJ4	100	AASYDFWGSYSH
81	IGKV2-28	IGKJ2	99.1	-----Y -	IGHV3-23	IGHD6-19	IGHJ4	87.4	ASGGPYGVAGNPFYDY
79	IGKV2-28	IGKJ4	100	-----Q -	IGHV1-3	IGHD4-23	IGHJ3	100	ARDDSGGYAFDI
10	IGKV3-15	IGKJ2	100	QQYNNWPPSYT	IGHV5-51	IGHD3-3	IGHJ4	100	ARRNYDFWSGYTPAPFDY
8	IGKV3-15	IGKJ2	100	-----L --	IGHV3-23	IGHD4-17	IGHJ4	99.5	ANGPRGDYVFPFDY

27	IGKV3-15	IGKJ2	94.7	-----F -	IGHV3-23	IGHD2-8	IGHJ4	89.6	AKASVLQWVSSTYYFDY
36	IGKV3-15	IGKJ2	96.3	---DT---R--	IGHV5-51	IGHD5-18	IGHJ6	93.5	VKGGYSYGYHYMDV
76	IGLV3-1	IGLJ2	96.5	QAWDTSTAADV	IGHV4-59	IGHD6-19	IGHJ4	90.0	ARDLRSSGWSPYFDY
19	IGLV3-1	IGLJ2	99.1	----S-- VV	IGHV3-33	IGHD3-3	IGHJ6	98.6	AKDAALTVSCMDV
45	IGLV3-1	IGLJ2	97.4	----S-- M-	IGHV4-39	IGHD2-21	IGHJ4	97.6	GRRYGDSTNAIFDY
73	IGLV3-1	IGLJ2	94.8	----S-- --	IGHV4-34	IGHD3-9	IGHJ3	95.5	ARKQGEDMALRYFDWSYAFDI
77	IGLV3-1	IGLJ1	94.8	----S--Y-	IGHV4-34	IGHD6-6	IGHJ6	94.3	ARVAGRPGYMDV
55	IGLV3-21	IGLJ3	99.6	QVWDSSSDHPWV	IGHV3-21	IGHD1-14	IGHJ6	96.9	ATDRNGMDV
56	IGLV3-21	IGLJ3	98.3	-----G-----	IGHV3-21	NA	IGHJ6	97.4	TR--T----
53	IGLV3-21	IGLJ3	99.4	-----	IGHV3-21	IGHD3-9	IGHJ6	97.0	-L-----
61	IGLV3-21	IGLJ3	99.6	-----	IGHV3-21	NA	IGHJ6	94.7	-A-----
54	IGLV3-21	IGLJ3	99.1	-----	IGHV3-21	NA	IGHJ6	96.1	-R-A-V---
38	IGLV3-21	IGLJ3	97.0	---E---Q---	IGHV3-21	NA	IGHJ6	97.8	-R-ESV---
52	IGLV3-21	IGLJ3	99.3	-----	IGHV3-15	IGHD4-17	IGHJ4	97.5	TADGDYRGTLRN
16	IGLV3-21	IGLJ3	100	-----	IGHV3-15	IGHD3-16	IGHJ5	99.4	TTDSFS
17	IGLV3-21	IGLJ1	99.1	-----L -	IGHV3-15	IGHD2-21	IGHJ4	98.6	TTDLHCGGDCNFDY
51	IGLV3-21	IGLJ1	98.7	-----LY-	IGHV1-24	IGHD3-10	IGHJ4	95.7	ATGGGSGSYTFDY
18	IGLV3-21	IGLJ3	99.6	-----L-M	IGHV3-23	IGHD6-25	IGHJ4	98.1	AKDLGDSSGYFDY
63	IGLV3-21	IGLJ1	98.7	-----Q-Y-	IGHV3-23	IGHD1-26	IGHJ4	97.2	--A-ELQ -----
58	IGLV3-21	IGLJ3	98.3	-----	IGHV3-23	IGHD5-24	IGHJ4	96.3	---P-SDGFKKN
57	IGLV3-21	IGLJ3	98.3	-----G-----	IGHV3-23	IGHD6-6	IGHJ4	97.4	--GD-T--H---
68	IGLV3-21	IGLJ2	98.9	-----VRG -	IGHV3-30	IGHD5-12	IGHJ6	100	ARDKGYDPSLGMDV
20	IGLV3-21	IGLJ3	99.1	-----	IGHV3-48	IGHD6-13	IGHJ4	98.2	ARVGSSTYFDY
23	IGLV3-21	IGLJ3	99.1	-----	IGHV3-48	IGHD1-26	IGHJ4	98.0	ATDGVGAPH
59	IGLV3-21	IGLJ3	99.6	-----	IGHV3-48	IGHD2-15	IGHJ4	97.1	ASDPHCSSGGSCYSYGY
64	IGLV3-21	IGLJ3	99.6	-----	IGHV3-48	IGHD4-17	IGHJ4	97.7	ARAVTTGGGC
42	IGLV3-21	IGLJ1	97.9	-----D--Y-	IGHV3-48	IGHD3-10	IGHJ6	97.6	ARDFGGETVQPPWNV
60	IGLV3-21	IGLJ3	99.1	-----	IGHV3-53	IGHD1-26	IGHJ4	96.9	ARGQGANDY
65	IGLV3-21	IGLJ1	98.7	-----G- -F	IGHV3-74	IGHD3-10	IGHJ3	95.4	ARDDSGGNSFDI
62	IGLV3-21	IGLJ1	99.6	-----L-Y-	IGHV3-74	IGHD5-18	IGHJ4	97.6	AGDLVDTAGGVFDY
22	IGLV3-21	IGLJ1	99.7	-----L Y-	IGHV4-59	IGHD2-8	IGHJ4	99.2	ARAPGDTGGYRIDY
66	IGLV3-21	IGLJ1	98.7	-----L-Y-	IGHV4-59	IGHD4-11	IGHJ3	97.4	AVWTVTDPFDI

4. The relationship between the somatic mutational status of the heavy and light chains

We observed a strong relationship between the mutational status of the heavy chain and the light chain in this cohort. In most cases, mutated IGHV rearrangements were associated with mutated IGK/IGL rearrangements and vice versa (Figure 2C) even though multiple light chain rearrangements were noted in some cases. For 3 cases co-expressing $Ig\kappa$ and $Ig\lambda$ light chains, the mutational status of the heavy and light chains were concordant in 2 cases. One expressed three unmutated rearrangements of heavy and light chains, and the other expressed one mutated and two unmutated rearrangements of heavy and light chains. Finally, the last case expressed a mutated heavy chain and two productive light chains with different mutational statuses.

Of the 34 $Ig\kappa$ -positive cases, the mutational statuses of the heavy and light chains were concordant in 30 (88.2%) cases with 16 double-unmutated and 14 double-mutated cases. One case expressed two productive heavy chains (one mutated and one unmutated) and two productive light chains (one mutated and one unmutated). Only 3 cases showed a discordant profile with a mutated heavy chain and an unmutated light chain.

Of the 43 Ig λ -positive cases, the mutational statuses of the heavy and light chains were concordant in 26 (60.5%) cases with 10 double-unmutated and 16 double-mutated cases. One patient (the only one in our whole series) expressed an unmutated heavy chain and a mutated light chain. All 16 other cases expressed a mutated heavy chain and an unmutated light chain. Strikingly, all 16 cases expressed an IGLV3-21 light chain with very similar CDR3 sequences (QVWDSSSDHPWV).

5. IGLV3-21^{R110} defines a new group of CLL

Then, we assessed all light chains and found that 31% (25/80) were IGLV3-21. The light chain in 25 IGLV3-21-CLL patients always shared a similar LCDR3 (Figure 3A), and most of them (24/25) had a point mutation, termed R110, between the variable and constant junction, named IGLV3-21^{R110}. All IGLV3-21^{R110} cases carried germline K16 residue, and 20/24 cases expressed the germline YDSD motif required for homotypic BCR-BCR interaction. The remaining 4 cases carried a motif changed by one residue, which has similar properties in all cases (three YDTD and one FDSD). In these 24 IGLV3-21^{R110}-CLL patients, 6 belonged to subset 2 (Figure 3B). Compared with other light chains, IGLV3-21^{R110}-CLL was more likely to exhibit a discordant mutational status of heavy and light chains (16/20 cases, Figure 3C). We divided patients into three groups based on the mutation status of the heavy chain and the point mutation of IGLV3-21: M-CLL non IGLV3-21^{R110}, IGLV3-21^{R110} and U-CLL non IGLV3-21^{R110}. When we compared the identification of the heavy chain in different groups, we found that IGLV3-21^{R110} was more likely to exhibit approximately 97-99% homology, representing borderline cases (Figure 3D). The rearrangements of the light chain were also different among subset 2 IGLV3-21^{R110}-CLL, IGLV3-21^{R110}-CLL and IGLV3-21-CLL patients (Figure 3E). All six subset 2 IGLV3-21^{R110}-CLLs expressed IGLV3-21*01 with IGLJ3*02, whereas only half of IGLV3-21^{R110}-CLL patients (9/18) showed the same rearrangements. The other half of IGLV3-21^{R110}-CLL patients (8/18) showed IGLV3-21*01 with IGLJ1*01. Only one patient with IGLV3-21 without the R110 mutation expressed IGLV3-21*01 with IGLJ2*01. The identification of the heavy chain and light chain showed a linear correlation in the non IGLV3-21^{R110} group ($p < 0.0001$) but not in the IGLV3-21^{R110} group ($p = 0.504$) (Figures 3F and 3G).

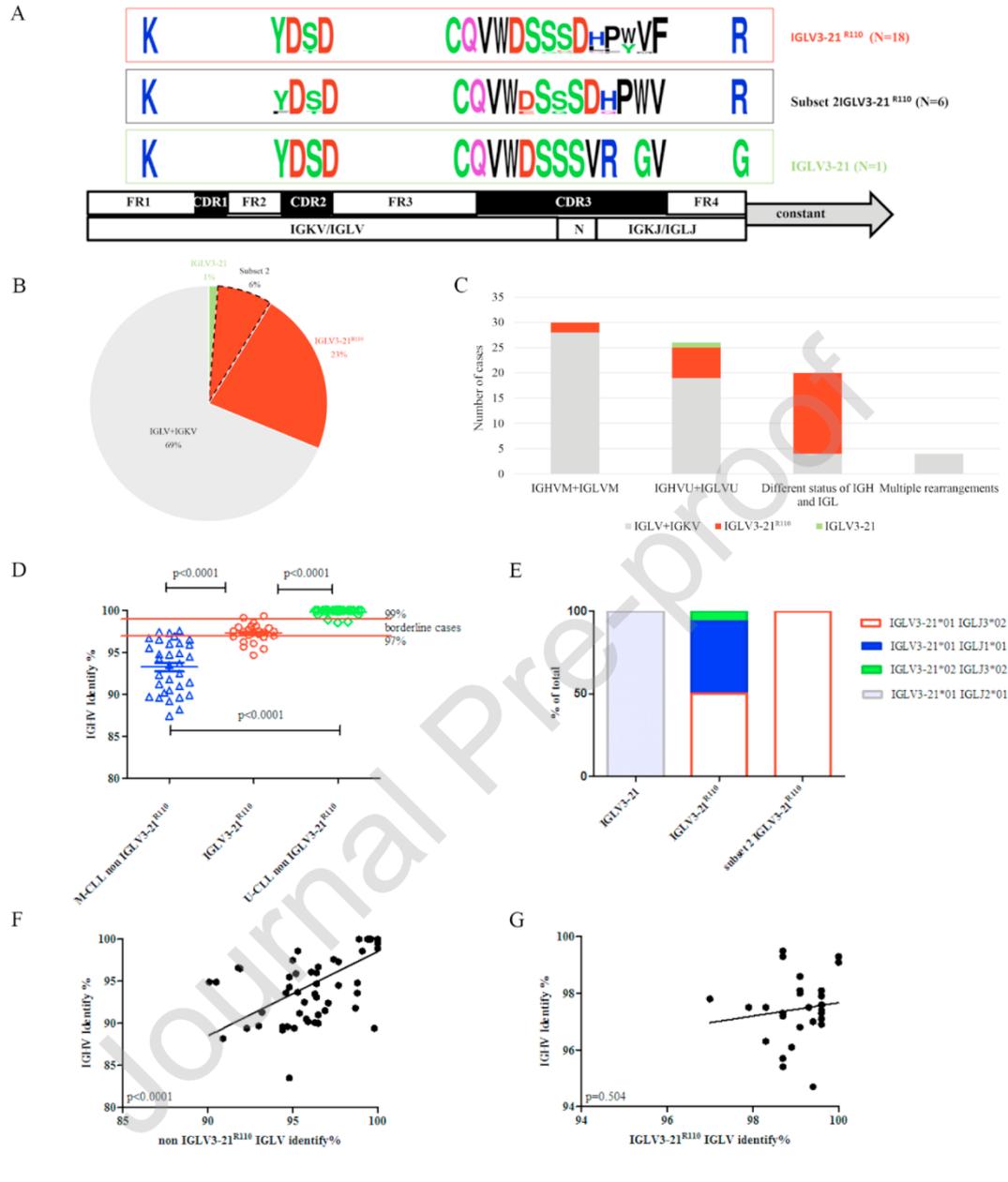


Figure 3 Value of the light chain.

(A) Sequence of all IGLV3-21 alleles, including the IGLV3-21, IGLV3-21^{R110} and subset 2 IGLV3-21^{R110} alleles. (B) Mutational relationship of IG heavy and light chains based on different types of light chains. (C) Distribution of IGLV3-21 in the total mutational status of heavy and light chains. (D) Identification of IGHV in M-CLL non-IGLV3-21^{R110}, IGLV3-21^{R110} and U-CLL non-IGLV3-21^{R110}. (E) Distribution of the J segment in all IGLV3-21 rearrangements. (F) Scatter plots of the identification of IGHV and IGLV in non-IGLV3-21^{R110} CLL (N=56). (G) Scatter plots of the identification of IGHV and IGLV in IGLV3-21^{R110} CLL (N=24).

6. The clinical value of 5' RACE and IGLV3-21^{R110}

We have summarized the clinical data of all the patients (Appendix 5). A significant difference in survival was noted between UM-CLL and M-CLL (Figures 4A and 4B). We divided patients into three groups as noted above (M-CLL non IGLV3-21^{R110}, IGLV3-21^{R110}-CLL and U-CLL non IGLV3-21^{R110}). We found that IGLV3-

21^{R110}-CLL and U-CLL non IGLV3-21^{R110} were more likely to occur in males than females ($p=0.042$), and these patients were more often classified as Binet stage B or C than A ($p=0.013$). Next, we analyzed the prognostic value of IGLV3-21^{R110} ($n=24$). Expectedly, M-CLL non IGLV3-21^{R110} ($n=33$) showed a longer OS and TTFT than UM-CLL non IGLV3-21^{R110} ($n=23$) (Figures 4C and D, $p=0.001$ and $p<0.0001$, respectively). Then, we compared the prognosis of IGLV3-21^{R110}-positive cases. Regardless of the mutational status of *IGHV*, IGLV3-21^{R110}-positive cases presented a similar poor prognosis compared with UM-CLL patients ($p=0.14$ for OS and $p=0.602$ for TTFT), even though most of them (18/24) presented mutated *IGHV* and showed a shorter survival time than M-CLL non IGLV3-21^{R110} ($p=0.05$ for OS and $p<0.0001$ for TTFT). Taken together, IGLV3-21^{R110}-CLL had a short time to first treatment and overall survival similar to unmutated CLL cases, whereas M-CLL non-IGLV3-21^{R110} had a good prognosis. IGLV3-21^{R110} defines a CLL subgroup representing a clinically aggressive group that shows a high-risk stage and is exclusively defined by LC identity regardless of *IGHV* family and mutational status.

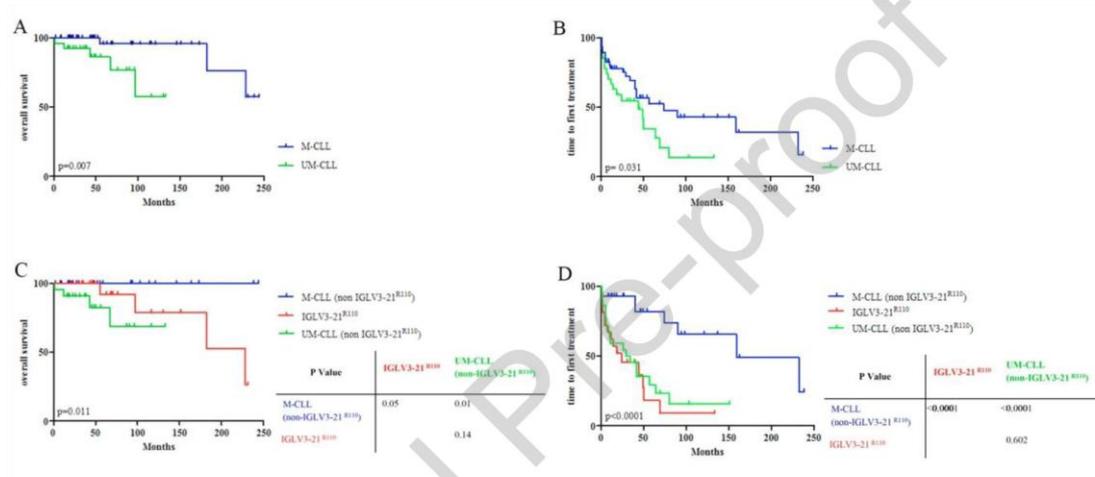


Figure 4 Survival analysis of the different groups.

(A) Analysis of overall survival in M-CLL and UM-CLL patients ($p=0.007$). (B) Analysis of time to first treatment in M-CLL and UM-CLL patients ($p=0.031$). (C) Analysis of overall survival in M-CLL non IGLV3-21^{R110}, UM-CLL non IGLV3-21^{R110} and IGLV3-21^{R110} ($p=0.011$). (D) Analysis of time to first treatment in M-CLL non IGLV3-21^{R110}, UM-CLL non IGLV3-21^{R110} and IGLV3-21^{R110} ($p<0.0001$).

DISCUSSION

CLL is one of the most frequent types of leukemia in Western countries. It generally occurs in elderly patients, and only 3.8% of patients with CLL were younger than 50 years in our study. More male than female patients were affected in this study, and the gender effect seems to be stable across several studies³³⁻³⁵. Over the past two decades, clinical and biological evidence has established that the BCR IG sequence is the ideal clonal marker for CLL. This marker is present from the birth of every CLL clone. In contrast to other markers, it remains stable over time and is unaffected by disease progression, which was also identified in this study. Recent evidence has reinforced the important prognostic and predictive value of immune genetic analysis in CLL^{3,4,36-38}; however, progress has also raised new questions, such as how to assess the prognosis of borderline cases and cases expressing multiple rearrangements, the value of the light chain, and the difference between IgG-CLL and IgM/D-CLL. These new problems highlight the need to move from a single determination of *IGHV* gene mutational status to a more comprehensive assessment of the *IGHV-D-J* gene rearrangement, including stereotyped subset assignment and light chain information, at least in certain cases^{14,39}.

The current standard method of interrogating the *IGHV* transcript requires analysis of clones with gel electrophoresis and 6 subsequent rounds of PCR amplification with consensus degenerate primers in the leader region^{10,40}. Although this method is relatively cheap and more directly presents the data based on the gel electrophoresis, there are also some shortcomings. Sanger sequencing does not provide insight into subclonal architecture and intraclonal diversity and cannot be combined with other assays in the same program. Even for

the target segment, the PCR-amplified transcripts are shorter, which leads to a high risk of inaccurate diagnosis of immunoglobulin and calculations of percent identity of the mutation. Finally, numerous manual manipulations are performed in the routine method. For example, the amplification products must then be extracted from the gel, purified and then sequenced, which can increase labor costs and the possibility of making mistakes. In practice, this method can also prove to be particularly difficult when several rearrangements are present. Therefore, this method is suitable for only a small number of cases.

Our method overcomes shortcomings. Because of the NGS workflow, 5' RACE makes it possible to expand the spectrum of utility of RNA sequencing experiments, opening the door to direct clinical application in CLL. With concordance levels greater than 90% concerning both the mutation status and the characterization of the V-D-J segments compared with routine methods, 5' RACE appears more stable and sensitive in handling multiple rearrangements and poor-quality RNA samples. In 5 discordant patients, four are caused by the poor quality of Sanger sequence and only 1 case is truly borderline case, which indicates that caution should be taken when reporting conclusions regarding the mutational status of *IGHV* in borderline cases with 97-98.9% homology. We also observed that different kinds of heavy and light chains were biased. The results of 5' RACE provide more detailed insight into subclonal architecture and intraclonal diversity. Given the increasing research on *IGH* in CLL, the need to detect more rearrangement is necessary⁴¹. In addition, this approach is extendable to immunoglobulin light chains as well as the T-cell receptor. Although the majority of studies investigating the prognostic significance of the BCR have focused exclusively on the heavy chain^{3,4}, emerging evidence indicates that the light chains might also play a crucial role in BCR function in CLL⁴². With the development of CLL studies, there are an increasing number of problematic cases¹⁵, and it is not easy to evaluate the treatment programs and prognosis of patients by simply determining the mutation status of the heavy chain. Through 5' RACE, we obtained more details of the Ig gene, which allowed us to better understand the value of the BCR pathway. When using NGS-based methods to determine *IGHV* gene SHM status, it is necessary to consider that this biomarker is not only useful for prognosis but also aids in guiding therapeutic decisions. All of the processes involved, including lab workflow, library preparation and bioinformatics data analyses, must be highly stable and reproducible. In our study, 5'RACE shows the stability and reproducibility in handling different timepoint samples of same patients, which represents a very reliable method for judging the treatment plan and patient prognosis.

Through 5' RACE, we found that most of the cases expressed identical mutation statuses of the heavy and light chains, except IGLV3-21. Interestingly, 92% (23/25) of IGLV3-21 light chains were unmutated, and the corresponding heavy chains were frequently mutated with 97-99% identity (16/25). Then, we analyzed the characteristics of IGLV3-21-positive cases, which were also predominantly associated with particular heavy chains. The IGHV3-21, IGHV3-48, and IGHV3-23 families represented 60% (15/25) of all IGHVs/IGLV3-21. This finding confirms that the BCR, including its light chain, is globally conserved and biased^{16,19-21}. Some reports found that gene expression profiles obtained from total RNA sequencing revealed genes that were differentially expressed between IgLV3-21-expressers and other patients⁴³⁻⁴⁵. With our RACE approach, we found that 94% (24/25) of IGLV3-21, which showed a similar CDR3 (QVWDS SSDHPWV), harbored a point mutation in the position between FR4 and the constant region termed R110. Combined with clinical data, we confirmed that IGLV3-21^{R110}-CLL presented a high-risk stage and worse prognosis than M-CLL non-IGLV3-21^{R110}, even though 75% of IGLV3-21^{R110}-CLL (18/24) cases harbored heavy chain mutations. It has been also reported that the allele IGLV3-21*01 is a risk factor for IGLV3-21^{R110} CLL. The crucial residues for the BCR-BCR interaction in CLL include residues R110 and K16 in one BCR and residues D50 and D52 of the YSD motif in a neighboring BCR⁴⁶. Recently, studies have also shown that IGLV3-21^{R110}-CLL, although composite in terms of IGHV mutational status, exhibits a similar clinical outcome to U-IGHV CLL^{22,24}. These findings suggest that the immunoglobulin light chain plays an important role in the clonal selection and somatic hypermutation (SHM) of Ig genes, and there is a greater need for improved methods for detecting these immunoglobulins.

There are also some limitations with 5' RACE that we need to mention. First, implementation of this high-throughput technology into routine clinical applications requires a network of laboratories with specialists that bring immune biological knowledge, technical experience with NGS methodology and immune informatics

expertise²⁵. Second, the data interpretation of NGS is more complex, and a standardized method does not currently exist. Thus, additional studies using NGS to define more accurately BCR features, especially in the light chain, are necessary.

In conclusion, the results suggest that 5' RACE could be used as an alternative to routine methods for the analysis of immunoglobulin rearrangements in research and clinical assessment of chronic lymphoid leukemia.

Acknowledgments:

This study was supported by a grant of the University of Rouen and INSERM.

Conflicts of Interest:

The authors declare no conflict of interest.

Appendix 1. RT Primers	
Primers RT	5'---3'
Primer_RT_IgM	AAGTGATGGAGTCGGGAAGG
Primer_RT_IgG	CGCCTGAGTTCCACGACA
Primer_RT_IgA	GGAAGAAGCCCTGGACCAG
Primer_RT_IgE	AGGTGGCATTGGAGGGAAT
Primer_RT_IgD	AGGGCTGTTATCCTTTGGGT
Primer_RT_IgK	CACACAACAGAGGCAGTTCC
Primer_RT_IgL123	TCAGAGGAGGGCGGGAAC
Primer_RT_IgL7	CCGGGTAGAAGTCACTTACGA

Appendix 2. Nested I2 Primers	
	Primers Nested
I2_Primer_Nested_IgM	AAAAGGGTTGGGGCGGAT
I2_Primer_Nested_IgG	GGAAGACCGATGGGCCCTTG
I2_Primer_Nested_IgA	GAGGCTCAGCGGGAAGAC
I2_Primer_Nested_IgE	AAGGGGAAGACGGATGGGCT
I2_Primer_Nested_IgD	ATGGGGAACACATCCGGAGC
I2_Primer_Nested_IgK	CAGATGGCGGGAAGATGAAG
I2_Primer_Nested_IgL1	AGTGACAGTGGGGTTGCCT
I2_Primer_Nested_IgL237	ACAGAGTGACCGAGGGGCA

Appendix 3. Primer of search pattern	
Primer	Search pattern
IgM	GAATTCTCACAGGAGACGAGGGGGAAAAGGGTTGGGGCGGATGCACTCC
IgG1234	GGAAGACCGATGGGCCCTTGGTGGAGG
IgA	GAGGCTCAGCGGGAAGACCTTGGGGCTGGTCGGGGATG

IgE	AAGGGGAAGACGGATGGGCTCTGTGTGGAGG
IgD	ATGGGGAACACATCCGGAGCCTTGGTGGGTG
IgK	CAGATGGCGGGAAGATGAAGACAGATGGTGCAGCCACAGTTC
IgL1	AGTGACAGTGGGGTTGGCCTTGGGCTGAC
IgL237	ACAGAGTGACCGAGGGGGCAGCCTTGGGCTGAC

Appendix 4. The results of patients tested by 5' RACE at different time.

No.	Date of sample	Delay between diagnosis and each evaluation	IGH					IGLK				
			Mut %	Type of IgH	V	D	J	Mut %	Type of IgL	V	J	
9	Diagnosis	-	100	M/D	IGHV3-33	IGHD3-10	IGHJ6	100	K	IGKV2-28	IGKJ3	
	Evaluation 1	3.03	100	M/D	IGHV3-33	IGHD3-10	IGHJ6	100	K	IGKV2-28	IGKJ3	
	Evaluation 2	5.26	100	M/D	IGHV3-33	IGHD3-10	IGHJ6	100	K	IGKV2-28	IGKJ3	
16	Diagnosis	-	99.37	M/D	IGHV3-15	IGHD3-16	IGHJ5	100	L	IgLV3-21	IgLJ3	
	Evaluation 1	10.89	99.34	M/D	IGHV3-15	IGHD3-16	IGHJ5	100	L	IgLV3-21	IgLJ3	
17	Diagnosis	-	98.56	M/D	IGHV3-15	IGHD2-21	IGHJ4	99.14	L	IgLV3-21	IgLJ1	
	Evaluation 1	2.12	98.56	M/D	IGHV3-15	IGHD2-21	IGHJ4	99.14	L	IgLV3-21	IgLJ1	
	Evaluation 2	3.74	98.56	M/D	IGHV3-15	IGHD2-21	IGHJ4	99.14	L	IgLV3-21	IgLJ1	
19	Diagnosis	-	98.58	M/D	IGHV3-33	IGHD3-3	IGHJ6	99.13	L	IgLV3-1	IgLJ2	
	Evaluation 1	3.32	98.58	M/D	IGHV3-33	IGHD3-3	IGHJ6	99.13	L	IgLV3-1	IgLJ2	
22	Diagnosis	-	99.15	M/D	IGHV4-59	IGHD2-8	IGHJ4	100	L	IgLV3-21	IgLJ1	
	Evaluation 1	7.09	99.15	M/D	IGHV4-59	IGHD2-8	IGHJ4	100	L	IgLV3-21	IgLJ1	
34	Diagnosis	-	97.46	M/D	IGHV4-34	IGHD4-17	IGHJ6	94.98	K	IGKV1-5	IGKJ2	
	Evaluation 1	7.52	96.18	M/D	IGHV4-34	IGHD4-17	IGHJ6	95.81	K	IGKV1-5	IGKJ2	
42	Diagnosis	-	97.55	M/D	IGHV3-48	IGHD3-10	IGHJ6	97.85	L	IgLV3-21	IgLJ1	
	Evaluation 1	1.38	97.55	M/D	IGHV3-48	IGHD3-10	IGHJ6	97.85	L	IgLV3-21	IgLJ1	
52	Diagnosis	-	97.47	M/D	IGHV3-15	IGHD4-17	IGHJ4	99.31	L	IgLV3-21	IgLJ3	
	Evaluation 1	3.6	97.18	M/D	IGHV3-15	IGHD4-17	IGHJ4	99.57	L	IgLV3-21	IgLJ3	
59	Diagnosis	-	97.06	M/D	IGHV3-48	IGHD2-15	IGHJ4	99.57	L	IgLV3-21	IgLJ3	
	Evaluation 1	2.29	97.06	M/D	IGHV3-48	IGHD2-15	IGHJ4	99.57	L	IgLV3-21	IgLJ3	
60	Diagnosis	-	96.85	M/D	IGHV3-53	IGHD1-26	IGHJ4	99.13	L	IgLV3-21	IgLJ3	
	Evaluation 1	1.54	96.64	M/D	IGHV3-53	IGHD1-26	IGHJ4	99.31	L	IgLV3-21	IgLJ3	
61	Diagnosis	-	94.71	M/D	IGHV3-21	NA	IGHJ6	99.44	L	IgLV3-21	IgLJ3	
	Evaluation 1	0.64	94.67	M/D	IGHV3-21	NA	IGHJ6	99.57	L	IgLV3-21	IgLJ3	
62	Diagnosis	-	97.58	M/D	IGHV3-74	IGHD5-18	IGHJ4	99.56	L	IgLV3-21	IgLJ1	
	Evaluation 1	2.8	97.1	M/D	IGHV3-74	IGHD5-18	IGHJ4	99.56	L	IgLV3-21	IgLJ1	
64	Diagnosis	-	97.27	G	IGHV3-13 ⁰	IGHD2-2	IGHJ4	99.57	L	IgLV3-21	IgLJ3	
			97.91	M/D	IGHV3-48	IGHD4-17	IGHJ4					

Evaluation 1	2.01	97.61	G	IGHV3-13 ^o	IGHD2-2	IGHJ4	99.31	L	IGLV3-21	IGLJ3
		97.33	M/D	IGHV3-48	IGHD4-17	IGHJ4				
Diagnosis	-	95.43	M/D	IGHV3-74	IGHD3-10	IGHJ3	98.67	L	IGLV3-21	IGLJ1
		99.34	M/D	IGHV6-1 ^o	IGHD6-6	IGHJ4				
Evaluation 1	9.18	96.73	M/D	IGHV3-74	IGHD3-10	IGHJ3	98.72	L	IGLV3-21	IGLJ1
		98.94	M/D	IGHV6-1 ^o	IGHD6-6	IGHJ4				

* Unproductive arrangement

Appendix 5. Clinical information of CLL patients							
	IgG-CLL N=13	IgM-CLL N=67	p	IGLV3-21 ^{R110} -CLL N=24	M-CLL (non-IGLV3-21 ^{R110}) N=33	U-CLL (non-IGLV3-21 ^{R110}) N=23	p
Age at diagnosis (years) mean ± SD	73.4±10.8	69.8±11.0	0.28	68.5±11.8	71.0±10.7	71.4±10.7	0.62
Gender (F:M)	4 : 9	16 : 51	0.86	3 : 21	13 : 20	4 : 19	0.042
Binet-no. (%)			0.90				0.013
A	6 (46.1%)	38 (56.7%)		9 (37.5%)	24 (72.7%)	11 (47.8%)	
B	3 (23.1%)	18 (26.9%)		7 (29.2%)	4 (12.1%)	10 (43.5%)	
C	2 (15.4%)	8 (11.9%)		6 (25.0%)	2 (6.1%)	2 (8.7%)	
unknown	2 (15.4%)	3 (4.5%)		2 (8.3%)	3 (9.1%)	0 (0.0)	
FISH aberrations, no. (%)							
Del 11q12 (ATM)	2 (15.4%)	6 (9.0%)	0.61	2 (8.3%)	1 (3.0%)	5 (21.7%)	0.17
Del 13q14 (D-LEU2)	5 (38.5%)	31 (46.3%)	0.83	13 (54.2%)	13 (39.4%)	10 (43.5%)	0.39
Trisomy 12 (MDM2)	2 (15.4%)	6 (9.0%)	0.61	1 (4.2%)	2 (6.1%)	5 (21.7%)	0.19
Del 17q13 (TP53)	0 (0.0)	2 (3.0%)	1.00	0 (0.0)	0 (0.0)	2 (8.7%)	0.18
Del 17q13 (TP53 exon 1)	0 (0.0)	1 (1.5%)	1.00	0 (0.0)	0 (0.0%)	1 (4.3%)	0.61
Trisomy 12 (POU6F1)	2 (15.4%)	4 (6.0%)	0.17	1 (4.2%)	2 (6.1%)	3 (13.0%)	0.66
Del 6q21 (SEC63)	2 (15.4%)	2 (3.0%)	0.08	1 (4.2%)	2 (6.1%)	1 (4.3%)	0.68
unknown	4 (30.8%)	7 (10.4%)		4 (16.7%)	6 (18.2%)	1 (4.3%)	
TP53 gene, no. (%)	1 (7.7)	4 (6.0)	0.53	0 (0.0)	2 (6.1%)	3 (13.0%)	0.22
Karyotype no. (%)							
Del11	2 (15.4%)	4 (6.0%)	0.28	2 (8.3%)	1 (3.0%)	3 (13.0%)	0.86
Tri12	3 (23.1%)	8 (12.0%)	0.61	0 (0.0)	4 (12.1%)	7 (30.4%)	0.011
Del13	2 (15.4%)	10 (15.0%)	1.00	6 (25.0%)	4 (12.1%)	2 (8.7%)	0.35
Del6	2 (15.4%)	2 (3.0%)	0.14	1 (4.2%)	1 (3.0%)	2 (8.7%)	1.0
unknown	2 (15.4%)	15 (22.4%)		3 (12.5%)	12 (36.4%)	2 (8.7%)	
SHM IGH (M:U)	9 : 4	42 : 25	0.98	18 : 6	33 : 0	0 : 23	
Same SHM IGH and IgL no. (%)	13 (100)	46 (68.7)	0.045	8 (33.3%)	29 (87.9%)	22 (95.7%)	<0.0001
Laboratory parameters at diagnosis							
Platelets (150-450 Giga/L) mean ± SD	276.0±49.8	218.7±49.8	0.002	215.0±56.2	212.8±67.7	252.2±97.3	0.31
Median (min, max)	271.0 (219.0, 330.0)	225.0 (80.0, 438.0)		212.0 (80.0, 320.0)	233.0 (85.0, 330.0)	250.0 (120.0, 438.0)	
LDH (240-480 UI/L) mean ± SD	415.4±67.0	367.3±167.2	0.013	335.9±90.1 325.0	396.2±233.6	388.0±106.0	0.18
Median (min, max)	387.0 (349.0, 523.0)	353.0 (125.0, 1288.0)		(177.0, 562.0)	366.0 (125.0, 1288.0)	376.0 (233.0, 670.0)	
sIgM (0.4-2.3 g/L) mean ± SD	0.56±0.27	0.52±0.32	0.71	0.38±0.30	0.60±0.31	0.61±0.35	0.025
Median (min, max)	0.5 (0.3, 1.0)	0.45 (0.1, 1.4)		0.3 (0.1, 0.9)	0.60 (0.1, 1.2)	0.60 (0.2, 1.4)	
Beta-2 microglobulin (0.8-2.4 mg/L) mean ± SD	2.1±0.64	2.61±1.13	0.53	2.63±1.07	2.03±0.43	3.16±1.43	0.92
Median (min, max)	1.9 (1.3, 3.0)	2.3 (1.4, 6.4)		2.4 (1.4, 5.0)	2.1 (1.3, 3.0)	2.9 (1.5, 6.4)	

Legend: two comparisons were performed: firstly, between the IgG-CLL group vs. the IgM-CLL group and secondly, between the three following groups: IGLV3-21^{R110}-CLL, M-CLL (non-IGLV3-21^{R110}) and U-CLL (non-IGLV3-21^{R110})

References

1. Linet MS, Schubauer-Berigan MK, Weisenburger DD, et al. Chronic lymphocytic leukaemia: an overview of aetiology in light of recent developments in classification and pathogenesis. *Br J Haematol.* 2007;139(5):672-686. doi:10.1111/j.1365-2141.2007.06847.x

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66(1):7-30. doi:10.3322/caac.21332
3. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood.* 1999;94(6):1848-1854.
4. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood.* 1999;94(6):1840-1847.
5. Delgado J, Doubek M, Baumann T, et al. Chronic lymphocytic leukemia: A prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *Am J Hematol.* 2017;92(4):375-380. doi:10.1002/ajh.24660
6. Thunberg U, Johnson A, Roos G, et al. CD38 expression is a poor predictor for VH gene mutational status and prognosis in chronic lymphocytic leukemia. *Blood.* 2001;97(6):1892-1894. doi:10.1182/blood.v97.6.1892
7. Maloum K, Davi F, Merle-Béral H, et al. Expression of unmutated VH genes is a detrimental prognostic factor in chronic lymphocytic leukemia. *Blood.* 2000;96(1):377-379.
8. Oscier DG, Gardiner AC, Mould SJ, et al. Multivariate analysis of prognostic factors in CLL: clinical stage, IGHV gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood.* 2002;100(4):1177-1184.
9. Kröber A, Seiler T, Benner A, et al. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood.* 2002;100(4):1410-1416.
10. van Dongen JJM, Langerak AW, Brüggemann 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(12):2257-2317. doi:10.1038/sj.leu.2403202
11. Davis Z, Forconi F, Parker A, et al. The outcome of Chronic lymphocytic leukaemia patients with 97% IGHV gene identity to germline is distinct from cases with <97% identity and similar to those with 98% identity. *Br J Haematol.* 2016;173(1):127-136. doi:10.1111/bjh.13940
12. Gocke CD, Gladstone DE. The absolute percent deviation of IGHV mutation rather than a 98% cut-off predicts survival of chronic lymphocytic leukaemia patients treated with fludarabine, cyclophosphamide and rituximab. *Br J Haematol.* 2018;180(1):7-8. doi:10.1111/bjh.15015
13. Morabito F, Shanafelt TD, Gentile M, et al. Immunoglobulin heavy chain variable region gene and prediction of time to first treatment in patients with chronic lymphocytic leukemia: Mutational load or mutational status? Analysis of 1003 cases. *Am J Hematol.* 2018;93(9):E216-E219. doi:10.1002/ajh.25206
14. Rosenquist R, Ghia P, Hadzidimitriou A, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia.* 2017;31(7):1477-1481. doi:10.1038/leu.2017.125
15. Langerak AW, Davi F, Ghia P, et al. Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. *Leukemia.* 2011;25(6):979-984. doi:10.1038/leu.2011.49
16. Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest.* 1998;102(8):1515-1525. doi:10.1172/JCI3009

17. Tobin G, Thunberg U, Karlsson K, et al. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood*. 2004;104(9):2879-2885. doi:10.1182/blood-2004-01-0132
18. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood*. 2007;109(1):259-270. doi:10.1182/blood-2006-03-012948
19. Agathangelidis A, Darzentas N, Hadzidimitriou A, et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood*. 2012;119(19):4467-4475. doi:10.1182/blood-2011-11-393694
20. Stamatopoulos K, Belessi C, Hadzidimitriou A, et al. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood*. 2005;106(10):3575-3583. doi:10.1182/blood-2005-04-1511
21. Shanafelt TD, Wang XV, Kay NE, et al. Ibrutinib-Rituximab or Chemoimmunotherapy for Chronic Lymphocytic Leukemia. *N Engl J Med*. 2019;381(5):432-443. doi:10.1056/NEJMoa1817073
22. Maity PC, Bilal M, Koning MT, et al. IGLV3-21*01 is an inherited risk factor for CLL through the acquisition of a single-point mutation enabling autonomous BCR signaling. *Proc Natl Acad Sci U S A*. 2020;117(8):4320-4327. doi:10.1073/pnas.1913810117
23. Stamatopoulos B, Smith T, Crompton E, et al. The Light Chain IgLV3-21 Defines a New Poor Prognostic Subgroup in Chronic Lymphocytic Leukemia: Results of a Multicenter Study. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2018;24(20):5048-5057. doi:10.1158/1078-0432.CCR-18-0133
24. Nadeu F, Royo R, Clot G, et al. IGLV3-21R110 identifies an aggressive biological subtype of chronic lymphocytic leukemia with intermediate epigenetics. *Blood*. 2021;137(21):2935-2946. doi:10.1182/blood.2020008311
25. Davi F, Langerak AW, de Septenville AL, et al. Immunoglobulin gene analysis in chronic lymphocytic leukemia in the era of next generation sequencing. *Leukemia*. 2020;34(10):2545-2551. doi:10.1038/s41375-020-0923-9
26. Calis JJA, Rosenberg BR. Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends Immunol*. 2014;35(12):581-590. doi:10.1016/j.it.2014.09.004
27. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131(25):2745-2760. doi:10.1182/blood-2017-09-806398
28. Aho AV, Corasick MJ. Efficient string matching: an aid to bibliographic search. *Commun ACM*. 1975;18(6):333-340. doi:10.1145/360825.360855
29. IgBLAST: an immunoglobulin variable domain sequence analysis tool | Nucleic Acids Research | Oxford Academic. Accessed June 20, 2022. <https://academic.oup.com/nar/article/41/W1/W34/1097536>
30. Lefranc MP, Giudicelli V, Kaas Q, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res*. 2005;33(Database issue):D593-597. doi:10.1093/nar/gki065
31. Li Y ze, Xu J jie, Qian H zhu, et al. High prevalence of HIV infection and unprotected anal intercourse among older men who have sex with men in China: a systematic review and meta-analysis. *BMC Infect Dis*. 2014;14:531. doi:10.1186/1471-2334-14-531
32. Zheng X, Yu S, Long J, et al. Comparison of the clinical characteristics of primary thyroid lymphoma and diffuse sclerosing variant of papillary thyroid carcinoma. *Endocr Connect*. 2022;11(1):e210364. doi:10.1530/EC-21-0364

33. Miranda-Filho A, Piñeros M, Ferlay J, Soerjomataram I, Monnereau A, Bray F. Epidemiological patterns of leukaemia in 184 countries: a population-based study. *Lancet Haematol*. 2018;5(1):e14-e24. doi:10.1016/S2352-3026(17)30232-6
34. Visentin A, Bonaldi L, Rigolin GM, et al. The combination of complex karyotype subtypes and IGHV mutational status identifies new prognostic and predictive groups in chronic lymphocytic leukaemia. *Br J Cancer*. 2019;121(2):150-156. doi:10.1038/s41416-019-0502-x
35. Hallek M, Shanafelt TD, Eichhorst B. Chronic lymphocytic leukaemia. *Lancet Lond Engl*. 2018;391(10129):1524-1537. doi:10.1016/S0140-6736(18)30422-7
36. Koehrer S, Burger JA. B-cell receptor signaling in chronic lymphocytic leukemia and other B-cell malignancies. *Clin Adv Hematol Oncol HO*. 2016;14(1):55-65.
37. Vardi A, Agathangelidis A, Sutton LA, Ghia P, Rosenquist R, Stamatopoulos K. Immunogenetic studies of chronic lymphocytic leukemia: revelations and speculations about ontogeny and clinical evolution. *Cancer Res*. 2014;74(16):4211-4216. doi:10.1158/0008-5472.CAN-14-0630
38. Tobin G. The immunoglobulin genes: structure and specificity in chronic lymphocytic leukemia. *Leuk Lymphoma*. 2007;48(6):1081-1086. doi:10.1080/10428190701342034
39. Dighiero G. CLL Biology and Prognosis. *Hematology*. 2005;2005(1):278-284. doi:10.1182/asheducation-2005.1.278
40. Ghia P, Stamatopoulos K, Belessi C, et al. ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia*. 2007;21(1):1-3. doi:10.1038/sj.leu.2404457
41. Stamatopoulos B, Timbs A, Bruce D, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia*. 2017;31(4):837-845. doi:10.1038/leu.2016.307
42. Hadzidimitriou A, Darzentas N, Murray F, et al. Evidence for the significant role of immunoglobulin light chains in antigen recognition and selection in chronic lymphocytic leukemia. *Blood*. 2009;113(2):403-411. doi:10.1182/blood-2008-07-166868
43. Yeomans A, Thirdborough SM, Valle-Argos B, et al. Engagement of the B-cell receptor of chronic lymphocytic leukemia cells drives global and MYC-specific mRNA translation. *Blood*. 2016;127(4):449-457. doi:10.1182/blood-2015-07-660969
44. Pozzo F, Bittolo T, Vendramini E, et al. NOTCH1-mutated chronic lymphocytic leukemia cells are characterized by a MYC-related overexpression of nucleophosmin 1 and ribosome-associated components. *Leukemia*. 2017;31(11):2407-2415. doi:10.1038/leu.2017.90
45. Wang L, Brooks AN, Fan J, et al. Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia. *Cancer Cell*. 2016;30(5):750-763. doi:10.1016/j.ccell.2016.10.005
46. Minici C, Gounari M, Übelhart R, et al. Distinct homotypic B-cell receptor interactions shape the outcome of chronic lymphocytic leukaemia. *Nat Commun*. 2017;8:15746. doi:10.1038/ncomms15746

Highlights

- The mutational status of *IGHV* is a crucial biomarker for CLL patients. We developed a 5'RACE NGS method, applicable in routine to determine IGHV mutational status and compare it to standard Biomed-

21

2 procedure. The agreement rate between the two methods was 93.8%. 5' RACE assay was more sensitive to identify unproductive and multiple rearrangements and provides more details regarding subclonal architecture and intraclonal diversity.

- The 5' RACE assay can also be used to simultaneously sequence light chains variables regions (*IGLV*). In most CLL cases, the mutational status of heavy and light chains is concordant but may guide therapeutic strategy and prognosis in borderline cases.
- By this approach, we identified a molecular subgroup characterized by *IGLV3-21* rearrangement harboring a single point mutation *IGLV3-21*^{R110} and confirm its unfavorable prognosis impact, emphasizing the importance to identify the *IGVL* chain in addition to the *IGHV* status.

Comprehensive analysis of the association of the expression of *TP53* mRNA by RT-MLPA high-throughput sequence with multiple genetic variants in chronic lymphocytic leukaemia

Xuan Lan^{1,2}, Philippe Ruminy¹, Elodie Bohers¹, Pascaline Etancelin, Vinciane Marchand¹, Mathieu Viennot¹, Pierre-Julien Viailly¹, Florian Bouclet, Stéphane Leprêtre, Sorina Mihailescu, Hervé Tilly¹ and Fabrice Jardin¹

1 INSERM U1245, Team Genetic and biomarkers in lymphomas and solid tumors, Centre Henri Becquerel, University of Normandie; Rouen, France

2. Department of molecular biology, Centre Henri Becquerel, Rouen, France

3. Department of clinical Hematology, Centre Henri Becquerel, Rouen, France

4. Clinical Research Unit, Centre Henri Becquerel, Rouen France

Abstract

Background: Chronic lymphocytic leukemia (CLL) is the most frequent leukemia in adults. *TP53*, mutated in 10% of cases, is one of the most crucial prognosis factors and treatment guides. Several p53 isoforms have been described in both physiological and pathological situations. Their clinical and biological relevance in CLL is still unclear.

Aim: To evaluate the interplay between *TP53* mRNA expressions and *TP53* genetic variants, including mutations and splice variants, and their relationship with other genetic factors in CLL collected in routine practice.

Methods: We collected clinical data and RNA samples from 114 CLL patients from our institution between 2002 and 2021. Expression of *TP53* mRNA was detected by Reverse-transcriptase multiplex ligation-dependent probe amplification assay (RT-MLPA) with high-throughput sequencing to determine a single and routinely applicable *TP53* mutation status, mRNA expression, and detect *TP53* splice variants.

Results: Unmutated heavy chain of immunoglobulin genes (*IgHV*), deletions of the short arm of chromosome 17 (del17p13), rs2909430, and lower level of hemoglobin were highly correlated with the *TP53* mutational status ($p=0.005$, $p<0.0001$, $p=0.036$ and 0.001 , respectively). Patients with *TP53* aberration and *SF3B1* mutations expressed a lower level of *TP53* exon4-5 and showed a shorter overall survival (OS) compared with wildtype ($p<0.0001$). Unmutated *IgHV* and del(6) showed higher expression of exon9-10 and lower expression of exon9-9b and exon9bg-10 and predicted lower OS and time to first treatment (TTFT) ($p=0.002$ and $p=0.015$, respectively). Finally, Combined unmutated *IgHV*, *TP53* aberration, *NOTCH1*mut, *SF3B1*mut, and deletion 6 indicated a layered model of OS, TTFT and correlated the expression of exon5-6, exon9-9b, and exon9bg-10.

Conclusion: The high-throughput RT-MLPA sequence technology is a suitable and reliable method for assessing multiple molecular parameters targeting the *TP53* gene. The different expression level of several exon junctions, which highlights the existence of various p53 isoforms, proved their interest in predicting the treatment and prognosis in CLL.

Keywords: chronic lymphocytic leukemia, RT-MPLA high-throughput sequencing, biomarkers, *TP53* gene, alternative splicing.

Introduction

CLL is the most frequent adult leukemia in the western world, especially in the elderly population¹, and has a highly heterogeneous clinical course. In the last 30 years, various biological and genetic markers have been reported, which expand novel therapeutic strategies and valuable prognostic landscape for the CLL²⁻⁶. Specific inhibitors were lately identified and approved as critical pathways of CLL cell survival and apoptosis while replacing standard chemoimmunotherapy in the first and second-line indications⁷⁻¹². Despite this progress, CLL remains incurable. Either the patient shares multiple gene mutation¹³, or the disease transforms into diffuse large B-cell lymphoma (DLBCL)^{14,15}, both situations associated with poor prognosis.

Therefore, understanding biological and genetic predictors of response to therapy and disease prognosis become crucial issues. The mutational status of *IgHV* is a powerful and stable predictor of treatment and OS in patients treated by chemoimmunotherapy (CIT)^{16,17}. Another increasingly significant predictor is tumor suppressor protein P53, responsible for the cell cycle and apoptosis. Although the rate of *TP53* mutation is around 10% in CLL, *TP53* aberrations, including mutations of the gene, especially in the DNA binding domain, and deletion of chromosome 17p, are still significant predictors of resistance to fludarabine-based chemotherapy and adverse prognostic factor in CLL^{18,19}. Therefore, detecting *IgHV* and *TP53* status is necessary to guide therapeutic strategies.

Deletion of 17p can be detected by fluorescence in situ hybridization (FISH). Next-generation sequencing (NGS) was defined as the gold standard for detecting variant allele frequencies (VAFs) of *TP53* mutation²⁰. Recent research also found that low-burden *TP53* mutation with VAFs between 0.1%-10% shows a significantly poor prognosis impact on OS²¹. Recently, studies highlighted the importance of p53 isoforms, especially the N-terminally truncated ones such as $\Delta 40p53$, $\Delta 133p53$, and $\Delta 160p53$, suggesting their involvement in carcinogenesis²². Due to the lack of method standardization and structural characterization, the clinical and biological relevance of the p53 isoforms is still unclear. The European Research Infrastructure Consortium (ERIC) has published guidelines recommending combining multiple methods for detecting *TP53* mutations²³.

Moreover, other genetic factors were necessary for CLL's prognosis and treatment decision. In the context of randomized, prospective clinical studies, CLL4 and CLL8 trials showed that *SF3B1* and *NOTCH1* mutations were predictive therapeutic markers^{19,24}. *NOTCH1* and *SF3B1* mutations are detected in 5% to 20% of patients, both occurring more frequently with unmutated *IgHV*. *NOTCH1* is a transmembrane receptor that plays a crucial role in cell differentiation and proliferation, and *NOTCH1* mutations can lead to oncogenic pathway activation in various malignancies. *SF3B1* mutations affect the splicing machinery as a novel pathogenic mechanism.

In this work, we described a new method that combined Reverse-transcriptase multiplex ligation-dependent probe amplification (RT-MLPA) with high-throughput sequencing technology to quantify the expression of different exon junctions of *TP53* in CLL patients and correlate these new molecular features to other CLL routine markers including NGS sequencing data, FISH, and CNV (copy number variations)

Material and Methods

Patients

We retrospectively collected data from previously 114 patients including 27 untreated and 87 treated patients, diagnosed with CLL from a single institution (Centre Henri Becquerel, Rouen, France) between January 1, 2002, and July 1, 2021. The diagnosis of CLL was determined by cytometry, according to the criteria of the recently updated WHO diagnosis guideline²⁵. This study was performed by the Declaration of Helsinki and local laws, and the Institutional Review Board approved the protocol of the Henri Becquerel Centre. Biological material was obtained after receiving patient consent. Inclusion criteria were available clinical data, follow-up, and tumor RNA at the time of diagnosis. Mutations of *TP53* were identified through NGS sequencing (covering exons 2 to 11) of DNA samples, allowing the detection of low-frequency variants down to 1% VAF.

RNA preparation

Mononuclear cells were isolated by density gradient centrifugation (lymphoprep, 15 min at 2000 g). RNA was isolated from peripheral blood mononuclear cells (PBMCs) using the RNA NOW kit (Biogenex, Seabrook, TX, USA) according to the instruction manual and stored at -80°C in water. RNA concentrations were determined using a nanodrop spectrophotometer.

QMPSF assays

CNV of targeted genes was determined using QMPSF (quantitative Multiplex PCR of Short Fluorescent Fragments) assays as previously reported^{26,27}. Briefly, seven genes located in the commonly deleted regions of chromosomes 6, 11, 13, 17 and the more frequently gained regions in trisomy 12 were selected as target genes, including *SEC63* at 6q21, *ATM* at 11q22, *DLEU2* at 13q14, and *TP53* at 17p13 and 17p13.1, *POU6F1* at 12q13 and *MDM2* at 12q15.

RT-MLPA high-throughput sequencing

The probes of MLPA were designed according to the procedure named "Designing synthetic MLPA probes" from MRC-Holland except for PCR primer sequences²⁸. MLPA probe pair for each *TP53* exon junction, including Left Probe Oligonucleotide (LPO) and Right Probe Oligonucleotide (RPO) (as presented in Appendix 1). To prepare sequencing libraries for NGS, we added forward/reverse primer binding sites as 5' universal adaptor sequences (5'GTGCCAGCAAGATCCAATCTAGA3') on LPO and 3' universal adaptor sequences (5'TCCAACCCTTAGGGAACCC3') on RPO, named Left Hybridizing Sequence (LHS) and Right Hybridizing Sequence (RHS) as sequencing primers. To eliminate PCR bias, we added a 7 N random sequence at the 5' end of LHS as Unique Molecular Identifier (UMI). As illustrated in Figure.2B, LHS and RHS are directly adjacent and exactly reverse-complementary to cDNA from the target RNA.

The RNA obtained from samples was reverse transcribed into cDNA, hybridized with LPOs and RPOs for each exon, and then ligated. The ligated probes were purified by Ampure magnetic beads and used as PCR templates to prepare libraries using P5 and barcode P7 adaptor primers. Different samples constituted libraries were pooled into a single sequencing assay, sequenced on an Illumina MiSeq System and then demultiplexed according to the barcode contained in the P7 adaptor primers.

Sequencing data were aligned to the hybridizing sequences of all gene probes and analyzed to calculate the number of sequence reads that contained both LHS and RHS. Relative expression levels of each exon junction were compared with the ratio of corresponding reads among the total reads of *TP53*.

Statistical analysis

The statistical analysis was performed using SPSS v17.0²⁹ and GraphPad Prism 5.0 software³⁰. Comparisons between variables were assessed by the Chi-square test (and Yates' correction), Fisher's exact test for qualitative variables, and t-tests or non-parametric Mann–Whitney test for quantitative data. A stepwise method developed a multiple linear regression model to describe the linear association between the expressions of each exon junction. Correlation between the variant allele ratio of p53R72P obtained by NGS and RT-MLPA high-throughput sequencing was established using Spearman's correlation test. Overall survival (OS) was measured from the time of diagnosis to the date of death from any cause. Time-to-first-treatment (TTFT) was calculated from the time of diagnosis to the date of the first treatment initiation. Survival probabilities were calculated using the Kaplan–Meier method. Log-rank tests were performed to evaluate the impact of mutational status on survival. A two-tailed p-value below 0.05 was considered statistically significant.

Results

1. Patient characteristics

This study included 114 CLL patients with their clinical data summarised in **Table 1**. The male-to-female ratio was 1.95 (76:39), and the mean age was 71.7 (range from 32 to 100 years old). The information of each gene was described in the figure. 1A. We found that unmutated *IgHV*, deletions of the short arm of chromosome 17 (del17p13) correlated with *TP53*, and *TP53* exon1 were highly correlated with the mutated *TP53* ($p=0.005$, $p<0.0001$ and $p<0.0001$, respectively). Regarding the results of NGS, the most frequent single nucleotide polymorphisms (SNP) were rs1042522 (101, 88.6%), rs162895 (18, 15.8%) and rs2909430 (18, 15.8%). The last two dbSNP were systematically present in the same patients and showed similar mutational frequency and a close correlation with *TP53* mutation ($p=0.001$). Patients with mutated *TP53* were more likely to have a lower hemoglobin level (mean of 12.1 g/dL with mutated *TP53*) than a mean of 13.2 g/dL with wide type *TP53*, $p=0.04$).

Variable	<i>TP53</i> MUT (n=18)	<i>TP53</i> WT (n=96)	P-value
Gender			0.58
Male	12 (66.7%)	63 (65.6%)	
Female	6 (33.3%)	33 (34.4%)	
Age at diagnosis (years) Median (min; max)	72.2 (56-100)	71.3 (32-94)	0.75
Accept treatment	10	77	0.051
Binet stage			0.87
A	11 (61.1%)	60 (62.5%)	
B	3 (16.7%)	19 (19.8%)	
C	4 (22.2%)	17 (17.7%)	
Laboratory parameters at diagnosis			
Platelets (150-450 Giga/L) mean \pm SD	227.7 \pm 88.8	201.1 \pm 77.0	0.25
Median (min, max)	216.0 (101.0, 292.0)	199.5 (2.0, 409.0)	
Haemoglobin (13.0-17.0 g/dL) mean \pm SD	12.1 \pm 2.5	13.2 \pm 2.1	0.04
Median (min, max)	12.6 (6.7, 15.8)	13.4 (6.7, 17.3)	
White blood cells (4.0-11.0 Giga/L) mean \pm SD	58.1 \pm 109.4	50.2 \pm 117.8	0.21
Median (min, max)	25.4 (6.6, 473.1)	18.8 (4.4, 996.3)	
LDH (240-480 UI/l) mean \pm SD	391.2 \pm 145/9	343.0 \pm 118.9	0.12
Median (min, max)	354.0 (117.0, 764.0)	339.0 (118.0, 650.0)	
Protein (66-87 g/L) mean \pm SD	68.2 \pm 6.3	67.5 \pm 4.3	0.59
Median (min, max)	68.0 (57.0, 79.0)	68.0 (56.0, 79.0)	
Albumin (40-49 g/L) mean \pm SD	41.1 \pm 5.0	41.5 \pm 3.8	0.97
Median (min, max)	42.0 (30.0, 51.0)	42.0 (26.0, 50.0)	
Beta-2 microglobulin (0.8-2.4 mg/L) mean \pm SD	3.4 \pm 2.4	2.9 \pm 1.6	0.16
Median (min, max)	2.7 (1.0, 10.8)	2.5 (1.0, 8.0)	
Karyotype	n=16	n=90	
Deletion11	4 (25.0%)	17 (18.9%)	0.82
Trisomy12	1 (6.3%)	21 (23/3%)	0.22
Deletion13	3 (18.8%)	11 (12.2%)	0.76
Deletion6	2 (12.5%)	6 (6.7%)	0.76
Trisomy19	0 (0.0%)	3 (3.3%)	1.00
QMPSF	n=18	n=95	
<i>ATM</i> (del11q12)	4 (22.2%)	17 (17.9%)	0.92
<i>D-LEU2</i> (del13q14)	10 (55.5%)	35 (36.8%)	0.14
<i>MDM2</i> (trisomy12)	2 (11.1%)	11 (11.6%)	1.00
<i>TP53</i> (del17p13)	6 (33.3%)	1 (1.1%)	<0.0001
<i>TP53</i> exon1 (del17p13)	6 (33.3%)	1 (1.1%)	<0.0001
<i>POU6F1</i> (trisomy12)	1 (5.6%)	12 (12.6%)	0.73
<i>SEC63</i> (del6q12)	1 (5.6%)	4 (4.2%)	0.59
<i>IgHV</i> Unmut (98% cut-off)	13 (92.9%, n=14)	44 (55.0%, n=80)	0.006

NGS dbSNP	n=18	n=96	
RS1042522	17 (94.4%)	84 (87.5%)	0.66
RS2909430	8 (44.4%)	10 (10.4%)	0.001
RS1625895	8 (44.4%)	10 (10.4%)	0.001
RS1800372	3 (16.6%)	4 (4.2%)	0.14
<i>SF3B1</i>	2 (11.1%)	6 (6.3%)	0.81
<i>NOTCH1</i>	2 (11.1%)	18 (18.8%)	0.66

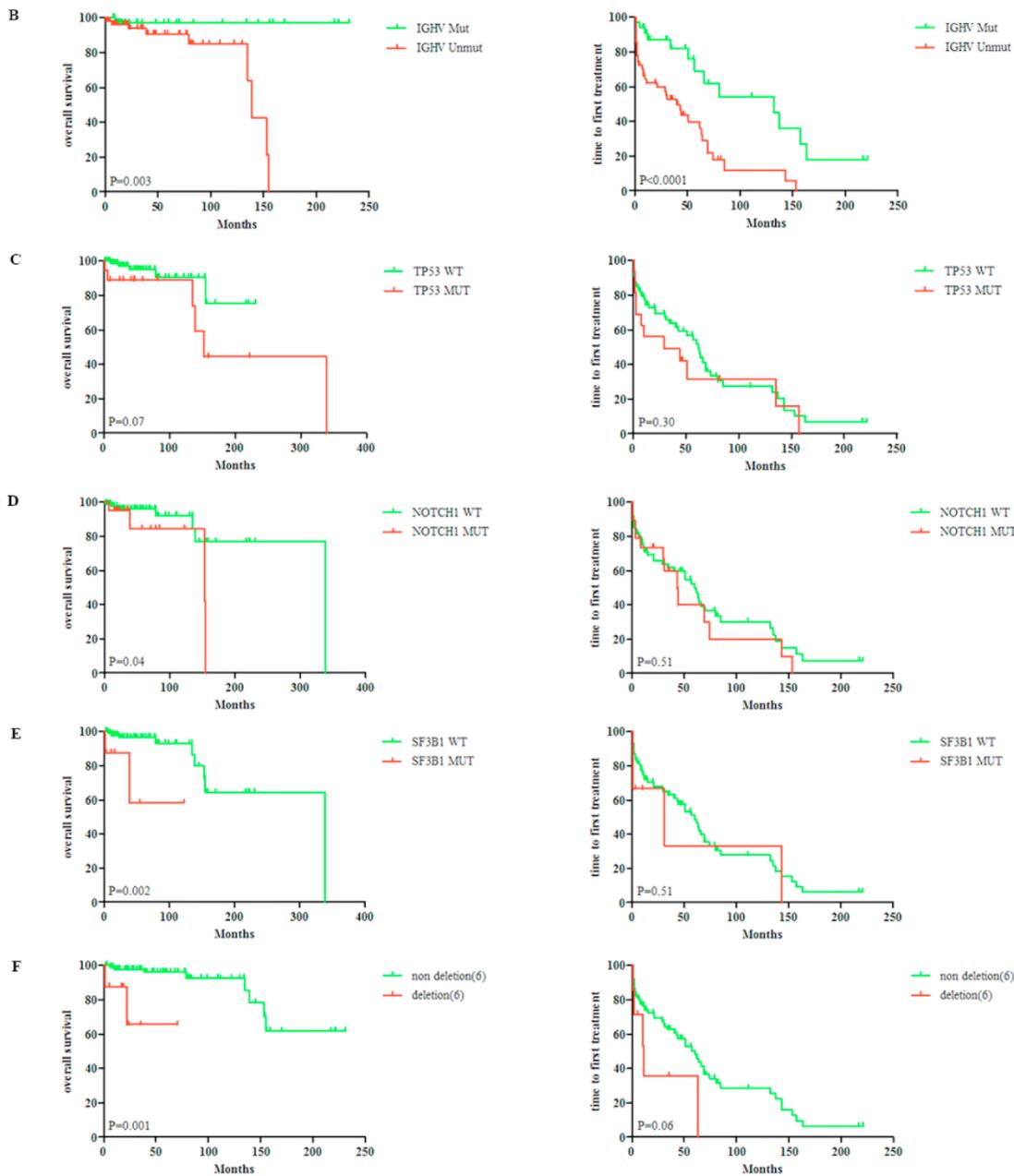
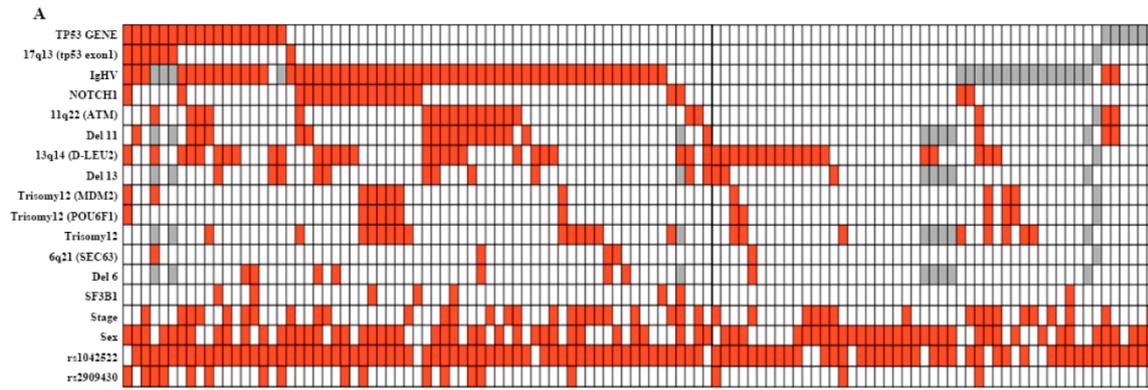


Figure 1 Characteristic of each critical genetic marker in CLL. (A) Summary of different genetic and clinical factors in the 114 CLL patients. Red presents unmutated *IgHV* and mutation for other genes, white presents mutated *IgHV* and wildtype for other genes, and grey presents not evaluated items. (B) The curve of OS for *IgHV* mutated (green) vs unmutated (red) ($p=0.003$) and the curve of TTFT for *IgHV* mutated (green) vs unmutated (red) ($p<0.0001$). (C) The curve of OS for *TP53* wildtype (green) vs mutated (red) ($p=0.07$) and the curve of TTFT for *TP53* wildtype (green) vs mutated (red) ($p=0.30$). (D) The curve of OS for *NOTCH1* wildtype (green) vs mutated (red) ($p=0.04$) and the curve of TTFT for *NOTCH1* wildtype (green) vs mutated (red) ($p=0.51$). (E) The curve of OS for *SF3B1* wildtype (green) vs mutated (red) ($p=0.002$) and the curve of TTFT for *SF3B1* wildtype (green) vs mutated (red) ($p=0.51$). (F) The curve of OS for chromosome 6 wildtype (green) vs deletion 6 (red) ($p=0.001$) and the curve of TTFT for chromosome 6 wildtype (green) vs deletion 6 (red) ($p=0.06$).

For survival analysis, as expected, unmutated *IgHV*, *NOTCH1*mut, *SF3B1*mut, and deletion 6q predicted shorter OS in CLL patients ($p=0.003$, 0.04, 0.002, 0.001, respectively). Unmutated *IgHV* also showed a lower TTFT ($p<0.0001$). Multivariate Cox proportional hazards regression analysis indicated that unmutated *IgHV* (RR=11.5, $p=0.03$), del(6) (RR=10.0, $p=0.021$) and *NOTCH1* mutation (RR=9.1, $p=0.017$) were associated with poor survival and unmutated *IgHV* (RR=2.7, $p=0.002$) and stage (RR=1.5, $p=0.021$) predicted for treatment to be initiated.

2. RT-MLPA and high-throughput sequencing

The structure of *TP53* is presented in Figure 2A. The combination of two promoters, named P1 and P2, and alternative splicing generate 12 major isoforms. The mRNA transcribed by P1 will translate either the full-length p53 (FLp53) or the $\Delta 40p53$ isoform, which started at codon 40. Promotor P2 can transcribe the isoforms $\Delta 133p53$ and $\Delta 160p53$, started at codon 133 or 160, respectively. Three variants in the C-terminus are also present due to alternative splicing of the exon9, named α , β , and γ . Exon9 β and γ have stop codons, leading to translational termination. Isoform α also translates exon10 and exon11. There are six functional domains of *TP53* which are depicted by different colors in Figure 2A. We have marked all the mutated sites on the *TP53* gene in our series (Figure 2A), all gathered on the DNA-binding domain, which is the most important functional domain of P53.

The probes of RT-MPLA were designed as illustrated in Figure 2B. We detected the expression of each exon junction to find more information in the mRNA and to determine the percent of variation on codon 72 to confirm the stability of this method (and the number of allele/ the zygosity?). More than 3000 reads of *TP53* in each patient among the 114 included were detected. Firstly, we evaluated the frequency of codon 72 variation detected by NGS versus RT-MLPA, illustrated in Figure 2C, showing a positive correlation ($p<0.0001$). The result means that RT-MLPA can also detect the mutation frequency in the coding site and reflects the situation of mRNA expression stably. The expression level of each *TP53* exon junction for the 114 CLL patients was shown in Figure 2D. Table 2 presents the linear association between the expression of exon9-10 (p53 α), exon9-9b (p53 β), exon9bg-10 (p53 β and γ), and a set of other expression exon junction, respectively, using a multiple linear regression model that was developed by a stepwise method. The expression of exon9-9b and exon9bg-10 had the most impact according to the value of standardized coefficients.

Model	B	t	R ²	F	p	Durbin-Watson
Exon9-10 (p53 α)	Constant	7.01	0.724	56.72	<0.0001	1.932
	Exon9bg-10	-3.26				
	Exon2-3	4.86				
	Exon7-8	2.81				
	Exon8-9	-2.76				
	Exon9-9b	-2.26				
Exon9-9b (p53 β)	Constant	2.93	0.841	593.03	<0.0001	2.057
	Exon9bg-10	24.35				
Exon9bg-10 (p53 β and γ)	Constant	5.45	0.881	271.23	<0.0001	2.171
	Exon9-9b	14.16				
	Exon6-7	-5.23				
	Exon9-10	-3.19				

Moreover, RT-MLPA can also reflect the frequency of a given splice site mutation. In our series, we found a patient that, after receiving 6 cycles FCR, presented a splice site mutation in intron 4 (c.376-2A>G) with a mutational frequency of 45% tested by NGS. Interestingly, when we tested this sample by RT-MLPA, this splice site mutation caused a decrease (47.4%) in the number of normal junctions between exons 4 and 5 and an increase (52.6%) in the number of junctions between exon 4 and exon5 truncated at Δ 133 site (exon5b) (Figure 2E1 and Figure 2E3 represented by the red lines), which correlates with the splice mutation found by NGS. After two years the patient relapsed and received 2-year treatment of ibrutinib (an irreversible inhibitor of Bruton's tyrosine kinase (BTK)), the expression of exon4-5b decreased to 5.9% (Figure 2E2 and Figure 2E3 represented by the green lines), while the mutational frequency of splice site detected by NGS was of 1.4%. Therefore, besides the fact that RT-MLPA can reflect the mutational frequency, it can also detect the change of mRNA expression caused by splice site mutation and treatment strategy.

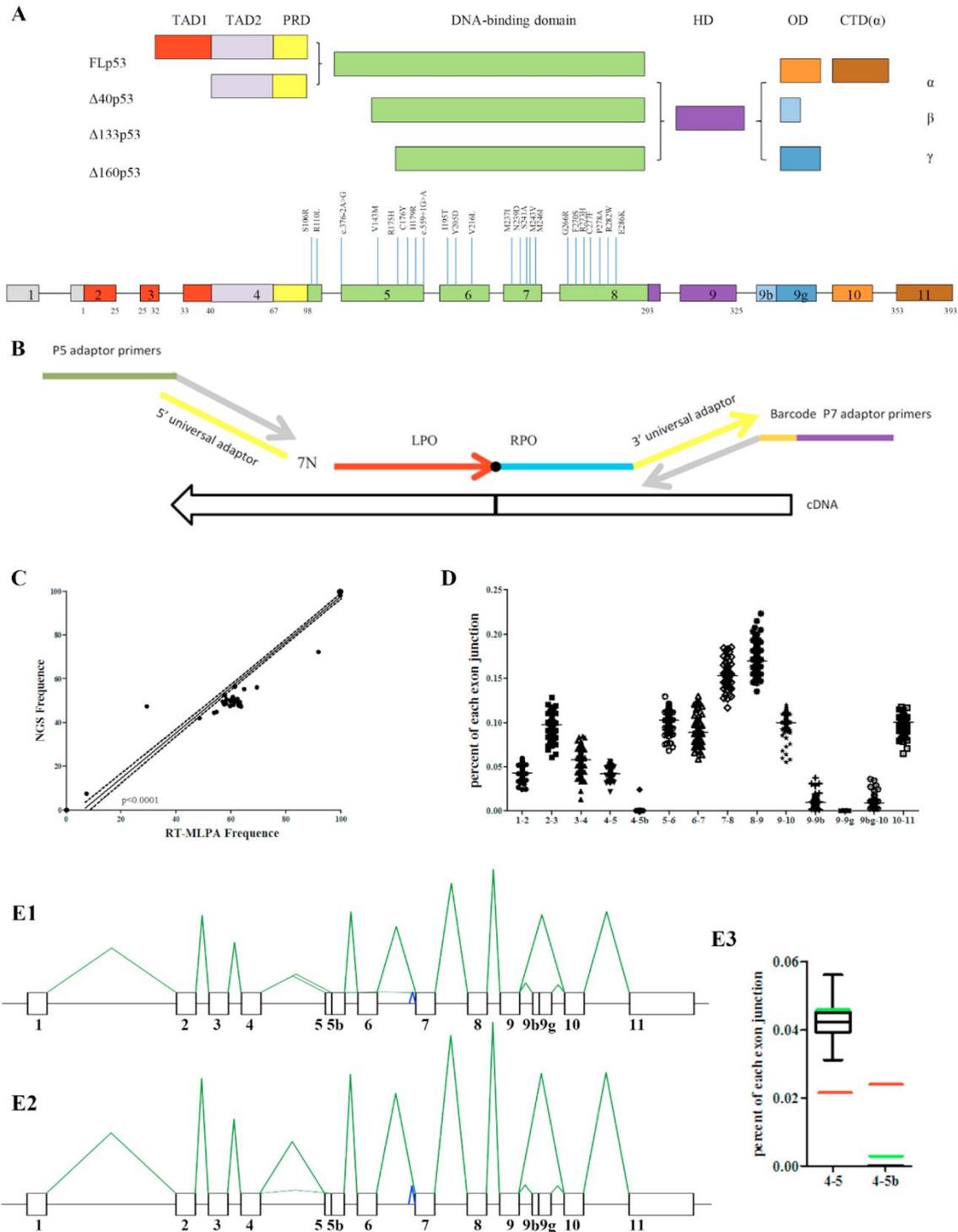


Figure 2 Analysis of *TP53* variants by NGS and RT-MLPA (A) The structure of *TP53*. (A) 11 exons make up six domains. The different colors represented different domains: two transactivation domains I and II (TAD I (red) and TAD II (light purple)), spanning residues 1–67; the proline-rich region (PRD, yellow) with residues 68–98; the DNA-binding domain (DBD, also called the core domain, green) with residues 94–292; the hinge domain (HD, dark purple), with residues 293–325; the oligomerization domain (OD, orange) with residues 326–353; and the carboxy-terminal regulatory domain (CTD, brown) spanning residues 353–393. Other variants of the C-terminal are exon9b (light blue) and 9g (dark blue). All the mutated sites were directed at the genetic picture, all gathered on the DNA-binding domain. (B) Principle of RT-MLPA. The probe pair for each *TP53* exon junction comprised left hybridizing sequence (LHS) and right hybridizing sequencing (RHS). LHS included a 5' universal adaptor sequence, a 7N random sequence (UMI), and a left probe

oligonucleotide (LPO) from 5' to 3'. RHS contained a right probe oligonucleotide (RPO) and a 3' universal adaptor sequence from 5' to 3'. After the probes are hybridized and ligated, we prepare libraries by P5 and barcode P7 adaptor primers to classify different samples. (C) Comparing the variant allele frequency of codon 72 on the exon 4 detected by NGS and RT-MLPA showed a positive relationship ($p < 0.0001$). (D) The expression of each exon junction in 114 CLL patients. (E) The RT-MLPA results of a patient presented splice site mutation between exon 4 to 5 (c.376-2A>G), E1 showed the expression of exon4-5 and exon4-5b when the patient just received 6 cycles of FCR and E2 presented the change of the expression of exon4-5 and exon4-5b after the 2-year treatment of Bruton's tyrosine kinase (BTK) inhibitor, which is a concordance to the result of NGS. E3 showed the total expression level of exon4-5 and exon4-5b in 113 other CLL patients (black) and patients with splice site mutation before (red line) and after (green line) treatment of BTK inhibitor.

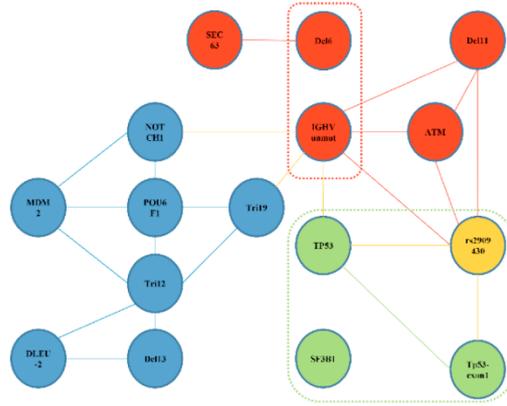
3. Relationship between the expression of *TP53* and multiple genes mutation

Although several studies investigate the clinical influence of *TP53* mutations, the role of *TP53* mRNA expression is controversial. Therefore, firstly, we analyzed the relationship between different genes (Figure 3A) and illustrated the significant positive correlations ($p < 0.05$) between them with lines. Secondly, we highlighted that the expression of some exon junctions was statistically different according to peculiar mutational profiles. For example, *TP53* mutated cases expressed a lower level of exon4-5, belonging only to FLp53 and $\Delta 40p53$ isoforms. The same situation was identified in del17p13 (*TP53*exon1), rs2909430, and *SF3B1*mut patients. Positive correlations were also detected between *TP53*mut, del17p13 (*TP53*exon1), and rs2909430 (Figure 3A, represented in green). Consequently, we divided patients into three groups according to the classification presented in Figure 3B (2). The three groups not only expressed different levels of exon4-5 but also showed a difference in OS, as illustrated in Figure 3B (3) and (4).

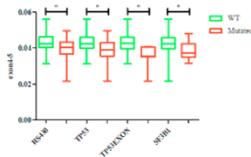
Additionally, we also compared the expression of *TP53* in other genetic groups. Unmutated *IgHV* cases expressed a lower level of exon9-9b and exon9bg-10, which may represent the p53 β and p53 γ isoforms (Figure 3C (1)). A similar situation also appeared in the deletion 6 group. According to Figure 3C (2), classification 2, we divided patients into three groups with increasingly different levels of expression; the higher points they scored, the higher level of exon9-10 and the lower level of exon9-9b and 9bg-10 they expressed (Figure 3C (3)). This trend was maintained in the OS and TTFT analysis (Figure 3C (4), $p < 0.0001$ and 0.001, respectively).

The last prognostic genetic molecular marker is mutations in *NOTCH1*. *NOTCH1* mutated patients expressed a higher level of exon10-11 (Figure 3D (1)), which represents another kind of C-terminal isoform of *TP53* (p53 α). When we combined all these 5 genetic factors, according to Figure 3D (2), classification 3, we found that higher risk groups expressed a lower level of exon9-9b and exon9bg-10 and a higher level of exon10-11 (Figure 3D (3)). The same trend could also be observed in the analysis of OS and TTFT (Figure 3D (4), $p = 0.0003$ and 0.004, respectively).

A



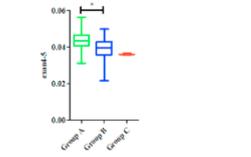
B (1)



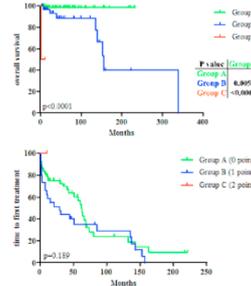
(2)

Classification 1 (green)	
Biomarker	Score
SF3B1	
Wildtype	0
Mutated	1
Tp53 net	
Wildtype	0
TP53mut	1
Del17p13	1
rs2909430	1
Risk group	
Group A	0
Group B	1
Group C	2

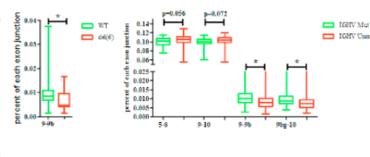
(3)



(4)



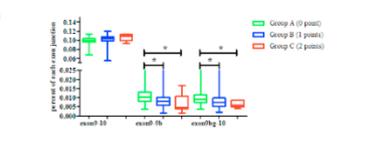
C (1)



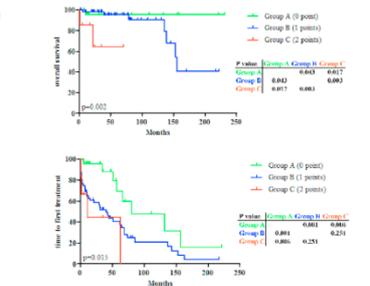
(2)

Classification 2 (red)	
Biomarker	Score
Deletion 6	
Wildtype	0
Deletion	1
IGHV	
Mutated	0
Unmutated	1
Risk group	
Group A	0
Group B	1
Group C	2

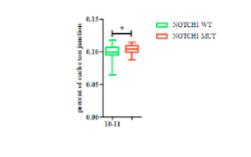
(3)



(4)



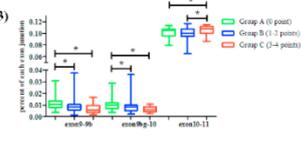
D (1)



(2)

Classification 3			
Biomarker	Score	Biomarker	Score
SF3B1			
Wildtype	0	Mutated	0
Mutated	1	Unmutated	1
Tp53 net		Deletion 6	
Wildtype	0	Wildtype	0
TP53mut	1	Deletion	1
Del17p13	1	NOTCH1	1
rs2909430	1	Wildtype	0
		Mutated	1
Risk group			
Group A	0	Group B	1-2
Group B	1	Group C	3-4
Group C	2		

(3)



(4)

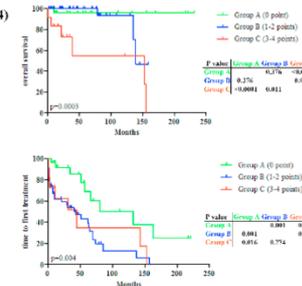


Figure 3 Relationship between expression of TP53 and different prognostic indexes.

(A) Correlation between different genetic alterations, the line connected two genes showed positive relation ($p < 0.05$). (B) (1) The index expressed a lower level of exon4-5, (2) according to classification 1, the genetic groups showed a significant difference in expression of exon4-5 (3) and OS (4). (C) (1) The index expressed a lower level of exon9-9b and exon9b-10, and (2) according to classification 2, the different groups showed different expressions of exon9-9b and exon9b-10 (3) and OS, TTFT (4) ($p < 0.0001$ and 0.001 , respectively). (D) (1) The different expression of exon10-11 between wildtype and mutated *NOTCH1*, and (2) according to classification 3 (including all genes showed prognostic value), the different groups showed different expression of exon9-9b, exon9b-10, and exon10-11 (3) and OS, TTFT (4) ($p = 0.0003$, 0.004 , respectively).

* Present the significant analysis ($p < 0.05$)

Discussion

Mutations of *TP53*, particularly in the DNA binding domain, have been regarded as the leading cause of p53 inactivation³¹. Although only 10% of patients showed a genetic mutation in CLL, *TP53* is still a significant predictor of treatment and prognosis. Among the clinical data of our CLL patients, mutational status of *IgHV*, *TP53* copy number variations detected by QMPFS, SNP rs29009430 and rs1625895 tested by NGS, and low level of hemoglobin had a close relationship with mutational status of *TP53*. Since it is well-known that the frequency of *TP53* mutation is consistently higher at chemo-refractoriness than at the time of diagnosis, it may decrease after the treatment of target inhibitors. Compared to the results of one patient in our series who received FCR and BTK inhibitor (ibrutinib), we found that RT-MLPA high-throughput sequence can also serve as a surveillance tool to monitor the change of the mutational frequency in *TP53*.

Traditionally, point mutations of the *TP53* gene have been considered to be responsible for cancer occurrence³². However, various truncated forms of p53 can also be responsible for its inactivation, especially the N-terminally truncated ones. Some researchers suggest the involvement of p53 isoforms in carcinogenesis²² since different expressions in normal and tumorigenic tissue in various types of cancer³³. The expression of p53 isoforms were dependent on the tissue types in normal body. For example, in normal breast tissues, P53, p53 β , p53 γ mRNA, but not Δ 133p53 α , Δ 133p53 β and Δ 133p53 γ were tested. However, in renal cell P53, p53 β , p53 γ Δ 133p53 α , and Δ 133p53 γ can detect except Δ 133p53 β . Several cancers that showed a low mutation rate of *TP53* have reported a significant change in the expression of p53 isoforms. For instance, breast cancer which shares a 25% mutation rate of *TP53*, expressed different levels of p53 isoforms. A decrease of the p53 β and p53 γ isoforms by 60%, contrasting with an increase of the Δ 133p53 by 40%, is observed in this setting³⁴. Similarly, The same situation is also observed in acute myeloid leukemia, which displayed only a 10% mutation rate but showed a large different expression level of the p53 β and p53 γ isoform³⁵. These studies have expanded the perspective to discover the mechanisms of the p53 isoforms and the improvement of cancer treatment.

In this study, RT-MLPA was designed to detect relative expression levels of mRNA. The detection procedure included hybridization probes to DNA templates, ligation of matched probe pairs, and amplification with universal PCR primers. Because each probe of *TP53* needed only about 30 bps, the length of the cDNA template may be small and the assay is suitable for degraded RNA samples from FFPE tissues. The other feature of our method was the high-throughput sequence for target RNA. NGS can analyze the specific nucleotide sequences in LHS and RHS and thus have no limit on the number of probes in one assay.

Furthermore, using the Illumina-compatible sequencing primers allowed us to have a more flexible pooling strategy when the RT-MLPA sequence needs extremely little sequencing space. The non-concordance between cytogenetic and molecular testing leads to a challenge in estimating these factors into prognosis models to guide the clinical decision. We combined the results of the RT-MLPA high-throughput sequence with other molecular factors to find more information on mRNA and the relationship among multiple genes alterations including *TP53* in CLL. It's the first time that the RT-MLPA high-throughput was used to sequence CLL patients to discover the transcriptional level of mRNA.

The different expressions of exon junction between several prognostic factors were systematically related to the exon4-5, exon5-6, exon9-9b, exon9bg-10, and exon10-11, which may represent the distribution of FLp53 and Δ 40p53, Δ 133p53 and Δ 160p53, p53 α , p53 β , and p53 γ isoforms, respectively. When we classified patients according to the factors that expressed a similar level of exon junction, different groups predicted the treatment and prognosis well in CLL. The low expression level of exon4-5, exon9-9b, and exon9bg-10 and high expression level of exon5-6, exon9-10, and exon10-11 indicated lower overall survival and highlighted the importance of treatment initiation in our study.

The principal limitation of our study is the single center design with a reduced number of patients included. We performed this research on 114 patients since January 2002, among which only 18 showed mutated *TP53*. Because

of limited samples, we could not compare the change of expression before and after the different treatments in each patient. Multicentre cooperation is required to achieve this goal.

In summary, high-throughput RT-MLPA technology is a robust and low-cost method to detect RNA expression levels in CLL samples, which can be used in both research and clinical practice. To our knowledge, this is the first study to evaluate the expression of *TP53* by testing the expression level of all exon junctions. Our findings are in concordance with data reported in other cancers since the short N-terminal truncated isoforms, including $\Delta 133$ and $\Delta 160$ p53, are more likely to be associated with aggressive disease, and the expression of p53 β and p53 γ isoforms may predict good prognosis²².

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66(1):7-30. doi:10.3322/caac.21332
2. Koehrer S, Burger JA. B-cell receptor signaling in chronic lymphocytic leukemia and other B-cell malignancies. *Clin Adv Hematol Oncol HO.* 2016;14(1):55-65.
3. Oppezzo P, Vuillier F, Vasconcelos Y, et al. Chronic lymphocytic leukemia B cells expressing AID display dissociation between class switch recombination and somatic hypermutation. *Blood.* 2003;101(10):4029-4032. doi:10.1182/blood-2002-10-3175
4. Pozzo F, Bittolo T, Vendramini E, et al. NOTCH1-mutated chronic lymphocytic leukemia cells are characterized by a MYC-related overexpression of nucleophosmin 1 and ribosome-associated components. *Leukemia.* 2017;31(11):2407-2415. doi:10.1038/leu.2017.90
5. Wang L, Brooks AN, Fan J, et al. Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia. *Cancer Cell.* 2016;30(5):750-763. doi:10.1016/j.ccell.2016.10.005
6. Oscier DG, Gardiner AC, Mould SJ, et al. Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood.* 2002;100(4):1177-1184.
7. Byrd JC, Hillmen P, Ghia P, et al. Acalabrutinib Versus Ibrutinib in Previously Treated Chronic Lymphocytic Leukemia: Results of the First Randomized Phase III Trial. *J Clin Oncol Off J Am Soc Clin Oncol.* 2021;39(31):3441-3452. doi:10.1200/JCO.21.01210
8. Cuneo A, Follows G, Rigolin GM, et al. Efficacy of bendamustine and rituximab as first salvage treatment in chronic lymphocytic leukemia and indirect comparison with ibrutinib: a GIMEMA, ERIC and UK CLL FORUM study. *Haematologica.* 2018;103(7):1209-1217. doi:10.3324/haematol.2018.189837
9. Jain N, Keating M, Thompson P, et al. Ibrutinib and Venetoclax for First-Line Treatment of CLL. *N Engl J Med.* 2019;380(22):2095-2103. doi:10.1056/NEJMoa1900574
10. Shanafelt TD, Wang XV, Kay NE, et al. Ibrutinib-Rituximab or Chemoimmunotherapy for Chronic Lymphocytic Leukemia. *N Engl J Med.* 2019;381(5):432-443. doi:10.1056/NEJMoa1817073
11. Liu E, Marin D, Banerjee P, et al. Use of CAR-Transduced Natural Killer Cells in CD19-Positive Lymphoid Tumors. *N Engl J Med.* 2020;382(6):545-553. doi:10.1056/NEJMoa1910607
12. Goede V, Fischer K, Busch R, et al. Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N Engl J Med.* 2014;370(12):1101-1110. doi:10.1056/NEJMoa1313984
13. Guièze R, Robbe P, Clifford R, et al. Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood.* 2015;126(18):2110-2117. doi:10.1182/blood-2015-05-647578

14. Al-Sawaf O, Robrecht S, Bahlo J, et al. Richter transformation in chronic lymphocytic leukemia (CLL)-a pooled analysis of German CLL Study Group (GCLLSG) front line treatment trials. *Leukemia*. 2021;35(1):169-176. doi:10.1038/s41375-020-0797-x
15. Klintman J, Appleby N, Stamatopoulos B, et al. Genomic and transcriptomic correlates of Richter transformation in chronic lymphocytic leukemia. *Blood*. 2021;137(20):2800-2816. doi:10.1182/blood.2020005650
16. Stamatopoulos B, Timbs A, Bruce D, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia*. 2017;31(4):837-845. doi:10.1038/leu.2016.307
17. Gocke CD, Gladstone DE. The absolute percent deviation of IGHV mutation rather than a 98% cut-off predicts survival of chronic lymphocytic leukaemia patients treated with fludarabine, cyclophosphamide and rituximab. *Br J Haematol*. 2018;180(1):7-8. doi:10.1111/bjh.15015
18. International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol*. 2016;17(6):779-790. doi:10.1016/S1470-2045(16)30029-8
19. Stilgenbauer S, Schnaiter A, Paschka P, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood*. 2014;123(21):3247-3254. doi:10.1182/blood-2014-01-546150
20. Brieghel C, Kinalis S, Yde CW, et al. Deep targeted sequencing of TP53 in chronic lymphocytic leukemia: clinical impact at diagnosis and at time of treatment. *Haematologica*. 2019;104(4):789-796. doi:10.3324/haematol.2018.195818
21. Al-Sawaf O, Fischer K. TP53 mutations in CLL: does frequency matter? *Blood*. 2021;138(25):2600-2601. doi:10.1182/blood.2021012343
22. Kim S, An SSA. Role of p53 isoforms and aggregations in cancer. *Medicine (Baltimore)*. 2016;95(26):e3993. doi:10.1097/MD.0000000000003993
23. Malcikova J, Tausch E, Rossi D, et al. ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia-update on methodological approaches and results interpretation. *Leukemia*. 2018;32(5):1070-1080. doi:10.1038/s41375-017-0007-7
24. Oscier DG, Rose-Zerilli MJ, Winkelmann N, et al. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood*. 2013;121(3):468-475. doi:10.1182/blood-2012-05-429282
25. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131(25):2745-2760. doi:10.1182/blood-2017-09-806398
26. Jardin F, Ruminy P, Kerckaert JP, et al. Detection of somatic quantitative genetic alterations by multiplex polymerase chain reaction for the prediction of outcome in diffuse large B-cell lymphomas. *Haematologica*. 2008;93(4):543-550. doi:10.3324/haematol.12251
27. Jardin F, Picquenot JM, Parmentier F, et al. Detection of gene copy number aberrations in mantle cell lymphoma by a single quantitative multiplex PCR assay: clinicopathological relevance and prognosis value. *Br J Haematol*. 2009;146(6):607-618. doi:10.1111/j.1365-2141.2009.07791.x
28. Designing Synthetic MLPA Probes - Downloads / MLPA / Synthetic Probe Design - MRC Holland Support. Accessed June 8, 2022. <https://support.mrcholland.com/downloads/files/designing-synthetic-mlpa-probes>
29. Li Y ze, Xu J jie, Qian H zhu, et al. High prevalence of HIV infection and unprotected anal intercourse among older men who have sex with men in China: a systematic review and meta-analysis. *BMC Infect Dis*. 2014;14:531. doi:10.1186/1471-2334-14-531

30. Zheng X, Yu S, Long J, et al. Comparison of the clinical characteristics of primary thyroid lymphoma and diffuse sclerosing variant of papillary thyroid carcinoma. *Endocr Connect.* 2022;11(1):e210364. doi:10.1530/EC-21-0364
31. Bouaoun L, Sonkin D, Ardin M, et al. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. *Hum Mutat.* 2016;37(9):865-876. doi:10.1002/humu.23035
32. Kasthuber ER, Lowe SW. Putting p53 in Context. *Cell.* 2017;170(6):1062-1078. doi:10.1016/j.cell.2017.08.028
33. Joruz SM, Bourdon JC. p53 Isoforms: Key Regulators of the Cell Fate Decision. *Cold Spring Harb Perspect Med.* 2016;6(8):a026039. doi:10.1101/cshperspect.a026039
34. Bourdon JC, Fernandes K, Murray-Zmijewski F, et al. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.* 2005;19(18):2122-2137. doi:10.1101/gad.1339905
35. Anensen N, Oyan AM, Bourdon JC, Kalland KH, Bruserud O, Gjertsen BT. A distinct p53 protein isoform signature reflects the onset of induction chemotherapy for acute myeloid leukemia. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2006;12(13):3985-3992. doi:10.1158/1078-0432.CCR-05-1970

Appendix 1 The probes of aim exon of <i>TP53</i>	
Probes	Seq
<i>TP53</i> Exon1 LPO	GACACGCTTCCCTGGATTGG
<i>TP53</i> Exon2 RPO	CAGCCAGACTGCCTTCCGG
<i>TP53</i> Exon2 LPO	TCAGGAAACATTTTCAGACCTATGGAAACT
<i>TP53</i> Exon3 RPO	ACTTCCTGAAAACAACGTTCTG
<i>TP53</i> Exon3 LPO	ACTTCCTGAAAACAACGTTCTG
<i>TP53</i> Exon4 RPO	TCCCCCTTGCCGTCCCAAGC
<i>TP53</i> Exon4 LPO	AGCCAAGTCTGTGACTTGCACG
<i>TP53</i> Exon5 RPO	TACTCCCCTGCCCTCAACAAGAT
<i>TP53</i> Exon5b RPO	ATGTTTTGCCAACTGGCCAAGA
<i>TP53</i> Exon5 LPO	AGCGTGCTCAGATAGCGATG
<i>TP53</i> Exon6 RPO	GTCTGGCCCCCTCCTCAGCAT
<i>TP53</i> Exon6 LPO	TGCCCTATGAGCCGCTGAG
<i>TP53</i> Exon7 RPO	GTTGGCTCTGACTGTACCACCATCC
<i>TP53</i> Exon7 LPO	CCATCATCACACTGGAAGACTCCAG
<i>TP53</i> Exon8 RPO	TGGTAATCTACTGGGACGGAACAGCT
<i>TP53</i> Exon8 LPO	CCCCAGGGAGCACTAAGCGAG
<i>TP53</i> Exon9 RPO	CACTGCCCAACAACACCAGCT
<i>TP53</i> Exon9 LPO	ACTGGATGGAGAATATTTACCCTTCAG
<i>TP53</i> Exon9b RPO	GACCAGACCAGCTTTCAAAAAGAAAATTGT
<i>TP53</i> Exon9g RPO	ATGCTACTTGACTTACGATGGTGTACTTC
<i>TP53</i> Exon9bg LPO	CTGATAAACTCGTCGTAAGTTGAAAATATT
<i>TP53</i> Exon10 RPO	ATCCGTGGGCGTGAGCGCTT
<i>TP53</i> Exon10 LPO	GGGAGCAGGGCTCACTCCAG
<i>TP53</i> Exon11 RPO	CCACCTGAAGTCCAAAAAGGGTC
<i>TP53</i> Exon4 codon72 mut RPO	GCGTGGCCCCCTGCACCAGCAG

<i>TP53</i> Exon4 codon72 LPO	ATGCCAGAGGCTGCTCCCC
<i>TP53</i> Exon4 codon72 wt RPO	CCGTGGCCCCTGCACCAGCAG

REFERENCES

1. Hallek M, Shanafelt TD, Eichhorst B. Chronic lymphocytic leukaemia. *Lancet Lond Engl*. 2018;391(10129):1524-1537. doi:10.1016/S0140-6736(18)30422-7
2. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131(25):2745-2760. doi:10.1182/blood-2017-09-806398
3. Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*. 2008;111(12):5446-5456. doi:10.1182/blood-2007-06-093906
4. Moreau EJ, Matutes E, A'Hern RP, et al. Improvement of the chronic lymphocytic leukemia scoring system with the monoclonal antibody SN8 (CD79b). *Am J Clin Pathol*. 1997;108(4):378-382. doi:10.1093/ajcp/108.4.378
5. Strati P, Shanafelt TD. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood*. 2015;126(4):454-462. doi:10.1182/blood-2015-02-585059
6. Solomon BM, Chaffee KG, Moreira J, et al. Risk of non-hematologic cancer in individuals with high-count monoclonal B-cell lymphocytosis. *Leukemia*. 2016;30(2):331-336. doi:10.1038/leu.2015.235
7. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood*. 1975;46(2):219-234.
8. Binet JL, Auquier A, Dighiero G, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer*. 1981;48(1):198-206. doi:10.1002/1097-0142(19810701)48:1<198::aid-cnrc2820480131>3.0.co;2-v
9. Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526(7574):525-530. doi:10.1038/nature15395
10. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999;94(6):1840-1847.
11. Hallek M, Fischer K, Fingerle-Rowson G, et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet Lond Engl*. 2010;376(9747):1164-1174. doi:10.1016/S0140-6736(10)61381-5
12. Crespo M, Bosch F, Villamor N, et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N Engl J Med*. 2003;348(18):1764-1775. doi:10.1056/NEJMoa023143
13. Rassenti LZ, Jain S, Keating MJ, et al. Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood*. 2008;112(5):1923-1930. doi:10.1182/blood-2007-05-092882

14. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475(7354):101-105. doi:10.1038/nature10113
15. Hallek M, Wanders L, Ostwald M, et al. Serum beta(2)-microglobulin and serum thymidine kinase are independent predictors of progression-free survival in chronic lymphocytic leukemia and immunocytoma. *Leuk Lymphoma*. 1996;22(5-6):439-447. doi:10.3109/10428199609054782
16. International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol*. 2016;17(6):779-790. doi:10.1016/S1470-2045(16)30029-8
17. Gentile M, Shanafelt TD, Mauro FR, et al. Predictive value of the CLL-IPI in CLL patients receiving chemo-immunotherapy as first-line treatment. *Eur J Haematol*. Published online July 24, 2018. doi:10.1111/ejh.13149
18. Lad DP, Tejaswi V, Jindal N, et al. Modified CLL International Prognostic Index (CLL-LIPI) using lymphocyte doubling time (LDT) in place of IgHV mutation status in resource-limited settings predicts time to first treatment and overall survival. *Leuk Lymphoma*. 2020;61(6):1512-1515. doi:10.1080/10428194.2020.1719099
19. Wierda WG, O'Brien S, Wang X, et al. Prognostic nomogram and index for overall survival in previously untreated patients with chronic lymphocytic leukemia. *Blood*. 2007;109(11):4679-4685. doi:10.1182/blood-2005-12-051458
20. Molica S, Mauro FR, Callea V, et al. The utility of a prognostic index for predicting time to first treatment in early chronic lymphocytic leukemia: the GIMEMA experience. *Haematologica*. 2010;95(3):464-469. doi:10.3324/haematol.2009.011767
21. Bulian P, Rossi D, Forconi F, et al. IGHV gene mutational status and 17p deletion are independent molecular predictors in a comprehensive clinical-biological prognostic model for overall survival prediction in chronic lymphocytic leukemia. *J Transl Med*. 2012;10:18. doi:10.1186/1479-5876-10-18
22. Tam CS, Seymour JF. A new prognostic score for CLL. *Blood*. 2014;124(1):1-2. doi:10.1182/blood-2014-05-575407
23. Pflug N, Bahlo J, Shanafelt TD, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood*. 2014;124(1):49-62. doi:10.1182/blood-2014-02-556399
24. Molica S, Giannarelli D, Mirabelli R, et al. Unavailability of thymidine kinase does not preclude the use of German comprehensive prognostic index: results of an external validation analysis in early chronic lymphocytic leukemia and comparison with MD Anderson Cancer Center model. *Eur J Haematol*. 2016;96(1):72-77. doi:10.1111/ejh.12550
25. Gentile M, Shanafelt TD, Rossi D, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood*. 2016;128(16):2093-2095. doi:10.1182/blood-2016-07-728261
26. Molica S, Shanafelt TD, Giannarelli D, 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):1090-1095. doi:10.1002/ajh.24493
27. Seymour EK, Ruterbusch JJ, Beebe-Dimmer JL, Schiffer CA. Real-world testing and treatment patterns in chronic lymphocytic leukemia: A SEER patterns of care analysis. *Cancer*. 2019;125(1):135-143. doi:10.1002/cncr.31738
 28. Muñoz-Novas C, Poza-Santaella M, González-Gascón Y Marín I, et al. The International Prognostic Index for Patients with Chronic Lymphocytic Leukemia Has the Higher Value in Predicting Overall Outcome Compared with the Barcelona-Brno Biomarkers Only Prognostic Model and the MD Anderson Cancer Center Prognostic Index. *BioMed Res Int*. 2018;2018:9506979. doi:10.1155/2018/9506979
 29. Nadeu F, Delgado J, Royo C, et al. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood*. 2016;127(17):2122-2130. doi:10.1182/blood-2015-07-659144
 30. Sutton LA, Young E, Baliakas P, et al. Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica*. 2016;101(8):959-967. doi:10.3324/haematol.2016.141812
 31. Shanafelt TD, Wang XV, Kay NE, et al. Ibrutinib-Rituximab or Chemoimmunotherapy for Chronic Lymphocytic Leukemia. *N Engl J Med*. 2019;381(5):432-443. doi:10.1056/NEJMoa1817073
 32. Byrd JC, Hillmen P, Ghia P, et al. Acalabrutinib Versus Ibrutinib in Previously Treated Chronic Lymphocytic Leukemia: Results of the First Randomized Phase III Trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 2021;39(31):3441-3452. doi:10.1200/JCO.21.01210
 33. Xu W, Yang S, Zhou K, et al. Treatment of relapsed/refractory chronic lymphocytic leukemia/small lymphocytic lymphoma with the BTK inhibitor zanubrutinib: phase 2, single-arm, multicenter study. *J Hematol Oncol J Hematol Oncol*. 2020;13(1):48. doi:10.1186/s13045-020-00884-4
 34. Molica S, Giannarelli D, Mirabelli R, Levato L, Shanafelt TD. The magnitude of improvement in progression-free survival with targeted therapy in relapsed/refractory chronic lymphocytic leukemia based on prognostic risk category: a systematic review and meta-analysis. *Leuk Lymphoma*. 2019;60(7):1644-1649. doi:10.1080/10428194.2018.1543882
 35. Kittai AS, Lunning M, Danilov AV. Relevance of Prognostic Factors in the Era of Targeted Therapies in CLL. *Curr Hematol Malig Rep*. 2019;14(4):302-309. doi:10.1007/s11899-019-00511-1
 36. Eichhorst B, Hallek M. Prognostication of chronic lymphocytic leukemia in the era of new agents. *Hematol Am Soc Hematol Educ Program*. 2016;2016(1):149-155. doi:10.1182/asheducation-2016.1.149
 37. Tausch E, Schneider C, Robrecht S, et al. Prognostic and predictive impact of genetic markers in patients with CLL treated with obinutuzumab and venetoclax. *Blood*. 2020;135(26):2402-2412. doi:10.1182/blood.2019004492

38. Dighiero G, Maloum K, Desablens B, et al. Chlorambucil in indolent chronic lymphocytic leukemia. French Cooperative Group on Chronic Lymphocytic Leukemia. *N Engl J Med*. 1998;338(21):1506-1514. doi:10.1056/NEJM199805213382104
39. Robak T, Dmoszynska A, Solal-Céligny P, et al. Rituximab plus fludarabine and cyclophosphamide prolongs progression-free survival compared with fludarabine and cyclophosphamide alone in previously treated chronic lymphocytic leukemia. *J Clin Oncol Off J Am Soc Clin Oncol*. 2010;28(10):1756-1765. doi:10.1200/JCO.2009.26.4556
40. Flinn IW, Neuberg DS, Grever MR, et al. Phase III trial of fludarabine plus cyclophosphamide compared with fludarabine for patients with previously untreated chronic lymphocytic leukemia: US Intergroup Trial E2997. *J Clin Oncol Off J Am Soc Clin Oncol*. 2007;25(7):793-798. doi:10.1200/JCO.2006.08.0762
41. Goede V, Fischer K, Engelke A, et al. Obinutuzumab as frontline treatment of chronic lymphocytic leukemia: updated results of the CLL11 study. *Leukemia*. 2015;29(7):1602-1604. doi:10.1038/leu.2015.14
42. Fischer K, Bahlo J, Fink AM, et al. Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial. *Blood*. 2016;127(2):208-215. doi:10.1182/blood-2015-06-651125
43. Döhner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000;343(26):1910-1916. doi:10.1056/NEJM200012283432602
44. O'Brien S, Jones JA, Coutre SE, et al. Ibrutinib for patients with relapsed or refractory chronic lymphocytic leukaemia with 17p deletion (RESONATE-17): a phase 2, open-label, multicentre study. *Lancet Oncol*. 2016;17(10):1409-1418. doi:10.1016/S1470-2045(16)30212-1
45. Burger JA, Keating MJ, Wierda WG, et al. Safety and activity of ibrutinib plus rituximab for patients with high-risk chronic lymphocytic leukaemia: a single-arm, phase 2 study. *Lancet Oncol*. 2014;15(10):1090-1099. doi:10.1016/S1470-2045(14)70335-3
46. Jain N, Keating M, Thompson P, et al. Ibrutinib and Venetoclax for First-Line Treatment of CLL. *N Engl J Med*. 2019;380(22):2095-2103. doi:10.1056/NEJMoa1900574
47. Roberts AW, Davids MS, Pagel JM, et al. Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. *N Engl J Med*. 2016;374(4):311-322. doi:10.1056/NEJMoa1513257
48. Stilgenbauer S, Eichhorst B, Schetelig J, et al. Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study. *Lancet Oncol*. 2016;17(6):768-778. doi:10.1016/S1470-2045(16)30019-5
49. Seymour JF, Ma S, Brander DM, et al. Venetoclax plus rituximab in relapsed or refractory chronic lymphocytic leukaemia: a phase 1b study. *Lancet Oncol*. 2017;18(2):230-240. doi:10.1016/S1470-2045(17)30012-8
50. Kater AP, Seymour JF, Hillmen P, et al. Fixed Duration of Venetoclax-Rituximab in Relapsed/Refractory Chronic Lymphocytic Leukemia Eradicates Minimal Residual Disease and Prolongs Survival: Post-Treatment Follow-Up of the MURANO Phase III Study. *J Clin Oncol Off J Am Soc Clin Oncol*. 2019;37(4):269-277. doi:10.1200/JCO.18.01580

51. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365(26):2497-2506. doi:10.1056/NEJMoa1109016
52. Rossi D, Rasi S, Fabbri G, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 2012;119(2):521-529. doi:10.1182/blood-2011-09-379966
53. Al-Sawaf O, Robrecht S, Bahlo J, et al. Richter transformation in chronic lymphocytic leukemia (CLL)-a pooled analysis of German CLL Study Group (GCLLSG) front line treatment trials. *Leukemia*. 2021;35(1):169-176. doi:10.1038/s41375-020-0797-x
54. Klintman J, Appleby N, Stamatopoulos B, et al. Genomic and transcriptomic correlates of Richter transformation in chronic lymphocytic leukemia. *Blood*. 2021;137(20):2800-2816. doi:10.1182/blood.2020005650
55. Puente XS, Beà S, Valdés-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519-524. doi:10.1038/nature14666
56. Villamor N, Conde L, Martínez-Trillos A, et al. NOTCH1 mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia*. 2013;27(5):1100-1106. doi:10.1038/leu.2012.357
57. Guièze R, Robbe P, Clifford R, et al. Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood*. 2015;126(18):2110-2117. doi:10.1182/blood-2015-05-647578
58. González-Gascón Y Marín I, Hernández-Sánchez M, Rodríguez-Vicente AE, et al. A high proportion of cells carrying trisomy 12 is associated with a worse outcome in patients with chronic lymphocytic leukemia. *Hematol Oncol*. 2016;34(2):84-92. doi:10.1002/hon.2196
59. Baliakas P, Puiggros A, Xochelli A, et al. Additional trisomies amongst patients with chronic lymphocytic leukemia carrying trisomy 12: the accompanying chromosome makes a difference. *Haematologica*. 2016;101(7):e299-302. doi:10.3324/haematol.2015.140202
60. Roos-Weil D, Nguyen-Khac F, Chevret S, et al. Mutational and cytogenetic analyses of 188 CLL patients with trisomy 12: A retrospective study from the French Innovative Leukemia Organization (FILO) working group. *Genes Chromosomes Cancer*. 2018;57(11):533-540. doi:10.1002/gcc.22650
61. Goy J, Gillan TL, Bruyere H, et al. Chronic Lymphocytic Leukemia Patients With Deletion 11q Have a Short Time to Requirement of First-Line Therapy, But Long Overall Survival: Results of a Population-Based Cohort in British Columbia, Canada. *Clin Lymphoma Myeloma Leuk*. 2017;17(6):382-389. doi:10.1016/j.clml.2017.04.001
62. Hernández JÁ, Hernández-Sánchez M, Rodríguez-Vicente AE, et al. A Low Frequency of Losses in 11q Chromosome Is Associated with Better Outcome and Lower Rate of Genomic Mutations in Patients with Chronic Lymphocytic Leukemia. *PLoS One*. 2015;10(11):e0143073. doi:10.1371/journal.pone.0143073
63. Chavez JC, Kharfan-Dabaja MA, Kim J, et al. Genomic aberrations deletion 11q and deletion 17p independently predict for worse progression-free and overall survival after allogeneic

- hematopoietic cell transplantation for chronic lymphocytic leukemia. *Leuk Res*. 2014;38(10):1165-1172. doi:10.1016/j.leukres.2014.04.006
64. Hallek M, Al-Sawaf O. Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures. *Am J Hematol*. 2021;96(12):1679-1705. doi:10.1002/ajh.26367
 65. Bagacean C, Tempescul A, Ternant D, et al. 17p deletion strongly influences rituximab elimination in chronic lymphocytic leukemia. *J Immunother Cancer*. 2019;7(1):22. doi:10.1186/s40425-019-0509-0
 66. Yuan YY, Zhu HY, Wu JZ, et al. The percentage of cells with 17p deletion and the size of 17p deletion subclones show prognostic significance in chronic lymphocytic leukemia. *Genes Chromosomes Cancer*. 2019;58(1):43-51. doi:10.1002/gcc.22692
 67. Hu B, Patel KP, Chen HC, et al. Association of gene mutations with time-to-first treatment in 384 treatment-naive chronic lymphocytic leukaemia patients. *Br J Haematol*. 2019;187(3):307-318. doi:10.1111/bjh.16042
 68. Juliusson G, Oscier DG, Fitchett M, et al. Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. *N Engl J Med*. 1990;323(11):720-724. doi:10.1056/NEJM199009133231105
 69. Baliakas P, Iskas M, Gardiner A, et al. Chromosomal translocations and karyotype complexity in chronic lymphocytic leukemia: a systematic reappraisal of classic cytogenetic data. *Am J Hematol*. 2014;89(3):249-255. doi:10.1002/ajh.23618
 70. Baliakas P, Jeromin S, Iskas M, et al. Cytogenetic complexity in chronic lymphocytic leukemia: definitions, associations, and clinical impact. *Blood*. 2019;133(11):1205-1216. doi:10.1182/blood-2018-09-873083
 71. Yeung CCS, Shadman M. How to Choose the Best Treatment and Testing for Chronic Lymphocytic Leukemia in the Tsunami of New Treatment Options. *Curr Oncol Rep*. 2019;21(8):74. doi:10.1007/s11912-019-0819-x
 72. Pekarsky Y, Balatti V, Croce CM. BCL2 and miR-15/16: from gene discovery to treatment. *Cell Death Differ*. 2018;25(1):21-26. doi:10.1038/cdd.2017.159
 73. Veronese A, Pepe F, Chiacchia J, et al. Allele-specific loss and transcription of the miR-15a/16-1 cluster in chronic lymphocytic leukemia. *Leukemia*. 2015;29(1):86-95. doi:10.1038/leu.2014.139
 74. Cimmino A, Calin GA, Fabbri M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A*. 2005;102(39):13944-13949. doi:10.1073/pnas.0506654102
 75. Raveche ES, Salerno E, Scaglione BJ, et al. Abnormal microRNA-16 locus with synteny to human 13q14 linked to CLL in NZB mice. *Blood*. 2007;109(12):5079-5086. doi:10.1182/blood-2007-02-071225
 76. Lia M, Carette A, Tang H, et al. Functional dissection of the chromosome 13q14 tumor-suppressor locus using transgenic mouse lines. *Blood*. 2012;119(13):2981-2990. doi:10.1182/blood-2011-09-381814

77. Ghia P, Transidico P, Veiga JP, et al. Chemoattractants MDC and TARC are secreted by malignant B-cell precursors following CD40 ligation and support the migration of leukemia-specific T cells. *Blood*. 2001;98(3):533-540. doi:10.1182/blood.v98.3.533
78. Bagnara D, Kaufman MS, Calissano C, et al. A novel adoptive transfer model of chronic lymphocytic leukemia suggests a key role for T lymphocytes in the disease. *Blood*. 2011;117(20):5463-5472. doi:10.1182/blood-2010-12-324210
79. Giza DE, Calin GA. microRNA and Chronic Lymphocytic Leukemia. *Adv Exp Med Biol*. 2015;889:23-40. doi:10.1007/978-3-319-23730-5_2
80. Davari N, Ahmadpour F, Kiani AA, Azadpour M, Asadi ZT. Evaluation of microRNA-223 and microRNA-125a expression association with STAT3 and Bcl2 genes in blood leukocytes of CLL patients: a case-control study. *BMC Res Notes*. 2021;14(1):21. doi:10.1186/s13104-020-05428-0
81. Sharma S, Pavlasova GM, Seda V, et al. miR-29 modulates CD40 signaling in chronic lymphocytic leukemia by targeting TRAF4: an axis affected by BCR inhibitors. *Blood*. 2021;137(18):2481-2494. doi:10.1182/blood.2020005627
82. Cerna K, Oppelt J, Chochola V, et al. MicroRNA miR-34a downregulates FOXP1 during DNA damage response to limit BCR signalling in chronic lymphocytic leukaemia B cells. *Leukemia*. 2019;33(2):403-414. doi:10.1038/s41375-018-0230-x
83. Mraz M, Chen L, Rassenti LZ, et al. miR-150 influences B-cell receptor signaling in chronic lymphocytic leukemia by regulating expression of GAB1 and FOXP1. *Blood*. 2014;124(1):84-95. doi:10.1182/blood-2013-09-527234
84. Balatti V, Rizzotto L, Miller C, et al. TCL1 targeting miR-3676 is codeleted with tumor protein p53 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2015;112(7):2169-2174. doi:10.1073/pnas.1500010112
85. Bichi R, Shinton SA, Martin ES, et al. Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. *Proc Natl Acad Sci U S A*. 2002;99(10):6955-6960. doi:10.1073/pnas.102181599
86. Cui B, Chen L, Zhang S, et al. MicroRNA-155 influences B-cell receptor signaling and associates with aggressive disease in chronic lymphocytic leukemia. *Blood*. 2014;124(4):546-554. doi:10.1182/blood-2014-03-559690
87. Baliakas P, Hadzidimitriou A, Sutton LA, et al. Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. *Lancet Haematol*. 2014;1(2):e74-84. doi:10.1016/S2352-3026(14)00005-2
88. Lezina L, Spriggs RV, Beck D, et al. CD40L/IL-4-stimulated CLL demonstrates variation in translational regulation of DNA damage response genes including ATM. *Blood Adv*. 2018;2(15):1869-1881. doi:10.1182/bloodadvances.2017015560
89. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. *Blood*. 2008;112(5):1570-1580. doi:10.1182/blood-2008-02-078071
90. Ghia P, ten Boekel E, Sanz E, de la Hera A, Rolink A, Melchers F. Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the

- rearrangement status of the immunoglobulin H and L chain gene loci. *J Exp Med*. 1996;184(6):2217-2229. doi:10.1084/jem.184.6.2217
91. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 2000;13(1):37-45. doi:10.1016/s1074-7613(00)00006-6
 92. Chi X, Li Y, Qiu X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology*. 2020;160(3):233-247. doi:10.1111/imm.13176
 93. Niemann CU, Wiestner A. B-cell receptor signaling as a driver of lymphoma development and evolution. *Semin Cancer Biol*. 2013;23(6):410-421. doi:10.1016/j.semcancer.2013.09.001
 94. Koehrer S, Burger JA. B-cell receptor signaling in chronic lymphocytic leukemia and other B-cell malignancies. *Clin Adv Hematol Oncol HO*. 2016;14(1):55-65.
 95. Robak T, Robak P. BCR signaling in chronic lymphocytic leukemia and related inhibitors currently in clinical studies. *Int Rev Immunol*. 2013;32(4):358-376. doi:10.3109/08830185.2013.786711
 96. Packham G, Krysov S, Allen A, et al. The outcome of B-cell receptor signaling in chronic lymphocytic leukemia: proliferation or anergy. *Haematologica*. 2014;99(7):1138-1148. doi:10.3324/haematol.2013.098384
 97. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999;94(6):1848-1854.
 98. Hacken E ten, Gounari M, Ghia P, Burger JA. The importance of B cell receptor isotypes and stereotypes in Chronic Lymphocytic Leukemia. *Leukemia*. 2019;33(2):287-298. doi:10.1038/s41375-018-0303-x
 99. Chiorazzi N. Implications of new prognostic markers in chronic lymphocytic leukemia. *Hematol Am Soc Hematol Educ Program*. 2012;2012:76-87. doi:10.1182/asheducation-2012.1.76
 100. Hamblin TJ, Davis ZA, Oscier DG. Determination of how many immunoglobulin variable region heavy chain mutations are allowable in unmutated chronic lymphocytic leukaemia - long-term follow up of patients with different percentages of mutations. *Br J Haematol*. 2008;140(3):320-323. doi:10.1111/j.1365-2141.2007.06928.x
 101. Lin K, Sherrington PD, Dennis M, Matrai Z, Cawley JC, Pettitt AR. Relationship between p53 dysfunction, CD38 expression, and IgV(H) mutation in chronic lymphocytic leukemia. *Blood*. 2002;100(4):1404-1409. doi:10.1182/blood-2001-11-0066
 102. Davis Z, Forconi F, Parker A, et al. The outcome of Chronic lymphocytic leukaemia patients with 97% IGHV gene identity to germline is distinct from cases with <97% identity and similar to those with 98% identity. *Br J Haematol*. 2016;173(1):127-136. doi:10.1111/bjh.13940
 103. Agathangelidis A, Darzentas N, Hadzidimitriou A, et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood*. 2012;119(19):4467-4475. doi:10.1182/blood-2011-11-393694

104. Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood*. 2003;101(12):4952-4957. doi:10.1182/blood-2002-11-3485
105. Bomben R, Dal Bo M, Capello D, et al. Comprehensive characterization of IGHV3-21-expressing B-cell chronic lymphocytic leukemia: an Italian multicenter study. *Blood*. 2007;109(7):2989-2998. doi:10.1182/blood-2006-10-051110
106. Stamatopoulos B, Smith T, Crompton E, et al. The Light Chain IgLV3-21 Defines a New Poor Prognostic Subgroup in Chronic Lymphocytic Leukemia: Results of a Multicenter Study. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2018;24(20):5048-5057. doi:10.1158/1078-0432.CCR-18-0133
107. Ntoufa S, Papakonstantinou N, Apollonio B, et al. B Cell Anergy Modulated by TLR1/2 and the miR-17~92 Cluster Underlies the Indolent Clinical Course of Chronic Lymphocytic Leukemia Stereotyped Subset #4. *J Immunol Baltim Md 1950*. 2016;196(10):4410-4417. doi:10.4049/jimmunol.1502297
108. Murray F, Darzentas N, Hadzidimitriou A, 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*. 2008;111(3):1524-1533. doi:10.1182/blood-2007-07-099564
109. Agathangelidis A, Hadzidimitriou A, Minga E, et al. Reappraising Immunoglobulin Repertoire Restrictions in Chronic Lymphocytic Leukemia: Focus on Major Stereotyped Subsets and Closely Related Satellites. *Blood*. 2016;128(22):4376. doi:10.1182/blood.V128.22.4376.4376
110. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood*. 2007;109(1):259-270. doi:10.1182/blood-2006-03-012948
111. Baliakas P, Agathangelidis A, Hadzidimitriou A, et al. Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. *Blood*. 2015;125(5):856-859. doi:10.1182/blood-2014-09-600874
112. Gounari M, Ntoufa S, Apollonio B, et al. Excessive antigen reactivity may underlie the clinical aggressiveness of chronic lymphocytic leukemia stereotyped subset #8. *Blood*. 2015;125(23):3580-3587. doi:10.1182/blood-2014-09-603217
113. Del Giudice I, Chiaretti S, Santangelo S, et al. Stereotyped subset #1 chronic lymphocytic leukemia: a direct link between B-cell receptor structure, function, and patients' prognosis. *Am J Hematol*. 2014;89(1):74-82. doi:10.1002/ajh.23591
114. Rossi D, Spina V, Bomben R, et al. Association between molecular lesions and specific B-cell receptor subsets in chronic lymphocytic leukemia. *Blood*. 2013;121(24):4902-4905. doi:10.1182/blood-2013-02-486209
115. Mansouri L, Sutton LA, Ljungström V, et al. Functional loss of IκBε leads to NF-κB deregulation in aggressive chronic lymphocytic leukemia. *J Exp Med*. 2015;212(6):833-843. doi:10.1084/jem.20142009

116. Strefford JC, Sutton LA, Baliakas P, et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia*. 2013;27(11):2196-2199. doi:10.1038/leu.2013.98
117. Marincevic M, Cahill N, Gunnarsson R, et al. High-density screening reveals a different spectrum of genomic aberrations in chronic lymphocytic leukemia patients with “stereotyped” IGHV3-21 and IGHV4-34 B-cell receptors. *Haematologica*. 2010;95(9):1519-1525. doi:10.3324/haematol.2009.021014
118. Rossi D, Spina V, Cerri M, et al. Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2009;15(13):4415-4422. doi:10.1158/1078-0432.CCR-08-3266
119. Langerak AW, Davi F, Ghia P, et al. Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. *Leukemia*. 2011;25(6):979-984. doi:10.1038/leu.2011.49
120. van Dongen JJM, Langerak AW, Brüggemann 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(12):2257-2317. doi:10.1038/sj.leu.2403202
121. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res*. 2008;36(Web Server issue):W503-508. doi:10.1093/nar/gkn316
122. Sayers EW, Beck J, Bolton EE, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2021;49(D1):D10-D17. doi:10.1093/nar/gkaa892
123. Stamatopoulos K, Belessi C, Hadzidimitriou A, et al. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood*. 2005;106(10):3575-3583. doi:10.1182/blood-2005-04-1511
124. Minici C, Gounari M, Übelhart R, et al. Distinct homotypic B-cell receptor interactions shape the outcome of chronic lymphocytic leukaemia. *Nat Commun*. 2017;8:15746. doi:10.1038/ncomms15746
125. Maity PC, Bilal M, Koning MT, et al. IGLV3-21*01 is an inherited risk factor for CLL through the acquisition of a single-point mutation enabling autonomous BCR signaling. *Proc Natl Acad Sci U S A*. 2020;117(8):4320-4327. doi:10.1073/pnas.1913810117
126. Lane DP. p53, guardian of the genome. *Nature*. 1992;358(6381):15-16. doi:10.1038/358015a0
127. Hainaut P. TP53: Coordinator of the Processes That Underlie the Hallmarks of Cancer. In: Hainaut P, Olivier M, Wiman KG, eds. *P53 in the Clinics*. Springer; 2013:1-23. doi:10.1007/978-1-4614-3676-8_1
128. Kaiser AM, Attardi LD. Deconstructing networks of p53-mediated tumor suppression in vivo. *Cell Death Differ*. 2018;25(1):93-103. doi:10.1038/cdd.2017.171
129. Powell DJ, Hrstka R, Candeias M, Bourougaa K, Vojteseka B, Fåhræus R. Stress-dependent changes in the properties of p53 complexes by the alternative translation product p53/47. *Cell Cycle*. 2008;7(7):950-959. doi:10.4161/cc.7.7.5626

130. Harris SL, Levine AJ. The p53 pathway: positive and negative feedback loops. *Oncogene*. 2005;24(17):2899-2908. doi:10.1038/sj.onc.1208615
131. Kruiswijk F, Labuschagne CF, Vousden KH. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nat Rev Mol Cell Biol*. 2015;16(7):393-405. doi:10.1038/nrm4007
132. Chen Y, Dey R, Chen L. Crystal Structure of the p53 Core Domain Bound to a Full Consensus Site as a Self-Assembled Tetramer. *Structure*. 2010;18(2):246-256. doi:10.1016/j.str.2009.11.011
133. Lazarian G, Cymbalista F, Baran-Marszak F. Impact of Low-Burden TP53 Mutations in the Management of CLL. *Front Oncol*. 2022;12:841630. doi:10.3389/fonc.2022.841630
134. Khoury MP, Bourdon JC. The Isoforms of the p53 Protein. *Cold Spring Harb Perspect Biol*. 2010;2(3):a000927. doi:10.1101/cshperspect.a000927
135. Yu H, Chen JK, Feng S, Dalgarno DC, Brauer AW, Schrelber SL. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell*. 1994;76(5):933-945. doi:10.1016/0092-8674(94)90367-0
136. The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific PIG3 gene transactivation and with transcriptional repression. *EMBO J*. 1998;17(16):4668-4679. doi:10.1093/emboj/17.16.4668
137. Chillemi G, Kehrlöesser S, Bernassola F, et al. Structural Evolution and Dynamics of the p53 Proteins. *Cold Spring Harb Perspect Med*. 2017;7(4):a028308. doi:10.1101/cshperspect.a028308
138. Vieler M, Sanyal S. p53 Isoforms and Their Implications in Cancer. *Cancers*. 2018;10(9):288. doi:10.3390/cancers10090288
139. Jorujiz SM, Bourdon JC. p53 Isoforms: Key Regulators of the Cell Fate Decision. *Cold Spring Harb Perspect Med*. 2016;6(8):a026039. doi:10.1101/cshperspect.a026039
140. Bourdon JC, Fernandes K, Murray-Zmijewski F, et al. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev*. 2005;19(18):2122-2137. doi:10.1101/gad.1339905
141. Haaland I, Hjelle SM, Reikvam H, et al. p53 Protein Isoform Profiles in AML: Correlation with Distinct Differentiation Stages and Response to Epigenetic Differentiation Therapy. *Cells*. 2021;10(4):833. doi:10.3390/cells10040833
142. Courtois S, Verhaegh G, North S, et al. Δ N-p53, a natural isoform of p53 lacking the first transactivation domain, counteracts growth suppression by wild-type p53. *Oncogene*. 2002;21(44):6722-6728. doi:10.1038/sj.onc.1205874
143. Weinberg RL, Veprintsev DB, Fersht AR. Cooperative Binding of Tetrameric p53 to DNA. *J Mol Biol*. 2004;341(5):1145-1159. doi:10.1016/j.jmb.2004.06.071
144. Maier B, Gluba W, Bernier B, et al. Modulation of mammalian life span by the short isoform of p53. *Genes Dev*. 2004;18(3):306-319. doi:10.1101/gad.1162404
145. Aoubala M, Murray-Zmijewski F, Khoury MP, et al. p53 directly transactivates Δ 133p53 α , regulating cell fate outcome in response to DNA damage. *Cell Death Differ*. 2011;18(2):248-258. doi:10.1038/cdd.2010.91

146. Xie N, Chen M, Dai R, et al. SRSF1 promotes vascular smooth muscle cell proliferation through a $\Delta 133p53/EGR1/KLF5$ pathway. *Nat Commun*. 2017;8(1):16016. doi:10.1038/ncomms16016
147. Cancer-specific mutations in p53 induce the translation of $\Delta 160p53$ promoting tumorigenesis. *EMBO Rep*. 2016;17(11):1542-1551. doi:10.15252/embr.201541956
148. Kim S, An SSA. Role of p53 isoforms and aggregations in cancer. *Medicine (Baltimore)*. 2016;95(26):e3993. doi:10.1097/MD.0000000000003993
149. Aho AV, Corasick MJ. Efficient string matching: an aid to bibliographic search. *Commun ACM*. 1975;18(6):333-340. doi:10.1145/360825.360855
150. IgBLAST: an immunoglobulin variable domain sequence analysis tool | Nucleic Acids Research | Oxford Academic. Accessed June 20, 2022. <https://academic.oup.com/nar/article/41/W1/W34/1097536>
151. Lefranc MP, Giudicelli V, Kaas Q, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res*. 2005;33(Database issue):D593-597. doi:10.1093/nar/gki065
152. Designing Synthetic MLPA Probes - Downloads / MLPA / Synthetic Probe Design - MRC Holland Support. Accessed June 8, 2022. <https://support.mrcholland.com/downloads/files/designing-synthetic-mlpa-probes>
153. Jardin F, Ruminy P, Kerckaert JP, et al. Detection of somatic quantitative genetic alterations by multiplex polymerase chain reaction for the prediction of outcome in diffuse large B-cell lymphomas. *Haematologica*. 2008;93(4):543-550. doi:10.3324/haematol.12251
154. Jardin F, Picquenot JM, Parmentier F, et al. Detection of gene copy number aberrations in mantle cell lymphoma by a single quantitative multiplex PCR assay: clinicopathological relevance and prognosis value. *Br J Haematol*. 2009;146(6):607-618. doi:10.1111/j.1365-2141.2009.07791.x
155. Li Y ze, Xu J jie, Qian H zhu, et al. High prevalence of HIV infection and unprotected anal intercourse among older men who have sex with men in China: a systematic review and meta-analysis. *BMC Infect Dis*. 2014;14:531. doi:10.1186/1471-2334-14-531
156. Zheng X, Yu S, Long J, et al. Comparison of the clinical characteristics of primary thyroid lymphoma and diffuse sclerosing variant of papillary thyroid carcinoma. *Endocr Connect*. 2022;11(1):e210364. doi:10.1530/EC-21-0364
157. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016;66(1):7-30. doi:10.3322/caac.21332
158. Delgado J, Doubek M, Baumann T, et al. Chronic lymphocytic leukemia: A prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *Am J Hematol*. 2017;92(4):375-380. doi:10.1002/ajh.24660
159. Oscier DG, Gardiner AC, Mould SJ, et al. Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood*. 2002;100(4):1177-1184.

160. Morabito F, Shanafelt TD, Gentile M, et al. Immunoglobulin heavy chain variable region gene and prediction of time to first treatment in patients with chronic lymphocytic leukemia: Mutational load or mutational status? Analysis of 1003 cases. *Am J Hematol.* 2018;93(9):E216-E219. doi:10.1002/ajh.25206
161. Rosenquist R, Ghia P, Hadzidimitriou A, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia.* 2017;31(7):1477-1481. doi:10.1038/leu.2017.125
162. Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest.* 1998;102(8):1515-1525. doi:10.1172/JCI3009
163. Tobin G, Thunberg U, Karlsson K, et al. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood.* 2004;104(9):2879-2885. doi:10.1182/blood-2004-01-0132
164. Nadeu F, Royo R, Clot G, et al. IGLV3-21R110 identifies an aggressive biological subtype of chronic lymphocytic leukemia with intermediate epigenetics. *Blood.* 2021;137(21):2935-2946. doi:10.1182/blood.2020008311
165. Davi F, Langerak AW, de Septenville AL, et al. Immunoglobulin gene analysis in chronic lymphocytic leukemia in the era of next generation sequencing. *Leukemia.* 2020;34(10):2545-2551. doi:10.1038/s41375-020-0923-9
166. Stamatopoulos B, Timbs A, Bruce D, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia.* 2017;31(4):837-845. doi:10.1038/leu.2016.307
167. Giudice ID, Foà R. Another step forward in the 20-year history of IGHV mutations in chronic lymphocytic leukemia. *Haematologica.* 2019;104(2):219-221. doi:10.3324/haematol.2018.207399
168. Ghia P, Stamatopoulos K, Belessi C, et al. ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia.* 2007;21(1):1-3. doi:10.1038/sj.leu.2404457
169. Hadzidimitriou A, Darzentas N, Murray F, et al. Evidence for the significant role of immunoglobulin light chains in antigen recognition and selection in chronic lymphocytic leukemia. *Blood.* 2009;113(2):403-411. doi:10.1182/blood-2008-07-166868
170. Pozzo F, Bittolo T, Vendramini E, et al. NOTCH1-mutated chronic lymphocytic leukemia cells are characterized by a MYC-related overexpression of nucleophosmin 1 and ribosome-associated components. *Leukemia.* 2017;31(11):2407-2415. doi:10.1038/leu.2017.90
171. Wang L, Brooks AN, Fan J, et al. Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia. *Cancer Cell.* 2016;30(5):750-763. doi:10.1016/j.ccell.2016.10.005
172. Zenz T, Benner A, Döhner H, Stilgenbauer S. Chronic lymphocytic leukemia and treatment resistance in cancer: the role of the p53 pathway. *Cell Cycle Georget Tex.* 2008;7(24):3810-3814. doi:10.4161/cc.7.24.7245

173. Klein U, Lia M, Crespo M, 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):28-40. doi:10.1016/j.ccr.2009.11.019
174. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 2011;44(1):47-52. doi:10.1038/ng.1032
175. Bouaoun L, Sonkin D, Ardin M, et al. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. *Hum Mutat*. 2016;37(9):865-876. doi:10.1002/humu.23035
176. Kasthuber ER, Lowe SW. Putting p53 in Context. *Cell*. 2017;170(6):1062-1078. doi:10.1016/j.cell.2017.08.028
177. Joruz SM, Bourdon JC. p53 Isoforms: Key Regulators of the Cell Fate Decision. *Cold Spring Harb Perspect Med*. 2016;6(8):a026039. doi:10.1101/cshperspect.a026039
178. Bourdon JC, Fernandes K, Murray-Zmijewski F, et al. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev*. 2005;19(18):2122-2137. doi:10.1101/gad.1339905
179. Anensen N, Oyan AM, Bourdon JC, Kalland KH, Bruserud O, Gjertsen BT. A distinct p53 protein isoform signature reflects the onset of induction chemotherapy for acute myeloid leukemia. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2006;12(13):3985-3992. doi:10.1158/1078-0432.CCR-05-1970
180. Gocke CD, Gladstone DE. The absolute percent deviation of IGHV mutation rather than a 98% cut-off predicts survival of chronic lymphocytic leukaemia patients treated with fludarabine, cyclophosphamide and rituximab. *Br J Haematol*. 2018;180(1):7-8. doi:10.1111/bjh.15015