MASTER THESIS

STUDY OF IGH SEQUENCES IN LOW-COUNT MONOCLONAL B-CELL LYMPHOCYTOSIS

José Manuel Serrano Lozano

Biology and Clinic of Cancer Master's Degree
Academic Year 2021/22

TUTOR: Prof. Dr. Jose Alberto Orfao De Matos Correira E Vale
CO-TUTOR: Dra. Arancha Rodríguez Caballero

University of Salamanca | Laboratory 11, Cancer Research Center

El Doctor D. Jose Alberto Orfao De Matos Correira E Vale, Catedrático del Departamento de Medicina de la Universidad de Salamanca y la Doctora Dña. María Aránzazu Rodríguez Caballero, investigadora posdoctoral del Laboratorio 11 del Centro de Investigación del Cáncer.

Certifican que:

El Trabajo de Fin de Máster realizado por el estudiante José Manuel Serrano Lozano del Máster Universitario en Biología y Clínica del Cáncer titulado **"Study of IGH sequences in low-count monoclonal B-cell lymphocytosis"** ha sido realizado bajo su dirección en el Laboratorio 11 del Centro de Investigación del Cáncer de la Universidad de Salamanca y reúne, a su juicio, originalidad y contenidos suficientes para que sea presentado ante el tribunal correspondiente.

Y para que conste, y a los efectos oportunos, expiden y firman el presente certificado a 10 de julio de 2022.

D. Jose Alberto Orfao De Matos Correira E Vale

RODRIGUEZ CABALLERO MARIA ARANTZAZU - 07977918T

Firmado digitalmente por RODRIGUEZ CABALLERO MARIA ARANTZAZU - 07977918T
Fecha: 2022.07.12 12:03:04 +02'00'

Dña. María Aránzazu Rodríguez Caballero

INDEX

ABSTRACT

Low-count monoclonal B-cell lymphocytosis (MBL^{lo}) is defined by the presence of $<0.5 \times 10^9$ clonal B-cells/L in blood in the absence of any other symptoms or signs of a neoplastic lymphoproliferative disorder. It is still unclear whether MBL^{lo} represents a pre-malignant state prior to chronic lymphocytic leukemia (CLL) or it relates to a physiological immunosenescent process. The molecular and biochemical characteristics of the B-cell receptor (BCR) including the specific immunoglobulin heavy-chain (IGH) sequences have been related to disease behavior in CLL. In MBL^{lo}, few studies have been reported of *IGH* sequences on the clonal B-cells, mostly due to their low counts in blood.

Here, we set up a workflow adapted to the low MBL^{lo} clonal B-cell numbers aimed at more efficiently characterize genetically and molecularly the IGHV rearrangements of the BCR of MBL^{lo} cases. We also compare genetic and molecular characteristics of the BCR of MBL^{lo} vs MBL^{hi} and CLL B-cell clones, and their impact on the behavior of MBL^{lo} clonal B-cells during follow-up.

MBL^{lo} clonal B-cells were isolated from blood of otherwise healthy individuals by high-sensitive flow cytometry. Two different multiplex PCR protocols for *IGHV* sequencing were applied, depending on the number of MBL^{lo} clonal B-cells available, using a threshold of 50,000 cells for protocol selection.

The new workflow set up allowed us to increase the number of MBL^{lo} cases included in out cohort (N=34), for subsequent comparison of *IGHV* sequences to two MBL^{hi} (N=73) and CLL (N=138) cohorts. The *IGHV3-7* gene sequence was more frequently found in MBL^{lo} subjects, whereas *IGHV3-23* was highly represented in MBL^{hi} patients. No significant differences were reported in D segments, while *IGHJ4* was significantly more frequent in MBL^{lo} and *IGHJ6* in CLL. Compared to other J-gene segments, clonal B-cells with *IGHJ6* show a more immature origin and present longer HCDR3 sequences. Nevertheless, this BCR characteristic showed no significant impact in the behavior of the MBL^{lo} clone during follow-up; however a tendency towards a greater size of MBL^{lo} clones was observed among *IGHJ6* cases.

The new workflow here proposed proved to be a useful tool to increase the efficiency of the characterization of the BCR sequences of MBL^{lo} cases with low clonal B-cells numbers revealing unique BCR sequences vs MBL^{hi} and CLL.

INTRODUCTION

Chronic lymphocytic leukemia (CLL) is defined as a B-cell lymphoproliferative disorder characterized by the presence of $>5 \times 10^9$ clonal B-cells/L in blood with a unique $CD19^+$, $CD20^{lo}$, $CD5^+$, $CD23^+$, $CD27^+$, $CD200^+$, $sIgM^{-/lo}$ immunophenotype and light chain restriction pattern. CLL represents ≈30% of all adult leukemias in Western countries, thus being the most prevalent hematological malignancy in this population. The mean of age at diagnosis of CLL is of 70 years, with a slight predominance in men vs women[1] and a 5-year survival rate of 84%[2], but a highly heterogenous clinical course and outcome, with a life expectancy that ranges from a few months to several decades. Despite the great advances in the early detection of the disease and the emergence of novel therapies in recent years, CLL remains an incurable disease[3].

After birth, normal B-cells are produced from hematopoietic stem cells in the bone marrow. After they are produced and matured, naïve B-lymphocytes reach the blood and secondary lymphoid tissues, expressing a non-autoreactive B-cell receptor (BCR)[4]. During maturation of B-cells, different stages can be defined depending on the status of rearrangement of VDJ/VJ heavy/light chain immunoglobulin genes (*IGH*/*IGL*)[5].

In bone marrow, pre-B-cells and immature B-cells undergo a positive and negative selection based on which only those B-cells that have a BCR that can signal inside the cell will survive (positive selection), only if they subsequently do not recognize self-antigens (negative selection)[6]. Naïve B-lymphocytes can recognize specific antigens in secondary lymphoid tissues -e.g. lymph nodes, spleen and mucosal-associated lymphoid tissues (MALTs)- and proliferate in the germinal centers[7,8]. In the germinal center, activated B-cells undergo a process known as somatic hypermutation (SHM), in order to produce Igs (BCR and antibodies) with higher affinity for the recognized antigen (i.e. maturation of BCR affinity). In parallel, activated B-lymphocytes may undergo Ig class switch, through which the firstly expressed IgM receptor are replaced by an IgG, IgA and, less frequently, IgE-based BCR [9–11]. Subsequently, activated B-cells differentiate to effector B-cells (i.e., plasma cells and memory B-cells), which are responsible for antibody production and for maintaining the pool of memory B-cells specific for previously encountered antigens. At present, the normal B-cell counterpart of CLL cells, still remains a matter of debate. However, evidences exist which point out towards two different normal B-cell compartments, that include pre-germinal center naïve B-cells and post-germinal center memory B -cells[12,13], depending on the mutational status of the BCR of CLL cells, respectively (unmutated vs mutated).

From the prognostic point of view, the mutational status of the variable region of the *IGH* gene and the hydropathy index of the heavy-chain complementarity-determinant region 3 (HCDR3) -quantified using the GRAVY score (GS)- have been both identified as prognostic markers for disease progression and patient survival[14,15]. Thus, based on the *IGHV* mutational status of BCR of CLL cells, two CLL patient risk groups can be identified: those with clonal B-cells that express a mutated BCR (M-CLL: ≥2% difference with germline IGHV sequence) and those with an unmutated BCR (U-CLL: <2% difference with germline IGHV sequence)[16]. From a pathogenic point of view, clonal B-cells that express a mutated IGHV sequence had undergone somatic hypermutation in the germinal centers and they are associated with more stable disease, whereas those that have not suffered this process express an unmutated IGHV sequence in their BCR, associated with a poorer prognosis (lower time to treatment and shortened overall survival vs M-CLL)[17]. Depending on the HCDR3 hydrophobicity index, M-CLL can be further divided into two subgroups: cells presenting negatively charged HCDR3 sequences and those with a neutral HCDR3 aminoacidic sequence. Patients with neutral HCDR3 sequences tend to have a higher rate of (and shorter time to) disease progression, in addition to a higher frequency of genetic alterations and shorter time to therapy, compared with those with negatively charged HCDR3 sequences[15].

Overall, the BCR IGHV sequences of CLL lymphocytes display a limited repertoire compared to normal B-cells of age-matched healthy subjects. Thus, some specific IGHV gene rearrangements are more frequently found in CLL than in normal B-cells from individuals of the general population, including "stereotyped" BCR with highly similar or even identical aminoacidic sequences in their HCDR3 regions[19]. Altogether, these findings could point out the potential involvement of specific antigens and the chronic stimulation of B-cells in the ontogeny of CLL, where the BCR would play a key role in the expansion of clonal B-cells [20,21].

Nowadays, it is well-accepted that every CLL is preceded by a silent state termed monoclonal B-cell lymphocytosis (MBL) [22], defined by the presence of $< 5 \times 10^9$ circulating clonal B-cells/L in blood, in the absence of any other symptoms or signs (e.g. hepatomegaly, splenomegaly or lymphadenopathy) of a mature peripheral B-cell malignant disease[23]. Depending on the number of clonal B-cells, MBL is divided into high-count MBL (MBL^{hi}) when there are $\geq 0.5 \times 10^9$ clonal B-cells/L and low-count MBL (MBL^{lo}) when $<0.5 \times 10^9$ clonal B-cells are identified. Progression from MBL^{hi} to CLL has been reported, at a rate of 1-2% of subjects per year, while for MBL^{lo} the rate of progression to MBL^{hi} would be of around 0.2%/year[22]. According to the immunophenotype of clonal B-cells, MBL^{lo} and MBL^{hi} cases can be further classified into: CLL-like MBL when clonal cells are phenotypically identical to typical CLL malignant cells, atypical CLL-like MBL characterized by a $CD5^+$, $CD20^{hi}$ and/or $CD23^-$ phenotype and non-CLL-like MBL, when clonal cells are $CD5^-$ independently of the pattern of expression of other markers [24,25].

Since individuals presenting with MBL^{lo} do not show any symptom or sign of disease and they present with very low numbers of clonal B-cells in blood, they are typically identified when high-sensitive flow cytometry methods are used to screen for this condition in the general population. Thus, the incidence of MBL^{lo} depends on factors such as the sensitivity of the technique used for its detection, the age of the population studied and the geographical area of study. Based on 2-4 color flow cytometry and the measurement of <500,000 cells per sample, early studies showed frequencies of MBL^{lo} in the general population of between 0.3% and 6.4%, [26–28]. In a more recent study based on a higher sensitivity, Nieto et al. found a prevalence in the general population of up to 15% of adults aged >40 years [29]. In fact, they hypothesized that clonal MBL cells might be found in virtually every subject >70 years, if enough volume of blood (>50 mL) would have been analyzed per sample [30].

Previous studies indicate that MBL^{lo} is associated with an immunodeficiency state due to an impaired humoral response, similar to what has been reported in CLL[31,32]. This results in a shortened life expectancy, due to a higher death rate because of (severe) infections and cancer[31]. These findings are consistent with previous studies that also showed higher rates and more severe infections in MBL^{hi} and CLL patients, among whom they represent one of the main causes of death[33,34].

Despite except for familial CLL, no progression from sporadic MBL^{lo} to CLL has been reported yet, it is hypothesized that those will be seen when cohorts with longer follow-up are studied. However, the precise mechanisms that drive the expansion of clonal B-cells and progression from MBL to CLL (or another lymphoproliferative disorder) remain unclear. Kikshuge et al. proposed a model to explain the relationship between MBL and CLL[35], which is based on their demonstration that differentiation towards the B lymphoid lineage increases after a stem cell transplant from CLL patients to immunodeficient mice. They proposed the existence of genetic alterations in hematopoietic cells, which lead to an oligoclonal expansion of B-cells with similar characteristics to CLL cells[35]. From this oligoclonal pool of distinct B-cell clones, a dominant B-cell clone may preferentially expand due to the acquisition of additional alterations, giving rise to a hematological malignancy. However, they fail to establish whether the specific factors involved in this clonal selection and expansion are exogenous (e.g. antigen stimuli, microenvironment)

and/or endogenous (e.g. mutations in cancer driver genes and other molecular/genetic alterations).

Characterization of the BCR of clonal MBL^{lo} B-cells at the molecular and genetic level currently remains challenging. This is due to the low number of aberrant clonal B-cells present in blood of most MBL^{lo} subjects and its low relative amount with respect to background polyclonal B-cells, including the co-existence of several MBL^{lo} clones in many individuals.

Despite the limited number of molecular studies about the BCR of MBL^{lo} clones, a potentially higher frequency of mutated IGHV sequences compared to CLL (53%) has been reported [36,37]. However, the specific incidence of MBL^{lo} with mutated IGHV genes varies among different cohorts. Vardi et al., in a study based on a pool of subjects from different geographical areas, reported that 75% of MBL^{lo} cases show mutated IGHV sequences[37]. Similarly, Henriques et al. also reported a higher frequency (vs CLL) of 67% mutated IGHV sequences in MBL^{lo} cases in a single geographical area in Spain [36]. Regarding the distribution of the *IGHV* gene repertoire, also some bias has been reported in MBL^{lo} with non-random usage of specific *IGHV* gene rearrangements, different from those more frequently found in CLL[36,38] and MBL^{hi} clonal B-cells[37]. These include a preferential usage for IGHV3-23 genes in MBL^{lo} together with a lower frequency of IGHV4-34 and IGHV1-69 genes compared to CLL [36,37]. Furthermore, stereotyped BCRs (HCDR3 sequences which are highly similar or even identical) are reported to be frequent in CLL, but not in MBL^{lo} [39]. Based on the available data and despite the high heterogeneous distribution of the reported BCR features in MBL^{lo}, MBL^{hi} and CLL, overall MBL^{hi} appears to be somewhat similar to CLL, whereas MBL^{lo} forms a separate group with distinct BCR molecular features[40].

The non-random distribution of the IGHV gene rearrangement repertoire would further support the hypothesis that clonal B-cell expansions in MBL and CLL may be driven by chronic stimulation of the BCR via specific antigen triggers. However, differences between the IGHV gene repertoire of the MBL^{lo} subjects and both MBL^{hi} and CLL cases[41] also point out that BCR stimulation might play a different role at different stages of the disease. Despite all the above, other BCR-associated features other than IGHV repertoire such as the HCDR3 hydrophobicity index[15] might also play a role in progression of MBL^{lo} to MBL^{hi} and CLL, which deserves further investigations.

HYPOTHESIS

At present it is well established that every chronic lymphocytic leukemia (CLL) is preceded by a silent state called monoclonal B-cell lymphocytosis (MBL), which can be divided into high-count MBL (MBL^{hi}) or low-count MBL (MBL^{lo}) depending on the number of clonal B-cells present in blood. Whereas MBL^{hi} progress to CLL at a rate of 1-2% per year, MBL^{lo} progress to MBL^{hi} at a lower rate. Previous studies provide evidences on the existence of an association between the genetic and molecular features of the BCR (including the mutational status of IGHV sequences and the HCDR3 hydropathy index) and both disease progression and the clinical outcome of CLL patients.

Here, we hypothesize that a fraction of MBL^{lo} clonal-B-cells might present *IGHV* sequences that share molecular and genetic characteristics commonly observed in MBL^{hi} and CLL, including poor-prognosis CLL patients. In order to accurately assess the molecular features of BCR of the MBL^{lo} clones, the methodology used for the molecular characterization of the BCR in MBL^{hi} and CLL patients requires optimization adapted to the lower (relative and absolute) numbers of clonal B-cells that can be obtained from blood of MBL^{lo} cases.

OBJECTIVES

Here, we set up an approach and a workflow to characterize BCR *IGH* sequences of MBL^{lo} clonal B-cells. Subsequently, we compared the BCR features of MBL^{lo} cells with those of B-cells from MBL^{hi} and CLL patients and investigated their potential association with the behavior of MBL^{lo} clones during follow-up. For this purpose, three specific objectives were pursued in this study:

1) To set up a methodological work-flow for the molecular characterization of *IGHV* gene rearrangements of the BCR adapted to the small B-cell clones from subjects presenting with MBL^{lo}.

2) To determine the similarities and differences of the molecular and biochemical characteristics of the BCR and its corresponding *IGHV* sequences of MBL^{lo} vs MBL^{hi} and CLL B-cell clones.

3) To investigate the impact of the molecular features of the BCR of MBL^{lo} clones on their behavior during follow-up.

MATERIAL AND METHODS

Subjects and samples

A total of 34 CLL-like MBL^{lo} subjects diagnosed according to the current WHO02/2022 criteria (persistent $<0.5x10^9$ clonal B-cells/L in blood in the absence of other signs or symptoms of a lymphoproliferative disorder) were included in this study[23]. Diagnosed MBL^{lo} subjects were followed at the Hematology Service of the University Hospital of Salamanca (Salamanca, Spain). From them, 25 had been previously characterize for their *IGHV* sequence.

As controls, a cohort of 73 MBL^{hi} and 138 CLL patients previously characterized were also studied. Clinical and laboratory data on these subjects had been previously obtained and recorded.

The study was approved by the local institutional Ethics Committee. Prior to entering the study each participant gave his/her written informed consent to participate, in agreement with the Declaration of Helsinki.

Clonal B-cells isolated from peripheral blood were FACS-purified and stored in PBS at -80ºC locally, and subsequently at the Spanish National DNA Bank (Carlos III, University of Salamanca, Salamanca, Spain) until analyzed.

Purification of clonal B-cells

EDTA-anticoagulated peripheral blood samples were obtained from each participant and stained with a panel of monoclonal antibodies that included CD5, CD10, CD19, CD20, CD23, CD27, CD38, CD79b, CD81, CD200, sIgKAPPA, sIglambda and sIgM markers. Subsequent immunophenotyping by high-sensitive multiparameter flow cytometry was performed, as previously described[36]. Stained samples were measured in FACSCanto II and BD LSRFortessa™ X-20 flow cytometers (BD Biosciences, San José, CA), calibrated according the EuroFlow Consortium instrument set up and calibration protocol available at www.euroflow.org. For flow cytometry data analysis, the INFINICYT™ software (Cytognos S.L., Salamanca, Spain) was used. Clonal B-cells present in blood were sorted in a fluorescent-activated cell sorter (FACS) FACSAria III flow cytometer (BD). Prior to sorting, CLL-like clonal B-cells were identified by its unique aberrant phenotype: $CD19^+$, $CD20^{lo}$, $CD5^{+/++}$, $CD23^+$, $CD27^+$, $CD200^+$, $sIgM^{-/lo}$ and either κ/λ light chain restriction.

Isolation of genomic DNA

Two different methods were used to purify genomic DNA from purified clonal B-cells, depending on the number of cells available for molecular analysis. Whenever >50,000 purified clonal B-cells were available, DNA was purified using the Puregene® Genomic DNA Purification Kit (Qiagene, Hilden, Germany). Briefly, purified clonal B-cells were lysed using the Cell Lysis Solution (Qiagene), an anionic detergent with DNA stabilizers, which limit the action of intra and extracellular DNases. Afterward, DNA was precipitated with 100% isopropanol containing 0.5 µL of glycogen per 300 µL of isopropanol in order to improve the yield of the DNA purification step. Then, the extracted DNA was washed using 70% ethanol and subsequently hydrated with TE1X Low-EDTA. Purified DNA was stored at -20ºC for subsequent studies.

In case there were <50,000 cells, genomic DNA was extracted using the Extract-N-Amp™ Blood PCR Kit (Sigma-Aldrich, St Louis, MO) protocol for DNA extraction followed by DNA amplification with slight modifications to improve the yield after the cell lysing and subsequent PCR amplification steps. Cells were lysed with 5 µL of Cell Lysis

Solution (Sigma Aldrich) for 5 minutes (room temperature); 45 µL of the Neutralization Solution (Sigma-Aldrich) was then added to stop further cell lysis and DNA degradation.

The Extract-N-Amp™ PCR ReadyMix™ for Blood reagent (Sigma-Aldrich) was used to amplify *IGHV* gene rearrangements in these latter samples. This mix contains buffer, salts, dNTPs and a *Taq* polymerase, as well as a JumpStart™ antibody, which blocks the polymerase at mild temperatures (35-40 ºC) and allows the PCR amplification step to start over 75ºC (hot-start PCR), in order to improve the specificity of the PCR amplification reaction[42].

IGHV gene rearrangement analysis

A multiplex PCR (polymerase chain reaction) was performed with purified DNA from clonal B-cells isolated from blood of MBL^{lo} subjects, to amplify the VDJs rearrangements of the *IGHV* genes. The PCR product was sequenced by Sanger sequencing in an ABI PRISM®3100 Genetic Analyzer sequencer (Applied Biosystems, Foster City, CA) and analyzed using the 4Peaks v1.8 Software program (Griekspoor and Groothuis; http://nucleobytes.com/index.php/4peaks).

Primers used were designed by the BIOMED consortium and the EuroClonality-NGS working group[43,44], keeping the possibility of using them in a multiplex format (i.e. similar melting temperature). Initially, IGHV, IGHD and IGHJ DNA sequences were amplified in two different multiplex PCRs, with fifteen forward primers -which anneal to different IGHD segments- grouped in two different sets of primers depending on their concentration (0.1 µM and 0.2 µM, respectively). Two consensus reverse primers mixed in a single set of primers were used with both forward sets.

PCR protocols were different for the different DNA purification methods. In case of precipitation with isopropanol and glycogen (>50,000 cells per sample), the PCR protocol described by Brüggemann et al. was followed[44]. When the Extract-N-Amp™ Blood PCR Kit was used to extract DNA (<50,000 clonal B-cells per sample), the PCR protocol recommended by the manufacturer was employed, except for the melting temperature that was adjusted to the primers used in this study (Table 1).

TABLE 1 | *PCR protocol for DNA samples prepared with the Extract-N-Amp™ Blood PCR Kit.*

	Initial	PCR cycles (N=40)			Final
	denaturation	Denaturation	Annealing	Extension	Extension
Temperature (ºC)	95	95	63	72	72
Time (s)	180	45	45	90	600

Results expressed in seconds

PCR products were purified using ExoProStart (GE Healthcare, Buckinghamshire, UK) and sequenced. Sequence analysis was performed using bioinformatic tools available at the web service of the European Bioinformatics Institute (EMBL-EBI) and the IMGT/V-Quest tool (http://www.imgt.org). Classification between mutated or unmutated *IGHV* sequences was based on previously defined criteria, using a cut-off value of 98% identity with the germline sequence[16]. In order to calculate the Gravy Score (GS), the ProtScale Tool from the ExPASy Bioinformatic Resource Portal was used (https://web.expasy.org/protscale/). Sequence analysis was performed following the ERIC protocols for immunoglobulin gene sequencing analysis[45].

Statistical methods

Conventional statistical analysis, were performed using SPSS 26.0 software (SPSS-IBM, Armonk, NY). The Kruskal-Wallis and Mann-Whitney *U* non-parametric tests were used to determine the statistical significance of differences among groups for continuous variables, whereas the Pearsons's chi-squared test and the Fisher's exact test were used for the comparisons involving categorical variables. Statistical significance was set at *p*-values <0.05.

RESULTS

1. Analysis of BCR sequences in MBL^{lo} cohort

Sensitivity of the primers. Sensitivity of forward primers was first tested. In order to determine the minimum quantity of DNA needed per sample based on which all primers were able to anneal and a specific PCR amplification could be carried out, decreasing quantities of clonal B-cell DNA ranging from 200 to 5 ng, were amplified in the presence of each of the fifteen forward primers in combination with both reverse consensus primers: 10 ng was the minimum amount of DNA with which all primers showed a specific PCR product (Table 2).

Afterward, sensitivity was tested in a multiplex PCR format for the Extract-N-Amp™ PCR Kit. In this multiplex PCR the results with decreasing number of cells (range: 5,000 to 100 cells) were also tested, *IGHV* gene rearrangement sequences being obtained with down to 200 cells assay (Table 2).

Primer specificity. We wanted to clarify if all the fifteen forward primers were able to amplify their respective rearrangements in a multiplex PCR context with no preferential amplification, comparing both protocols of genomic DNA extraction: purification with isopropanol-glycogen (high number of cells per sample) and cell lysis with Extract-N-Amp™ PCR Kit (low number of cells per sample). One previously characterized sample was amplified with both forward sets (0.1 µM and 0.2 µM) and sequenced afterwards, obtaining a productive rearrangement with primers corresponding to the same *IGHV* rearrangement previously identified for the sample (Table 3).

TABLE 2 | *Assay sensitivity for the different sets of primers used in an individual format or a multiplex format*

	Primer identification	Primer sequence	IGHV rearrangement	Individual format with Brüggeman et al. protocol[44] (10 ng DNA =1,667 cells)	Multiplex format with Extract-N-Amp™ PCR Kit (200 cells=1.2 ng DNA)
SET 0,1 µM	B1	GCAGTCTGGAGCAGAGGTGAAAA	5-51*01	✓	
	E1	GAGGTGCAGCTGTTGGAGTC	3-23*01	✓	
	G1	CAGTGGGGCGCAGGACTGTT	4-34*01	✓	
	H1	CCAGGACTGGTGAAGCCTCC	4-4*01	✓	
	K1	CCTCAGTGAAGGTTTCCTGCAAGG	1-45*01	✓	✓
	L1	AAACCCACAGAGACCCTCACGCTGAC	2-26*01	✓	
	M1	CTGGGGGGTCCCTGAGACTCTCCTG	3-16*01	✓	
	N1	CTTCACAGACCCTGTCCCTCACCTG	4-31*01	✓	
	O1	TCGCAGACCCTCTCACTCACCTGTG	6-1*01	✓	
SET REVERSE	JA1	CTTACCTGAGGAGACGGTGACC	IGHJ6*01		
	JB1	CTCACCTGAGGAGACAGTGACC	IGHJ2*01		
SET 0,2 µM	A1	CTGGGGCTGAGGTGAAGAAG	1-2*01	✓	
	C1	TCACCTTGAAGGAGTCTGGTCC	2-5*01	✓	
	D1	AGGTGCAGCTGGTGGAGTC	3-7*01	✓	✓
	F1	CCAGGACTGGTGAAGCCTTC	4-4*01	✓	
	I1	GTACAGCTGCAGCAGTCAGG	6-1*01	✓	
	J1	GCTGGTGCAATCTGGGTCTG	7-4-1*01	✓	

✓ → *Positive PCR amplification*

TABLE 3 | *Comparison of IGHV sequences results obtained from the same sample with the two different protocols used in this study for DNA extraction and PCR amplification*

Sample identification	Protocol	IGHV	IGHD	IGHJ	Mutational status (% identity with germline)	HCDR3	HCDR3 length	Gravy Score	Group
HM0084	gDNA purification (isopropanol/glycogen)	3_7	3_10	4	Mutated (95.6%)	CARDLYDS GSSDYW	12	-1.23	M_MBL_II
	Extract-N-Amp PCR Kit protocol (17,000 cells)	3_7	3_10	4	Mutated (96.8%)	CARDLYDS GSSDYW	12	-1.23	M_MBL_II

M_MBL_I → Monoclonal B-cell lymphocytosis with a mutated IGHV sequence and neutral HCDR3
M_MBL_II → Monoclonal B-cell lymphocytosis with a mutated IGHV sequence and negatively-charged HCDR3

Proposed workflow results. Once the primer sensitivity and specificity for the two assays had been confirmed, a workflow was set to subsequent process MBL^{lo} samples. Thus, MBL^{lo} samples containing purified clonal B-cells were split depending on whether ≥50,000 cells were sorted or there were lower numbers of purified B-cells available (<50,000 cells). In the former group of samples DNA was extracted and precipitated using the isopropanol-glycogen assay, which had shown a better yield than other extraction methods, above the 50,000 cells threshold (based on the DNA concentration reached (Supplementary Table 1). Three samples had >50,000 cells and were thereby processed by this method, while the other 15 were evaluated using the Extract-N-Amp™ Blood PCR Kit.

Subsequently, a multiplex PCR using the two sets of previously described forward primers was applied. At this step 5/18 samples showed PCR-product amplification. Sequencing performed showed 3/5 productive *IGHV* rearrangements, while in 1/5 samples sequences corresponding to >1 *IGHV* rearrangement were recognized and in 1/5 samples an incomplete *IGHV* gene rearrangement was found, the latter two samples thus requiring another conventional PCR using a subset of individual forward primers (Figure 1), which allowed identification of the *IGHV* sequences in both of these sequences.

For those 13/18 samples for which no PCR product was detected at this step, reduced number of forward primers was then considered, by grouping them by *IGHV* families. At this stage, PCR-amplification was obtained in another 3/18 MBL^{lo} samples, one of which showed sequences corresponding to more than one *IGHV* sequences requiring another conventional PCR using one selected forward primer.

No PCR amplification or no sequence was obtained in this second step in 10/18 samples. In such cases, several PCRs with individual forward primers were done, with a positive result (*IGHV* sequence identified) in 1/10 samples, plus the three re-sequenced samples (derived from the previous steps).

Thus, the proposed workflow allowed full characterization of *IGH-VDJ* sequences of 9/18 (50%) samples. Data on the specific HCDR3 aminoacidic sequence and its hydropathy index obtained are shown in Table 4 and Supplementary Table 2.

From those 9/18 samples for which no *IGHV* sequences were identified, this was due to DNA degradation observed in an agarose gel electrophoresis done after the PCR -2/9 (22,2%)-, the presence of sequences corresponding to >1 IGHV gene rearrangement in

a single forward primer PCR in another 2/9 samples (22,2%) and the absence of a PCR-amplified product in the remaining five samples (55,6%).

The 9 *IGHV* sequenced cases were then merged with another group of 25 MBL^{lo} cases plus 73 MBL^{hi} and 138 CLL patients which had been sequenced previously.

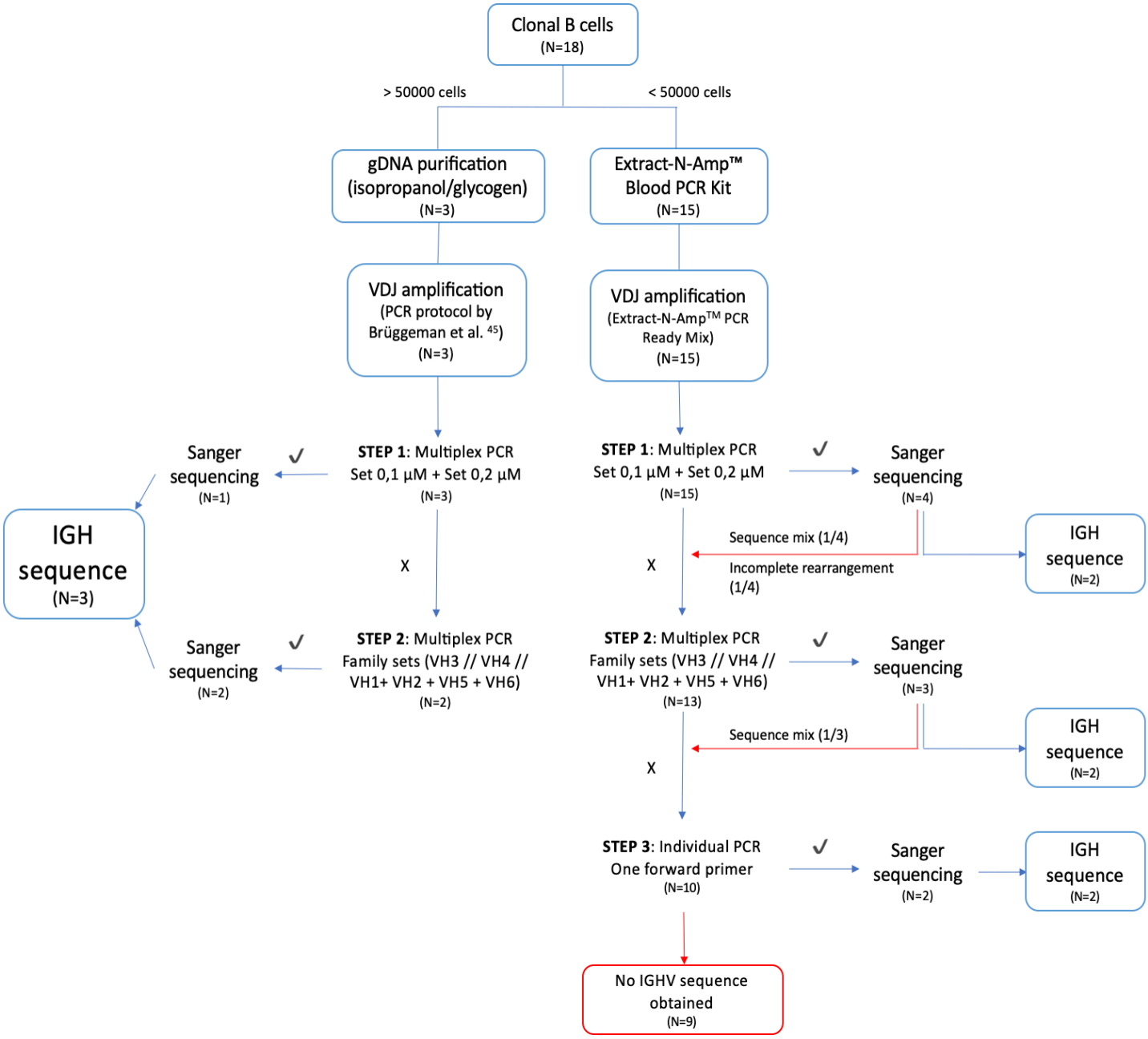

Figure 1 | *Workflow developed and followed in this study to sequence IGH-VDJ rearrangements from blood purified clonal B-cells from MBL^{lo} subjects.*

TABLE 4 | *MBL^{lo} cases characterized for their IGH-VDJ rearrangements from blood purified clonal B-cells.*

Identification	Cell number	Protocol	IGHV	IGHJ	IGHD	Group
HM0069	664	Cell lysis	3-7*03	4*02	6-19*01	U_MBL_II
HM0037	1,613	Cell lysis	2-5*01	5*02	3-10*02	M_MBL_II
HM0086	3,000	Cell lysis	3-15*01	4*02	2-15*01	M_MBL_I
HM0075	3,000	Cell lysis	No IGH sequence obtained (DNA degradation)			
HM0027	3,000	Cell lysis	No IGH sequence obtained (no PCR amplification)			
HM0074	3,000	Cell lysis	No IGH sequence obtained (no PCR amplification)			
HM0049	4,000	Cell lysis	No IGH sequence obtained (no PCR amplification)			
HM0032	8,000	Cell lysis	3-7*01	4*02	3-3*01	M_MBL_I
HM0059	10,000	Cell lysis	No IGH sequence obtained (no PCR amplification)			
HM0084	17,000	Cell lysis	3-7*01	4*01	3-10*01	M_MBL_II
HM0008	20,000	Cell lysis	3-23*04	4*02	6-25*01	M_MBL_II
HM0056	24,000	Cell lysis	No IGH sequence obtained (no PCR amplification)			
HM0077	25,000	Cell lysis	No single IGH sequence obtained (multiple sequences)			
HM0058	30,000	Cell lysis	No single IGH sequence obtained (multiple sequences)			
HM0063	30,000	Cell lysis	No IGH sequence obtained (DNA degradation)			
HM0044	125,000	gDNA purification	3-72*01	6*02	2-8*01	U_MBL
HM0034	140,000	gDNA purification	4-4*02	5*02	3-16*01	M_MBL_I
HM0030	1,000,000	gDNA purification	3-7*01	4*02	6-13*01	M_MBL_I

M_MBL_I → Monoclonal B-cell lymphocytosis with a mutated IGHV sequence and neutral HCDR3
M_MBL_II → Monoclonal B-cell lymphocytosis with a mutated IGHV sequence and negatively-charged HCDR3

U_MBL_I → Monoclonal B-cell lymphocytosis with an unmutated IGHV sequence and neutral HCDR3
U_MBL_II → Monoclonal B-cell lymphocytosis with an unmutated IGHV sequence and negatively-charged HCDR3

2. BCR sequences in MBL^{lo} vs MBL^{hi} and CLL

The *IGHV*, *IGHD* and *IGHJ* gene segment usage was compared among the three different diagnostic categories of MBL^{lo}, MBL^{hi} and CLL cases. Among the three gene segments, *IGHV* sequences emerged as the most variable ones with up to 40 different IGHV gene segments being identified in our study cohort. Because of this, only the most frequently detected *IGHV* gene rearrangements (>5% of cases in a specific diagnostic group) were selected for subsequent analysis. These included 6 *IGHV* gene segments

IGHV1_69, IGHV3_7, IGHV3_23, IGHV3_21, IGHV4_34 and *IGHV4_4* (Figure 2 and Table 6).

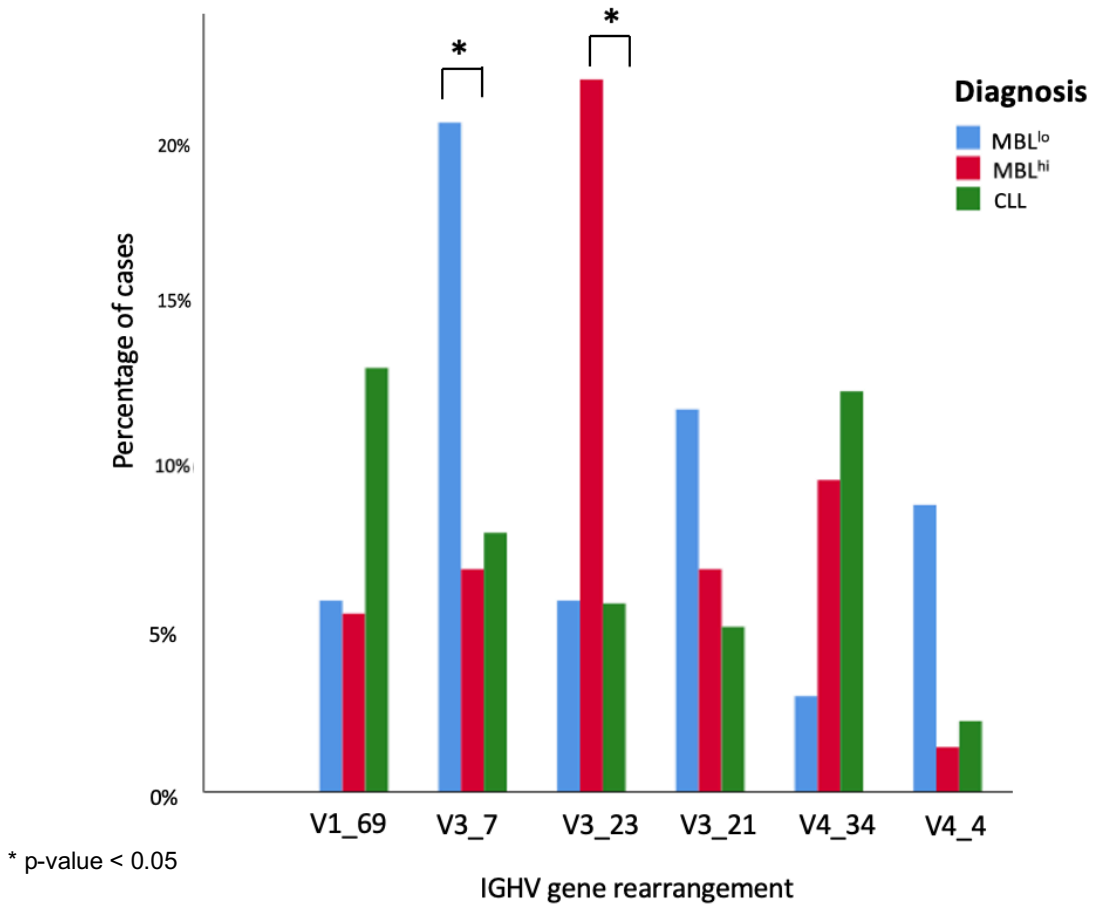

Figure 2 | *Distribution of the most frequent IGHV rearrangements encountered (>5%) for each diagnostic category (MBL^{lo}, MBL^{hi} and CLL).*

From these six gene segments, two (*IGHV3-7* and *IGHV3-23*) were found at significantly different frequencies in the distinct diagnostic groups. While the *IGHV3-7* gene rearrangement was characteristic of MBL^{lo} (p=0.049), *IGHV3-23* was more frequently found in MBL^{hi} than in MBL^{lo} and CLL (p=0.001). For other gene rearrangements a tendency towards a greater prevalence of *IGHV1-69* in CLL, *IGHV4-4* in MBL^{lo} and *IGHV4-34* both in MBL^{hi} and CLL, was observed.

TABLE 6 | *Relative frequency of the most prevalent IGHV gene segments used (>5% cases) in MBL^{lo}, MBL^{hi} and CLL cases*

IGHV rearrangement	MBL^{lo} (N=34)	MBL^{hi} (N=73)	CLL (N=138)	p-value
IGHV 1_69	2/34 (5.9%)	4/73 (5.5%)	18/138 (13.0%)	0.623^{a} / 0.196^{b} / *0.064*^{c}
IGHV3_7	7/34 (20.6%)	5/73 (6.8%)	11/138 (8.0%)	**0.049**^{a} / *0.054*^{b} / 0.502^{c}
IGHV 3_23	2/34 (5.9%)	16/73 (21.9%)	8/138 (5.8%)	*0.051*^{a} / 0.624^{b} / **0.001**^{c}
IGHV3_21	4/34 (11.8%)	5/73 (6.8%)	7/138 (5.1%)	0.461^{a} / 0.231^{b} / 0.756^{c}
IGHV 4_34	1/34 (2.9%)	7/73 (9.6%)	17/138 (12.3%)	0.211^{a} / *0.090*^{b} / 0.364^{c}
IGHV 4_4	3/34 (8.8%)	1/73 (1.4%)	3/138 (2.2%)	*0.094*^{a} / *0.092*^{b} / 0.570^{c}
Other	15/34 (44.1%)	35/73 (47.9%)	74/138 (53.6%)	

Results expressed as number of cases (percentage)

^{a} MBL^{lo} vs MBL^{hi}; ^{b} MBL^{lo} vs CLL; ^{c} MBL^{hi} vs CLL

Regarding the D-gene segments of the immunoglobulin heavy-chain sequences identified, a total of 7 families were detected (Figure 3), with a tendency towards a lower *IGHD5* frequency in MBL^{lo} vs CLL (p=0.052).

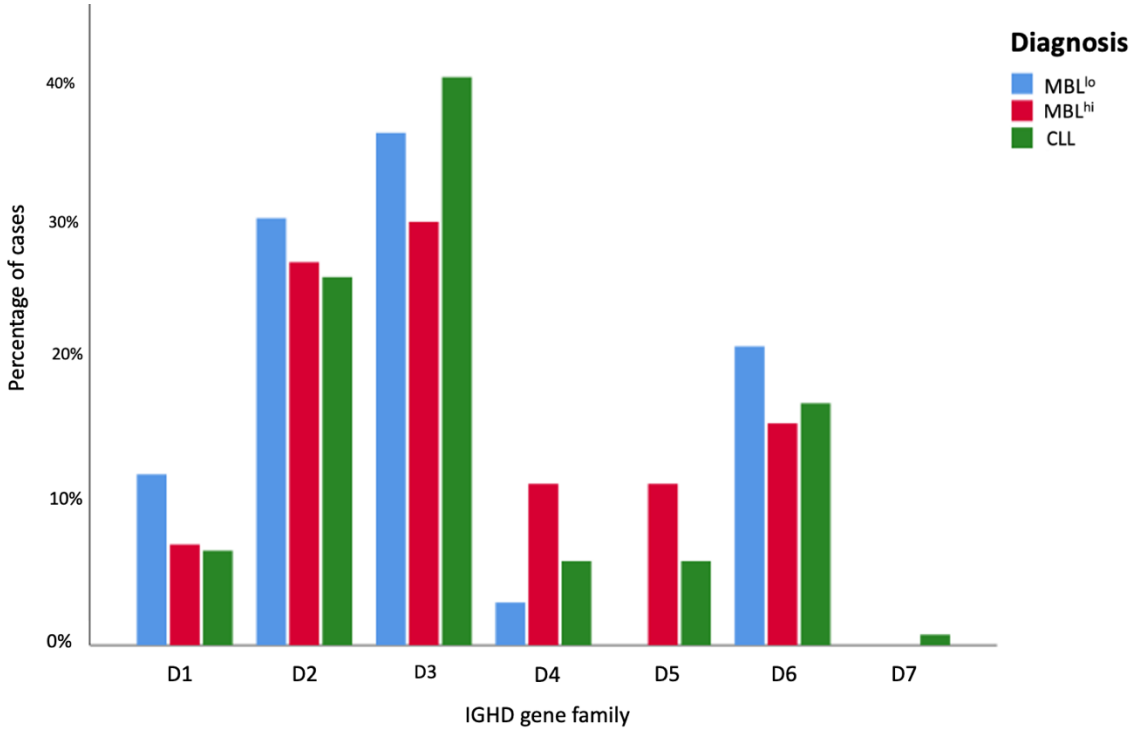

Figure 3 | *Distribution of the different IGHD families identified by diagnostic category (MBL^{lo}, MBL^{hi} and CLL)*

More differences were observed as regards the usage of different J-segments of the *IGH* genes among the different diagnostic groups (Figure 4). Thus, *IGHJ4* was more frequently found in MBL^{lo} (61.8%) compared to CLL (35.9%; p=0.033, while the frequency of *IGHJ6* gene usage was higher in CLL (37%) vs both MBL^{lo} (14.7%; p=0.014) and MBL^{hi} (23.3%; p=0.046) (Table 7).

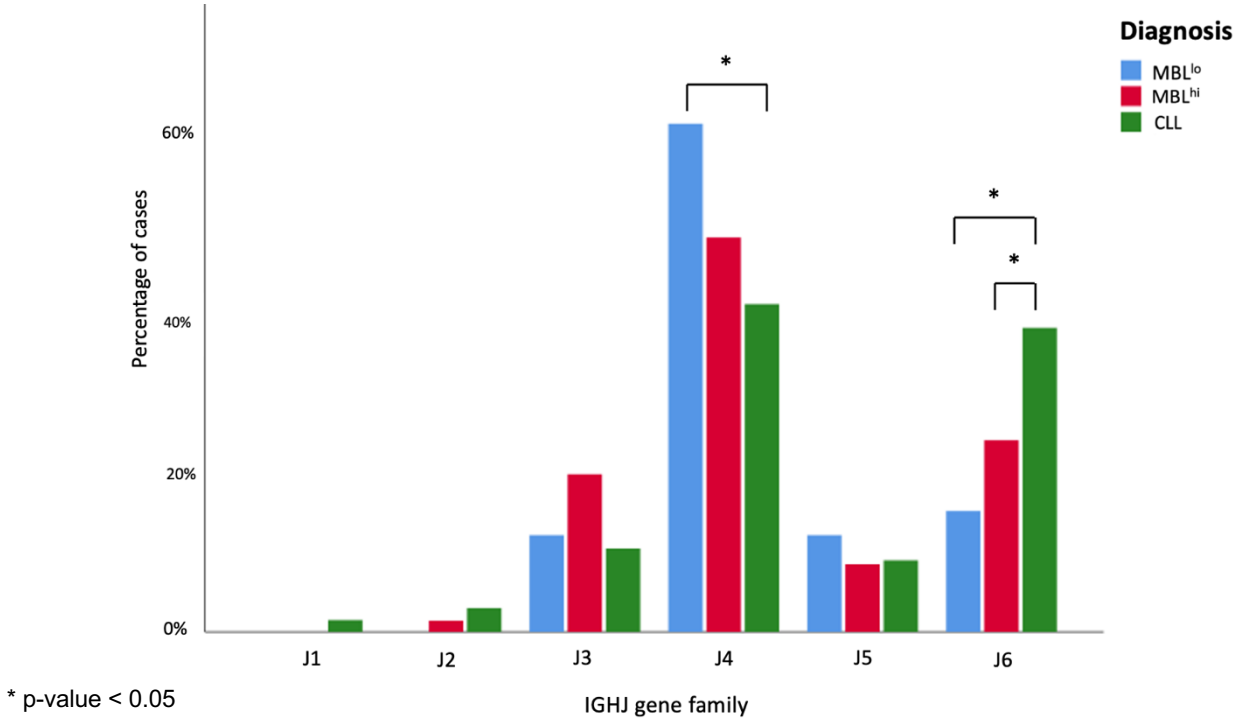

Figure 4 | *Distribution of the different IGHJ families identified by diagnostic category (MBL^{lo}, MBL^{hi} and CLL).*

TABLE 7 | *Relative frequency of different IGHJ families identified in MBL^{lw}, MBL^{hi} and CLL cases*

IGHJ rearrangment family	MBL^{lo} (N=34)	MBL^{hi} (N=73)	CLL (N=138)	p-value
IGHJ4	21/34 (61.8%)	35/73 (47.9%)	55/138 (35.9%)	0.216^{a} / **0.033**^{b} / 0.306^{c}
IGHJ6	5/34 (14.7%)	17/73 (23.3%)	51/138 (37.0%)	0.442^{a} / **0.014**^{b} / **0.046**^{c}
Other	8/34 (23.5%)	21/73 (28.8%)	32/138 (23.1%)	

Results expressed as number of cases (percentage)
^{a} MBL^{lo} vs MBL^{hi}; ^{b} MBL^{lo} vs CLL; ^{c} MBL^{hi} vs CLL

Apart from the biased distribution of the *VDJ* gene segments described above, differences in the mutational status, the length and the hydropathy index of HCDR3 sequences were also observed between MBL^{lo}, MBL^{hi} and CLL(Table 8). In this regard, the HCDR3 length was significantly higher (p=0.017 vs MBL^{lo} and p=0.004 vs MBL^{hi}) in CLL (18 amino acids) compared to MBL^{lo} (16 amino acids) and MBL^{hi} (16 amino acids), but with a similar distribution of hydrophobic and hydrophilic aminoacidic HCDR3 sequences among all three diagnostic categories as evaluated by a neutral vs negative Gravy Score (Table 8).

TABLE 8 | *Comparison of BCR characteristics clustered by MBL^{lo}, MBL^{hi} and CLL*

Patients feature	MBL^{lo} (N=34)	MBL^{hi} (N=73)	CLL (N=138)	*p-value*
HCDR3 length	16 (9-23)	16 (7-27)	18 (7-28)	0.765[a] **/ 0.017[b] / 0.004[c]**
HCDR3 Gravy Score				
Neutral (GS ≥ -0,5)	17/34 (50%)	36/73 (49.3%)	65/138 (47.1%)	1.000[a] / 0.849[b] / 0.774[c]
Negatively charged (GS <-0,5)	17/34 (50%)	37/73 (50.7%)	73/138 (52.9%)	
IGHV Mutational Status				
Mutated	22/34 (65%)	64/73 (88%)	78/138 (57%)	**0.007[a] /** 0.441[b] / **<0.001[c]**
Unmutated	12/34 (35%)	9/73 (12%)	60/138 (43%)	

Results expressed as number of cases (percentage)
[a] MBL^{lo} vs MBL^{hi}; [b] MBL^{lo} vs CLL; [c] MBL^{hi} vs CLL

3. Association between BCR sequences and the behavior of MBL^{lo} clones

According to their *IGHJ* gene usage, MBL^{lo} cases we clustered into two subgroups including cases presenting J-segments upstream *IGHJ4* (N=25) and those who presented downstream *IGHJ5* and *IGHJ6* gene rearrangements (N=9). Both groups of MBL^{lo} cases showed similar distributions per sex and the IGHV mutational status (Table 9). Despite this, slight differences were found between both groups as regards a lower age (61 years vs 68 years) and a higher frequency of cytogenetic alterations among IGHJ(5-6) cases (60% vs 40% respectively). In addition, clonal B-cells showed higher counts among the *IGHJ*(5-6)-positive group both at diagnosis and at follow-up, and displayed longer HCDR3 sequences than *IGHJ1-4* cases.

Altogether, these results suggest that usage of *IGHJ5-6* by MBL^{lo} cases might be associated with a tendency towards a higher similarity to CLL aberrant B-cells, further studies in larger group of cases being required to confirm this hypothesis.

TABLE 9 | *Biological and clinical features of MBL^{lo} subjects grouped according to the IGHJ gene family usage*

Patients feature	IGHJ(1-4) (N=25)	IGHJ(5-6) (N=9)	*p-value*
Age at diagnosis	68 (43-88)	61 (44-76)	0.175
Sex			
Men/Women	8/17 (32%/68%)	6/3 (67%/33%)	0.166
HCDR3 length	15 (10-20)	17 (9-23)	0.130
IGHV Mutational Status			
Mutated	17/25 (68%)	5/9 (56%)	0.687
Unmutated	8/25 (32%)	4/9 (44%)	
Gravy Score			
Neutral	13/25 (52%)	4/9 (44%)	1.000
Negatively charged	12/25 (48%)	5/9 (56%)	
Cytogenetic Alteration			
None	6/10 (60%)	2/5 (40%)	0.593
Del(13q14) (*D13S25)*	3/10 (30%)	3/5 (60%)	0.329
Trisomy 12	1/10 (10%)	0/5 (0%)	1.000
Clone size at diagnosis (nº of cells/µL)	6.14 (0.08-65.62)	10.54 (0.28-64.40)	0.397
Clone size at follow-up (nº of cells/µL)	37 (0.05-325.74)	106.1 (0.31-807.82)	0.393
Growth rate (size at diagnosis/size at follow-up)	7.6 (0.14-82.1)	6.7 (0.14-37.9)	0.495

Results expressed as median (range) or as number of cases(percentage).
IGHJ: inmunoglobulin heavy-chain J-segment; HCDR3: heavy-chain complementarity-determinant region; GS: Gravy Score

DISCUSSION

Molecular studies in MBL^{lo} are challenging due to the very limited number and frequency of clonal B-cells present in blood of subjects and the relatively high frequency of oligoclonal expansions that may hamper *IGH* sequencing of aberrant MBL^{lo} B-cells. In order to enlarge our MBL^{lo} cohort with characterized *IGHV* sequences of clonal B-cells, here we retrieved MBL^{lo} samples of purified clonal B-cells which had been previously stored for several years and that could not be analyzed at MBL^{lo} diagnosis due to the low number of cells purified. For this purpose here we designed a new workflow algorithm and applied a new methodology for analyzing samples for which <50,000 clonal B-cells had been obtained, based on the use of the Extract-N-Amp^{TM} Blood PCR Kit and. This new approach allowed us to retrieve the *IGHV* sequences of 9 new cases out of 18

processed that could not be sequenced previously based on other protocols requiring higher cell counts. Despite the fact that we demonstrated a higher sensitivity and similar specificity of the new protocol compared to conventional approaches, the new methodology still failed to provide *IGHV* sequences in a substantial fraction of the cases. However, it might well be that such high rate of failure is related to the poor quality of the stored purified clonal B-cell samples from MBL^{lo} subjects, which had been prepared and stored several years ago, optimal storage not being completely guaranteed prior to the storage at the Spanish National DNA Bank. In line with this possibility, a fraction of the samples showed DNA degradation. In such cases that showed no PCR amplification, NGS is not a feasible alternative because, as it is described in the NGS protocols for *IGHV* sequencing[44], the first step consists of a multiplexed PCR (similar to the one here performed) with the same primer mixes used here except for the addition of a complementary sequence required to add the adaptors and barcodes in the second PCR of the NGS protocol. Thus, if no PCR amplification has been obtained in these samples (due to DNA degradation or other causes), no NGS results would also be expected. Further studies on freshly processed purified B-cell clones are required to understand the exact causes for the lack of enough PCR-amplified DNA in these cases.

Nevertheless, the use of NGS approaches for *IGH* gene sequencing might provide more information than that obtained with the protocol used in this study, since it can identify the dominant (most frequent) *IGHV* gene rearrangement, but also oligoclonal expansions that might frequently coexist within and in parallel in MBL^{lo} cases[23]. This information would be specially relevant in longitudinal sequential studies, to identify potentially different kinetics for distinct clones showing different *IGHV* gene rearrangements, particularly in cases showing progression to MBL^{hi} and CLL. However, application of NGS protocols would also be associated with higher costs per sample with respect to the protocol used here: around 200€/sample in NGS vs 25€/sample in our protocol, respectively.

At present, it is well established that CLL cells show a biased and distinct *IGHV* repertoire compared to B-cells from healthy donors, associated with a high frequency of the so-called "stereotyped" BCRs[21,46,47]. In contrast, limited data has been reported on the repertoire of MBL^{lo} clones. In this regard, Vardi et al. have found different immunogenetic signatures in MBL^{hi} vs MBL^{lo}, with specific IGHV gene rearrangements typical of each of these conditions: *IGHV4-59/61* were highly represented in MBL^{lo} while *IGHV1-69, IGHV2-9, IGHV3-33, IGHV3-48* and *IGHV4-34* were more frequently found in MBL^{hi} [37]. In parallel, Henriques et al., in a cohort consisting of MBL^{lo}, MBL^{hi} and CLL cases, more frequently found *IGHV1-69* and *IGHV4-34* gene rearrangements in CLL while *IGHV3-23* was characteristic in MBL^{hi} [36]. In line with these results, here we found greater prevalence of *IGHV3-23* gene rearrangements in MBL^{hi} and *IGHV1-69* gene rearrangements in CLL. In addition, here we report for the first time a greater frequency of *IGHV3-7* rearrangements in MBL^{lo}, in contrast to the data reported by Vardi et al., who found similar *IGHV3-7* frequencies in MBL^{lo}, MBL^{hi} and early stage CLL (9.7%, 8.1% and 9.9% respectively)[37] at frequencies half than those reported here for our MBL^{lo} cases. Such apparent discrepancy might relate to the fact that here we investigated MBL^{lo} cases with very low clonal B-cell counts in blood, compared to Vardi et al, whose MBL^{lo} cases showed absolute clonal B-cell counts much closer to those of MBL^{hi}. However, further studies in larger MBL^{lo} cohorts are needed to confirm our findings.

Most interestingly, here we found a significant bias in the distribution of J-gene segments among MBL^{low}, MBL^{hi} and CLL, with higher usage of *IGHJ4* in MBL^{lo} vs MBL^{hi} and CLL cases who more frequently had *IGHJ6* gene rearrangements. Of note, among normal B-lymphocytes *IGJH4* segments are related to more mature B cells[48] and shorter HCDR3 aminoacidic sequences, in contrast to BCR carrying *IGHJ6* gene rearrangements that show longer HCDR3 aminoacidic sequences in association with VDJ rearrangement patterns of viral specific B-cells (e.g. SARS-CoV2, HIV and Ebola[49-51]).

The greater usage of *IGHJ4* gene segments (characteristic of more mature B-cells) in MBL^{lo} vs MBL^{hi} and CLL clonal B-cells, may support the notion that MBL^{lo} clones are more likely derived from immunosenescent B lymphocytes, as also supported by the previously reported differences in the *IGHV* repertoire between MBL^{lo} and CLL, suggesting a lower growth potential vs that of clonal B-cells with *IGHJ6* gene rearrangements.

Despite all the above, MBL^{lo} clones with *IGHJ(1-4)* gene segments showed similar features to those expressing the *IGHJ5-IGHJ6* regions in their BCR but with a tendency towards a slightly different biological behavior of both groups of MBL^{lo} cases. Further studies in larger MBL^{lo} cohorts are needed to clarify the potential implication of *IGHJ* gene segments selection in the expansion and malignant transformation of B-cell clones, independently (or not) of the *IGHV* mutational status.

CONCLUSIONS

1. Here we propose a new approach for an increased detection of *IGHV* sequences in MBL^{lo} subjects for which very limited number of purified clonal B-cells are obtained, which requires further validation in freshly-purified samples including direct comparison with NGS.
2. The frequency of distinct *IGHV* gene rearrangements in MBL^{lo} cases differs from that of MBL^{hi} and CLL patients, with some V-gene (*IGHV3-7)* being more frequently observed in the former cases compared with the latter groups. Similarly, *IGHJ4* gene segments are more frequently found among MBL^{lo} clonal B-cells, while *IGHJ6* gene rearrangements are more prevalent in MBL^{hi} and CLL. Altogether, these findings support a potential role for BCR stimulation on the ontogeny of CLL.
3. Despite the differences observed on the *JH*-gene segment of the BCR repertoire of MBL^{lo} vs CLL cases, no significantly different molecular, biological and demographic features were found among MBL^{lo} cases expressing different JH-gene rearrangements, although a tendency towards higher and more frequently (cytogenetically) altered clones was observed among *IGHJ5-6* MBL^{lo} cases. Further studies in larger cohorts of MBL^{lo} cases being required to confirm these observations.

REFERENCE LIST

1. Cancer Statistics Review, 1975-2011 - Previous Version - SEER Cancer Statistics Review. Available at: https://seer.cancer.gov/archive/csr/1975_2011/. (Accessed: 18th May 2022)
2. Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2019. *CA. Cancer J. Clin.* **69**, 363–385 (2019).
3. Scarfò, L., Ferreri, A. J. M. & Ghia, P. Chronic lymphocytic leukaemia. *Crit. Rev. Oncol. Hematol.* **104**, 169–182 (2016).
4. Lebien, T. W. & Tedder, T. F. B lymphocytes: how they develop and function. *Blood* **112**, 1570–1580 (2008).
5. Melchers, F. *et al.* Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B-cells. *Immunol. Rev.* **175**, 33–46 (2000).
6. Pelanda, R. & Torres, R. M. Central B-cell tolerance: where selection begins. *Cold Spring Harb. Perspect. Biol.* **4**, a007146 (2012).
7. Coffey, F., Alabyev, B. & Manser, T. Initial clonal expansion of germinal center B-cells takes place at the perimeter of follicles. *Immunity* **30**, 599–609 (2009).
8. Natkunam, Y. The biology of the germinal center. *Hematol. Am. Soc. Hematol. Educ. Progr.* 210–215 (2007). doi:10.1182/ASHEDUCATION-2007.1.210
9. Allen, C. D. C., Okada, T. & Cyster, J. G. Germinal-center organization and cellular dynamics. *Immunity* **27**, 190–202 (2007).
10. Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).
11. Mesin, L., Ersching, J. & Victora, G. D. Germinal Center B-cell Dynamics. *Immunity* **45**, 471–482 (2016).
12. Seifert, M. & Küppers, R. Human memory B-cells. *Leukemia* **30**, 2283–2292 (2016).
13. Seifert, M. *et al.* Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J. Exp. Med.* **209**, 2183–2198 (2012).
14. Gladstone, D. E. *et al.* Importance of immunoglobulin heavy chain variable region mutational status in del(13q) chronic lymphocytic leukemia. **10**, 1873–71 (2011).
15. Rodríguez-Caballero, A. *et al.* The Hydropathy Index of the HCDR3 Region of the B-Cell Receptor Identifies Two Subgroups of IGHV-Mutated Chronic Lymphocytic Leukemia Patients With Distinct Outcome. *Front. Oncol.* **11**, 723722 (2021).
16. Fais, F. *et al.* Chronic lymphocytic leukemia B-cells express restricted sets of mutated and unmutated antigen receptors. *J. Clin. Invest.* **102**, 1515–1525 (1998).
17. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig VH Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia. *Blood* **94**, 1848–1854 (1999).
18. Sutton, L. A. *et al.* Immunoglobulin genes in chronic lymphocytic leukemia: key to understanding the disease and improving risk stratification. *Haematologica* **102**, 968–971 (2017).
19. Ghia, P. & Caligaris-Cappio, F. Monoclonal B-cell lymphocytosis: right track or red herring? *Blood* **119**, 4358–4362 (2012).
20. Rosén, A., Murray, F., Evaldsson, C. & Rosenquist, R. Antigens in chronic lymphocytic leukemia--implications for cell origin and leukemogenesis. *Semin. Cancer Biol.* **20**, 400–409 (2010).
21. Ten Hacken, E., Gounari, M., Ghia, P. & Burger, J. A. The importance of B-cell receptor isotypes and stereotypes in chronic lymphocytic leukemia. *Leukemia* **33**, 287–298 (2019).

22. Landgren, O. *et al.* B-cell clones as early markers for chronic lymphocytic leukemia. *N. Engl. J. Med.* **360**, 659–667 (2009).

23. Maitre, E. & Troussard, X. Monoclonal B-cell lymphocytosis. *Best Pract. Res. Clin. Haematol.* **32**, 229–238 (2019).

24. Swerdlow, S. H. et al. *World Health Organization Classification of Tumors of Haematopoietic and Lymphoid Tissues 2017 Revised Edition*. (IARC, 2017).

25. Kalpadakis, C. *et al.* New insights into monoclonal B-cell lymphocytosis. *Biomed Res. Int.* **2014**, 258917 (2014).

26. Rawstron, A. C. *et al.* Monoclonal B lymphocytes with the characteristics of 'indolent' chronic lymphocytic leukemia are present in 3.5% of adults with normal blood counts. *Blood* **100**, 635–639 (2002).

27. Ghia, P. *et al.* Monoclonal CD5+ and CD5- B-lymphocyte expansions are frequent in the peripheral blood of the elderly. *Blood* **103**, 2337–2342 (2004).

28. Shim, Y. K. *et al.* Monoclonal B-cell lymphocytosis in healthy blood donors: An unexpectedly common finding. *Blood* **123**, 1319–1326 (2014).

29. Nieto, W. G. *et al.* Increased frequency (12%) of circulating chronic lymphocytic leukemia-like B-cell clones in healthy subjects using a highly sensitive multicolor flow cytometry approach. *Blood* **114**, 33–37 (2009).

30. Almeida, J. *et al.* CLL-like B-lymphocytes are systematically present at very low numbers in peripheral blood of healthy adults. *Leukemia* **25**, 718–722 (2011).

31. Criado, I. *et al.* Low-count monoclonal B-cell lymphocytosis persists after seven years of follow up and is associated with a poorer outcome. *Haematologica* **103**, 1198–1208 (2018).

32. Hauswirth, A. W. *et al.* Monoclonal B-cell lymphocytosis (MBL) with normal lymphocyte counts is associated with decreased numbers of normal circulating B-cell subsets. *Am. J. Hematol.* **87**, 721–724 (2012).

33. Moreira, J. *et al.* Infectious complications among individuals with clinical monoclonal B-cell lymphocytosis (MBL): a cohort study of newly diagnosed cases compared to controls. *Leukemia* **27**, 136–141 (2013).

34. Forconi, F. & Moss, P. Perturbation of the normal immune system in patients with CLL. *Blood* **126**, 573–581 (2015).

35. Kikushige, Y. *et al.* Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell* **20**, 246–259 (2011).

36. Henriques, A. *et al.* Combined patterns of IGHV repertoire and cytogenetic/molecular alterations in monoclonal B lymphocytosis versus chronic lymphocytic leukemia. *PLoS One* **8**, e67751 (2013).

37. Vardi, A. *et al.* Immunogenetics shows that not all MBL are equal: The larger the clone, the more similar to CLL. *Blood* **121**, 4521–4528 (2013).

38. Dagklis, A. *et al.* The immunoglobulin gene repertoire of low-count chronic lymphocytic leukemia (CLL)-like monoclonal B lymphocytosis is different from CLL: diagnostic implications for clinical monitoring. *Blood* **114**, 26–32 (2009).

39. Agathangelidis, A. *et al.* Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood* **119**, 4467–4475 (2012).

40. Agathangelidis, A. *et al.* Infrequent 'chronic lymphocytic leukemia-specific' immunoglobulin stereotypes in aged individuals with or without low-count monoclonal B-cell lymphocytosis. *Haematologica* **106**, 1178–1181 (2021).

41. Henriques, A. *et al.* Molecular and cytogenetic characterization of expanded B-cell clones from multiclonal versus monoclonal B-cell

chronic lymphoproliferative disorders. *Haematologica* **99**, 897–907 (2014).

42. Sharkey, D. J., Scalice, E. R., Christy, K. G., Atwood, S. M. & Daiss, J. L. Antibodies as thermolabile switches: high temperature triggering for the polymerase chain reaction. *Biotechnology. (N. Y).* **12**, 506–509 (1994).

43. van Dongen, J. J. 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* **17**, 2257–2317 (2003).

44. Brüggemann, M. *et al.* Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* **33**, 2241–2253 (2019).

45. Rosenquist, R. *et al.* Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia* **31**, 1477–1481 (2017).

46. Stamatopoulos, K. *et al.* Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood* **109**, 259–270 (2007).

47. Hengeveld, P. J., Levin, M. D., Kolijn, P. M. & Langerak, A. W. Reading the B-cell receptor immunome in chronic lymphocytic leukemia: revelations and applications. *Exp. Hematol.* **93**, 14–24 (2021).

48. Huang, C., Keith Stewart, A., Schwartz, R. S. & David Stollar, B. Immunoglobulin heavy chain gene expression in peripheral blood B lymphocytes. *J. Clin. Invest.* **89**, 1331–1343 (1992).

49. Wang, H. *et al.* Antibody heavy chain CDR3 length-dependent usage of human IGHJ4 and IGHJ6 germline genes. *Antib. Ther.* **4**, 101–108 (2021).

50. Zhang, Y. *et al.* Analysis of B-cell Receptor Repertoires Reveals Key Signatures of the Systemic B-cell Response after SARS-CoV-2 Infection. *J. Virol.* **96**, (2022).

51. Tian, X. *et al.* Characterization of human IgM and IgG repertoires in individuals with chronic HIV-1 infection. *Virol. Sin.* **37**, (2022).

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE 1 | *DNA purification results with the isopropanol-glycogen DNA precipitation protocol vs the Machery-Nagel precipitating protocol (N=12)*

Sample code	Extraction protocol	Cell number	Expected DNA (ng)	[DNA] (ng/µL)	Total purified DNA (ng)
63188	Isopropanol-glycogen	50,000	300	5.36	107.2
15504-Precip	Isopropanol-glycogen	10,000	60	2.46	49.2
TK1378	Isopropanol-glycogen	10,000	60	1.06	21.2
TK1405	Isopropanol-glycogen	10,000	60	0.68	13.6
63140	Machery-Nagel	25,000	150	0.0033	0.066
63253	Machery-Nagel	25,000	150	0.0796	1.58
38258	Machery-Nagel	18,000	108	0.191	3.82
26082	Machery-Nagel	16,000	96	0.145	2.9
63281	Machery-Nagel	16,000	96	0.019	0.38
15504-MN	Machery-Nagel	10,000	60	0.224	4.48
63366	Machery-Nagel	10,000	60	0.073	1.46
TK1378-8000	Machery-Nagel	8,000	48	0.244	4.48

SUPLEMENTARY TABLE 2| *MBL^{low} cases characterized in this study for their IGH-VDJ sequence of blood purified clonal B-cells*

Sample code	Cell number	Protocol	IGHV	IGHJ	IGHD	HCDR3 sequence	Mutational status	GRAVY Score	Group
HM0069	664	Cell lysis	3-7*03	4*02	6-19*01	CARRGWLF DYW	Unmutated (98,95%)	-0.73	U_MBL_II
HM0037	1,613	Cell lysis	2-5*01	5*02	3-10*02	CAHSNTETT AGEFDPW	Mutated (87,11%)	-1.11	M_MBL_II
HM0086	3,000	Cell lysis	3-15*01	4*02	2-15*01	CSAGLGQT DTDCW	Mutated (93,28%)	-0.492	M_MBL_I
HM0075	3,000	Cell lysis	No IGH sequence obtained (DNA degradation)						
HM0027	3,000	Cell lysis	No IGH sequence obtained (absence of PCR amplification)						
HM0074	3,000	Cell lysis	No IGH sequence obtained (absence of PCR amplification)						
HM0049	4,000	Cell lysis	No IGH sequence obtained (absence of PCR amplification)						
HM0032	8,000	Cell lysis	3-7*01	4*02	3-3*01	CARETFWS GYFDFW	Mutated (96,36%)	-0.47	M_MBL_I
HM0059	10,000	Cell lysis	No IGH sequence obtained (absence of PCR amplification)						
HM0084	17,000	Cell lysis	3-7*01	4*01	3-10*01	CARDLYDSG SSDYW	Mutated (96,84%)	-1.23	M_MBL_II
HM0008	20,000	Cell lysis	3-23*04	4*02	6-25*01	CAKEAGATR LYFDYW	Mutated (93,01%)	-0.53	M_MBL_II
HM0056	24,000	Cell lysis	No IGH sequence obtained (absence of PCR amplification)						
HM0077	25,000	Cell lysis	No single IGH sequence obtained (multiple sequences)						
HM0058	30,000	Cell lysis	No single IGH sequence obtained (multiple sequences)						
HM0063	30,000	Cell lysis	No IGH sequence obtained (DNA degradation)						
HM0044	125,000	gDNA purification	3-72*01	6*02	2-8*01	CARAYCADT VCSLRSYYG MDVW	Unmutated (99,32%)	0.08	U_MBL
HM0034	140,000	gDNA purification	4-4*02	5*02	3-16*01	CASSIASTKG CFDPW	Mutated (96,97%)	0.08	M_MBL_I
HM0030	1,000,000	gDNA purification	3-7*01	4*02	6-13*01	CARDYAGG AIDFW	Mutated (95,10%)	-0.1	M_MBL_I

M_MBL_I → Monoclonal B-cell lymphocytosis with a mutated IGHV sequence and neutral HCDR3
M_MBL_II → Monoclonal B-cell lymphocytosis with a mutated IGHV sequence and negatively-charged HCDR3

U_MBL_I → Monoclonal B-cell lymphocytosis with an unmutated IGHV sequence and neutral HCDR3
U_MBL_II → Monoclonal B-cell lymphocytosis with an unmutated IGHV sequence and negatively-charged HCDR3