



## OPEN ACCESS

## EDITED BY

Francesco Bertolini,  
European Institute of Oncology  
(IEO), Italy

## REVIEWED BY

Andrea Nicola Mazzarello,  
University of Genoa, Italy  
Bartosz Puła,  
Institute of Hematology and  
Transfusiology (IHT), Poland

## \*CORRESPONDENCE

Kostas Stamatopoulos  
kostas.stamatopoulos@certh.gr

## SPECIALTY SECTION

This article was submitted to  
Cancer Genetics,  
a section of the journal  
Frontiers in Oncology

RECEIVED 25 October 2022

ACCEPTED 28 November 2022

PUBLISHED 14 December 2022

## CITATION

Sofou E, Zaragoza-Infante L,  
Pechlivanis N, Karakatsoulis G,  
Notopoulou S, Stavroyianni N,  
Psomopoulos F, Georgiou E,  
de Septenville AL, Davi F,  
Agathangelidis A, Chatzidimitriou A  
and Stamatopoulos K (2022) Evidence  
of somatic hypermutation in the  
antigen binding sites of patients with  
CLL harboring IGHV genes with 100%  
germline identity.  
*Front. Oncol.* 12:1079772.  
doi: 10.3389/fonc.2022.1079772

## COPYRIGHT

© 2022 Sofou, Zaragoza-Infante,  
Pechlivanis, Karakatsoulis, Notopoulou,  
Stavroyianni, Psomopoulos, Georgiou,  
de Septenville, Davi, Agathangelidis,  
Chatzidimitriou and Stamatopoulos. This  
is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction  
in other forums is permitted, provided  
the original author(s) and the  
copyright owner(s) are credited and  
that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is  
permitted which does not comply  
with these terms.

# Evidence of somatic hypermutation in the antigen binding sites of patients with CLL harboring IGHV genes with 100% germline identity

Electra Sofou<sup>1,2</sup>, Laura Zaragoza-Infante<sup>1</sup>,  
Nikolaos Pechlivanis<sup>1</sup>, Georgios Karakatsoulis<sup>1</sup>,  
Sofia Notopoulou<sup>1</sup>, Niki Stavroyianni<sup>3</sup>, Fotis Psomopoulos<sup>1</sup>,  
Elisavet Georgiou<sup>2</sup>, Anne Langlois de Septenville<sup>4</sup>,  
Frederic Davi<sup>4</sup>, Andreas Agathangelidis<sup>5</sup>,  
Anastasia Chatzidimitriou<sup>1,6</sup> and Kostas Stamatopoulos<sup>1,6\*</sup>

<sup>1</sup>Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece,

<sup>2</sup>Laboratory of Biological Chemistry, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece, <sup>3</sup>Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece, <sup>4</sup>Department of Hematology, APHP, Hôpital Pitié-Salpêtrière and Sorbonne University, Paris, France, <sup>5</sup>Department of Biology, School of Science, National and Kapodistrian University of Athens, Athens, Greece, <sup>6</sup>Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

Classification of patients with chronic lymphocytic leukemia (CLL) based on the somatic hypermutation (SHM) status of the clonotypic immunoglobulin heavy variable (IGHV) gene has established predictive and prognostic relevance. The SHM status is assessed based on the number of mutations within the IG heavy variable domain sequence, albeit only over the rearranged IGHV gene excluding the variable heavy complementarity determining region 3 (VH CDR3). This may lead to an underestimation of the actual impact of SHM, in fact overlooking the most critical region for antigen-antibody interactions, i.e. the VH CDR3. Here we investigated whether SHM may be present within the VH CDR3 of cases bearing 'truly unmutated' IGHV genes (i.e. 100% germline identity across VH FR1-VH FR3) employing Next Generation Sequencing. We studied 16 patients bearing a 'truly unmutated' CLL clone assigned to stereotyped subsets #1 (n=12) and #6 (n=4). We report the existence of SHM within the germline-encoded 3'IGHV, IGHD, 5'IGHJ regions of the VH CDR3 in both the main IGHV-IGHD-IGHJ gene clonotype and its variants. Recurrent somatic mutations were identified between different patients of the same subset, supporting the notion that they represent true mutational events rather than technical artefacts; moreover, they were located adjacent to/within AID hotspots, pointing to SHM as the underlying mechanism. In conclusion, we provide immunogenetic evidence for intra-VH CDR3

variations, attributed to SHM, in CLL patients carrying 'truly unmutated' IGHV genes. Although the clinical implications of this observation remain to be defined, our findings offer a new perspective into the immunobiology of CLL, alluding to the operation of VH CDR3-restricted SHM in U-CLL.

#### KEYWORDS

CLL (chronic Lymphocytic Leukemia), B cell receptor, antigen binding, somatic hypermutation, immunoglobulin genes

## Introduction

The somatic hypermutation (SHM) status of the clonotypic, rearranged immunoglobulin heavy variable (IGHV) gene is a cornerstone for risk stratification of patients with chronic lymphocytic leukemia (CLL) (1–4). Depending on the SHM burden, i.e. the number of mutations within the sequence of the rearranged IGHV gene, cases are classified in two categories, namely unmutated CLL (U-CLL) and mutated CLL (M-CLL), with different biological background and clinical behavior (5–8).

The established approach for determining the SHM burden relies on the robust identification of nucleotide changes across the sequence of the rearranged IGHV gene, excluding the heavy variable complementarity determining region 3 (VH CDR3). This is mostly due to the difficulty in discriminating between actual SHM and random nucleotides added in the junction between the recombined IGHV, IGHD and IGHJ genes (9–11). However, this approach may result in the underestimation of the actual impact of SHM, in fact overlooking the most critical region for antigen-antibody interactions, i.e. the VH CDR3.

Here we investigated the possibility that SHM may also be present in CLL cases bearing 'truly unmutated' clonotypic IGHV genes (i.e. those with 100% germline identity across the VH FR1-VH FR3). To that end, we focused on two well characterized major stereotyped subsets: subset #1 (clan I IGHV genes/IGHD6-19/IGHJ4) and subset #6 (IGHV1-69/IGHD3-16/IGHJ3), displaying germline-encoded amino acid (aa) motifs QWL and YDYVWGSY within the respective VH CDR3 that originate from the IGHD6-19 and the IGHD3-16 gene, respectively (12). However, as reported by previous Sanger-based studies, patients assigned to both subsets can exhibit variations in these motifs that could potentially represent SHM events (12).

In order to address our starting question, we studied the IG gene repertoire of cases assigned to subsets #1 and #6 utilizing next generation sequencing (NGS), which enables the assessment of the subclonal architecture of antigen receptor gene repertoires at a high resolution (13, 14). We report VH CDR3-focused variations, very likely attributed to SHM, in CLL patients carrying 'truly unmutated' IGHV genes. While the small

number of our cohort limits our capacity to safely predict the clinical relevance of this observation, our findings highlight the possible need to reappraise definitions regarding the characterization of the SHM status in CLL.

## Materials and methods

### Patient cohort

The study group comprised 12 patients assigned to the stereotyped subset #1 and 4 patients assigned to stereotyped subset #6. Detailed information on the immunogenetic features of the clonotypic IGHV-IGHD-IGHJ gene rearrangements is provided in [Supplemental Table 1](#). All 16 patients were selected for exhibiting a 'truly-unmutated' IGHV status i.e. 100% germline IGHV identity as determined by Sanger sequencing. The study was approved by the local Ethics Review Boards of the participating institutions and was conducted in accordance with the declaration of Helsinki.

### IGHV-IGHD-IGHJ amplification and high-throughput sequencing methodology

Total RNA was isolated from Peripheral Blood Mononuclear Cells (PBMCs) and 1 µg was reverse transcribed to cDNA with the SuperScript™ II Reverse Transcriptase (Invitrogen, UK). IGHV-IGHD-IGHJ rearrangements were PCR amplified from 40 nanograms of cDNA using the Platinum™ Taq DNA Polymerase (Invitrogen, UK). IGHV subgroup-specific forward primers annealing to the leader region of the respective IGHV gene (depending on the case), and reverse primers annealing either to the IGHJ gene or the IGHC genes were utilized, the latter in order to amplify isotype-specific transcripts (15, 16). All amplicons were gel-purified with the Monarch® DNA Gel Extraction Kit (New England Biolabs, USA) and 85 ng were used for library preparation. NGS libraries were prepared with a Dual Indexing sequencing

approach according to the manufacturer's instructions (NEBNext<sup>®</sup> Ultra<sup>™</sup> II FS DNA Library Prep Kit, New England Biolabs, USA). Library quantification was performed with Qubit<sup>™</sup> (ThermoFisher Scientific), whereas purity and size were checked with the Fragment Analyzer<sup>™</sup> Automated CE System (Agilent Technologies, USA). Paired-end NGS was carried out on a MiSeq Benchtop Sequencer (MiSeq reagent kit v3, 2x300 bp, Illumina Inc.)

## Bioinformatics analysis

Base calling, adapter trimming, and demultiplexing were performed by the Illumina signal processing software. Quality control of the raw NGS data was performed with a purpose-built, in-house pipeline. Briefly, we performed length and quality filtering, excluding low quality reads, then merged the paired-end NGS reads, followed again by length and quality filtering of the joined, full-sequence reads. Finally, to ensure high sequence quality in the VH CDR3 region, we included an additional filter that ensured a Qscore $\geq$ 30 for nucleotides in the 30-45 nucleotide stretch ahead of the GXG motif in the FR4 region (Supplemental Table 2). High-quality sequences were then annotated with IMG/HighV-QUEST (<https://www.imgt.org>) and meta-data analysis was performed with tripr (17) and custom scripts in R. To further ensure that biases due to sequencing errors would not be taken into account, we used the OLGEN coverage limit calculator (18) (<http://app.olgen.cz/clc>) in order to determine the minimum number of reads per variant that should be considered for analysis, based on the sequencing error induced during the NGS process. In our case, for a sequencing error of 2.4%, determined by the percentage of sequences aligned to PhiX, we set the minimum cutoff of reads at 66.

## Definitions

Clonotypes were defined as unique IGHV-IGHD-IGHJ nucleotide gene rearrangement sequences. The clonotype accounting for the majority of reads in a given sample was characterized as the main variant. All clonotypes utilizing the same IGHV gene and bearing a VH CDR3 amino acid sequence of same length as the main variant but exhibiting a maximal difference of up to two amino acids were defined as subclonal variants. Subclonal variants displaying different isotypes, namely IGHG or IGHA, were defined as switched variants.

## Statistical analysis and data visualization

Descriptive statistics for clonotype computation included counts and frequency distributions. For quantitative variables, we calculated the mean, median and minimum/maximum

values. The Mann-Whitney U test was used to compare the levels of subclonal heterogeneity between subsets #1 and #6. The Kruskal-Wallis test was used to assess the presence of any amino acid positions in the VH CDR3 germline-encoded regions with significant differences in mutation frequencies. The Wilcoxon rank sum test was then used for *post-hoc* comparisons, in order to identify the exact amino acid positions for which the aforementioned significance was observed. P-values were adjusted using the Holm-Bonferroni correction. A significance level of  $\alpha=0.05$  was set. Data visualization was performed with custom R scripts.

## Results

### Overview of the NGS output

Forty PCR amplicons were sequenced, in particular: 36 amplicons corresponded to the mu, gamma and alpha transcripts of 12 cases assigned to subset #1, while 4 amplicons corresponded to mu transcripts of 4 cases assigned to subset #6.

In total, 8,955,886 raw reads were generated (median 215,546 reads/sample, range 109,639-466,729) corresponding to 5,855,795 merged, full-sequence reads (median 142,719 reads/sample, range 63,298-341,654), of which 5,329,188 (median 123,323/sample, range 55,282-320,999) represented productive IGHV-IGHD-IGHJ gene rearrangements. As the scope of the study was to examine the impact of SHM within the VH CDR3 region of 'truly unmutated' gene rearrangements, we excluded from downstream analysis any subclonal variants that exhibited nucleotide changes across the germline VH FR1 - VH FR3 part of the VH domain. Hence, we ended up with a total of 2,409,663 'truly-unmutated' IGHV-IGHD-IGHJ gene rearrangement sequences (median 109 sequences/sample, range 33-198,699) as defined by their 100% identity to the respective IGHV germline gene.

### Clonal distribution and subclonal heterogeneity of 'truly-unmutated' IGHV-IGHD-IGHJ gene rearrangements in the BcR IG repertoire

After identifying all 'truly-unmutated' clonotypic IGHV-IGHD-IGHJ mu transcripts in each patient, we computed the number of the corresponding clonotypes as well as the number of the respective subclonal variants. In subset #1, the main CLL variant was present at a median frequency of 63.9% (range 56.2%-71.2%), whereas subclonal variants were identified in all cases (median: 22, range: 7-39 subclonal variants) accounting for a minor fraction of the respective repertoire (median frequency: 0.1%, range: 0.02%-2.27%). In subset #6, the main CLL variant

was present at a median frequency of 54.0% (range: 41.1%-63.7%), while subclonal variants were again identified in all cases (median 7, range 1-10 subclonal variants) again accounting for a minor fraction of the total repertoire (median frequency: 0.1%, range: 0.03-0.21%). The remaining clonotypic background consisted of: (i) clonotypes unrelated to the main CLL clonotype, (ii) CLL-related subclonal variants with <100% IGHV identity, and (iii) subclonal variants of very small frequencies, which were discarded due to the high possibility of representing artefacts. (Figure 1).

## Evidence for SHM within the VH CDR3 in CLL cases assigned to stereotyped subset #1

Stereotyped subset #1 cases carry a 13 amino acid-long VH CDR3 with a highly conserved QWL motif at positions 4-6 encoded by the IGHD6-19 gene in reading frame 1. Our findings from the present NGS analysis were in line with this, since all subset #1 patients carried this QWL motif at the exact same positions. However, NGS also revealed subclonal variants in both this motif but also in other IGHV and IGHD germline-encoded codons within the VH CDR3. The topological analysis of these variations revealed a preferential targeting of VH CDR3

amino acid positions 2, 4 and 12 versus the remaining codons ( $p < 0.05$ ).

In more detail: (i) all 12 patients carried a subclonal R>G aa substitution at position 2 of the VH CDR3 (IGHV-encoded; median frequency: 0.12%, range: 0.08-0.13%); (ii) 11/12 patients (91.6%) carried a subclonal Q>R aa substitution at position 4 within the QWL motif (IGHD-encoded; median frequency: 0.1%, range: 0.09-0.12%); and, finally, (iii) all 12 patients carried a subclonal D>G aa substitution at position 12 (IGHJ-encoded; median frequency: 0.11%, range: 0.08-0.13%). Notably, all three aforementioned subclonal substitutions resulted from an a>g nucleotide transition at the respective codons, with the resulting amino acid changes conferring different physicochemical properties from the germline-encoded ones (Figure 2).

Next, in order to further characterize recurrent SHMs identified within the VH CDR3 of subset #1 cases we investigated whether these SHMs were also present in switched gamma and alpha clonotypic variants. The main switched variants were 100% identical to the main mu variant in all cases. To ensure that this finding reflects actual immunogenetic relatedness, the respective primer sets used for the amplification of gene rearrangements of different isotypes were different, while the respective PCR amplicons were prepared and sequenced in different batches. Indeed, 8/12 (75%) patients shared the R>G aa substitution at VH CDR3

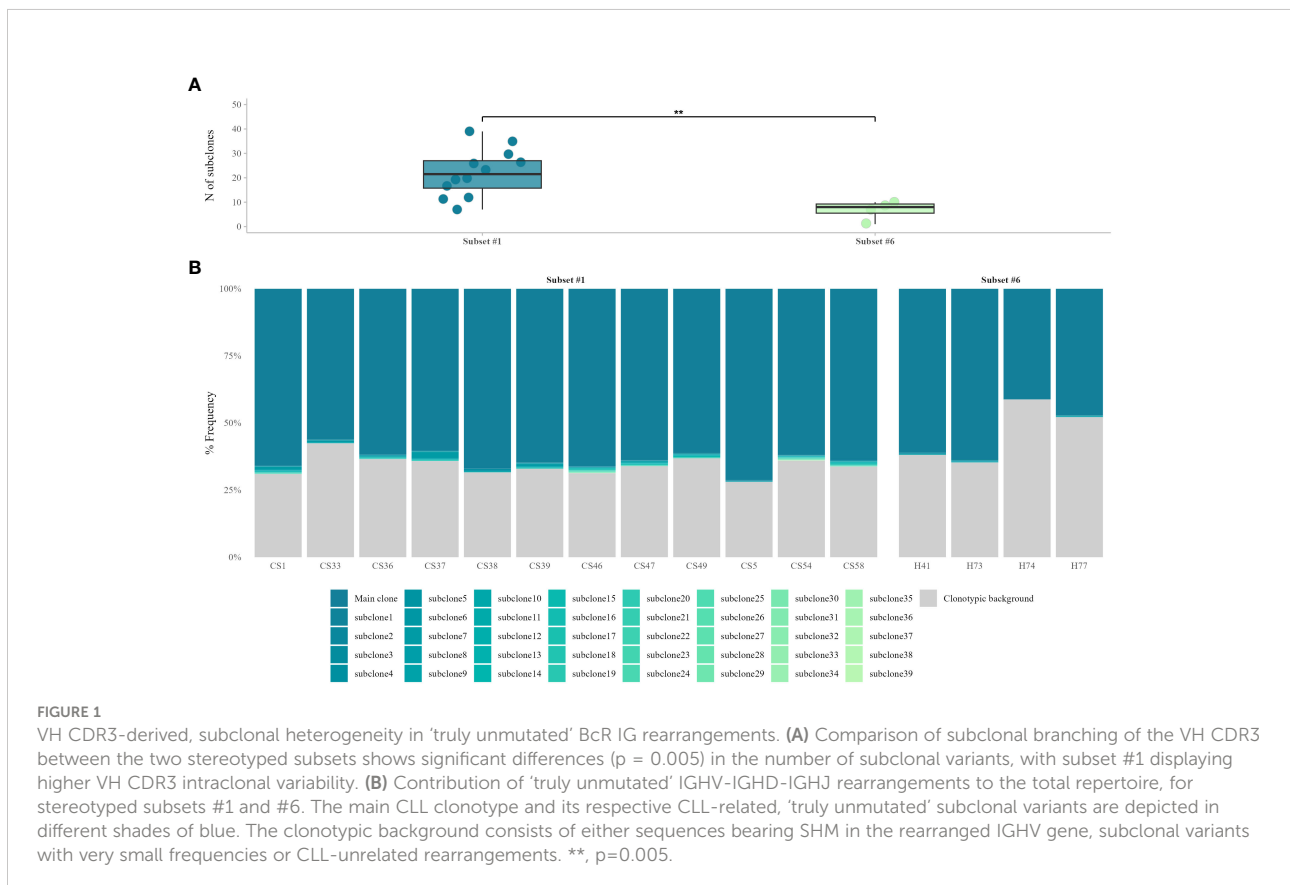




FIGURE 2

VH CDR3 amino acid sequence logos for subset #1 patients showing in color the positions with the most frequent substitutions in the VH CDR3 of the dominant mu and the switched gamma or alpha variants, as well as the most frequent substitutions at these positions. Each amino acid is colored according to its physicochemical properties, using the IMGT coloring system (<https://www.imgt.org/IMGTEducation/>), namely the basic (red) arginine as well as the acidic (yellow) aspartic residues mutate to the nonpolar (green) uncharged glycine residue, whereas the uncharged and polar (dark yellow) glutamine is replaced by a basic arginine residue (red).

position 2 in either the gamma or the alpha switched variants or both (median frequency: 0.11% for both transcripts, range: 0.05–1.08%, median frequency for gamma and alpha transcripts: 0.08 and 0.14%, respectively). Moreover, the Q>R aa substitution at VH CDR3 position 4 was shared by 8/12 (75%) patients in at least one gamma (median frequency: 0.08%, range: 0.05%–0.1%) or alpha switched variant (median frequency: 0.09%, range: 0.05–0.13%). Finally, the D>G aa substitution at VH CDR3 position 12 was shared by 7/12 (66%) patients in at least one gamma (median frequency: 0.1%, range: 0.09–0.13%) or alpha switched variant (median frequency: 0.09%, range: 0.05–0.12%). Cases where we did not document switched variants with the aforementioned replacement SHMs exhibited overall lower levels of VH CDR3-derived subclonal heterogeneity in the switched variants compared to the main, mu-expressing CLL variant (median 22 subclonal variants for the main variant vs. median 15 and 8 subclonal variants for gamma, alpha switched variants respectively); this explains, at least in part, the absence of VH CDR3-focused SHM in some of these cases.

### Evidence for SHM within the VH CDR3 in CLL cases assigned to stereotyped subset #6

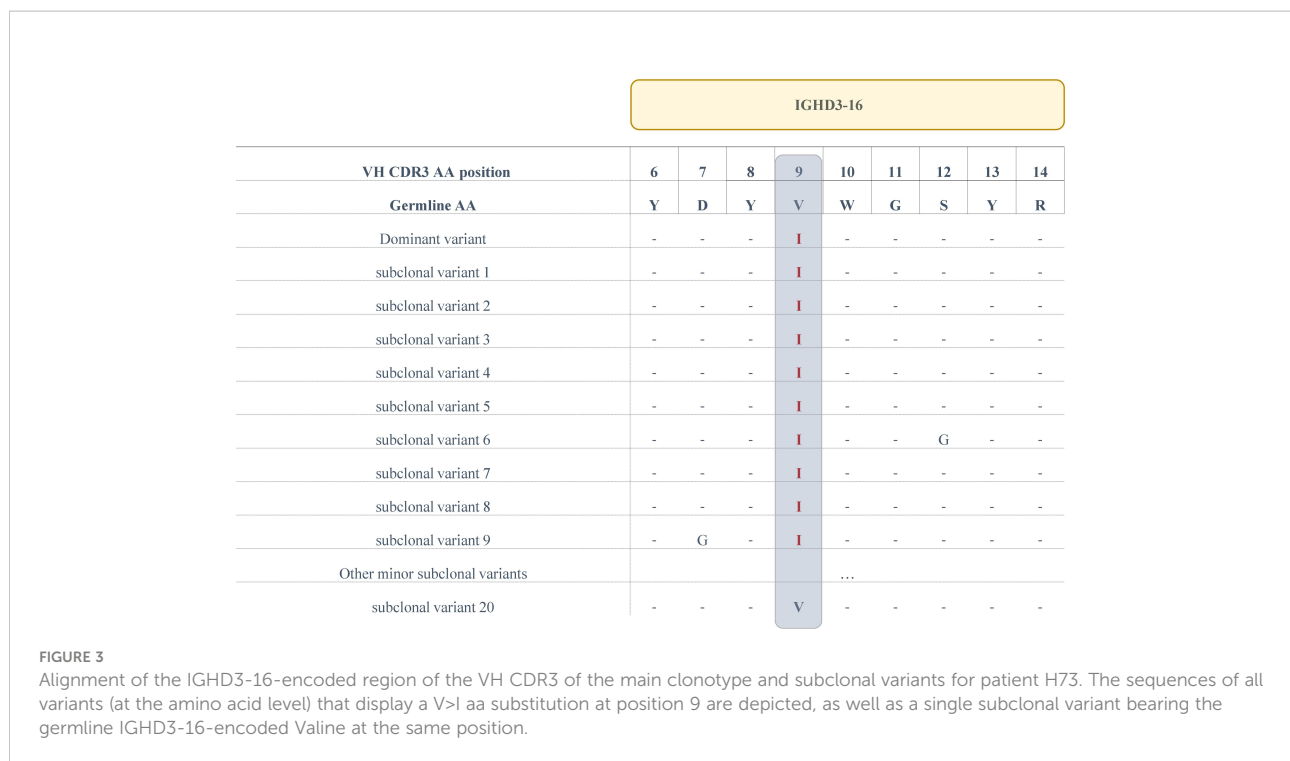
Stereotyped subset #6 is characterized by the expression of a 21 amino acid-long VH CDR3 with high conservation regarding not only the IGHV-, IGHD- or IGHJ-encoded codons but also

the IGHV-IGHD and IGHD-IGHJ gene junctions. However, VH CDR3 codon 9, which is IGHD-encoded, has been reported to display variation (12): indeed, a significant fraction of cases analyzed by Sanger sequencing bear an Isoleucine (I) residue rather than the Valine (V) residue encoded by the germline sequence of the IGHD3-16 gene in reading frame 2.

In the present series, 2/4 (50%) subset #6 patients were known by previous Sanger sequencing to carry a clonal V>I aa substitution at VH CDR3 codon 9. Results from the current NGS experiments confirmed this finding; in more specific, ‘truly unmutated’ clonotypic rearrangement sequences bearing this change represented 64.7% of the total repertoire of patient H73 and 41.2% of patient H74, respectively. The V>I aa substitution was also identified in all subclonal variants of these patients with a median frequency of 0.1% (range: 0.08%–0.21%). Unlike subset #1, switched variants are not detected in subset #6 (19), therefore we could not assess recurrence of the documented V>I aa substitution in a similar manner to that of subset #1. That notwithstanding, it is worth noting that we detected a minor variant (33 reads) carrying the germline-encoded V residue at codon 9 in patient H73 (Figure 3), which we report despite being below the adopted cutoff of 66 reads on the grounds that V is the germline residue. On this evidence, it becomes apparent that the clonal V>I substitution in patient 73 represents a true SHM result rather than a sequencing artefact.

Similar to subset #1, NGS analysis disclosed additional recurrent SHMs at the subclonal level, clustered at particular VH CDR3 codons. In more detail, 3/4 (75%) subset #6 patients





carried a R>G aa substitution at VH CDR3 position 2 (IGHV-encoded; median frequency: 0.11%, range: 0.09%-0.13%), whereas all 4 patients (100%) carried a F>V aa substitution resulting from a t>g transversion at VH CDR3 codon 19 (IGHJ-encoded; median frequency: 0.2%, range: 0.13%-0.21%).

## SHM topology in relation to AID hotspots

In order to obtain more supportive evidence for the origin of the identified variations, we scanned the IGHV, IGHD, IGHJ germline nucleotide sequences of the VH CDR3 for the topological overall between the identified variants and hotspots of the activation-induced deaminase (AID). Starting from subset #1, all three mutations that were subclonally detected in the majority of patients resulted from a>g transitions. Such mutations usually arise after U:G mismatch repair *via* the mismatch repair machinery pathway (MMR) and are adjacent to the deaminated cytosine of the AID WRC/GYW hotspots (W=A/T, R=A/G, Y=T/C). Indeed, the 3' IGHV-encoded arginine (R) residue, as well as the IGHJ4-encoded aspartic acid (D) residue were localized in close proximity with two of these hotspots, namely TGC/GCA in the IGHV gene, and AAC/GTT in the IGHJ gene. The Q>R aa substitution found within the IGH6-19-encoded QWL motif and detected in 91.6% of subset #1 patients, also resulted from an a>g transition. Interestingly, in 9/11 patients in which a detected subclonal variant was found to bear the Q>R aa substitution, we observed

generation of an overlapping WGCW SHM hotspot or a WRC motif (AGCA or AAC) either at the N1-IGHD junction, or directly at the IGHV-IGHD junction, adjacent to the codon encoding for the glutamine (Q) residue. These findings indicate that there may be selective pressure for SHM targeting at this particular position.

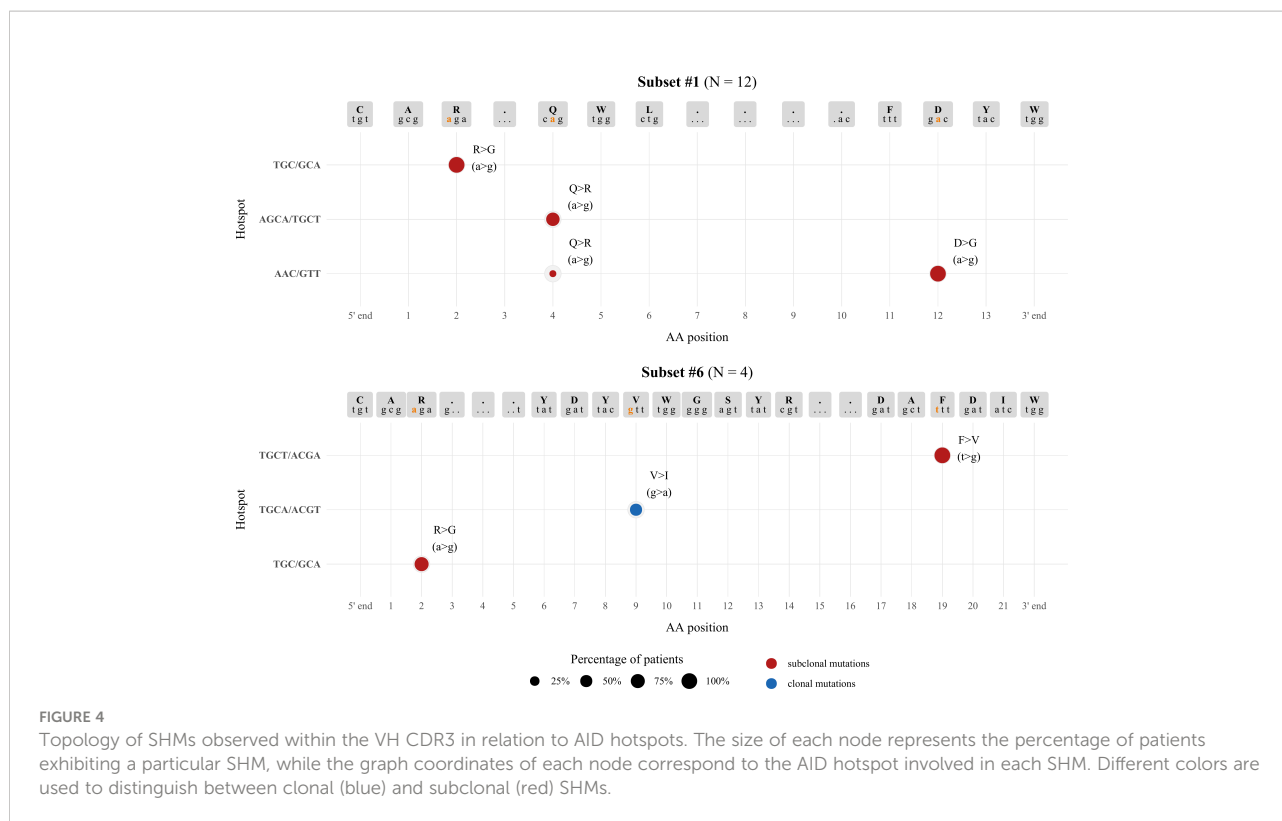
For subset #6, the clonal V>I aa substitution results from a g>a transversion located within an overlapping motif present in both strands (AGCT/ACGT), likely generated after mismatch repair from the replication system (Figure 4).

Altogether, this analysis supports the notion that the mutations occurring within the VH CDR3 likely represent *bona fide* SHM rather than NGS artefacts.

## Discussion

Determination of the SHM status in CLL is key to disease prognostication and prediction of clinical outcome (20). Going beyond the binary distinction between U-CLL versus M-CLL, increasing evidence supports the theory that BcR IG stereotypy may assist in refining risk stratification, given that cases belonging to the same stereotyped subset share several clinical and biological features, including recurrent SHMs (12, 21–23).

The study of SHM in various B cell malignancies, including CLL, has offered valuable insight into disease ontogeny and evolution, particularly as it concerns derivation and interactions with antigens (24, 25). Particularly for U-CLL, which is the focus



of the present work, the prevailing view is that U-CLL likely derives from a cell differentiating independently from a germinal center reaction, yet the cell(s) of origin remain to be conclusively defined (26–29). That notwithstanding, U-CLL is defined based on the 98% germline identity cutoff, thus it represents an assortment of cases with varying SHM status, ranging from limited to none, the latter referred to as ‘truly-unmutated’. This definition implies a complete absence of SHM, which may be misleading considering that the prognostically relevant determination of SHM status is confined to the sequence of the rearranged IGHV gene. Hence, the VH CDR3, i.e. the most diverse part of the BcR IG and most relevant for antigen recognition, is completely ignored.

In the present work we sought to obtain evidence for the existence of SHM in IGHV-IGHD-IGHJ gene rearrangements from ‘truly unmutated’ CLL clones. To that end, we focused on U-CLL stereotyped subsets #1 and #6 and selected cases expressing IGHV genes with 100% identity to the respective germline IGHV gene, as previously documented by Sanger Sequencing. We applied NGS in order to obtain a comprehensive view of the (sub)clonal architecture of the BcR IG gene repertoire (30). Moreover, given our choice to study ‘truly unmutated’ rearrangements, we applied a series of stringent filters to our NGS data in order to ensure the absence of SHM in the VH FR1 to VH FR3 as well to exclude potential sequencing artefacts.

Intraclonal diversification analysis revealed the presence of subclonal branching in both subsets #1 and #6. This is in keeping

with previous studies reporting the existence of subclonal heterogeneity within the clonotypic CLL BcR IG gene rearrangements likely in the context of ongoing antigen interactions (31–33). That said, the aforementioned studies focused on the IGHV-encoded regions (VH FR1 to VH FR3) and, moreover, examined mainly M-CLL cases. Hence, our results show not only that subclonal heterogeneity is present in U-CLL, but also that this diversification can be the result of specific SHM targeting of the VH CDR3, even if the clonotypic IGHV gene exhibits complete lack of SHM. The observed differences in the levels of subclonal branching between stereotyped subsets #1 and #6 are not entirely surprising since: (i) subset #6 is known for very high levels of conservation in terms of the aa composition of the VH CDR3 (12), which could be likely reflected at the subclonal level, and (ii) each stereotyped subset represents a distinct entity, with notable differences in the aa composition and, thus, structure of the BcR IG, arguably affected by distinct factors/stimuli driving (sub)clonal evolution (34–36).

In regard to stereotyped subset #1, we detected recurrent subclonal events within the IGHD-encoded QWL motif as well as in the IGHV- and IGHJ- encoded parts of the VH CDR3. Such recurrence combined with the presence of the same events in the respective switched variants strengthens the argument that these substitutions are indeed true SHM-induced mutational events, likely attributed to ongoing antigenic interactions post-transformation in subset #1. Turning to

subset #6, both the F>V aa substitution subclonally detected in the IGHJ3-encoded FDIW motif of all patients, as well as the clonal V>I aa substitution observed in half the cases resulted in a conservative substitution that is not anticipated to affect significantly the overall physicochemical properties of the respective regions, maintaining their nonpolar profile.

Similar to normal B cells (37), SHM in CLL preferentially clusters within certain hotspot motifs which represent targets of the AID enzyme (38). Of note, all VH CDR3 nucleotide changes identified in subset #1 as well as the respective changes in the subclonal variants of subset #6 patients were located at a/t sites adjacently to AID hotspots, consistent with a non-canonical SHM mechanism (39). These findings strengthen our hypothesis that the mutations reported in the present work are indeed results of the SHM process. The clonal V>I aa substitution in subset #6 resulted from a g>a transversion at the respective codon, which is located at an overlapping WGCW hotspot, pointing to canonical SHM as the underlying mechanism.

Overall, the herein reported VH CDR3-focused SHM, present in all patients of our cohort, may imply that U-CLL BcR IG may diversify post-transformation through ongoing antigen interactions critically mediated *via* the VH CDR3 region. These results question the existence of ‘truly-unmutated’ CLL, with important ontogenetic implications. Their clinical relevance cannot be determined presently, especially given the limited cohort size, and will require formal investigation in large, well characterized clinical cohorts.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: The NGS data generated for this study have been deposited and are publicly available in the European Nucleotide Archive (ENA of EMBL-EBI) under accession number PRJEB56468 (ERP141418).

## Ethics statement

The studies involving human participants were reviewed and approved by The Bioethics committees of Centre for Research and Technology, Hellas and Aristotle University of Thessaloniki. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

ES performed experiments, analyzed data and wrote the manuscript. LZ-I, NP, FP assisted with bioinformatics analysis

and data visualization. GK assisted with statistical analysis. SN assisted with experiments. NS provided clinical samples. EG, AA, FD, AC assisted in the interpretation of results. KS designed and supervised the study and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

The research work was supported in part by the Hellenic Foundation for Research and Innovation (HFRI) under the 3rd Call for HFRI PhD Fellowships (Fellowship Number: 5845); the project ODYSSEAS (Intelligent and Automated Systems for enabling the Design, Simulation, and Development of Integrated Processes and Products), implemented under the “Action for the Strategic Development on the Research and Technological Sector,” funded by the Operational Programme “Competitiveness, Entrepreneurship, and Innovation” (NSRF 2014-2020), and cofinanced by Greece and the European Union with grant agreement no. MIS 5002462; the project “BBMRI: Biobanking and Biomolecular Resources Research Infrastructure,” implemented under the action “Reinforcement of the Research and Innovation Infrastructure,” funded by the Operational Programme “Competitiveness, Entrepreneurship, and Innovation” (NSRF 2014-2020), and cofinanced by Greece and the European Union (European Regional Development Fund) with grant agreement no. MIS5028275; the project “Employing NGS technology for improved, non-invasive early detection, staging and prediction of progression in lymphoma patients” – TRANSCAN NOVEL, funded by the ERANET on Translational Cancer Research JTC2016 program.

## Acknowledgments

The authors would like to thank Mr. Konstantinos Pasentsis for his valuable insights into the experimental procedure and NGS methodology.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated



organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Fais F, Ghiotto F, Hashimoto S, Sellars B, Valetto A, Allen SL, et al. Chronic lymphocytic leukemia b cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest* (1998) 102(8):1515–25. doi: 10.1172/JCI3009
- Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated ig VH genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* (1999) 94(6):1848–54. doi: 10.1182/blood.V94.6.1848
- Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* (2018) 131(25):2745–60. doi: 10.1182/blood-2017-09-806398
- Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* (1999) 94(6):1840–7. doi: 10.1182/blood.V94.6.1840
- Visentin A, Facco M, Gurrieri C, Pagnin E, Martini V, Imbergamo S, et al. Prognostic and predictive effect of IGHV mutational status and load in chronic lymphocytic leukemia: Focus on FCR and BR treatments. *Clin Lymphoma Myeloma Leuk* (2019) 19(10):678–85.e4. doi: 10.1016/j.clml.2019.03.002
- Bulian P, Rossi D, Forconi F, del Poeta G, Bertoni F, Zucca E, 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(1):18. doi: 10.1186/1479-5876-10-18
- Lin KI, Tam CS, Keating MJ, Wierda WG, O'Brien S, Lerner S, et al. Relevance of the immunoglobulin VH somatic mutation status in patients with chronic lymphocytic leukemia treated with fludarabine, cyclophosphamide, and rituximab (FCR) or related chemoimmunotherapy regimens. *Blood* (2009) 113(14):3168–71. doi: 10.1182/blood-2008-10-184853
- Baliakas P, Moysiadis T, Hadzidimitriou A, Xochelli A, Jeromin S, Agathangelidis A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica* (2019) 104(2):360–9. doi: 10.3324/haematol.2018.195032
- Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stilgenbauer S, Stevenson F, 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
- Rosenquist R, Ghia P, Hadzidimitriou A, Sutton LA, Agathangelidis A, Baliakas P, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia* (2017) 31(7):1477–81. doi: 10.1038/leu.2017.125
- Agathangelidis A, Chatzidimitriou A, Chatzikonstantinou T, Tresoldi C, Davis Z, Giudicelli V, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: the 2022 update of the recommendations by ERIC, the European research initiative on CLL. *Leukemia* (2022) 36(8):1961–8. doi: 10.1038/s41375-022-01604-2
- Agathangelidis A, Darzentas N, Hadzidimitriou A, Brochet X, Murray F, Yan XJ, 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):4467–75. doi: 10.1182/blood-2011-11-393694
- Hom JR, Tomar D, Tipton CM. Exploring the diversity of the b-cell receptor repertoire through high-throughput sequencing. *Methods Mol Biol* (2022) 2421:231–41. doi: 10.1007/978-1-0716-1944-5\_16
- Stamatopoulos B, Timbs A, Bruce D, Smith T, Clifford R, Robbe P, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia* (2017) 31(4):837–45. doi: 10.1038/leu.2016.307
- Potter KN, Mockridge CI, Neville L, Wheatley I, Schenk M, Orchard J, et al. Structural and functional features of the b-cell receptor in IgG-positive chronic lymphocytic leukemia. *Clin Cancer Res* (2006) 12(6):1672–9. doi: 10.1158/1078-0432.CCR-05-2164
- Agathangelidis A, Rosenquist R, Davi F, Ghia P, Belessi C, Hadzidimitriou A, et al. Immunoglobulin gene analysis in chronic lymphocytic leukemia. *Methods Mol Biol* (2019) 1881:51–62. doi: 10.1007/978-1-4939-8876-1\_5
- Kotouza MTH, Gemenetzi K, Galigalidou C, Vlachonikola E, Pechlivanis N, Agathangelidis A, et al. TRIP - T cell receptor/immunoglobulin profiler. *BMC Bioinf* (2020) 21(1):422. doi: 10.1186/s12859-020-03669-1
- Petrackova A, Vasinek M, Sedlarikova L, Dyskova T, Schneiderova P, Novosad T, et al. Standardization of sequencing coverage depth in NGS: Recommendation for detection of clonal and subclonal mutations in cancer diagnostics. *Front Oncol* (2019) 9:851. doi: 10.3389/fonc.2019.00851
- Xochelli A, Stalika E, Sutton LA, Douka V, Karypidou M, Iskas M, et al. Distinct profiles of *in vivo* class switch recombination in chronic lymphocytic leukemia subsets with stereotyped b cell receptors, suggestive of distinct modes of activation by antigen. *Blood* (2012) 120(21):1777–7. doi: 10.1182/blood.V120.21.1777.1777
- The 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):779–90.
- Agathangelidis A, Chatzidimitriou A, Gemenetzi K, Giudicelli V, Karypidou M, Plevova K, et al. Higher-order connections between stereotyped subsets: implications for improved patient classification in CLL. *Blood* (2021) 137(10):1365–76.
- Stamatopoulos K, Belessi C, Moreno C, Boudjogh M, Guida G, Smilevska T, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood* (2007) 109(1):259–70.
- Baliakas P, Hadzidimitriou A, Sutton LA, Minga E, Agathangelidis A, Nichelatti M, 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
- Tobin G, Thunberg U, Karlsson K, Murray F, Laurell A, Willander 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):2879–85.
- Murray F, Darzentas N, Hadzidimitriou A, Tobin G, Boudjogra M, Scielzo C, 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):1524–33.
- 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):4211–6. doi: 10.1158/0008-5472.CAN-14-0630
- Seifert M, Sellmann L, Bloehdorn J, Wein F, Stilgenbauer S, Dürig J, et al. Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J Exp Med* (2012) 209(12):2183–98. doi: 10.1084/jem.20120833
- Chiorazzi N, Ferrarini M. Cellular origin(s) of chronic lymphocytic leukemia: Cautionary notes and additional considerations and possibilities. *Blood* (2011) 117(6):1781–91. doi: 10.1182/blood-2010-07-155663
- Stamatopoulos K, Agathangelidis A, Rosenquist R, Ghia P. Antigen receptor stereotypy in chronic lymphocytic leukemia. *Leukemia* (2017) 31(2):282–91. doi: 10.1038/leu.2016.322
- Agathangelidis A, Vlachonikola E, Davi F, Langerak AW, Chatzidimitriou A. High-throughput immunogenetics for precision medicine in cancer. *Semin Cancer Biol* (2022) 84:80–8. doi: 10.1016/j.semcancer.2021.10.009
- Gemenetzi K, Psomopoulos F, Carriles A, Gounari M, Minici C, Plevova K, et al. Higher order immunoglobulin repertoire restrictions in clt: the illustrative case of stereotyped subsets 2 and 169. *Blood* (2020) 137(14):1895–904. doi: 10.1182/blood.2020005216
- Bagnara D, Tang C, Brown JR, Kasar S, Fernandes S, Colombo M, et al. Post-transformation IGHV-IGHD-IGHJ mutations in chronic lymphocytic leukemia b cells: Implications for mutational mechanisms and impact on clinical course. *Front Oncol* (2021) 11. doi: 10.3389/fonc.2021.640731
- Sutton LA, Kostareli E, Hadzidimitriou A, Darzentas N, Tsafaris A, Anagnostopoulos A, et al. Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors:

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.1079772/full#supplementary-material>

implications for ongoing interactions with antigen. *Blood* (2009) 114(20):4460–8. doi: 10.1182/blood-2009-05-221309

34. Sutton LA, Rosenquist R. The complex interplay between cell-intrinsic and cell-extrinsic factors driving the evolution of chronic lymphocytic leukemia. *Semin Cancer Biol* (2015) 34:22–35. doi: 10.1016/j.semcancer.2015.04.009

35. Kanduri M, Marincevic M, Halldórsdóttir AM, Mansouri L, Junevik K, Ntoufa S, et al. Distinct transcriptional control in major immunogenetic subsets of chronic lymphocytic leukemia exhibiting subset-biased global DNA methylation profiles. *Epigenetics* (2012) 7(12):1435–42. doi: 10.4161/epi.22901

36. Ntoufa S, Vardi A, Papakonstantinou N, Anagnostopoulos A, Aleporou-Marinou V, Belessi C, et al. Distinct innate immunity pathways to activation and tolerance in subgroups of chronic lymphocytic leukemia with distinct

immunoglobulin receptors. *Mol Med* (2012) 18(9):1281–91. doi: 10.2119/molmed.2011.00480

37. Tang C, Bagnara D, Chiorazzi N, Scharff MD, MacCarthy T. AID overlapping and pol $\eta$  hotspots are key features of evolutionary variation within the human antibody heavy chain (IGHV) genes. *Front Immunol* (2020) 11.

38. Yuan C, Chu CC, Yan XJ, Bagnara D, Chiorazzi N, MacCarthy T. The number of overlapping AID hotspots in germline IGHV genes is inversely correlated with mutation frequency in chronic lymphocytic leukemia. *PLoS One* (2017) 12(1):1–17. doi: 10.1371/journal.pone.0167602

39. Oppezio P, Navarrete M, Chiorazzi N. AID in chronic lymphocytic leukemia: Induction and action during disease progression. *Front Oncol* (2021) 11. doi: 10.3389/fonc.2021.634383