Video Article

Immunoglobulin Gene Sequence Analysis In Chronic Lymphocytic Leukemia: From Patient Material To Sequence Interpretation

Andreas Agathangelidis*[1], Lesley Ann Sutton*[2,3], Anastasia Hadzidimitriou[1], Cristina Tresoldi[4], Anton W. Langerak[5], Chrysoula Belessi[6], Frederic Davi[7], Richard Rosenquist[2,3], Kostas Stamatopoulos[1,2], Paolo Ghia[8]

[1]Institute of Applied Biosciences, Centre for Research and Technology Hellas

[2]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University

[3]Department of Molecular Medicine and Surgery, Karolinska Institutet

[4]Division of Immunology, Transplantation and Infectious, IRCCS San Raffaele Scientific Institute

[5]Department of Immunology, Laboratory for Medical Immunology, Erasmus University Medical Center

[6]Hematology Department, Nikea General Hospital

[7]Assistance publique - Hôpitaux de Paris (AP-HP), Hopital Pitié-Salpêtrière, Department of Hematology, and UPMC University Paris 06, UMRS 1138

[8]Division of Experimental Oncology, IRCCS Istituto Scientifico San Raffaele and Università Vita-Salute San Raffaele

*These authors contributed equally

Correspondence to: Andreas Agathangelidis at aagathangelidis@gmail.com, Lesley Ann Sutton at Lesley-ann.sutton@ki.se

Abstract

During B cell maturation, the complex process of immunoglobulin (IG) gene V(D)J recombination coupled with somatic hypermutation (SHM) gives rise to a unique DNA sequence within each individual B cell. Since B cell malignancies result from the clonal expansion of a single cell, IG genes represent a unique molecular signature common to all the malignant cells within an individual patient; thus, IG gene rearrangements can be used as clonal markers. In addition to serving as an important clonal identifier, the IG gene sequence can act as a 'molecular timeline' since it is associated with specific developmental stages and hence reflects the history of the B cell involved in the neoplastic transformation. Moreover, for certain malignancies, in particular chronic lymphocytic leukemia (CLL), the IG gene sequence holds prognostic and potentially predictive capabilities. That said, extrapolating meaningful conclusions from IG gene sequence analysis would be impossible if robust methods and tools were not available to aid in their analysis. This article, drawing on the vast experience of the European Research Initiative on CLL (ERIC), details the technical aspects and essential requirements necessary to ensure reliable and reproducible IG gene sequence analysis in CLL, a test that is now recommended for all CLL patients prior to treatment. More specifically, the various analytical stages are described ranging from the identification of the clonotypic IG gene rearrangement and the determination of the nucleotide sequence to the accurate clinical interpretation of the IG gene sequence data.

Video Link

The video component of this article can be found at https://www.jove.com/video/57787/

Introduction

Chronic lymphocytic leukemia (CLL), the most common form of leukemia in adults in western countries, is characterized by the clonal expansion of mature neoplastic B cells[1]. From a clinical perspective, the disease course is extremely variable with some patients experiencing an aggressive disease, requiring treatment very early after diagnosis, and often relapsing or being refractory to therapy. This is in stark contrast to a substantial proportion of patients (~30%) who present with an indolent disease, never require treatment, and have a life expectancy similar to healthy age-matched individuals[2]. This clinical heterogeneity is reflected in the diversity of the molecular abnormalities found in CLL patients which ultimately drive the pathogenesis and progression of this disease[3].

Establishing an accurate CLL diagnosis is usually straightforward, however, the aforementioned clinical heterogeneity may hinder the effective management of CLL patients and underscores the need for prognostic and predictive markers that could assist in treatment decision-making[2]. Numerous studies have attempted to refine the prognostication of CLL, culminating in an abundance of novel clinical and biological markers being proposed[4]. The mutational status of the clonotypic immunoglobulin heavy variable (IGHV) gene is one of the most robust prognostic markers in CLL, largely due to the fact that (i) it remains stable over time and as the disease progresses, and (ii) it is independent of other clinical

and biological parameters[5]. That notwithstanding, very few, if any markers, including IGHV mutational status, are currently applied in clinical routine at the time of diagnosis.

Initial reports on the clinical utility of IG gene sequencing in CLL date as far back as 1999, when 2 independent studies reported that patients with no or a minimal somatic hypermutation (SHM) load (unmutated CLL, U-CLL) have a worse prognosis than patients carrying a higher SHM burden (mutated CLL, M-CLL)[6,7]. More specifically, the U-CLL group consisted of cases harboring clonotypic IGHV genes with few or no SHM and hence a high percent identity to the closest germline IGHV gene (≥98%), whereas M-CLL consisted of cases with a higher mutational load (percent identity to the closest germline IGHV gene <98%). Since these early studies, it has been continually demonstrated that U-CLL cases display a shorter time-to-first treatment (TTFT) and overall survival (OS) compared to M-CLL.

In hindsight, these seminal studies highlighted the pivotal role of the B cell receptor (BcR) IG in CLL pathobiology, thus paving the way for extensive research into CLL-microenvironmental interactions which, in turn, led to a more comprehensive appreciation of the biological heterogeneity of this disease[8,9]. More recent studies have provided additional support for the importance of immunogenetic analyses in CLL by revealing that individual cases may cluster in subsets due to sharing (quasi)identical BcR IG gene sequences, a phenomenon termed BcR IG stereotypy[10,11,12,13]. Accumulating evidence supports the notion that patients assigned to the same stereotyped subset harbor similar clinicobiological properties that clearly separate them from other CLL patients within the same SHM category but with different IG gene sequences; it has indeed been reported that the categorization of CLL patients based on BcR IG stereotypy further refines the bulk segregation of CLL patients into U-CLL or M-CLL groups[14,15,16,17,18,19,20].

From a clinical perspective, it is noteworthy that the SHM status of the clonotypic IGHV gene seems to correlate with a specific response to chemoimmunotherapy. In particular, M-CLL patients treated with a combination of fludarabine/cyclophosphamide/rituximab (FCR), the gold standard treatment for medically fit CLL patients lacking *TP53* gene defects, exhibited significantly longer progression-free survival (PFS) and OS compared to U-CLL patients who received the same treatment[21,22,23]. Therefore, identification of the SHM status has become important in particular for assisting in the therapeutic management of CLL patients *i.e.,* a predictive marker, rather than for assessing the clinical course of the disease *i.e.,* a prognostic marker. In fact, according to the recent update of the NCI-sponsored guidelines from the International Workshop on CLL, the SHM status of the rearranged IGHV genes should be determined in all CLL cases prior to treatment initiation in both general practice and clinical trials[24]. Moreover, owing to the recent approval of novel treatment agents, and the growing number of drugs in trials, the SHM status of the IG molecule may play an even greater role in the therapeutic management of patients with CLL in the near future[5].

This report details all aspects of IG gene sequence analysis, beginning with choosing the appropriate material and culminating with the clinical interpretation of the sequencing results. In-depth assessment of these steps is important in diagnostic routine for the production of reliable results and to ensure accurate patient stratification within clinical trials. Protocols are provided to aid in the harmonization of IG gene sequence analysis, thus ensuring that results are comparable amongst different laboratories.

Protocol

The research protocol was approved by the Ethics Committee of the San Raffaele Scientific Institute, Milan, Italy.

1. Material Selection

1. **Choice of starting material**
   NOTE: In most cases of CLL the tumor cell count is high at diagnosis (>80-90%). Thus, cell sorting or enrichment of the leukemic cells is usually not necessary.
   1. If using Peripheral blood (PB), the most common material of choice for IG gene sequence analysis, use a blood volume between 10-20 mL.
   2. Alternatively, in cases with a low CLL cell burden, use cell sorting of PB samples or material from other tissues, like bone marrow (BM) or lymph nodes (LN).

2. **Sampling time**
   1. Sampling time is not crucial given that the IG SHM status is stable overtime; that said, avoid sampling at certain periods, such as immediately after or during therapy, since the low number of CLL cells may complicate the analysis and lead to unreliable results.

3. **Material collection and storage**
   NOTE: Collection and storage of starting material depends strongly on the material choice.
   1. For PB or BM samples, use EDTA or CPT (citrate-tris-pyridossalphosphate) collection tubes to prevent inhibition of the PCR amplification process; alternatively, use heparin tubes.
   2. For LN samples, collection options are either cell suspensions, biopsies from fresh frozen tissue or, in rare instances formalin-fixed paraffin-embedded (FFPE) tissues.

2. Density Gradient Separation

NOTE: If PB or BM is the starting material of choice, which is the most frequent scenario, mononuclear cell isolation using density gradient separation is recommended.

1. Add 200 µL of EDTA (0.5 M EDTA/ pH 8.0) to a 50 mL sterile collection tube. Add 20 mL of PB and 15 mL of RPMI. Mix by pipetting (2-3 times is sufficient).
2. Add 15 mL of density gradient medium to a new 50 mL collection tube.
   NOTE: In case the PB volume is less than 20 mL, the volumes of RPMI (previous step) and density gradient should be modified accordingly.

3. Slowly layer in the solution from step 2.1 so that it does not disrupt the gradient. Centrifuge at 800 x g for 20 minutes with the centrifuge brake off.
4. Collect the mononuclear cell ring that has formed between the upper plasma/platelet layer and the density gradient medium layer using a sterile Pasteur pipette, then transfer to a sterile 15 mL collection tube.
5. Add RPMI to a final volume of 12 mL and mix. Centrifuge at 750 x g for 8 minutes. Discard the supernatant.
6. Resuspend the cell pellet using 10 mL of RPMI and repeat step 2.5.
7. Dissolve the cell pellet in 1 mL of PBS (DPBS, no calcium, no magnesium). Perform a cell count using a Neubauer plate by mixing 20 µL of sample and 80 µL of 1:10 Türk's solution.
8. If downstream processing of the sample is not scheduled for the same day, centrifuge the sample at 750 x g for 8 minutes. Discard the supernatant and store the cell pellet at -80° C.

3. Nucleic Acid Extraction

1. Decide on a substrate for analysis of the SHM status of IG genes. Genomic DNA (gDNA) is the most popular choice as there is no need for a reverse transcription step; alternatively, the use of RNA/complementary DNA (cDNA) ensures that, at least at the transcriptional level, the productive, functional clonotypic IG rearrangement is the "favored" target.
2. Use one of numerous commercially available kits suitable for gDNA or total RNA extraction and the synthesis of cDNA (see **Table of Materials**).
3. Assess the integrity of the DNA or RNA using sensitive nucleic acid quantification systems. If using FFPE material for the analysis, assess the quality and quantity of the extracted gDNA/cDNA by PCR amplification of a known housekeeping gene, *e.g.* retinoic acid receptor alpha (*RARa*), beta-actin, *etc.*.

4. Amplification and Sequencing of IGHV-IGHD-IGHJ Gene Rearrangements

1. **IGH PCR primer selection**
   1. Preferably, perform multiplex PCR with IGHV subgroup specific 5' primers and IGHJ gene specific 3' primers[25,26]. If results are sub-optimal, prepare separate PCR mixes using single IGHV subgroup-specific primers.
   2. Use 5' primers that anneal to either the leader region located immediately upstream of the IG rearrangement or within the coding region (*e.g.* framework 1, FR1) of the IGHV gene.
      1. Use leader primers that allow for the amplification of the entire rearranged IGHV-IGHD-IGHJ gene sequence, which in turn will enable the accurate estimation of SHM levels. Thus, IGHV leader primers are recommended by ERIC in their updated guidelines for IG gene sequence analysis[27].
      2. In instances where leader primers fail to amplify the clonotypic IG rearrangement, use IGHV FR1 primers. However, framework primers do not amplify the entire clonotypic IG gene rearrangement, which may lead to inaccurate determination of the level of SHM. For this scenario, clearly state in the clinical report that the use of IGHV FR1 primers may overestimate the true percent identity to the closest germline gene.
   3. For the 3' end, use consensus primers targeting the IGHJ genes, multiplex sets of IGHJ gene-specific primers or isotype-specific primers, depending on the substrate. Primer sequences are provided in Table 1, whereas **Figure 1** depicts the various locations for primer annealing.

2. **PCR amplification of IGHV-IGHD-IGHJ gene rearrangements**
   1. Prepare the primer mix using equimolar quantities of each subgroup primer(s) to ensure unbiased amplification.
      1. For example, when using leader primers, two different primers are used to amplify rearrangements belonging to the IGHV1 subgroup. Thus, use a quantity of these 2 primers (IGHV1L and IGHV1bL) that is half compared to IGHV4L, which is the only primer for IGHV4 subgroup rearrangements.
      2. Apply the opposite approach in the case of the IGHJ1-2 and IGHJ4-5 primers. As these primers target two different IGHJ gene subgroups, double the quantity compared to the IGHJ3 and IGHJ6 primers.
      3. Use **Table 1** to determine the exact primer quantities when preparing relevant primer mixes.

   2. Mix PCR reagents as listed in **Table 2**. After combining all the PCR reagents in a PCR tube, add 0.5 µL of Taq polymerase and 100 ng of gDNA or 2 µL of cDNA (cDNA should be prepared using 1 µg of RNA).
      NOTE: A Taq enzyme with proofreading activity is not essential as the amplification error rate is extremely low and will therefore not affect the level of SHM.
   3. Run the thermal PCR protocol detailed in **Table 3**.

3. **Clonality assessment**
   NOTE: A critical next step involves evaluating the clonality pattern of the PCR product. When assessing clonality in CLL patients, the most common finding is a single, monoclonal peak/band.
   1. Use methods with high-sensitivity, such as fragment or heteroduplex analysis, for clonality assessment[28,29].
   2. Fragment analysis
      1. Perform multiplex PCRs using 5'-labeled IGHV leader or FR1 subgroup specific PCR primers; this will yield PCR products that can be separated by capillary electrophoresis, thereby permitting clonality assessment of the IGH PCR products based on their IGHV complementarity determining region 3 (CDR3) lengths.
      2. Use primers, reagents and thermal conditions identical to those previously mentioned (see Tables 1-3). Custom 5´- labeled primers are fluorescently labeled oligos with a choice of dyes on the 5' end. Examples of reporter dyes include 6-FAM, HEX, NED or TET. Thus, 2 separate reactions are required based on the use of 3 different dyes per reaction.

3. Assess the size of the PCR products on a six percent (6%) polyacrylamide gel (this percentage results in optimal resolution), using 1x TBE as a running buffer. Run at 80 V for 45 min.
   NOTE: It is recommended that PCR products are allowed to migrate on the gel for sufficient time so that monoclonal samples can be separated from the polyclonal background and the DNA size marker can separate adequately. Factors such as the size and thickness of the gel can affect migration so no single setting (voltage and time) can guarantee an optimal result That said, setting the voltage at 80 V for 45 minutes is a good starting point as it minimizes the risk of the sample migrating too quickly and 'running' off the gel.
4. Check for PCR products of approximately 500 base pairs when using IGHV leader primers and 350 base pairs when using the IGHV FR1 primer mixes.
   NOTE: In rare instances, CLL cases have been found to carry double, monoclonal products and in even rarer instances, cases may exhibit an oligoclonal pattern. An intercalating agent such as Ethidium bromide (EtBr) or less hazardous and more sensitive commercially available DNA stains can be used for the visualization of DNA on acrylamide gels using UV excitation or blue-light.

4. **PCR product purification**
   NOTE: PCR products are purified to remove any polyclonal background originating from normal B cells within the CLL sample as well as primer dimers that may lead to suboptimal sequencing. Use any method that removes unincorporated dNTPs and primers for PCR clean-up, such as an enzymatic treatment, *e.g.,* Sap (shrimp alkaline phosphatase)/Exo (Exonuclease I), or column-based purification.
   1. Load the total volume of the PCR product onto a 3% low-melting agarose gel.
   2. Allow the PCR products to run on the gel until they separate from the background.
   3. Excise the sharp, prominent PCR bands, purify and elute. If two rearrangements are detected, both bands should be excised and sequenced seperately.
   4. To perform Sap/Exo clean-up, follow the manufacturer's instructions regarding the specific volumes to use; however a typical reaction may contain 1 µL Sap, 0. 5 µL Exo and 2 µL 5x incubation buffer per 5-10 µL PCR reaction. Mix by gently vortexing each sample and incubate on a PCR block at 37 °C for 30 minutes. Inactivate the enzymes by heating to 85 °C for 15 minutes.
   5. Load a small quantity of the PCR products on a 3% agarose gel to assess the purity of the product.

5. Sanger Sequencing

NOTE: Several sequencing strategies exist, however, regardless of the exact approach, direct sequencing of both strands is mandatory to ensure a reliable result.

1. **General sequencing protocols**
   1. If amplification has been performed using single primers, proceed with sequencing using the appropriate IGHV- and IGHJ- or constant region-specific primers. This approach can also be adopted if multiplexed fluorescently labeled primers have been used and the PCR product was assessed using fragment analysis (sequencing primers should not be labeled).
   2. If multiplexed primers have been used and clonality was assessed on a gel, then use a downstream primer at the first step of sequencing *e.g.* a consensus IGHJ primer with the sequence being 5'-GTGACCAGGGTNCCTTGGCCCCAG-3'. Alternatively, a CH primer can be used if cDNA was used as substrate. Next, use subgroup-specific IGHV leader or FR1 primers for the second sequencing step.

2. **Example sequencing protocol**
   NOTE: Several sequencing protocols exist and an example protocol is detailed below.
   1. Dilute 33 ng of the PCR product in a total volume of 11.5 µL using, *e.g.,* sterile water. Denature the PCR product by heating to 95 °C for 5 minutes. Immediately place on ice for 10 minutes to prevent formation of secondary structures.
   2. Add 8 µL of master mix to each tube along with 0.5 µL (corresponding to 5 pmol) of the primer.
   3. As a control, prepare a tube containing 8 µL of master mix, 0.5 µL pUC18 control template, 2 µL (-)47 sequencing primer and 9.5 µL sterile water. The final total volume in each tube should be 20 µL.
   4. Use thermal cycling conditions for the sequencing reaction as detailed in **Table 4**.
   5. Prepare the Stop Solution by mixing 2 µL of sodium acetate 3M (pH 5.2), 2 µL EDTA 100 Mm (pH 8.0) and 1 µL of glycogen (concentration 20 mg/mL), then add 5 µL to each sample. Transfer the product to a new microcentrifuge tube.
   6. Add 60 µL of 100% ethanol (-20 °C) and gently mix. Centrifuge at 14,000 rpm for 15 minutes (4 °C). Discard the supernatant.
   7. Add 100 µL of 70% ethanol (-20 °C) and gently mix. Centrifuge at 14,000 rpm for 10 minutes (4 °C). Discard the supernatant. Repeat once.
   8. Dry the pellet for 90 minutes at room temperature. Add 40 µL of sodium dodecyl sulfate (SDS) solution to each sample.
   9. Transfer the sample to a sequencing plate, cover with strip caps or tape seals, load to the machine and perform Sanger sequencing. The sequencing plate is usually a 96-well, v-bottom, full-skirt PCR plate compatible with the sequencer. The plates may be barcoded depending on whether the sequencing is to be performed in-house, at a commercial company or at an academic sequencing center/ platform.

3. After sequencing, 'stitch together' the 2 reads generated from the individual sequencing steps to form a complete IGHV gene rearrangement *i.e.,* the consensus sequence.

6. Sequence Analysis

NOTE: The following sections focus on the output obtained from the IMGT/V-QUEST tool[30] when analyzing rearranged IG sequences, and IMGT/JunctionAnalysis[31], both of which are particularly relevant for clinical reporting and research studies. IMGT/V-QUEST is an alignment tool that compares user submitted IG sequences with the IMGT reference directory sets[30]. The tool output includes several sections within which the user can define certain parameters.

1. Specify the species, *e.g., Homo sapiens* (human), and receptor type or locus (*e.g.* IGH, IGK or IGL).
2. Paste sequences to be analysed, in FASTA format and including headers, into the text area (in batches of up to 50 sequences per run). Alternatively, an option to enter the path to a local file containing the FASTA sequences is provided.
3. Choose one of the following 3 display options for the results:
    1. Detailed view: choose this option in order to get the results for each sequence individually.
    2. Synthesis view: choose this option to compile a summary table ordered by IGHV gene and allele name or by sequence input order. This option also provides an alignment of sequences that, in any submitted batch, are assigned to the same IGHV gene and allele.
    3. Excel file: choose this option to provide all output files as spreadsheets incorporated into a single file. All viewing options have a large selection of basic and advanced parameters to choose from, however only the factors most pertinent to CLL IG sequence analysis are discussed below in the 'Representative Results' section.

7. ARResT/AssignSubsets Tool

1. To investigate BcR IG stereotypy within CLL (additional details provided below in the 'Representative results' section), submit IGHV-IGHD-IGHJ FASTA sequences to the ARResT/AssignSubsets tool[32]. The tool reports the assignment to major CLL stereotyped subsets. The subset-assignment tool together with detailed instructions are available on the ARResT/AssignSubset website[32] and also on the ERIC website under the services tab.

Representative Results

Examples provided below illustrate results obtained from IG gene sequencing analysis following the steps detailed above. Although DNA and/or RNA extraction are routine procedures for most clinical/research laboratories, a first critical step involves checking the quality/quantity of gDNA and/or RNA using a spectrophotometer or other suitable technology with high sensitivity. The next crucial step involves PCR amplification of the clonotypic IGHV-IGHD-IGHJ gene rearrangement. As mentioned above, only the use of IGHV leader primers enables amplification of the full length of the rearranged IGHV gene sequence, thus permitting the true SHM load to be determined; therefore, IGHV leader primers are the recommended primer choice (**Figure 2**). In rare instances where IGHV leader primers cannot amplify the clonotypic IG rearrangement, IGHV FR1 primers may be used. As evidenced in **Figure 3**, several PCR products/bands may be observed, especially when gDNA is the starting material; these bands should be excised and sequenced seperately. If both IGHV leader and IGHV FR1 primers fail to yield results, the analysis should be repeated using a new patient sample (when possible). The last checkpoint prior to sequencing is determining the clonality of the sample using fragment analysis (**Figure 4**).

Sequences obtained should be analyzed using the IMGT/V-QUEST tool[30]. User-selected parameters include species, receptor type or locus, search for insertion, and deletion option, *etc.* and are listed together with the number of sequences analyzed. Each FASTA sequence is displayed and the V-DOMAIN analyzed by IMGT/V-QUEST is highlighted in green. Additionally, if the input sequence was in the opposite orientation (antisense), the program automatically provides the complementary reverse sequence and states this above the FASTA sequence.

The Result summary table is provided directly below the FASTA sequence, at the top of the 'Detailed View' results page, and contains information important for the clinical reporting of IG gene sequence analysis in CLL. Each row of this table is explained below.

Result summary. The first row in the results table states the functionality of the input sequence: an IG rearranged sequence can be either productive or unproductive (**Figure 5A & 5B**). An IG gene rearrangement is unproductive if stop codons are present within the V-D-J-REGION (for VH) or within the V-J-REGION (for VL) and/or if the rearrangement is out-of-frame due to insertions/deletions. A third output is provided when the functionality cannot be determined, *i.e.,* if the IGHV-IGHD-IGHJ junction could not be identified; in this instance, the result summary would read as 'No rearrangement found' or 'No junction found'. It is important to note that only productive and, thus, functional rearrangements should be analyzed further. If the sole amplified IG rearrangement is not functional (unproductive or 'no rearrangement found'), the PCR amplification process should be repeated. If the result remains negative a new patient sample should be obtained.

V-GENE and allele. The 'V-GENE and allele' row indicates the closest germline IGHV gene and allele with its alignment score and the percent identity. If several gene(s) and alleles give the same high score, all are listed. The percent identity between the submitted sequence and the closest IGHV gene and allele is of particular importance since it determines the SHM status. This value is calculated from the first nucleotide of the V-REGION to the 3' end of the V-REGION excluding the CDR3. If IGHV FR1 primers are used, the sequence corresponding to that of the primer should always be removed to ensure as accurate a result as possible. It should be noted that a low identity percentage (<85%) may result from an insertion or deletion (discussed further below) and, should this be the case a 'Warning' note will appear in the 'Result summary' row to alert the user.

J-GENE and allele. As described for the V-gene and allele above, the 'J-GENE and allele' row indicates the closest germline IGHJ gene and allele with its alignment score and the percent identity. However, as the IGHJ gene is rather short and its 5' end may have been trimmed due to exonuclease activity, alternative choices are also searched by the tool, based on the highest number of consecutive identical nucleotides.

D-GENE and allele by IMGT/JunctionAnalysis. The 'D-GENE and allele by IMGT/JunctionAnalysis' indicates the closest germline IGHD gene and allele with the D-REGION reading frame identified by the tool in the user sequence. IMGT/JunctionAnalysis[31] provides a detailed and accurate analysis of the junction and has been integrated into the IMGT/V-QUEST tool interface. This tool manages all aspects of difficulty linked to IGHD gene and allele identification *i.e.,* (i) the short length of the D-REGION; (ii) use of 2 or even 3 reading frames; (iii) exonuclease trimming at both ends of the gene; and, (iv) the presence of mutations.

FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION. This row provides the length of the IGHV FRs and CDRs shown within brackets and separated by dots and the amino acid (AA) junction. The AA sequence of the junction illustrates (i) an in- or out-of-frame rearrangement (out-of-frame positions are indicated by the '#' sign), (ii) the presence or absence of stop codons (a stop codon is shown by an asterisk '*'), and (iii) the presence or absence of the anchors of the VH CDR3-IMGT: C (2nd-CYS 104) and W (J-TRP 118).

Somatic hypermutations predominantly consist of single nucleotide changes, however, small insertions and deletions within the V-region can be observed, albeit at a much lower frequency[33]. When using the default IMGT/V-QUEST parameters, insertions and deletions are not automatically detected; however, strong indications that a sequence may carry such alterations, *i.e.,* a low percent identity and/or different IGHV CDR1-IMGT and CDR2-IMGT lengths between the submitted user sequence and the closest germline gene and allele, are detected by the tool and a warning alert appears below the Results summary table (**Figure 5C**).

IMGT provides an option called 'Search for insertions and deletions' in the 'Advanced parameters' section located below the 'Display results' section. In cases where relevant warnings appear, it is advisable to use this functionality. When insertions and deletions are detected, comprehensive details of the finding are provided in the 'Result summary' section including (i) where the insertion or deletion is located *i.e.,* VH FR-IMGT or CDR-IMGT; (ii) the number of inserted/deleted nucleotides; (iii) the sequence (only for insertions); (iv) whether a frameshift has occurred; (v) the V-REGION codon from which the insertion or deletion starts; and (vi) the affected nucleotide position in the submitted sequence (**Figure 5D**). For insertions, the tool removes the insertion(s) from the submitted user sequence and then re-analyzes the sequence using the standard IMGT/V-QUEST parameters. If deletions are detected, the tool adds gaps, represented by dots, to replace the deleted nucleotides before repeating the analysis.

As for every diagnostic or prognostic test performed in clinical laboratories, stringent standards and a high level of reproducibility are of utmost importance. The IMGT/V-QUEST output provides much of the information necessary for reporting the results of IG gene sequence analysis in CLL, *i.e.,* (i) the IGHV, IGHD and IGHJ genes and alleles utilised; (ii) the functionality of the clonotypic IGHV-IGHD-IGHJ gene rearrangement *i.e.,* whether the rearrangement is productive or unproductive; and (iii) the percent identity of the rearranged IGHV gene and allele in comparison to its closest germline IGHV gene and allele. For comprehensive details regarding the clinical reporting of IG genes in CLL, the interpretation of problematic or technically challenging cases and recommendations for the accurate and robust determination of IGHV gene SHM status in CLL, the reader is directed to recently updated ERIC guidelines and reports[27,34].

Finally, owing to our growing knowledge concerning BcR IG stereotypy in CLL, an additional recommended requirement for the clinical reporting of IG gene rearrangements in CLL relates to the assignment of cases to major stereotyped subsets. It is now recommended to state in the clinical report whether the analyzed productive IGHV gene rearrangement can be assigned to one of the major stereotyped subsets, namely subsets #1, #2, #4 or #8[12,27]. Assignment to major CLL stereotyped subsets should be performed with the use of the ARResT/AssignSubsets tool (**Figure 6**)[32].

5' IGHV FR1 primers	Primer sequence	Quantity for 60 µL of primer mix
IGHV1	CAGGTGCAGCTGGTGCAGTCTGG	10
IGHV2	CAGGTCAACTTAAGGGAGTCTGG	10
IGHV3	GAGGTGCAGCTGGTGGAGTCTGG	10
IGHV4	CAGGTGCAGCTGCAGGAGTCGGG	10
IGHV5	GAGGTGCAGCTGTTGCAGTCTGC	10
IGHV6	CAGGTACAGCTGCAGCAGTCAGG	10
5' IGHV leader primers		
IGHV1L	AAATCGATACCACCATGGACTGGACCTGGAGG	5
IGHV1bL	AAATCGATACCACCATGGACTGGACCTGGAG(C/A)	5
IGHV2aL	AAATCGATACCACCATGGACACACTTTGCT(A/C)AC	5
IGHV2bL	AAATCGATACCACCATGGACATACTTTGTTCCAC	5
IGHV3aL	AAATCGATACCACCACCATGGAGTTTGGGCTGAGC	5
IGHV3bL	AAATCGATACCACCACCATGGA(A/G)(C/T)T(G/T)(G/T)G(G/A)CT(G/C/T)(A/C/T)GC	5
IGHV4L	AAATCGATACCACCATGAAACACCTGTGGTTCTT	10
IGHV5L	AAATCGATACCACCATGGGGTCAACCGCCATC	10
IGHV6L	AAATCGATACCACCATGTCTGTCTCCTTCCTC	10
3' IGHJ primers		
IGHJ1-2	TGAGGAGACGGTGACCAGGGTGCC	20
IGHJ3	TGAAGAGACGGTGACCATTGTCCC	10
IGHJ4-5	TGAGGAGACGGTGACCAGGGTTCC	20
IGHJ6	TGAGGAGACGGTGACCGTGGTCCC	10

Table 1. Primer sequences and quantities for the PCR amplification of the clonotypic IGHV-IGHD-IGHJ gene rearrangement.

Reagent	Volume (µL)
RB x 10 buffer	5
MgCl$_2$ (50 mM)	3
dNTPs (10 mM)	2
Primer IGHV leader/FR1 mix (10 µM)	3
Primer IGHJ mix (10 µM)	3
H$_2$O	31.5
Total volume	47.5

Table 2. Reagents for the PCR amplification of the clonotypic IGHV-IGHD-IGHJ gene rearrangement.

Temperature	Duration	Cycle number	Description
94 °C	5 minutes	1	polymerase activation
94 °C	1 minute	39	denaturation
59 °C	1 minute		annealing
72 °C	1.5 minutes		extension
72 °C	10 minutes	1	final extension
18 °C	∞	1	preservation

Table 3. Thermal cycling conditions for the PCR amplification of the clonotypic IGHV-IGHD-IGHJ gene rearrangement.

Temperature	Duration	Cycle number	Description
94 °C	5 minutes	1	polymerase activation
96 °C	20 seconds	30	denaturation
50 °C	20 seconds		annealing
60 °C	4 minutes		extension
4 °C	∞	1	preservation

Table 4. Thermal cycling conditions for the IG gene sequencing reaction.
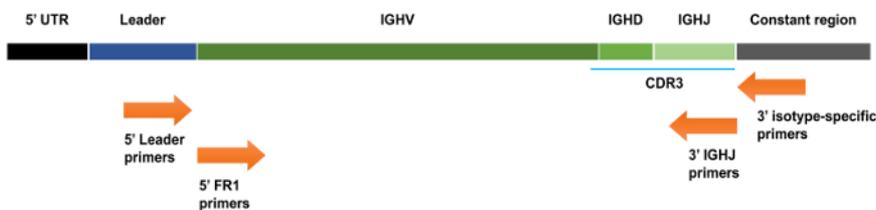
Figure 1. Schematic representation of an IGHV-IGHD-IGHJ gene rearrangement with various primer annealing locations indicated. 5' IGHV primers anneal to the leader sequence that is located upstream of the IGHV coding sequence. 5' IGHV framework (FR) primers anneal to the start of the FR1 region, which is located within the rearranged IGHV gene, whereas the 3' IGHJ primers anneal to the end of the rearranged IGHJ gene. UTR: untranslated region. Please click here to view a larger version of this figure.
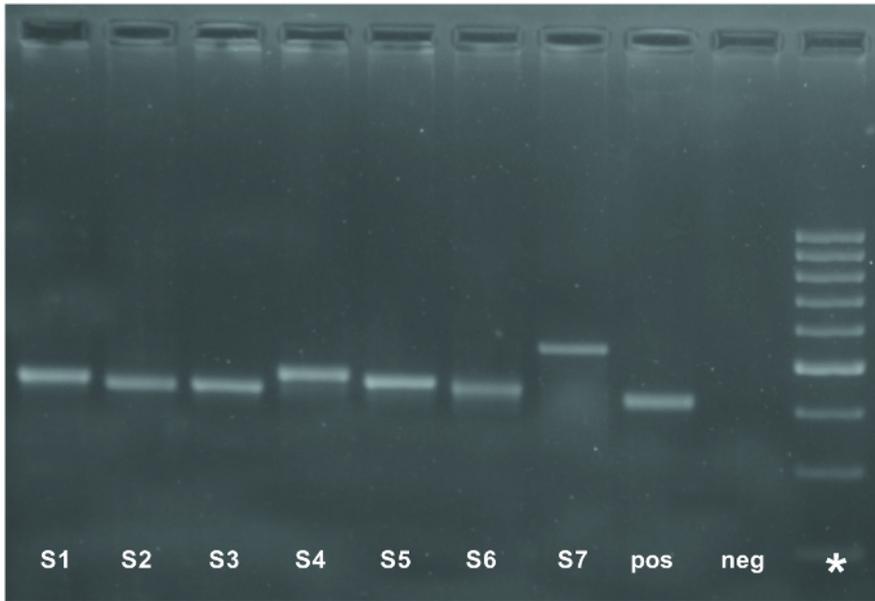
Figure 2. PCR amplification of the IGHV-IGHD-IGHJ gene rearrangement in 7 patient samples using 5' IGHV leader primers. The eighth lane represents a positive control whereas the negative control for the PCR was loaded into the ninth lane. The expected PCR product is approximately 500 base pairs in size when using IGHV leader primers. The asterisk in the last column of the gel indicates the 100 bp DNA ladder. Please click here to view a larger version of this figure.
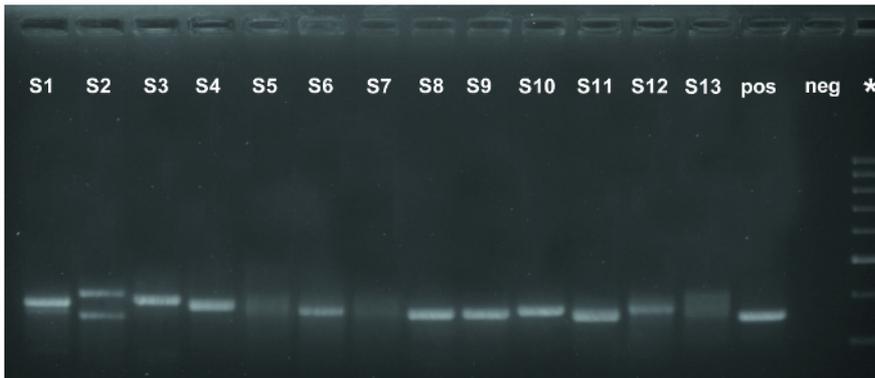

Figure 3. PCR amplification of the IGHV-IGHD-IGHJ gene rearrangement in 13 patient samples using IGHV FR1 primers. The fourteenth lane contains the positive control whilst the negative control was loaded into the fifteenth lane. As evidenced in lane 2, for which DNA was the starting material, several PCR bands are observed, which should be excised and sequenced seperately. Samples in lanes 5, 7, and 13 yielded polyclonal results (evidenced by the smear in the gel) and, thus, the PCR step should be repeated. If a second PCR fails to amplify the rearranged IG gene(s) the analysis should be performed on a new sample (if possible) or using a different primer set. The expected PCR product is approximately 350 base pairs in size when using IGHV FR1 primer mixes. The asterisk in the last column of the gel indicates the 100 bp DNA ladder. Please click here to view a larger version of this figure.
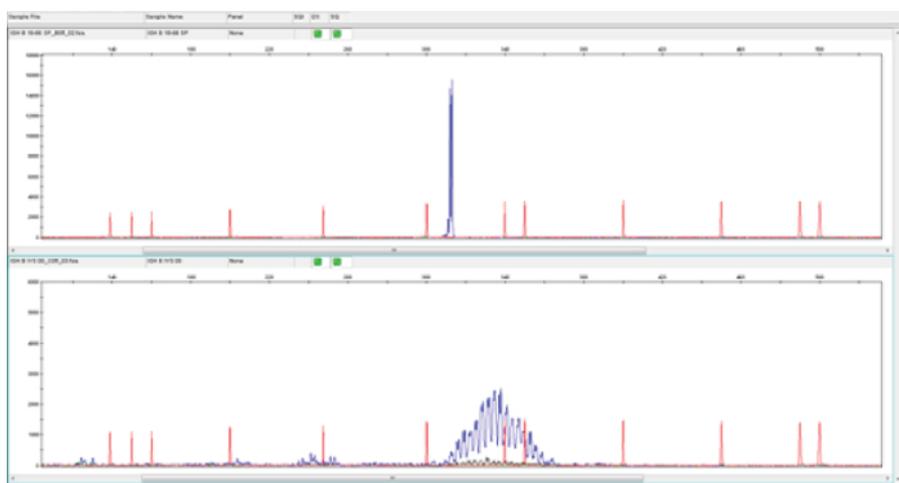
Figure 4. Assessment of clonality using genescan analysis. In genescan analysis, monoclonal PCR products give rise to a dominant peak of fluorescent products of identical size (upper panel), whereas polyclonal PCR products result in a more normal distribution of product sizes (lower panel). The red traces are peaks from a fluorescently-labeled DNA ladder that is loaded in the same capillary injection as the sample being analysed. The size standard fragments are subjected to the same electrophoretic force as the sample and are therefore exposed to the same injection conditions. The uniform spacing of the size standard fragments confirms precise size calling. The blue traces represent the monoclonal (top panel) or polyclonal (bottom panel) nature of the samples. Please click here to view a larger version of this figure.

Figure 5. Examples of the result summary tables provided by IMGT/V-QUEST. (A) Result summary table for a productive IGHV-IGHD-IGHJ gene rearrangement (no stop codon(s) and an in-frame sequence junction). The V-GENE and allele and the J-GENE and allele identified in the user sequence by IMGT/V-QUEST in this case are IGHV4-34*01 and IGHJ6*02 (on the basis of the highest alignment score evaluation). The percent identity is 99.65% due to a single somatic mutation. The D-GENE and allele identified by IMGT/JunctionAnalysis is the IGHD3-3*01 in reading frame 1. The lengths of the 4 framework regions (FRs) as well as the 3 complementarity determining regions (CDRs) are indicated in brackets. The amino acid (AA) sequence of the IGHV-D-J junction is also provided. In this example the junction comprised 22 AAs. **(B)** Result summary table for an IGHV-IGHD-IGHJ gene rearrangement sequence that is unproductive due to an out-of-frame junction. An X for the CDR3-IMGT length indicates that for this sequence the length could not be determined. The '#' signs observed in the AA junction sequence indicate a frameshift. **(C)** Result summary table warning of low V-REGION identity (60.94%) and indicating that the 'insertions and deletions' option in the 'Advanced parameters' section should be used. **(D)** Result summary table for the case with germline low percent identity illustrated in (C) after repeating the sequence analysis using the 'Search for insertions and deletions' option. The correct identity percent is displayed in brackets and is calculated considering the insertion as a single mutational event. Please click here to view a larger version of this figure.

Figure 6. Subset assignment report table generated by the ARResT/AssignSubsets tool. The FASTA IG sequences from 10 CLL patients (A1-A10) were submitted to the ARResT/AssignSubsets tool. Case A4 was assigned to CLL stereotyped subset #2 (patients within this group are known to have a poor prognosis irrespective of SHM load) with high confidence. Seven of the other cases/sequences were not assigned to a major stereotyped subset and hence classed as unassigned. Two of the submitted sequences (A7 and A8) could not be used for subset assignment and were instead classed as skipped/unhealthy due to issues with the sequence, *i.e.,* due to an out-of-frame junction or the presence of stop codons. A comprehensive explanation of the table headings and the tool is available at the ARResT/AssignSubsets website[32]. Please click here to view a larger version of this figure.

Discussion

The study of IG genes in CLL has been vital not only for providing a better understanding into disease pathobiology but also for improving patient risk stratification. It is therefore paramount that the multi-step procedure of IG gene sequence analysis and subsequent interpretation of the sequence data is performed in a standardized manner. The availability of suitable and sufficient patient material and the timing of sampling should not impact on the ability to perform IG gene mutational analysis since the test can be performed on PB and on any sample with a high CLL tumor cell count. The separation of blood mononuclear cells using gradient separation methods is routinely performed in diagnostic laboratories, and as always, care should be taken when adding the blood/RPMI solution to the tube containing the gradient; this task should be performed in a slow, gentle manner to avoid disturbing the gradient and prevent the loss of cells.

Following cell separation, nucleic acids are extracted. Both gDNA and cDNA are suitable for SHM analysis of the clonotypic IGH rearrangement, however, there are advantages and disadvantages for the use of either material. The laboratory workflow proceeds quicker when using gDNA since no reverse transcription has to be performed prior to PCR amplification. That said, using gDNA as the substrate for PCR may result in the amplification of both productive and unproductive rearrangements which, in turn, can lead to additional hands-on laboratory work, *i.e.,* purification or gel excision of multiple PCR products and individual sequencing of all rearrangements. When comparing the gDNA and cDNA approaches, for the latter a reverse transcription step must be performed before proceeding with PCR amplification. However, in this instance, for most cases, only productive IG rearrangements will be amplified and subsequently sequenced. Thus, only a single PCR product will be purified and sequenced, while results interpretation is more straightforward.

The efficiency of amplification largely depends on the quality of the gDNA/cDNA, which should be assessed using a sensitive quantification method, and the primers utilized. The primers used are critical to the entire analytical procedure because the accurate determination of the SHM level can only be achieved by amplifying the entire sequence of the rearranged IGHV gene. A full-length rearrangement can only be assessed if 5' IGHV leader primers are used, thereby justifying the recent recommendations by ERIC for the use of leader primers[27]. The use of alternative primers, leading to the amplification of incomplete IGHV-IGHD-IGHJ gene rearrangements, is not appropriate for IGHV gene mutational analysis, and hence strongly advised against. The sole exception pertains to cases which repeatedly produce negative results when IGHV leader primers are used and analysis of new patient material is either not possible or also does not yield results. If use of a different patient sample and/or material (gDNA/cDNA) and various primers sets still fail to produce a PCR product, possible reasons could be that the tumor cell burden is too low or that lymphocytosis is not due to a CLL clone (in which case the immunophenotype should be re-evaluated). Primers used for the analysis should always be stated in the clinical report.

As CLL results from the accumulation of monoclonal B cells bearing the same IGHV-IGHD-IGHJ gene rearrangement, for the vast majority of cases clonality assessment will reveal a unique monoclonal peak, *i.e.,* a single clone. On rare occasions, double rearrangements or even oligoclonal patterns may be observed[35]. In such cases, gel electrophoresis is not the preferred technique for clonality assessment and more sensitive methods such as fragment analysis should be applied. Once clonality has been confirmed, the final step prior to sequencing relates to the purification of the PCR product to ensure that sequences of high quality are obtained. Finally, the IMGT/V-QUEST tool is the resource recommended for IG sequence analysis by ERIC. As the IMGT databases are constantly updated and report results in a consistant and well annotated manner, this tool ensures standardized analysis and facilitates comparative analysis amongst different clinical or academic facilities.

As immunogenetic analysis in CLL has prognostic and, in particular, predictive implications, robust analysis of the IGHV-IGHD-IGHJ gene rearrangement including the assignment to major CLL stereotyped subsets, is no longer only desirable but of key importance. This is particularly evident in this era of novel therapeutics and the potential that IG gene analysis has to guide treatment decisions[5,21,22,23,36,37,38], as officially indicated by the recent update of the NCI-sponsored guidelines from the International Workshop on CLL[24]. That said, rigorous standards are warranted, especially as determination of the SHM status of the clonotypic rearranged IGHV gene in CLL patients is not a trivial matter and involves a multi-step process. Importantly, certain experimental steps, most notably the choice of primers, yield superior results when specific options and/or approaches are adhered to, other technical aspects are amenable to some degree of flexibility, such as material selection or the method for assessing clonality, due to the fact that they lead to subtle differences, if any, with regard to the final results.

Over the last decade, ERIC has strived to promote the standardization and harmonization of various methods pertinent to CLL diagnostics and prognostication. This is reflected in the published recommendations and guidelines for immunogenetic analysis[27,34], numerous organized educational workshops and events, the establishment of an IG Network, as well as an expert online forum to discuss and provide guidance on IG gene sequence interpretation in CLL. The overall goal of ERIC through these actions is to promote optimal clinical management and patient care. Despite intense efforts, this task is far from complete, and in fact has become more complex owing to the development of novel high-throughput sequencing aimed at immune profiling. Nevertheless, although challenges persist, intensive efforts for the robust standardization of IG gene analysis in CLL continue and are currently the focus of ongoing activities within ERIC.

Disclosures

The authors have no relevant conflict of interest to disclose.

Acknowledgements

References

1. Swerdlow, S. H. *et al*. *WHO classification of tumors of haematopoietic and lymphoid tissues.* Vol. 2. IARC publications (2017).
2. Baliakas, P., Mattsson, M., Stamatopoulos, K., & Rosenquist, R. Prognostic indices in chronic lymphocytic leukaemia: where do we stand how do we proceed? *Journal of Internal Medicine.* **279** (4), 347-357, (2016).
3. Sutton, L. A., & Rosenquist, R. The complex interplay between cell-intrinsic and cell-extrinsic factors driving the evolution of chronic lymphocytic leukemia. *Seminars in Cancer Biology.* (2015).
4. Mina, A. *et al.* Using prognostic models in CLL to personalize approach to clinical care: Are we there yet? *Blood Reviews.* (2017).
5. Sutton, L. A. *et al.* Immunoglobulin genes in chronic lymphocytic leukemia: Key to understanding the disease and improving risk stratification. *Haematologica.* **102** (6), 968-971, (2017).
6. Damle, R. N. *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood.* **94** (6), 1840-1847, (1999).
7. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G., & Stevenson, F. K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood.* **94** (6), 1848-1854, (1999).
8. Burger, J. A., & Chiorazzi, N. B cell receptor signaling in chronic lymphocytic leukemia. *Trends in Immunology.* **34** (12), 592-601, (2013).
9. Packham, G. *et al.* The outcome of B-cell receptor signaling in chronic lymphocytic leukemia: proliferation or anergy. *Haematologica.* **99** (7), 1138-1148, (2014).
10. Messmer, B. T. *et al.* Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *The Journal of Experimental Medicine.* **200** (4), 519-525, (2004).
11. Stamatopoulos, K. *et al.* Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood.* **109** (1), 259-270, (2007).
12. 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** (19), 4467-4475, (2012).
13. Stamatopoulos, K., Agathangelidis, A., Rosenquist, R., & Ghia, P. Antigen receptor stereotypy in chronic lymphocytic leukemia. *Leukemia.* **31** (2), 282-291, (2017).
14. Baliakas, P. *et al.* Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: A retrospective multicentre study. *The Lancet Haematology.* **1** (2), 74-84, (2014).
15. Gounari, M. *et al.* Excessive antigen reactivity may underlie, the clinical aggressiveness of chronic lymphocytic leukemia stereotyped subset #8. *Blood.* **125** (23), 3580-3587, (2015).
16. Ntoufa, S. *et al.* B cell anergy modulated by TLR1/2 and the miR-17 approximately 92 cluster underlies the indolent clinical course of chronic lymphocytic leukemia stereotyped subset #4. *Journal of Immunology.* **196** (10), 4410-4417, (2016).
17. Mansouri, L. *et al.* Functional loss of IkappaBepsilon leads to NF-kappaB deregulation in aggressive chronic lymphocytic leukemia. *The Journal of Experimental Medicine.* **212** (6), 833-843, (2015).
18. Navrkalova, V. *et al.* ATM mutations in major stereotyped subsets of chronic lymphocytic leukemia: enrichment in subset #2 is associated with markedly short telomeres. *Haematologica.* **101** (9), e369-373, (2016).
19. Rossi, D. *et al.* Association between molecular lesions and specific B-cell receptor subsets in chronic lymphocytic leukemia. *Blood.* **121** (24), 4902-4905, (2013).
20. Sutton, L. A. *et al.* Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica.* **101** (8), 959-967, (2016).
21. Fischer, K. *et al.* Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: Updated results of the CLL8 trial. *Blood.* **127** (2), 208-215, (2016).

22. Rossi, D. *et al.* Molecular prediction of durable remission after first-line fludarabine-cyclophosphamide-rituximab in chronic lymphocytic leukemia. *Blood.* **126** (16), 1921-1924, (2015).

23. Thompson, P. A. *et al.* Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia. *Blood.* **127** (3), 303-309, (2016).

24. Hallek, M. *et al.* Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia. *Blood.* (2018).

25. van Dongen, J. J. *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** (12), 2257-2317, (2003).

26. Campbell, M. J., Zelenetz, A. D., Levy, S., & Levy, R. Use of family specific leader region primers for PCR amplification of the human heavy chain variable region gene repertoire. *Molecular Immunology.* **29** (2), 193-203, (1992).

27. Rosenquist, R. *et al.* Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia.* **31** (7), 1477-1481, (2017).

28. Linke, B. *et al.* Automated high resolution PCR fragment analysis for identification of clonally rearranged immunoglobulin heavy chain genes. *Leukemia.* **11** (7), 1055-1062, (1997).

29. Gonzalez, M. *et al.* Heteroduplex analysis of VDJ amplified segments from rearranged IgH genes for clonality assessments in B-cell non-Hodgkin's lymphoma. A comparison between different strategies. *Haematologica.* **84** (9), 779-784, (1999).

30. Brochet, X., Lefranc, M. P., & Giudicelli, V. IMGT/V-QUEST: The highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research.* **36** (Web Server issue), W503-508, (2008).

31. Yousfi Monod, M., Giudicelli, V., Chaume, D., & Lefranc, M. P. IMGT/JunctionAnalysis: The first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics.* **20 Suppl 1** i379-385, (2004).

32. Bystry, V. *et al.* ARResT/AssignSubsets: A novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy. *Bioinformatics.* **31** (23), 3844-3846, (2015).

33. Belessi, C. J. *et al.* IGHV gene insertions and deletions in chronic lymphocytic leukemia: "CLL-biased" deletions in a subset of cases with stereotyped receptors. *European Journal of Immunology.* **36** (7), 1963-1974, (2006).

34. Langerak, A. W. *et al.* Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. *Leukemia.* **25** (6), 979-984, (2011).

35. Heyman, B., Volkheimer, A. D., & Weinberg, J. B. Double IGHV DNA gene rearrangements in CLL: Influence of mixed-mutated and -unmutated rearrangements on outcomes in CLL. *Blood Cancer Journal.* **6** (7), e440, (2016).

36. Farooqui, M. Z. *et al.* Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: A phase 2, single-arm trial. *The Lancet Oncology.* **16** (2), 169-176, (2015).

37. Furman, R. R. *et al.* Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *The New England Journal of Medicine.* **370** (11), 997-1007, (2014).

38. Burger, J. A. *et al.* Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia. *The New England Journal of Medicine.* **373** (25), 2425-2437, (2015).