A novel next-generation sequencing capture-based strategy to report somatic hypermutation status using genomic regions downstream to immunoglobulin rearrangements

Neil McCafferty,[1,2] James Peter Stewart,[1] Nikos Darzentas,[3] Jana Gazdova,[1] Mark Catherwood,[4] Kostas Stamatopoulos,[5] Anton W. Langerak[6] and David Gonzalez[1]

[1]Patrick G Johnston Centre for Cancer Research, Queens University Belfast, Belfast, UK; [2]Barts Cancer Institute, Queen Marys University, London, UK; [3]Department of Hematology, University Hospital Schleswig-Holstein, Kiel, Germany; [4]Belfast City Hospital, Belfast, UK; [5]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece and [6]Laboratory of Medical Immunology, Department of Immunology, Erasmus University Medical Center, Rotterdam, The Netherlands

Correspondence: D. Gonzalez
D.GonzalezdeCastro@qub.ac.uk

Received: August 15, 2022.
Accepted: December 15, 2022.
Early view: December 29, 2022.

https://doi.org/10.3324/haematol.2022.281928

Abstract

The somatic hypermutation (SHM) status of the clonotypic, rearranged immunoglobulin heavy variable (IGHV) gene is an established prognostic and predictive marker in chronic lymphocytic leukemia (CLL). Analysis of SHM is generally performed by polymerase chain reaction (PCR)-amplification of clonal IGHV-IGHD-IGHJ gene rearrangements followed by sequencing to identify IGHV gene sequences and germline identity. Targeted-hybridization next-generation sequencing (NGS) can simultaneously assess clonality and other genetic aberrations. However, it has limitations for SHM analysis due to sequence similarity between different IGHV genes and mutations introduced by SHM, which can affect alignment efficiency and accuracy. We developed a novel SHM assessment strategy using a targeted-hybridization NGS approach (EuroClonality-NDC assay) and applied it to 331 samples of lymphoproliferative disorder (LPD). Our strategy focuses on analyzing the sequence downstream to the clonotypic, rearranged IGHJ gene up to the IGHM enhancer (IGHJ-E) which provides more accurate alignment. Overall, 84/95 (88.4%) CLL cases with conventional SHM data showed concordant SHM status, increasing to 91.6% when excluding borderline cases. Additionally, IGHJ-E mutation analysis in a wide range of pre- and post-germinal center LPD showed significant correlation with differentiation and lineage status, suggesting that IGHJ-E analysis is a promising surrogate marker enabling SHM to be reported using NGS-capture strategies and whole genome sequencing.

Introduction

Somatic hypermutation (SHM) is a key part of physiological B-cell development that underlies immunoglobulin (IG) affinity maturation and diversity in rearranged IG genes. Activation-induced cytidine deaminase (AID) enzyme introduces single nucleotide variants (SNV) in transcribed IG gene rearrangements.[1,2] AID-mediated SHM is restricted to a narrow window of B-cell development, predominantly in germinal centers (GC), controlling the frequency of genomic mutations.[3,4] Lymphoproliferative disorders (LPD) associated with various stages of GC development therefore present variable SHM levels; disorders originating from mature memory B and plasma cells present higher SHM levels.

Somatic hypermutation status in immunoglobulin heavy chain variable (IGHV) genes is a robust prognostic and predictive marker in chronic lymphocytic leukemia (CLL),[5] and can define two groups: mutated-CLL (M-CLL), with a long time to first treatment and overall survival,[6,7] and unmutated-CLL (U-CLL), a more aggressive disease, less sensitive to chemoimmunotherapy, and with reduced overall survival.[8-10] IGHV SHM status testing is recommended prior to treatment for all patients by the European Research Initiative on CLL (ERIC) and the International Workshop on CLL (iwCLL).[11-13]

Somatic hypermutation analysis is generally performed by polymerase chain reaction (PCR) amplification of clonal

IGHV-IGHD-IGHJ gene rearrangements followed by Sanger sequencing.[14] Rearranged IGH genes and germline identity of the clonal sequence are determined by comparison to IG germline databases, such as ImMunoGeneTics (IMGT).[15,16] SHM status in CLL employs a 2% mutational threshold (i.e., 98% identity to the closest germline gene allele) to account for potential polymorphic variation from unknown IGHV alleles, according to the first reports.[17-19] Almost two and a half decades later, IMGT have substantially expanded the reference dataset of the polymorphic alleles, challenging the accuracy of this empirical threshold.[20] In addition, borderline IGHV identity (97-97.9%, according to ERIC guidelines) presents variable outcomes, warranting caution in clinical decision making.[20-22] More recently, a wider borderline group of 97-98.9% IGHV identity was established, as cases with 98-98.9% SHM appear indistinguishable from 97-97.9%,[20,21] in contrast to earlier reports.[22] In this paper, borderline SHM status refers to the wider 97-98.9% identity.

Polymerase chain reaction amplification of clonal IGHV-IGHD-IGHJ gene rearrangements and Sanger sequencing is well-validated, highly standardized, and can be adapted for a small number of cases, but it has limited scalability, provides no insight on subclonal architecture nor assessment of other molecular risk factors.[23] Recent next-generation sequencing (NGS) applications for IGHV SHM analysis mainly rely on amplicon-based enrichment and present similar limitations, yet allow analysis of intraclonal variation.[24-26] Alternatively, NGS targeted-hybridization/capture applications can simultaneously assess IG/T-cell receptor gene rearrangements alongside other molecular risk factors, enabling a transition from multiple clinical investigations to a single assay.[27,28] However, so far, SHM status has not been widely reported using NGS-capture methods.

Current SHM assessment is restricted to the IGHV gene due to technical challenges imposed by the most common PCR methods, but also to the inherent difficulty in accurately distinguishing between SHM and random, non-templated nucleotides inserted at the ends of the recombining V, D and J genes. NGS-capture assessment of clonotypic IGHV genes is challenging due to insufficient and inaccurate read alignment of rearranged genes, particularly in the presence of a high level of SHM. Interestingly, high AID activity continues beyond rearranged IGHV genes (1.5-2kb downstream of IGHM enhancer) and could become a potential surrogate SHM marker (Figure 1A).[29,30] Despite multiple studies looking at these regions in mouse models, to the best of our knowledge no studies in CLL or other clinical entities in humans have considered SHM in alternative AID-targeted regions such as IGHJ genes or in the region between the VDJ junction and the IGHM enhancer (IGHJ-E). In this paper, we describe a novel method using targeted NGS-capture data generated with the EuroClonality-NDC assay to report SHM status using IGHJ-E in LPD. This method aims to provide a new alternative for SHM assessment using NGS data without reliance on clonal amplification of VDJ junctions, allowing integrated analysis of SHM alongside clinically relevant genomic alterations in a single assay.

Methods

Patient samples

A total of 331 LPD cases were studied for NGS IGHJ-E SHM status. All samples were collected according to local institutional review board approval and/or policies and in accordance with the Declaration of Helsinki. Patient samples consisted of genomic DNA (gDNA) from three cohorts. Cohort 1 was made up of 73 T-cell LPD (negative controls, presumed to lack IG rearrangements and SHM); cohort 2: 197 B-cell LPD comprising a wide range of clinical entities with different levels of maturation and including 34 CLL cases with SHM data; and cohort 3: whole peripheral blood from 61 CLL samples with available SHM data (*Online Supplementary Table S1*). Cohorts 1 and 2 are part of the EuroClonality-NDC validation study[28] and consisted of 184 high-molecular weight (HMW) and 86 formalin-fixed paraffin-embedded (FFPE) samples. Samples were processed as previously described[28] apart from cohort 3 samples, where 500 ng of gDNA were used for PCR-free library preparation.

NGS-capture assessment of clonality and sequence variation

The EuroClonality-NDC ARResT/Interrogate bioinformatics pipeline,[31,32] capable of detecting B- and T-cell receptor rearrangements from NGS capture data with >95% sensitivity and specificity, was used for the detection of IGHV-IGHD-IGHJ rearrangements. We used the published threshold of ≥6 unique rearranged clonal fragments to assign IGH clonality.[28] Variant calling for somatic sequence variants in IGHJ-E was performed using Pisces (v5.1.3.60) (Illumina, San Diego, CA, USA).[33,34] Aligned BAM files were visually assessed using the Integrated Genomic Viewer (IGV v2.5; Broad Institute, Cambridge, MA, USA)[35] and ≥6 unique rearranged reads to report IGHV-IGHD-IGHJ rearrangements.

IGHJ-E somatic hypermutation reporting strategy

Next-generation sequencing-capture SHM status was determined using an integrated analytical workflow that combines IGHV-IGHD-IGHJ gene rearrangements and SNV analysis (Figure 1). Rearranged IGHJ genes were identified by ARResT/Interrogate (*http://bat.infspire.org/arrest/interrogate/*), providing a sample-specific analysis target region covering 1500 bp from the rearranged IGHJ towards the en-

hancer region (IGHJ-E). The assessed IGHJ-E region started from the 3' end of the rearranged IGHJ gene to minimize alignment artefacts associated with reads falling on the VDJ junction. SNV detected in the 1500 bp IGHJ-E target region were assessed for somatic status (Figure 1B and C).
Distinction of SHM from non-somatic variants (i.e., polymorphic variants and sequencing/alignment artefacts) was achieved using progressive variant filtering and ROC/AUC analysis using the original clinical IGHV SHM status in CLL cases (see *Online Supplementary Methods*).

IGHV gene somatic hypermutation analysis by Sanger sequencing

Ninety-five CLL cases from cohorts 2 and 3 were assessed for IGHV gene SHM status using PCR and Sanger sequencing according to ERIC guidelines.[12] The IMGT database was used to assign IGH genes and IGHV identity.[16]

IGHJ-E somatic hypermutation analysis in a wide range of lymphoproliferative disorders

Lymphoproliferative disorder subtypes were categorized into three groups based on the expected SHM frequency. The 'High' category included 103 mature LPD where >90% cases are expected to have a significant level of SHM, namely Burkitt's lymphoma (BL), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), mucosa-associated lymphoid tissue (MALT) lymphoma, and plasma cell myeloma (PCM). An 'Intermediate' category (n=139) was defined for entities in which up to 30-60% of cases contain SHM, i.e., CLL, mantle cell lymphoma (MCL), and marginal zone lymphomas (MZL). The 'Low' category (n=89) was reserved for entities with no expected SHM, such as immature B-cell malignancies like B-cell precursor acute lymphoblastic leukemia (BCP-ALL) plus the control T-cell LPD cohort.

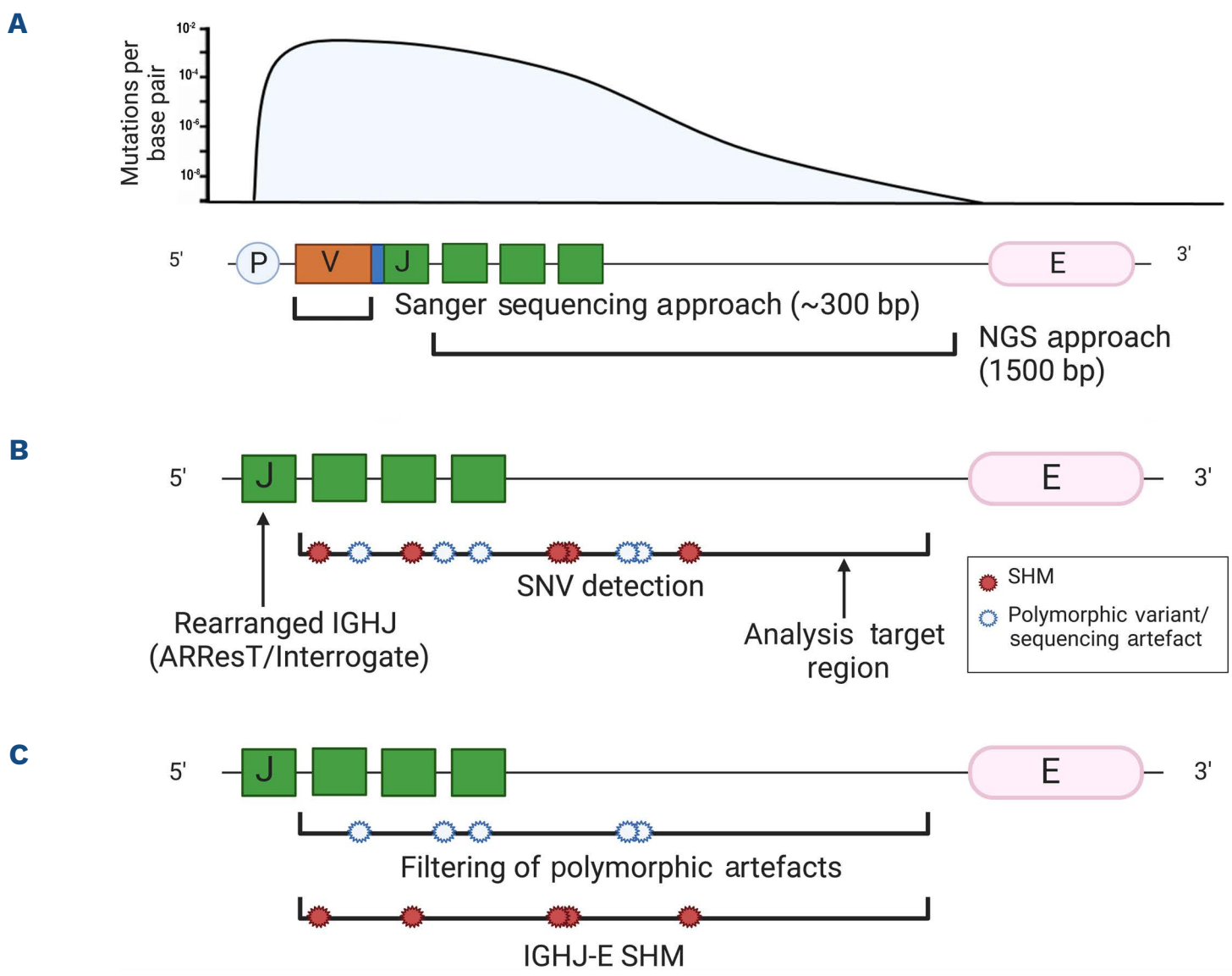

Figure 1. Somatic hypermutation frequency over the IGH genes and approaches to report somatic hypermutation status using Sanger sequencing and a novel next-generation sequencing (NGS)-capture approach. (A) Somatic hypermutation (SHM) occurs from the promoter region (P), and the frequency of mutations decreases towards the conserved enhancer region (E). The Sanger sequencing approach targets the clonal immunoglobulin heavy variable (IGHV) gene to compare to a database of known germline sequences and a mutational threshold of 98% is implemented to account for polymorphic variation and distinguish between mutated chronic lymphocytic leukemia (CLL) M-CLL (≤98%) and unmutated CLL (U-CLL) (>98%). SHM testing by NGS-capture targets the 1500 bp between the 3' end of the rearranged IGHJ and the IGHM enhancer, referred to as IGHJ-E identity. (B) The clonotypic rearranged IGHJ gene is identified by ARResT/Interrogate and all single nucleotide variants (SNV) detected are reported in the case-specific IGHJ-E. (C) SNV are inspected for non-somatic variants, separating meaningful SHM from any artefacts and polymorphisms. The frequency of true somatic mutations is calculated and IGHJ-E identity is reported as a percentage.

Results

Detection of rearranged IGHV genes by next-generation sequencing-capture in chronic lymphocytic leukemia

Sanger sequencing showed 27 different rearranged IGHV genes among the cohort of 95 CLL cases, in which *IGHV1-69, IGHV3-21*, and *IGHV4-34* were most common: n=13, n=9, and n=9, respectively (*Online Supplementary Table S2*). IGHV identity results showed 50 M-CLL (IGHV identity <98%) and 45 U-CLL (IGHV identity ≥98%); a total of 12 cases (7 M-CLL and 5 U-CLL) showed borderline IGHV identity (i.e., 97-98.9% identity).

Visual assessment of rearranged IGHV genes was concordant with clonotypic IGHV from the original Sanger sequencing IGHV data in 53/95 (55.8%) cases, all of which were also concordant by ARResT/Interrogate. Of note, M-CLL made up 32/42 (76.2%) discordant cases by visual assessment, and so discordant cases also showed significantly lower IGHV identities compared to concordant cases (*Online Supplementary Figure S1A*; Wilcoxon test, *P*<0.01). Ten U-CLL cases showed discordant rearranged IGHV genes by visual analysis, including four indetermined IGHV genes and four also undetected by ARResT/Interrogate. Reads in clonally rearranged IGHV genes showed poor alignment of mutated NGS reads, resulting in poor detection of SNV in highly mutated cases (*Online Supplementary Figure S2*). Taken together, this suggests IGHV analysis by capture-based assays and conventional alignment algorithms is not reliable without significant post-processing manipulation.

Detection of IGHV rearranged genes by ARResT/Interrogate analysis was concordant with clonotypic IGHV in 86/95 (90.5%) cases. Four of the cases with IGHV genes that are not specifically included in the NGS-capture assay were concordant with ARResT/Interrogate, suggesting those IGHV genes were captured due to sequence homology to other IGHV genes included in the assay. No significant difference in percentages of IGHV identity from the original data was shown between ARResT/Interrogate discordant and concordant cases (Wilcoxon test, *P*=0.9) (*Online Supplementary Figure S1B*).

Comparison of NGS-capture IGHJ-E and Sanger sequencing-based IGHV analysis in chronic lymphocytic leukemia

We evaluated the frequency of mutations in the 1500 bp IGH sequence from the 3' end of the rearranged IGHJ gene to derive the NGS IGHJ-E mutation status. Progressive non-somatic variant filtering was performed using the population GnomAD database (v3, minor allele frequency >0.02%), recurrent polymorphic artefacts from the test samples (TF>10%), and a minimum frequency filter for SNV (variant allele frequency [VAF] >5%) prior to visual IGV assessment of all SNV to confirm somatic mutations (*Online Supplementary Figures S3-S7*). Receiver operating characteristics analysis of the IGHJ-E region performed against clinical IGHV SHM status in CLL cases showed an area under the curve of 0.94 with an optimal IGHJ-E germline identity threshold of 99.8%, i.e., IGHJ-E identity >99.8% was used to classify samples as U-CLL and ≤99.8% as M-CLL (*Online Supplementary Figure S3*). IGHJ-E analysis showed significant correlation to clinical IGHV identity (Spearman's test, *P*<0.01) (Figure 2A). M-CLL cases showed significantly lower IGHJ-E identity compared to U-CLL (Spearman's test, *P*<0.01) (Figure 2B). Furthermore, borderline cases showed a distinct IGHJ-E identity that was significantly higher than M-CLL and lower than U-CLL (Wilcoxon test, *P*<0.01) (Figure 2B).

Employing a 99.8% IGHJ-E and 98% IGHV identity threshold, IGHJ-E SHM status was concordant with IGHV gene SHM status in 84/95 (88.4%) CLL cases (Figure 2C). A positive predictive value (PPV) of 86.5% to report M-CLL and a negative predictive value (NPV) of 90.7% to report U-CLL was found in 95 CLL cases. Borderline comprised 12/95 cases and showed the lowest concordance between IGHJ-E and IGHV gene SHM status, with only 8/12 (66.7%), compared to 39/42 (92.9%) M-CLL and 37/41 (90.2%) U-CLL (Figure 2D). Excluding clinically ambiguous borderline cases, overall concordance increased to 91.6% (76/83), with a PPV of 90.7% to report M-CLL and an NPV of 92.5% to report U-CLL.

Assessment of IGHJ-E somatic hypermutation status in lymphoproliferative disorders besides chronic lymphocytic leukemia

IGHJ-E SHM status was assessed in 331 LPD cases encompassing various lymphoid (B- and T-cell) malignancies categorized by expected SHM frequency into three groups. The High SHM category (103 cases from predominantly [post-]GC malignancies) showed significantly lower IGHJ-E identity than both Intermediate (*P*<0.01) and Low (*P*<0.01) categories (Figure 3). Intermediate cases (139 cases from malignancies where 30-60% carry SHM, such as CLL, MCL and MZL) also showed significantly lower IGHJ-E identity than the Low SHM category (89 cases from immature B-cell malignancies and T-cell LPD) (*P*<0.01) (Figure 3).

Using the 99.8% threshold, the High SHM category showed 80/103 (77.7%) mutated IGHJ-E sequences, while five showed at least one somatic mutation and 18 reported no mutations. The Intermediate SHM category showed 73/139 (52.5%) mutated IGHJ-E sequences. The Low SHM category showed 88/89 cases (98.9%) mutated IGHJ-E sequences; one angioimmunoblastic T-cell lymphoma (AITL) case was IGHJ-E mutated (mean VAF=0.12) with clonal IGHV-IGHD-IGHJ gene rearrangements (ARResT/Interrogate; *http://bat.infspire.org/arrest/inter-*

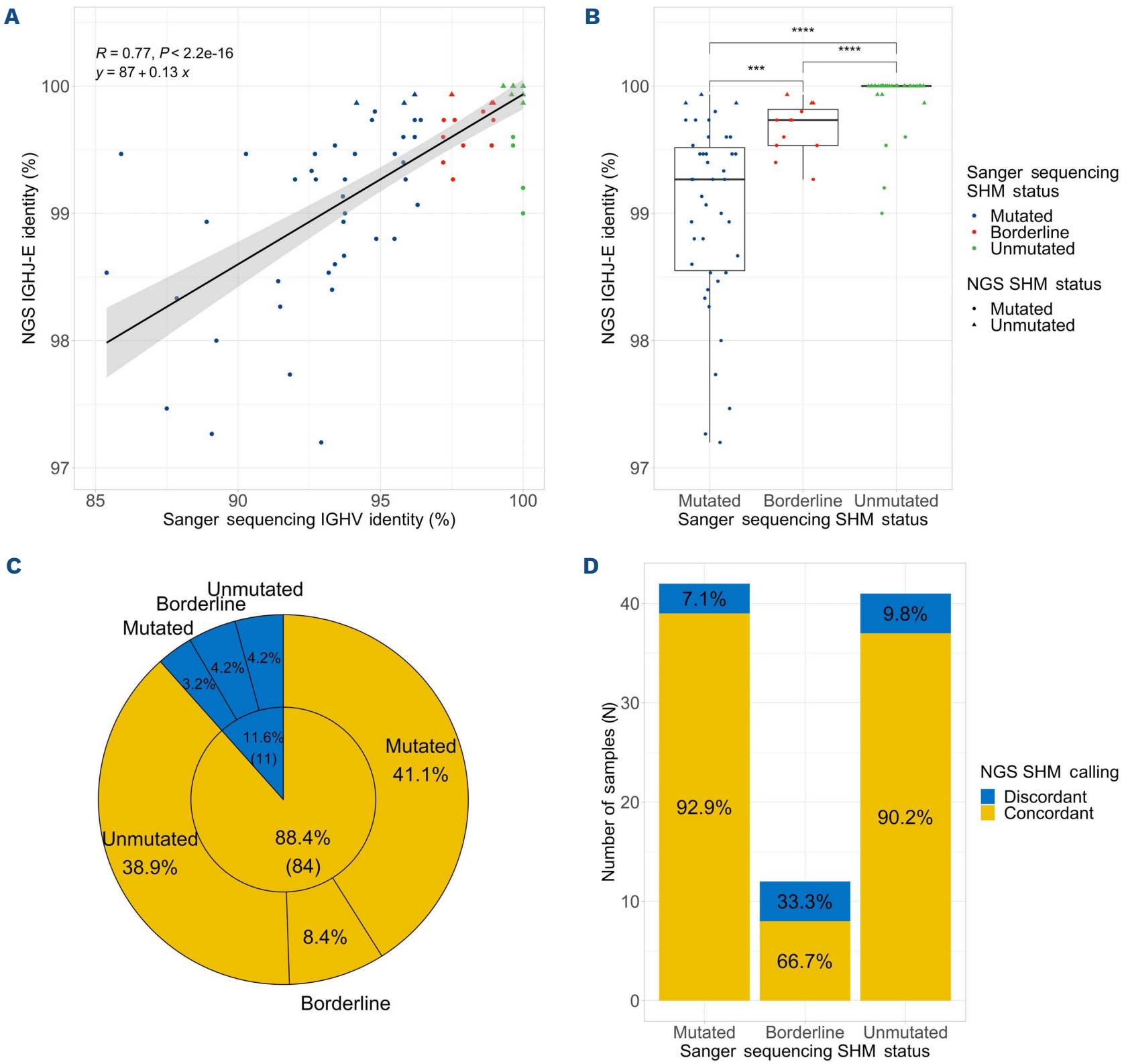

Figure 2. Comparison of Sanger sequencing immunoglobulin heavy variable gene to next-generation sequencing IGHJ gene enhancer with regards to identity and somatic hypermutation status in 95 chronic lymphocytic leukemia (CLL) samples from B-cell malignancies. (A) Spearman's correlation of Sanger sequencing immunoglobulin heavy variable (IGHV) gene compared to next-generation sequencing (NGS) rearranged IGHJ and the IGHM enhancer (IGHJ-E). (B) Comparison of IGHJ-E identity between cases from mutated (<97%), borderline (97-98.9%) and unmutated (>99%) subgroups reported by Sanger sequencing. Wilcoxon signed-rank test, ***P≤0.001, ****P≤0.0001. (C) Pie chart of the NGS IGHJ-E somatic hypermutation (SHM) status concordance to Sanger sequencing. (D) Bar chart of NGS SHM status concordance at mutated (<97%), borderline (97-98.9%) and unmutated (>99%) subgroups reported by Sanger sequencing.

rogate/), consistent with the presence of accompanying clonal B-cell population.

Discussion

This study aimed to develop a novel analytical approach for NGS-capture and whole genome sequencing (WGS) methods as a surrogate for canonical SHM status assessment, complicated by aligning shotgun library preparations to highly mutated rearranged IGHV genes. The NGS-capture IGHJ-E SHM assessment strategy applies four key differences to traditional PCR-based IGHV SHM testing: (i) IGHJ-E as a surrogate SHM marker for IGHV; (ii) a longer assessable IGH sequence (i.e., 1500 bp *vs.* 300 bp); (iii) exclusion of most polymorphic germline variants; and (iv) a stringent 99.8% mutational threshold. Altogether, 331 LPD cases were assessed for IGHJ-E SHM

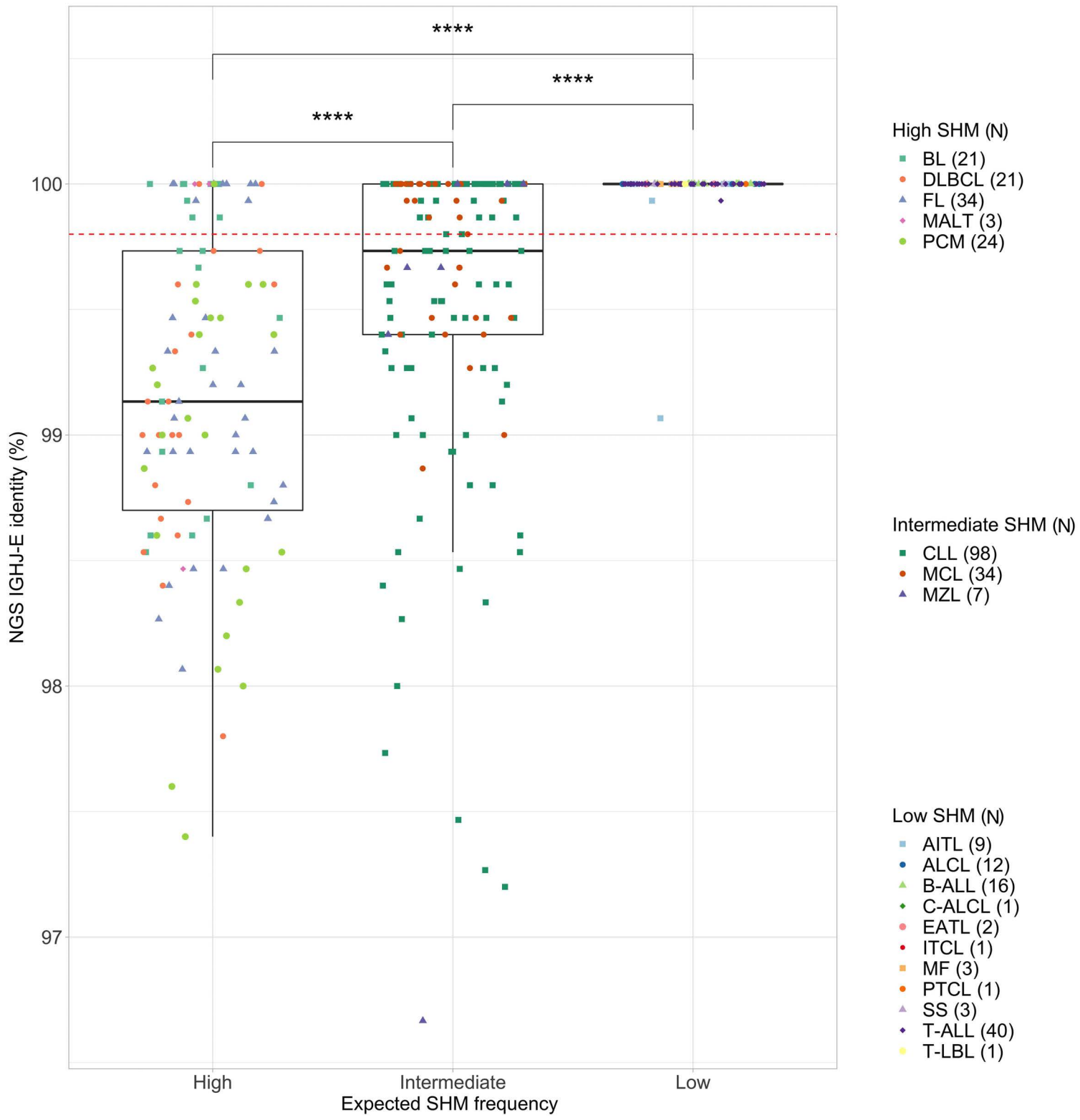

Figure 3. Comparison of next-generation sequencing rearranged IGHJ and enhancer gene identity (IGHJ-E) for 331 cases with different expected somatic hypermutation status from B- and T-cell malignancies. Expected somatic hypermutation (SHM) frequency were categorized as High, Intermediate, and Low. High (blue): germinal center (GC) or post-GC B-cell malignancies from Burkitt's lymphoma (BL), diffuse large B-cell lymphoma. (DLBCL), follicular lymphoma (FL), mucosa-associated lymphoid tissue (MALT) lymphoma, and multiple myeloma (MM). Intermediate (orange): heterogeneous-GC B-cell malignancies from chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL), and marginal zone lymphomas (MZL). Low (green): pre-GC B-cell from B-cell acute lymphocytic leukemia (B-ALL) and T-cell malignancies from angioimmunoblastic T-cell lymphoma (AITL), anaplastic large cell lymphoma (ALCL), cutaneous anaplastic large cell lymphoma (C-ALCL), enteropathy-associated T-cell lymphoma (EATL), intestinal T-cell lymphoma (ITCL), mycosis fungoides (MF), peripheral T-cell lymphoma (PTCL), Sézary syndrome (SS), T-cell acute lymphocytic leukemia (T-ALL), and T-cell lymphoblastic lymphoma (T-LBL). Red line: next-generation sequencing (NGS)-capture CLL 99.8% mutational threshold. Wilcoxon signed-rank test, ****$P \leq 0.0001$. NGS: next-generation sequencing; PCM: plasma cell myeloma; N: number.

status using this strategy. IGHJ-E NGS showed 88% overall concordance in SHM status to traditional assessment in CLL, which rose to 91.6% when clinically ambiguous borderline cases were disregarded.

Next-generation sequencing-capture to report IGHV gene SHM status was initially considered. However, visual assessment of sequencing reads identified clonotypic rearranged IGHV genes in only 56% cases, mostly because of SHM inhibiting alignment to the reference genome, as shown by significantly higher IGHV identity in the concordant cases. Consequently, IGHV reads containing potential SHM are unavailable for SHM assessment in the absence of significant post-processing manipulation of unaligned reads. Conversely, ARResT/Interrogate analysis

(*http://bat.infspire.org/arrest/interrogate/*), which does not rely on alignment to the reference genome, detected the clonotypic IGHV gene in 89% of cases, suggesting that poor alignment rather than probe hybridization is responsible for the absence of IGHV rearrangements (particularly somatic hypermutated cases) from the aligned files.

Polymerase chain reaction amplification of IGHV-IGHD-IGHJ limits assessment to ~300-360 bp of IGHV genes,[29,30] while analysis of IGHJ-E provides a larger 1500 bp region with fewer homologous sequences, hence improving read alignment.[24,36] One key limitation of IGHJ-E analysis is the assumption that only one allele is contributing to the SHM category, and while this is certainly true in most cases, there is evidence that discordant SHM can be found in 2 alleles in up to 1.5% of CLL.[12,13]

Chronic lymphocytic leukemia SHM status testing employs a 98% identity threshold for IGHV genes, although functionally, a single SHM could significantly influence B-cell receptor affinity and clonal selection.[20] Continuous IGHV-based SHM values have been reported to significantly impact overall survival[37] and recent retrospective analysis of historical studies showed that up to 10% of cases can change between M-CLL, U-CLL, and borderline categories using the latest ERIC guidelines and IMGT databases.[21] While the 98% threshold validity can be questioned, it still clearly defines CLL subgroups with independent prognostic and treatment response.[20,37-39]

The NGS-capture IGHJ-E strategy used a 99.8% mutational threshold to stratify cases with high correlation to conventional IGHV SHM status. We believe that a stringent NGS mutational threshold is required due to reduced AID activity towards the IGH enhancer leading to greater germline homology compared to rearranged IGHV,[40] and filtering SNV accounting for most common germline variations. IGHJ-E assessment may provide more clarity for risk-stratification in borderline cases, given the lower concordance with IGHV gene germline identity.

Incidence and burden of SHM differs significantly by LPD subtype, based primarily on the postulated cell of origin and its stage of development in relation to GC.[41] B-cell malignancies thought to originate from GC cells display high SHM rates, whereas T-cell and progenitor B-cell malignancies (e.g., B-cell acute lymphocytic leukemia, B-ALL) are not expected to display SHM. CLL, MCL and MZL constitute an intermediate group where the stage of differentiation is heterogenous, displaying variable SHM status.[42] LPD with expected high SHM frequency showed at least one mutation in 81% of cases, with a significantly lower IGHJ-E identity compared to B-ALL and T-cell LPD, providing biological validation of the approach. A single AITL case from the Low SHM category showed mutated IGHJ-E, and since clonal IG gene rearrangements have been reported in up to one-third of AITL cases,[43-45] we hypothesize that the observed SHM in this case is due to infiltrating mature B-cell clones. Cases with an expected Intermediate SHM frequency formed a significantly distinct subgroup between the High and Low SHM categories, as anticipated.

The novel IGHJ-E strategy correlates with CLL SHM status by conventional IGHV sequencing, and detects biologically meaningful SHM in other LPD, offering a strong proof of principle. While conventional IGHV SHM analysis in CLL provides critical prognostic and predictive value, NGS-capture can improve risk stratification by reducing polymorphic interference and evaluating prognostic markers including mutations, translocations and copy number analysis.[28]

A critical advantage of this new strategy is its potential application in IGKJ and IGLJ genes, providing a more comprehensive view of the SHM status than current methods. This NGS strategy can also be applied to WGS data, something that has not been reported thus far.

The main limitation of our study is the small CLL sample size and lack of clinical outcome analysis. Nonetheless, the IGHJ-E method showed a level of discrepancies between SHM categories similar to recent studies using newer versions of IMGT.[21] Studies in larger CLL cohorts with long-term follow up are warranted to evaluate the clinical prognostic value, particularly in ambiguous borderline cases. Further investigation into IGHJ-E SHM may confirm its clinical significance and facilitate an integrated next-generation analysis of SHM assessment alongside other genomic risk factors. Additional research may also consider this method in the analysis of WGS data to seamlessly incorporate SHM status in CLL and other LPD.

Disclosures

No conflicts of interest to declare.

Contributions

NMC and JG performed research. NMC, PS and ND performed data analysis. MC, KS and AWL provided clinical samples and data. NMC and DG wrote the manuscript. DG designed and supervised the study. All authors reviewed the manuscript and approved the final version for publication.

Acknowledgments

We would like to acknowledge the EuroClonality-NGS working group for access to cohort data.

Data-sharing statement

The raw data are not available for sharing as no specific consent for this purpose was available.

References

1. Goossens T, Klein U, Kuppers R. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. Proc Natl Acad Sci U S A. 1998;95(5):2463-2468.
2. Liu M, Duke JL, Richter DJ, et al. Two levels of protection for the B cell genome during somatic hypermutation. Nature. 2008;451(7180):841-845.
3. Liu M, Schatz DG. Balancing AID and DNA repair during somatic hypermutation. Trends Immunol. 2009;30(4):173-181.
4. Di Noia JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem. 2007;76(1):1-22.
5. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. Lancet Oncol. 2016;17(6):779-790.
6. Fischer K, Hallek M. Optimizing frontline therapy of CLL based on clinical and biological factors. Hematology Am Soc Hematol Educ Program. 2017;2017(1):338-345.
7. Thompson PA, Tam CS, O'Brien SM, et al. Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia. Blood. 2016;127(3):303-309.
8. Stevenson FK, Krysov S, Davies AJ, Steele AJ, Packham G. B-cell receptor signaling in chronic lymphocytic leukemia. Blood. 2011;118(16):4313-4320.
9. Wiestner A. Emerging role of kinase-targeted strategies in chronic lymphocytic leukemia. Hematology Am Soc Hematol Educ Program. 2012;2012(1):88-96.
10. Hallek M. Chronic lymphocytic leukemia: 2017 update on diagnosis, risk stratification, and treatment. Am J Hematol. 2017;92(9):946-965.
11. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood. 2018;131(25):2745-2760.
12. Rosenquist R, Ghia P, Hadzidimitriou A, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. Leukemia. 2017;31(7):1477-1481.
13. Agathangelidis A, Chatzidimitriou A, Chatzikonstantinou T, 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(1):1961-1968.
14. Gupta SK, Viswanatha DS, Patel KP. Evaluation of somatic hypermutation status in chronic lymphocytic leukemia (CLL) in the era of next generation sequencing. Front Cell Dev Biol. 2020;8(1):357.
15. Crombie J, Davids MS. IGHV mutational status testing in chronic lymphocytic leukemia. Am J Hematol. 2017;92(12):1393-1397.
16. Lefranc MP. Immunoglobulins: 25 years of immunoinformatics and IMGT-ONTOLOGY. Biomolecules. 2014;4(4):1102-1139.
17. Oscier DG, Thompsett A, Zhu D, Stevenson FK. Differential rates of somatic hypermutation in VH genes among subsets of chronic lymphocytic leukemia defined by chromosomal abnormalities. Blood. 1997;89(11):4153-4160.
18. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood. 1999;94(6):1840-1847.
19. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94(6):1848-1854.
20. Langerak AW, Davi F, Stamatopoulos K. Immunoglobulin heavy variable somatic hyper mutation status in chronic lymphocytic leukaemia: on the threshold of a new era? Br J Haematol. 2020;189(5):809-810.
21. Raponi S, Ilari C, Della Starza I, et al. Redefining the prognostic likelihood of chronic lymphocytic leukaemia patients with borderline percentage of immunoglobulin variable heavy chain region mutations. Br J Haematol. 2020;189(5):853-859.
22. Davis Z, Forconi F, Parker A, et al. The outcome of chronic lymphocytic leukaemia patients with 97% IGHV gene identity to germline is distinct from cases with <97% identity and similar to those with 98% identity. Br J Haematol. 2016;173(1):127-136.
23. Davi F, Langerak AW, de Septenville AL, et al. Immunoglobulin gene analysis in chronic lymphocytic leukemia in the era of next generation sequencing. Leukemia. 2020;34(10):2545-2551.
24. McClure R, Mai M, McClure S. High-throughput sequencing using the Ion Torrent personal genome machine for clinical evaluation of somatic hypermutation status in chronic lymphocytic leukemia. J Mol Diagn. 2015;17(2):145-154.
25. Stamatopoulos B, Timbs A, Bruce D, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. Leukemia. 2017;31(4):837-845.
26. Rodriguez-Vicente AE, Bikos V, Hernandez-Sanchez M, Malcikova J, Hernandez-Rivas JM, Pospisilova S. Next-generation sequencing in chronic lymphocytic leukemia: recent findings and new horizons. Oncotarget. 2017;8(41):71234-71248.
27. Wren D, Walker BA, Brüggemann M, et al. Comprehensive translocation and clonality detection in lymphoproliferative disorders by next-generation sequencing. Haematologica. 2017;102(2):e57-e60.
28. Stewart JP, Gazdova J, Darzentas N, et al. Validation of the EuroClonality-NGS DNA capture panel as an integrated genomic tool for lymphoproliferative disorders. Blood Adv. 2021;5(16):3188-3198.
29. Chandra V, Bortnick A, Murre C. AID targeting: old mysteries and new challenges. Trends Immunol. 2015;36(9):527-535.
30. Lebecque SG, Gearhart PJ. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. J Exp Med. 1990;172(6):1717-1727.
31. Bystry V, Reigl T, Krejci A, et al. ARResT/Interrogate: an interactive immunoprofiler for IG/TR NGS data. Bioinformatics. 2017;33(3):435-437.
32. Knecht H, Reigl T, Kotrova M, et al. Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. Leukemia. 2019;33(9):2254-2265.
33. Dunn T, Berry G, Emig-Agius D, et al. Pisces: an accurate and versatile variant caller for somatic and germline next-generation sequencing data. Bioinformatics. 2019;35(9):1579-1581.
34. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32(8):1220-1222.
35. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-26.
36. Lefranc MP, Giudicelli V, Duroux P, et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. Nucleic Acids Res. 2015;43(1):D413-422.
37. Jain P, Nogueras González GM, Kanagal-Shamanna R, et al. The

absolute percent deviation of IGHV mutation rather than a 98% cut-off predicts survival of chronic lymphocytic leukaemia patients treated with fludarabine, cyclophosphamide and rituximab. Br J Haematol. 2018;180(1):33-40.
38. Baliakas P, Moysiadis T, Hadzidimitriou A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. Haematologica. 2019;104(2):360-369.
39. Morabito F, Shanafelt TD, Gentile M, et al. Immunoglobulin heavy chain variable region gene and prediction of time to first treatment in patients with chronic lymphocytic leukemia: mutational load or mutational status? Analysis of 1003 cases. Am J Hematol. 2018;93(9):E216-E219.
40. Maul RW, Gearhart PJ. Controlling somatic hypermutation in immunoglobulin variable and switch regions. Immunol Res. 2010;47(1-3):113-122.
41. Seifert M, Scholtysik R, Küppers R. Origin and pathogenesis of B cell lymphomas. Methods Mol Biol. 2013;971(1):1-25.
42. Seifert M, Sellmann L, Bloehdorn J, et al. Cellular origin and pathophysiology of chronic lymphocytic leukemia. J Exp Med. 2012;209(12):2183-2198.
43. Carbone A, Gloghini A, Cabras A, Elia G. The germinal centre-derived lymphomas seen through their cellular microenvironment. Br J Haematol. 2009;145(4):468-480.
44. de Leval L. Approach to nodal-based T-cell lymphomas. Pathology. 2020;52(1):78-99.
45. Bruggemann M, White H, Gaulard P, et al. Powerful strategy for polymerase chain reaction-based clonality assessment in T-cell malignancies: report of the BIOMED-2 Concerted Action BHM4 CT98-3936. Leukemia. 2007;21(2):215-221.