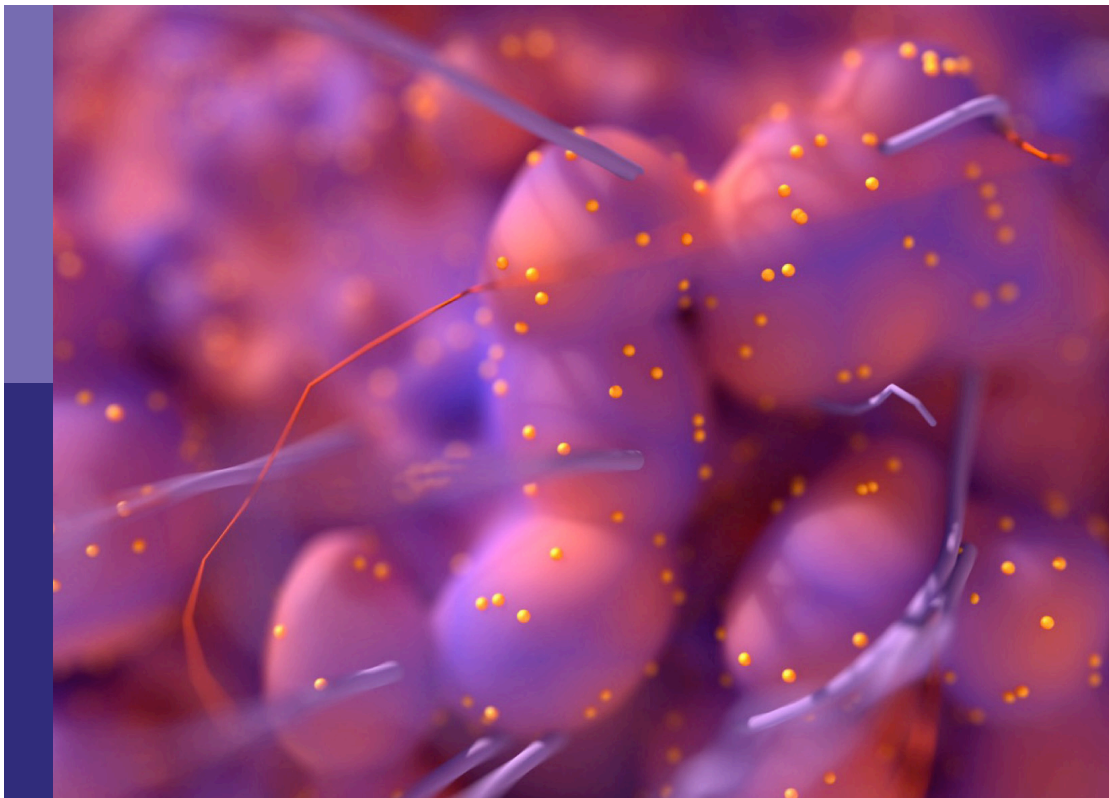
frontiers | Research Topics

The promise of immunogenetics for precision oncology

Edited by
Richard Rosenquist, Anastasia Chatzidimitriou and Anton W. Langerak

Published in
Frontiers in Oncology
Frontiers in Immunology



ISSN 1664-8714
ISBN 978-2-8325-3322-2
DOI 10.3389/978-2-8325-3322-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

The promise of immunogenetics for precision oncology

Topic editors

Richard Rosenquist — Karolinska Institutet (KI), Sweden
Anastasia Chatzidimitriou — Institute of Applied Biosciences (INAB), Greece
Anton W. Langerak — Laboratory of Medical Immunology, Department of Immunology, Erasmus Medical Center, Netherlands

Citation

Rosenquist, R., Chatzidimitriou, A., Langerak, A. W., eds. (2023). *The promise of immunogenetics for precision oncology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3322-2

Table of contents

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 09 August 2023
DOI 10.3389/fonc.2023.1252189

OPEN ACCESS

EDITED AND REVIEWED BY
Claudio Sette,
Catholic University of the Sacred Heart, Rome, Italy

*CORRESPONDENCE
Anastasia Chatzidimitriou
achatzidimitriou@certh.gr

RECEIVED 03 July 2023
ACCEPTED 24 July 2023
PUBLISHED 09 August 2023

CITATION
Vlachonikola E, Langerak AW, Rosenquist R and Chatzidimitriou A (2023) Editorial: The promise of immunogenetics for precision oncology.
Front. Oncol. 13:1252189.
doi: 10.3389/fonc.2023.1252189

Editorial: The promise of immunogenetics for precision oncology

Elisavet Vlachonikola[1], Anton W. Langerak[2], Richard Rosenquist[3,4] and Anastasia Chatzidimitriou[1,3]*

[1]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, [2]Department of Immunology, Laboratory Medical Immunology, Erasmus MC, Rotterdam, Netherlands, [3]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, [4]Clinical Genetics, Karolinska University Hospital, Solna, Sweden

KEYWORDS

immunogenetics, precision oncology, immune receptors, tumor microenvironment, next generation (deep) sequencing (NGS)

Editorial on the Research Topic
The promise of immunogenetics for precision oncology

The concept of precision medicine (PM) was introduced as early as the times of Hippocrates and later emphasized by Sir William Osler, both fostering the perception that "It is much more important to know what sort of patient has a disease than what sort of disease a patient has". In the last 20 years, rapid advances in diagnostic technologies along with the development of novel targeted therapies have paved the way for the implementation of PM approaches in cancer treatment and prevention, based on the individual's tumor characteristics, such as the genomic profile and the composition of the microenvironment of the malignant cells (1).

In the context of hematological malignancies, the distinctive features of the patients' immune repertoires, namely the unique B-cell receptor (BcR) and/or T-cell receptor (TR) genetic rearrangements within their adaptive immune system, have evolved as highly promising tools for PM in patient stratification and disease monitoring. In the clinical setting, immunogenetics is the basis for clonality assessment which allows to chartacterize at the molecular level the architecture of a lymphoproliferation (2, 3). Additionally, immunogenetics has been proposed as a means of reliably assessing measurable residual disease (MRD) (4). Turning to translational research, immunogenetic analysis of immune repertoires has shed light on the complex biological heterogeneity of lymphoid malignancies, which is reflected in the observed clinical variability in everyday practice. Furthermore, accumulating evidence on intraclonal diversification of the immunoglobulin genes has further enhanced our understanding of the mechanisms driving disease ontogeny (5).

In this Research Topic of Frontiers in Oncology, a collection of ten papers provides different perspectives on the promise of immunogenetics in hemato-oncology, highlighting on the one hand the contribution of immunogenetics to the deeper and more precise understanding of the

natural history of lymphoid malignancies, and on the other hand its value for tracking individual immunopathogenetic trajectories in hematological malignancies (Figure 1).

In more detail, Michaeli et al. identify distinct mutational patterns of the clonotypic immunoglobulin heavy variable (IGHV) genes in longitudinal biopsies from patients with follicular lymphoma (FL) versus transformed aggressive FL (t-FL), compared also with healthy control germinal center (GC) samples. The authors highlight different signatures of clonal evolution and somatic hypermutation (SHM) in the FL tumor clones in sequential biopsies from the same patient, in which SHM leads to the accumulation of novel potential N-glycosylation sites. Along similar lines, two independent papers by Sofou et al. and Neuman et al. describe distinct selection processes that shape the malignant clone in patients with chronic lymphocytic leukemia (CLL) bearing unmutated or mutated IGHV genes, respectively. The findings discussed in both papers strongly attribute this feature to alterations in SHM mechanisms. Moreover, for the first time, intra-VH CDR3 variations due to SHM were documented within cases bearing 'truly unmutated' IGHV genes i.e. those lacking any SHM over the sequence of the rearranged, clonotypic IGHV gene. (Sofou et al.) As described by Gkoliou et al. distinct features in the IGHV gene repertoires and SHM profiles have also been observed in cases with multiple myeloma expressing IgA or IgG isotype, alluding to unique immune trajectories to disease development.

Considering the technical aspects of immunogenetics analysis, several publications within this Research Topic present different methodologies for capturing the immune repertoire diversity, while also offering novel insights into the value of early detection of clonal events in different hematological malignancies and their diagnostic implications. Groenen et al. provide a detailed description of the next-generation sequencing (NGS)-based clonality assay applied across a wide spectrum of entities, spanning from immunodeficiency and autoimmunity to lymphoma and solid tumors. The authors emphasize the contribution of NGS-based immunogenetics to the sensitive diagnosis and early distinction of malignant lymphoproliferations in complex clinical and histopathological contexts.

In a similar vein, Kolijn et al. document that immunogenetic profiling of pre-diagnostic samples could predict the transformation of primary Sjögren's syndrome (pSS) lesions to B-cell lymphoma. They report that clonotypes of extranodal marginal zone lymphoma (eMZL) are detectable prior to overt eMZL diagnosis in the repertoire of activated B cells from pSS patients. Darzentas et al. present novel tools for the identification and characterization of clonal evolution of the malignant cells in B cell precursor acute lymphoblastic leukemia (BCP-ALL) with direct application for initial marker identification and MRD monitoring. As an alternative methodology for MRD detection in hematological malignancies, overcoming the demanding validation experiments and laboratory accreditation requirements of the NGS-based clinical tests, Schilhabel et al. present a robust and reproducible monitoring assay based on patient-specific, allele-specific oligonucleotide PCR, targeting the patient- and tumor-restricted clonally rearranged IGHV genes.

In-depth immunogenetic analysis of bystander immune cells within a tumor also provides valuable information regarding cell communication within the tumor microenvironment, thereby contributing to the development of patient-specific immune-based therapeutic interventions. In this Research Topic, Vlachonikola et al. describe restrictions in the TR gene repertoire in patients with CLL, likely arising from distinct genomic aberrations. This evidence supports the design of strategies to guide immune recognition by CLL-specific T cells as a complementary or alternative treatment in CLL.

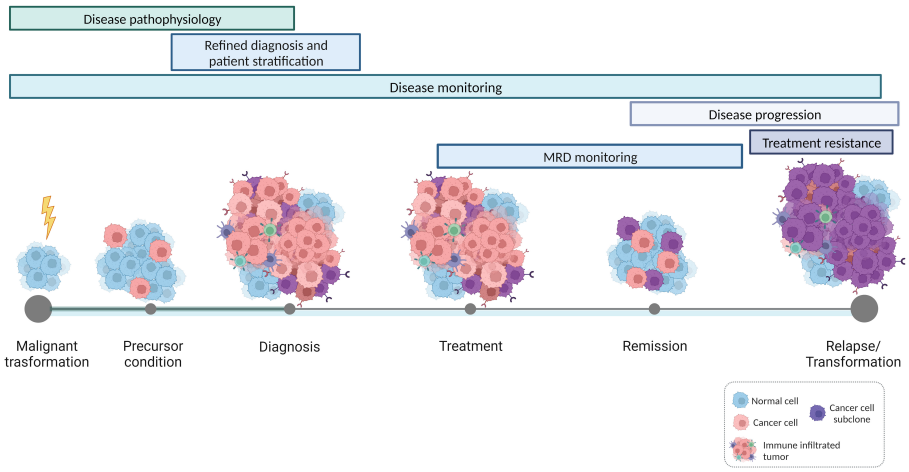


FIGURE 1
Applications of immunogenetics for precision hemato-oncology. High-throughput studies on the features and dynamics of antigen receptor gene repertoires contribute to a refined understanding of the natural history of lymphoproliferative disorders, shedding light on the mechanisms that drive malignant transformation and disease progression. In parallel, immunogenetics also offers valuable insights into the interactions occurring within the tumor microenvironment and the intricate communication between bystander immune cells and tumor cells. This, in turn, establishes the basis for patient-specific immunotherapeutic interventions.

While these manuscripts underline the huge potential of immunogenetics to transform patient management and cancer treatment, it is important to also acknowledge the limitations of the available techniques. Although NGS-based approaches offer great analytical depth, immunogenetic analyses typically rely on a single biopsy or a small sample, which may not fully capture the entire tumor's complexity. To overcome this pitfall, high-resolution mapping at the single-cell level has been proposed as the upcoming breakthrough strategy, enabling the identification of novel biomarkers for disease monitoring and response to treatment. In this Research Topic, a review article focusing on the employment of single-cell technologies in CLL by Oder et al. describes in detail the current status of single-cell transcriptomics, genomics, epigenomics, immunogenetics and profiling of cellular subpopulations, of both malignant and non-malignant cells within the tumor microenvironment. Single-cell sequencing technology has already shown its potential in dissecting the complex landscape of CLL, with similar applications in other immune-mediated malignancies, however, as highlighted by the authors, several challenges still exist that must be addressed through an integrated multi-omics approach.

In summary, the articles in this Research Topic illustrate the vital role of immunogenetics in precision hemato-oncology by providing insights into the tumor's intratumor heterogeneity and immunogenicity; the interaction between the tumor and immune system; and, potential targets for immunotherapies. By harnessing this knowledge, researchers and clinicians are entering a new era with more tailored and effective stratification and treatment schemes for patients with cancer.

Author contributions

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Acknowledgments

The authors would like to acknowledge Kostas Stamatopoulos for his critical revision of the manuscript.

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

References

1. Rosenquist R, Fröhling S, Stamatopoulos K. Precision medicine in cancer: a paradigm shift. *Semin Cancer Biol* (2022) 84:1–2. doi: 10.1016/J.SEMCANCER.2022.05.008

2. van Bladel DAG, van den Brand M, Rijntjes J, Pamidimarri Naga S, Haacke DLCM, Luijks JACW, et al. Clonality assessment and detection of clonal diversity in classic Hodgkin lymphoma by next-generation sequencing of immunoglobulin gene rearrangements. *Mod Pathol* (2022) 35:757–66. doi: 10.1038/s41379-021-00983-8

3. Wang HW, Raffeld M. Molecular assessment of clonality in lymphoid neoplasms. *Semin Hematol* (2019) 56:37–45. doi: 10.1053/J.SEMINHEMATOL.2018.05.008

4. Brüggemann M, Kotrová M, Knecht H, Bartram J, Boudjogrha M, Bystry V, et al. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* (2019) 33:2241–53. doi: 10.1038/S41375-019-0496-7

5. 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

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 10 November 2022
DOI 10.3389/fonc.2022.1029995

OPEN ACCESS

EDITED BY
Richard Rosenquist,
Karolinska Institutet (KI), Sweden

REVIEWED BY
Davide Bagnara,
Universitàdi Genova,
Italy
Huilai Zhang,
Tianjin Medical University, China

*CORRESPONDENCE
Ramit Mehr
ramit.mehr@biu.ac.il

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

RECEIVED 28 August 2022
ACCEPTED 24 October 2022
PUBLISHED 10 November 2022

CITATION
Michaeli M, Carlotti E, Hazanov H, Gribben JG and Mehr R (2022) Mutational patterns along different evolution paths of follicular lymphoma. *Front. Oncol.* 12:1029995. doi: 10.3389/fonc.2022.1029995

Mutational patterns along different evolution paths of follicular lymphoma

Miri Michaeli[1], Emanuela Carlotti[2], Helena Hazanov[1], John G. Gribben[2] and Ramit Mehr[1]*

[1]The Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, Ramat Gan, Israel, [2]Center for Haemato-Oncology, Barts Cancer Institute – a CR-UK Centre Of Excellence, Queen Mary University of London, London, United Kingdom

Follicular lymphoma (FL) is an indolent disease, characterized by a median life expectancy of 18-20 years and by intermittent periods of relapse and remission. FL frequently transforms into the more aggressive diffuse large B cell lymphoma (t-FL). In previous studies, the analysis of immunoglobulin heavy chain variable region (IgHV) genes in sequential biopsies from the same patient revealed two different patterns of tumor clonal evolution: direct evolution, through acquisition of additional IgHV mutations over time, or divergent evolution, in which lymphoma clones from serial biopsies independently develop from a less-mutated common progenitor cell (CPC). Our goal in this study was to characterize the somatic hypermutation (SHM) patterns of IgHV genes in sequential FL samples from the same patients, and address the question of whether the mutation mechanisms (SHM targeting, DNA repair or both), or selection forces acting on the tumor clones, were different in FL samples compared to healthy control samples, or in late relapsed/transformed FL samples compared to earlier ones. Our analysis revealed differences in the distribution of mutations from each of the nucleotides when tumor and non-tumor clones were compared, while FL and transformed FL (t-FL) tumor clones displayed similar mutation distributions. Lineage tree measurements suggested that either initial clone affinity or selection thresholds were lower in FL samples compared to controls, but similar between FL and t-FL samples. Finally, we observed that both FL and t-FL tumor clones tend to accumulate larger numbers of potential N-glycosylation sites due to the introduction of new SHM. Taken together, these results suggest that transformation into t-FL, in contrast to initial FL development, is not associated with any major changes in DNA targeting or repair, or the selection threshold of the tumor clone.

KEYWORDS

somatic hypermutation, high-throughput sequencing, follicular lymphoma, clonal evolution, B lymphocytes

Introduction

Follicular lymphoma (FL) is the second most common non-Hodgkin lymphoma. It is an indolent disease, clinically characterized by intermittent relapses and remissions (1) with about a third of cases transforming into a more aggressive lymphoma, most commonly diffuse large B cell lymphoma (t-FL) (2–4).

Previous analysis of immunoglobulin heavy chain variable region (IgHV) genes performed on sequential biopsies from the same patient revealed at least two different patterns of clonal evolution: direct evolution through acquisition of additional somatic mutations over time, and divergent evolution, in which later FL clones come from a less-mutated common progenitor cell (CPC) (5–9), that has escaped treatment and given rise to new diverse tumors.

High-throughput sequencing (HTS) has a great advantage over classical sequencing methods in the field of immunoglobulin (Ig) gene research, as it enables us to simultaneously analyze and compare many samples at a great depth (10–19). By analyzing the qualitative and quantitative pattern of SHM we were able to understand whether changes in the mutation pathways, including the creation of U:G mismatches by the enzyme activation-induced cytidine deaminase (AID), and their correction by error-prone DNA repair mechanisms, may be responsible for some transformation events [reviewed in (20)]. More recently, lineage tree-based mutation analysis has proven to provide better mutation counts than the sequence-based analysis, allowing to count only once the mutations shared between different progeny cells and enabling us to identify reversal mutations (21, Neumann et al., Front. Immunol., *in press*). In a previous lineage tree-based analysis, performed on 40 indolent and 39 aggressive lymphomas, we showed that lymphoma trees were more branched and had longer trunks – features of higher intraclonal diversification and a longer mutational history – compared to those from controls (21). However, tumor clones exhibited similar mutation frequencies (numbers of mutations per sequence) and identical SHM motifs to those observed in not-tumor B cells. These results suggested that the observed differences were probably a consequence of the longer diversification times of lymphoma clones rather than changes in their mutation rates (numbers of mutations per sequence per cell division). FL, which is considered a less aggressive lymphoma, displayed higher intraclonal diversity than Diffuse Large B Cell Lymphoma (DLBCL) and higher numbers of recent diversification events, confirming that the most aggressive lymphoma diversifies the least as it usually has less time to diversify until it is discovered and treated.

Our goal in the present study was to characterize at a greater depth the SHM patterns of FL tumor clones from sequential relapsed/transformed samples. One of SHM outcomes is the creation of N-glycosylation sites (22). N-glycosylation sites are rare in germline (GL) sequences from healthy individuals (23), but FL clones have been proven to acquire N-glycosylation sites on the heavy or light chains of the immunoglobulin gene (24). Recently, N-glycosylation sites have been also described in some autoimmune diseases (25). Thus, N-glycosylation sites, whether in the germline or acquired, may have a critical role in the development and selection of malignant clones.

Materials and methods

Samples, RNA extraction, amplification and sequencing

Biopsies were obtained after written informed consent in accordance with the Declaration of Helsinki and approval from the North East London Research Committee. Three patients were included in the study: Patient no.1 (Pt1) had three samples, one t-FL and two FL; Patients no. 2 and 3 (Pt2 and Pt3) had 2 FL samples each. All tumor samples carried an IgH-VH3 rearranged major tumor clone; they were selected, and RNA extracted and amplified, as previously described (26). Briefly, 37 libraries (seven from the whole lymph node biopsies corresponding to the three patients, and the rest from flow-sorted B cell sub-populations) were prepared in the Gribben lab using JH consensus and VH3-FR1 primers (5) containing unique molecular identification (MID) tags for sample identification. Libraries were sequenced using the Roche 454 Life Sciences Genome Sequencer FLX following the manufacturer's instructions for the Titanium series (454 Life Science, Roche).

In addition, 32 FL samples (FL-S) containing 772 sequences (21) and nine healthy germinal center (GC) samples from spleen and Peyer's patches containing 129 sequences, previously analyzed by the Mehr lab (27, 28), were also included in the analysis and used as controls. Comparisons were performed between tumor and non-tumor clones from the three FL patients of this study (FL-HTS), and between them and the previously studied FL-S and healthy GC clones. An intra-patient analysis was also performed by comparing tumor clones from sequential samples from the same patient, collected from different anatomical sites (Pt3) or at different stages of the malignancy or treatment (Pt1 and Pt2, FL/t-FL, Table S1).

Data pre-processing

Reads of the tumor clone in each sample were first processed as described (26). Briefly, data were identified by BLAST against the sequences obtained by homo/heteroduplex analysis, separated by their sample molecular identification (MID) tags, and filtered to remove reads of length ≤60 nucleotides (excluding MID tags and primers) or reads captured only once. Reads captured only twice were examined manually and included in the study only if the pyrograms showed high quality throughout. Remaining paired-end reads were assembled, annotated by

SoDA (29) and aligned by ClustalW2 (30, 31) before and again after removal of insertions and/or deletions suspected to be artifacts.

In order to discard artifactual insertions and/or deletions (henceforth called indels, typically introduced during the 454 Roche sequencing) we used our program Ig-Indel-Identifier (Ig Insertion – Deletion Identifier) (32). We assumed that the two other datasets (32 FL samples and 9 GC samples), generated by using the Sanger method, did not contain any such artifacts and were therefore excluded from this initial analysis. In order to include as large a number of reads as possible in the analysis, we ran the Ig-Indel-Identifier program with the following, permissive parameter values: the minimum homopolymer tract (HPT, a stretch of identical nucleotides) length that must be checked was set to 2 nucleotides, the minimum number of sequences from the same B cell clone that must share the same indel or a low quality score point mutation for this indel or mutation to be considered legitimate was set to 1, and no exclusion of low-quality point mutations. Only unique sequences, which differ from all other sequences by one or more mutations, were kept for further analyses. A total of 2,381 unique sequences without suspected artifact indels were used for further analyses (Table 1).

Germline VDJ segment identification and assignment into clones

Clonally related sequences were defined as reads having identical V,D, and J segments according to SoDA (29); if there were two or more clones with the same V, D and J segments in a sample (as shown in Table S2), only sequence groups with highly homologous sequences of complementarity determining region 3 (CDR3) were considered as clones, as confirmed by visual examination of the alignments after sequences with the same V, D and J segments were aligned using ClustalW2 (30, 31). Since libraries were prepared using VH3-FR1 primers, VH3 genes from non-tumor B cells (NT) were also amplified and sequenced. Tumor-related reads (T) were defined as the reads identical to the dominant sequence plus those sequences that, based on the SHM pattern of IgH-VH, were clonally related to the dominant tumor sequence. Clones with the same V(D)J segments but with completely different CDR3 sequences (no shared nucleotides) and no shared mutations elsewhere were considered as non-tumor clones. Table 1 shows the numbers of clones, unique sequences and mutations in the tumor and non-tumor clones in each sample, in each patient.

Mutational analyses

Ig lineage tree analyses

Clonally-related Ig gene sequences were used to generate lineage trees (Figure S1) using our program IgTree© (33), as previously described (21, 34). The lineage trees were then measured using our program MTree© (35, 36). In a previous study, a thorough statistical analysis performed on simulated data has established the quantitative relationships between lineage tree characteristics and the parameters characterizing

TABLE 1 Numbers of clones, unique sequences and mutations in tumor and non-tumor clones in each sample, in each patient.

Patient	Sample*	Clones**	No. of clones	No. of unique sequences***	No. of mutations****			
					Total	Min	Max	Med
Pt1	1	T	1	166	263	NA	NA	NA
		NT	21	51	548	3	44	34
	2	T	1	141	229	NA	NA	NA
		NT	11	22	373	11	42	37
	3	T	1	235	368	NA	NA	NA
		NT	34	99	731	3	55	19.5
Pt2	1	T	2	108	257	33	224	128.5
		NT	9	13	279	27	39	30
	2	T	1	407	622	NA	NA	NA
		NT	7	23	206	4	44	30
Pt3	1	T	2	345	438	209	229	219
		NT	13	42	443	5	49	33
	2	T	2	721	813	406	407	406.5
		NT	4	8	175	33	51	45.5

*Samples are numbered chronologically. **T, tumor clones; NT, non-tumor clones; clones are defined as described in the methods. ***Unique sequences are sequences that differ from all other sequences by one or more mutations. ****The numbers of mutations were calculated from the lineage trees. This way, we counted each mutation only once if it happened earlier in the clone. We present the total, minimum (Min), maximum (Max) and median (Med) numbers of mutations. The V(D)J gene segment combinations detected in all samples analyzed on the 454 Roche sequencer are given in Supplementary Table S2.
NA, Not Applicable.

affinity maturation dynamics (proliferation, differentiation and mutation rates, initial affinity of the Ig to the antigen, and selection thresholds); seven specific characteristics (the minimum root to leaf path length, the average distance from a leaf to the first split node/fork, the average outgoing degree, that is the average number of branches coming out of any node, the root's outgoing degree, the minimum distance between adjacent split nodes/forks, the length of the tree's trunk and the minimum distance from the root to any split node/fork) were the most informative (37). The comparison between lineage tree characteristics from different patients or between different datasets was done using the non-parametric Mann-Whitney U-test, as these characteristics are not always normally distributed. To correct for multiple comparisons, we used the false detection rate (FDR) correction method (38).

Mutation distributions

The analysis of mutation distributions (or mutation spectra, together with targeting motif analysis described in the next section) enables us to characterize the SHM mechanisms operating in the B cell clones. The numbers of mutations from A, C, G, and T were counted for each sample and expressed as percentages of the total number of mutations detected in each sample. When different samples were compared, the *expected* numbers of mutations from A, C, G, and T in each sample were calculated as the observed number of mutations from either A, C, G, or T in that sample, multiplied by the total number of mutations in that sample and divided by the total number of mutations of the two samples. A χ^2 analysis was then performed on all mutation numbers, comparing between the sets of observed and expected mutation numbers. In addition, the ratios of the percentages of transition and transversion mutation (from the total number of mutations for each group) were examined in each sample. A χ^2 test was performed to compare the tumor and non-tumor transition and transversion percentage ratios in all patients and samples.

SHM targeting motif analysis

It is established that SHMs occur at higher frequency in specific sequence motifs (39, 40). Identification of SHM targeting motifs around mutated positions was performed as described (41) to further examine the mechanism of SHM. This analysis was based on a previous published work by Spencer and Dunn-Walters (42). Briefly, the base composition at positions flanking a mutation (three nucleotides on either side) was determined and then, for each nucleotide, a χ^2 test was performed to check whether the frequency of each type of mutation was statistically significant compared with the background frequency observed in the germline (GL) sequence. The F-test was used to compare the base compositions surrounding the mutations from different datasets.

N-glycosylation analysis

We first analyzed the potential glycosylation sites in the GL sequences, and then the acquired glycosylation sites (AGS) introduced by SHM in the mutated sequences. The N-glycosylation motif included in the analysis was Asn–X–Ser/Thr, where X is any amino acid except Pro, Asp or Glu. To analyze the potential glycosylation sites in the GLs, we counted the number of occurrences of the full motif, and – separately – the occurrences of motifs which differ by one nucleotide from the full motif. We compared the two numbers between tumor vs. non-tumor clones, in order to establish whether N-glycosylation may have affected clonal dynamics. The χ^2 test was used to compare the groups; in clones with less than five GLs and more than one glycosylation site, for which the χ^2 test did not apply, a likelihood ratio was used instead. In the analysis of AGS, we compared the number of clones with AGS and the number of AGS per clone between tumor and non-tumor clones in each sample and patient.

Results

SHM mechanisms in FL clones may be different from those in normal clones

Our hypothesis is that changes in mutational mechanisms, including AID targeting and the subsequent error-prone correction by DNA repair mechanisms, may be responsible for some transformation events. Several lines of evidence in this study show that in some cases a change in SHM mechanisms may have occurred in the tumor clone between biopsies. First, we observed that the mutation spectra of tumor clones were different from that of non-tumor clones at both the patient and the sample levels; the only exception was sample number 2 of Pt2, displaying similar mutation spectra in both tumor and non-tumor clones. The mutation spectra of tumor clones from all patients were similar (Figure 1A), while the non-tumor clones from all patients presented unusual and highly variable mutation spectra, mostly in the non-tumor clones of sample 2 from Pt3 and samples 1, 2 and 3 from Pt1; this variability may be due, in part, to the relatively low numbers of unique sequences detected in these NT clones, combined with the intrinsic randomness of SHM. Tumor clones from all three Pt1 samples were similar in their mutation spectra (Figure 1B), suggesting that the mutation mechanisms did not change over the elapsed time. On the other hand, mutations from G to any other nucleotide were found to be more frequent than mutations from C to any other nucleotide in tumor clones of the second (later) biopsies of both patients 2 and 3 than in the earlier samples, implying that some mutational mechanisms (possibly the targeting) may have changed between the two consecutive biopsies of each of these patients.

Second, the transition-transversion mutation ratios greatly varied between patients and samples. Transition mutations are point mutations that replaces a purine nucleotide by another

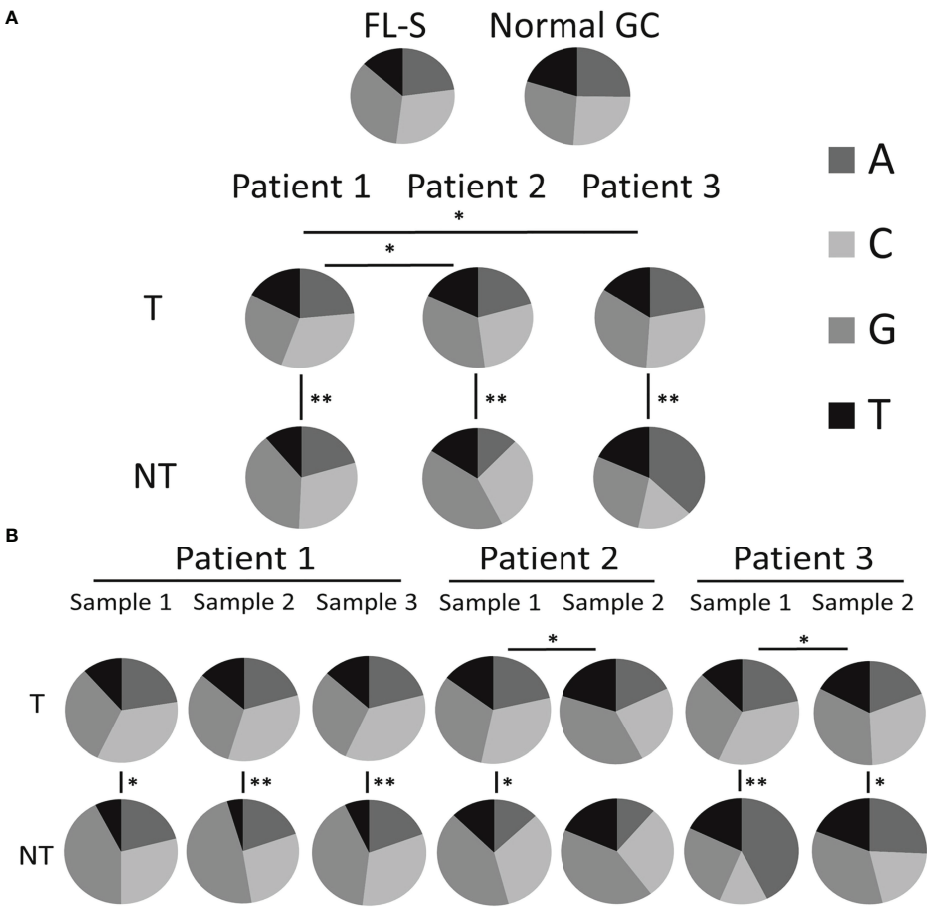


FIGURE 1
Mutation spectra (distribution among nucleotides). The number of mutations from each nucleotide, presented as a fraction out of the total number of mutations, for **(A)** tumor and non-tumor clones in each patient analyzed (also including the reference FL samples from a previous study and control GC samples) and **(B)** for tumor and non-tumor clones in each sample analyzed. Significant differences are indicated with lines between pie-charts (*p-value < 0.05; **p-value < 0.0005). Data regarding the numbers of mutations from each nucleotide and p-values of the comparisons between the different groups are shown in Tables S3, S4 and Tables S5, S6, respectively.

purine or a pyrimidine by another pyrimidine, while a transversion is a replacement of a purine by a pyrimidine or vice-versa. A transition-transversion ratio larger than 1 means that there were more transition mutations than transversion, and vice versa for a ratio smaller than 1; in normal B cell clones, this ratio is larger than 1, as in the healthy control GCs used in this study and various others [e.g. (43)]. Pt2 had transition-transversion ratios larger than 1 in all clones in both samples except the tumor clone in sample 1, while Pt3 presented transition-transversion ratios smaller than 1 (Figure 2) in all clones. In non-tumor clones from Pt1 the ratios were ~1:1, while they were <1 in the tumor clones. Note that, when samples were combined together to look at a more complete picture for each clone, the lineage tree structure of each clone – and hence the characteristics of some mutations – may have changed. Significant differences between tumor and non-tumor clones were found only in the second FL sample of Pt1 and in the first FL sample of Pt2. These results raise the question of whether DNA repair mechanisms are altered with transformation in these FL cases. According to a previously published paper, the polymerase Rev1 may promote transversions at C:G pairs, while the low-fidelity polymerase θ can introduce both transitions and transversions at abasic sites (44). Alternatively, BCR-based selection may be impaired, if it operates at all, in FL clones, as in other lymphomas (21, 43).

Third, all FL-HTS samples presented different mutation targeting motifs for mutations from G relative to the reported motif (Figure 3). The healthy GCs samples, used as controls, presented the reported motif but not the new motifs, supporting the hypothesis that a possible change in the SHM mechanism occurred in at least one FL case. When we examined how many positions in the motif for each mutation contained significant differences between the tumor and non-tumor clones in each

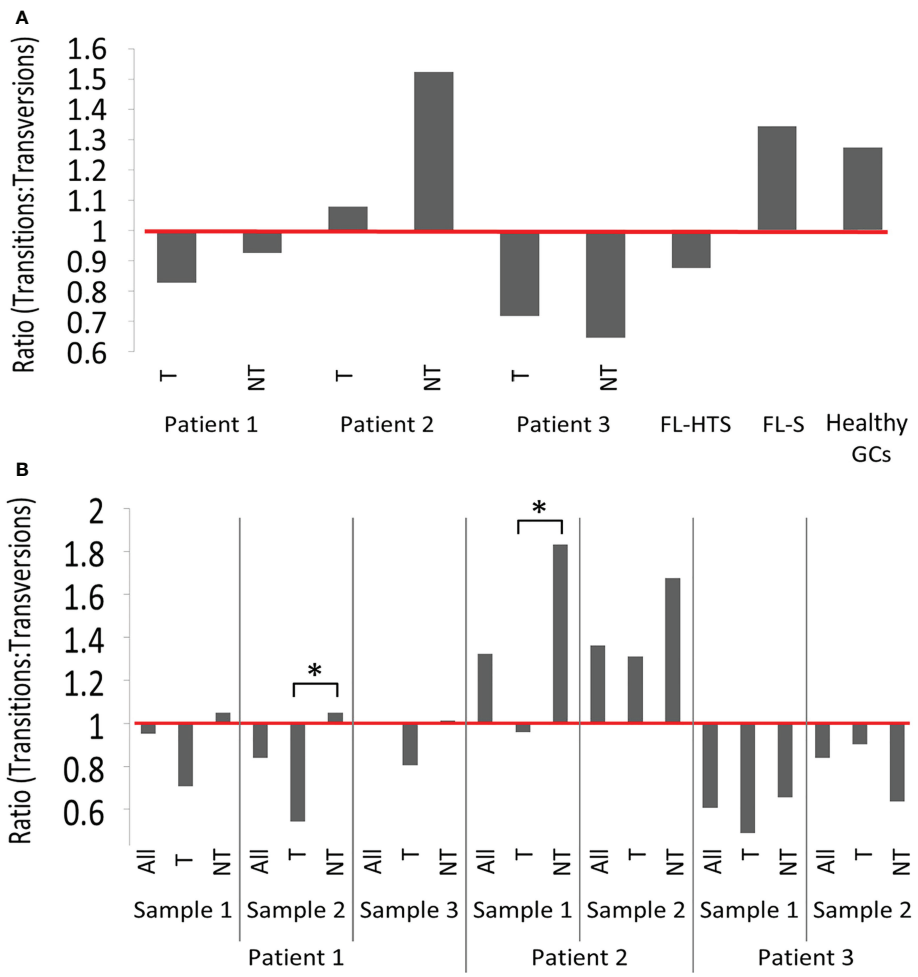


FIGURE 2
Transition-transversion mutation ratios. **(A)** In tumor and non-tumor clones per patient analyzed and **(B)** In tumor and non-tumor clones per sample. Additional information regarding the numbers of transition and transversion mutations in tumor and non-tumor clones of FL patients, and in healthy GCs are provided in the Tables S7 and S8. According to the χ^2 test, no significant differences were observed between tumor and non-tumor clones across different patients. Significant differences were found when tumor and non-tumor clones from sample 2 from patient 1 and sample 1 from patient 2 (*p-value < 0.005) were compared. (See also Tables S7 and S8).

patient, we observed that T and NT clones from Pt1 were the most similar in terms of motif usage. In addition, consecutive tumor samples from the same patient did not have statistically significantly different mutation targeting motifs (Tables S9-S12). Taken together, the differences presented above suggest that SHM mechanisms (targeting, DNA repair or both) may be different in FL compared to normal B cell clones.

Tumor clones acquire more new potential N-glycosylation sites than non-tumor clones

It is known that only a minority of GL V segments and normally-developed memory B cell V region genes contain potential N-glycosylation sites (PGS) (23). In contrast, human B-cell malignancies, and FL in particular, are characterized by an extremely variable incidence of acquired N-glycosylation sites (AGS) in their Ig variable region sequences (23, 24, 45–47). Hence, we examined the potential and acquired N-glycosylation motifs in GL sequences and in tumor and non-tumor clones in order to determine whether potential N-glycosylation sites are more frequent in FL and t-FL than in healthy B cell clones.

As shown in Figure 4A, Pt1 had no GL sequences with existing N-glycosylation motifs, while Pt2 and Pt3 had no more than one GL sequence each that contained such motifs. However, in Pt1 and Pt3 there were, on average, 12 motifs that were only one mutation away from becoming a potential AGS. In all patients, these average numbers were similar both in tumor and non-tumor clones. After examination of the tumor

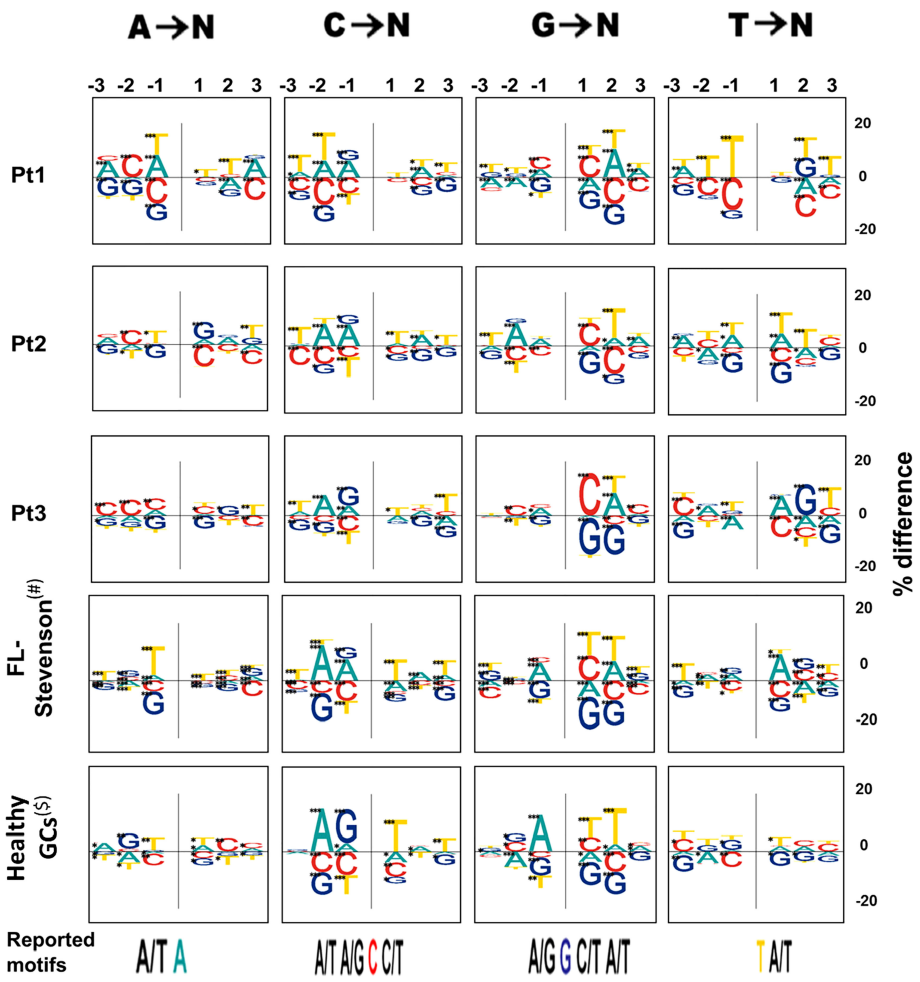


FIGURE 3
Mutation targeting motifs. On the top we show the 3 nucleotides examined upstream and downstream, for each mutated nucleotide, denoted as -1, -2, and -3, for the 3 positions flanking the mutation upstream and 1, 2, and 3 for those flanking the mutation downstream. The positive and negative sides of the Y axis denote excess or paucity of the indicated nucleotide in that position, respectively. The size of each letter is given by the "% difference", calculated as percentage of each base at each position flanking a particular mutation, minus the percentage composition of the GL sequence at that position. Asterisks represent levels of significance (*p-value < 0.05; **p-value < 0.005; ***p-value < 0.0005). Previously reported motifs are shown at the bottom of the figure; the mutated nucleotide is colored and the flanking nucleotides are shown for reported positions.

and non-tumor clones from each sample (Figure 4B), we observed similar average numbers of PGS in clonal GL sequences, with motifs that were only one mutation away from a potential AGS in both tumor and non-tumor samples of Pt1. Pt2 presented the highest average numbers in both tumor and non-tumor GLs samples (p-value < 0.005 for both patients 1 and 3 compared to Pt2, Table S13).

The percentages of clones that acquired at least one new potential glycosylation site were calculated and found significantly higher in Pt2 as compared to other patients, in both tumor and non-tumor clones. Moreover, all tumor clones and 92% of non-tumor clones of Pt2 acquired at least one AGS. This is not surprising, as Pt2 had the highest percentages of clonal GL sequences with motifs that were only one mutation away from AGS. In patients 1 and 3 there were more tumor clones than non-tumor clones that acquired AGS (Figure 4C). Although the average numbers of motifs that are only one mutation away from becoming a potential AGS were similar between tumor and non-tumor clones in all patients (Figure 4A), there were more AGS per clone in tumor clones than in non-tumor clones in each patient (Figure 4D). Interestingly, the highest numbers of AGS per clone were found in the latter samples in each patient (Figure 4E). AGS in all tumor clones were present along the sequence, ranging from CDR1 to CDR3. In both Pt2 and Pt3, the highest numbers of AGS in the later tumors correlated with the highest number of mutations in the

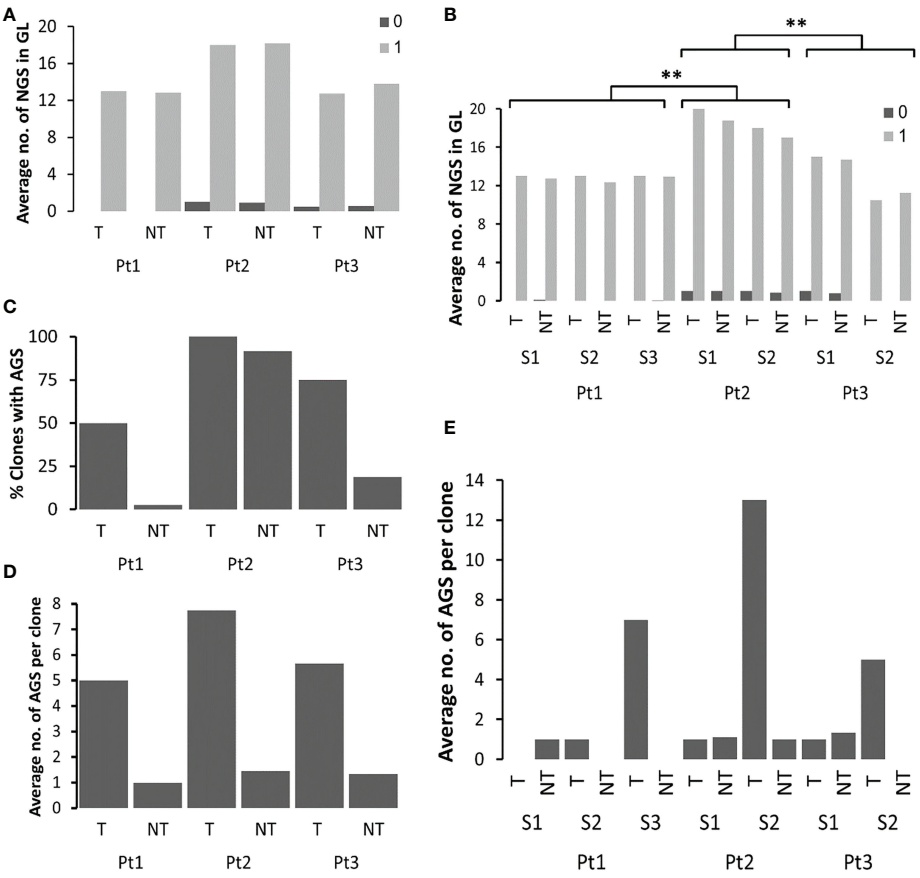


FIGURE 4
Analysis of PGS and AGS in FL samples. **(A)** Average numbers of existing potential PGS (0) and number of motifs that are only one mutation away from becoming an AGS (1) in clonal GL sequences, in tumor and non-tumor clones in each patient, and **(B)** in each sample in each patient. In case there was only one clonal GL sequence, the average number is the actual number. **(C)** Percentages of clones with AGS out of all clones in each patient. **(D)** Average numbers of AGS in tumor and non-tumor clones from each patient, and **(E)** from each sample in each patient. **p-value < 0.005.

later tumors (Table 1). However, in Pt1, the numbers of SHMs observed in the third sample was not significantly higher than that detected in the two former samples, while the number of AGS in the third sample was much higher than the rest, implying that in addition to the influence of the number of SHMs on the number of AGS (22), there might be additional selection for N-glycosylated sequences in FL cases (or impairment of selection against them).

FL tumor clones display more branched lineage trees and may have had lower initial affinities and selection thresholds than non-tumor and healthy GC clones

In order to further quantify the differences between the dynamics of SHM and antigen-driven selection in healthy GCs, FL-S and FL-HTS, we performed a quantitative analysis of lineage tree topologies, using our program MTree© (35, 36). Tumor clones from the three FL-HTS patient samples presented significantly larger average outgoing degree (OD-avg, that is, number of children per node), which is a branching measure, when compared to the non-tumor clones detected with HTS (Figure 5A). Non-tumor clones presented values around 1, indicating that most non-tumor trees were not highly branched. According to simulations (37), when trees are not highly branched, it suggests that either the initial affinity of the clone's B-cell receptor to the driving antigen was not very high, or that antigen-driven selection was rather stringent. As in the present study we analyzed fewer sequences than in the simulations, many of the non-tumor clones contained only one sequence and (creating "sticks" rather than branched trees), although some of those may have been part of larger, branched clones that were undetected. However, we believe that this could not account for the large, significant differences in the degree of branching between tumor and non-tumor clones, as all

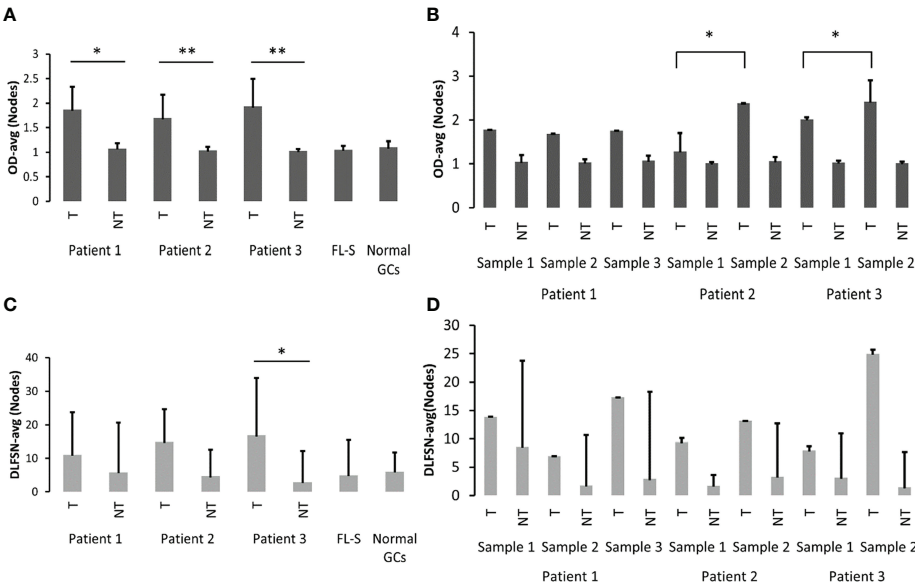


FIGURE 5
Lineage tree topologies. **(A)** Outgoing degree (OD-avg) of tumor and non-tumor clones of the three FL-HTS patients, FL-S and healthy GC samples. **(B)** OD-avg of tumor and non-tumor clones in each sample from all three FL-HTS patients. **(C)** The average distance from a leaf to the first split node/fork (DLFSN-avg) of tumor and non-tumor clones of the three FL-HTS patients, FL-S and healthy GC samples. **(D)** DLFSN-avg of tumor and non-tumor clones in each sample in each of the three FL-HTS patients. *p-value < 0.05; **p-value < 0.005.

our other studies show that OD-avg is always close to 1 in non-tumor clones (Mehr lab, unpublished data). Furthermore, tumor clones from the later biopsies of patients 2 and 3 showed significantly larger OD-avg values than those of tumor clones from the earlier biopsy in each case (Figure 5B). This may imply that the later tumors are more diversified than the earlier tumor clones, possibly due to weakening of the selection forces operating on the tumor clone with time.

We also observed differences in tree length measures: trunk length, minimum root to leaf path length (PL-min) and minimum distance from the root to any split node/fork (DRSN-min) (37). However, PL-min and DRSN-min mostly include the trunk length, which in turn includes the clone's history of mutations, some of which may have been acquired before the transformation event into FL (or t-FL). Thus, in Figure 5 we only show the average distance from a leaf back to the first split node/fork (DLFSN-avg), that is, the paths to leaves without the trunks, which were found in simulations to be correlated with lower initial affinity and selection thresholds (37). Tumor clones presented larger DLFSN-avg values than non-tumor clones, which supports the suggestion that transformation (and possibly also relapse) decreases the sensitivity of the clone to selection. This result was significant only in Pt3; although the same trend also appeared in patients 1 and 2 (Figures 5C, D), it was not statistically significant; this could stem from the low number of non-tumor sequences compared with tumor sequences. Overall, the larger branching and length tree measurements presented by tumor clones from the three FL-HTS samples indicate larger trees, and thus more diversification than those in healthy subject GCs.

Discussion

Our goal in this study was to characterize the clonal evolution and SHM mechanisms of FL tumor clones, across sequential LN biopsies from the same patient (26). We observed large, highly branched lineage trees with long trunks that, together with the mutation patterns, clearly support the GC origin ascribed to FL. FL tumor clones presented more branched trees than healthy GC samples and even non-tumor clones in the same patients, with lineage tree topological measures indicative of lower initial affinities and/or selection thresholds. Moreover, these measures were similar in FL and t-FL samples from Pt1, but increased between biopsies of patients 2 and 3. Indeed, it has previously been shown that antigen-driven selection may persist after transformation and participate in diversification and progression (48–50).

We identified both direct and divergent clonal evolution patterns in the studied samples, and this was supported by mutation analyses. The similar mutation patterns in the FL/t-FL tumor clones of Pt1 may fit the hypothesis of direct evolution, while the different mutation patterns in consecutive biopsies from patients 2 and 3 may point at the existence of a CPC, different descendant clones of which were sampled in each biopsy.

Moreover, the similar tree topological properties of the FL and t-FL samples in Pt1 were consistent with direct evolution, while different and, in particular, increased tree measures (that suggest a possible decrease in selection) between two consecutive biopsies in pt2 and in pt3 are also in line with the existence of a CPC.

Previous studies have shown that the analysis of DNA motifs around mutated nucleotides in the Ig genes can reveal many aspects concerning the targeting of SHM mechanism induced by AID to conserved sequence motifs (39, 40, 42, 51–57), while the analysis of mutation spectra can reveal changes in repair mechanisms (58, 59). We thus analyzed the mutation characteristics, including targeting motifs, of the seven sequential FL samples collected from the three patients, in search for evidence of such changes. Mutation distributions of tumor and non-tumor clones were different, while FL and t-FL tumor clones had similar mutation distributions. This may imply that there was no change in the SHM mechanism between t-FL and FL tumors, and that the latter transformation event did not affect these mutations. In addition, as samples from Pt1 were taken after different treatments, we may speculate that the observed SHM patterns were intrinsic to FL B cells and were not affected by the therapy. In contrast, tumor clones from consecutive samples of patients 2 and 3 differed in their mutation frequencies, suggesting that either the SHM mechanisms have changed, or there was a decrease in the tumor cell sensitivity to selection (9). Compared to a recently published analysis (60), our analysis of mutations is, on one hand, more precise, as it is done on lineage trees so that each mutation is defined relative to the closest known or deduced ancestor (Neumann et al., Front. Immunol., *in press*); and, on the other hand, it was limited to IgH coding regions, so we have no findings on non-Ig regions.

The biased frequencies of mutations from G over C we observed in FL may suggest that there was a bias for generating these mutations on only one strand during the second phase of SHM. Moreover, an elevated number of mutations from G was ascribed to DNA mismatch repair protein MutS homolog 2 (MSH2), uracil-DNA glycosylase (UNG) or DNA repair protein REV1 deficiency (61). This mutation pattern also suggests the possible intervention of a reverse transcription step in fixing the pattern in DNA (62). Furthermore, the difference in transition-transversion mutation ratios between samples raised the question of whether DNA repair mechanisms are altered in FL. Healthy replication over abasic sites after U removal by base excision repair (BER) followed by UNG can lead to G/C targeted transversion mutations (62). Thus, UNG overexpression or enhanced activity may cause transition-transversion ratios smaller than 1 in FL. Thus, SHM mechanisms in FL have to be more thoroughly examined by gene expression or proteomics for detecting enzyme expression levels.

Finally, the number of AGS was higher in tumor clones than in non-tumor clones in all FL-HTS samples, implying it is likely that AGS have some role in FL development, at least in the initial stage, as was previously suggested (47). In addition, because the later samples from all patients had the highest numbers of AGS per clone, we may conclude that the tumors accumulate AGS over time. This fits with the model of a B cell tumor population entrapped in the germinal center that keep undergoing SHM, with selection against AGS – and possibly other potentially harmful mutations – being impaired in the tumor B cells.

To summarize the mutation analyses, we found differences in the mutation distributions from each of the nucleotides, in initial clone affinity and in selection thresholds between tumor and non-tumor clones, but no differences between FL and t-FL clones. Additionally, we observed that tumor clones tend to accumulate larger numbers of potential N-glycosylation sites due to SHM. Taken together, these results suggest that transformation from FL into t-FL, in contrast to the initial transformation to FL, is not characterized by any major changes in DNA repair mechanisms, SHM, or shape of lineage trees, although the possibility of subtle changes in enzyme expression or activity should still be investigated. On the other hand, selection – at least against accumulation of AGS – seems to be impaired in FL and t-FL. This study also shows that even a few samples can provide many interesting insights, provided that these samples contain sufficient numbers of sequences and mutations.

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: GenBank, OP426445-OP428638.

Ethics statement

The studies involving human participants were reviewed and approved by The North East London Research Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MM developed some of the computational tools, performed all the computational analyses and wrote the

manuscript, which was part of her PhD thesis in Bar-Ilan University. EC performed all the experiments and wrote the manuscript. HH developed some of the computational tools and helped with the analysis. JG supervised the experimental research. RM designed and supervised the research and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by US-Israel Binational Science Foundation (BSF) grant number 20130432 and Israel Science Foundation grant number 270/09 (to RM). MM and HH were supported by Bar-Ilan University President's PhD Scholarships.

Acknowledgments

The authors are indebted to Dr. Meirav Kedmi for critical reading of an early version of the manuscript, and to Hadas Neuman for help in manuscript preparation.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1029995/full#supplementary-material

References

1. Gribben JG. How I treat indolent lymphoma. *Blood* (2007) 109:4617–26. doi: 10.1182/blood-2006-10-041863

2. Montoto S, Davies AJ, Matthews J, Calaminici M, Norton AJ, Amess J, et al. Risk and clinical implications of transformation of follicular lymphoma to diffuse large b-cell lymphoma. *J Clin Oncol* (2007) 25:2426–33. doi: 10.1200/JCO.2006.09.3260

3. Lossos IS, Gascoyne RD. Transformation of follicular lymphoma. *Best Pr Res Clin Haematol* (2011) 24:147–63. doi: 10.1016/j.beha.2011.02.006

4. Pasqualucci L, Khiabanian H, Fangazio M, Vasishtha M, Messina M, Holmes AB, et al. Genetics of follicular lymphoma transformation. *Cell Rep* (2014) 6:130–40. doi: 10.1016/j.celrep.2013.12.027

5. Carlotti E, Wrench D, Matthews J, Iqbal S, Davies A, Norton A, et al. Transformation of follicular lymphoma to diffuse large b-cell lymphoma may occur by divergent evolution from a common progenitor cell or by direct evolution from the follicular lymphoma clone. *Blood* (2009) 113:3553–7. doi: 10.1182/blood-2008-08-174839

6. Eide MB, Liestøl K, Lingjaerde OC, Hystad ME, Kresse SH, Meza-Zepeda L, et al. Genomic alterations reveal potential for higher grade transformation in follicular lymphoma and confirm parallel evolution of tumor cell clones. *Blood* (2010) 116:1489–97. doi: 10.1182/blood-2010-03-272278

7. Aarts WM. Variable heavy-chain gene analysis of follicular lymphomas: subclone selection rather than clonal evolution over time. *Blood* (2001) 98:238–40. doi: 10.1182/blood.V98.1.238

8. Matolcsy A, Schattner EJ, Knowles DM, Casali P. Clonal evolution of b cells in transformation from low- to high-grade lymphoma. *Eur J Immunol* (1999) 29:1253–64. doi: 10.1002/(SICI)1521-4141(199904)29:04<1253::AID-IMMU1253>3.0.CO;2-8

9. Ruminy P, Jardin F, Picquenot J-M, Parmentier F, Contentin N, Buchonnet G, et al. S(mu) mutation patterns suggest different progression pathways in follicular lymphoma: early direct or late from FL progenitor cells. *Blood* (2008) 112:1951–9. doi: 10.1182/blood-2007-11-124560

10. Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Sci (80-)* (2009) 324:807–10. doi: 10.1126/science.1170020

11. Boyd SD, Gaëta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* (2010) 184:6986–92. doi: 10.4049/jimmunol.1000445

12. Briney BS, Willis JR. B a McKinney; JE crowe. high-throughput antibody sequencing reveals genetic evidence of global regulation of the naïve and memory repertoires that extends across individuals. *Genes Immun* (2012) 13:469–73. doi: 10.1038/gene.2012.20

13. Wu Y-C, Kipling D, Leong HS, Martin V, Ademokun A, Dunn-Walters DK. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory b-cell populations. *Blood* (2010) 116:1070–8. doi: 10.1182/blood-2010-03-275859

14. Warren EH. F a matsen; J chou. high-throughput sequencing of b- and T-lymphocyte antigen receptors in hematology. *Blood* (2013) 122:19–22. doi: 10.1182/blood-2013-03-453142

15. Dekosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* (2013) 31:166–9. doi: 10.1038/nbt.2492

16. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends Immunol* (2015) 36:738–49. doi: 10.1016/j.it.2015.09.006

17. Chaudhary N, Wesemann DR. Analyzing immunoglobulin repertoires. *Front Immunol* (2018) 9:1–18. doi: 10.3389/fimmu.2018.00462

18. Polak P, Mehr R, Yaari G. *Immunoglobulin Clonotype and Ontogeny Inference*. Encyclopedia of Bioinformatics and Computational Biology (2019). Eds. Ranganathan S., Nakai K., Schönbach C.Gribskov M. (Oxford: Elsevier) 2:972–983.

19. Smakaj E, Babrak L, Ohlin M, Shugay M, Briney B, Tosoni D, et al. Benchmarking immunoinformatic tools for the analysis of antibody repertoire sequences. *Bioinformatics* (2020) 36:1731–9. doi: 10.1093/bioinformatics/btz845

20. Casellas R, Yamane A, Kovalchuk AL, Potter M. Restricting activation-induced cytidine deaminase tumorigenic activity in b lymphocytes. *Immunology* (2009) 126:316–28. doi: 10.1111/j.1365-2567.2008.03050.x

21. Zuckerman NS, McCann KJ, Ottensmeier CH, Barak M, Shahaf G, Edelman H, et al. Ig gene diversification and selection in follicular lymphoma, diffuse large b cell lymphoma and primary central nervous system lymphoma revealed by lineage tree and mutation analyses. *Int Immunol* (2010) 22:875–87. doi: 10.1093/intimm/dxq441

22. Dunn-Walters D, Boursier L, Spencer J. Effect of somatic hypermutation on potential n-glycosylation sites in human immunoglobulin heavy chain variable regions. *Mol Immunol* (2000) 37:107–13. doi: 10.1016/S0161-5890(00)00038-9

23. Sabouri Z, Schofield P, Horikawa K, Spierings E, Kipling D, Randall KL, et al. Redemption of autoantibodies on anergic b cells by variable-region glycosylation and mutation away from self-reactivity. *Proc Natl Acad Sci U.S.A.* (2014) 111:E2567–75. doi: 10.1073/pnas.1406974111

24. Zhu D, McCarthy H, Ottensmeier CH, Johnson PWM, Hamblin TJ, Stevenson FK. Acquisition of potential n-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. *Blood* (2002) 99:2562–8. doi: 10.1182/blood.V99.7.2562

25. Goulabchand R, Vincent T, Batteux F, Eliaou J-F, Guilpain P. Impact of autoantibody glycosylation in autoimmune diseases. *Autoimmun Rev* (2014) 13:742–50. doi: 10.1016/j.autrev.2014.02.005

26. Carlotti E, Wrench D, Rosignoli G, Marzec J, Sangaralingam A, Hazanov L, et al. High throughput sequencing analysis of the immunoglobulin heavy chain gene from flow-sorted b cell sub-populations define the dynamics of follicular lymphoma clonal evolution. *PloS One* (2015) 10:1–20. doi: 10.1371/journal.pone.0134833

27. Banerjee M, Mehr R, Belelovsky A, Spencer J, Dunn-Walters DK. Age- and tissue-specific differences in human germinal center b cell selection revealed by analysis of IgVH gene hypermutation and lineage trees. *Eur J Immunol* (2002) 32:1947–57. doi: 10.1002/1521-4141(200207)32:7<1947::AID-IMMU1947>3.0.CO;2-1

28. Zuckerman NS, Howard WA, Bismuth J, Gibson K, Edelman H, Berrih-Aknin S, et al. Ectopic GC in the thymus of myasthenia gravis patients show characteristics of normal GC. *Eur J Immunol* (2010) 40:1150–61. doi: 10.1002/eji.200939914

29. Volpe JM, Cowell LG, Kepler TB. SoDA: Implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. *Bioinformatics* (2006) 22:438–44. doi: 10.1093/bioinformatics/btk004

30. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. *Bioinformatics* (2007) 23:2947–8. doi: 10.1093/bioinformatics/btm404

31. Michaeli M, Barak M, Hazanov L, Noga H, Mehr R. Automated analysis of immunoglobulin genes from high-throughput sequencing: Life without a template. *J Clin Bioinforma* (2013) 3:1. doi: 10.1186/2043-9113-3-15

32. Michaeli M, Noga H, Tabibian-Keissar H, Barshack I, Mehr R. Automated cleaning and pre-processing of immunoglobulin gene sequences from high-throughput sequencing. *Front Immunol* (2012) 3:386. doi: 10.3389/fimmu.2012.00386

33. Barak M, Zuckerman NS, Edelman H, Unger R, Mehr R. IgTree©: Creating immunoglobulin variable region gene lineage trees. *J Immunol Methods* (2008) 338:67–74. doi: 10.1016/j.jim.2008.06.006

34. Tabibian-Keissar H, Zuckerman NS, Barak M, Dunn-Walters DK, Steiman-Shimony A, Chowers Y, et al. B-cell clonal diversification and gut-lymph node trafficking in ulcerative colitis revealed using lineage tree analysis. *Eur J Immunol* (2008) 38:2600–9. doi: 10.1002/eji.200838333

35. Dunn-Walters DK, Belelovsky A, Edelman H, Banerjee M, Mehr R. The dynamics of germinal centre selection as measured by graph-theoretical analysis of mutational lineage trees. *Dev Immunol* (2002) 9:233–43. doi: 10.1080/10446670310001593541

36. Dunn-Walters DK, Edelman H, Mehr R. Immune system learning and memory quantified by graphical analysis of b-lymphocyte phylogenetic trees. *Biosystems* (2004) 76:141–55. doi: 10.1016/j.biosystems.2004.05.011

37. Shahaf G, Barak M, Zuckerman NS, Swerdlin N, Gorfine M, Mehr R. Antigen-driven selection in germinal centers as reflected by the shape characteristics of immunoglobulin gene lineage trees: A large-scale simulation study. *J Theor Biol* (2008) 255:210–22. doi: 10.1016/j.jtbi.2008.08.005

38. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

39. Dörner T, Brezinschek HP, Brezinschek RI, Foster SJ, Domiati-Saad R, Lipsky PE. Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. *J Immunol* (1997) 158:2779–89.

40. Dörner T, Foster SJ, Brezinschek HP, Lipsky PE. Analysis of the targeting of the hypermutational machinery and the impact of subsequent selection on the distribution of nucleotide changes in human VHDJH rearrangements. *Immunol Rev* (1998) 162:161–71. doi: 10.1111/j.1600-065X.1998.tb01439.x

41. Zuckerman NS, Hazanov H, Barak M, Edelman H, Hess S, Shcolnik H, et al. Somatic hypermutation and antigen-driven selection of b cells are altered in autoimmune diseases. *J Autoimmun* (2010) 35:325–35. doi: 10.1016/j.jaut.2010.07.004

42. Spencer J, Dunn-Walters DK. Hypermutation at a-T base pairs: the a nucleotide replacement spectrum is affected by adjacent nucleotides and there is no reverse complementarity od sequences flanking mutated a and T nucleotides. *J Immunol* (2005) 175:5170–7. doi: 10.4049/jimmunol.175.8.5170

43. Iosselevitch I, Tabibian-Keissar H, Barshack I, Mehr R. Gastric DLBCL clonal evolution as function of patient age. *Front Immunol* (2022) 13:957170. doi: 10.3389/fimmu.2022.957170

44. Teng G, Papavasiliou FN. Immunoglobulin somatic hypermutation. *Annu Rev Genet* (2007) 41:107–20. doi: 10.1146/annurev.genet.41.110306.130340

45. Jankovic M, Nussenzweig A, Nussenzweig MC. Antigen receptor diversification and chromosome translocations. *Nat Immunol* (2007) 8:801–8. doi: 10.1038/ni1498

46. Weill J-C, Reynaud C-A. DNA Polymerases in adaptive immunity. *Nat Rev Immunol* (2008) 8:302–12. doi: 10.1038/nri2281

47. Zhu D, Ottensmeier CH, Du M-Q, McCarthy H, Stevenson FK. Incidence of potential glycosylation sites in immunoglobulin variable regions distinguishes between subsets of burkitt's lymphoma and mucosa-associated lymphoid tissue lymphoma. *Br J Haematol* (2003) 120:217–22. doi: 10.1046/j.1365-2141.2003.04064.x

48. Zelenetz A. Histologic transformation of follicular lymphoma to diffuse lymphoma represents tumor progression by a single malignant b cell. *J Exp Med* (1991) 173:197–207. doi: 10.1084/jem.173.1.197

49. Zelenetz AD. Clonal expansion in follicular lymphoma occurs subsequent to antigenic selection. *J Exp Med* (1992) 176:1137–48. doi: 10.1084/jem.176.4.1137

50. Bahler DW, Zelenetz AD, Chen TT, Levy R. Antigen selection in human lymphomagenesis. *Cancer Res* (1992) 52:5547s.

51. Smith DS, Creadon G, Jena PK, Portanova JP, Kotzin BL, Wysocki LJ. Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive b cells. *J Immunol* (1996) 156:2642–52.

52. Cowell LG, Kepler TB. The nucleotide-replacement spectrum under somatic hypermutation exhibits microsequence dependence that is strand-symmetric and distinct from that under germline mutation. *J Immunol* (2000) 164:1971–6. doi: 10.4049/jimmunol.164.4.1971

53. Shapiro GS, Ellison MC, Wysocki LJ. Sequence-specific targeting of two bases on both DNA strands by the somatic hypermutation mechanism. *Mol Immunol* (2003) 40:287–95. doi: 10.1016/S0161-5890(03)00101-9

54. Feng J, Shaw DA, Minin VN, Simon N. FA 4th matsen. survival analysis of DNA mutation motifs with penalized proportional hazards. *Ann Appl Stat* (2019) 13:1268–94. doi: 10.1214/18-AOAS1233

55. Feng Y, Seija N, Di Noia JM, Martin A. JM d:i noia; a martin. AID in antibody diversification: There and back again. *Trends Immunol* (2020) 41:586–600. doi: 10.1016/j.it.2020.04.009

56. Elhanati Y, Sethna Z, Marcou Q, Callan CGJ, Mora T, Walczak AM. Inferring processes underlying b-cell repertoire diversity. *Philos Trans R Soc Lond B Biol Sci* (2015) 370(1676):20140243. doi: 10.1098/rstb.2014.0243

57. Cui A, Di Niro R, Vander Heiden JA, Briggs AW, Adams K, Gilbert T, et al. A model of somatic hypermutation targeting in mice based on high-throughput ig sequencing data. *J Immunol* (2016) 197:3566–74. doi: 10.4049/jimmunol.1502263

58. Dolle MET, Snyder WK, Gossen JA, Lohman PHM, Vijg J. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc Natl Acad Sci* (2000) 97:8403–8. doi: 10.1073/pnas.97.15.8403

59. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature* (2007) 446:153–8. doi: 10.1038/nature05610

60. Ye X, Ren W, Liu D, Li X, Li W, Wang X, et al. Genome-wide mutational signatures revealed distinct developmental paths for human b cell lymphomas. *J Exp Med* (2021) 218(2):e20200573. doi: 10.1084/jem.20200573

61. Kim N, Storb U. The role of DNA repair in somatic hypermutation of immunoglobulin genes. *J Exp Med* (1998) 187:1729–33. doi: 10.1084/jem.187.11.1729

62. Steele EJ. Mechanism of somatic hypermutation: critical analysis of strand biased mutation signatures at A:T and G:C base pairs. *Mol Immunol* (2009) 46:305–20. doi: 10.1016/j.molimm.2008.10.021

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 14 December 2022
DOI 10.3389/fonc.2022.1079772

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

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

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

[1]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, [2]Laboratory of Biological Chemistry, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece, [3]Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece, [4]Department of Hematology, APHP, Hôpital Pitié-Salpêtrière and Sorbonne University, Paris, France, [5]Department of Biology, School of Science, National and Kapodistrian University of Athens, Athens, Greece, [6]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® Ultra™ II FS DNA Library Prep Kit, New England Biolabs, USA). Library quantification was performed with Qubit™ (ThermoFisher Scientific), whereas purity and size were checked with the Fragment Analyzer™ 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>=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 IMGT/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 α=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 IGHJ 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

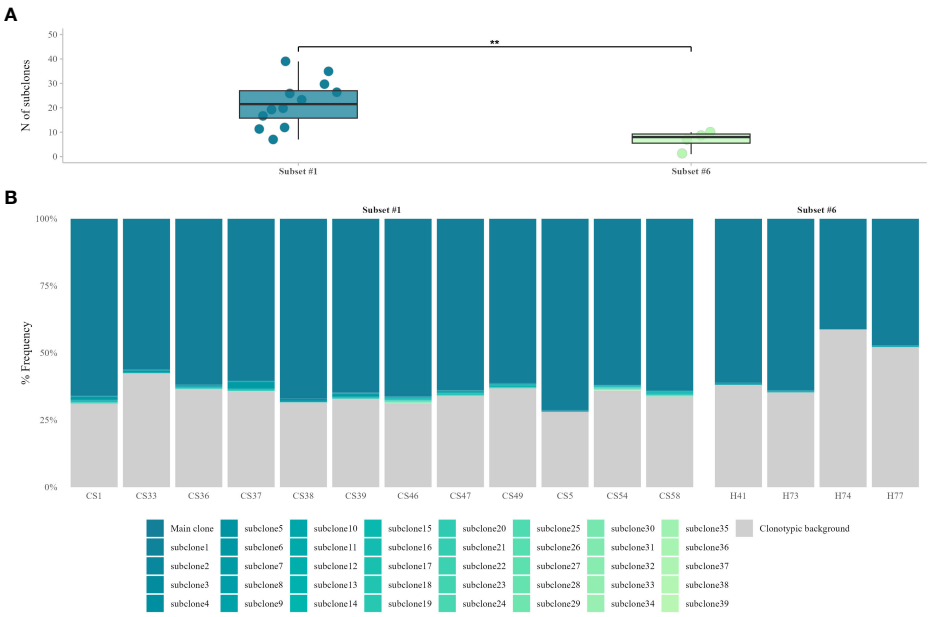


FIGURE 1
VH CDR3-derived, subclonal heterogeneity in 'truly unmutated' BcR IG rearrangements. **(A)** Comparison of subclonal branching of the VH CDR3 between the two stereotyped subsets shows significant differences (p = 0.005) in the number of subclonal variants, with subset #1 displaying higher VH CDR3 intraclonal variability. **(B)** Contribution of 'truly unmutated' IGHV-IGHD-IGHJ rearrangements to the total repertoire, for stereotyped subsets #1 and #6. The main CLL clonotype and its respective CLL-related, 'truly unmutated' subclonal variants are depicted in different shades of blue. The clonotypic background consists of either sequences bearing SHM in the rearranged IGHV gene, subclonal variants with very small frequencies or CLL-unrelated rearrangements. **, p=0.005.

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).

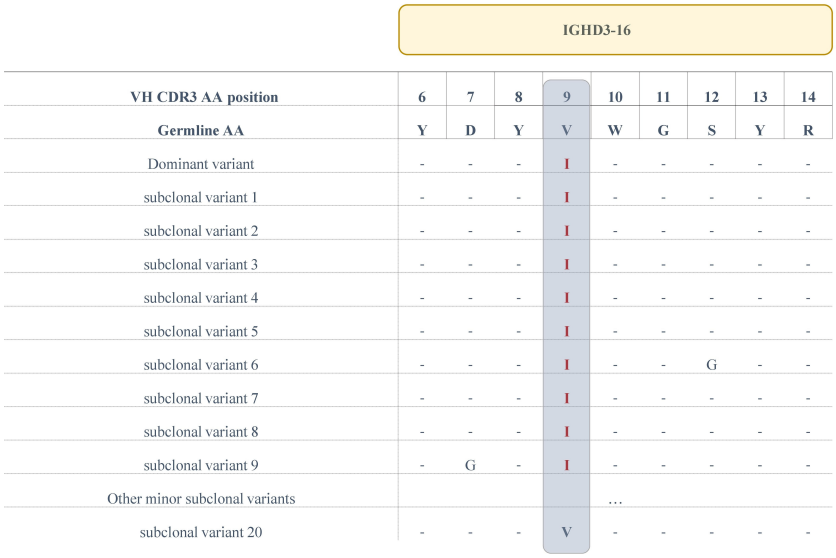
position 2 in either the gamma or the alpha switched variants or both (median frequency: 0.11% for both transcripts, range: 0.05-1.08%, median frequency for gamma and alpha trasncripts: 0.08 and 0.14%, respectively). Moreover, the Q>R aa substitution at VH CDR3 position 4 was shared by 8/12 (75%) patients 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



	IGHD3-16								
VH CDR3 AA position	6	7	8	9	10	11	12	13	14
Germline AA	Y	D	Y	V	W	G	S	Y	R
Dominant variant	-	-	-	I	-	-	-	-	-
subclonal variant 1	-	-	-	I	-	-	-	-	-
subclonal variant 2	-	-	-	I	-	-	-	-	-
subclonal variant 3	-	-	-	I	-	-	-	-	-
subclonal variant 4	-	-	-	I	-	-	-	-	-
subclonal variant 5	-	-	-	I	-	-	-	-	-
subclonal variant 6	-	-	-	I	-	-	G	-	-
subclonal variant 7	-	-	-	I	-	-	-	-	-
subclonal variant 8	-	-	-	I	-	-	-	-	-
subclonal variant 9	-	G	-	I	-	-	-	-	-
Other minor subclonal variants					...				
subclonal variant 20	-	-	-	V	-	-	-	-	-

FIGURE 3
Alignment of the IGHD3-16-encoded region of the VH CDR3 of the main clonotype and subclonal variants for patient H73. The sequences of all variants (at the amino acid level) that display a V>I aa substitution at position 9 are depicted, as well as a single subclonal variant bearing the germline IGHD3-16-encoded Valine at the same position.

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 IGHD6-19-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 occuring 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 stereotypd 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

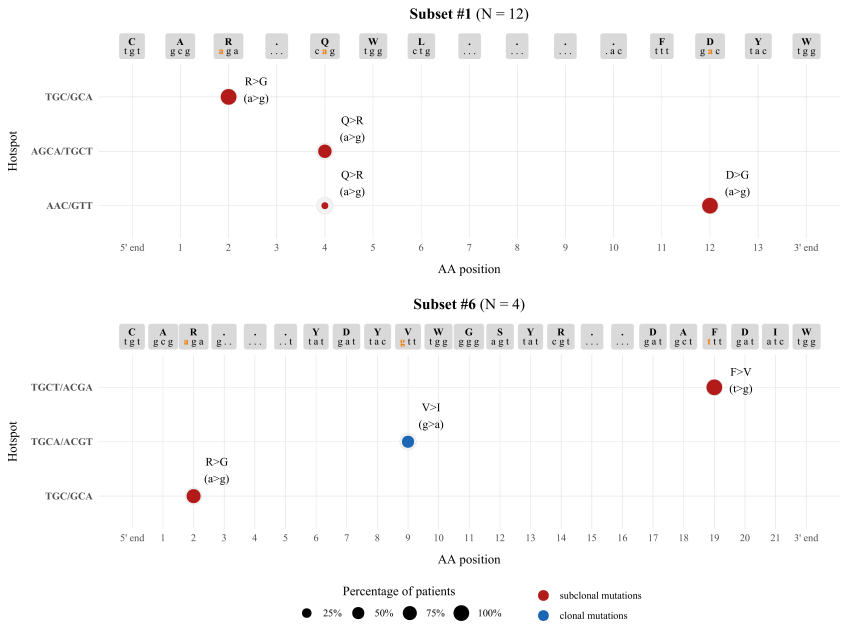


FIGURE 4
Topology of SHMs observed within the VH CDR3 in relation to AID hotspots. The size of each node represents the percentage of patients exhibiting a particular SHM, while the graph coordinates of each node correspond to the AID hotspot involved in each SHM. Different colors are used to distinguish between clonal (blue) and subclonal (red) SHMs.

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.

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

References

1. 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

2. 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

3. 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

4. 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

5. 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

6. 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

7. 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

8. 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

9. 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

10. 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

11. 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

12. 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

13. 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

14. 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

15. 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

16. 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

17. 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

18. 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

19. 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

20. 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.

21. 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.

22. Stamatopoulos K, Belessi C, Moreno C, Boudjograh 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.

23. 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

24. 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.

25. 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.

26. 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

27. 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

28. 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

29. 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

30. 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

31. Gemenetzi K, Psomopoulos F, Carriles A, Gounari M, Minici C, Plevova K, et al. Higher order immunoglobulin repertoire restrictions in cll: the illustrative case of stereotyped subsets 2 and 169. *Blood* (2020) 137(14):1895–904. doi: 10.1182/blood.2020005216

32. 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

33. Sutton LA, Kostareli E, Hadzidimitriou A, Darzentas N, Tsaftaris A, Anagnostopoulos A, et al. Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors:

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 polη 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. Oppezzo 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

frontiers | Frontiers in Oncology

TYPE Methods
PUBLISHED 16 January 2023
DOI 10.3389/fonc.2022.1111209

OPEN ACCESS

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

REVIEWED BY
Fabio Guolo,
San Martino Hospital (IRCCS), Italy
Azza Mahmoud Kamel,
Cairo University, Egypt

*CORRESPONDENCE
Anke Schilhabel
anke.schilhabel@uksh.de

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

RECEIVED 29 November 2022
ACCEPTED 28 December 2022
PUBLISHED 16 January 2023

CITATION
Schilhabel A, Szczepanowski M, van Gastel-Mol EJ, Schillalies J, Ray J, Kim D, Nováková M, Dombrink I, van der Velden VHJ, Boettcher S, Brüggemann M, Kneba M, van Dongen JJM, Langerak AW and Ritgen M (2023) Patient specific real-time PCR in precision medicine – Validation of IG/TR based MRD assessment in lymphoid leukemia.
Front. Oncol. 12:1111209.
doi: 10.3389/fonc.2022.1111209

Patient specific real-time PCR in precision medicine – Validation of IG/TR based MRD assessment in lymphoid leukemia

Anke Schilhabel[1*], Monika Szczepanowski[1], Ellen J. van Gastel-Mol[2], Janina Schillalies[1], Jill Ray[3], Doris Kim[3], Michaela Nováková[4], Isabel Dombrink[1], Vincent H. J. van der Velden[2], Sebastian Boettcher[5], Monika Brüggemann[1], Michael Kneba[1], Jacques J. M. van Dongen[6], Anton W. Langerak[2] and Matthias Ritgen[1]

[1]Hämatologie Labor Kiel, Medical Department II, Hematology and Oncology, University Medical Center Schleswig-Holstein, Kiel, Germany, [2]Laboratory Medical Immunology, Department of Immunology, Erasmus Medical Center (MC), University Medical Center Rotterdam, Rotterdam, Netherlands, [3]Oncology Biomarker Development, Genentech, Inc., South San Francisco, CA, United States, [4]Childhood Leukemia Investigation Prague (CLIP)-Department of Pediatric Hematology and Oncology, Second Medical Faculty, Charles University and University Hospital Motol, Prague, Czechia, [5]Department of Medicine III Hematology, Oncology and Palliative Care, University Hospital, Rostock, Germany, [6]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, Netherlands

Detection of patient- and tumor-specific clonally rearranged immune receptor genes using real-time quantitative (RQ)-PCR is an accepted method in the field of precision medicine for hematologic malignancies. As individual primers are needed for each patient and leukemic clone, establishing performance specifications for the method faces unique challenges. Results for series of diagnostic assays for CLL and ALL patients demonstrate that the analytic performance of the method is not dependent on patients' disease characteristics. The calibration range is linear between 10^{-1} and 10^{-5} for 90% of all assays. The detection limit of the current standardized approach is between 1.8 and 4.8 cells among 100,000 leukocytes. RQ-PCR has about 90% overall agreement to flow cytometry and next generation sequencing as orthogonal methods. Accuracy and precision across different labs, and above and below the clinically applied cutoffs for minimal/measurable residual disease (MRD) demonstrate the robustness of the technique. The here reported comprehensive, IVD-guided analytical validation provides evidence that the personalized diagnostic methodology generates robust, reproducible and specific MRD data when standardized protocols for data generation and evaluation are used. Our approach may also serve as a guiding example of how to accomplish analytical validation of personalized in-house diagnostics under the European IVD Regulation.

KEYWORDS

MRD, RQ-PCR, IG rearrangement, TR rearrangement, personalized diagnostics, IVDR, method validation, EuroMRD

1 Introduction

Allele-specific oligonucleotide real-time quantitative PCR (ASO-PCR) is an accepted method in the field of specialized diagnostics for hematologic malignancies to analyze minimal residual disease (MRD). It is being used for several decades, has been established as a prognostic marker (1–4) and clinical standard in many different lymphatic neoplasms (5–8). The test results guide clinical decisions in the therapy of acute lymphoblastic leukemia (ALL) and highlight the influence of analytical methods in precision medicine. MRD is the term used for small numbers of malignant cells that remain in the peripheral blood (PB) or bone marrow (BM) during or after treatment. Its detection is based on the junctional region of rearranged immunoglobulin (IG) heavy and light chain genes and T-cell receptor (TR) genes, which are fingerprint-like sequences that can be used as clone-specific PCR targets for the vast majority of B- and T-cell neoplasias (9, 10). In contrast to methods that detect a common target in different samples, for each patient and each individual PCR target an allele-specific oligonucleotide (ASO-primer) is designed. The ASO primer is combined with an adequate primer/probe system to generate a specific real-time quantitative PCR assay, which detects the tumor-related junctional region (CDR3) identified by combination of multiplex PCR and nucleic acid sequencing at the time point of diagnosis or relapse.

High analytical standardization of this methodology has been reached by an international collaboration, the EuroMRD Consortium, which published a data evaluation and interpretation guideline for ASO-PCR (11). The current level of analytical standardization has been reached by the quality objectives of the EuroMRD Consortium organizing round robin testing, training meetings and carefully evaluating the results and pitfalls of the method, as well as by addressing the biology of the targeted molecules and the principles of PCR amplification (11–13). Due to the patient-specific nature of the test, the required validation evidence for a diagnostic assay has only been assessed in a limited way for ASO-PCR. Repeated analysis under different conditions using a defined number of samples to address precision, accuracy or robustness of each individual ASO-PCR assay would require a major investment in time and costs to establish such a test for each patient. Additionally, due to the lack of sufficient reference materials, medical samples for e.g. ALL will only be available by additional bone marrow aspirations from patients.

Several scientific reports from different fields (14–17), regulatory guidelines (18, 19) or standards from different sources (Clinical Laboratory Improvement Amendments (CLIA), Clinical Laboratory Standards Institute (CLSI), International Organization for Standardization (ISO)) exist, addressing the scope and amount of validation activities. There is considerable variation in the suggested scope of a validation and it is necessary to adapt them to the specific analytical method and its intended use. Nevertheless, validation of a method is a formal requirement, both in applicable international standards for the quality management of analytical laboratories, and in regulations from health authorities. It is also included in the European IVD regulation (20) as part of Article 5.5, which besides other declarations and justifications force the laboratories' documentation to adequately address analytical and clinical performance of the used methods.

The current report summarizes retrospective and prospective analytic studies for the validation of the ASO-PCR method. It shows that despite the need for patient-specific reagents, sensitive, precise and accurate results can be obtained, even in the context of the latest developments in MRD techniques. The described approach may serve as a practical guide to laboratories, which have to adequately and reliably align their methods with regulatory requirements.

2 Materials and methods

The scope of the validation was defined from the frequency of the possible PCR targets evaluated from actively managed assays for B-cell chronic lymphocytic leukemia (CLL) (520 assays) and ALL patients (2110 assays B-ALL, 614 assays T-ALL, Figure 1). Preliminary results and further planned activities have been presented to the US Food and Drug Administration (FDA) at several meetings in the application process for ASO-PCR as a drug development tool for CLL clinical trials. After collecting the performance data for CLL, a comparable data set was selected for actively used ALL ASO-PCR assays considering the proportion of the different rearrangement types determined during the preparation of the work and B- and T-ALL frequencies (21–23).

Materials: The validation approach included analytical data from peripheral blood samples collected during the clinical trials CLL11 (NCT01010061), CLL14 (NCT02242942) and bone marrow samples collected during GMALL trials 07/03 (NCT00198991), 08/13 (NCT02881086) and GMALL registry, as well as prospectively analyzed blood samples evaluated in two EuroMRD laboratories. Informed consent was obtained from all subjects involved in the study. Samples for retrospective analysis were randomly selected from a data pool that fulfilled the necessary criteria according to the EuroMRD guidelines (11).

Methods: ASO-PCR of the patient-specific and the control gene assays was performed using the BCR master and detection kit (Roche Diagnostics GmbH, Penzberg, Germany) for CLL, the Lightcycler 480 Probe Master buffer (Roche Diagnostics GmbH, Penzberg, Germany) or QPCR Mastermix Plus (Eurogentec, Seraing, Belgium) for ALL as described by Cazzaniga et al. (24). Assay results were evaluated according to the EuroMRD guidelines (11). MRD values were calculated as the ratio of tumor cells per number of analyzed nucleated cells as assessed by RQ-PCR of the albumin single copy housekeeping gene. Whenever possible the guidelines and standards from the Clinical Standards Laboratory Institute (CLSI) (25–29) were used to derive an appropriate experimental set up and the recommended statistical analysis. Two EuroMRD laboratories were involved in the validation activities. Statistical computing was performed using R (30) and R studio (31).

The considered parameters, the way these were experimentally addressed, as well as the method and standards used for statistical analysis are summarized in Supplementary Table 1. For easier reading throughout the report the term Cp will be used for results obtained by the different data evaluation techniques for RQ-PCR (Cp representing the cycle crossing point of the second derivative of the measured fluorescence intensities during the PCR amplification and Ct representing the cycle threshold set by the user at which exponential amplification during the PCR reaction starts).

10.3389/fonc.2022.1111209

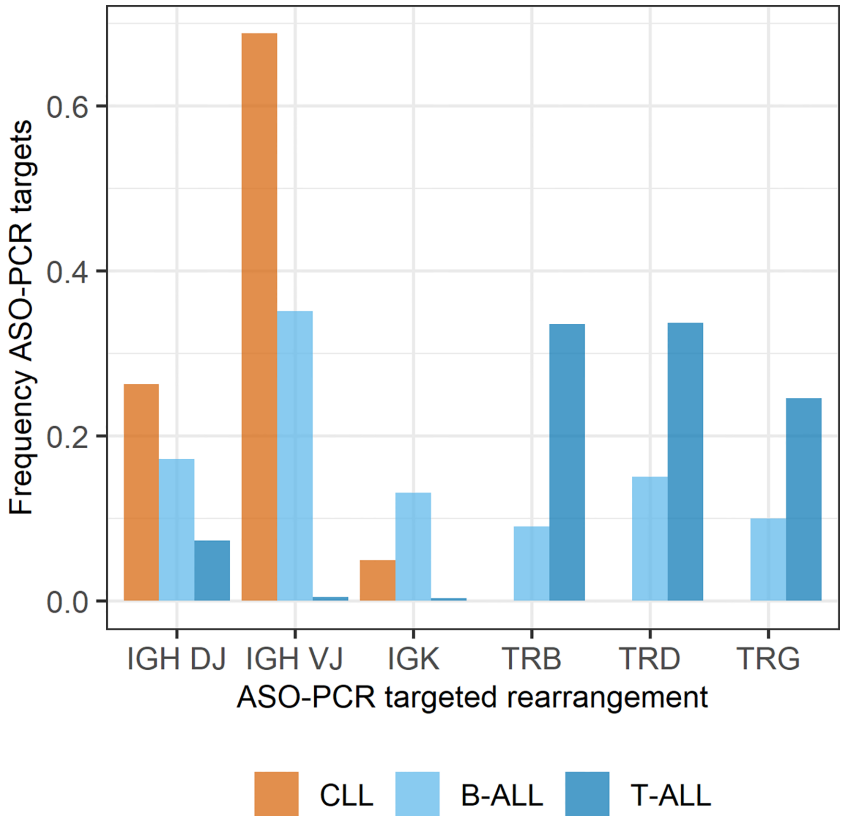


FIGURE 1
Frequency of IG and TR rearrangements as ASO-PCR targets in actively used MRD assays of patients with CLL (N=520), B-ALL (N=1078), and T-ALL (N=322). In accordance with the literature (21), 23% of the ALL patients in our assessment had a T-ALL diagnosis and 77% were diagnosed with B-ALL. For 88% of the patients two MRD assays are established, and ~10% of patients have only one assay for MRD assessment in both ALL sub-entities, thereby allowing assessments on the distribution of ASO-PCR targets from 520, 2110, and 614 assays for CLL, B-ALL and T-ALL, respectively.

3 Results

Critical performance parameters in the process of MRD assessments concerned the primer design (resulting in success and ability of specifically quantifying the rearrangement of interest), the precision, and the accuracy of the method. These were then addressed in prospective validation activities for CLL, which is characterized by the expansion of the malignant clone in peripheral blood, and thus no additional bone marrow samples from patients were required. Additional performance parameters evaluated from retrospective data were the linearity and the lower limits of the calibration range for detection and quantitation, the stability of the medical specimen and the reagents, the interference from endogenous sample ingredients that influence DNA quality or PCR amplification, and the recovery of the amount of malignant cells from a given sample when different DNA extraction methods are used.

3.1 Analytical specificity of ASO-PCR is influenced by non-specific amplification

Analytical specificity is defined by the primer sequence derived from the CDR3 sequence of the different IG and TR gene rearrangements. Although, specificity of a primer has to be determined by sequence comparisons to known IG or TR sequences prior to assay establishment, non-specific amplification might still occur and is addressed according to the EuroMRD criteria using the negative control from PB buffy coat DNA. Variability in non-specific amplification has been reported to depend on the type of target, the type of sample (BM or PB) and the time point during or after therapy for ALL (32). For CLL restriction of the IG receptor repertoire is well known (33), and stereotyped sequences as well as biased somatic hypermutation patterns can be a source of non-specific amplification. When analyzed with one or more non-patient-specific assay, 8.4% (7/83) of the technical replicates of 22 MRD samples of 6 CLL patients showed non-specific amplification. Applying the clinical cut-off $<10^{-4}$ for MRD negativity in CLL, four of the replicates were recorded as MRD negative, and only the remaining three replicates were recorded as false positives due to non-specific amplification. The observed overall ASO-PCR specificity is 96.4% (confidence interval, CI: 89.9; 98.8).

3.2 Accuracy of ASO-PCR is high

Comparison of the accuracy of ASO-PCR to an orthogonal method is influenced by the sensitivity (cells or cell equivalents used for testing) of both compared methods. Currently, two additional methods are routinely used to assess MRD. The first method, multiparametric flow cytometry, relies on the detection of

an aberrant immunophenotype on vital cells, whereas the second approach, high-throughput next generation sequencing (NGS), detects PCR amplified rearranged IG or TR genes in a patients PB or BM DNA. The accuracy of ASO-PCR was therefore determined from comparisons with both methods, flow cytometry and NGS (Table 1).

Compared to flow cytometry the overall qualitative agreement (OA) of ASO-PCR at a MRD cut-off $1x10^{-4}$ (used in CLL to score MRD negativity of patients and assign them to the MRD low risk group in survival assessments in clinical trials, and often used to attest an MRD response in ALL (34)) is about or above 90% for ALL and CLL. Additionally, the agreement for samples with positive MRD status in both methods (positive agreement) as well as the agreement for the negative samples in both methods (negative agreement) at a $1x10^{-4}$ cut-off is balanced at levels >90% (Table 1) for ALL and CLL.

For ALL the agreement of ASO-PCR to NGS (OA 94.3%) is comparable to the agreement reported by Svaton et al. (35) (80.6%) for both methods, as well as to the one observed for flow cytometry (OA 94.1%). The CLL samples show lower agreement to NGS than to flow cytometry, with only 76% OA. This potentially results from the high number of MRD low positive samples (25/62) in the CLL data set, which were all confirmed as MRD positive by the NGS method. These samples are scored as MRD negative at the applied cut-off. Additionally the sensitivity of the NGS analysis is about three times lower compared to ASO-PCR, as the MRD is derived from a single measurement from a pool of 0.25 - 1.5 million cells compared to the mean of a triplicate measurement of up to 150,000 cells for ASO-PCR.

For both entities and comparator methods the Bland-Altman plots show an underestimation of the MRD by ASO-PCR. For CLL this underestimation is 7.6% compared to NGS and 16.3% compared to flow cytometry (Figure 2, flow cytometry yellow, NGS blue). In ALL MRD is underestimated at 1.4% compared to NGS and 26.2% compared to flow cytometry. No systematic bias depending on the MRD level of the samples was detected, although the slope of the linear regressions were between 0.95 and 0.68 for CLL (R^2 = 0.9 and 0.64, flow and NGS, respectively) and 0.81 and 1.02 for ALL (R^2 = 0.86 and 0.8; flow and NGS, respectively). About 6.0% (72/1205 flow cytometry; 3/49 NGS) of the CLL samples, and 2.3% to 4.7% (1/44 NGS; 8/168 flow cytometry) of the ALL samples were outside the 2-sigma range.

An inter-lab comparison between our laboratories showed 93.5% and 86.4% OA (CLL and ALL, respectively) and details are summarized in Table 2. Values of positive MRD samples showed high concordance (R^2 = 0.94 ALL; R2 = 0.97 CLL), which underline the specificity of the different primers for a given biomarker, and the accuracy of the MRD values assessed by two different, albeit technically identical, personalized diagnostic tests.

3.3 Precision of the ASO-PCR is acceptable despite multiple sources of variability

According to the currently applied EuroMRD criteria, precision of ASO-PCR is estimated from the Cp- differences observed in replicate measurements of the standards or samples. This criterion (1.5 Cp) has been derived from the variation of the reporting signals mean to the minimum and maximum reporting signals from >100 samples (10). No estimates for the variance for the calculated MRD values have been published so far. Therefore an experiment was set up to investigate the influence of different sources of variation on the MRD assessment by ASO-PCR. A spike-in approach using 3 CLL patient samples was used to evaluate the effect of random factors on the variance of the method at different MRD levels. Intermediate precision for the MRD range $\geq 10^{-4}$ was <50% at both laboratories, and increased further for MRD levels of $\sim 10^{-5}$ due to the exponential character of the conversion of Cps to copy numbers (Table 3). Considering the data of both labs, none of the tested random factors can be determined as having an equal impact on the total variance of the method across the complete MRD range. Within-day precision (including the within-run precision) contributes evenly and significantly to the overall precision. The MRD status determined from the 144 spiked samples in the two laboratories were highly concordant with an OA of 89.2% (CI: 83.5; 93.1) using the clinical MRD negativity cut-off 10^{-4}, which does not consider qualitative and quantitative RQ-PCR results at concentrations $<10^{-4}$. The OA increased to 97.0% (CI: 93.0; 98.7) when the limit of detection was used for the definition of negativity.

Results of repeated measurements to evaluate intermediate precision for the assessment of clinical specimen can be found in the Supplementary Material (Supplementary Table 2). The precision estimate for the clinical samples are in good accordance to the precision estimates determined from the spike-in experiment, demonstrating the applicability of the spike-in approach using only a very small number of patient samples to establish a precision estimate for ASO-PCR.

TABLE 1 Accuracy of ASO-PCR at the MRD cut-off 10^{-4}.

Comparator method	Entity	Median Sensitivity ASO-PCR	Median Sensitivity comparator	Samples	Patients	Overall agreement [%]	Positive agreement [%]	Negative agreement [%]
Flow cytometry*	CLL	7.90E-06	4.70E-06	2233 PB	304	92.7 (91.5/93.7)	92.7 (90.7/94.2)	92.7 (91.2/93.9)
NGS**	CLL	7.80E-06	2.70E-06	62 PB	23	75.8 (63.8/84.8)	84.1 (70.6/92.1)	55.6 (33.7/75.4)
Flow cytometry*	ALL	1.30E-05	5.00E-06	357 (104 PB, 253 BM)	137	94.1 (91.2/96.1)	94.2 (89.7/96.8)	94.0 (89.6/96.6)
NGS	ALL	4.60E-06	4.60E-06	105 (24 PB, 81 BM)	14	94.3 (88.1/97.4)	91.3 (73.2/97.6)	95.1 (88.1/98.1)

* 4-color or 8-color flow cytometry, no bulk lysis performed, sensitivity was calculated from a threshold of 20 positive events per measured number of nucleated cells.
** high number of ASO-PCR low positive samples ($<1x10^{-4}$) included in data set.
95% confidence interval is given in parentheses.

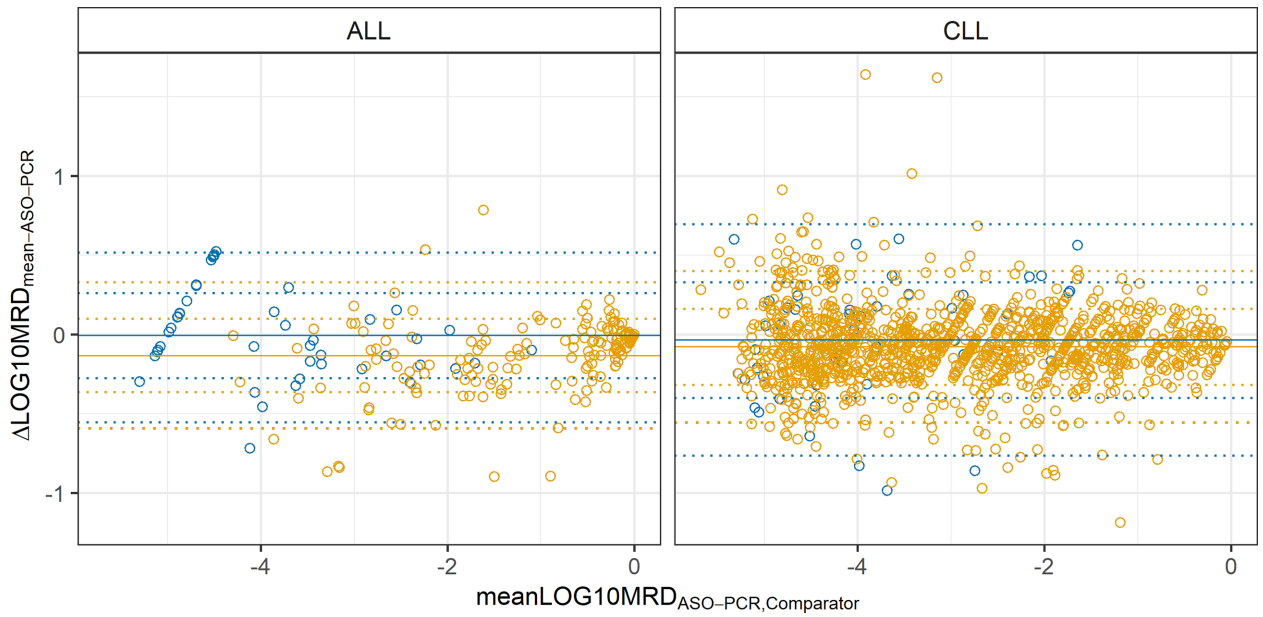


FIGURE 2
Bland-Altman plots of concordant positive MRD samples analyzed by ASO-PCR and 4-color or 8-color flow cytometry (yellow) and NGS (blue). 1205 CLL samples and 169 ALL samples were compared using flow cytometry, and 49 CLL samples and 44 ALL samples were compared using NGS. Bias (solid line) and 1S and 2S intervals (dotted lines) are shown.

3.4 Linearity and LOD of ASO-PCR easily reach the 10^{-4} level and even beyond

The linearity and limits of the calibration range might be affected by a number of factors that influence the amplification of the target during the PCR reaction like starting concentration, reagent quality, amplification efficiency (influenced by factors like mutation status of the target), or primer design. Linearity and limit of detection (LOD) were evaluated from a total of 90 assays (60 CLL, 30 ALL). Deviations from linearity can be visually recognized from a number of standard curves (Supplementary Figure 1), and statistical analysis of the linear and nonlinear models revealed significant nonlinearity for 26/60 CLL assays and 21/30 ALL assays ($p<0.05$). The deviation from linearity for each of the technical replicates from these assays was tested (2^{nd} and 3^{rd} order models) against the set acceptance limit of ≤1.32 Cp deviation from linearity. This acceptance limit is based on the estimated maximum Cp difference that would be observed if the slope differs at acceptable levels (EuroMRD: -3.1 to -3.9 (11),) from the theoretical slope -3.3 for a PCR amplification efficiency of 2. Thus, in total 91.9% (55/60) of CLL assays and 96.7% (29/30) of ALL assays show a linear calibration range from 10^{-1} to 10^{-5}.

Due to the methodological limitations of the assays, leading to measurements without fluorescence signals in the PCR amplification of low concentrated samples and no template controls (blanks), the limit of blank (LOB) for ASO-PCR was determined non-parametrically from the measured Cps using the signal reporting limit of 45 Cp. Based on the Cp, the LOB for CLL were 45 and 40.8 for the two participating laboratories, respectively. Using a probit approach with a hit rate of 95%, where a hit is defined by Cp<LOB, the corresponding LODs were calculated at $5.1x10^{-5}$ and $6.3x10^{-5}$. For ALL the Cp for LOB was determined at 43.5, and LOD was $5.8x10^{-5}$.

3.5 Sample quality and reagent stability effects on ASO-PCR MRD measurements are minor

Usually, medical samples are processed within a short time after sample collection, whereas reagents are used over a longer time period according to the shelf life given by the manufacturer, but storage time of samples and reagents, or repeated freeze-thawing of reagents might influence the quality of MRD assessments. In a global survey to assess

TABLE 2 Personalized ASO-PCR assays independently developed in the two participating laboratories are highly accurate and allow quantitation of MRD samples across different laboratories.

Entity	Study design*	Samples	Patients	OA [%](CI)	PPA [%](CI)	PNA [%](CI)
ALL	EuroMRD QA task 2, 3	44	13	86.4 (73.3/93.6)	85.7 (68.5/94.3)	87.5 (64.0/96.5
CLL	equal to EuroMRD task 3	46	10	93.5 (82.5/97.8)	95.2 (84.2/98.7)	75.0 (30.0/95.4)

* The EuroMRD quality assurance program task 3 provides the biomarker sequences to the participating labs, which have to establish the ASO-PCR assay (design and test the primers, and define LOD and LOQ) before analyzing the follow-up samples. Task 2 includes the identification of the marker from the diagnostic sample.
OA, overall agreement; PPA, percent positive agreement; PNA, percent negative agreement; CI 95%, confidence interval.

TABLE 3 Precision estimates derived from a mixed effects regression model of a spike-in experiment of different patient assays (n=3) for different random factors.

nominal MRD	mean MRD	SD Lot	SD Operator	SD Day	SD Repeat	SD_{total}	CV_{total} [%]
1.00E-02	1.08E-02	3.10E-07	1.93E-03	1.47E-07	3.64E-04	3.97E-03	36.79
1.00E-03	1.13E-03	6.49E-08	2.17E-04	6.56E-05	1.04E-09	5.23E-04	46.32
3.20E-04	3.58E-04	1.97E-08	6.39E-05	2.39E-05	1.10E-11	1.70E-04	47.44
1.00E-04	1.04E-04	7.12E-09	1.78E-05	1.11E-05	1.85E-11	6.33E-05	60.91
3.20E-05	4.02E-05	6.27E-09	9.02E-06	9.05E-06	3.89E-12	4.29E-05	106.8
1.00E-05	2.41E-05	4.58E-06	7.46E-06	8.92E-06	1.97E-09	4.42E-05	183.4

Assays have been performed using different reagent lots (n=2) by different operators (n=4, 2 at each laboratory), and repeated on different days (n=3) including technical replicates (n=3) per analysis to evaluate influence on the variance of MRD measurements.

the short-term stability of medical samples based on the DNA quality of >17,000 samples received for MRD analysis, a significant drop in DNA quality is seen after 6-7 days after sample collection (Figure 3A), but 89.4% of samples are received within 3 days. Additionally, no significant MRD drift upon storage of initial whole blood samples of 3 CLL patients for up to 5 or 6 days on ambient temperature was observed (Figure 3B). A similar experiment with bone marrow samples from either ALL or CLL patients was not performed, as the sample amount required to provide enough material for repeated DNA extractions and subsequent analysis in general exceeds the available surplus of samples from daily routine.

Results for the short-term stability of reagents subjected to repeated freeze-thaw cycles can be found in the Supplementary Material (Supplementary Figure 2). Laboratory supplied reagents kept under quality control are not affected by freeze-thawing during short time periods and seem to be highly stable for up to at least ~3-4 years. From the standard dilution series of 10 CLL assays, 0.04% (1/264) of the repeated standard measurements in the range of 10^{-1} to 10^{-4} showed a $\Delta Cp \geq \pm 1.5$ compared to the mean Cp determined during first analysis using the same DNA dilution and the same primer stock solution (Figure 3C). For ALL this ratio was 1.3% (6/458) using the standards from 24 assays of 14 patients. Signal drift over time, albeit not statistically significant ($p>0.05$), was observed for most standards. Although 10-15% of the standards had a significant signal drift (6/60 CLL; 7/48 ALL), it was observed, that only single standards from individual assays but no complete dilution series were affected.

3.6 Effects of DNA extraction methods and inhibitory effects of endogenous substances on ASO-PCR MRD measurements are minor

The recovery of MRD values from clinical samples was determined from spike-in samples at MRD levels of 10^{-2}, 10^{-3} and $2x10^{-4}$. The recovered mean MRD values were twice as high as expected (Supplementary Table 3, automated DNA extraction using Qiasymphony). The constant overestimation of MRD values in the samples is probably derived from a systematic error introduced by the use of frozen CLL cells for the preparation of the spike-ins, but may also be related to deviations in the flow cytometric CLL count used to prepare the spike-in samples. An underestimation of MRD has already been reported from a comparison of median MRD values obtained by flow cytometric and RQ-PCR based MRD detection (36), but dilution factors between the different spike-in levels indicate that the experimental design established the correct level differences compared to the highest CLL/PBMC ratio (nominal MRD value). Results for the influence of endogenous substances on the PCR amplification are summarized in Supplementary Table 4.

4 Discussion

ASO-PCR has so far been validated by means of ongoing round robin testing in international reference laboratories and the results of these and regular training meetings of the participating laboratories have been used by the EuroMRD Consortium to highly standardize the methodology to the best laboratory practice. The study reported here demonstrates how to use retrospective and prospective data to establish general technical performance parameters of the technique. During the preparation of the validation plan the lack of official international standards or any specific published guideline for the validation of patient-specific diagnostics was noted. Validation of a method is a formal requirement included in both the applicable international standards for the quality management of certified analytical laboratories, and the national as well as international legislations like e.g. the European IVDR (20). Neither publications by the scientific community (15, 16) nor applicable official guidelines provided satisfactory guidance to accept and qualify a patient-specific biomarker.

Additionally, because patient material is very limited for patient-specific diagnostics, it is mandatory for clinically relevant diagnostics like ASO-PCR to conclude on the type of experiments and number of samples or patients needed at an early time point in the validation process. The general performance parameters of ASO-PCR could be established using FDA recommendations resulting from the pro-active dialog between the authors and health authorities. Although, only runs previously accepted according to the EuroMRD criteria were included in the data set, the results showed that the validation method and the acceptance criteria for ASO-PCR evaluation as developed from the technical understanding of the methodology and the EuroMRD harmonization efforts (10, 11) are equally effective for the assessment of the assay parameters. Whereas

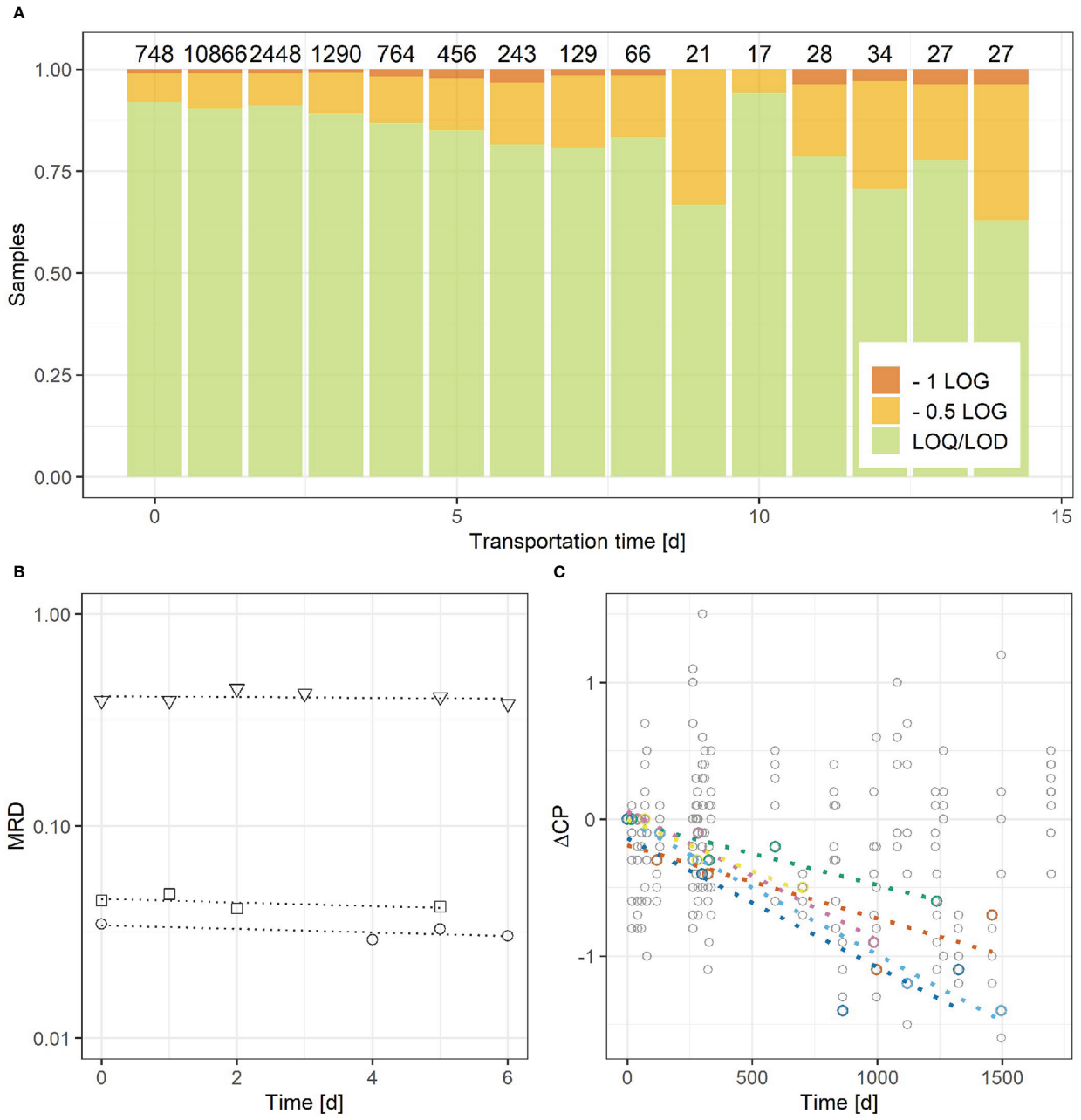


FIGURE 3
Assessment of short-term stability of medical samples **(A, B)** and long-term stability of reagents **(C)**. **(A)** Short-term stability was evaluated based on the need to adapt the analytic sensitivity of an MRD assessment due to a lower number of cell equivalents per ng of DNA in a test. If ≥50% of the required cell equivalents could be tested no adaptation of analytic performance was required (LOQ/LOD, green). LOQ and LOD is reduced by 1 LOG or 0.5 LOG level if <10% or <50% of the required cells would be tested, respectively. Cell equivalents were calculated from the copy number of the albumin gene determined by RQ-PCR according to (9). **(B)** PB was used to evaluate the potential drift of the MRD value from 3 CLL samples up to 6 days ambient storage after sample collection to prospectively assess short-term stability of medical samples. **(C)** Long-term stability was assessed from the Cp signal drift of a total of 60 primer and standard combinations from 10 CLL assays and during time periods of up to 4 years (only standards with significant Cp-drift are colored).

EuroMRD criteria are applied to an individual assay, the results described here enabled to establish the general performance of this diagnostic method across different patients, entities and clinical time points. Still, there are parameters that need to be evaluated and validated on an individual patient's basis, like the primer specificity and specific LOD. Inter- and intra-patient variability has also been discussed as limitation for the molecular gene fusion *BCR::ABL* as a surrogate endpoint in clinical trials of CML (37). As no patient specific reagents must be provided for the detection of *BCR::ABL*, this method performs at a different level than ASO-PCR. The FDA premarket clearance of a MRD detection kit for CML in 2016 pushed the method from a laboratory developed test to a standardized IVD kit. The validation of ASO-PCR to IVD standards has not been the focus of this work, but results reported here may be used to adapt the routine evaluation practice for ASO-PCR. E.g. acceptance or rejection of standards or samples which is currently evaluated by the Cp

differences of the technical replicates could be assessed based on the precision of the individual standard or sample measurements as in common procedures of clinical biochemistry (38).

One limitation of our study is that even though we did address the effect of reagent stability on MRD measurements, it has not been possible to precisely evaluate the influence of long term storage of the laboratory supplied reagents using identical reagent lots for the time span needed for a clinical study. Shelf life and turnover times of purchased reagents were too short to cover repeated analysis within 3 years using the same reagent lot, and the amount of patient material was limited. The slow, although not significant increase of the Cp observed for most of the standards could result from the variability caused by the use of different reagent lots over the required time span or by other causes that affect the stability of the reagents over time such as freeze-thawing or chemical processes occurring upon prolonged storage. However, most of these causes would affect standards, samples and albumin measurements in the same way, and should therefore have only minor influence on the MRD results.

First evidence has been published showing that additional MRD scoring at concentrations $<10^{-4}$ would show a benefit for prediction of progression free or overall survival in patients with ALL (1, 39, 40) and CLL (41, 42). Nevertheless, a difference in relapse-free survival of patients being clearly MRD negative compared to patients being MRD positive $\leq 10^{-4}$ has already been observed at early treatment time points using ASO-PCR (43, 44) and flow cytometry (2) for childhood and adult ALL. Low level positivity in ASO-PCR is strongly influenced by the data interpretation using the PCR signals of polyclonal healthy control samples. An investigation of the specificity of the reaction products from ASO-PCR of healthy control samples reported 30-40% (IG) and even up to 90% (TR) nonspecific positive signals, which can affect 10-65% of the measured samples depending on the time point the sample was collected (32). Fronkova et al. reported, that 15-20% of clinical samples can give nonspecific positive RQ-PCR signals when tested with a non-patient-specific assay (44), while NGS confirmed low-level positive ASO-PCR results in 61% of cases at treatment week 16 (40). Additionally, the definition of MRD negativity is for both reasons, specificity and sensitivity, under continuous discussion, especially for those cases, where the MRD evaluation has been implemented in the patient care and is no longer merely a research tool (12). The pros and cons of ASO-PCR and flow cytometry to assess MRD levels in CLL patients within the boundaries of clinical studies but also in clinical routine have also been extensively discussed (45–47) and the necessary regulatory requirements were included in the European Medicines Agency guideline on the evaluation of anticancer medicinal products in man effective since July 2016 (48).

As Wendtner (49) pointed out in a comment to the article by Thompson et al. (41) it needs sophisticated standards to provide reliable MRD data at the currently aimed sensitivity level of one in a million cells, independently of the investigated entity. Multicolor-flow cytometry or next generation sequencing being equally or more sensitive or less time consuming than ASO-PCR for MRD assessment, are currently either on the technical level or on the clinical level not validated and standardized to the same degree as ASO-PCR. Especially the NGS methodology which is strongly pushed to diagnostic use in the field of hematology and precision medicine (35, 41, 42, 50–52) and could be used to quantify MRD without the use of personalized standards and reagents, requires profound validation and experience from clinical trials to circumvent false negative results and to deliver valuable disease prognosis. This is also underlined by the accreditation requirements of the laboratory standard released from the College of American Pathologists for NGS-based clinical tests (53), which does not include requirements for quantitative aspects of molecular oncology and emphasizes that a consensus on practice must be built from professional experience.

In summary, our results demonstrate that comprehensive technical performance level validation according to regulatory requirements can also be achieved for the patient specific ASO-PCR approach.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by CLL11: University Cologne Ethical Review Committee, Medical Faculty University Cologne, Cologne, Germany CLL14: University Cologne Ethical Review Committee, Medical Faculty University Cologne, Cologne, Germany GMALL07/03: Goethe University Frankfurt Research Ethics Board, Medical Faculty, Johann Wolfgang Goethe-University, Frankfurt/Main, Germany GMALL08/13: Goethe University Frankfurt Research Ethics Board, Medical Faculty, Johann Wolfgang Goethe-University, Frankfurt/Main, Germany. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Research Design AS, JR, AL, JD, MR. Formal analysis, investigation and data curation AS, MS, MN, DK, AL, VV, EG-M, JS, SB, MR. Resources: MB, MK, JD. Writing - original draft preparation: AS. Writing - review and editing: ID, MS, MR, MB, SB, MK, MN, VV, AL, JD, DK, JR, EG-M, JS. All authors have read and agreed to the published version of the manuscript.

Funding

Parts of this work was supported by the Deutsche José Carreras Leukämie-Stiftung (grants DJCLS R 15/11 to MB and DJCLS 06R/2019 to MS and MB). MN was supported by the National Institute for Cancer Research Program EXCELES, ID Project No. LX22NPO5102) funded by the European Union - Next Generation EU.

Acknowledgments

The authors are indebted to the GMALL Trial Center (R. Reutzel, C. Fuchs, N. Gökbuget) and participating hospitals for patient

recruitment, care and logistics and to the members of the Hematology Lab Kiel and the Laboratory Medical Immunology Rotterdam for their excellent technical support. We also kindly acknowledge the research funding to Erasmus MC and UKSH Kiel for parts of the work during the application of the method as drug development tool at the FDA by F. Hoffmann-La Roche, and the contribution of Yuda Zhu (formerly Genentech team member) to the data analysis.

Conflict of interest

JR, DK: employee of Genentech. MB: Honoraria, Speakers Bureau Amgen, Celgene, Janssen, Pfizer; Membership on an entity's Board of Directors or advisory committees Amgen, Janssen; Research Funding Affimed, Amgen, Celgene, Regeneron; Consultancy Amgen, Incyte, PRMA. JD: Founding chairman of EuroMRD 2001-2016, current Board member of EuroMRD and Founding Chairman of EuroFlow 2006-now; Educational Services Agreement from BD Biosciences San José, CA and Scientific Advisor Agreement with Cytognos Salamanca, ES; all related fees and honoraria are for dept. IHB at Leiden University Medical Center LUMC. AL: Research Funding Roche, Genentech; Advisory Committee AbbVie, Speakers Bureau Gilead, Janssen. VV: Research Funding Agilent, BD Biosciences, Pfizer, Navigate, Janssen, Speakers Bureau Amgen. SB: Research funding AbbVie, Celgene, Janssen, and Roche, consultancy or advisory role AbbVie, AstraZeneca, Celgene, Roche, Amgen, and Janssen. MR: Honoraria AbbVie, Roche, Membership on an entity's Board of Directors or advisory committees AbbVie, Roche, BMS, Research Funding AbbVie, Roche; Other: travel support, Research.

The remaining 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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1111209/full#supplementary-material

References

1. Van Dongen JJM, Seriu T, Panzer-Grümayer ER, Biondi A, Pongers-Willemse MJ, Corral L, et al. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet* (1998) 352(9142):1731–8. doi: 10.1016/S0140-6736(98)04058-6

2. Borowitz MJ, Wood BL, Devidas M, Loh ML, Raetz EA, Salzer WL, et al. Prognostic significance of minimal residual disease in high risk b-ALL: A report from children's oncology group study AALL0232. *Blood.* (2015) 126(8):964–71. doi: 10.1182/blood-2015-03-633685

3. Böttcher S, Ritgen M, Fischer K, Stilgenbauer S, Busch RM, Fingerle-Rowson G, et al. Minimal residual disease quantification is an independent predictor of progression-free and overall survival in chronic lymphocytic leukemia: A multivariate analysis from the randomized GCLLSG CLL8 trial. *J Clin Oncol* (2012) 30(9):980–8. doi: 10.1200/JCO.2011.36.9348

4. Langerak AW, Ritgen M, Goede V, Robrecht S, Bahlo J, Fischer K, et al. Prognostic value of MRD in CLL patients with comorbidities receiving chlorambucil plus obinutuzumab or rituximab. *Blood.* (2019) 133(5):494–7. doi: 10.1182/blood-2018-03-839688

5. Hallek M, Al-Sawaf O. Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures. *Am J Hematol* (2021) 96(12):1679–705. doi: 10.1002/ajh.26367

6. Hoelzer D, Bassan R, Dombret H, Fielding A, Ribera JM, Buske C, et al. Acute lymphoblastic leukaemia in adult patients: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* (2016) 27(April):v69–82. doi: 10.1093/annonc/mdw025

7. De Haas V, Ismaila N, Advani A, Arber DA, Dabney RS, Patel-Donelly D, et al. Initial diagnostic work-up of acute leukemia: ASCO clinical practice guideline endorsement of the college of American pathologists and American society of hematology guideline. *J Clin Oncol* (2019) 37(3):239–53. doi: 10.1200/JCO.18.01468

8. Heuser M, Freeman SD, Ossenkoppele GJ, Buccisano F, Hourigan CS, Ngai LL, et al. 2021 Update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD working party. *Blood* (2021) 138(26):2753–67. doi: 10.1182/blood.2021013626

9. Pongers-Willemse M, Verhagen O, Tibbe G, Wijkhuijs A, de Haas V, Roovers E, et al. Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes. *Leukemia.* (1998) 12:2006–14. doi: 10.1038/sj.leu.2401246

10. van der Velden VHJ, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJM. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: Principles, approaches, and laboratory aspects. *Leukemia.* (2003) 17(6):1013–34. doi: 10.1038/sj.leu.2402922

11. van der Velden VHJ, Cazzaniga G, Schrauder A, Hancock J, Bader P, Panzer-Grumayer ER, et al. Analysis of minimal residual disease by Ig/TCR gene rearrangements: Guidelines for interpretation of real-time quantitative PCR data. *Leukemia.* (2007) 21(4):604–11. doi: 10.1038/sj.leu.2404586

12. Van Dongen JJM, van der Velden VHJ, Brüggemann M, Orfao A. Minimal residual disease diagnostics in acute lymphoblastic leukemia: Need for sensitive, fast, and standardized technologies. *Blood.* (2015) 125(26):3996–4009. doi: 10.1182/blood-2015-03-580027

13. van der Velden VHJ, Panzer-Grümayer ER, Cazzaniga G, Flohr T, Sutton R, Schrauder A, et al. Optimization of PCR-based minimal residual disease diagnostics for childhood acute lymphoblastic leukemia in a multi-center setting. *Leukemia.* (2007) 21(4):706–13. doi: 10.1038/sj.leu.2404535

14. Burd EM. Validation of laboratory-developed molecular assays for infectious diseases. *Clin Microbiol Rev* (2010) 23(3):550–76. doi: 10.1128/CMR.00074-09

15. Mattocks CJ, Morris MA, Matthijs G, Swinnen E, Corveleyn A, Dequeker E, et al. A standardized framework for the validation and verification of clinical molecular genetic tests. *Eur J Hum Genet* (2010) 18(12):1276–88. doi: 10.1038/ejhg.2010.101

16. Raymaekers M, Smets R, Maes B, Cartuyvels R. Checklist for optimization and validation of real-time PCR assays. *J Clin Lab Anal* (2009) 23(3):145–51. doi: 10.1002/jcla.20307

17. Linnet K. Necessary sample size for method comparison studies based on regression analysis. *Clin Chem* (1999) 45(6 I):882–94. doi: 10.1093/clinchem/45.6.882

18. EMEA Committee for Medicinal Products in Human Use. *Guideline on bioanalytical method validation. EMEA/CHMP/EWP/192217/2009 rev. 1 corr. 2.* London: Euorpean Medicines Agency (2011) p. 1–23.

19. Booth BP, Simon WC. Analytical method validation. *New Drug Dev Regul Paradig Clin Pharmacol Biopharm.* (2016), 138–59.

20. European Commission. IVDR - regulation (EU) 2017/746 on in-vitro diagnostic medical devices. *Off J Eur Union.* (2017) 60(April 2014):1–175.

21. Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A, et al. Proposals for the immunological classification of acute leukemias. In: *European Group for the immunological characterization of leukemias (EGIL)*, vol. Vol. 9. . England: Leukemia (1995). p. 1783–6.

22. Van Der Velden VHJ, Van Dongen JJM. MRD detection in acute lymphoblastic leukemia patients using Ig/TCR gene rearrangements as targets for real-time quantitative PCR. *Methods Mol Biol* (2009) 538:115–50. doi: 10.1007/978-1-59745-418-6_7

23. Flohr T, Schrauder A, Cazzaniga G, Panzer-Grümayer R, van der Velden V, Fischer S, et al. Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements

in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia.* (2008) 22(4):771–82. doi: 10.1038/leu.2008.5

24. Cazzaniga G, Songia S, Biondi A. Pcr technology to identify minimal residual disease. *Methods Mol Biol* (2021) 2185:77–94. doi: 10.1007/978-1-0716-0810-4_6

25. CLSI. *Evaluation of stability of in vitro diagnostic reagents; approved guideline. CLSI document EP25-a. clinical and laboratory standards institute*. Wayne, PA, USA: Clinical and Laboratory Standards Institute (2009).

26. CLSI. *User protocol for evaluation of qualitative test performance; approved guideline–second edition. CLSI document EP12-A2. CLSI/NCCLS document EP12-A2, approved guideline. 2nd Edition*. Wayne, PA, USA: Clinical and Laboratory Standards Institute (2008).

27. CLSI. *Evaluation of precision of quantitative measurement procedures; approved guideline–third edition. CLSI document EP05-A3*. Wayne, PA, USA: Clinical and Laboratory Standards Institute (2014).

28. NCCLS. *Evaluation of the linearity of quantitative measurement procedures: A statistical approach, approved guideline. NCCLS document EP6-a*. Wayne, PA, USA: National Committee for Clinical Laboratory Standards (2003).

29. NCCLS. *Protocols for determination of limits of detection and limits of quantitation; approved guideline. NCCLS document EP17-a*. Wayne, PA, USA: National Committee for Clinical Laboratory Standards (2004).

30. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (2021). Available at: https://www.r-project.org/.

31. RStudio Team. *RStudio: Integrated development for r*. Boston, MA: RStudio, PBC (2020). Available at: http://www.rstudio.com/.

32. van der Velden VHJ, Wijkhuijs JM, van Dongen JJM. Non-specific amplification of patient-specific Ig/TCR gene rearrangements depends on the time point during therapy: Implications for minimal residual disease monitoring. *Leukemia.* (2008) 22(3):641–4. doi: 10.1038/sj.leu.2404925

33. Stamatopoulos K, Belessi C, Moreno C, Boudjograh 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. doi: 10.1182/blood-2006-03-012948

34. Bruggemann M, Schrauder A, Raff T, Pfeifer H, Dworzak M, Ottmann OG, et al. Standardized MRD quantification in European all trials: Proceedings of the second international symposium on MRD assessment in Kiel, Germany, 18-20 September 2008. *Leukemia.* (2010) 24(3):521–35. doi: 10.1038/leu.2009.268

35. Svaton M, Skotnicova A, Reznickova L, Rennerova A, Valova T, Kotrova M, et al. NGS-based MRD quantitation: An alternative to qPCR validated on a Large consecutive cohort of children with ALL. *Blood.* (2021) 138(Supplement 1):1314–4. doi: 10.1182/blood-2021-152971

36. Böttcher S, Stilgenbauer S, Busch R, Brüggemann M, Raff T, Pott C, et al. Standardized MRD flow and ASO IGH RQ-PCR for MRD quantification in CLL patients after rituximab-containing immunochemotherapy: A comparative analysis. *Leukemia.* (2009) 23(11):2007–17. doi: 10.1038/leu.2009.140

37. Guilhot J, Preudhomme C, Mahon FX, Guilhot F. Analyzing molecular response in chronic myeloid leukemia clinical trials: Pitfalls and golden rules. *Cancer.* (2015) 121 (4):490–7. doi: 10.1002/cncr.29053

38. Thompson M, Wood R. Harmonized guidelines for internal quality control in analytical chemistry laboratories. *Pure Appl Chem* (1995) 67(4):649–66. doi: 10.1351/pac199567040649

39. Kotrova M, Muzikova K, Mejstrikova E, Novakova M, Bakardjieva-Mihaylova V, Fiser K, et al. The predictive strength of next-generation sequencing MRD detection for relapse compared with current methods in childhood ALL. *Blood.* (2015) 126:1045–7. doi: 10.1182/blood-2015-07-655159

40. Kotrová M, Koopmann J, Trautmann H, Alakel N, Beck J, Nachtkamp K, et al. Prognostic value of low-level MRD in adult acute lymphoblastic leukemia detected by low- and high-throughput methods. *Blood Adv* (2022) 6(10):3006–10. doi: 10.1182/bloodadvances.2021006727

41. Thompson PA, Srivastava J, Peterson C, Strati P, Jorgensen JL, Hether T, et al. Minimal residual disease undetectable by next-generation sequencing predicts improved outcome in CLL after chemoimmunotherapy. *Blood.* (2019) 134(22):1951–9. doi: 10.1182/blood.2019001077

42. Hengeveld PJ, van der Klift MY, Kolijn PM, Davi F, Kavelaars FG, de Jonge E, et al. Detecting measurable residual disease beyond 10-4 through an IGHV leader-based NGS approach improves prognostic stratification in CLL. *Blood.* (2022) blood.2022017411. doi: 10.1182/blood.2022017411

43. Raff T, Gökbuget N, Lüschen S, Reutzel R, Ritgen M, Irmer S, et al. Molecular relapse in adult standard-risk ALL patients detected by prospective MRD monitoring during and after maintenance treatment: Data from the GMALL 06/99 and 07/03 trials. *Blood.* (2007) 109(3):910–5. doi: 10.1182/blood-2006-07-037093

44. Fronkova E, Muzikova K, Mejstrikova E, Kovac M, Formankova R, Sedlacek P, et al. B-cell reconstitution after allogeneic SCT impairs minimal residual disease monitoring in children with ALL. *Bone Marrow Transplant.* (2008) 42(3):187–96. doi: 10.1038/bmt.2008.122

45. Uhrmacher S, Erdfelder F, Kreuzer KA. Flow cytometry and polymerase chain reaction-based analyses of minimal residual disease in chronic lymphocytic leukemia. *Adv Hematol* (2010) 2010:272517. doi: 10.1155/2010/272517

46. Thompson PA, Wierda WG. Eliminating minimal residual disease as a therapeutic end point: Working toward cure for patients with CLL. *Blood.* (2016) 127(3):279–86. doi: 10.1182/blood-2015-08-634816

47. Böttcher S, Hallek M, Ritgen M, Kneba M. The role of minimal residual disease measurements in the therapy for CLL: is it ready for prime time? *Hematol Oncol Clin North Am* (2013) 27(2):267–88. doi: 10.1016/j.hoc.2013.01.005

48. European Medicines Agency. *Appendix 4 to the guideline on the evaluation of anticancer medicinal products in man*. London, UK: European Medicines Agency (2015).

49. Wendtner CM. CLL: Deep dive for residual cells by NGS matters. *Blood.* (2019) 134(22):1883–4. doi: 10.1182/blood.2019003244

50. Medina A, Puig N, Flores-Montero J, Jimenez C, Sarasquete ME, Garcia-Alvarez M, et al. Comparison of next-generation sequencing (NGS) and next-generation flow (NGF) for minimal residual disease (MRD) assessment in multiple myeloma. *Blood Cancer J* (2020) 10(10):108. doi: 10.1038/s41408-020-00377-0

51. Alonso CM, Llop M, Sargas C, Pedrola L, Panadero J, Hervás D, et al. Clinical utility of a next-generation sequencing panel for acute myeloid leukemia diagnostics. *J Mol Diagnostics* (2019) 21(2):228–40. doi: 10.1016/j.jmoldx.2018.09.009

52. Galimberti S, Genuardi E, Mazziotta F, Iovino L, Morabito F, Grassi S, et al. The minimal residual disease in non-hodgkin's lymphomas: From the laboratory to the clinical practice. *Front Oncol* (2019) 9:1–15. doi: 10.3389/fonc.2019.00528

53. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, et al. College of American pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch Pathol Lab Med* (2015) 139(4):481–93. doi: 10.5858/arpa.2014-0250-CP

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fonc.2023.1097942

OPEN ACCESS

EDITED BY
Silvia Jiménez-Morales,
National Institute of Genomic Medicine (INMEGEN), Mexico

REVIEWED BY
Deepshi Thakral,
All India Institute of Medical Sciences, India
Adolfo Martinez,
General Hospital of Mexico, Mexico

*CORRESPONDENCE
Anastasia Chatzidimitriou
achatzidimitriou@certh.gr

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

RECEIVED 14 November 2022
ACCEPTED 17 January 2023
PUBLISHED 01 February 2023

CITATION
Vlachonikola E, Pechlivanis N, Karakatsoulis G, Sofou E, Gkoliou G, Jeromin S, Stavroyianni N, Ranghetti P, Scarfo L, Österholm C, Mansouri L, Notopoulou S, Siorenta A, Anagnostopoulos A, Ghia P, Haferlach C, Rosenquist R, Psomopoulos F, Kouvatsi A, Baliakas P, Stamatopoulos K and Chatzidimitriou A (2023) T cell receptor gene repertoire profiles in subgroups of patients with chronic lymphocytic leukemia bearing distinct genomic aberrations. *Front. Oncol.* 13:1097942. doi: 10.3389/fonc.2023.1097942

T cell receptor gene repertoire profiles in subgroups of patients with chronic lymphocytic leukemia bearing distinct genomic aberrations

Elisavet Vlachonikola[1,2], Nikolaos Pechlivanis[1,2], Georgios Karakatsoulis[1,3], Electra Sofou[1,4], Glykeria Gkoliou[1,5], Sabine Jeromin[6], Niki Stavroyianni[7], Pamela Ranghetti[8], Lydia Scarfo[8], Cecilia Österholm[9], Larry Mansouri[9], Sofia Notopoulou[1], Alexandra Siorenta[10], Achilles Anagnostopoulos[7], Paolo Ghia[8], Claudia Haferlach[6], Richard Rosenquist[9,11], Fotis Psomopoulos[1], Anastasia Kouvatsi[2], Panagiotis Baliakas[12], Kostas Stamatopoulos[1,9] and Anastasia Chatzidimitriou[1,9]*

[1]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, [2]Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle, University of Thessaloniki, Thessaloniki, Greece, [3]Department of Mathematics, School of Sciences, University of Ioannina, Ioannina, Greece, [4]Laboratory of Biological Chemistry, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece, [5]Department of Medicine, Democritus University of Thrace, Alexandroupolis, Greece, [6]MLL - Munich Leukemia Laboratory, Munich, Germany, [7]Hematology Department and Hematopoietic Cell Transplantation (HCT) Unit, G. Papanicolaou Hospital, Thessaloniki, Greece, [8]Division of Experimental Oncology, Università Vita-Salute San Raffaele and Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Ospedale San Raffaele, Milan, Italy, [9]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, [10]Immunology Department and National Tissue Typing Center, General Hospital of Athens "G. Gennimatas", Athens, Greece, [11]Clinical Genetics, Karolinska University Hospital, Solna, Sweden, [12]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Background: Microenvironmental interactions of the malignant clone with T cells are critical throughout the natural history of chronic lymphocytic leukemia (CLL). Indeed, clonal expansions of T cells and shared clonotypes exist between different CLL patients, strongly implying clonal selection by antigens. Moreover, immunogenic neoepitopes have been isolated from the clonotypic B cell receptor immunoglobulin sequences, offering a rationale for immunotherapeutic approaches. Here, we interrogated the T cell receptor (TR) gene repertoire of CLL patients with different genomic aberration profiles aiming to identify unique signatures that would point towards an additional source of immunogenic neoepitopes for T cells.

Experimental design: TR gene repertoire profiling using next generation sequencing in groups of patients with CLL carrying one of the following copy-number aberrations (CNAs): del(11q), del(17p), del(13q), trisomy 12, or gene mutations in *TP53* or *NOTCH1*.

Results: Oligoclonal expansions were found in all patients with distinct recurrent genomic aberrations; these were more pronounced in cases bearing CNAs, particularly trisomy 12, rather than gene mutations. Shared clonotypes were found both within and across groups, which appeared to be CLL-biased based on extensive comparisons against TR databases from various entities. Moreover, *in silico* analysis identified TR clonotypes with high binding affinity to neoepitopes predicted to arise from *TP53* and *NOTCH1* mutations.

Conclusions: Distinct TR repertoire profiles were identified in groups of patients with CLL bearing different genomic aberrations, alluding to distinct selection processes. Abnormal protein expression and gene dosage effects associated with recurrent genomic aberrations likely represent a relevant source of CLL-specific selecting antigens.

KEYWORDS

T cell receptor (TR) gene repertoire, chronic lymphocytic leukemia (CLL), recurrent genomic aberrations, neoepitopes, T cell based immunotherapies

Introduction

In recent years, novel agents targeting critical processes and pathways such as microenvironmental interactions (e.g. BTK inhibitors) and apoptotic cell death (e.g. BCL-2 inhibitors) have revolutionized the management of chronic lymphocytic leukemia (CLL), leading to high overall response rates (often exceeding 90%) and superior progression-free and, at least in some trials, overall survival compared to classic chemoimmunotherapy (1). Nevertheless, relapses still occur and CLL remains an incurable malignancy, highlighting the need for alternative therapeutic approaches aiming at more effective disease control (2). Immunotherapy that engages immune responses against targets on the malignant cells could, at least in principle, represent such an approach also for CLL. However, despite initial promising results, it has become apparent that not all patients with CLL benefit from (different forms of) immunotherapy, highlighting the need for further research into the underlying mechanisms, particularly focusing on T cells (3, 4).

Multiple lines of evidence support that T cells present in the tumor microenvironment (TME) are implicated in the natural history of CLL, albeit in rather contrasting ways (5). On the one hand, T cells provide trophic signals for CLL cell survival. This is evidenced by the fact that transfer of autologous activated T cells is required for the successful engraftment of CLL cells in murine models; and that efficient proliferation of CLL cells was observed after the engraftment of $CD4^+$ T cells bearing a T cell receptor (TR) with CLL-unrelated specificity (6–8). On the other hand, CLL cells induce various changes in the T cell compartment, leading to functional exhaustion as a consequence of persistent antigenic stimulation, which underlies tumor evasion and immune suppression (9–12).

Molecular studies by us and others revealed skewing of the TR repertoire and clonal expansions of T cells in CLL, supporting antigen drive (13–15). Moreover, different patients were found to share TR clonotypes, many of which were 'CLL-biased' i.e. not present in other settings (16, 17). Altogether, these findings strongly imply ongoing antigenic triggering in a CLL-specific context (16, 17). The cognate antigens remain to be determined and could arguably be those selecting the malignant clone or unrelated ones, including tumor-derived antigens.

Particularly regarding the latter, we have previously demonstrated that immunogenic epitopes can be isolated from the heavy complementarity-determining region 3 (VH CDR3) sequence of both human and murine CLL bearing stereotyped [i.e. (quasi) identical] B cell receptors (BcR), efficiently processed and presented by CLL cells and effectively recognized by specific T cells (18). Of note, immunization of Eμ-TCL1 CLL mice model reduced leukemia development and increased overall survival of the animals (18). Subsequently, we provided evidence that the targeted somatic hypermutation operating on the BcR immunoglobulin (BcR IG) repertoire may produce idiotypic targets for the cognate T cells (19).

Taken together, the clonotypic BcR IG expressed by CLL cells can be envisioned as a source of neoepitopes selecting T cells in the TME. That said, recurrent genomic aberrations associated with distinct abnormal expression profiles could represent an alternative, non-mutually exclusive, pool of such potent immunogenic neoepitopes. Judging from a great variety of other malignancies shown to harbor reactive T cells against epitopes arisen from recurrent genomic aberrations, this may prove clinically relevant, especially if one considers the promising results of immunotherapy in e.g. patients with melanoma or lung cancer exhibiting a high mutational load (20–22).

On these grounds, here we interrogated the TR gene repertoire in groups of patients with CLL carrying specific single genomic aberrations, one of del(11q), del(17p), del(13q), trisomy 12, *TP53* or *NOTCH1* mutations. We report oligoclonality in all groups of patients with distinct recurrent genomic aberrations, albeit more pronounced in cases bearing copy number aberrations (CNAs), particularly trisomy 12. This, combined with the existence of group-specific clonotypes, suggest that abnormal protein expression and gene dosage effects can lead to the emergence of CLL-specific selecting epitopes that likely restrict the TR gene repertoire.

Materials and methods

Patient group

The study cohort included 44 patients with CLL from 4 different centers in Thessaloniki, Greece; Milan, Italy; Munich, Germany; and Stockholm, Sweden. The local Ethics Review Committees approved the study, while written informed consent was obtained from all study participants, in accordance with the Declaration of Helsinki.

Patients were selected on the basis of carrying one of the following genomic aberrations: del(11q) (n=10), del(13q) (n=7), trisomy 12 (n=17), *NOTCH1* mutation (n=5) or *TP53* mutation (n=5). The possible confounding effects of multiple aberrations were minimized, as we previously established that the analyzed patients carried only one of the aforementioned aberrations through comprehensive genomic characterization (including FISH, SNP arrays, gene panels and WES; for methodological details, see Supplementary Methods). Demographic and clinicobiological characteristics are provided in Supplementary Table S1. In all patients, samples were collected before the initiation of any treatment and there was no evidence of infection (either signs or symptoms), or recent vaccination at sampling that could bias the results.

Next-generation sequencing: Library preparation, analysis and interpretation

T cell receptor beta chain (TRB) gene rearrangements were amplified on either genomic DNA or complementary DNA (cDNA) isolated from peripheral blood mononuclear cells (PBMCs). In all samples, the starting absolute PBMC count was set at $5x10^6$ cells in order to determine actual repertoire diversity while avoiding to over-amplify the same TRBV-TRBD-TRBJ gene rearrangements, hence further normalizing the different samples. PCR amplification of the TRBV-TRBD-TRBJ gene rearrangements, library construction and next generation sequencing (NGS) were performed as previously described (23). A paired-end sequencing protocol was followed in order to achieve double coverage in the TRB complementarity-determining region 3 (TRB CDR3) for each amplicon, thus increasing the accuracy of the results (for methodological details, see Supplementary Methods).

Paired-end read merging and strict length/quality filtering was performed by a purpose-built bioinformatics pipeline, as previously described; details are provided in Supplementary Methods (17, 24). Annotation of the TRBV-TRBD-TRBJ gene rearrangements was carried out by the IMGT/HighV-QUEST tool (25). Meta-data analysis was performed using the T cell Receptor/Immunoglobulin Profiler (TRIP) tool (26).

For the interpretation of the NGS results, the term "clonotypes" as used in this study refers to TRBV-TRBD-TRBJ gene rearrangements with unique pairs of TRBV genes and identical TRB CDR3 amino acid (aa) sequences within a sample. For the computation of clonotypes, we assessed only productive TRBV-TRBD-TRBJ gene rearrangements. Rearrangements carrying TRBV genes with <95% germline identity were discarded, being considered as sequences with unacceptable error rate. Details of the overall metrics regarding the NGS data are given in Supplementary Table S2.

The 10 clonotypes with the highest frequency within a sample are herein referred to as "major". The relative frequency of each clonotype/sample was calculated as the number of TRBV-TRBD-TRBJ gene rearrangements corresponding to this particular clonotype divided by the total number of productive, filtered-in TRBV-TRBD-TRBJ gene rearrangements of the sample.

To overcome empirical assumptions regarding the fraction of clonal expansions that was represented at a meaningful frequency, hence potentially claiming biological relevance, "significantly expanded clonotypes" were further defined based on a data-driven statistical analysis. The frequency distribution of clonotypes from all samples were transformed to z-scores; clonotypes with z-scores greater that 2 were selected. The minimum frequency of those clonotypes was set as the per sample threshold. The median value of all individual thresholds led to the identification of 0.216% as the discerning frequency above which a given clonotype would be considered as significantly expanded (27). A table listing the number of significantly expanded clonotypes per sample in each group is provided in Supplementary Table S3.

Repertoire diversity and clonality

In order to characterize the complexity of the repertoire in each group, two different metrics were estimated: (i) repertoire diversity, describing the number of the different clonotypes present in each sample, and (ii) repertoire clonality, referring to skewed clone size distribution in a given sample or group of samples.

Regarding the former, repertoire diversity was described by Hill numbers (^{1}D) using Recon for the calculations (28, 29). This algorithm is based on a modified maximum-likelihood method that calculates diversity not only by counting the different number of clonotypes identified per sample but also by evaluating the clonotype-size distribution, and further reconstructs the overall diversity of the initial samples. To assess the sensitivity of the calculated diversity value comparing the rare versus the abundant clonotypes, ^{1}D numbers were considered for the analysis, meaning that each clonotype was exactly weighted by its proportional abundance (30). Through this approach, sampling biases, experimental errors, as well as the variable number of T cells that were captured in different blood samples can be normalized (29). Moreover, comparability of the results between different samples was reinforced through utilizing identical numbers of isolated cells as starting material for nucleic acid isolation, identical quantity of genomic DNA or cDNA for PCR amplification of TRBV-TRBD-TRBJ gene rearrangements, and similar number of sequencing reads for downstream bioinformatics analysis.

Regarding the assessment of repertoire clonality, the following values were determined: (i) the cumulative frequency (CF-10) of the major (top-10) clonotypes within a given sample; (ii) the median CF-10 value of all the samples within a given group (MCF-10) (17); (ii) the cumulative frequency of the significantly expanded clonotypes (CFEx) per sample; and (iv) the median CFEx value of all the samples within a given group (ex-MCF).

TRBV gene repertoire

In order to evaluate the relative frequency of each TRB gene that partakes in the clonotype formation, clonotypes rather than individual TRBV-TRBD-TRBJ gene rearrangements were considered. In more detail, TRBV gene frequencies within a sample were calculated as the number of clonotypes using particular TRBV genes over the total number of clonotypes. For comparisons of the TRBV gene repertoire between the significantly expanded clonotypes and the remaining polyclonal background, an approach similar to differential expression analysis was implemented using *limma*, an R/Bioconductor software package (31).

Clonotype comparisons

Extensive comparisons of the clonotypes between patients within a given group have been undertaken in order to identify common clonotypes in patients sharing the same genomic aberration ("group-specific" clonotypes). Furthermore, in order to identify common clonotypes regardless of the background of genomic aberrations ("public" clonotypes) comparisons between patients in different groups have been also pursued. Finally, cross comparisons have been performed against: (i) a well-annotated database of TR clonotypes with known antigen specificities (n=47,107 TR clonotypes), and (ii) TR clonotypes from our published NGS studies in monoclonal B-cell lymphocytosis (a potential precursor state to CLL), chronic idiopathic neutropenia and benign ethnic neutropenia (n=1,033,310) (32–34).

Prediction of neoepitopes and search for neoepitope-specific clonotypes

In the case of *NOTCH1* and *TP53* mutations, each mutant nucleotide sequence was translated *in silico* into a protein sequence that was used as input for the TAP 1.0 tool (35). Based on the prediction model implemented by the tool, each examined epitope was characterized either as a tumor or non-tumor antigen with a given probability score. This score was estimated by TAP 1.0 based on the selected datasets of human tumor and non-tumor antigens that were used for the development and training of the tool (35).

For the prediction of possible neoepitopes deriving from the mutant p53 and Notch1 proteins in samples of the studied cohort, a region of 15aa length stretching in both sides of each particular mutated position was dissected into k-mers whose immunogenicity was calculated by TAP 1.0. In cases where an insertion/deletion occurred and led to a frameshift, the complete altered aa sequence downstream of this position was used for epitope prediction. All k-mers that were characterized as tumor antigens by TAP 1.0 with high probability score (over 0.9 in order to ensure stringency), were selected ('filtered-in') for further analysis. The lists of the filtered-in tumor-derived epitopes for each case are given in Supplementary Table S4.

Each group of possible tumor epitopes produced from a particular sample was used as input for ERGO-II (pEptide tcR matchinG predictiOn), a highly specific and generic TR-peptide binding predictor, allowing the *in silico* identification of specific TR clonotypes with high affinity for these epitopes (36). To that purpose, all the TR clonotypes of a given sample were checked for their binding affinity against all the respective filtered-in tumor-derived epitopes. The complete list of the MHC alleles (Supplementary Table S5) expressed in each sample was used as an additional feature in ERGO-II in order to refine the predicted interactions. Pairs of clonotypes and epitopes - MHC alleles with high binding affinity greater than 0.8, were further analyzed.

HLA genotyping

Typing of the HLA-A, -B, -C (low resolution) and -DRB1 (allelic level high resolution determination) loci was performed by reverse PCR-SSOP (sequence specific oligonucleotide probe) using a commercially available kit (LABtype ® RSSO, One Lambda, Thermo Fisher, CA, USA).

Statistical analysis

Descriptive statistics (median, mean, min, max) were computed to characterize gene counts and clonotype frequency distribution. The non-parametric Kruskal Wallis and Mann–Whitney tests were used in order to identify differences between the studied variables at the distinct groups, while ANOVA was applied to compare diversity distributions amongst the different groups. Correction for multiple comparisons was obtained using the Benjamini-Hochberg *post hoc* test and the significance level was set at $\alpha = 0.05$. All statistical analyses were performed with the statistical Package GraphPad Prism version 8 (GraphPad Software, Inc., San Diego, CA, USA) and R version 4.1.3.

Results

Distinct TR gene repertoire profiles in different groups of patients with CLL with distinct genomic aberrations

A total of 12,607,219 sequences were obtained by the sequencing experiments (median: 293,122/sample). After strict quality filtering, synthesis of the paired-end sequencing reads and annotation with the IMGT/HighV-QUEST tool, we finally evaluated 8,989,297 high-quality sequences corresponding to productive TRBV-TRBD-TRBJ rearrangements (median: 185,420/sample). Overall, we identified 465,401 distinct clonotypes with a median of 10,608 clonotypes/sample (Supplementary Table S2).

In general, cases carrying CNAs tended to display less diverse TR repertoires than cases carrying mutations in the *NOTCH1* and *TP53* genes, though not reaching statistical significance (α=0.06) likely due to the relatively small numbers of cases in each group. In more detail, the average Hill diversity numbers (^{1}D) per group were 850, 1301 and 636 for cases carrying isolated del(11q), del(13q) or trisomy 12,

respectively, versus 2051 and 1917 for cases carrying isolated *NOTCH1 or TP53* mutations, respectively. (Supplementary Figure 1).

Prominent oligoclonal expansions in cases with copy number aberrations

All groups of the present study displayed repertoire restriction characterized by oligoclonal expansions. More precisely, the cumulative frequency of the major clonotypes (CF-10) ranged between 11.5-46.3, 17.3-37.1, 17.7-66.5 for cases carrying isolated del(11q), del(13q) or trisomy 12, respectively, versus 8.4-20.9 and 7.4-53.7 for cases carrying isolated NOTCH1 or TP53 mutations, respectively (Figure 1A). We detected similar patterns when assessing the expanded clonotypes (range: 5-31 expanded clonotypes/sample; median: 17 expanded clonotypes/sample) (Figure 1B).

The only difference concerned the *TP53*-mutation group, where the median cumulative frequency of the expanded clonotypes (MCF-Ex) was larger compared to the del(13q) group (Figure 2). The Kruskal-Wallis test documented a significant difference in the clonality profiles of the various groups of our study. More particularly, cases bearing CNAs displayed a significantly more oligoclonal repertoire compared to cases carrying *TP53* or *NOTCH1* mutations, irrespective of which repertoire fraction was tested (MCF-10, α=0.05; ex-MCF: α=0.03). Interestingly, *post-hoc* comparisons using Benjamini-Hochberg correction revealed that significant statistical differences (α=0.03) held mainly for the MCF-10 values of the trisomy 12 versus the *NOTCH1* mutation group (Figure 2).

Distinct TRBV gene repertoires in groups of patients with CLL with distinct genomic aberrations

The TRBV gene repertoires of all groups were restricted (Figure 3). In more detail, TRBV12-3, TRBV29-1, TRBV19, TRBV5-1 and TRBV6-5 accounted for one-third of the total repertoire in all groups, except the *TP53*-mutation group, where the TRBV29-1 gene was under-represented (α=0.002) and, in contrast, the TRBV11-3 and TRBV6-9 genes were over-represented (α=0.01).

However, when we focused only on the major (top-10) clonotypes, differences in the TRBV gene rank became more obvious: TRBV29-1 was the most frequent TRBV gene in the major clonotypes of all groups, except for the *TP53*-mutation group, where TRBV12-3 predominated. In general, cases carrying *NOTCH1* and *TP53* mutations demonstrated more prominent, 'group-biased' differences in the usage of particular TRBV genes in the repertoire of major clonotypes (α<0.05) (Figure 4).

Skewing was also noted in the TRBV gene repertoire of the significantly expanded clonotypes in different groups. In more detail, TRBV29-1 appeared with increased frequency (α<0.05) in the repertoire of the significantly expanded clonotypes in the del (11q), trisomy 12 and *NOTCH1*-mutation groups when compared to the remaining polyclonal background (Figure 5A). On the other hand, the significantly expanded clonotypes of the *TP53*-mutation group displayed increased frequency in TRBV12-3 (α=0.02) (Figure 5A). In addition, differential expression analysis highlighted several genes in each group that were significantly depressed in the repertoire of the

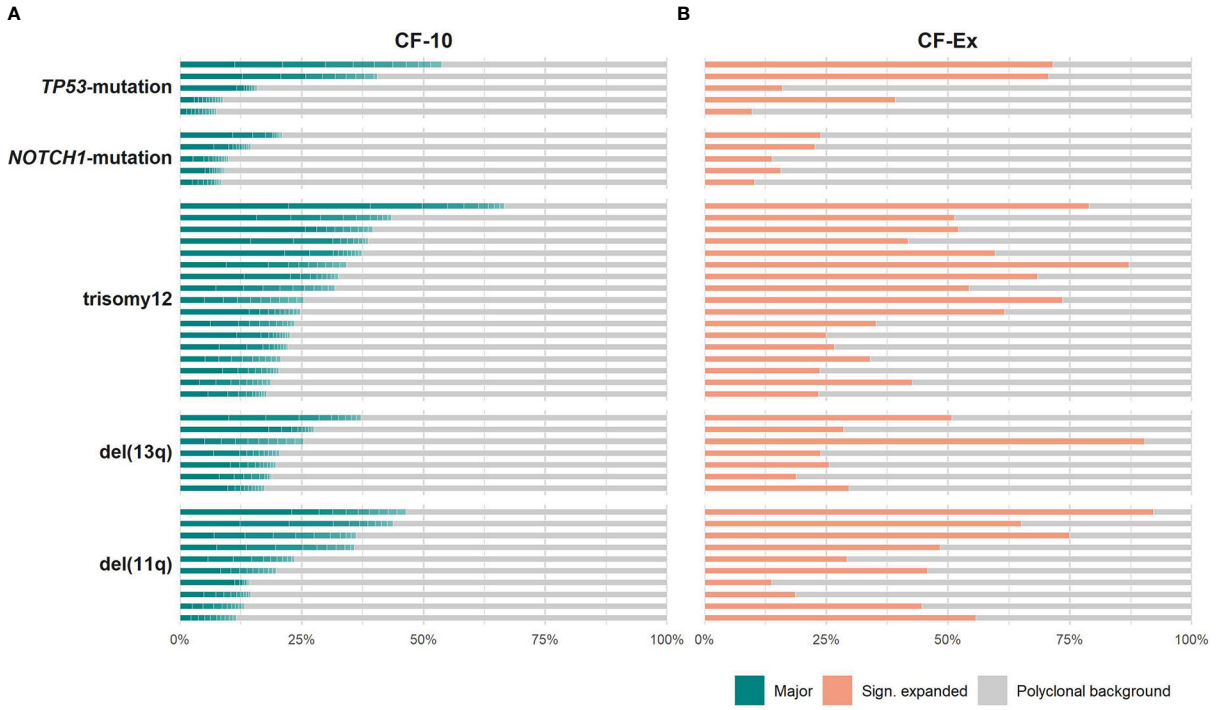


FIGURE 1
Clonality profiles of individual cases with CLL bearing distinct genomic aberrations. **(A)** Frequency (%) of the major clonotypes per sample of a given group. **(B)** Cumulative frequency (%) of the significantly expanded clonotypes per sample of a given group. Colored tiles depict each of the 10 major clonotypes within a sample and the fraction occupied by the significantly expanded clonotypes, respectively; the remaining clonotypes (polyclonal background) are depicted in grey.

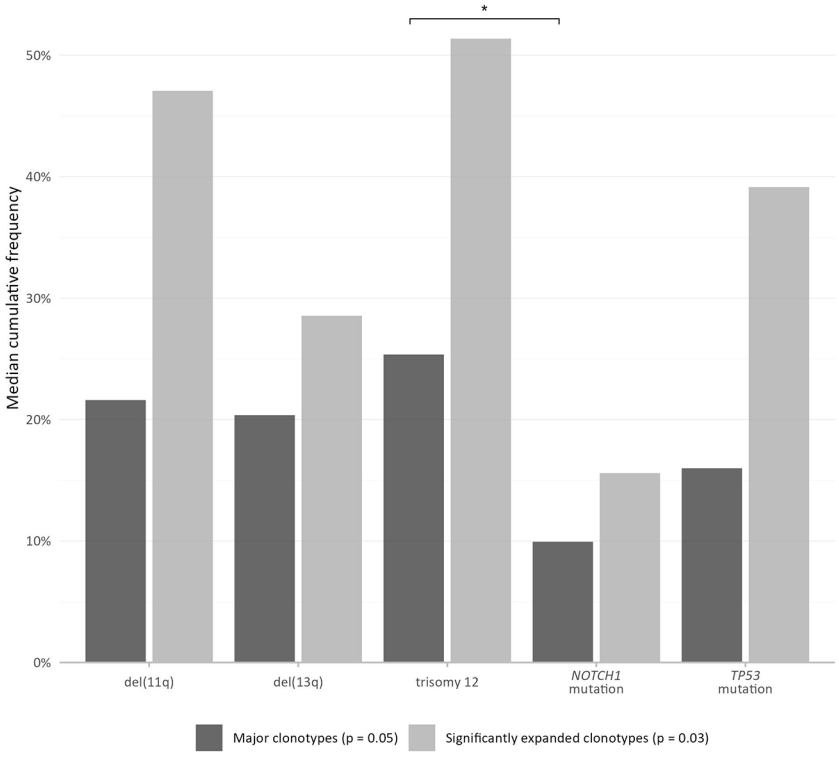


FIGURE 2
More pronounced TR repertoire skewing in cases bearing copy number aberrations, particularly trisomy 12, versus gene mutations. The bars represent the median cumulative frequency of the major clonotypes in each group (MCF-10) in dark grey color, and the median cumulative frequency of the significantly expanded clonotypes in each group (ex-MCF), in light grey color. The asterisk (*) refers to statistical significance at the level of α=0.03 following *post-hoc* comparisons using Benjamini-Hochberg correction.

significantly expanded clonotypes versus the remaining clonotypes (Figure 5A and Supplementary Table S6).

Finally, comparisons of the repertoires of the significantly expanded clonotypes in the different groups highlighted asymmetric usage of certain TRBV genes (Figure 5B). In more detail, TRBV3-1 and TRBV25-1 were found exclusively in cases carrying del(13q) (α<0.0002). Additionally, TRBV2 was found exclusively in cases bearing CNAs, while TRBV7-4 prevailed only in cases with trisomy 12 (α<0.0003).

Clonotype sharing: "Public" and "group-specific"

Comparisons of the clonotype repertoires of the various groups identified a total of 1,252 "public" clonotypes (0.26% of all analyzed clonotypes) that were shared by 2-7 samples of the present cohort regardless the background of genomic aberrations. Cross-comparisons against TRB gene rearrangement sequences from various other entities documented that none of these "public" clonotypes had been previously described in any other context, thus, they could be deemed as "CLL-biased". Interestingly, patients of a given group also displayed a number of exclusive shared clonotypes ("group-specific"). In more detail, we identified 321 clonotypes exclusively shared between cases with del(11q), 55 in the del(13q) group, 180 in the trisomy 12 group, and, finally, 70 and 56 clonotypes in the *NOTCH1*- and *TP53*-mutant groups, respectively.

In silico prediction reveals putative neoepitopes derived from *NOTCH1* and *TP53* gene mutations in CLL

The search for putative specific epitopes is far more straightforward in cases bearing point mutations or small insertions/deletions compared to cases with large genomic aberrations, since the latter involve many genes and affect many different pathways. On these grounds, we focused on the *NOTCH1*-mutation and *TP53*-mutation groups in order to explore the hypothesis that mutations in these genes could lead to the expression of neoepitopes that might select specific TR clones.

To that purpose, the mutant amino acid sequences of the Notch1 and p53 proteins from cases in the respective groups were dissected into epitopes of different length (k-mers of 9-15 amino acids) containing the mutant positions and all the produced epitopes were checked for their immunogenicity as described in the Materials and Methods section. In cases where the mutation resulted to a frameshift, the entire abnormal part of the protein was used for T-cell class I epitope prediction (Figure 6) Interestingly, mutations in the *TP53* gene led to the identification of greater number of possibly

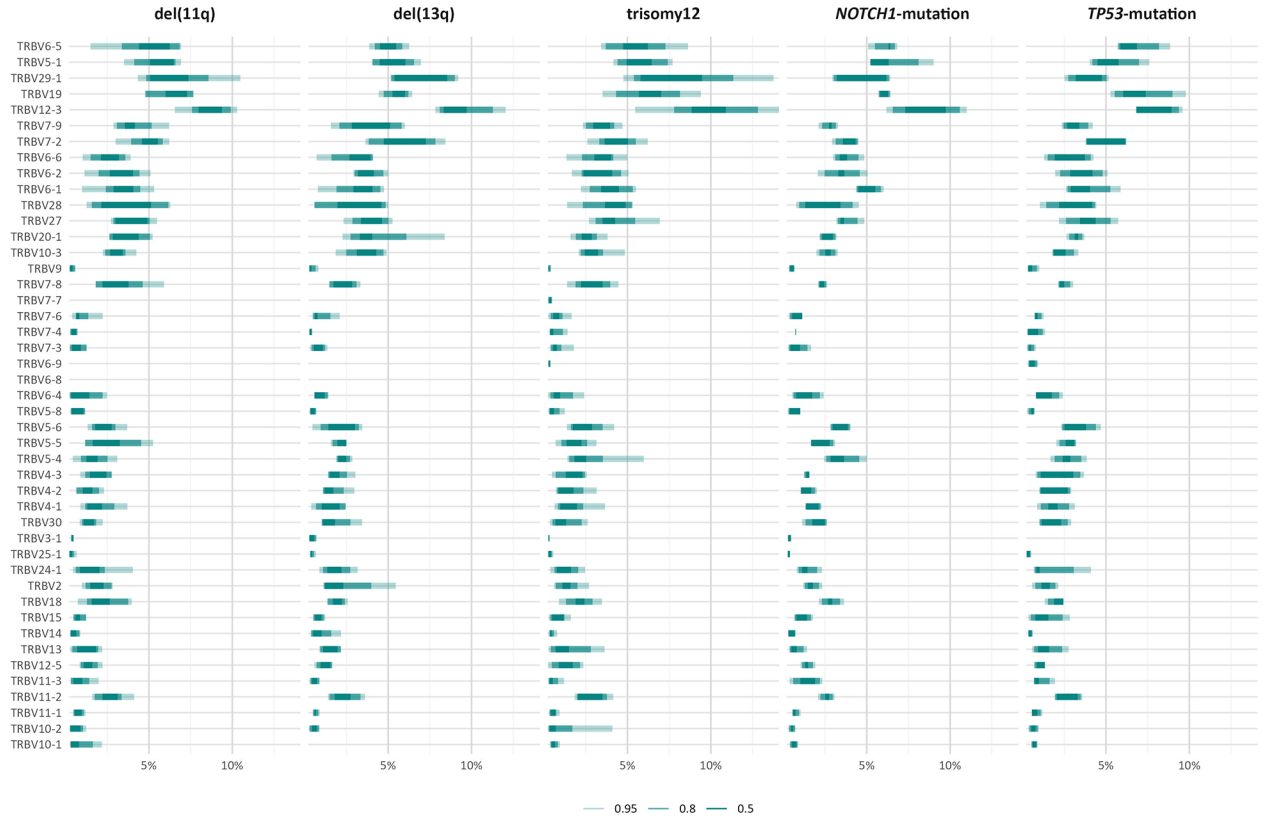


FIGURE 3
TRBV gene usage in different groups of patients with CLL carrying distinct genomic aberrations. The area including the frequency values of 50% of all observations in a given group is depicted in darker shade.

immunogenic neoepitopes, in sharp contrast to the cases with a 2-base pair frameshift deletion in *NOTCH1*: 1-182 predicted putative neoepitopes (median: 20 neoepitopes) in cases with mutations in *TP53* versus 3 predicted immunogenic neoepitopes derived from the 2-base pair deletion in *NOTCH1* (Supplementary Table S4). All examined cases are listed in Table 1, which also gives information about the identified mutations, their postulated molecular consequences and the number of filtered-in putative tumor-derived neoepitopes.

TR clonotypes from each of these cases along with the respective MHC alleles and the predicted putative neoepitopes were used as an input for the ERGO-II tool in order to predict their binding affinity. A different fraction of the repertoire (0.31-26.79% of all identified clonotypes) in each case corresponded to TR clonotypes identified

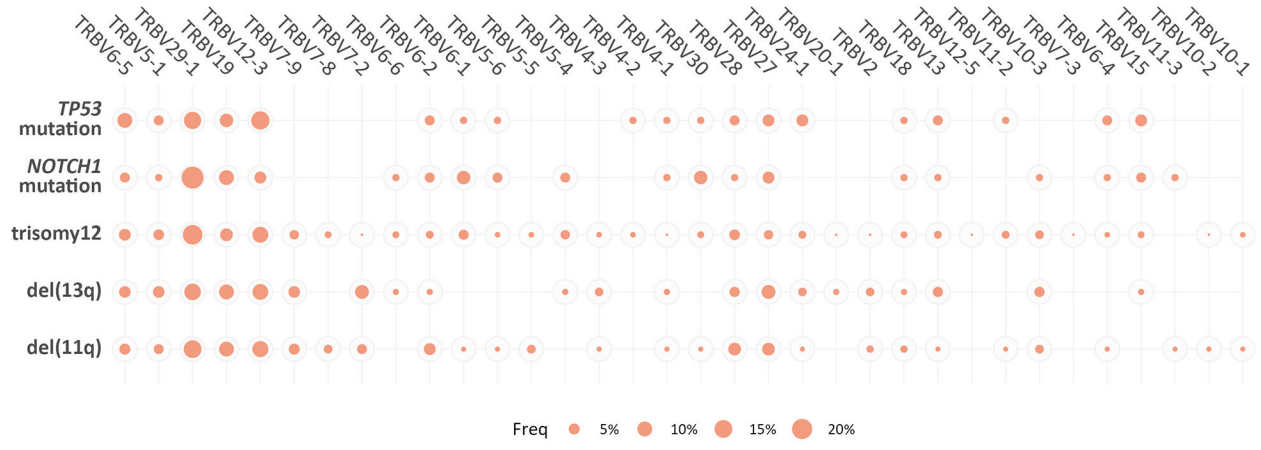


FIGURE 4
Biased TRBV gene usage in the major clonotypes in groups of patients with CLL bearing distinct genomic aberrations. The size of each bubble is proportional to the mean frequency (%) of each TRBV gene in the repertoire of the major clonotypes of all patients within a given group.

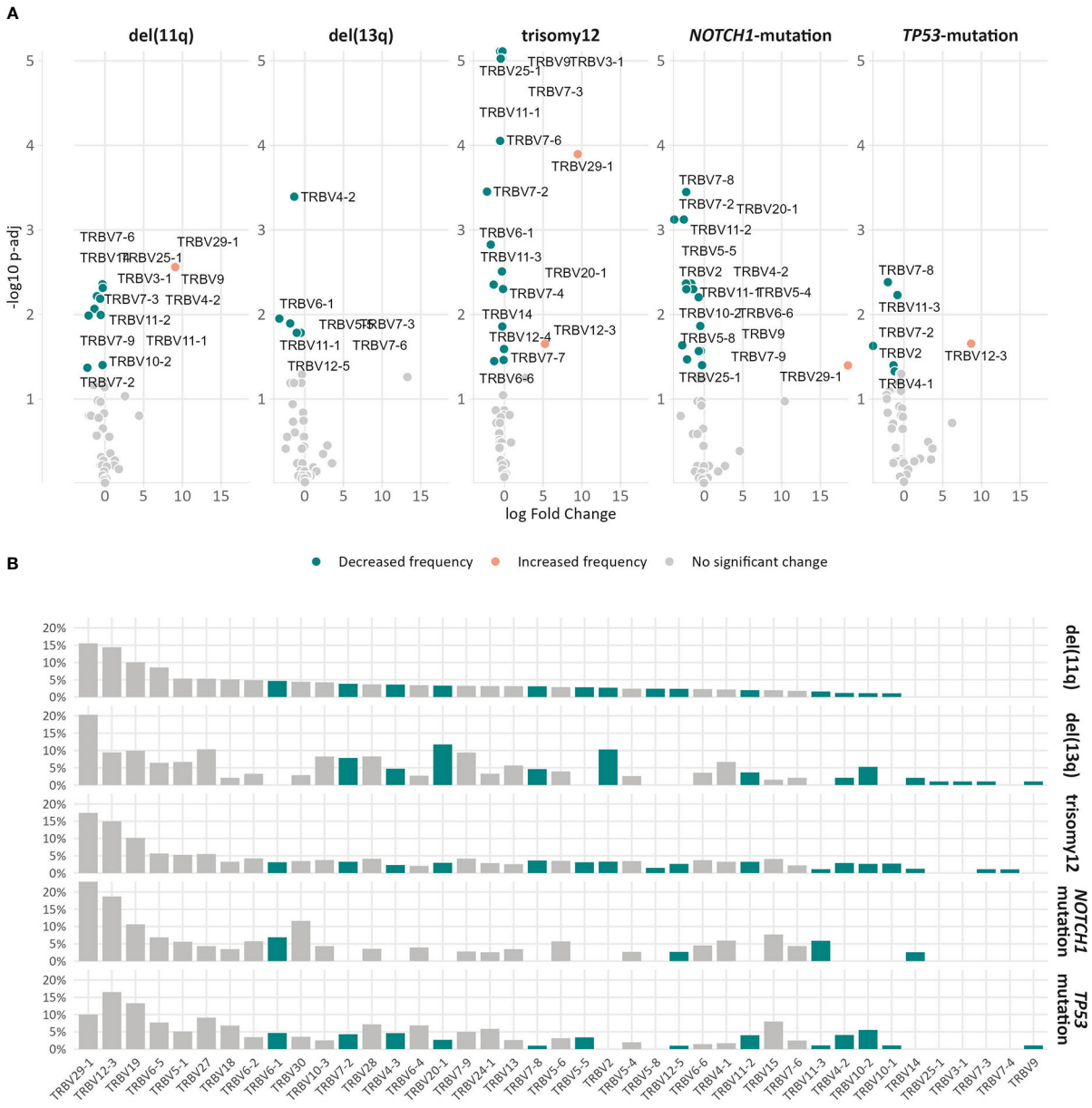


FIGURE 5
Differential usage of TRBV genes in the repertoires of the expanded clonotypes of the various groups in the study. **(A)** Volcano plots depict the significant variation in the TRBV gene usage in the repertoire of expanded clonotypes' compared to the remaining polyclonal background within a given group. TRBV genes with decreased frequency in the repertoire of expanded clonotypes are depicted as petrol bubbles, whereas TRBV genes with increased frequency are depicted as orange bubbles. **(B)** TRBV gene frequencies in the repertoire of expanded clonotypes in each group. The colored bars represent TRBV genes with statistically significant differences between the different groups.

by ERGO-II as displaying high binding affinity against putative neoepitopes derived from a particular genomic lesion (Table 1). Such postulated neoepitope-specific clonotypes ranged significantly in size from immunodominant to singletons i.e. clonotypes called out from a single sequencing read. Interestingly though, the most immunodominant clonotype (clonotype frequency: 11.1%) of Pt4 was characterized as neoepitope-specific displaying great binding affinity (score=0.85) against a 10-aa mutant epitope of p53 bound to HLA-B*27 based on ERGO-II. TR clonotypes identified by ERGO-II as high-affinity against putative neoepitopes were cross-compared against the complete clonotype repertoires of cases in the *TP53*-mutation and *NOTCH1*-mutation groups. This comparison did not identify identical clonotypes and/or identical CDR3s shared between the examined cases that belong to the same group. Interestingly, within the *NOTCH1*-mutation group, despite the fact that the examined cases bore the same 2-base pair frameshift deletion leading to same predicted neoepitopes, only 2 putative 'epitope-specific' TR clonotypes (TRBV12-3 - CASSLKSAGRNNSPLHF | TRBV29-1 - CSVVATVSGNTIYF) were shared between 2 of 3 examined cases. Moreover, no statistically significant differences were identified regarding the CDR3 length distribution and TRBV gene repertoire of the predicted neoepitope-specific TR clonotypes versus the remaining clonotypes of either the *TP53* or the *NOTCH1* mutant groups. Particularly regarding TRBV gene repertoire, we

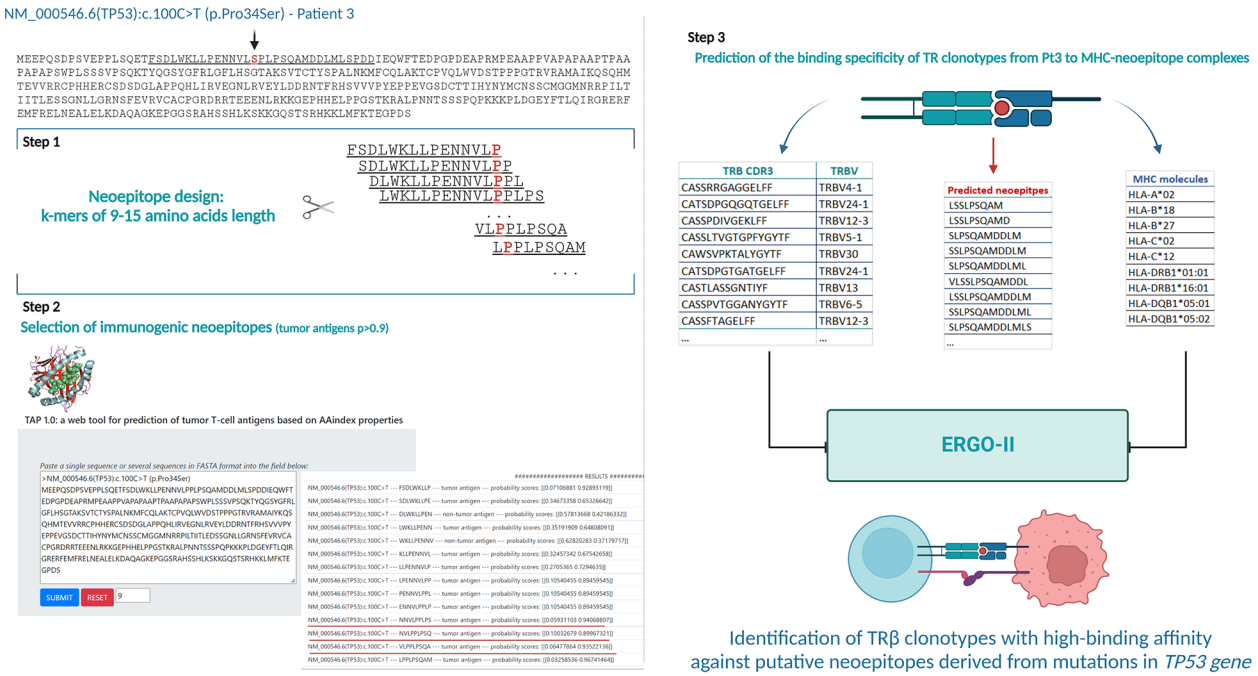


FIGURE 6
Analytical workflow for the prediction of neoepitopes and search for neoepitope-specific clonotypes (Illustrative case of Pt3). Identification of putative tumor-derived epitopes on TAP 1.0 tool using the mutant protein sequence of each case (Step 1 and 2). Lists of the TR clonotypes, *in silico* predicted neoepitopes and MHC molecules from each patient used for prediction of their binding affinity in ERGO-II (Step 3). Created with BioRender.com.

noted that the predicted neoepitope-specific TR clonotypes displayed the same group-specific restrictions (as documented above), with TRBV genes retaining almost the same rank in the repertoire when gene frequencies were considered.

Discussion

The advent of NGS coupled with robust bioinformatics tools have revolutionized the field of immunogenetics towards precision medicine approaches. High-throughput studies allow for in-depth characterization of clonal dynamics of complex repertoires in various clinical contexts, including cancer, offering important insight in the implicated processes and mechanisms (37). In CLL, NGS immunogenetics has documented repertoire restrictions consistent with antigenic drive in the TR repertoire of the bystander T cells, alluding to dynamic interactions operating within the TME that arguably shape clonal behavior (16, 37–39). However, the exact nature of the selecting antigens remains to be determined (17, 40).

In previous studies from our group, we have provided evidence that at least a fraction of T cells in patients with CLL may specifically recognize leukemia-associated antigens, with the clonotypic BcR IG expressed by the malignant cells emerging as a potential source of neoepitopes selecting T cells (17, 19, 24). In the present study, we extended our analyses and examined common recurrent genomic aberrations of CLL cells as another possible source of immunogenic neoepitopes for T cells. More precisely, we assessed by NGS the TR gene repertoire in groups of patients with CLL carrying as an isolated genomic aberration one of del(11q), del(13q), trisomy 12, *TP53* or *NOTCH1* gene mutations.

In keeping with the literature, all cases of the present study displayed an oligoclonal TR repertoire, regardless of the underlying genomic background (13, 14, 16, 17). However, cases bearing CNAs, particularly those with trisomy 12, presented with more pronounced repertoire skewing compared to cases bearing isolated *TP53* or *NOTCH1* mutations. On these grounds, we argue that the observed differences may be linked to the particular genomic profile of each group. The possible confounding effect of shared (stereotyped) BcR IG was excluded since cases within a given group expressed unique clonotypic BcR IG. Another possible confounding factor referring to the well-known aged-related TR repertoire restriction was minimized given that the studied groups were aged-balanced (41).

TRBV gene repertoires were biased in all groups, albeit with no major differences between groups, possibly excepting the *TP53*-mutation group. The herein reported overrepresentation of the TRBV12-3, TRBV29-1, TRBV19, TRBV5-1 and TRBV6-5 genes has been previously described in CLL and can be explained by naturally occurring convergence also observed in the TR gene repertoires from healthy T cells (16, 42). Nonetheless, significant differences between the groups emerged when the analysis was restricted to either the major clonotypes or the significantly expanded clonotypes. In this case, we noted asymmetric usage of certain TRBV genes, prompting us to speculate that the respective clones may represent the most biologically relevant ones, possibly expanding in response to CLL-derived neoepitopes.

A notable finding of the present study concerns the identification of ‘public’ clonotypes i.e. clonotypes shared by different patients, which, however, had not been previously described in other contexts beyond CLL: on these grounds, such clonotypes may be deemed as “CLL-biased”. Moreover, certain “public” clonotypes were exclusive to a given group of our study (“group-specific”), supporting the

10.3389/fonc.2023.1097942

TABLE 1 Putative neoepitopes and predicted neoepitope-specific TR clonotypes in cases with *TP53* or *NOTCH1* mutations.

Case ID	Gene	Variant name	Molecular consequence	Structural motif	No of putative neoepitopes	No of neoepitope specific TR clonotypes	Percentage (%) of neoepitope specific TR clonotypes
Pt3	*TP53*	NM_000546.6: c.100C>T (p.Pro34Ser)	Substitution - Missense	N-term/ Transactivation	19	789	9.21
Pt14	*TP53*	- NM_000546.6: c.733G>A (p.Gly245Ser); - NM_000546.6: c.1146del (p.Lys382fs)	Substitution - Missense; Deletion /Insertion - Frameshift	L2/L3 | C - terminal	182	6,427	26.79
Pt21	*TP53*	NM_000546.6: c.464C>A (p.Thr155Asn)	Substitution - Missense	NDBL/beta-sheets	28	766	6.36
Pt22	*TP53*	NM_000546.6: c.607G>C (p.Val203Leu)	Substitution - Missense	NDBL/beta-sheets	1	42	1.19
Pt23	*NOTCH1*	NM_017617.5: c.7541_7542del (p.Pro2514fs)	Deletion - Frameshift	C-terminal heterodimerization and PEST domains	3	31	0.31
Pt24	*NOTCH1*	NM_017617.5: c.7541_7542del (p.Pro2514fs)	Deletion - Frameshift	C-terminal heterodimerization and PEST domains	3	124	1.64
Pt26	*TP53*	NM_000546.6: c.721T>C (p.Ser241Pro)	Substitution - Missense	L2/L3	20	194	3.24
Pt27	*NOTCH1*	NM_017617.5: c.7541_7542del (p.Pro2514fs)	Deletion - Frameshift	C-terminal heterodimerization and PEST domains	3	69	0.70

hypothesis of shared immunogenic neoepitopes originating from a particular genomic aberration.

To validate this hypothesis, peptides derived from mutant variants of p53 and Notch1 displayed by the patients of the respective groups, were used along with their MHC alleles for the *in silico* identification of neoepitope-specific TR clonotypes. Indeed, patients with *TP53* or *NOTCH1* mutations presented a fraction of TR clonotypes, reaching up to 26.8% of the total number of clonotypes per case, that were identified *in silico* to bind with high affinity to immunogenic neoepitopes predicted to arise from each respective mutation in each case. Additional TRBV gene repertoire analysis on these TR clonotypes revealed that the group-specific restrictions identified in the particular fractions of the total repertoire (i.e. major and significantly-expanded clonotypes) were retained also in the fraction of neoepitope-specific TR clonotypes.

Postulated high affinity, neoantigen-specific clonotypes were ranging in frequency from immunodominant to clonotypes corresponding to a single sequencing read. This further highlights the value of the great analytical depth achieved by NGS, while also providing novel information regarding the diversity and frequency of such clonotypes that could contribute to therapeutic benefit. Arguably, the significant repertoire restrictions described in cases bearing CNAs may arise from dosage changes: increase of gene expression levels progressively leads to cellular effects due to protein aggregation, proteotoxicity and stress responses, that further shape T cell-mediated immunity, as described elsewhere (43, 44). Tracking the abundance and the relative clonal size of these neoantigen-specific clonotypes in the lymphoid tissues will offer valuable insights into the clonal dynamics in the different anatomical sites hosting microenvironmental interactions of the T cells with the CLL and antigen-presenting cells.

The strict inclusion criteria applied in the present study regarding the presence of isolated genomic aberrations led to a limited number of patients in each group, which is unsurprising considering that CLL clones are often characterized by multiple aberrations. Despite this limitation, our findings support our starting hypothesis and represent the first evidence that abnormalities in gene expression as well as gene dosage alterations caused by recurrent genomic aberrations in CLL may actively shape the TR gene repertoire. Admittedly, the presented immunogenetic evidence would require future formal validation, whereby the herein reported library of predicted neoantigens could serve as input for further experimentation using peptide-MHC-I multimers aiming to detect neoantigen-specific T cells circulating on the blood that could be engaged in anti-tumor responses against CLL cells. Arguably, relevant knowledge can meaningfully contribute to increasing the efficacy of T cell-based immunotherapeutic approaches driven by the ability to select and guide immune recognition by CLL-specific T cells.

Data availability statement

The data presented in the study are deposited in the ArrayExpress repository (https://www.ebi.ac.uk/arrayexpress/), accession number E-MTAB-12216.

Ethics statement

The studies involving human participants were reviewed and approved by CERTH's ethics committee: 21/03/2019 Approval document: ETH.COM-45. The patients/participants provided their written informed consent to participate in this study.

Author contributions

EV, AC, KS, and PB conceived of the presented idea. EV planned and carried out the experiments, performed the analysis and wrote the manuscript. ES and GG contributed to the experiments. NP and GK contributed on the statistical analysis and the design of the figures. SJ, NS, PR, LS, CO, and LM provided samples for the study. All authors provided critical feedback and helped shape the manuscript.

Funding

This study was supported in part by: The Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No 2285. The Lion's Cancer Research Foundation, Uppsala. The Swedish Cancer Society (19 0425 Pj 01 H), the Swedish Research Council (2020-01750), the Knut and Alice Wallenberg Foundation (2016.0373), Region Stockholm (ALF/FoUI-962423), and Radiumhemmets Forskningsfonder (194133), Stockholm. The Special Program on Metastatic Disease – 5 per mille 2018 (ID #21198), by Associazione Italiana per la Ricerca sul Cancro – AIRC, Milano, Italy. The ERA-NET TRANSCAN-2 Joint Transnational Call for Proposals JTC 2016 (project #179 NOVEL). 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.

Acknowledgments

We thank Konstantinos Pasentsis for expert technical assistance.

Conflict of interest

CH declares part ownership of Munich Leukemia Laboratory MLL. SJ is employed by the MLL. RR received honoraria and is a member on the advisory board of Abbvie, AstraZeneca, Janssen, Illumina and Roche. PG received honoraria and is a member on the advisory board of AbbVie, Acerta/AstraZeneca, Adaptive, ArQule/MSD, BeiGene, CelGene/Juno, Gilead, Janssen, Loxo/Lilly, Sunesis; and also receives research funding from AbbVie, Gilead, Janssen, Novartis, Sunesis. LS received honoraria and is a member on the advisory board of AbbVie, AstraZeneca, BeiGene, Janssen; and also received travel grants from BeiGene, Janssen; Speaker bureau: Octapharma. KS received research funding from Abbvie, AstraZeneca and Abbvie and is a member on the advisory board of AbbVie, AstraZeneca, Gilead, Janssen, and Bristol Myers Squibb. AC received research funding from Abbvie, Novartis and Janssen.

The remaining 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1097942/full#supplementary-material

References

1. Burger JA. Treatment of chronic lymphocytic leukemia. *New Engl J Med* (2020) 383 (5):460–73. doi: 10.1056/NEJMra1908213

2. Wierda WG, Brown J, Abramson JS, Awan F, Bilgrami SF, Bociek G, et al. NCCN guidelines® insights: Chronic lymphocytic Leukemia/Small lymphocytic lymphoma, version 3.2022. *J Natl Compr Canc Netw* (2022) 20(6):622–34. doi: 10.6004/jnccn.2022.0031

3. Vlachonikola E, Stamatopoulos K, Chatzidimitriou A. T Cell defects and immunotherapy in chronic lymphocytic leukemia. *Cancers (Basel)* (2021) 13(13):3255. doi: 10.3390/cancers13133255

4. Perutelli F, Jones R, Griggio V, Vitale C, Coscia M. Immunotherapeutic strategies in chronic lymphocytic leukemia: Advances and challenges. *Front Oncol* (2022) 12. doi: 10.3389/fonc.2022.837531

5. Vlachonikola E, Stamatopoulos K, Chatzidimitriou A. T Cells in chronic lymphocytic leukemia: A two-edged sword. *Front Immunol* (2021) 11:612244. doi: 10.3389/fimmu.2020.612244

6. Bagnara D, Kaufman MS, Calissano C, Marsilio S, Patten PEM, Simone R, et al. A novel adoptive transfer model of chronic lymphocytic leukemia suggests a key role for T lymphocytes in the disease. *Blood* (2011) 117(20):5463–72. doi: 10.1182/blood-2010-12-324210

7. Gorgun G, Ramsay AG, Holderried TAW, Zahrieh D, Le Dieu R, Liu F, et al. Eμ-TCL1 mice represent a model for immunotherapeutic reversal of chronic lymphocytic leukemia-induced T-cell dysfunction. *Proc Natl Acad Sci U S A* (2009) 106(15):6250–5. doi: 10.1073/pnas.0901166106

8. Grioni M, Brevi A, Cattaneo E, Rovida A, Bordini J, Bertilaccio MTS, et al. CD4+ T cells sustain aggressive chronic lymphocytic leukemia in eμ-TCL1 mice through a CD40L-independent mechanism. *Blood Adv* (2021) 5(14):2817–28. doi: 10.1182/bloodadvances.2020003795

9. Arruga F, Gyau BB, Iannello A, Deaglio S, Vitale N, Vaisitti T. Immune response dysfunction in chronic lymphocytic leukemia: Dissecting molecular mechanisms and microenvironmental conditions. *Int J Mol Sci* (2020) 21(5):1825-47. doi: 10.3390/ijms21051825

10. Riches JC, Davies JK, McClanahan F, Fatah R, Iqbal S, Agrawal S, et al. T Cells from CLLpatients exhibit features of T-cell exhaustion but retain capacity for cytokine production. *Blood* (2013) 121(9):1612–21. doi: 10.1182/blood-2012-09-457531

11. Nicholas NS, Apollonio B, Ramsay AG. Tumor microenvironment (TME)-driven immune suppression in b cell malignancy. *Biochim Biophys Acta Mol Cell Res* (2016) 1863 (3):471–82. doi: 10.1016/j.bbamcr.2015.11.003

12. Riches J C, Ramsay A G, G. Gribben J. Immune dysfunction in chronic lymphocytic leukemia: The role for immunotherapy. *Curr Pharm Des* (2012) 18 (23):3389–98. doi: 10.2174/138161212801227023

13. Farace F, Orlanducci F, Dietrich PY, Gaudin C, Angevin E, Courtier MH, et al. T Cell repertoire in patients with b chronic lymphocytic leukemia: Evidence for multiple *in vivo* T cell clonal expansions. *J Immunol* (1994) 153(9):4281–90. doi: 10.4049/jimmunol.153.9.4281

14. Goolsby CL, Kuchnio M, Finn WG, Peterson L. Expansions of clonal and oligoclonal T cells in b-cell chronic lymphocytic leukemia are primarily restricted to the CD3(+)CD8(+) T-cell population. *Cytometry* (2000) 42(3):188–95. doi: 10.1002/1097-0320(20000615)42:3<188::AID-CYTO5>3.0.CO;2-Q

15. Zaborsky N, Holler C, Geisberger R, Asslaber D, Gassner FJ, Egger V, et al. B cell receptor usage correlates with the sensitivity to CD40 stimulation and the occurrence of CD4+ T cell clonality in chronic lymphocytic leukemia. *Haematologica* (2015) 100 (8):307–10. doi: 10.3324/haematol.2015.124719

16. Vardi A, Agathangelidis A, Stalika E, Karypidou M, Siorenta A, Anagnostopoulos A, et al. Antigen selection shapes the T-cell repertoire in chronic lymphocytic leukemia. *Clin Cancer Res* (2016) 22(1):167–74. doi: 10.1158/1078-0432.CCR-14-3017

17. Vardi A, Vlachonikola E, Karypidou M, Stalika E, Bikos V, Gemenetzi K, et al. Restrictions in the T-cell repertoire of chronic lymphocytic leukemia: High-throughput immunoprofiling supports selection by shared antigenic elements. *Leukemia* (2017) 31 (7):1555–61. doi: 10.1038/leu.2016.362

18. Rovida A, Maccalli C, Scarfò L, Dellabona P, Stamatopoulos K, Ghia P. Exploiting b-cell receptor stereotypy to design tailored immunotherapy in chronic lymphocytic leukemia. *Clin Cancer Res* (2021) 27(3):729–39. doi: 10.1158/1078-0432.CCR-20-1632

19. Vardi A, Agathangelidis A, Gkagkaridou S, Schaap-Johansen AL, Karipidou M, Boukla A, et al. The clonotypic BCR IG of CLL patients contain predicted T-cell class I epitopes with shared structural properties. *Blood* (2021) 138(Supplement 1):1540–0. doi: 10.1182/blood-2021-152422

20. de Mattos-Arruda L, Vazquez M, Finotello F, Lepore R, Porta E, Hundal J, et al. Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the ESMO precision medicine working group. *Ann Oncol* (2020) 31(8):978–90. doi: 10.1016/j.annonc.2020.05.008

21. Hersey P, Menzies SW, Coventry B, Nguyen T, Farrelly M, Collins S, et al. Phase I/II study of immunotherapy with T-cell peptide epitopes in patients with stage IV melanoma. *Cancer Immunol Immunother* (2005) 54(3):208–18. doi: 10.1007/s00262-004-0587-8

22. Neumann A, Hörzer H, Hillen N, Klingel K, Schmid-Horch B, Bü HJ, et al. Identification of HLA ligands and T-cell epitopes for immunotherapy of lung cancer. *Cancer Immunol, Immunother* (2013) 62(9):1485–97. doi: 10.1007/s00262-013-1454-2

23. Vlachonikola E, Vardi A, Stamatopoulos K, Hadzidimitriou A. High-throughput sequencing of the T-cell receptor beta chain gene repertoire in chronic lymphocytic leukemia. *Methods Mol Biol* (2019) 1881:355–63. doi: 10.1007/978-1-4939-8876-1_24

24. Vardi A, Vlachonikola E, Papazoglou D, Psomopoulos F, Kotta K, Ioannou N, et al. T-Cell dynamics in chronic lymphocytic leukemia under different treatment modalities. *Clin Cancer Res* (2020) 26(18):4958–69. doi: 10.1158/1078-0432.CCR-19-3827

25. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT® tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol (Clifton NJ)* (2012) 882:569–604. doi: 10.1007/978-1-61779-842-9_32

26. Kotouza MT, 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

27. Shiffler RE. Maximum z scores and outliers. *Am Stat* (1988) 42(1):79–80.

28. Alberdi A, Gilbert MTP. A guide to the application of hill numbers to DNA-based diversity analyses. *Mol Ecol Resour* (2019) 19(4):804–17. doi: 10.1111/1755-0998.13014

29. Kaplinsky J, Arnaout R. Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. *Nat Commun* (2016) 7(1):11881. doi: 10.1038/ncomms11881

30. Tuomisto H. A diversity of beta diversities: straightening up a concept gone awry. part 1. defining beta diversity as a function of alpha and gamma diversity. *Ecography* (2010) 33(1):2–22. doi: 10.1111/j.1600-0587.2009.05880.x

31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47–7. doi: 10.1093/nar/gkv007

32. Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res* (2020) 48(D1):D1057–62. doi: 10.1093/nar/gkz874

33. Mastrodemou S, Stalika E, Vardi A, Gemenetzi K, Spanoudakis M, Karypidou M, et al. Cytotoxic T cells in chronic idiopathic neutropenia express restricted antigen receptors. *Leuk Lymphoma* (2017) 58(12):2926–33. doi: 10.1080/10428194.2017.1324154

34. Agathangelidis A, Galigalidou C, Scarfò L, Moysiadis T, Rovida A, Vlachonikola E, et al. High-throughput analysis of the T cell receptor gene repertoire in low-count monoclonal b cell lymphocytosis reveals a distinct profile from chronic lymphocytic leukemia. *Haematologica* (2020) 105(10):e515. doi: 10.3324/haematol.2019.221275

35. Herrera-Bravo J, Herrera Belén L, Farias JG, Beltrán JF. TAP 1.0: A robust immunoinformatic tool for the prediction of tumor T-cell antigens based on AAindex properties. *Comput Biol Chem* (2021) 91:107452. doi: 10.1016/j.compbiolchem.2021.107452

36. Springer I, Besser H, Tickotsky-Moskovitz N, Dvorkin S, Louzoun Y. Prediction of specific TCR-peptide binding from Large dictionaries of TCR-peptide pairs. *Front Immunol* (2020) 11:1803. doi: 10.3389/fimmu.2020.01803

37. 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

38. Hengeveld PJ, Levin MD, Kolijn PM, Langerak AW. Reading the b-cell receptor immunome in chronic lymphocytic leukemia: revelations and applications. *Exp Hematol* (2021) 93:14–24. doi: 10.1016/j.exphem.2020.09.194

39. Rodríguez-Vicente AE, Bikos V, Hernández-Sánchez M, Malcikova J, Hernández-Rivas JM, Pospisilova S. Next-generation sequencing in chronic lymphocytic leukemia: recent findings and new horizons. *Oncotarget* (2017) 8(41):71234–48. doi: 10.18632/oncotarget.19525

40. Rosén A, Murray F, Evaldsson C, Rosenquist R. Antigens in chronic lymphocytic leukemia–implications for cell origin and leukemogenesis. *Semin Cancer Biol* (2010) 20 (6):400–9. doi: 10.1016/j.semcancer.2010.09.004

41. Yoshida K, Cologne JB, Cordova K, Misumi M, Yamaoka M, Kyoizumi S, et al. Aging-related changes in human T-cell repertoire over 20 years delineated by deep sequencing of peripheral T-cell receptors. *Exp Gerontol* (2017) 96:29–37. doi: 10.1016/j.exger.2017.05.015

42. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res* (2009) 19 (10):1817–24. doi: 10.1101/gr.092924.109

43. Basilicata MF, Keller Valsecchi CI. The good, the bad, and the ugly: Evolutionary and pathological aspects of gene dosage alterations. *PloS Genet* (2021) 17(12):e1009906. doi: 10.1371/journal.pgen.1009906

44. Wheeler MC, Rizzi M, Sasik R, Almanza G, Hardiman G, Zanetti M. KDEL-retained antigen in b lymphocytes induces a proinflammatory response: A possible role for endoplasmic reticulum stress in adaptive T cell immunity. *J Immunol* (2008) 181 (1):256–64. doi: 10.4049/jimmunol.181.1.256

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 08 February 2023
DOI 10.3389/fonc.2023.1123029

OPEN ACCESS

EDITED BY
Richard Chahwan,
University of Zurich, Switzerland

REVIEWED BY
Richard J. Bende,
Academic Medical Research (AMR),
Netherlands
Linda B. Baughn,
Mayo Clinic, United States

*CORRESPONDENCE
Andreas Agathangelidis
agathan@biol.uoa.gr

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

RECEIVED 13 December 2022
ACCEPTED 24 January 2023
PUBLISHED 08 February 2023

CITATION
Gkoliou G, Agathangelidis A, Karakatsoulis G, Lalayanni C, Papalexandri A, Medina A, Genuardi E, Chlichlia K, Hatjiharissi E, Papaioannou M, Terpos E, Jimenez C, Sakellari I, Ferrero S, Ladetto M, Sanz RG, Belessi C and Stamatopoulos K (2023) Differences in the immunoglobulin gene repertoires of IgG versus IgA multiple myeloma allude to distinct immunopathogenetic trajectories. *Front. Oncol.* 13:1123029. doi: 10.3389/fonc.2023.1123029

Differences in the immunoglobulin gene repertoires of IgG versus IgA multiple myeloma allude to distinct immunopathogenetic trajectories

Glykeria Gkoliou[1,2], Andreas Agathangelidis[3]*, Georgos Karakatsoulis[1,4], Chrysavgi Lalayanni[5], Apostolia Papalexandri[5], Alejandro Medina[6], Elisa Genuardi[7], Katerina Chlichlia[2], Evdoxia Hatjiharissi[8], Maria Papaioannou[8], Evangelos Terpos[9], Cristina Jimenez[6], Ioanna Sakellari[5], Simone Ferrero[7], Marco Ladetto[7], Ramon Garcia Sanz[6], Chrysoula Belessi[10] and Kostas Stamatopoulos[1,11]

[1]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, [2]Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupoli, Greece, [3]Department of Biology, School of Science, National and Kapodistrian University of Athens, Athens, Greece, [4]Department of Mathematics, School of Sciences, University of Ioannina, Ioannina, Greece, [5]Hematology Department and HCT Unit, G. Papanikolaou Hospital, Thessaloniki, Greece, [6]Hematology Department, University Hospital of Salamanca, Salamanca, Spain, [7]Department of Molecular Biotechnologies and Health Sciences, Hematology Division, University of Turin, Turin, Italy, [8]First Department of Internal Medicine, AHEPA University Hospital, Aristotle University of Thessaloniki, Thessaloniki, Greece, [9]Department of Clinical Therapeutics, National and Kapodistrian University of Athens, Athens, Greece, [10]Hematology Department, Nikea General Hospital, Piraeus, Greece, [11]Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

The analysis of the immunogenetic background of multiple myeloma (MM) has proven key to understanding disease ontogeny. However, limited information is available regarding the immunoglobulin (IG) gene repertoire in MM cases carrying different heavy chain isotypes. Here, we studied the IG gene repertoire in a series of 523 MM patients, of whom 165 and 358 belonged to the IgA and IgG MM groups, respectively. IGHV3 subgroup genes predominated in both groups. However, at the individual gene level, significant ($p<0.05$) differences were identified regarding IGHV3-21 (frequent in IgG MM) and IGHV5-51 (frequent in IgA MM). Moreover, biased pairings were identified between certain IGHV genes and IGHD genes in IgA versus IgG MM. Turning to the imprints of somatic hypermutation (SHM), the bulk of rearrangements (IgA: 90.9%, IgG: 87.4%) were heavily mutated [exhibiting an IGHV germline identity (GI) <95%]. SHM topology analysis disclosed distinct patterns in IgA MM versus IgG MM cases expressing B cell receptor IG encoded by the same IGHV gene: the most pronounced examples concerned the IGHV3-23, IGHV3-30 and IGHV3-9 genes. Furthermore, differential SHM targeting was

also identified between IgA MM versus IgG MM, particularly in cases utilizing certain IGHV genes, alluding to functional selection. Altogether, our detailed immunogenetic evaluation in the largest to-date series of IgA and IgG MM patients reveals certain distinct features in the IGH gene repertoires and SHM. These findings suggest distinct immune trajectories for IgA versus IgG MM, further underlining the role of external drive in the natural history of MM.

KEYWORDS

multiple myeloma, immunogenetics, immunoglobulin isotypes, immunoglobulin a, immunoglobulin g, immunoglobulin gene repertoire, somatic hypermutation (SHM), n-glycosylation

Introduction

Multiple myeloma (MM) is a malignancy characterized by the accumulation of terminally differentiated clonal plasma cells (PCs) in the bone marrow (BM). It accounts for 10% of hematologic malignancies and 0.8% of all cancer diagnoses, respectively, with 130,000 new cases every year worldwide (1). The main feature of MM is the secretion of a specific monoclonal immunoglobulin (IG) by the malignant PCs which can be detected in the serum and/or the urine of patients with MM. MM is classified into different types depending on the isotype of the IG heavy and light chain. The predominant myeloma type is IgG (52%), followed by IgA (21%), light chain (16%), bi-clonal (2%), and IgM (0.5%), while IgD and IgE myelomas are rather infrequent (2).

Immunogenetic analysis of MM has proven key to understanding disease ontogenesis. Indeed, relevant studies have reported: (i) clonal relationship between switch variants expressed by the BM PCs and myeloma progenitors in the BM and blood (3, 4); (ii) virtual absence of the inherently autoreactive IGHV4-34 gene (5–7); and, (iii) patterns of somatic hypermutation (SHM) suggestive of a post-germinal center (GC) derivation (8, 9). Yet, limited evidence exists about the architecture of the IG gene repertoire in different types of MM characterized by the expression of different heavy chain isotypes, in particular IgA versus IgG. This is relevant for the ontogeny of MM, especially in the context of conflicting evidence regarding the immunogenetic profiles of IgA versus IgG memory B cells in non neoplastic conditions. Indeed, a higher SHM burden has been reported for $CD27^{+}IgA^{+}$ versus $CD27^{+}IgG^{+}$ normal memory B cells, perhaps due to the distinct microenvironment of the respective immune responses, especially given that IgA class switching occurs mostly in mucosa-associated lymphoid tissues (10). On the other hand, studies in IgA and IgG expressing B cells in different settings (11–13) have found an overall similar burden of SHM; finally, BM IgG PC may display higher SHM rates compared to circulating IgG memory B cells (14). At the clinical level, patients with IgA MM exhibit a higher incidence of t(4;14) (40% in IgA versus 13% in IgG), shorter progression-free survival (PFS) and worse overall survival (OS) compared to patients with IgG MM (15, 16).

Here, we investigated potential differences in the immunogenetic profiles of IgA versus IgG MM in a large multi-institutional cohort. We report distinct profiles particularly regarding the individual frequencies of particular IGH genes, the topology and molecular features of SHM and the characteristics of the heavy variable complementarity determining region 3 (VH CDR3). These findings allude to different antigen exposure trajectories and/or affinity maturation processes for MM subtypes, i.e. IgA and IgG MM, further underscoring the key role of microenvironmental interactions in the pathogenesis of MM.

Materials and methods

Study group

All patients were diagnosed with MM following the IMWG criteria (17). Collected samples derived from collaborating Institutions in Greece (n=176), Italy (n=72) and Spain (n=201) and the IMGT/LIGM-DB public database (18) (n=74). The sequence datasets from the Italian and Spanish groups have been reported previously (references (19) and (20), respectively). The study was approved by the local Ethics Review Committees of each participating Institution and was conducted in accordance with the Declaration of Helsinki.

Sample collection and sequencing analysis of IGHV-IGHD-IGHJ gene rearrangements

BM aspirates were collected from 449 patients with MM with a PC infiltration of at least 20%. Available immunogenetic data for the remaining 74 cases were retrieved from the IMGT/LIGM-DB public database (18). On the basis of serum immunoelectrophoresis/immunoblot analysis, 165 and 358 cases expressed clonotypic IG gene rearrangements belonging to the IgA and IgG isotypes, respectively.

BM mononuclear cells (BMMCs) were isolated by Ficoll-Hypaque density centrifugation. Genomic DNA (gDNA) extraction, total cellular RNA isolation and cDNA synthesis were performed as previously described (7, 21, 22). IGHV-IGHD-IGHJ gene rearrangements were PCR-amplified using Leader and IGHJ

primers as previously described (7). PCR products were subjected to bidirectional Sanger sequencing.

IGHV-IGHD-IGHJ gene rearrangement sequence analysis

IGHV-IGHD-IGHJ gene rearrangement sequences were annotated with the use of the IMGT/V-QUEST tool (23, 24). Only productive rearrangements were subjected to further analysis. IGHV, IGHD, and IGHJ gene usage, SHM molecular features and topology, and the characteristics (amino acid length and composition) of the VH CDR3 were extracted for each sequence.

IGHV-IGHD-IGHJ gene rearrangements were analyzed for VH CDR3 stereotypy through the use of our validated bioinformatics pipeline (25). In detail, a set of immunogenetic criteria were applied on the clustering process: (i) utilization of IGHV genes originating from the same phylogenetic clan, (ii) ≥ 50% amino acid identity and ≥ 70% similarity of the VH CDR3, (iii) same VH CDR3 length and, (iv) same offset of the common amino acid pattern.

The NetNglyc tool was used for the prediction of N-Glycosylation sites across the IGHV-IGHD-IGHJ gene rearrangement sequences, based on the application of artificial neural networks for the examination of the sequence context of Asn-Xaa-Ser/Thr sequons (26).

Statistical analysis and data visualization tools

Descriptive statistics for qualitative variables included counts and frequency distributions. To examine potential differences in terms of immunogenetic characteristics between the two patient groups (IgA MM versus IgG MM), the chi-square test and Fisher's exact test were applied. For all comparisons a significance level of α=0.05 was set. All statistical analyses were performed in R v.4.1.1.

Data visualization was performed in the R environment, with the open-source data visualization framework RawGraphs (https://rawgraphs.io/) and the Circos software package (http://mkweb.bcgsc.ca/tableviewer/).

Results

The IGHV gene repertoires of IgA and IgG MM patients display both shared but also unique characteristics

All 7 IGHV gene subgroups were represented in our cohort. In specific, IGHV3 predominated at cohort level (279/523 rearrangements; 53.3%), followed by IGHV4 (96/523 rearrangements; 18.4%) (Supplemental Table 1). At the individual IGHV gene level, 47 distinct IGHV genes were identified. Seven IGHV genes were defined as frequent (individual frequency ≥4%), collectively accounting for 41.9% of the total repertoire: IGHV3-30, IGHV3-23, IGHV3-9, IGHV3-33, IGHV4-59, IGHV2-5 and IGHV1-69 (Supplemental Table 2).

Subsequently, IGHV gene repertoires were studied separately in IgG MM and IgA MM. Predominance of the IGHV3 subgroup was evident in both groups, accounting for 181/358 rearrangements (50.6%) in the IgG group and 98/165 rearrangements (59.4%) in the IgA group. Eight IGHV genes were characterized as frequent in the IgG group and 9 in the IgA group, respectively. Of these, only 4 IGHV genes, namely IGHV3-23, IGHV3-30, IGHV3-9 and IGHV4-59, were characterized as frequent in both groups. The remaining frequent IGHV genes showed variations in utilization frequency among the 2 MM types, with some differences reaching statistical significance. The most prominent differences concerned the IGHV3-21 gene, which exhibited a frequency of 4.2% in IgG MM versus only 0.6% in IgA MM (p-value <0.05), and the IGHV5-51 gene, accounting for 4.7% in IgG MM versus 1.2% in IgA MM (p-value=0.05) (Supplemental Table 2) (Figure 1).

Shared and distinct VH CDR3 characteristics in IgA versus IgG multiple myeloma

IGHD gene repertoires

At cohort level, IGHD3 subgroup genes predominated (195/523 cases; 37.3%) followed by IGHD2 (99/523 cases; 18.3%) and IGHD6 genes (86/523 cases; 16.4%) (Supplemental Table 3). In total, 25 distinct IGHD genes were identified in the present cohort (Supplemental Table 4); the most frequent was IGHD3-10 (54/523

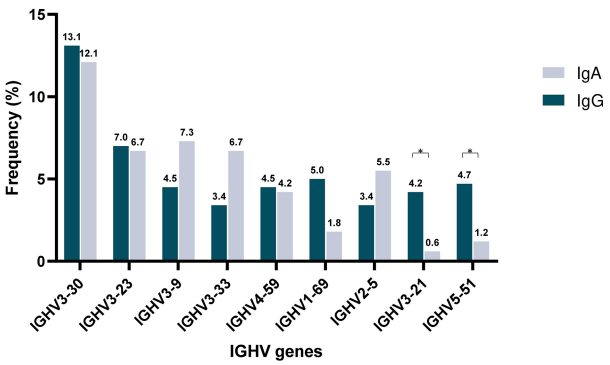


FIGURE 1
Shared and unique IGHV gene repertoires in IgA versus IgM multiple myeloma. Comparison of the most frequently used IGHV genes in IgA (blue) and IgG (gray) MM patients. Asterisks indicate the statistically significant differences between the two groups (p-value ≤ 0.05(*)).

cases; 10.3%), followed by IGHD3-3 (45/523 cases; 8.6%) and IGHD3-22 (44/523 cases; 8.4%).

IGHD gene repertoire analysis was also performed in relation to heavy chain isotype expression, which confirmed the predominance of IGHD3 subgroup genes in both MM groups. However, the second most common IGHD gene subgroup differed, being IGHD2 in IgG MM versus IGHD6 in IgA MM; moreover, IGHD1 genes were twice as frequent in IgG versus IgA MM (p-value <0.05) (Supplemental Table 3). At the individual IGHD gene level, the only significant difference concerned the IGHD5-12 gene that was more frequent in IgA MM compared to IgG MM (6.7% versus 2.5%, respectively; p-value <0.05) (Supplemental Table 4) (Supplemental Figure 1).

Analysis of the IGHD gene reading frames (RF) revealed an overrepresentation of RF2 in both IgA MM and IgG MM (46.6% in both groups) (Supplemental Table 5). RF1 and RF3 displayed similar relative frequencies in IgA MM (45 cases; 27.2% and 43 cases; 26%, respectively). In contrast, in IgG MM RF2 was followed by RF3 (118 cases, 32.9%), while RF1 displayed the smallest relative frequency (73 cases, 20.3%).

Subsequently, the distribution of RFs was compared at the individual IGHD gene level. Focusing on IGHD genes that were frequent in both MM types, we noted that the utilization of the IGHD3-10 gene in RF2 was significantly more frequent in IgG MM compared to IgA MM (57.9% versus 33.3%, p-value <0.05), whereas RF3 predominated in cases expressing the IGHD5-12 gene belonging to IgA MM versus IgG MM (63.6% versus 57.1%, p-value <0.05) (Supplemental Table 6).

IGHJ gene repertoires

The IGHJ4 gene was identified in 252/523 rearrangements of the total cohort (48.2%), followed by IGHJ6 (103/523 rearrangements; 19.7%) (Supplemental Table 7). At the heavy chain isotype level, the IGHJ4 gene was the most common in both IgG and IgA MM, followed in both cases by the IGHJ6 gene (Supplemental Table 7).

Biased pairings of IGHV/IGHD/IGHJ genes

Distinct patterns of associations of certain IGHV genes with IGHD and IGHJ genes were identified between IgA MM versus IgG MM (Figure 2). The most striking cases concerned: (i) IGHV3-30 in IgA MM, where 4/20 cases (20%) recombined with IGHD3-10, (ii) IGHV4-59 in IgG MM, where 5/16 cases (31%) recombined with IGHD2-21.

VH CDR3 properties

The median VH CDR3 length at cohort level was 15 amino acids (aa) (range: 6-29), with no differences between IgA MM versus IgG MM. However, statistically significant differences were identified regarding the distribution of IgA and IgG cases based on the VH CDR3 length. More particularly, a higher frequency of IgG cases expressed 14 aa long VH CDR3 (58/358 IgG cases; 16.2%) compared to IgA cases (15/165 cases; 9.1%) (p-value <0.05). The opposite trend was observed for cases carrying VH CDR3 of 19 amino acids, which were enriched in the IgA group (16/165 cases; 9.7%) compared to the IgG group (17/358 cases; 4.7%) (p-value <0.05) (Figure 3) (Supplemental Table 8).

All productive IgA and IgG MM IGHV-IGHD-IGHJ gene rearrangements of the present cohort were subjected to analysis for VH CDR3 stereotypy using our established bioinformatics (25). However, this analysis did not identify clusters of immunogenetically related sequences, supporting that VH CDR3 stereotypy does not occur (or is very infrequent) in MM.

Differential imprints of somatic hypermutation in subgroups of IgA versus IgG multiple myeloma

Cases were classified following the germline identity % (GI%) into the following subgroups: (i) truly unmutated [=100%], (ii) minimally mutated [>=99% and <100%], (iii) borderline mutated [>=97% and <99%], (iv) mutated [>=95% and <97%] and (v) heavily mutated [<95%]). The vast majority of rearrangements (471/523 cases, 90.1%) were heavily mutated (Supplemental Table 9), with a median GI of 92%; there was no difference in the median percent GI values burden between IgA versus IgG MM (91.3 GI% in IgA MM versus 92% in IgG MM). Only a single rearrangement from the IgA group was truly unmutated.

The analysis of SHM targeting at the individual IGHV gene level initially focused on IGHV genes that were characterized as frequent in both the IgG and IgA MM groups, namely IGHV3-23, IGHV3-30, IGHV3-9 and IGHV4-59. Cases expressing the IGHV3-23 and IGHV4-59 genes displayed significantly higher R:S ratios in VH CDRs versus VH FRs in IgG MM (p-value <0.001 and p-value < 0.01, respectively). When looking at individual regions of the VH domain, both the IgG and IgA groups displayed higher R:S ratios in VH CDR1 and VH CDR2 versus the VH FR regions (Figure 4). In contrast, significant differences were identified regarding the R:S ratios in the VH CDR1 domain between IgA and IgG patients carrying the IGHV1-18, IGHV3-48 and IGHV5-51 genes. In detail, the R:S ratio in the VH CDR1 of IGHV1-18 and IGHV5-51 gene rearrangements in patients with IgG MM was significantly higher compared to the respective IgA MM patients (10.5 in IgG versus 0.7 in IgA for IGHV1-18 rearrangements and 2.5 in IgG versus 0 in IgA for IGHV5-51 rearrangements, respectively; p-values <0.05 in both cases). The opposite trend was observed in cases expressing the IGHV3-48 gene, where the R:S ratio in the VH CDR1 in patients with IgG MM was significant lower compared to the respective ratio in IgA MM patients (0 in IgG versus 1.3 in IgA, respectively; p-value <0.05) (Supplemental Table 10).

Using the NetNglyc tool we detected N-glycosylation sequons in 93/523 IGHV-IGHD-IGHJ gene rearrangements (17.8%), with no significant differences between IgA MM (23/165 cases, 13.9%) versus IgG MM (70/358 cases, 19.5%) (p-value = 0.15). Of these, 38 cases carried novel N-glycosylation sequons introduced by SHM, again with no significant differences in frequency between IgA MM (8/23 cases, 34.8%) versus IgG MM (30/70 cases, 42.8%) (Supplemental Table 11).

Removal of germline encoded N-glycosylation sequons due to SHM was also observed in 18/430 cases (4.2%), again with a similar

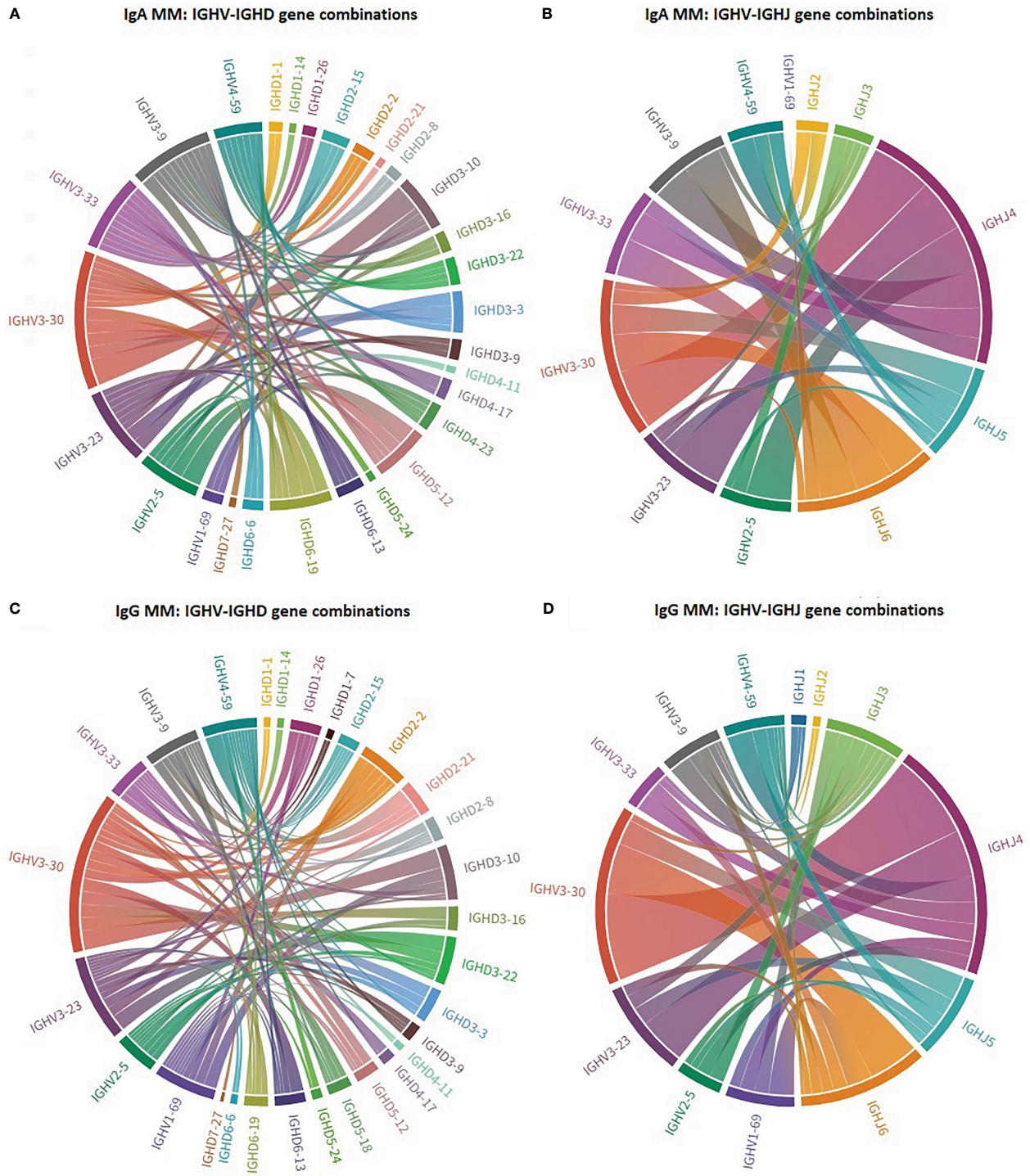


FIGURE 2
Biased IG gene pairings in multiple myeloma. Circos plot presenting IGHV-IGHD-IGHJ combinations of the 7 most frequent IGHV genes in IgA MM and IgG MM. The Circos software package was used to depict the predominant combinations between IGHV and IGHD genes [**(A)** IgA and **(C)** IgG patients] as well as between IGHV and IGHJ genes [**(B)** IgA and **(D)** IgG patients] in both MM groups.

distribution among the 2 MM types [7/142 IgA cases (4.9%) versus 11/288 IgG cases (3.8%)]. Interestingly, 5/18 cases where such removal was documented expressed the IGHV4-39 gene; in 4/5 cases (2 each concerning IgA or IgG MM) this was due to a recurrent Proline for Lysine substitution at position VH FR3 69 (Supplemental Table 12).

Discussion

In the present study we assessed the impact of antigen selection in shaping the IG gene repertoire of MM through an in-depth analysis of the clonotypic IGHV-IGHD-IGHJ gene rearrangements in patients expressing alpha or gamma isotypes.

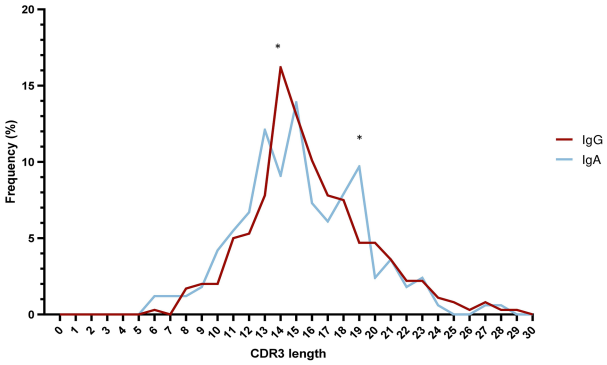


FIGURE 3
Skewed VH CDR3 length distribution in multiple myeloma. Distribution of VH CDR3 length of IgA MM (blue) and IgG MM (red). Significant differences between the two MM types are observed for VH CDR3 lengths of 14 aa and 19 aa. Asterisks indicate the statistically significant differences between the two groups (p-value ≤ 0.05 (*)).

Overall, the IGH gene repertoire in the present series was similar to those reported in previous reports (6, 7, 19), with the IGHV3-23, IGHV3-30, IGHV3-9 and IGHV4-59 genes predominating. However, certain IGHV or IGHD genes (e.g. IGHV2-5 and IGHD5-12) are herein reported as frequent for the first time, highlighting the added value of our study in offering a more comprehensive view of the immunogenetic landscape of MM. In keeping with the literature (19, 20), IGHV3 subgroup genes predominated in both IgA MM and IgG MM. Such predominance has also been noted in B cell lymphoproliferations, including monoclonal gammopathy of undetermined significance (MGUS) (27–29), chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL) (30, 31). A significant difference concerned IGHV4-34 that is conspicuously infrequent (<1%) in MM, yet it is very frequent in both hematologic malignancies, such as CLL and MCL (25, 31). Moreover, our cohort was notable for underrepresentation of the IGHV1-69, IGHV3-7 and IGHV1-8 genes, which have been reported as frequent in CLL and/or MCL (30, 31).

Studies in memory B cells have not identified differences in IG gene repertoires between IgA versus IgG memory B cells (10, 12, 32, 33). On these grounds, the herein reported distinctive features between IgA versus IgG MM do not appear to reflect the normal B cell differentiation, but rather allude to distinct selection forces shaping, at least in part, the respective BcR IG repertoires. This argument was reinforced by our finding of differential patterns of associations of certain IGHV genes with certain IGHD and IGHJ genes in IgG MM versus IgA MM types.

Overall, the present cohort of MM cases resembled normal B cells in terms of VH CDR3 length distribution and amino acid

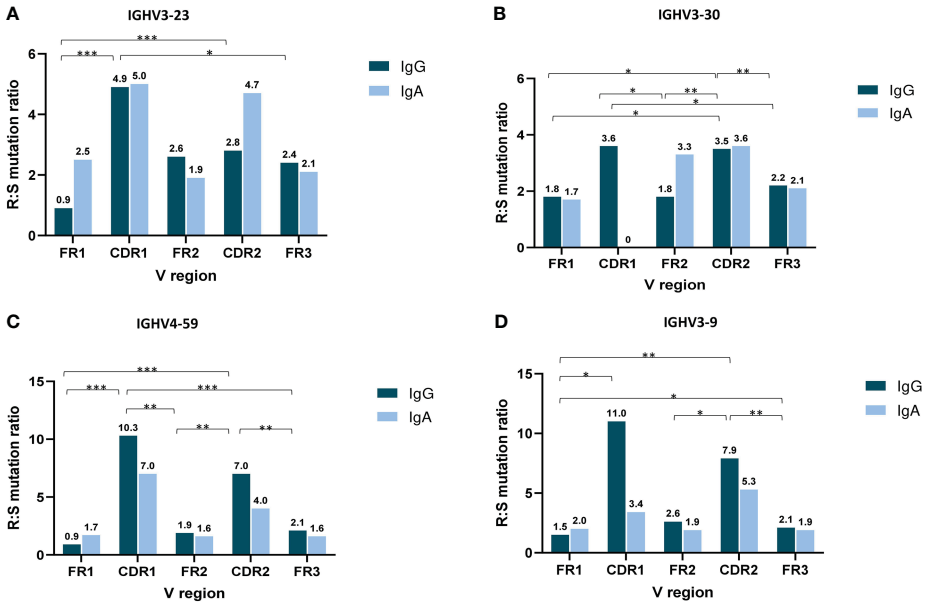


FIGURE 4
Topology of somatic hypermutation in IgA versus IgG multiple myeloma. Ratio of replacement to silent mutations (R:S) in the VH FRs and VH CDRs in cases expressing the 4 commonly frequent IGHV genes of MM patients with IgG and IgA isotypes: IGHV3-23 **(A)**, IGHV3-30 **(B)**, IGHV4-59 **(C)** and IGHV3-9 **(D)**. Asterisks indicate the statistically significant differences between the VH FR and VH CDR regions (p-value ≤ 0.05(*), p-value ≤ 0.01 (**) and p-value ≤ 0.001 (***)).

composition (34, 35), in sharp contrast to other malignancies of mature B cells e.g. chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL) or marginal zone lymphoma (MZL) (31, 36, 37). In keeping with the literature (19, 20) and in contrast with CLL, MCL or splenic marginal zone lymphoma (SMZL) (31, 38–40), VH CDR3 stereotypy was not observed in our cohort. Of note, however, when considering IgG MM and IgA MM cases separately, skewing to particular CDR3 lengths was noted, in some instances reaching statistical significance. Considering the fundamental role of the VH CDR3 in antigen recognition, at least in the pre-immune repertoire (41), this observation supports differences at the level of selection of the corresponding MM progenitor cells, particularly regarding the type of selecting elements, the timing and the precise location of antigenic interactions within the secondary lymphoid tissues.

The vast majority of cases in our cohort bore a significant number of SHMs, in line with previous studies (6, 7, 42), supporting the notion that the MM transforming events concern post-germinal center B cell progenitors (8, 43). This finding, along with SHM profiles characteristic of encounter with classical antigen(s), typified by preferential targeting of replacement mutations in VH CDRs rather than VH FRs, further corroborates the notion that the MM progenitor cells were selected by antigens during their process of differentiation into plasma cells (9, 44). A noteworthy novel finding concerned the differences observed in at least some subgroups of IgA MM versus IgG MM cases expressing BcR IG with the same IGHV gene. On these grounds, we argue that not only antigen exposure trajectories but also affinity maturation processes may differ for IgA MM versus IgG MM, further underlining the role of microenvironmental interactions in the pathogenesis of the disease. This notion is further supported by a recent study on transgenic mice reporting the existence of similar SHM patterns between two B cell subpopulations located at different tissues, i.e. Peyer's Patches germinal center B cells and splenic B cells (45). In this context, the identification of different SHM imprints in IgA compared to IgG MM plasma cells may allude to distinct modes of antigen exposure of MM progenitors in different microenvironmental niches.

N-glycosylation concerns the modification of IGs (or other proteins) by the addition of N-glycans to specific sequons; this modification can have a significant impact on the structure, stability, and biological function of the glycosylated molecules (46). Novel N-glycosylation sites introduced by SHM in the clonotypic IG are a very frequent feature in follicular lymphoma (FL) (~80%) and diffuse large B-cell lymphoma (DLCL) (51% in germinal-center B-cell–like DLBCL and 6% in activated B-cell–like DLBCL) (47–50). Furthermore, N-glycosylation of the clonotypic BcR IG has also been reported, albeit less frequently, in plasma cell disorders and B cell malignancies, such as AL amyloidosis (51–53). Prompted by evidence that SHM-induced changes in N-glycosylation of the IG variable domains is a mechanism redeeming B cells away from autoreactivity (54), here we sought to explore whether SHM within the clonotypic VH domain may lead to the creation of novel N-glycosylation sites or the disruption of germline-encoded ones. This is of potential clinical relevance if one also considers the generally low incidence of autoreactive phenomena in MM. Interestingly, examination of the particular SHMs leading to the creation and/or disruption of N-glycosylation sequons revealed the existence of recurrent amino acid substitutions at particular positions in certain IGHV genes (e.g. IGHV1-46 and IGHV4-39), albeit with no distinguishing features between IgA MM versus IgG MM.

In conclusion, in-depth analysis of the IG gene repertoire in the largest so far series of MM patients offers evidence supporting a role for antigens in disease pathogenesis. Distinct immunogenetic profiles in IgA MM versus IgG MM allude to distinct ontogenetic trajectories for particular subtypes of MM with potential impact on clonal behavior and clinical outcome.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

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

Author contributions

GG performed research, analyzed data and wrote the paper. AA supervised research and wrote the paper; GK performed statistical analysis. CL, AP, AM, EG performed research. KC, EH, MP, ET, CJ, IS, SF, ML, RS, CB provided samples and supervised research. KS designed the study, supervised research and wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This project was supported in part by the "Quant-ALL" project funded by Greece and the European Union with grant agreement no. MIS 5053938 and 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 co-financed by Greece and the European Union with grant agreement no. MIS 5002462. This research was also co-financed by Greece and the European Union (European Regional Development Fund) through the Operational Programme "Competitiveness, Entrepreneurship, and Innovation" (NSRF 2014-2020) in the context of the project "BBMRI: Biobanking and Biomolecular Resources Research Infrastructure", implemented under the action "Reinforcement of the Research and Innovation Infrastructure" (MIS5028275).

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.

KS has received grant support from Abbvie: Honoraria, Travel expenses, Janssen: Honoraria, Travel expenses, Research funding. ET has received grant support from GSK: Honoraria, Research funding, Novartis: Honoraria, Genesis: Honoraria, Research funding, BMS: Honoraria, Amgen: Honoraria, Travel expenses, Research funding, EUSA Pharma: Honoraria, Travel expenses, Takeda: Honoraria, Travel expenses, Research funding, Janssen: Honoraria, Research funding; Sanofi: Honoraria, Research funding. SF has received grant support from Janssen: Consultancy, Honoraria, Research funding, Advisory board, EUSA Pharma: Consultancy, Honoraria, Advisory board, Abbvie: Consultancy, Sandoz: Consultancy, Incyte: Advisory board, Itfarmaco: Advisory board, Clinigen: Advisory board, Morphosys: Research funding, Gilead: Research funding, Beigene: Research funding; Gentili: Honoraria.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1123029/full#supplementary-material

References

1. Chesi M, Bergsagel PL. Molecular pathogenesis of multiple myeloma: Basic and clinical updates. *Int J Hematol* (2013) 97(3):313–23. doi: 10.1007/s12185-013-1291-2

2. Bal S, Giri S, Godby KN, Costa LJ. Revisiting the impact of immunoglobulin isotypes in multiple myeloma. *Ann Hematol* (2022) 101(4):825–9. doi: 10.1007/s00277-022-04783-1

3. Taylor BJ, Kriangkum J, Pittman JA, Mant MJ, Reiman T, Belch AR, et al. Analysis of clonotypic switch junctions reveals multiple myeloma originates from a single class switch event with ongoing mutation in the isotype-switched progeny. *Blood* (2008) 112 (5):1894–903. doi: 10.1182/blood-2008-01-129221

4. Kosmas C, Stamatopoulos K, Stavroyianni N, Belessi C, Viniou N, Yataganas X. Molecular analysis of immunoglobulin genes in multiple myeloma. *Leuk Lymphoma* (1999) 33(3–4):253–65. doi: 10.3109/10428199909058425

5. Frøyland M, Thompson KM, Thorpe SJ, Sahota SS, Gedde-Dahl T, Bogen B. A VH4-34+ myeloma protein with weak autoreactivity. *Haematologica* (2007) 92(5):690–3. doi: 10.3324/haematol.10850

6. González D, González M, Balanzategui A, Sarasquete ME, López-Pérez R, Chillón MC, et al. Molecular characteristics and gene segment usage in IGH gene rearrangements in multiple myeloma. *Haematologica* (2005) 90(7):906–13.

7. Hadzidimitriou A, Stamatopoulos K, Belessi C, Lalayianni C, Stavroyianni N, Smilevska T, et al. Immunoglobulin genes in multiple myeloma: Expressed and non-expressed repertoires, heavy and light chain pairings and somatic mutation patterns in a series of 101 cases. *Haematologica* (2006) 91(6):781–7.

8. Cowan G, Weston-Bell NJ, Bryant D, Seckinger A, Hose D, Zojer N, et al. Massive parallel IGHV gene sequencing reveals a germinal center pathway in origins of human multiple myeloma. *Oncotarget* (2015) 6(15):13229–40. doi: 10.18632/oncotarget.3644

9. Bakkus MHC, Heirman C, Van Riet I, Van Camp B, Thielemans K. Evidence that multiple myeloma ig heavy chain VDJ genes contain somatic mutations but show no intraclonal variation. *Blood* (1992) 80(9):2326–35. doi: 10.1182/blood.V80.9.2326.2326

10. Berkowska MA, Driessen GJA, Bikos V, Grosserichter-Wagener C, Stamatopoulos K, Cerutti A, et al. Human memory b cells originate from three distinct germinal center-dependent and -independent maturation pathways. *Blood* (2011) 118(8):2150–8. doi: 10.1182/blood-2011-04-345579

11. Benckert J, Schmolka N, Kreschel C, Zoller MJ, Sturm A, Wiedenmann B, et al. The majority of intestinal IgA+ and IgG+ plasmablasts in the human gut are antigen-specific. *J Clin Invest* (2011) 121(5):1946–55. doi: 10.1172/JCI44447

12. IJspeert H, van Schouwenburg PA, van Zessen D, Pico-Knijnenburg I, Driessen GJ, Stubbs AP, et al. Evaluation of the antigen-experienced b-cell receptor repertoire in healthy children and adults. *Front Immunol* (2016) 7(OCT):1–13. doi: 10.3389/fimmu.2016.00410

13. Schickel J-N, Glauzy S, Ng Y-S, Chamberlain N, Massad C, Isnardi I, et al. Self-reactive VH4-34-expressing IgG b cells recognize commensal bacteria. *J Exp Med* (2017) 214(7):1991–2003. doi: 10.1084/jem.20160201

14. Scheid JF, Mouquet H, Kofer J, Yurasov S, Nussenzweig MC, Wardemann H. Differential regulation of self-reactivity discriminates between IgG+ human circulating memory b cells and bone marrow plasma cells. *Proc Natl Acad Sci USA* (2011) 108 (44):18044–8. doi: 10.1073/pnas.1113395108

15. Sato S, Kamata W, Okada S, Tamai Y. Clinical and prognostic significance of t (4; 14) Translocation in multiple myeloma in the era of novel agents. *Int J Hematol* (2021) 113(2):207–13. doi: 10.1007/s12185-020-03005-6

16. Chang H, Sloan S, Li D, Zhuang L, Yi Q-L, Chen CI, et al. The t (4;14) is associated with poor prognosis in myeloma patients undergoing autologous stem cell transplant. *Br J Haematol* (2004) 125(1):64–8. doi: 10.1111/j.1365-2141.2004.04867.x

17. Kumar S, Paiva B, Anderson KC, Durie B, Landgren O, Moreau P, et al. International myeloma working group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncol* (2016) 17(8):e328–46. doi: 10.1016/S1470-2045(16)30206-6

18. Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, et al. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res* (2006) 34(Database issue):781–4. doi: 10.1093/nar/gkj088

19. Ferrero S, Capello D, Svaldi M, Boi M, Gatti D, Drandi D, et al. Multiple myeloma shows no intra-disease clustering of immunoglobulin heavy chain genes. *Haematologica* (2012) 97(6):849–53. doi: 10.3324/haematol.2011.052852

20. Medina A, Jiménez C, Sarasquete ME, González M, Chillón MC, Balanzategui A, et al. Molecular profiling of immunoglobulin heavy-chain gene rearrangements unveils new potential prognostic markers for multiple myeloma patients. *Blood Cancer J* (2020) 10(2). doi: 10.1038/s41408-020-0283-8

21. Trudel S, Ghamlouch H, Dremaux J, Delette C, Harrivel V, Marolleau JP, et al. The importance of an in-depth study of immunoglobulin gene rearrangements when ascertaining the clonal relationship between concomitant chronic lymphocytic leukemia and multiple myeloma. *Front Immunol* (2016) 7(DEC):1–7. doi: 10.3389/fimmu.2016.00625

22. 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) 111:1524–33. doi: 10.1182/blood-2007-07-099564

23. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: The highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* (2008) 36(Web Server issue):503–8. doi: 10.1093/nar/gkn316

24. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, et al. IMGT®, the international ImMunoGeneTics information system®. *Nucleic Acids Res* (2009) 37(SUPPL. 1):1006–12. doi: 10.1093/nar/gkn838

25. 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. doi: 10.1182/blood.2020007039

26. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* (2002), 310–22.

27. Varettoni M, Zibellini S, Capello D, Arcaini L, Rossi D, Pascutto C, et al. Clues to pathogenesis of waldenström macroglobulinemia and immunoglobulin m monoclonal gammopathy of undetermined significance provided by analysis of immunoglobulin heavy chain gene rearrangement and clustering of b-cell receptors. *Leuk Lymphoma* (2013) 54(11):2485–9. doi: 10.3109/10428194.2013.779689

28. Martín-Jiménez P, García-Sanz R, Balanzategui A, Alcoceba M, Ocio E, Sanchez ML, et al. Molecular characterization of heavy chain immunoglobulin gene rearrangements in waldenström's macroglobulinemia and IgM monoclonal gammopathy of undetermined significance. *Haematologica* (2007) 92(5):635–42. doi: 10.3324/haematol.10755

29. Sahota SS, Leo R, Hamblin TJ, Stevenson FK. Ig VH gene mutational patterns indicate different tumor cell status in human myeloma and monoclonal gammopathy of undetermined significance. *Blood* (1996) 87(2):746–55. doi: 10.1182/blood.V87.2.746.bloodjournal872746

30. Stamatopoulos K, Belessi C, Moreno C, Boudjograh 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. doi: 10.1182/blood-2006-03-012948

31. Hadzidimitriou A, Agathangelidis A, Darzentas N, Murray F, Delfau-Larue MH, Pedersen LB, et al. Is there a role for antigen selection in mantle cell lymphoma? immunogenetic support from a series of 807 cases. *Blood* (2011) 118(11):3088–95. doi: 10.1182/blood-2011-03-343434

32. Wu YCB, Kipling D, Dunn-Walters DK. The relationship between CD27 negative and positive b cell populations in human peripheral blood. *Front Immunol* (2011) 2 (DEC). doi: 10.3389/fimmu.2011.00081

33. Kitaura K, Yamashita H, Ayabe H, Shini T, Matsutani T, Suzuki R. Different somatic hypermutation levels among antibody subclasses disclosed by a new next-generation sequencing-based antibody repertoire analysis. *Front Immunol* (2017) 8 (MAY):1–11. doi: 10.3389/fimmu.2017.00389

34. Shi B, Ma L, He X, Wang X, Wang P, Zhou L, et al. Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST. *Theor Biol Med Model* (2014) 11:30. doi: 10.1186/1742-4682-11-30

35. Soto C, Bombardi RG, Branchizio A, Kose N, Matta P, Sevy AM, et al. High frequency of shared clonotypes in human b cell receptor repertoires. *Nature* (2019) 566 (7744):398–402. doi: 10.1038/s41586-019-0934-8

36. Agathangelidis A, Hadzidimitriou A, Rosenquist R, Stamatopoulos K. Unlocking the secrets of immunoglobulin receptors in mantle cell lymphoma: implications for the origin and selection of the malignant cells. *Semin Cancer Biol* (2011) 21(5):299–307. doi: 10.1016/j.semcancer.2011.09.009

37. Darzentas N, Hadzidimitriou A, Murray F, Hatzi K, Josefsson P, Laoutaris N, et al. A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. *Leukemia* (2010) 24(1):125–32. doi: 10.1038/leu.2009.186

38. 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

39. 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

40. Zibellini S, Capello D, Forconi F, Marcatili P, Rossi D, Rattotti S, et al. Stereotyped patterns of b-cell receptor in splenic marginal zone lymphoma. *Haematologica* (2010) 95 (10):1792–6. doi: 10.3324/haematol.2010.025437

41. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* (2000) 13(1):37–45. doi: 10.1016/S1074-7613(00)00006-6

42. Rettig MB, Vescio RA, Cao J, Wu CH, Lee JC, Han E, et al. VH gene usage in multiple myeloma: Complete absence of the VH4.21 (VH4-34) gene. *Blood [Internet]* (1996) 87(7):2846–52.

43. González D, van der Burg M, García-Sanz R, Fenton JA, Langerak AW, González M, et al. Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. *Blood* (2007) 110(9):3112–21. doi: 10.1182/blood-2007-02-069625

44. Vescio RA, Cao J, Hong CH, Lee JC, Wu CH, Der Danielian M, et al. Myeloma ig heavy chain V region sequences reveal prior antigenic selection and marked somatic mutation but no intraclonal diversity. *J Immunol* (1995) 155(5):2487–97. doi: 10.4049/jimmunol.155.5.2487

45. Yeap LS, Hwang JK, Du Z, Meyers RM, Meng FL, Jakubauskaite A, et al. Sequence-intrinsic mechanisms that target AID mutational outcomes on antibody genes. *Cell* (2015) 163(5):1124–37. doi: 10.1016/j.cell.2015.10.042

46. Esmail S, Manolson MF. Advances in understanding n-glycosylation structure, function, and regulation in health and disease. *Eur J Cell Biol* (2021) 100(7–8). doi: 10.1016/j.ejcb.2021.151186

47. McCann KJ, Johnson PWM, Stevenson FK, Ottensmeier CH. Universal n-glycosylation sites introduced into the b-cell receptor of follicular lymphoma by somatic mutation: a second tumorigenic event? *Leukemia England;* (2006) 20:530–4. doi: 10.1038/sj.leu.2404095

48. Odabashian M, Carlotti E, Araf S, Okosun J, Spada F, Gribben JG, et al. IGHV sequencing reveals acquired n-glycosylation sites as a clonal and stable event during follicular lymphoma evolution. *Blood* (2020) 135(11):834–44. doi: 10.1182/blood.2019002279

49. Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. Acquisition of potential n-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. *Blood* (2002) 99(7):2562–8. doi: 10.1182/blood.V99.7.2562

50. Chiodin G, Allen JD, Bryant DJ, Rock P, Martino EA, Valle-Argos B, et al. Insertion of atypical glycans into the tumor antigen-binding site identifies DLBCLs with distinct origin and behavior. *Blood [Internet]* (2021) 138(17):1570–82. doi: 10.1182/blood.2021012052

51. Dispenzieri A, Larson DR, Rajkumar SV, Kyle RA, Kumar SK, Kourelis T, et al. N-glycosylation of monoclonal light chains on routine MASS-FIX testing is a risk factor for MGUS progression. *Leukemia* (2020) 34(10):2749–53. doi: 10.1038/s41375-020-0940-8

52. Kourelis T, Murray DL, Dasari S, Kumar S, Barnidge D, Madden B, et al. MASS-FIX may allow identification of patients at risk for light chain amyloidosis before the onset of symptoms. *Am J Hematol* (2018) 93(11):E368–70. doi: 10.1002/ajh.25244

53. Mellors PW, Dasari S, Kohlhagen MC, Kourelis T, Go RS, Muchtar E, et al. MASS-FIX for the detection of monoclonal proteins and light chain n-glycosylation in routine clinical practice: a cross-sectional study of 6315 patients. *Blood Cancer J [Internet]* (2021) 11(3). doi: 10.1038/s41408-021-00444-0

54. Sabouri Z, Schofield P, Horikawa K, Spierings E, Kipling D, Randall KL, et al. Redemption of autoantibodies on anergic b cells by variable-region glycosylation and mutation away from self-reactivity. *Proc Natl Acad Sci USA* (2014) 111(25):E2567–75. doi: 10.1073/pnas.1406974111

frontiers | Frontiers in Oncology

TYPE Perspective
PUBLISHED 08 February 2023
DOI 10.3389/fonc.2023.1107171

OPEN ACCESS

EDITED BY
Richard Rosenquist,
Karolinska Institutet (KI), Sweden

REVIEWED BY
Lorenzo Leoncini,
University of Siena, Italy

*CORRESPONDENCE
Patricia J. T. A. Groenen
Patricia.Groenen@radboudumc.nl

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

RECEIVED 24 November 2022
ACCEPTED 19 January 2023
PUBLISHED 08 February 2023

CITATION
Groenen PJTA, van den Brand M, Kroeze LI, Amir AL and Hebeda KM (2023) Read the clonotype: Next-generation sequencing-based lymphocyte clonality analysis and perspectives for application in pathology.
Front. Oncol. 13:1107171.
doi: 10.3389/fonc.2023.1107171

Read the clonotype: Next-generation sequencing-based lymphocyte clonality analysis and perspectives for application in pathology

Patricia J. T. A. Groenen[1,2*], Michiel van den Brand[1,3], Leonie I. Kroeze[1,2], Avital L. Amir[1] and Konnie M. Hebeda[1]

[1]Department of Pathology, Radboud University Medical Center, Nijmegen, Netherlands, [2]Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, Netherlands, [3]Pathology-DNA, Location Rijnstate Hospital, Arnhem, Netherlands

Clonality assessment using the unique rearrangements of immunoglobulin (IG) and T-cell receptor (TR) genes in lymphocytes is a widely applied supplementary test for the diagnosis of B-cell and T-cell lymphoma. To enable a more sensitive detection and a more precise comparison of clones compared with conventional clonality analysis based on fragment analysis, the EuroClonality NGS Working Group developed and validated a next-generation sequencing (NGS)-based clonality assay for detection of the IG heavy and kappa light chain and TR gene rearrangements for formalin-fixed and paraffin-embedded tissues. We outline the features and advantages of NGS-based clonality detection and discuss potential applications for NGS-based clonality testing in pathology, including site specific lymphoproliferations, immunodeficiency and autoimmune disease and primary and relapsed lymphomas. Also, we briefly discuss the role of T-cell repertoire of reactive lymphocytic infiltrations in solid tumors and B-lymphoma.

KEYWORDS

immunoglobulin E, T-cell receptor, gene rearrangement, clonality assessment, NGS, pathology

1 Introduction

Assessment of the clonality of the immunoglobulin (IG) and/or T-cell receptor (TR) genes is an important aid in the diagnosis of lymphoproliferative diseases. Lymphoma cells originate from a single transformed lymphoid cell and therefore all malignant cells have the same IG gene rearrangements in B-cell lymphoma (BCL) or TR rearrangements in T-cell lymphoma (TCL). Clonality testing uses this feature and facilitates the discrimination between clonally expanded cells and reactive cells with diverse IG and/or TR rearrangements. The BIOMED-2/EuroClonality assays for clonality testing have been used worldwide for more than a decade now. The strength of these assays is the complementarity of different PCR targets, which results in an unprecedented high detection rate (1–4).

Clinical clonality assessment is mainly used when discrimination between a lymphoma and a reactive lymphoid infiltrate is uncertain. Examples are low-grade lymphoma, including Bcl2-negative follicular lymphomas (5), B- or T-cell proliferations at specific sites, such as the skin, or in the context of immunodeficiency. Also lymphoproliferations with a combination of atypical B- and T-cells, such as angioimmunoblastic TCL, or lymphomas of unclear lineage, like some anaplastic large cell lymphoma, can benefit from clonality studies. In case of small biopsies clonality assessment can guide further diagnostics. It is, however, not always suitable to detect Hodgkin lymphoma, as the traditional EuroClonality/BIOMED-2 protocols may not be sufficiently sensitive to detect small numbers of malignant cells in a background of lymphoid cells, as is the case in most classical Hodgkin lymphomas (6–8). Also, it is important to note that lineage determination in acute lymphoblastic leukemia cannot be based on clonality assays alone, since neoplastic precursor B and T cells often show cross-lineage rearrangements (9). In addition, monoclonality does not always imply malignancy since some reactive processes contain clonal lymphocytic populations (10).

2 NGS-amplicon based clonality assessment

For the conventional clonality assays multiple IG targets, notably IGHV-IGHD-IGHJ (in FR 1,2,3), IGHD-IGHJ, IGKV-IGKJ, IGKV-KDE and Intron RSS-KDE, are analyzed. Likewise, TRGV-TRGJ, as well as TRBD-TRBJ and TRBV-TRBD-TRBJ are assessed for T-cell clonality (1). A polyclonal population harboring different V(D)J rearrangements gives rise to a range of differently sized PCR, resulting in a Gaussian distribution in fragment analysis (GeneScan) (Figure 1A). In case of a clonal population, there will be one or two dominant PCR products of a given size per target with GeneScan (Figure 1B). Guidelines for the uniform technical scoring of the individual PCR targets and the molecular conclusion of the entire sample, which is deduced from the results of the individual targets, have been developed (11). Finally, the result of the clonality assessment has to be interpreted in the context of the clinical presentation, other molecular studies, and the pathological findings.

Despite the good performance and the world-wide usage of these assays, they have some weaknesses that potentially yield (mainly) false-negative results. The BIOMED-2/EuroClonality assays have been designed for high-quality DNA samples generating amplicons in the range of 150–400 bp (1) (Figure 1). However, formalin-fixed paraffin-embedded (FFPE) tissue specimens generally used in a diagnostic setting, are affected by DNA crosslinking and fragmentation, resulting in DNA samples of suboptimal quality with short DNA fragments. Clonal rearrangements with longer amplicons may potentially go unnoticed in samples of suboptimal DNA quality. In addition, the detection of a small clonal rearrangement in a background of polyclonal B-cells is dependent on the position of the clonal product within the Gaussian curve; it can be entirely hidden in the polyclonal background. Furthermore, for clonal comparison only the size of the PCR fragments is used (Figure 1). Since different rearrangements may result in the same size PCR fragments, this may hamper interpretation, especially in cases in which a single rearranged target is detected.

To tackle these issues, the EuroClonality NGS-working group has developed next-generation sequencing (NGS)-based clonality assays for detection of IG and TR gene rearrangements (12, 13), which can be analyzed with the bioinformatics tool ARResT/Interrogate (14). They are based on the use of gene-specific primers and importantly,

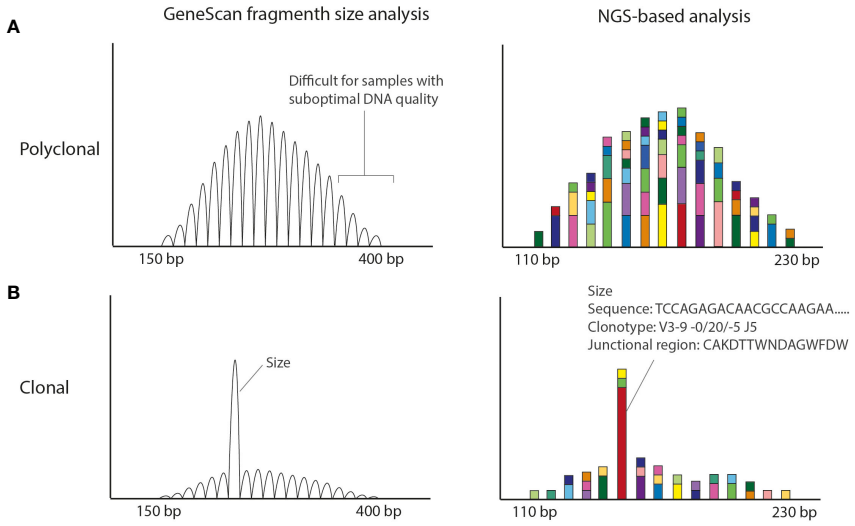


FIGURE 1
Conventional EuroClonality/BIOMED-2 fragment size vs NGS-amplicon based clonality analysis. **(A)** Polyclonal pattern observed in Genescan and NGS-based analysis. The size of the amplicons is smaller in the NGS-based approach making the technique more suitable for smaller DNA fragments obtained from FFPE material. **(B)** Clonal result (with low polyclonal background). Using GeneScan analysis only the size of the rearrangement is known, whereas with NGS-amplicon based clonality also the nucleotide sequence is obtained. The Bioinformatic program ARResT/Interrogate processes the nucleotide sequences into clonotypes, in which the used V, D and J genes as well as the junctional region in amino acids are defined.

on the generation of shorter amplicon sizes, which makes them more suitable for clonality detection in samples of suboptimal DNA-quality, like FFPE-samples. NGS-based clonality assays provide the nucleotide sequences of all IG and/or TR rearrangements, both from the malignant lymphoid cells and from the non-malignant background cells (Figure 1).

Bioinformatic software analyzes each sequence for the presence of V, (D) and J genes and the junctional region containing the complementary determining region 3, which is represented as an amino acid sequence. With this information the sequences are attributed to clonotypes. A clonotype is characterized by the same V and J gene and junctional region. Therefore reliable detection of minor clonal rearrangements is possible, resulting in a high sensitivity (12, Figure 2).

NGS-based IG clonality assessment showed a very good performance in a cohort of BCL, the vast majority of which were FFPE samples (15). In diagnostic FFPE samples of classical Hodgkin lymphoma (cHL), NGS-based detection of IG rearrangements was more accurate and sensitive to detect clonal rearrangements compared to the conventional BIOMED-2/EuroClonality assay (16). Because of the sequence information and the high sensitivity, the NGS-based clonality assays are very valuable to solve complex rearrangement patterns (17). They allow detailed comparison of sequential lesions or multiple lymphomas at different locations within a patient (18). The technical improvements due to the advent of NGS will allow higher sensitivity, more detailed analysis and broader applications, as will be discussed below.

3 Potential clinical applications

In various clinical situations the advantages of NGS-based clonality analysis can be exploited (Figure 2).

3.1 Inflammatory skin diseases

One important use of clonality analysis for the differentiation between inflammatory disease and lymphoma is in the diagnostic workup of mycosis fungoides (MF), the most common cutaneous TCL (19, 20) with a rising incidence (21). The diagnosis of MF can be difficult due to morphological overlap with various inflammatory skin diseases (ISD). Especially the early-stage of MF is both clinically and histologically difficult to distinguish from ISDs such as eczema, psoriasis and cutaneous lupus erythematosus (19, 20, 22–24), resulting in a time to diagnosis of 3 to 4 years after the first lesions appear.

Immunohistochemistry can be very helpful in the diagnosis of MF, but is not always decisive. When the diagnosis of MF is uncertain based on clinical picture, histology and immunohistochemistry, TR clonality assessment can be used to help discriminate between MF and ISDs. However, especially in early-stage MF, the lesions often contain a relatively small number of neoplastic T-cells admixed with a relatively high number of polyclonal reactive T-cells. In the conventional clonality assay a small monoclonal population can be difficult or impossible to detect in a high polyclonal background, potentially hampering the diagnosis of early-stage MF (25–27). Since

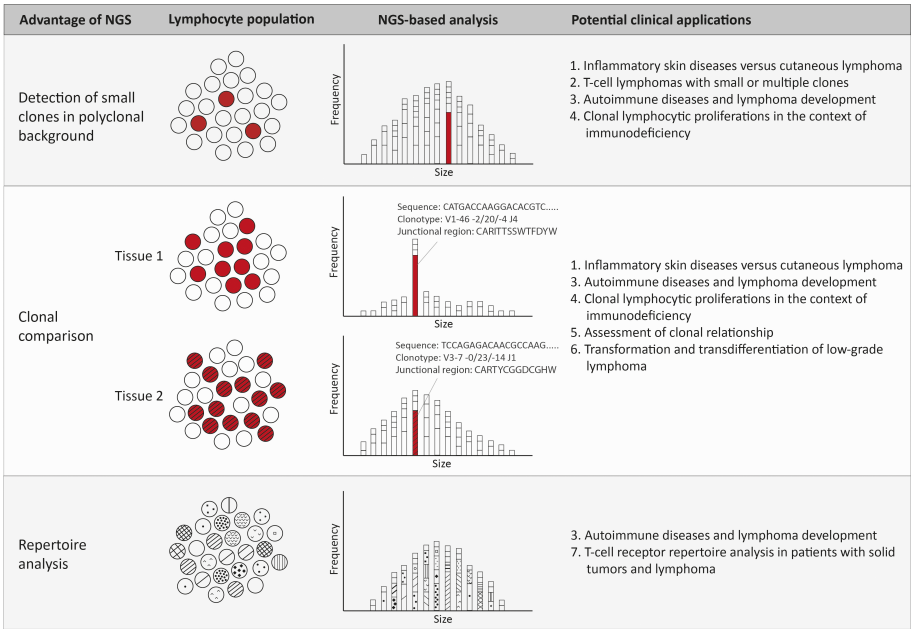


FIGURE 2
Advantages and potential clinical applications of NGS-amplicon based clonality analysis. Using NGS-amplicon based clonality analysis, small clones in a polyclonal background, or even hidden in the polyclonal background, can be more confidently detected. A second advantage of NGS is the sequence information and therefore the clonotype information, which is very useful for rapid and reliable clonal comparison. The clonotype obtained from the nucleotide sequence reveals the V, (D) and J gene as well as the junctional region that allows quick and efficient comparison of clonotypes of two or more specimen. Suspected ongoing somatic hypermutation in one of the specimens requires in depth investigation of the nucleotide sequences. The clonotypes also enable detection of the repertoire of the lymphocytes, including early detection of possibly malignant clonotypes.

biopsies of ISDs can also contain a clonal population in a polyclonal background (28), the ultimate proof of MF is considered to be the presence of an identical dominant clonal population in two biopsies derived from different anatomical locations (29). In the conventional clonality assay, comparison of the clonality between two different samples is based on the fragment size of the detected clonal peak, which can be challenging and does not always represent the same rearrangement.

NGS-based clonality analysis can most likely overcome these obstacles, since the sequence information of the TR gene rearrangements can be used to identify small relevant clones amidst a polyclonal background and may identify early stage disease (30, 31). Another advantage of analyzing the sequences is comparison of the clones in biopsies from different lesions and multiple time points for identification of recurrent MF (30). The high sensitivity of NGS-based clonality also carries a risk of false-positive results. Therefore studies are warranted to determine the cut-off points that separate MF from reactive conditions before NGS-based clonality can be introduced in the diagnostic workup of suspected MF.

3.2 T-cell lymphomas

The diagnosis of TCL by histology is often straightforward, but certain TTCL subtypes regularly cause diagnostic problems. The above mentioned cutaneous lymphomas are one example, but another example is the group of TCL that are derived from CD4 positive follicular helper T-cells that normally are resident in germinal centers (32). This group includes angioimmunoblastic TCL (AITL), the most frequent nodal mature TCL in elderly patients in western countries (33). AITL is clinically characterized by immune dysregulation and autoimmunity related symptoms. Histology usually shows lymph nodes with a prominent reactive background of CD4 and CD8 positive T-cells and proliferating follicular dendritic cell meshworks, accompanied by a variable proliferation of B- and plasma cells, often EBV-driven (32, 34). In this situation a T-cell clone combined with a B-cell clone is frequently seen (35). In case of a small T-cell clone or several T- and B-cell clones, which is in our experience and according to literature (36) not uncommon in AITL, NGS-based tests will be more informative than conventional clonality analysis, since NGS can better identify small clones that are hidden in a polyclonal background of T-cells. Bi-clonal and oligoclonal AITL were identified in a significant number of cases in a study applying NGS-based clonality testing, reflecting AITL evolution from a mutated hematopoietic progenitor pool and the subsequent exposure to the complex and dynamic environment of the germinal center (37). This knowledge of the biology of AITL can help to unravel the composition of the infiltrate and support the diagnosis in challenging cases of AITL.

3.3 Autoimmune diseases

Autoimmune diseases are associated with a significantly increased risk of lymphoma development, as reviewed recently (38). One group with a very high risk is patients with Sjögrens syndrome (SjS), who suffer from chronic lymphoid infiltration in the salivary glands, causing destruction of glandular structures leading to atrophy. In this context, biopsies are notorious for the difficult discrimination between lymphoepithelial sialadenitis and the onset of marginal zone lymphoma (39). Indeed, the ectopic formation of lymphoid tissue with functional germinal centers in the salivary glands is related to both Sjs disease activity, and the risk for lymphoma development (40). But the mere detection of clonal B-cell proliferations in salivary gland biopsies does not correlate with the subsequent development of BCL, as was shown in a retrospective series of 49 Sjs patients with [n=21] or without [n=28] lymphoma development (41). 18% of the patients without subsequent BCL development showed clonal Ig rearrangements in the minor salivary glands by conventional clonality analysis.

More promising is the analysis of the IG repertoire in the inflamed tissues by NGS-based clonality analysis. This gives insight in the composition of the infiltrate and enables quantification of large numbers of IG clonotypes, including usage of V and J genes. Mainly high-affinity stereotypic rheumatoid factor producing B-cells from the autoreactive, often oligoclonal to polyclonal environment, show clonal evolution to BCL (42–44). Early detection and monitoring of these clones in the background of inflammation, by NGS-based clonality testing, is expected to be useful in the management of Sjs patients at high risk of lymphoma development.

3.4 Immunodeficiency

Lymphoid proliferations occurring in the background of immunodeficiency (ID) can involve virtually all tissues and range from benign lymphoid proliferations to full-blown lymphomas (45, 46). The main cause is T-cell suppression or dysfunction, caused by a germline defect in primary ID or induced by infection or therapy in secondary ID. Because of the morphological overlap between infections, lymphoma and non-infectious reactive lymphoproliferations in biopsies of patients with ID, clonality assessment can be helpful in predicting outcome (47) and guiding therapy (48). ID can cause complex proliferations of both T- and B-cells as a result of varying stimuli and transforming events, in which EBV plays an important role (49). This can result in oligoclonal expansions and subclones, which can develop into clonally related lymphomas (50). Since treatment is tailored to the immune status, the underlying trigger, the aggressiveness of the lymphoproliferation and whether it is a recurrent disease, it is important to unravel clonal relationships between simultaneous or subsequent biopsies, which can be facilitated by NGS-based clonality assessment.

3.5 Assessment of clonal relationship

In patients with synchronous or metachronous lymphomas, the treatment can depend on whether these lymphomas are clonally related. Comparison of the morphology and immunophenotype of the primary lymphoma and the relapse cannot always be used as a surrogate for clonal comparison, as clonally related relapses can show a different morphology or immunophenotype (51, 52) and clonally unrelated relapses can look similar to the primary lymphoma. In cHL, unrelated relapse can be suspected by a change of histologic subtype or an altered EBV association, but this is not absolute (53). Therefore, clonal comparison is an advised in lymphoma relapse with a long interval.

With conventional clonality analysis, the sizes of the clonal products in at least two targets are compared to assess clonal relationship. If a single clonal product is available for assessment the result remains uncertain (54), since two different rearrangements can coincidentally have the same size. Contrary, differences in peak sizes can be a result of somatic hypermutation within a clone, rather than a different clonal origin (55). NGS-based clonality analysis markedly improves the assessment of clonal relationship as it allows comparison of the actual sequence, enabling easy assessment of a clonal relationship even from a single target.

The incidence of second primary lymphomas is not well known since the studies are often small with a focus on specific types of lymphoma and with a selection for recurrences after a long interval. In some larger studies investigating diffuse large B-cell lymphoma (DLBCL) with a long interval between diagnosis and relapse up to 25% of relapsed DLBCL were actually unrelated (56–58).

In cHL it is even more difficult to investigate clonal relationship between primary and relapsed lymphoma due to the scarcity of tumor cells. A study of 20 patients with relapsed cHL showed 40% clonally unrelated tumors (53). In this study, the samples were enriched for tumor cells with laser microdissection, a laborious technique which is not suitable for routine diagnostics. NGS-based clonality analysis is more suitable to detect clonal products in cHL in comparison with conventional clonality analysis without laser microdissection (16, 59). This opens up possibilities to evaluate the clonal relationship in relapsed cHL in routine diagnostics.

3.6 Transformation and transdifferentiation

In patients with low-grade BCL who develop a high-grade lymphoma, it can be important for treatment decisions and assessment of prognosis to determine if the transformed lymphoma is indeed related to the low-grade lymphoma, or whether it is a *de novo* high-grade lymphoma. This has been most extensively studied in chronic lymphocytic leukemia (CLL) where a limited percentage of patients develops a secondary aggressive lymphoma, usually DLBCL. Transformation to cHL is infrequent. DLBCL development is more often clonally related, showing identical IG rearrangement to the CLL, than cHL (~70% vs 40%) (60, 61). Patients with unrelated lymphoma have a better prognosis and are therefore treated differently than patients with clonally related lymphomas (62). The investigation of the clonal relationship is thus of clinical importance. As discussed above, the currently used clonality analysis has limitations in its ability to demonstrate clonal relationship, which can be overcome by NGS-based clonality.

In follicular lymphoma, transformation usually results in DLBCL or high-grade B-cell lymphoma with *MYC* and *BCL2* rearrangements (63). Transformation to cHL is rare (64). A diagnosis of transformed FL is usually made based on an assumed clonal relationship. Indeed, a clonal relationship between the initial FL and the transformation has only been established in small case series, but large studies are lacking (65–68).

In transdifferentiation, or lineage reprogramming, lymphoma cells acquire additional genetic aberrations leading to loss of B- or T-cell specific transcription factors. The resulting tumors show expression of a myeloid differentiation program. This causes a change of the morphology and phenotype from a lymphoma cell to a myeloid cell, most commonly resulting in a histiocytic or dendritic sarcoma (69). Since these sarcomas usually retain the original IG or TR rearrangements they can be clonally linked to the previous lymphoma.

3.7 TR repertoire analysis in patients with solid tumors and lymphoma

Immune escape represents an important mechanism in cancer development. Immune-checkpoint inhibition therapies directed against inhibitory checkpoint molecules such as PD-1 and CTLA-4 have become standard-of-care for several types of tumors, such as stage III or IV melanoma (70, 71), resectable lung cancer (72, 73), and colorectal cancer (74, 75). They exploit re-activation of T-cells that target neo-antigens arising from mutations in tumor cells. Analyzing the TR repertoire in a sample of interest can be performed using NGS (76). The dynamics of tumor and treatment related T-cell clones are an areas of intense research. In melanoma, breast and colon cancer, expanded T-cell clones are associated with response to immunotherapy treatment (77–79). In BCL, a restricted TR repertoire is found to be associated with poor outcome in DLBCL treated without immune-checkpoint inhibition (80) and in high grade B-cell lymphomas (81). Obviously, more studies are needed for different lymphoma types and treatments to understand the full potential of TR repertoire analysis in BCL.

4 Concluding remarks

The advent of NGS-based clonality analysis opens new possibilities for pathologists to define malignant lymphoproliferations in challenging clinical and histological situations and to discover clonal relationships between populations of lymphocytes in diverse infiltrates. B- and T-cell repertoire analysis in tissues in the context of immunodeficiency, autoimmune disease, lymphoma or solid tumors is a yet to be explored field with the potential to enable early detection of lymphoma development or prediction of therapy response.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

PG and KH designed and coordinated the manuscript. PG, MvdB, AA, and KH wrote the manuscript, LK made the figures and edited the manuscript. All co-authors take responsibility for the integrity of the literature study and all co-authors critically revised the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

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

References

1. van Dongen JJ, Langerak AW, Brüggemann M, Evans ,PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* (2003) 17:2257–317. doi: 10.1038/sj.leu.2403202

2. Evans PA, Pott C, Groenen PJ, Salles G, Davi F, Berger F, et al. Significantly improved PCR-based clonality testing in b-cell malignancies by use of multiple immunoglobulin gene targets. Report of the BIOMED-2 concerted action BHM4-CT98-3936. *Leukemia* (2007) 21:207–14. doi: 10.1038/sj.leu.2404479

3. Brüggemann M, White H, Gaulard P, Garcia-Sanz R, Gameiro P, Oeschger S, 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) 21:215–21. doi: 10.1038/sj.leu.2404481

4. van Krieken JH, Langerak AW, Macintyre EA, Kneba M, Hodges E, Sanz RG, et al. Improved reliability of lymphoma diagnostics *via* PCR-based clonality testing: Report of the BIOMED-2 concerted action BHM4-CT98-3936. *Leukemia* (2007) 21:201–6. doi: 10.1038/sj.leu.2404467

5. Schraders M, de Jong D, Kluin P, Groenen P, van Krieken H. Lack of bcl-2 expression in follicular lymphoma may be caused by mutations in the BCL2 gene or by absence of the t(14;18) translocation. *J Pathol* (2005) 205:329–35. doi: 10.1002/path.1689

6. Chute DJ, Cousar JB, Mahadevan MS, Siegrist KA, Silverman LM, Stoler MH. Detection of immunoglobulin heavy chain gene rearrangements in classic Hodgkin lymphoma using commercially available BIOMED-2 primers. *Diagn Mol Pathol* (2008) 17:65–72. doi: 10.1097/PDM.0b013e318150d695

7. Hebeda KM, Van Altena MC, Rombout P, Van Krieken JH, Groenen PJ. PCR clonality detection in Hodgkin lymphoma. *J Hematop* (2009) 2:34–41. doi: 10.1007/s12308-009-0024-1

8. Burack WR, Laughlin TS, Friedberg JW, Spence JM, Rothberg PG. PCR assays detect b-lymphocyte clonality in formalin-fixed, paraffin-embedded specimens of classical hodgkin lymphoma without microdissection. *Am J Clin Pathol* (2010) 134:104–11. doi: 10.1309/AJCPK6SBE0XOODHB

9. Dawidowska M, Jółkowska J, Szczepański T, Derwich K, Wachowiak J, Witt M. Implementation of the standard strategy for identification of Ig/TCR targets for minimal residual disease diagnostics in b-cell precursor ALL pediatric patients: Polish experience. *Arch Immunol Ther Exp (Warsz)* (2008) 56(6):409–18. doi: 10.1007/s00005-008-0045-y

10. Langerak AW, Molina TJ, Lavender FL, Pearson D, Flohr T, Sambade C, et al. Polymerase chain reaction-based clonality testing in tissue samples with reactive lymphoproliferations: Usefulness and pitfalls. a report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* (2007) 21:222–9. doi: 10.1038/sj.leu.2404482

11. Langerak AW, Groenen PJ, Brüggemann M, Beldjord K, Bellan C, Bonello L, et al. EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. *Leukemia* (2012) 26:2159–71. doi: 10.1038/leu.2012.246

12. Scheijen B, Meijers RWJ, Rijntjes J, van der Klift MY, Möbs M, Steinhilber J, et al. EuroClonality-NGS working group: Next-generation sequencing of immunoglobulin gene rearrangements for clonality assessment: A technical feasibility study by EuroClonality-NGS. *Leukemia* (2019) 33:2227–40. doi: 10.1038/s41375-019-0508-7

13. Brüggemann M, Kotrová M, Knecht H, Bartram J, Boudjogrha M, Bystry V, et al. EuroClonality-NGS working group. standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* (2019) 33:2241–53. doi: 10.1038/s41375-019-0496-7

14. Bystry V, Reigl T, Krejci A, Demko M, Hanakova B, Grioni A, et al. EuroClonality-NGS: ARResT/Interrogate: An interactive immunoprofiler for IG/TR NGS data. *Bioinformatics* (2017) 33:435–7. doi: 10.1093/bioinformatics/btw634

15. van den Brand M, Rijntjes J, Möbs M, Steinhilber J, van der Klift MY, Heezen KC, et al. Next-generation sequencing-based clonality assessment of ig gene rearrangements: A multicenter validation study by EuroClonality-NGS. *J Mol Diagn* (2021) 23:1105–15. doi: 10.1016/j.jmoldx.2021.06.005

16. van Bladel DAG, van den Brand M, Rijntjes J, Pamidimarri Naga S, Haacke DLCM, Luijks JACW, et al. Clonality assessment and detection of clonal diversity in classic Hodgkin lymphoma by next-generation sequencing of immunoglobulin gene rearrangements. *Mod Pathol* (2022) 35:706–7. doi: 10.1038/s41379-021-00986-5

17. Leenders AM, Kroeze LI, Rijntjes J, Luijks J, Hebeda KM, Darzentas N, et al. Multiple immunoglobulin κ gene rearrangements within a single clone unraveled by next-generation sequencing-based clonality assessment. *J Mol Diagn* (2021) 23:1097–104. doi: 10.1016/j.jmoldx.2021.05.002

18. Kroeze LI, Scheijen B, Hebeda KM, Rijntjes J, Luijks JACW, Evers D, et al. PAX5 P80R-mutated b-cell acute lymphoblastic leukemia with transformation to histiocytic sarcoma: Clonal evolution assessment using NGS-based immunoglobulin clonality and mutation analysis. *Virchows Arch* (2022) 20230124. doi: 10.1007/s00428-022-03428-y

19. Hristov AC, Tejasvi T, A Wilcox R. Cutaneous T-cell lymphomas: 2021 update on diagnosis, risk-stratification, and management. *Am J Hematol* (2021) 96(10):1313–28. doi: 10.1002/ajh.26299

20. Hodak E, Amitay-Laish I. Mycosis fungoides: A great imitator. *Clin Dermatol* (2019) 37(3):255–67. doi: 10.1016/j.clindermatol.2019.01.004

21. Ottevanger R, de Bruin DT, Willemze R, Jansen PM, Bekkenk MW, de Haas ERM, et al. Incidence of mycosis fungoides and sézary syndrome in the Netherlands between 2000 and 2020. *Br J Dermatol* (2021) 185(2):434–5. doi: 10.1111/bjd.20048

22. Martínez-Escala M,E, González BR. Mycosis fungoides variants. *Surg Pathol Clin* (2014) 7(2):169–89. doi: 10.1016/j.path.2014.02.003

23. Girardi M, Heald PW, Wilson LD. The pathogenesis of mycosis fungoides. *N Engl J Med* (2004) 350(19):1978–88. doi: 10.1056/NEJMra032810

24. Zackheim HS, McCalmont TH. Mycosis fungoides: the great imitator. *J Am Acad Dermatol* (2002) 47(6):914–8. doi: 10.1067/mjd.2002.124696

25. Jawed SI, Myskowski PL, Horwitz S, Moskowitz A, Querfeld C. Primary cutaneous T-cell lymphoma (mycosis fungoides and sézary syndrome): Part i. diagnosis: clinical and histopathologic features and new molecular and biologic markers. *J Am Acad Dermatol* (2014) 70(2):205.e1–16. doi: 10.1016/j.jaad.2013.07.049

26. Ponti R, Fierro MT, Quaglino P, Lisa B, Paola di Celle F, Michela O. TCRgamma-chain gene rearrangement by PCR-based GeneScan: diagnostic accuracy improvement and clonal heterogeneity analysis in multiple cutaneous T-cell lymphoma samples. *J Invest Dermatol* (2008) 128(4):1030–8. doi: 10.1038/sj.jid.5701109

27. Ritz N, Sahar D, Bergman R. T-Cell receptor gene rearrangement studies using the GeneScan technique as an adjunct to the histopathological diagnosis of mycosis fungoides. *Am J Dermatopathol* (2015) 37(3):210–3. doi: 10.1097/DAD.0000000000000204

28. Plaza JA, Morrison C, Magro CM. Assessment of TCR-β clonality in a diverse group of cutaneous T-cell infiltrates. *J Cutaneous Pathol* (2008) 35(4):358–65. doi: 10.1111/j.1600-0560.2007.00813.x

29. Thurber SE, Zhang B, Kim YH, Schrijver I, Zehnder J, Kohler S. T-Cell clonality analysis in biopsy specimens from two different skin sites shows high specificity in the diagnosis of patients with suggested mycosis fungoides. *J Am Acad Dermatol* (2007) 57(5):782–90. doi: 10.1016/j.jaad.2007.06.004

30. Sufficool KE, Lockwood CM, Abel HJ, Hagemann IS, Schumacher JA, Kelley TW, et al. T-Cell clonality assessment by next-generation sequencing improves detection sensitivity in mycosis fungoides. *J Am Acad Dermatol* (2015) 73(2):228–36.e2. doi: 10.1016/j.jaad.2015.04.030

31. de Masson A, O'Malley JT, Elco CP, Garcia SS, Divito SJ, Lowry EL, et al. High-throughput sequencing of the T cell receptor β gene identifies aggressive early-stage mycosis fungoides. *Sci Transl Med* (2018) 10(440):eaar5894. doi: 10.1126/scitranslmed.aar5894

32. Attygalle AD, Cabeçadas J, Gaulard P, Jaffe ES, de Jong D, Ko YH, et al. Peripheral T-cell and NK-cell lymphomas and their mimics; taking a step forward - report on the lymphoma workshop of the XVIth meeting of the European association for haematopathology and the society for hematopathology. *Histopathology* (2014) 64(2):171–99. doi: 10.1111/his.12251

33. Vose J, Armitage J, Weisenburger D. International T-cell lymphoma project. international peripheral T-cell and natural killer/T-cell lymphoma study: Pathology findings and clinical outcomes. *J Clin Oncol* (2008) 26(25):4124–30. doi: 10.1200/JCO.2008.16.4558

34. Xie Y, Jaffe ES. How I diagnose angioimmunoblastic T-cell lymphoma. *Am J Clin Pathol* (2021) 156(1):1–14. doi: 10.1093/ajcp/aqab090

35. Tan BT, Warnke RA, Arber DA. The frequency of b- and T-cell gene rearrangements and epstein-barr virus in T-cell lymphomas: A comparison between angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified with and without associated b-cell proliferations. *J Mol Diagn* (2006) 8(4):466–75. doi: 10.2353/jmoldx.2006.060016

36. Smith JL, Hodges E, Quin CT, McCarthy KP, Wright DH. Frequent T and b cell oligoclones in histologically and immunophenotypically characterized angioimmunoblastic lymphadenopathy. *Am J Pathol* (2000) 156(2):661–9. doi: 10.1016/S0002-9440(10)64770-0

37. Yao WQ, Wu F, Zhang W, Chuang SS, Thompson JS, Chen Z, et al. Angioimmunoblastic T-cell lymphoma contains multiple clonal T-cell populations

derived from a common TET2 mutant progenitor cell. *J Pathol* (2020) 250(3):346–57. doi: 10.1002/path.5376

38. Hemminki K, Huang W, Sundquist J, Sundquist K, Ji J. Autoimmune diseases and hematological malignancies: Exploring the underlying mechanisms from epidemiological evidence. *Semin Cancer Biol* (2020) 64:114–21. doi: 10.1016/j.semcancer.2019.06.005

39. Carbone A, Gloghini A, Ferlito A. Pathological features of lymphoid proliferations of the salivary glands: Lymphoepithelial sialadenitis versus low-grade b-cell lymphoma of the malt type. *Ann Otol Rhinol Laryngol* (2000) 109(12 Pt 1):1170–5. doi: 10.1177/000348940010901217

40. Pontarini E, Coleby R, Bombardieri M. Cellular and molecular diversity in sjogren's syndrome salivary glands: Towards a better definition of disease subsets. *Semin Immunol* (2021) 58, 101547. doi: 10.1016/j.smim.2021.101547

41. Johnsen SJ, Berget E, Jonsson MV, Helgeland L, Omdal R, Jonsson R, et al. Evaluation of germinal center-like structures and b cell clonality in patients with primary sjögren syndrome with and without lymphoma. *J Rheumatol* (2014) 41(11):2214–22. doi: 10.3899/jrheum.131527

42. Dong L, Masaki Y, Takegami T, Jin ZX, Huang CR, Fukushima T, et al. Umehara h. clonality analysis of lymphoproliferative disorders in patients with sjögren's syndrome. *Clin Exp Immunol* (2007) 150(2):279–84. doi: 10.1111/j.1365-2249.2007.03486.x

43. Bende RJ, Slot LM, Hoogeboom R, Wormhoudt TA, Adeoye AO, Guikema JE, et al. Stereotypic rheumatoid factors that are frequently expressed in mucosa-associated lymphoid tissue-type lymphomas are rare in the labial salivary glands of patients with sjögren's syndrome. *Arthritis Rheumatol* (2015) 67(4):1074–83. doi: 10.1002/art.39002

44. Broeren MGA, Wang JJ, Balzaretti G, Groenen PJTA, van Schaik BDC, Chataway T, et al. Proteogenomic analysis of the autoreactive b cell repertoire in blood and tissues of patients with sjögren's syndrome. *Ann Rheum Dis* (2022) 81(5):644–52. doi: 10.1136/annrheumdis-2021-221604

45. Bagg A, Dunphy CH. Immunosuppressive and immunomodulatory therapy-associated lymphoproliferative disorders. *Semin Diagn Pathol* (2013) 30(2):102–12. doi: 10.1053/j.semdp.2012.08.005

46. Elshiekh M, Naresh KN. Lymphoproliferative disorders and lymphoreticular malignancies in the setting of immunodeficiency. *Diagn Histopathol* (2018) 24(7):246–56. doi: 10.1016/j.mpdhp.2018.05.008

47. van der Velden WJ, Nissen L, van Rijn M, Rijntjes J, de Haan A, Venkatraman L, et al. Identification of IG-clonality status as a pre-treatment predictor for mortality in patients with immunodeficiency-associated Epstein-Barr virus-related lymphoproliferative disorders. *Haematologica* (2015) 100(4):e152–154. doi: 10.3324/haematol.2014.116780

48. Major A, Kamdar M. Management of non-diffuse Large b cell lymphoma post-transplant lymphoproliferative disorder. *Curr Treat Options Oncol* (2018) 19:33. doi: 10.1007/s11864-018-0549-6

49. Natkunam Y, Goodlad JR, Chadburn A, de Jong D, Gratzinger D, Chan JK, et al. EBV-positive b-cell proliferations of varied malignant potential: 2015 SH/EAHP workshop report-part 1. *Am J Clin Pathol* (2017) 47(2):129–52. doi: 10.1093/ajcp/aqw214

50. Hwang YY, Au-Yeung R, Leung RY, Tse E, Kwong YL. Clonal heterogeneity of polymorphic b-cell lymphoproliferative disease, EBV-positive, iatrogenic/immune senescence: Implications on pathogenesis and treatment. *Hematology* (2022) 27(1):684–90. doi: 10.1080/16078454.2022.2081299

51. Aung PP, Climent F, Muzzafar T, Curry JL, Patel KP, Servitje O, et al. Immunophenotypic shift of CD4 and CD8 antigen expression in primary cutaneous T-cell lymphomas: A clinicopathologic study of three cases. *J Cutan Pathol* (2014) 41(1):51–7. doi: 10.1111/cup.12252

52. Liu H, Shen Q, Chang CC, Hu S. Case report: Phenotypic switch in high-grade b-cell lymphoma with *MYC* and *BCL6* rearrangements: A potential mechanism of therapeutic resistance in lymphoma? *Front Oncol* (2021) 11:795330. doi: 10.3389/fonc.2021.795330

53. Obermann EC, Mueller N, Rufle A, Menter T, Mueller-Garamvoelgyi E, Cathomas G, et al. Clonal relationship of classical hodgkin lymphoma and its recurrences. *Clin Cancer Res* (2011) 17(16):5268–74. doi: 10.1158/1078-0432.CCR-10-1271

54. Nishiuchi R, Yoshino T, Teramoto N, Sakuma I, Hayashi K, Nakamura S, et al. Clonal analysis by polymerase chain reaction of b-cell lymphoma with late relapse: A report of five cases. *Cancer* (1996) 77(4):757–62. doi: 10.1002/(SICI)1097-0142(19960215)77:4<757::AID-CNCR23>3.0.CO;2-Z

55. Lee SE, Kang SY, Yoo HY, Kim SJ, Kim WS, Ko YH. Clonal relationships in recurrent b-cell lymphomas. *Oncotarget* (2016) 7(11):12359–71. doi: 10.18632/oncotarget.7132

56. de Jong D, Glas AM, Boerrigter L, Hermus MC, Dalesio O, Willemse E, et al. Very late relapse in diffuse large b-cell lymphoma represents clonally related disease and is marked by germinal center cell features. *Blood* (2003) 102(1):324–7. doi: 10.1182/blood-2002-09-2822

57. Geurts-Giele WR, Wolvers-Tettero IL, Dinjens WN, Lam KH, Langerak AW. Successive b-cell lymphomas mostly reflect recurrences rather than unrelated primary lymphomas. *Am J Clin Pathol* (2013) 140(1):114–26. doi: 10.1309/AJCPI14GXNWASVUZ

58. Juskevicius D, Lorber T, Gsponer J, Perrina V, Ruiz C, Stenner-Liewen F. Distinct genetic evolution patterns of relapsing diffuse large b-cell lymphoma revealed by genome-wide copy number aberration and targeted sequencing analysis. *Leukemia* (2016) 30(12):2385–95. doi: 10.1038/leu.2016.135

59. van Bladel DAG, Stevens WBC, van den Brand M, Kroeze LI, Groenen PJTA, van Krieken JHJM, et al. Novel approaches in molecular characterization of classical Hodgkin lymphoma. *Cancers* (2022) 14(13):3222. doi: 10.3390/cancers14133222

60. Mao Z, Quintanilla-Martinez L, Raffeld M, Richter M, Krugmann J, Burek C. IgVH mutational status and clonality analysis of richter's transformation: Diffuse large b-cell lymphoma and Hodgkin lymphoma in association with b-cell chronic lymphocytic leukemia (B-CLL) represent 2 different pathways of disease evolution. *Am J Surg Pathol* (2007) 31(10):1605–14. doi: 10.1097/PAS.0b013e31804bdaf8

61. Xiao W, Chen WW, Sorbara L, Davies-Hill T, Pittaluga S, Raffeld M, et al. Hodgkin Lymphoma variant of Richter transformation: morphology, Epstein-Barr virus status, clonality, and survival analysis-with comparison to Hodgkin-like lesion. *Hum Pathol* (2016) 55:108–16. doi: 10.1016/j.humpath.2016.04.019

62. Rossi D, Spina V, Deambrogi C, Rasi S, Laurenti L, Stamatopoulos K, et al. The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood* (2011) 117(12):3391–401. doi: 10.1182/blood-2010-09-302174

63. Maeshima AM, Taniguchi H, Ida H, Hosoba R, Fujino T, Saito Y, et al. Non-diffuse large b-cell lymphoma transformation from follicular lymphoma: A single-institution study of 19 cases. *Hum Pathol* (2020) 102:33–43. doi: 10.1016/j.humpath.2020.06.001

64. Trecourt A, Mauduit C, Szablewski V, Fontaine J, Balme B, Donzel M, et al. Plasticity of mature b cells between follicular and classic Hodgkin lymphomas: A series of 22 cases expanding the spectrum of transdifferentiation. *Am J Surg Pathol* (2022) 46(1):58–70. doi: 10.1097/PAS.0000000000001780

65. Zelenetz AD, Chen TT, Levy R. Histologic transformation of follicular lymphoma to diffuse lymphoma represents tumor progression by a single malignant b cell. *J Exp Med* (1991) 173(1):197–207. doi: 10.1084/jem.173.1.197

66. Carlotti E, Wrench D, Matthews J, Iqbal S, Davies A, Norton A, et al. Transformation of follicular lymphoma to diffuse large b-cell lymphoma may occur by divergent evolution from a common progenitor cell or by direct evolution from the follicular lymphoma clone. *Blood* (2009) 113(15):3553–7. doi: 10.1182/blood-2008-08-174839

67. Eide MB, Liestøl K, Lingjaerde OC, Hystad ME, Kresse SH, Meza-Zepeda L, et al. Genomic alterations reveal potential for higher grade transformation in follicular lymphoma and confirm parallel evolution of tumor cell clones. *Blood* (2010) 116(9):1489–97. doi: 10.1182/blood-2010-03-272278

68. Kridel R, Chan FC, Mottok A, Boyle M, Farinha P, Tan K, et al. Histological transformation and progression in follicular lymphoma: A clonal evolution study. *PloS Med* (2016) 13(12):e1002197. doi: 10.1371/journal.pmed.1002197

69. Egan C, Lack J, Skarshaug S, Pham TA, Abdullaev Z, Xi L, et al. The mutational landscape of histiocytic sarcoma associated with lymphoid malignancy. *Mod Pathol* (2021) 34(2):336–47. doi: 10.1038/s41379-020-00673-x

70. Eggermont AMM, Blank CU, Mandala M, Long GV, Atkinson V, Dalle S, et al. Adjuvant pembrolizumab versus placebo in resected stage III melanoma. *N Engl J ed* (2018) 2018;378:1789–801. doi: 10.1056/NEJMoa1802357

71. Carlino MS, Larkin J, Long GV. Immune checkpoint inhibitors in melanoma. *Lancet* (2021) 398(10304):1002–14. doi: 10.1016/S0140-6736(21)01206-X

72. Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med* (2018) 378:2078–92. doi: 10.1056/NEJMoa1801005

73. Forde PM, Spicer J, Lu S, Provencio M, Mitsudomi T, Awad MM, et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N Engl J Med* (2022) 386(21):1973–85. doi: 10.1056/NEJMoa2202170

74. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz HJ, Morse MA, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): An open-label, multicentre, phase 2 study. *Lancet Oncol* (2017) 18(9):1182–91. doi: 10.1016/S1470-2045(17)30422-9

75. Overman MJ, Lonardi S, Wong KYM, Lenz HJ, Gelsomino F, Aglietta M, et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-Deficient/Microsatellite instability-high metastatic colorectal cancer. *J Clin Oncol* (2018) 36(8):773–9. doi: 10.1200/JCO.2017.76.9901

76. Dash P, Fiore-Gartland A, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* (2017) 547:89–93. doi: 10.1038/nature22383

77. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* (2017) 171(4):934–949.e16. doi: 10.1016/j.cell.2017.09.028

78. Chun B, Pucilowska J, Chang S, Kim I, Nikitin B, Koguchi Y. Changes in T-cell subsets and clonal repertoire during chemoimmunotherapy with pembrolizumab and paclitaxel or capecitabine for metastatic triple-negative breast cancer. *J Immunother Cancer* (2022) 10(1):e004033. doi: 10.1136/jitc-2021-004033

79. Chalabi M, Fanchi LF, Dijkstra KK, van den Berg JG, Aalbers AG, Sikorska K, et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat Med* (2020) 26:566–76. doi: 10.1038/s41591-020-0805-8

80. Keane C, Gould C, Jones K, Hamm D, Talaulikar D, Ellis J. The T-cell receptor repertoire influences the tumor microenvironment and is associated with survival in aggressive b-cell lymphoma. *Clin Cancer Res* (2017) 23(7):1820–8. doi: 10.1158/1078-0432.CCR-16-1576

81. Olschewski V, Witte HM, Bernard V, Steinestel K, Peter W, Merz H, et al. Systemic inflammation and tumour-infiltrating T-cell receptor repertoire diversity are predictive of clinical outcome in high-grade b-cell lymphoma with *MYC* and *BCL2* and/or *BCL6* rearrangements. *Cancers (Basel)* (2021) 13(4):887. doi: 10.3390/cancers130408877

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 16 March 2023
DOI 10.3389/fonc.2023.1115361

OPEN ACCESS

EDITED BY
Richard Rosenquist,
Karolinska Institutet (KI), Sweden

REVIEWED BY
Jonathan C. Strefford,
University of Southampton,
United Kingdom
Jeroen EJ Guikema,
Academic Medical Center, Netherlands

*CORRESPONDENCE
Ramit Mehr
ramit.mehr@biu.ac.il

†These authors have contributed equally to this work and share senior authorship

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

RECEIVED 03 December 2022
ACCEPTED 03 March 2023
PUBLISHED 16 March 2023

CITATION
Neuman H, Arrouasse J, Benjamini O, Mehr R and Kedmi M (2023) B cell M-CLL clones retain selection against replacement mutations in their immunoglobulin gene framework regions.
Front. Oncol. 13:1115361.
doi: 10.3389/fonc.2023.1115361

B cell M-CLL clones retain selection against replacement mutations in their immunoglobulin gene framework regions

Hadas Neuman[1], Jessica Arrouasse[1], Ohad Benjamini[2,3], Ramit Mehr[1*†] and Meirav Kedmi[1,2,3†]

[1]The Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, Ramat Gan, Israel, [2]Division of Hematology and Bone Marrow Transplantation, Chaim Sheba Medical Center, Ramat-Gan, Israel, [3]Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

Introduction: Chronic lymphocytic leukemia (CLL) is the most common adult leukemia, accounting for 30–40% of all adult leukemias. The dynamics of B-lymphocyte CLL clones with mutated immunoglobulin heavy chain variable region (IgHV) genes in their tumor (M-CLL) can be studied using mutational lineage trees.

Methods: Here, we used lineage tree-based analyses of somatic hypermutation (SHM) and selection in M-CLL clones, comparing the dominant (presumably malignant) clones of 15 CLL patients to their non-dominant (presumably normal) B cell clones, and to those of healthy control repertoires. This type of analysis, which was never previously published in CLL, yielded the following novel insights.

Results: CLL dominant clones undergo – or retain – more replacement mutations that alter amino acid properties such as charge or hydropathy. Although, as expected, CLL dominant clones undergo weaker selection for replacement mutations in the complementarity determining regions (CDRs) and against replacement mutations in the framework regions (FWRs) than non-dominant clones in the same patients or normal B cell clones in healthy controls, they surprisingly retain some of the latter selection in their FWRs. Finally, using machine learning, we show that even the non-dominant clones in CLL patients differ from healthy control clones in various features, most notably their expression of higher fractions of transition mutations.

Discussion: Overall, CLL seems to be characterized by significant loosening – but not a complete loss – of the selection forces operating on B cell clones, and possibly also by changes in SHM mechanisms.

KEYWORDS

antibody, B lymphocytes, chronic lymphocytic leukemia (CLL), high-throughput sequencing (HTS), immunoglobulin (Ig), lineage trees, somatic hypermutation (SHM), machine learning (ML)

1 Introduction

Chronic lymphocytic leukemia (CLL) is the most common adult leukemia and stands for 30–40% of all adult leukemia cases (1, 2), and 7% of newly diagnosed cases of non-Hodgkin's lymphoma (3). B-CLL (henceforth referred to as simply CLL) is a chronic B-cell malignancy, which typically affects elderly people, progresses gradually over many years, and involves substantial innate and adaptive immune system perturbations. Adaptive response impairments include down-regulation of T-cell function and defects in antibody-dependent cellular cytotoxicity, and in B cells – hypogammaglobulinemia and alterations in cell-cell contact and cytokine release, all of which may contribute to the overall immune suppression observed in patients (4). Indeed, during the COVID-19 pandemic, fatality rates for CLL patients were 16.5-fold more than the median population fatality rates reported worldwide, and even higher in older patients (5).

It has long been known that CLL genomes show heterogeneity between patients (6, 7), and that CLL clinical manifestations range from very indolent to aggressive disease (1). One partitioning of CLL is based on "stereotypic BCRs", identified by the IgHV gene CDR3 region amino acid sequence; stereotypic BCRs can be assigned to 30% of CLL cases, and were associated with prognosis (8, 9). More importantly, CLL tumors are classified into two subgroups based on the presence of somatic hypermutations in their IgHV, where CLL patients with little to no SHM (98% IgHV sequence homology to germline) are defined as unmutated CLL (U-CLL), and CLL with SHM (less than 98% IgHV sequence homology) are defined as mutated CLL (M-CLL) (10). M-CLL patients have a better prognosis than those with U-CLL, as U-CLL is considerably more aggressive and less susceptible to chemo-immunotherapy (2, 8). This manuscript focuses solely on M-CLL (henceforth referred to simply as CLL). Although the mutational imprint on CLL cell IgHV genes has first been considered static, there is now clear evidence that, in a subgroup of cases, rearranged Ig genes are subject to ongoing mutational pressure (8). In such cases, the study of CLL clonal dynamics using Ig gene high-throughput sequencing (HTS) can yield important insights.

Since 2008, Adaptive Immune Receptor Repertoire HTS (AIRR-seq) has generated data sets of up to billions of reads (11, 12), and has, indeed, led to new insights into affinity maturation. BCR-seq has many applications (13), including broadly neutralizing antibody identification (14), vaccine response studies (15), B-cell migration and development tracking within the body (16) and disease diagnosis (17). In particular, Stamatopoulos and colleagues used HTS to sequence more than 200 CLL patient repertoires and demonstrated that one quarter of the CLL patients include multiple clones with unrelated, productively rearranged IgHV genes (18). The extensive amount of data that arise from AIRR-seq can also be analyzed using machine learning (ML) methods, e.g. for classification of B cell subpopulations, "public" vs. "private" clones, and more (19–21).

A B cell clone is a cell lineage that includes all the descendants of a founder B cell, all of which share a unique IgHV rearrangement; clonal diversification is best modeled by lineage trees. IgHV gene SHM and selection – including those in malignant clones, if any – are more precisely analyzed on IgHV gene lineage trees, because mutations are more correctly defined relative to the closest known ancestors, and thus mutation counts – and all the analyses relying on them, including selection analysis, in which CDR3s can only be included if using lineage trees – and lineage tree topologies are more correct on lineage trees (22). Using lineage trees, Abraham and colleagues found evidence of intraclonal diversification of characteristic clones in light chain amyloidosis patients, concluding the pathogenic plasma cells are probably derived from a precursor population in which SHM is ongoing (23). Zuckerman et al. used lineage tree-based mutation analysis to find that follicular lymphoma (FL), diffuse large B cell lymphoma (DLBCL), and primary central nervous system lymphoma repertoires have similar mutation frequencies and do not undergo positive selection *for* replacement mutations in their CDRs (24), using the focused binomial test (25) rather than relying on previously published tests for selection, which have all been shown to generate false positives (26, 27). The transformation of FL into DLBCL has been followed using clonal lineage trees to show that, in some cases, therapy eradicates a DLBCL clone but a new one develops from remnants of the original FL clone (28, 29). Lineage tree analysis of dominant clones from mucosa-associated lymphoid tissue lymphoma showed higher diversification and longer mutational histories compared with chronic gastritis or with gastric DLBCL (30); gastric DLBCL may originate from gastritis, mucosa-associated lymphoid tissue lymphoma or *de novo*, and, like CLL, may sometimes contain more than one dominant clone (31). Green et al. used lineage trees to distinguish early versus late genetic events in follicula lymphoma (32). Béguelin and colleagues used lineage tree analysis to show evidence of reduced efficacy of affinity maturation in mice with EZH2 mutations, which initiate lymphomagenesis (33). Finally, Kedmi et al. showed that the use of lineage trees is necessary for detection of minimal residual disease in a DLBCL patient, prior to its detection by PET-CT (34). In this work, we aimed to study the SHM and selection (if any) mechanisms that operate on CLL clones using IgHV gene lineage tree-based analyses and machine learning methods, which to the best of our knowledge have never been previously applied to CLL. Such analysis can yield novel insights, as demonstrated by our most important finding, i.e. that while CLL dominant clones undergo weaker selection *for* replacement mutations in their CDRs, they *retain* some selection (albeit weaker than that in healthy controls and non-dominant clones) *against* replacement mutations in their FWRs.

2 Methods

2.1 Datasets

IgHV gene sequences from peripheral blood samples of 16 M-CLL patients were obtained for routine diagnosis of mutated vs. unmutated CLL cases. Only M-CLL samples were chosen for this study. Sample data are summarized in Table 1. Buffy coats were

10.3389/fonc.2023.1115361

TABLE 1 CLL patient dataset.

Sample name	# Sequences in raw data	# Unique sequences after processing	# Clones	# Clones with 2 or more sequences
INDEX2_S1	1,040,682	9,373	182	100
INDEX3_S2	489,773	8,456	365	343
INDEX4_S3	763,818	4,014	53	17
INDEX5_S4	546,554	4,654	40	24
INDEX6_S5	1,147,666	15,969	684	652
INDEX7_S6*	657,167	25,750	2,473	1,840
INDEX8_S7	1,421,821	11,620	52	38
INDEX9_S8	382,498	2,330	21	12
INDEX10_S9	1,167,134	56,998	6,797	6,390
INDEX12_S10	736,105	9,554	96	87
INDEX13_S11	967,315	6,056	86	70
INDEX14_S12	686,510	8,563	26	16
INDEX15_S13	726,865	6,306	70	58
INDEX16_S14	547,240	2,688	243	129
INDEX18_S15	749,263	3,817	20	15
INDEX19_S16	945,151	4,896	70	55
Overall	12,975,562	181,044	11,278	9,846

* This sample contained several large clones, so that the dominant clone could not be identified with certainty; hence this sample was omitted from the study.

taken, and DNA was extracted directly from the buffy coat of each sample. IgHV gene libraries were produced using the LymphoTrack® kit (Dx IGHFR1 Assay Panel for MiSeq, Catalog #91210039, *In vivo*scribe, San Diego, CA, USA). Sequencing was performed using the MiSeq V300 kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. The use of the resulting IgHV sequence data (without any clinical or other identifying data) was approved by the Sheba Medical Center and Israeli Ministry of Health review boards.

For comparison with CLL patient repertoires, we used IgHV sequences from three blood samples of healthy individuals (35), which are publicly available, and were downloaded by us as part of a different study. Since CLL is most common in elderly patients, we chose the samples of the three eldest healthy individuals for this comparison; healthy control (HC) sample data are summarized in Table 2. For negative controls in the selection analysis (see below), we used lineage trees composed only of sequences containing a frame shift, taken from the same CLL patients, as these sequences most likely represent non-productively rearranged, non-expressed alleles.

2.2 Data processing steps

We preprocessed the sequences using pRESTO version 0.5.13 (36). The preprocessing included assembly of paired ends and quality filtering by (i) trimming low quality edges, (ii) filtering out reads with an average Phred score lower than 25, and (iii) masking bases with Phred scores lower than 20. Sequences with more than 10 masked or missing bases were removed. Since the sequencing kit manufacturer does not consent to reveal the primer sequences, we removed 30 nucleotides from both ends of each sequence. Next, identical sequences were collapsed, and only sequences with two copies or more were selected for analysis; this is standard practice meant to reduce the chance of including PCR and sequencing errors in cases such as this, where unique molecular

TABLE 2 Healthy repertoire samples.

Sample name	Sex	Age	# Unique sequences after processing	# Clones	# Clones with 2 or more sequences
H45_3	F	45	169,243	67,346	24,841
H45_4	F	45	257,571	171,620	35,946
H50_7	F	50	118,472	50,453	18,404
Overall			545,286	289,419	79,191

identifiers UMIs were not used. Further precautions we took to minimize such errors were: (a) Using only one copy of each set of identical sequences in the lineage tree analysis; sequence copy numbers weren't used in any of our analyses. (b) Omitting "clones" that contain only one unique sequence (regardless of its copy number) from the analysis.

We further processed the selected sequences using Change-O version 0.4.6 (37) and in-house custom scripts. The processing included annotation of the sequences with the IMGT/GENE-DB (38) reference germline sequences from July 1, 2021, and removal of sequences annotated as non-functional (those with frame-shifts or stop codons); dynamic clonal assignment according to V and J segment annotation and junction (CDR3) similarity (the numbers of clones in each sample are given in Table 1); and assessments of sampling depth and of the clonal size distributions of each repertoire. Putative germline sequences for each clone were created based on the same IMGT/GENE-DB database and the clonal consensus in junction regions, and clones with more than two unique sequences were sent to IgTree© (39) for lineage tree construction. Sample Lineage trees are shown in Figure 1 and Figure S1. Note that the only times a lineage tree node may represent more than one sequence is when these unique sequences differ by mutation(s) that fall in sequence margins, and these margins were further trimmed by IgTree© because one or more sequences in the clone lacked information on those margins.

To focus on the malignant clones in CLL patient repertoires, we separated the largest (dominant) clone from each repertoire, assuming it is the malignant clone. As internal controls, we used the non-dominant clones from the same patients, under the assumption that these are normal B cell clones (although they may be reactive to the tumor itself). This assumption was based on the knowledge that all B cell populations are composed of clones; even naïve B cells divide a few times before settling into a resting state, and may later perform homeostatic cell divisions (41). One sample included several large clones, and hence was omitted from the study (Table 1), to avoid the possibility of including a second CLL clone in the "non-dominant" control group. The healthy control repertoires served as external controls; for the sake of studying SHM and antigen-driven selection, if any, clones that were reactive at the time of sampling are the most valuable controls.

2.3 Lineage tree-based analyses

2.3.1 Tree-based mutation analyses

Lineage tree-based mutation analyses were performed using our program IgTreeZ (22), based on the linkage of tree nodes to their corresponding sequences. IgTreeZ traverses all tree nodes, counts all the observed mutations, and characterizes each mutation by its sequence location (CDR/FWR, based on IMGT region definitions (42)) and type (source nucleotide, transition/transversion, replacement/silent); if it was a replacement mutation, the program also characterized the pre- and post-mutation amino acids based on IMGT physicochemical amino acid classes (43, 44).

2.3.2 Selection analysis

Selection analysis was done using ShazaM (37, 45), which is based on the focused binomial test (25). The numbers of silent and replacement mutations in the CDRs and FWRs for all sequences in each tree received from IgTreeZ were sent to ShazaM, together with the corresponding clonal germline sequence and the CDR3 length of each tree. Using ShazaM, we calculated the expected mutation frequency in each region of each sequence, estimated the selection strength for each tree, and compared the selection scores of the different lineage tree repertoires. CDR3s were included in the analysis by modifying ShazaM's region definition parameter according to each tree's CDR3 length and calculating the

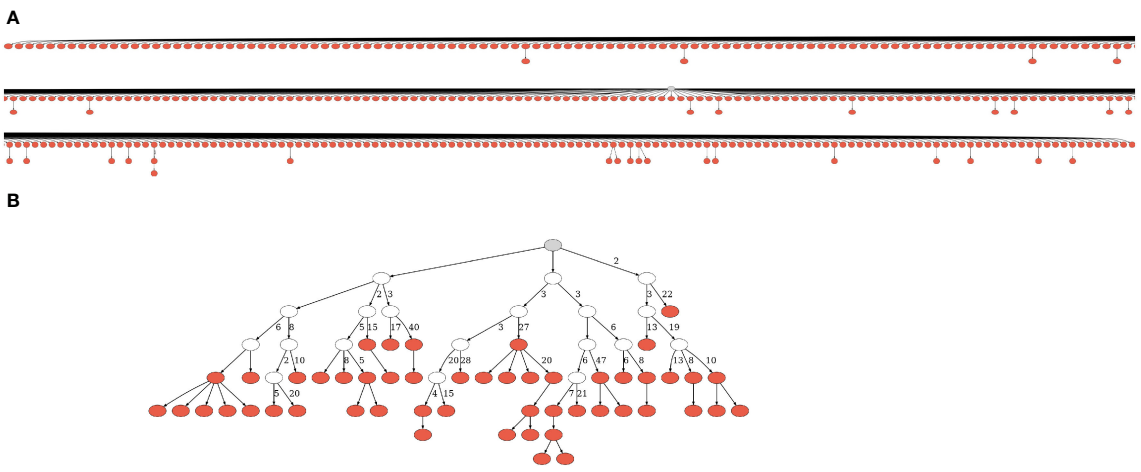


FIGURE 1
Lineage trees from CLL patients. **(A)** One of the smallest trees from expanded, dominant clones. Due to its size, we had to split the figure into partly overlapping segments. **(B)** One of the largest trees from presumably normal, non-dominant clones. A gray node represents the root, and a white node – a hypothetical split node. Numbers next to edges denote numbers of mutations; edges with no adjacent numbers represent one mutation. The trees were drawn using IgTreeZ (22) and Graphviz (40). More representative trees of all sizes are given in Figure S1.

expected mutation frequency for each clonal germline sequence separately.

2.3.3 Tree topology analysis

Seven graphical shape properties of IgV gene lineage trees were found to be most strongly influenced with B cell response parameters, such as activation, division, mutation and death rates and selection thresholds (46). The seven tree shape properties are: (i) trunk length (the number of mutations from the root node, which represents the pre-mutation sequence, to the first split node), (ii) the minimum root to leaf path (i.e. the minimum number of mutations per leaf), (iii) the minimum root to split node path (which equals the trunk when there is one), (iv) the number of children emerging from the root (a node's "children" are defined here as those representing sequences that differ from the parent node by a single mutation), (v) the average number of children per node, (vi) the average distance from the first split node to any leaf, and (vii) the minimum fork to fork distance (that is, the distance between two consecutive splits on the same path). IgTreeZ (22) calculates these variables for each tree and enables us to and compared the results between groups.

2.3.4 Tree drawing

To visually illustrate lineage tree shapes (Figure 1 and Figure S1), we created drawings using the graph description language DOT, as implemented in Graphviz (40). Node (sequence) names were omitted for better tree visualization.

2.3.5 Tree trunk removal

To exclude as much as possible of the pre-transformation mutation and selection history of each lymphoma clone from some of the analyses, we removed the trunks from the trees in all groups, and assigned the first split node of each "trunkless" tree to be the new root node. Trees that originally had no trunks were removed from the trunkless analyses, so the data are not biased, as such trees did not contain enough information regarding their diversification history. However, since the latter step left only three trees for analysis, we performed most analyses both with and without tree trunks and compared the results.

2.4 Statistical analyses

Comparisons between lymphoma lineage tree characteristics against those of healthy repertoires, which included more than 50,000 trees, were done based on the average measurements per patient/subject, to overcome the bias of the healthy control dataset being so much larger (in terms of numbers of trees) than the other datasets. For each comparison, the assumptions of normal data distribution and variance homogeneity were tested using the Shapiro test and the Levene test, correspondingly. If the data were normally distributed and had homogenous variances, Student's t-test or its paired version were used. Otherwise, the non-parametric Mann–Whitney U-test, or the Wilcoxon test for paired comparisons, were used. To correct for multiple comparisons, we used Benjamini and Hochberg's False detection rate (FDR) method (47). Only differences with p-values lower than the FDR-corrected α were considered to be significant.

2.5 Machine learning classification models

We used all the results of lineage tree-based mutation analyses of the CLL non-dominant and healthy control clones as input for our ML models. Data in all columns which included simple mutation counts were normalized by dividing them by the total number of mutations in each tree, to receive the *frequency* of each mutation type. Columns listing median and average replacement distances and CDR3 lengths were not normalized. We also excluded FWR1 mutation counts from the analysis, as it may be influenced by the sequencing. Since we had almost tenfold more healthy control clones than non-dominant clones in CLL patient samples (Tables 1, 2), the dataset was balanced using the SMOTetomek algorithm (48) – a combination of oversampling the CLL data by synthesizing new examples based on the structure and composition of the real non-dominant clones using SMOTE (49), and under-sampling of the healthy control data using the TOMEK algorithm (50). Three ML models were built using Python's Scikit-learn package (51) – a Support Vector Machine (SVM), a Random Forest and an XGBoost model. The F1-score, which is the harmonic mean of model precision and recall, was used as a model performance metric, in order to account for both measures.

3 Results

3.1 Dominant CLL clones undergo, or retain, more replacement mutations that alter amino acid physical properties

To examine CLL clone diversification, we first compared trees of dominant and non-dominant clones in CLL samples (each group separately) with trees of healthy controls, and found that dominant CLL clones include significantly more mutations per clone than non-dominant clones in the same patients (Figure 2A, $p < 0.01$, Wilcoxon paired test and FDR correction), or than clones from healthy control repertoires ($p < 0.01$, Mann Whitney test and FDR correction). Since tumor clone lineage tree 'trunks' may contain mutations that had occurred prior to malignant transformation, we also performed all analyses on the trees after trunk removal, as described in the Methods section. The above-described differences were also found in the trunkless tree analysis (Figure 2B, $p < 0.05$ for both comparisons, Wilcoxon paired test and FDR correction); the higher p-values in trunkless analysis vs. analysis with trunks may result from the decreased numbers of data points due to the exclusion of original trunkless trees. In contrast, when we compared the numbers of mutations *per sequence*, we found that dominant clones have fewer mutations per sequence than non-dominant ($p < 0.01$, Wilcoxon paired test and FDR correction) and healthy repertoire clones ($p < 0.01$, Mann Whitney test and FDR correction). These differences were also found in the trunkless trees,

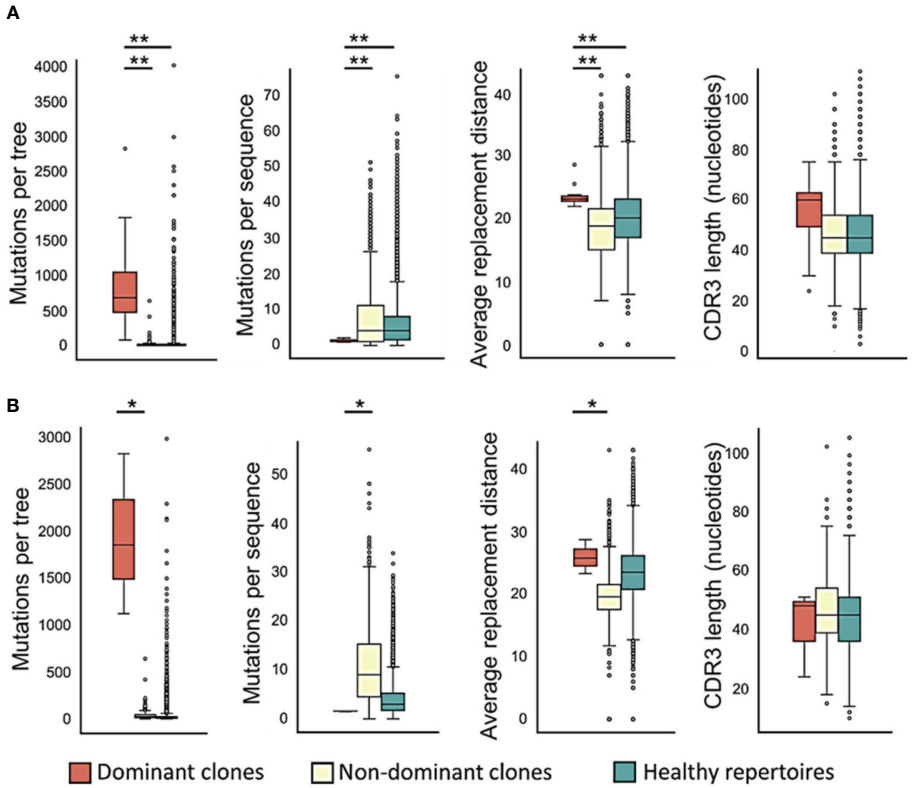


FIGURE 2
Dominant CLL clones undergo – or retain – more mutations, in particular replacement mutations, than non-dominant or healthy control clones. **(A)** Trees with trunks; **(B)** trunkless trees. The average physico-chemical distance was calculated between pre- and post- replacement mutation amino acids based on Sneath's index (52). The paired T-test or the Wilcoxon paired test were used when comparing between dominant and non-dominant clones in the same patients, and Student's T-test or Mann-Whitney test – between patient and healthy control clones, depending on whether the data were distributed normally or not. $*p < 0.05$, $**p < 0.01$.

with larger p-values. This results in the highly branched rather than "long" shape of the CLL trees (Figure 1 and Figure S1), which, in our experience, is typical not only in CLL but also in other B cell GC-derived lymphomas. We hypothesize that the combination of high numbers of mutations *per tree* with low numbers of mutations *per sequence* result from having a population of malignant cells constantly dividing and generating new mutants, which do not get to mutate further because the cells still retain *some* selection against deleterious mutations, as further investigated and discussed below.

The average physico-chemical distance between pre- and post-replacement mutation amino acid, measured by Sneath's index (52), was larger in dominant clones than in non-dominant or healthy repertoire clones (both $p < 0.01$, Mann Whitney test and FDR correction). Indeed, comparisons of several individual components of the Sneath index – that is, the frequencies of changes in several different amino acid properties – revealed that dominant clones tend to undergo, or retain, more replacement mutations that alter the amino acid charge, volume, and/or hydropathy more often than non-dominant clones and healthy repertoires (Figure 3).

The excluded, original trunkless dominant trees tend to have significantly more mutations per tree ($p < 0.01$, paired t test), than the dominant clone trees with trunks (Figure S2). The numbers of mutations per tree were also higher in originally trunkless dominant trees compared to trunk-including trees ($p < 0.01$, Mann Whitney test). Finally, the average physico-chemical distance between pre- and post-mutation amino acids in replacement mutations was higher in originally trunkless dominant trees compared to trunk-including trees ($p < 0.05$, Mann Whitney test). The latter differences may be due to the time it took for each CLL clone to develop until the sample was taken. Since every mutation requires cell replication to be completed, slower-growing clones, whether normal, pre-malignant or tumor clones, will gather fewer mutations. In addition, as long as the cells are sensitive to some level of selection, cells with harmful BCR mutations will eventually die, and thus such cells will produce fewer progeny overall. Slower-growing tumors are also likely to be detected after growing for a longer time, as it would take longer for symptoms to manifest in the patient. As a result of all these considerations, we assume that earlier branches of slow-growing clones have a lower chance of being picked up in the sample, and thus slower-growing clones are more likely to have both longer lineage tree trunks. Overall, the results presented in this section demonstrate that M-CLL tumors have very heterogenous diversification histories, and the presence of trunks in most lineage trees of these clones suggests that they may have been subject to some degree of selection against harmful BCR mutations, not only before the malignant transformation but also following it, even up to the time of sampling.

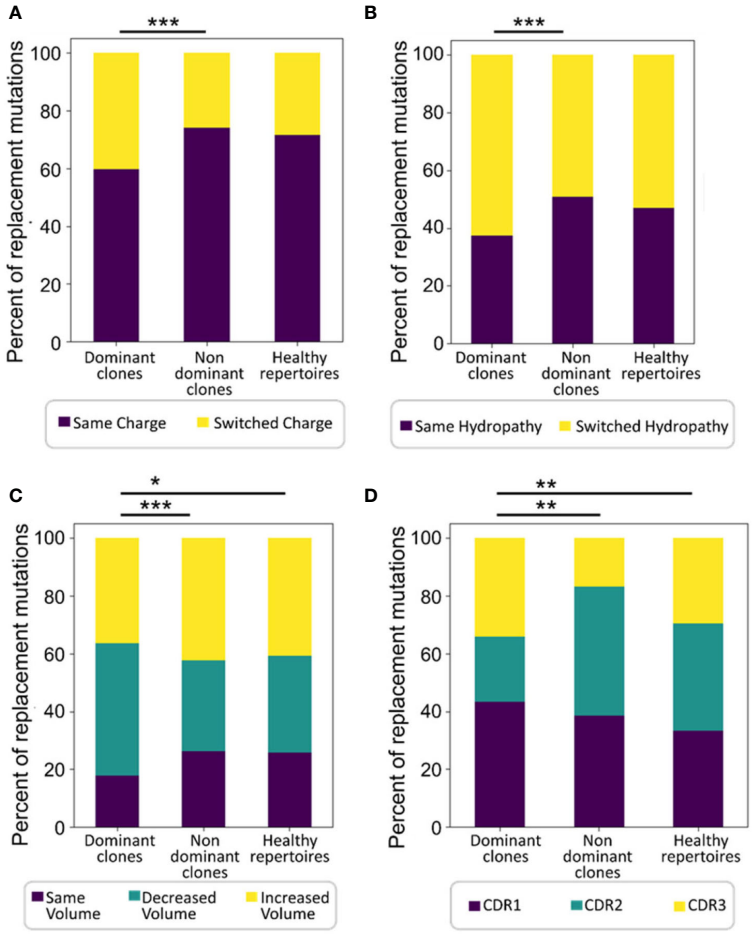


FIGURE 3
Dominant clones undergo or retain more replacement mutations that alter amino acid properties. Shown are percentages of replacement mutations in all trees that change the amino acid **(A)** charge, **(B)** hydropathy, or **(C)** volume, and **(D)** the distributions of mutations among CDRs in all trees. Significant differences were also found in mutations that change amino acid polarity, chemical group, and the tendency to donate and accept hydrogen (not shown). The Wilcoxon paired test was used when comparing between dominant and non-dominant clones in the same patients, and the Mann-Whitney test – between patient and normal healthy controls, as the data were not normally distributed. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

3.2 Lineage tree topologies suggest that CLL dominant clones retain some sensitivity to selection

Next, lineage tree topologies were studied, as another way to examine clonal diversification; here, we only present lineage tree shape properties for which significant differences between groups were detected. Trees from dominant CLL clones were found to have significantly shorter trunks than trees from non-dominant clones in the same patients (Figure S3A; $p < 0.01$, Wilcoxon paired sample test and FDR correction) and from healthy controls clones ($p < 0.01$, Mann-Whitney t test and FDR correction). The minimum root to leaf path (i.e., the minimum number of mutations per leaf) and the minimum root to fork path were significantly shorter in trees from dominant clones than in those from non-dominant ($p < 0.01$ for both, Wilcoxon paired sample test and FDR correction) or healthy control clones ($p < 0.05$ and $p < 0.01$, respectively, Mann-Whitney t test and FDR correction). In the original simulation study described in the methods section on which our interpretations are based (46), these lineage tree "length" measures were inversely influenced by initial clone affinity and selection strength, which makes intuitive sense, because (a) the higher the initial affinity, the fewer mutations are needed (if at all) for the BCR to reach the optimal shape for binding its cognate antigenic epitope (i.e. where any mutation would decrease the affinity, see also (53)); and (b) the more stringent antigen-driven selection is, the fewer mutations will survive. It is harder to interpret the shapes of tumor clones; however, their shorter branches suggest that CLL cells retain *some* sensitivity to selection.

CLL clone lineage trees are not only shorter but also much more branched, as demonstrated by the following findings. The numbers of leaves (branch endpoint nodes) per tree were significantly larger in dominant clones, with a median of 537 leaves per tree, rather than 1 in non-dominant and healthy control clonal trees, as most normal B cell clones are represented in the peripheral blood by one or very few sequences. The numbers of children emerging from the tree root, and the average number of children per node, were significantly larger in trees from dominant clones than in those

from non-dominant (Figure S3B; $p < 0.01$, Wilcoxon paired sample test and FDR correction) or healthy control clones ($p < 0.05$, Mann-Whitney t test and FDR correction). The median number of children emerging from the root was 342 descendant nodes in dominant trees and 1 descendant node in non-dominant and healthy control trees. In the original simulation study (46), these lineage tree "branching" measures were directly influenced by initial clone affinity – the higher the initial affinity, the more success in forming additional branches, as explained above – and the average number of children per node was inversely influenced by selection strength, again because selection would "trim" lower-affinity branches. To interpret the shapes of CLL clones, we should ignore initial (presumably pre-transformation) clonal affinity, and only refer to the highly branched shapes of the observed clonal trees. These shapes suggest that whatever selection acts on the IgHV mutants is weak enough to allow a constantly dividing and mutating tumor cell population to continuously replenish the dominant clone cells in the blood. Finally, the trunkless analysis showed similar trends to those in the trunk-including analysis (Figures S3C, D), though with lower statistical significance due to the smaller group sizes.

3.3 CLL dominant clones undergo weaker selection *for* replacement mutations in the CDRs, but retain selection *against* replacement mutations in the FWRs

To directly test which, if any, type of selection has been acting on the mutated CLL and control clones and to what extent, IgTreeZ mutation counts in the FWR and CDR regions were used as input for the ShazaM R package (37, 45). We also created a cohort of non-selected control clones by assigning all sequences in each repertoire – functional and non-functional – into clones, and constructing lineage trees from the clones that included only out-of-frame IgHV sequences, presumably representing un-productively rearranged, non-expressed IgHV alleles. Selection scores measured on all four clonal repertoires show that dominant clones undergo the weakest selection – or none at all – *for* replacement mutations in the CDRs, similar to the non-productive clones (Figures 4A, B), compared to that in non-dominant or healthy control clones ($p < 0.001$, Student's T-test with FDR correction for multiple comparisons). In contrast, in the FWRs, dominant clones clearly undergo selection *against* replacement mutations (as their selection scores significantly differ from those of the non-selected clones; the latter have scores that do not significantly differ from the case of no selection, depicted by the zero line), although it is weaker than the same selection observed in non-dominant clones and in healthy repertoires ($p < 0.001$, Student's T-test with FDR correction for multiple comparisons). Selection scores in non-dominant clones were similar in both CDRs and FWRs to those in healthy repertoires. Overall, these results suggest that the selection that operates on CLL clones is not completely abolished, but is certainly different from that in normal repertoires. The selection *against* replacement mutations in the FWRs may represent a need for (at least partial) maintenance of the structural integrity of the B cell receptor, as discussed below.

Dominant CLL clone trees with trunks removed still seem to undergo selection *for* replacement mutations in the CDRs, and selection *against* replacement mutations in the FWRs, similarly to trees from healthy repertoires (Figures 4C, D). Dominant trees that originally had no trunks, however, undergo weak – if any – selection in the CDRs, with indistinguishable scores from those measured on the fully non-productive clones. These results further illustrate CLL tumor heterogeneity, and emphasize the need for trunk removal from tumor clone trees, as many replacement mutations in tree trunks must have been selected (for or against), so including the pre-transformation mutation history in lymphoma clone analysis may confound the results.

3.4 Machine learning reveals potential SHM impairments even in *non*-dominant patient clones

In the past, we have shown that IgTreeZ extensive mutation counts can be used as input for ML models, to further elucidate the mutation mechanism in DLBCL clones (22). In the current study, we compared only patient *non*-dominant clones to healthy control clones; malignant clone data were not included in the ML models, as the purpose of the ML models was to identify traces of potential CLL patient-specific (rather than tumor-specific) impairments in SHM or antigen-driven selection, rather than to distinguish between patient and healthy control clones.

To perform the most unbiased analysis we could, we first normalized the mutation counts by dividing the specific mutation counts by the total number of mutations in each tree to receive the *frequency* of each mutation type. Second, since our dataset was extremely imbalanced, with almost tenfold healthy control clones than non-dominant clones in CLL patient samples (Tables 1, 2), we balanced the dataset using the SMOTetomek algorithm (48). Third, we built three different machine learning models – a Support Vector Machine (SVM), a Random Forest and an XGBoost model – to classify the revised datasets.

All three classification models exhibited very high accuracy; Random Forest presented the best performance with F1-scores of 0.962 and 0.961 for HC and non-dominant trees, respectively, and SVM the worst, with F1-scores of 0.831 and 0.836 for HC and non-dominant trees (Figure 5A). XGBoost performed almost as well as Random Forest (Figure 5B). To assess the relevance of specific input parameters to this classification – and thus to learn which features of SHM are specific to non-dominant clones from CLL patients rather than to their tumors – we calculated the feature importance scores of the Random Forest and XGBoost models. The transition mutation frequency was found to be the best predictor, accounting for 0.08 of the separation in Random Forest (Figure 5C) and 0.15 of the separation in XGBoost (Figure 5D). Indeed, transition mutation frequencies in the CLL non-dominant clones tended to be higher than those in healthy controls (Figure 5E). Overall, these results suggest either that the presence of CLL malignant clone(s)

10.3389/fonc.2023.1115361

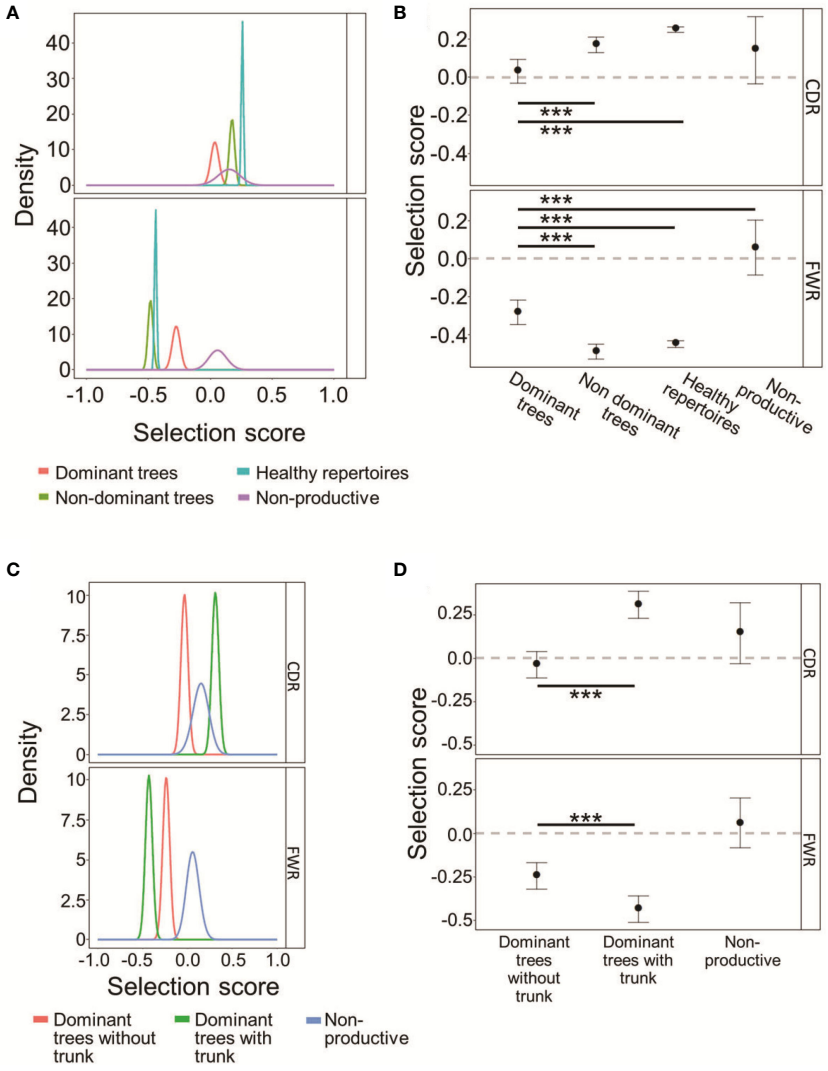


FIGURE 4
CLL dominant clones retain some selection *against* replacement mutations in the FWRs. **(A)** The probability density functions of the selection scores for dominant CLL clones in comparison to non-dominant clones in the same patients, or healthy donor clones, and to fully non-productive clones, calculated on the lineage tree-based mutation counts of the same data. Positive values indicate selection for, and negative values – selection against, replacement mutations. **(B)** Means and 95% confidence intervals of the selection scores plotted in **(A)**. **(C, D)** CLL dominant clones without trunks exhibit weaker selection than dominant clones with trunks, both for replacement mutations in the CDRs and against replacement mutations in the FWRs. **(C)** Same as **(A)** for *trunkless* trees. **(D)**. Means and 95% confidence intervals of selection scores of the selection scores plotted in **(C)**. The line at Selection Score=0 is shown to indicate when the results are indistinguishable from the case of no selection operating on the clones. Both graphs were plotted using ShazaM (37, 45) based on the focused binomial test (25). ***p < 0.001, Student's T-test with FDR correction for multiple comparisons.

influences SHM or selection of non-dominant B cell clones, or that some slight impairments in one of these mechanisms were present prior to malignancy detection, and may have even contributed to malignant transformation.

4 Discussion

CLL is a chronic disease, and M-CLL tumor clones may accumulate mutations in their IgHV genes for many years. For these reasons, we assumed that dominant clones would show different mutation characteristics than healthy control clones. Messmer and colleagues, who performed sequence-based mutational analysis of representative CLL IgHV gene sequences from the dominant clones of 172 CLL patients, found that dominant CLL sequences include more mutations than non-dominant ones (54); Petrova et al. used isotype-resolved BCR sequencing and indicated a distinct evolution of malignant CLL clones relative to clones from healthy volunteers (55). However, neither study characterized these mutations. Using IgHV lineage tree-based analyses, we found that dominant CLL clones undergo – or retain – more IgHV replacement mutations that alter amino acid physico-chemical properties than non-dominant or healthy control clones. Supporting the mutation analysis results, dominant CLL clone lineage trees possess tumor-typical, highly branched topologies, which correlate with weaker – but present – selection.

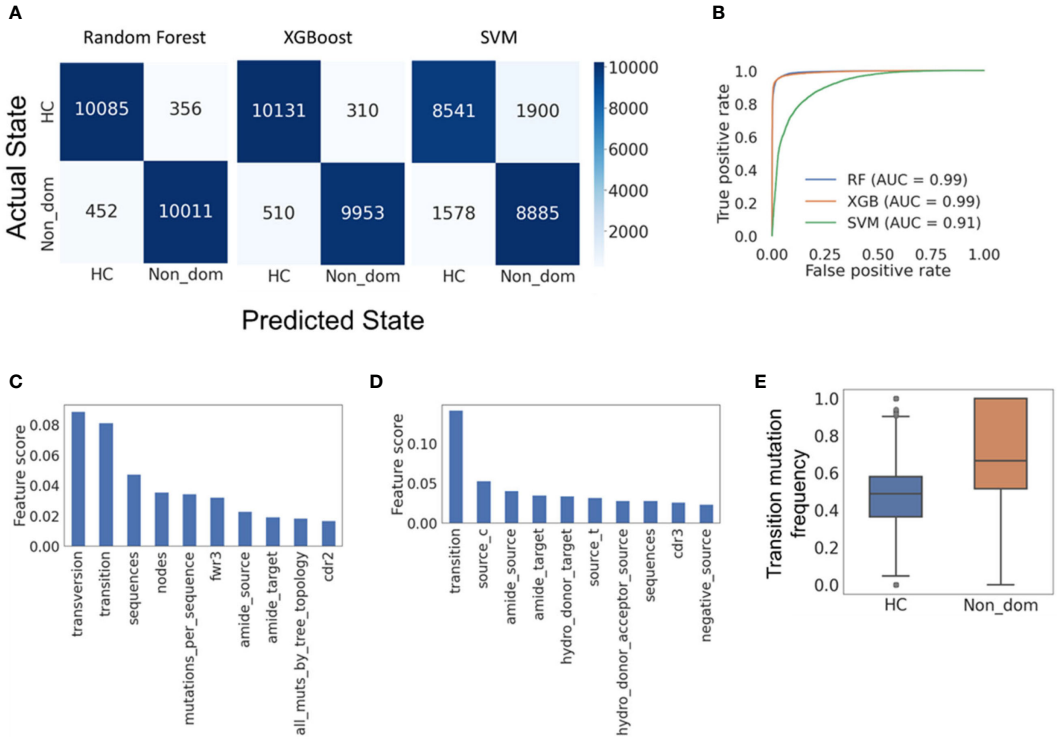


FIGURE 5
Machine learning models reveal features distinguishing CLL non-dominant from healthy control clones. **(A)** Confusion matrices for the Random forest, XGBoost and SVM models, respectively, all showing high accuracy in classification of the clones. **(B)** Receiver operating characteristic (ROC) curves for the three models. Such plots give the true-positive rate (a.k.a. sensitivity, recall or probability of detection) vs. the false-positive rate (a.k.a. the probability of false alarm, or 1 – the specificity). The larger the area under the curve (AUC), the better the model performance is. **(C)** Random forest feature importance scores (proportions of the influence of each feature out of the summed influences of all features), showing which model features are responsible for most of the separation between data groups. **(D)** XGBoost feature importance scores. **(E)** The frequencies of transition mutations in the two data groups; a p-value could not be determined, as the group size was larger than 5000. Confusion matrices and ROC curves were created using Python's scikit-learn package. HC – Healthy controls, Non_dom – Non-dominant CLL clones. RF – Random Forest. XG – XGBoost, SVM – support vector machine. Transversion, Transition – transversion or transition mutation frequency, respectively. Fwr3 – the frequency of mutations in the FWR3 region out of all mutations. Amide_source (target) – the frequency of amide amino acid in pre-(post-) replacement mutation amino acids. Cdr2 (cdr3) – the frequency of mutations in the CDR2 (CDR3) region out of all mutations. Source_c(t) – the frequency of cytosine (thymine) among all pre-mutation nucleotides. Hydro_donor_target – the frequency of hydrogen donating amino acids among post-mutation amino acids in replacement mutations. Hydro_donor_acceptor_source – the frequency of hydrogen donating and accepting amino acids among pre-mutation amino acids in replacement mutations. Negative_source – the frequency of negative amino acids among pre-mutation amino acids in replacement mutations.

Since it is difficult to distinguish between the effects of impairments in SHM *vs.* selection, we used the focused binomial test, which to our knowledge is the only correct test for selection used on lymphomas to date, and found that CLL dominant clones undergo almost no selection *for* replacement mutations in their IgHV gene CDRs. However, dominant clones clearly maintain some selection *against* replacement mutations in their FWRs, although this selection is weaker than that observed in normal healthy controls. Similar alterations in IgHV selection were also found in our studies of other Ig gene mutating B cell malignancies (24, 29, 30). Our finding that CLL clones retain the selection against replacement mutations in their IgHV FWRs indicates a need for IgH transcription, translation, and proper protein folding, and agrees with previous studies showing that CLL tumor clones depend on some type of signals from the BCR complex (56–59).

Several IgHV repertoire studies used ML for classification – to discriminate between IgHV in tumors and those in normal tissues (60), to discriminate between IgHV from celiac patients and healthy individuals (61), or to classify relapsing-remitting multiple sclerosis IgHV CDR3 data from other neurological disease data (62). Here, we used an extensive list of lineage tree-based mutation characteristics to build ML models that could identify minor differences between non-dominant (presumed non-malignant) clones in CLL patients and healthy control trees. Ignoring the dominant clone data, we used ML to look for CLL patient-specific (rather than tumor-specific) impairments in SHM or antigen-driven selection; such information may yield targets for molecular research into what pre-disposes people for CLL and possible other lymphomas. The best ML model classified the non-dominant and healthy control trees with high accuracy, and indicated that CLL non-dominant clones have more transition mutations relative to healthy control clones. Messmer et al. indicated in 2004 that dominant clone CLL IgHV sequences show preference for transitions over transversions (54); our analysis shows for the first time that this preference exists even in the CLL *non*-dominant sequences. Although the non-dominant clones we included in this

study were all small, we cannot exclude the possibility that some of the non-dominant clones we did include were also malignant. There could either be branches of the main tumor that have mutated so far away from it that our algorithm couldn't identify them as related to the main clone, or unrelated CLL or different tumors in the same patient. However, at least some of the non-dominant clones must have been from normal B cells, so we believe that this finding is worth looking into. It is possible that the balance between transitions, which are created *via* simple replication over AID-introduced uracils, and transversions, which are created by several other DNA repair mechanisms, is disrupted in CLL patients (63–65), and that this disruption is somehow linked to the malignancy.

In summary, we present IgHV sequence lineage tree-based analysis of 15 M-CLL patient tumors, in comparison with the same patients' non-dominant and with healthy control B cell clones, and show for the first time that (a) selection *against* replacement mutations is impaired in, but not completely abolished in the FWRs of, CLL dominant clones; SHM mechanisms may also be impaired in some way in CLL clones. (b) Even the non-dominant clones in CLL patients differ from those of healthy controls in various ways, the most notable being that they express higher fractions of transition mutations than healthy control clones. Performing a similar but larger scale study will allow a better understanding of IgHV SHM and selection in M-CLL, and may shed light on the clinical significance of the heterogeneity of M-CLL; the same methods would also be useful for studying any other tumor-related evolutionary processes that can be studied using lineage trees.

Data availability statement

Raw sequence data used for analysis in this study are publicly available at the NCBI Sequencing Read Archive (www.ncbi.nlm.nih.gov/sra) under BioProject number PRJNA887723.

Author contributions

HN designed the research, analyzed the data and wrote the manuscript. JA analyzed the healthy control data. OB and MK contributed the experimental data. RM and MK designed and supervised the research and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by US-Israel Binational Science Foundation (BSF) grant number 20130432 (to RM). The sequencing was funded by Janssen for diagnostic purposes (to MK and OB). HN was supported by a Bar-Ilan University President's Scholarship.

Acknowledgments

The authors are grateful to Dr. Maya Sasson at the Sheba Medical Center for the sequencing.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1115361/full#supplementary-material

References

1. Gemenetzi K, Agathangelidis A, Zaragoza-Infante L, Sofou E, Papaioannou M, Chatzidimitriou A, et al. B cell receptor immunogenetics in B cell lymphomas: Immunoglobulin genes as key to ontogeny and clinical decision making. *Front Oncol* (2020) 10:67. doi: 10.3389/fonc.2020.00067

2. Kikushige Y. Pathogenesis of chronic lymphocytic leukemia and the development of novel therapeutic strategies. *J Clin Exp Hematop* (2020) 60:146–58. doi: 10.3960/jslrt.20036

3. Zelenetz AD, Gordon LI, Wierda WG, Abramson JS, Advani RH, Andreadis CB, et al. Chronic lymphocytic leukemia/small lymphocytic lymphoma, version 1.2015. *JNCCN J Natl Compr Cancer Netw* (2015) 13:326–62. doi: 10.6004/jnccn.2015.0045

4. Arruga F, Gyau BB, Iannello A, Deaglio S, Vitale N, Vaisitti T. Immune response dysfunction in chronic lymphocytic leukemia: Dissecting molecular mechanisms and microenvironmental conditions. *Int J Mol Sci* (2020) 21:1825. doi: 10.3390/ijms21051825

5. Arellano-Llamas AA, Vela-Ojeda J, Hernandez-Caballero A. Chronic lymphocytic leukemia in the SARS-CoV-2 pandemic. *Curr Oncol Rep* (2022) 24:209–13. doi: 10.1007/s11912-022-01198-z

6. 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) 94:1848–54. doi: 10.1182/blood.v94.6.1848

7. 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) 94:1840–7. doi: 10.1182/blood.v94.6.1840

8. Hengeveld PJ, Levin MD, Kolijn PM, Langerak AW. Reading the B-cell receptor immunome in chronic lymphocytic leukemia: Revelations and applications. *Exp Hematol* (2021) 93:14–24. doi: 10.1016/j.exphem.2020.09.194

9. Gerousi M, Laidou S, Gemenetzi K, Stamatopoulos K, Chatzidimitriou A. Distinctive signaling profiles with distinct biological and clinical implications in aggressive CLL subsets with stereotyped B-cell receptor immunoglobulin. *Front Oncol* (2021) 11:771454. doi: 10.3389/fonc.2021.771454

10. Rai KR, Jain P. Chronic lymphocytic leukemia (CLL)–then and now. *Am J Hematol* (2016) 91:330–40. doi: 10.1002/ajh.24282

11. Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci USA* (2008) 105:13081–6. doi: 10.1073/pnas.0801523105

12. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires. *Front Immunol* (2018) 9:224. doi: 10.3389/fimmu.2018.00224

13. Hoehn KB, Fowler A, Lunter G, Pybus OG. The diversity and molecular evolution of B-cell receptors during infection. *Mol Biol Evol* (2016) 33:1147–57. doi: 10.1093/molbev/msw015

14. Sajadi MM, Dashti A, Rikhtegaran Tehrani Z, Tolbert WD, Seaman MS, Ouyang X, et al. Identification of near-pan-neutralizing antibodies against HIV-1 by deconvolution of plasma humoral responses. *Cell* (2018) 173:1783–1795.e14. doi: 10.1016/j.cell.2018.03.061

15. Bedognetti D, Zoppoli G, Massucco C, Zanardi E, Zupo S, Bruzzone A, et al. Impaired response to influenza vaccine associated with persistent memory B cell depletion in non-hodgkin's lymphoma patients treated with rituximab-containing regimens. *J Immunol* (2011) 186:6044–55. doi: 10.4049/jimmunol.1004095

16. Meng W, Zhang B, Schwartz GW, Rosenfeld AM, Ren D, Thome JJC, et al. An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol* (2017) 35:879–86. doi: 10.1038/nbt.3942

17. Arnaout RA, Prak ETL, Schwab N, Rubelt F, Dunn-walters DK, Nemazee D, et al. The future of blood testing is the immunome. *Front Immunol* (2021) 12:626793. doi: 10.3389/fimmu.2021.626793

18. 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) 31:837–45. doi: 10.1038/leu.2016.307

19. Ghraichy M, von Niederhäusern V, Kovaltsuk A, Galson JD, Deane CM, Trück J. Different B cell subpopulations show distinct patterns in their IgH repertoire metrics. *Elife* (2021) 10:e73111. doi: 10.7554/eLife.73111

20. Greiff V, Weber CR, Palme J, Bodenhofer U, Miho E, Menzel U, et al. Learning the high-dimensional immunogenomic features that predict public and private antibody repertoires. *J Immunol* (2017) 199:2985–97. doi: 10.4049/jimmunol.1700594

21. Ostrovsky-berman M, Frankel B, Polak P, Yaari G. Immune2vec: Embedding B / T cell receptor sequences in $\mathbb{R}^N$ using natural language processing. *Front Immunol* (2021) 12:680687. doi: 10.3389/fimmu.2021.680687

22. Neuman H, Arrouasse J, Kedmi M, Cerutti A, Magri G, Mehr R. IgTreeZ, a toolkit for immunoglobulin gene lineage tree-based analysis, reveals CDR3s are crucial for selection analysis. *Front Immunol* (2022) 13:822834. doi: 10.3389/fimmu.2022.822834

23. Abraham RS, Manske MK, Zuckerman NS, Sohni A, Edelman H, Shahaf G, et al. Novel analysis of clonal diversification in blood B cell and bone marrow plasma cell clones in immunoglobulin light chain amyloidosis. *J Clin Immunol* (2007) 27:69–87. doi: 10.1007/s10875-006-9056-9

24. Zuckerman NS, McCann KJ, Ottensmeier CH, Barak M, Shahaf G, Edelman H, et al. Ig gene diversification and selection in follicular lymphoma, diffuse large B cell lymphoma and primary central nervous system lymphoma revealed by lineage tree and mutation analyses. *Int Immunol* (2010) 22:875–87. doi: 10.1093/intimm/dxq441

25. Hershberg U, Uduman M, Shlomchik MJ, Kleinstein SH. Improved methods for detecting selection by mutation analysis of Ig V region sequences. *Int Immunol* (2008) 20:683–94. doi: 10.1093/intimm/dxn026

26. Dunn-Walters DK, Spencer J. Strong intrinsic biases towards mutation and conservation of bases in human IgVH genes during somatic hypermutation prevent statistical analysis of antigen selection. *Immunology* (1998) 95:339–45. doi: 10.1046/j.1365-2567.1998.00607.x

27. Bose B, Sinha S. Problems in using statistical analysis of replacement and silent mutations in antibody genes for determining antigen-driven affinity selection. *Immunology* (2005) 116:172–83. doi: 10.1111/j.1365-2567.2005.02208.x

28. Carlotti E, Wrench D, Rosignoli G, Marzec J, Sangaralingam A, Hazanov L, et al. High throughput sequencing analysis of the immunoglobulin heavy chain gene from flow-sorted B cell sub-populations define the dynamics of follicular lymphoma clonal evolution. *PloS One* (2015) 10:e0134833. doi: 10.1371/journal.pone.0134833

29. Michaeli M, Carlotti E, Hazanov H, Gribben JG, Mehr R. Mutational patterns along different evolution paths of follicular lymphoma. *Front Oncol* (2022) 12:1029995. doi: 10.3389/fonc.2022.1029995

30. Michaeli M, Tabibian-keissar H, Schiby G, Shahaf G, Pickman Y, Hazanov L, et al. Immunoglobulin gene repertoire diversification and selection in the stomach - from gastritis to gastric lymphomas. *Front Immunol* (2014) 5:264. doi: 10.3389/fimmu.2014.00264

31. Iosselevitch I, Tabibian-Keissar H, Barshack I, Mehr R. Gastric DLBCL clonal evolution as function of patient age. *Front Immunol* (2022) 13:957170. doi: 10.3389/fimmu.2022.957170

32. Green MR, Gentles AJ, Nair RV, Irish JM, Kihira S, Liu CL, et al. Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* (2013) 121:1604–11. doi: 10.1182/blood-2012-09-457283

33. Béguelin W, Teater M, Meydan C, Hoehn KB, Phillip JM, Soshnev AA, et al. Mutant EZH2 induces a pre-malignant lymphoma niche by reprogramming the immune response. *Cancer Cell* (2020) 37:655–673.e11. doi: 10.1016/j.ccell.2020.04.004

34. Kedmi M, Neuman H, Bitansky G, Nagar M, Scheinert-Shenhav G, Barshack I, et al. Identifying a malignant B-cell lymphoma clone in peripheral blood using immunoglobulin high-throughput sequencing and lineage tree analysis. *Int J Lab Hematol* (2022) 44:e239–42. doi: 10.1111/ijlh.13906

35. Ghraichy M, Galson JD, Kovaltsuk A, von Niederhäusern V, Pachlopnik Schmid J, Recher M, et al. Maturation of the human immunoglobulin heavy chain repertoire with age. *Front Immunol* (2020) 11:1734. doi: 10.3389/fimmu.2020.01734

36. Vander Heiden JA, Yaari G, Uduman M, Stern JNH, O'connor KC, Hafler DA, et al. PRESTO: A toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* (2014) 30:1930–2. doi: 10.1093/bioinformatics/btu138

37. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: A toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* (2015) 31:3356–8. doi: 10.1093/bioinformatics/btv359

38. Giudicelli V, Chaume D, Lefranc M-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res* (2005) 33(Database issue):D256–D61. doi: 10.1093/nar/gki010

39. Barak M, Zuckerman NS, Edelman H, Unger R, Mehr R. IgTree©: Creating immunoglobulin variable region gene lineage trees. *J Immunol Methods* (2008) 338:67–74. doi: 10.1016/j.jim.2008.06.006

40. Gansner ER, North SC. Open graph visualization system and its applications to software engineering. *Softw - Pract Exp* (2000) 30:1203–33. doi: 10.1002/1097-024X(200009)30:11<1203

41. Sablitzky F, Wildner G, Rajewsky K. Somatic mutation and clonal expansion of B cells in an antigen-driven immune response. *EMBO J* (1985) 4:345–50. doi: 10.1002/j.1460-2075.1985.tb03635.x

42. Lefranc M-P. Antibody informatics: IMGT, the international ImMunoGeneTics information system. *Microbiol Spectr* (2014) 2. doi: 10.1128/microbiolspec.AID-0001-2012

43. Pommié C, Levadoux S, Sabatier R, Lefranc G, Lefranc MP. IMGT standardized criteria for statistical analysis of immunoglobulin V-region amino acid properties. *J Mol Recognit* (2004) 17:17–32. doi: 10.1002/jmr.647

44. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* (1982) 157:105–32. doi: 10.1016/0022-2836(82)90515-0

45. Yaari G, Uduman M, Kleinstein SH. Quantifying selection in high-throughput immunoglobulin sequencing data sets. *Nucleic Acids Res* (2012) 40:10–2. doi: 10.1093/nar/gks457

46. Shahaf G, Barak M, Zuckerman NS, Swerdlin N, Gorfine M, Mehr R. Antigen-driven selection in germinal centers as reflected by the shape characteristics of immunoglobulin gene lineage trees: A large-scale simulation study. *J Theor Biol* (2008) 255:210–22. doi: 10.1016/j.jtbi.2008.08.005

47. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

48. Sain H, Purnami SW. Combine sampling support vector machine for imbalanced data classification. *Proc Comput Sci* (2015) 72:59–66. doi: 10.1016/j.procs.2015.12.105

49. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* (2002) 16:321–57. doi: 10.1613/jair.953

50. Tomek I. Two modifications of CNN. *IEEE Trans Syst Man Cybern* (1976) SMC-6:769–72. doi: 10.1109/TSMC.1976.4309452

51. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* (2011) 12:2825–30.

52. Sneath PHA. Relations between chemical structure and biological activity in peptides. *J Theor Biol* (1966) 12:157–95. doi: 10.1016/0022-5193(66)90112-3

53. Shannon M, Mehr R. Reconciling repertoire shift with affinity maturation: The role of deleterious mutations. *J Immunol* (1999) 162:3950–6. doi: 10.4049/jimmunol.162.7.3950

54. Messmer BT, Albesiano E, Messmer D, Chiorazzi N. The pattern and distribution of immunoglobulin VH gene mutations in chronic lymphocytic leukemia B cells are consistent with the canonical somatic hypermutation process. *Blood* (2004) 103:3490–5. doi: 10.1182/blood-2003-10-3407

55. Petrova VN, Muir L, McKay PF, Vassiliou GS, Smith KGC, Lyons PA, et al. Combined influence of B-cell receptor rearrangement and somatic hypermutation on B-cell class-switch fate in health and in chronic lymphocytic leukemia. *Front Immunol* (2018) 9:1784. doi: 10.3389/fimmu.2018.01784

56. Burger JA. Nurture versus nature: The microenvironment in chronic lymphocytic leukemia. *Hematol 2010 Am Soc Hematol Educ Program Book* (2011) 1:96–103. doi: 10.1182/asheducation-2011.1.96

57. Stevenson FK, Krysov S, Davies AJ, Steele AJ, Packham G. B-cell receptor signaling in chronic lymphocytic leukemia. *Blood* (2011) 118:4313–20. doi: 10.1182/blood-2011-06-338855

58. Woyach JA, Johnson AJ, Byrd JC. The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood* (2012) 120:1175–84. doi: 10.1182/blood-2012-02-362624

59. Efremov DG, Wiestner A, Laurenti L. Novel agents and emerging strategies for targeting the B-cell receptor pathway in CLL. *Mediterr J Hematol Infect Dis* (2012) 4: e2012067. doi: 10.4084/MJHID.2012.067

60. Konishi H, Komura D, Katoh H, Atsumi S, Koda H, Yamamoto A, et al. Capturing the differences between humoral immunity in the normal and tumor environments from repertoire-seq of B-cell receptors using supervised machine learning. *BMC Bioinf* (2019) 20:267. doi: 10.1186/s12859-019-2853-y

61. Shemesh O, Polak P, Lundin KEA, Sollid LM, Yaari G. Machine learning analysis of naïve B-cell receptor repertoires stratifies celiac disease patients and controls. *Front Immunol* (2021) 12:627813. doi: 10.3389/fimmu.2021.627813

62. Ostmeyer J, Christley S, Rounds WH, Toby I, Greenberg BM, Monson NL, et al. Statistical classifiers for diagnosing disease from immune repertoires: A case study using multiple sclerosis. *BMC Bioinf* (2017) 18:401. doi: 10.1186/s12859-017-1814-6

63. Chi X, Li Y, Qiu X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: Mechanism and regulation. *Immunology* (2020) 160:233–47. doi: 10.1111/imm.13176

64. Young C, Brink R. The unique biology of germinal center B cells. *Immunity* (2021) 54:1652–64. doi: 10.1016/j.immuni.2021.07.015

65. Peled JU, Kuang FL, Iglesias-Ussel MD, Roa S, Kalis SL, Goodman MF, et al. The biochemistry of somatic hypermutation. *Annu Rev Immunol* (2008) 26:481–511. doi: 10.1146/annurev.immunol.26.021607.090236

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fonc.2023.1130686

OPEN ACCESS

EDITED BY
Richard Chahwan,
University of Zurich, Switzerland

REVIEWED BY
Efstathia K. Kapsogeorgou,
National and Kapodistrian University of Athens, Greece
Caroline Grönwall,
Karolinska Institutet (KI), Sweden
Joanne Reed,
Westmead Institute for Medical Research, Australia

*CORRESPONDENCE
Anton W. Langerak
a.langerak@erasmusmc.nl

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

RECEIVED 23 December 2022
ACCEPTED 06 March 2023
PUBLISHED 23 March 2023

CITATION
Kolijn PM, Huijser E, Wahadat MJ, van Helden-Meeuwsen CG, van Daele PLA, Brkic Z, Rijntjes J, Hebeda KM, Groenen PJTA, Versnel MA, Thurlings RM and Langerak AW (2023) Extranodal marginal zone lymphoma clonotypes are detectable prior to eMZL diagnosis in tissue biopsies and peripheral blood of Sjögren's syndrome patients through immunogenetics.
Front. Oncol. 13:1130686.
doi: 10.3389/fonc.2023.1130686

Extranodal marginal zone lymphoma clonotypes are detectable prior to eMZL diagnosis in tissue biopsies and peripheral blood of Sjögren's syndrome patients through immunogenetics

P. Martijn Kolijn[1], Erika Huijser[2], M. Javad Wahadat[2,3], Cornelia G. van Helden-Meeuwsen[2], Paul L. A. van Daele[2,4], Zana Brkic[4], Jos Rijntjes[5], Konnie M. Hebeda[5], Patricia J. T. A. Groenen[5], Marjan A. Versnel[2], Rogier M. Thurlings[6] and Anton W. Langerak[1]*

[1]Department of Immunology, Laboratory Medical Immunology, Erasmus MC, Rotterdam, Netherlands, [2]Department of Immunology, Erasmus MC, Rotterdam, Netherlands, [3]Department of Paediatric Rheumatology, Sophia Children's Hospital, Erasmus MC, Rotterdam, Netherlands, [4]Department of Internal Medicine, Division of Clinical Immunology, Erasmus MC, Rotterdam, Netherlands, [5]Department of Pathology, Radboudumc, Nijmegen, Netherlands, [6]Department of Rheumatology, Radboudumc, Nijmegen, Netherlands

Introduction: Activated B cells play a key role in the pathogenesis of primary Sjögren's syndrome (pSS) through the production of autoantibodies and the development of ectopic germinal centers in the salivary glands and other affected sites. Around 5-10% of pSS patients develop B-cell lymphoma, usually extranodal marginal zone lymphomas (eMZL) of the mucosa-associated lymphoid tissue (MALT). The aim of the current study is to investigate if the eMZL clonotype is detectable in prediagnostic blood and tissue biopsies of pSS patients.

Methods/Results: We studied prediagnostic tissue biopsies of three pSS patients diagnosed with eMZL and four pSS controls through immunoglobulin (IG) gene repertoire sequencing. In all three cases, we observed the eMZL clonotype in prediagnostic tissue biopsies. Among controls, we observed transient elevation of clonotypes in two pSS patients. To evaluate if eMZL clonotypes may also be detected in the circulation, we sequenced a peripheral blood mononuclear cell (PBMC) sample drawn at eMZL diagnosis and two years prior to eMZL relapse in two pSS patients. The eMZL clonotype was detected in the peripheral blood prior to diagnosis in both cases. Next, we selected three pSS patients who developed eMZL lymphoma and five additional pSS patients who remained lymphoma-free. We sequenced the IG heavy chain (IGH) gene repertoire in PBMC samples taken a median of three years before eMZL diagnosis. In two out of three eMZL patients, the dominant clonotype in the prediagnostic PBMC samples matched the eMZL clonotype in the diagnostic biopsy. The eMZL clonotypes observed consisted of

stereotypic IGHV gene combinations (IGHV1-69/IGHJ4 and IGHV4-59/IGHJ5) associated with rheumatoid factor activity, a previously reported feature of eMZL in pSS.

Discussion: In conclusion, our results indicate that eMZL clonotypes in pSS patients are detectable prior to overt eMZL diagnosis in both tissue biopsies and peripheral blood through immunogenetic sequencing, paving the way for the development of improved methods of early detection of eMZL.

KEYWORDS

immunogenetics, Sjögren's syndrome, early detection, lymphoma, lymphomagenesis

Introduction

Primary Sjögren's syndrome (pSS) is a systemic autoimmune disease, characterized by impaired secretion of exocrine glands. Activated B cells play a key role in the pathogenesis of pSS, through the production of autoantibodies and the development of ectopic germinal centers in the salivary glands and other affected sites (1, 2). The produced autoantibodies cover a wide spectrum, including antinuclear, anti-Ro/SS-A, anti-La/SS-B and rheumatoid factor (RF) antibodies (3). In addition, 5-10% of pSS patients develop B-cell lymphoma, typically extranodal marginal zone lymphomas (eMZL) of the mucosa-associated lymphoid tissue (MALT), of which around 70% in the salivary glands (4). Lymphoma is the main cause of a decreased survival in pSS (5). Strikingly, parotid eMZL in pSS frequently express antibodies with RF activity, suggesting autoreactivity is an early driver during lymphoma development (6, 7). RF clones were previously shown to be enriched in inflamed tissues, particularly during pSS-related eMZL development (6). Therefore, a putative model for parotid eMZL lymphomagenesis has been proposed where RF clones organize in ectopic germinal center-like structures in a salivary gland, stimulating somatic hypermutation, proliferation and accumulation of driver mutations (8–10).

During their care for patients with pSS clinicians are confronted with a number of dilemmas. Patients are monitored for development of a complicated disease course by rheumatologists or immunologists, with lymphoma considered to be one of the most severe complications (11). The time to development of lymphoma varies with the highest incidence occurring after >8 years follow-up (12). Nonetheless, 50-70% of patients do not develop a complicated disease course, resulting in unnecessary costs for specialist care and anxiety for patients. Another dilemma occurs in the event of a clinical suspicion of lymphoma in patients with persistent general swelling of salivary glands or lymph nodes, or an unexplained mass lesion in other organs. A confident histological diagnosis can be difficult to make as eMZL in pSS develops in a background of chronic inflammation. This can result in uncertain circumstances in which a clear diagnosis cannot be reached, especially in small biopsies. Mortality and morbidity is increased in older patients with co-morbidities, with stage III/IV disease and in those with progression into a more aggressive lymphoma (11). Additionally, lymphoma treatment depends on staging and varies from wait-and-see policy, immunotherapy alone or mild chemotherapy in early stages to intensive chemo-immunotherapy in late stages and progression to diffuse large B-cell lymphoma.

Previously reported risk factors for lymphoma development in pSS include a combination of epidemiological, clinical, laboratory and pathological features. These prognostic markers include male gender, permanent parotid enlargement, lymphadenopathy, mixed monoclonal cryoglobulinemia, leukopenia, RF autoantibodies, low complement levels and an extensive lymphocytic infiltrate in a salivary gland biopsy (termed a high focus score) (4, 11, 13, 14). Some of the other risk factors may emerge as a reflection of pre-lymphomatous conditions, particularly parotid enlargement and cryoglobulinemia. Nonetheless, the specificity and sensitivity of these risk factors for lymphoma development is limited, and these features are not restricted to pSS patients developing eMZL. A composite score of lymphoma biomarkers improved specificity at the cost of sensitivity (71.8% sensitivity and 79% specificity at a score of ≥2) (4, 15). Unfortunately, the lack of clarity resulting from the limited specificity and sensitivity of existing markers can result in a significant diagnostic delay for eMZL diagnosis in pSS patients in the order of months to years, alongside a monitoring burden for pSS patients who never progress to eMZL. Altogether, there is a clear need for novel methods for early detection in pSS patients with high risk features for lymphoma.

Previously, we have shown that chronic lymphocytic leukemia (CLL), an indolent subtype of lymphoma, can be detected over 16 years prior to clinical diagnosis by sequencing the B-cell receptor immunoglobulin (BCR IG) gene repertoire in the peripheral blood (16). Our findings showed promise for early detection of lymphoma and for the use of immunogenetic sequencing for precision oncology. We hypothesize that patient groups at increased risk of lymphoma (such as pSS patients) in particular would benefit from a sensitive and specific method of early detection through immunogenetic sequencing.

Hence, the aims of the current study on pSS patients are: 1) to investigate the presence of the eMZL clonotype in tissue biopsies taken prior to eMZL diagnosis. 2) to perform an unbiased study of IG heavy chain (IGH) gene repertoire dynamics in the peripheral

blood prior to eMZL and 3) to validate the presence of a dominant clonotype using the eMZL clonotype in the eMZL diagnostic biopsy.

Material and methods

Patient selection and sampling

We selected three pSS cases with prediagnostic tissue biopsies dating up to seven years before eMZL diagnosis and 4 pSS controls who did not develop eMZL and sequenced the immunoglobulin (IG) gene repertoire in all tissues. The selected tissue biopsies originated either from the moment of pSS diagnosis (parotid or labial biopsies) or from biopsies taken due to unexplained symptoms, such as a swollen lymph node or parotid gland. Tissue type of diagnostic eMZL biopsy varied based on dissemination of eMZL at diagnosis (parotid, lymph node, liver, bronchial or breast). We then explored the technical feasibility of detection of the eMZL clonotype in the circulation by sequencing the peripheral blood of two pSS patients who had developed eMZL, either at the time of lymphoma diagnosis or 2.3 years prior to eMZL relapse. For the retrospective study on early detection of eMZL in the peripheral blood, we designed a case-control study. We selected three pSS patients who were diagnosed with eMZL and five pSS patient controls matched for age, sex and blood sample availability. All peripheral blood mononuclear cell (PBMC) samples were drawn after pSS diagnosis and prior to eMZL diagnosis. On average, disease duration was longer for controls (12-25 years) than for cases (6-8 years) at time of sampling. A total of sixteen longitudinal PBMC samples (seven from cases, nine from controls) were included in the study, drawn at a median of three years before eMZL diagnosis (interquartile range two years). For two of the three pSS patients with prediagnostic PBMC samples who developed eMZL, a matched diagnostic tissue biopsy was available. A schematic overview of the study design can be found in Supplementary Figure 1. Biopsies were stored in fresh frozen (FF) or formalin fixed and paraffin embedded (FFPE) form. Collection and usage of samples was approved by the IRB of both Erasmus MC and Radboudumc (MEC2011-116; MEC2019-484; MEC2015-1721) and studies were performed in accordance with the Declaration of Helsinki.

DNA isolation and immunogenetic sequencing

Genomic DNA was isolated from tissue biopsies using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO) or from PBMCs using the GenElute™ Blood Genomic DNA Kit (Sigma-Aldrich). A leader-based PCR was utilized to amplify the IGH repertoire of the PBMC samples and the FF tissue biopsies (17). 500 ng of gDNA was given as input for the multiplex PCR for the PBMC samples and 20-40 ng of gDNA for the tissue biopsies. The PCR product was sequenced on the Illumina Miseq platform. For the FFPE samples, an adapted protocol was used to account for fragmentation resulting from DNA crosslinking. DNA from FFPE samples was sequenced using IG clonality protocol through Illumina sequencing (18). The sequencing data were then annotated through the ARResT/Interrogate immunoprofiler (19). A clonotype was defined as a rearrangement with an identical HCDR3 amino acid sequence using the same IGHV-gene and IGHJ-gene. Clonotypes utilizing an IGHV-gene and IGHJ gene combination associated with RF stereotypy (IGHV4-59/IGHJ2, IGHV4-59/IGHJ5, IGHV1-69/IGHJ4, IGHV3-7/IGHJ3) were considered to present with RF-like features. None of the clonotypes with RF-like features in the current study were an identical match to RF clonotypes from literature. A quality control table for the sequencing results can be found in the supplement (Supplementary Table 1).

Data analysis

During data analysis, the abundance of the largest ("dominant") clonotype was contrasted to the background clonotypes by calculating a ratio, i.e. by dividing the abundance of the dominant clonotype by the mean of the abundance of the clonotypes ranked 3^{rd}-7^{th}, as previously reported (20). Subsequently, the ratio was compared between cases and matched controls. Dominant clonotypes identified in the diagnostic biopsies were considered to reflect the malignant clonotype and were tracked back in the prediagnostic IGH repertoire in the peripheral blood. Intraclonal diversification was evaluated using IgIDivA after processing the data through the T cell receptor/immunoglobulin profiler TRIP (21, 22).

Results

eMZL clonotypes are detectable in tissue biopsies of pSS patients before eMZL diagnosis

In order to investigate the prediagnostic presence of eMZL clonotypes, we sequenced prediagnostic tissue biopsies of three pSS patients developing eMZL and four pSS controls who did not develop eMZL. We detected the eMZL clonotype in prediagnostic tissue biopsies for all three pSS patients (Figure 1; Supplementary Table 2). For two patients (pSS1 and pSS2), the eMZL clonotype was detected in the IGH gene repertoire, while for the last patient (pSS3) we detected it in the immunoglobulin kappa-deleting element (IGK-Kde) repertoire, a rearrangement well recognized as a clonal marker in lymphoma (23, 24). In prediagnostic tissue biopsies where the eMZL clonotype was present, the eMZL clonotype was observed with abundances ranging from 4.2%-55% for all patients (Figure 1A). Clonal proliferations exceeding 10% were also detected for two out of four controls, pSS6 and pSS7, though the clonal proliferations did not remain present in the follow up biopsy for pSS6, suggesting dominant clonal proliferations in pSS patients who do not develop eMZL may be transient in nature. Determination of the abundance and longevity

10.3389/fonc.2023.1130686

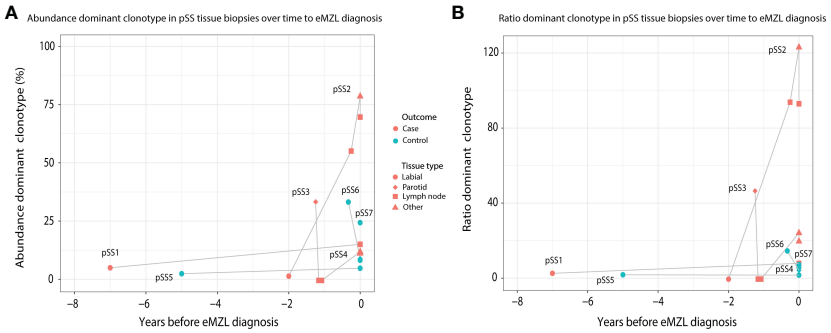


FIGURE 1
Abundance of eMZL clonotype in tissue biopsies obtained from pSS patients up to eMZL development. **(A)** The abundance of the dominant clone as a percentage of total reads in the IG gene repertoire (IGH for pSS1 and pSS2, IGK-Kde for pSS3, IGH for pSS4-7) is shown in various tissue samples of SS patients developing eMZL and pSS controls. **(B)** The ratio of the abundance of the dominant clonotype clonotype compared to the clonotypes ranked 3rd-7th is depicted for both cases and controls. Tissue biopsies are further described in Supplementary Table 2.

of clonal proliferations in pSS patients is essential to evaluate the potential for early detection of eMZL through immunogenetic sequencing. Calculating the ratio of the abundance of the eMZL clonotype compared to the clonotypes ranked 3rd-7th preserved longitudinal dynamics and appears to enhance differentiation between cases and controls in the current cohort, as previously described for other lymphoma subtypes, though our sample size is insufficient to establish a meaningful cutoff (Figure 1B) (20). In all 3 cases, the dominant clonotype observed in the earlier tissue biopsy matched the clonotype observed at eMZL diagnosis (Supplementary Table 2). Notably, for patient pSS3, the eMZL clonotype was only observed in the parotid gland biopsy, while two prediagnostic lymph node biopsies taken several months later showed no dominant clonotype, indicating limited dissemination of the eMZL clonotype. Interestingly, both patients with a detectable prediagnostic eMZL clonotype in the IGH gene repertoire (pSS1 and pSS2) presented with RF-like features (IGHV1-69/IGHJ4), while none of the dominant clonal proliferations in tissue biopsies of controls presented with RF-like features. The dominant eMZL clonotype was detected up to 7 years before diagnosis for patient pSS1.

The eMZL clonotype is detectable in the peripheral blood at diagnosis and prior to relapse

To evaluate the potential for a less invasive approach of early detection of eMZL using peripheral blood of pSS patients, we first sequenced the IGH gene repertoire of PBMC samples of two pSS patients (pSS11 and pSS12) diagnosed with eMZL, in patient pSS12 at the time of eMZL diagnosis and for patient pSS11 2.3 years prior to eMZL relapse. In both patients, we observed skewing of the IGH gene repertoire in the peripheral blood in both abundance and ratio (Figures 2A, B), though the skewing was more pronounced for the ratio. We additionally sequenced diagnostic tissue biopsies for these patients. For the patient with the relapsed eMZL (pSS11), the dominant clonotype identified in the peripheral blood 2.3 years prior to relapse was also an exact match to the malignant clonotype identified at relapse. For the sample drawn at diagnosis of eMZL (pSS12), the 2nd ranked clonotype matched the diagnostic eMZL clonotype (Supplementary Table 3), while the 1st ranked clonotype did not. Neither the 1st ranked clonotype nor the eMZL clonotype displayed RF stereotypy in patient pSS12. In summary, our data indicates that the detection of the eMZL clonotype is feasible in the peripheral blood at eMZL diagnosis and prior to relapse.

The eMZL clonotype is detectable in the peripheral blood prior to diagnosis of eMZL

Based on these results, we hypothesized that the eMZL clonotype may be detected in the circulating cells in the blood prior to diagnosis as well. Therefore, we designed a retrospective case-control pilot study screening the peripheral blood of pSS patients prior to eMZL diagnosis. For two patients (pSS9 and pSS10) a matched diagnostic biopsy at eMZL diagnosis was available, allowing identification of the eMZL clonotype. In both cases, the dominant clonotype in the prediagnostic blood sample matched the eMZL clonotype (Supplementary Table 3). These eMZL clonotypes consisted of IGHV/IGHJ rearrangements with RF-like features (IGHV1-69/IGHJ4 and IGHV4-59/IGHJ5). Furthermore, the eMZL clonotype observed in one of these two patients (pSS10) was already present seven years prior to eMZL diagnosis in the peripheral blood (Figures 2A–C). For the third pSS patient developing eMZL lymphoma (pSS8), no overt prediagnostic skewing was detected whilst the actual eMZL clonotype could not be determined due to the lack of availability of a diagnostic biopsy.

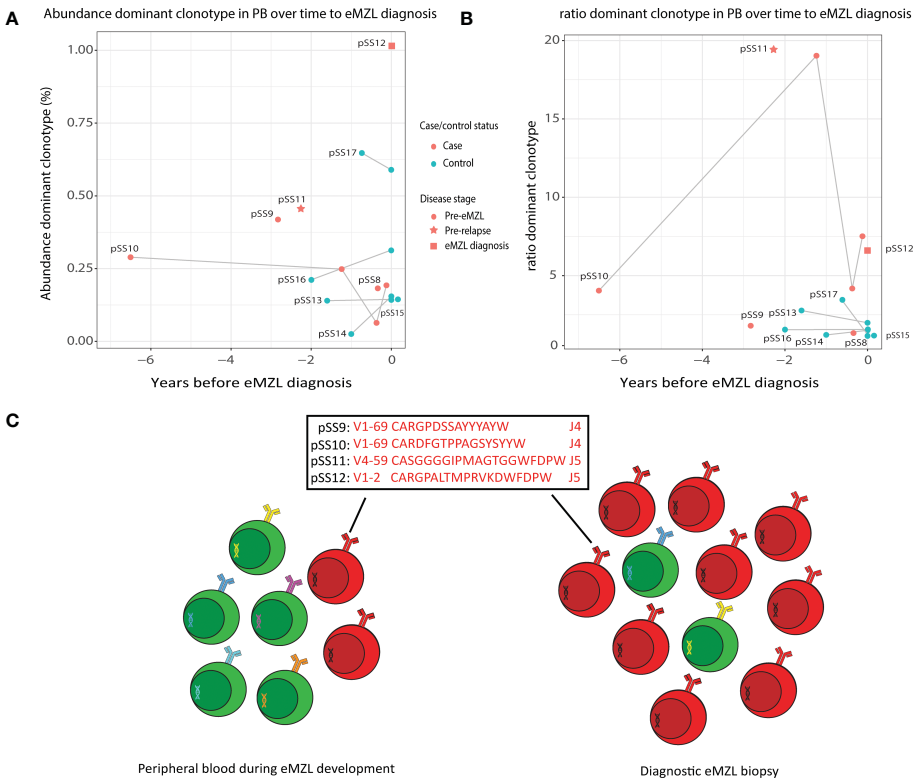


FIGURE 2
Abundance of eMZL clonotype in peripheral blood of pSS patients. **(A)** The abundance (% of total reads in the IG gene repertoire) of the eMZL clonotype in pSS patients is shown over time until eMZL diagnosis in comparison to matched pSS controls. For controls, the abundance of the dominant clonotype is shown. **(B)** The ratio (#1 clonotype divided by the mean of clonotype #3-#7) in pSS patients is shown over time to eMZL diagnosis in comparison to matched pSS controls. PBMC samples and matching diagnostic biopsies are further described in Supplementary Table 3. **(C)** Schematic overview of early detection of premalignant eMZL clonotypes. The eMZL clonotypes identified in both tissue biopsy and the peripheral blood in the current study are highlighted.

The absolute abundance of the eMZL clonotype in the peripheral blood does not significantly differ from the controls but the ratio of the dominant clonotype over the clonotypes ranked 3rd-7th is increased among pSS cases at eMZL diagnosis

The absolute abundance of the eMZL clonotypes in the peripheral blood remained relatively low, at less than 1% of the total IGH gene repertoire and no significant difference was observed in the absolute abundance of the dominant clonotype of cases and controls (Figure 2A). This observation indicates a fundamental difference between prediagnostic tissue biopsies in pSS patients *vs.* the peripheral blood. In prediagnostic tissue biopsies where the eMZL clonotype was present, a dominant clonal proliferation could be observed with abundances ranging from 4.2%-55% for all patients (Figure 1; Supplementary Table 2). Interestingly, similarly to the eMZL clonotypes observed at eMZL diagnosis and prior to relapse in patients pSS11 and pSS12, the eMZL clonotypes were the highest ranked (most abundant) clonotypes present in the peripheral blood of patients pSS9 and pSS10 prior to eMZL diagnosis. To quantify the magnitude of the increase in relative abundance, we calculated a ratio between the abundance of the dominant clonotype in a sample divided by the mean abundance of the clonotypes ranked 3rd-7th (20). We observed a marked increased abundance of the eMZL clonotype compared to the clonotypes ranked 3rd to 7th at diagnosis of eMZL and prior to eMZL diagnosis in patients pSS10, pSS11 and pSS12 (Figure 2B). Thus, the peripheral blood IGH gene repertoire of pSS patients developing lymphoma may not be characterized by a pronounced increase in the absolute abundance of the clonotype of interest, as previously observed for CLL (16), but instead by an increase in the ratio or relative abundance of the eMZL clonotype in comparison to the mean of the abundance of the clonotypes ranked 3rd-7th (Figures 2A, B).

Abundant and persistent clonotypes with RF-like features may have potential as a marker for eMZL development in pSS patients in tissue biopsies and peripheral blood

Based on these interesting findings, we then evaluated whether the abundance of clonotypes with RF-like features (IGHV1-69/IGHJ4, IGHV4-59/IGHJ2, IGHV4-59/IGHJ5 or IGHV3-7/IGHJ3) may have additional potential as a marker for eMZL development. In tissue biopsies, elevation of clonotypes with RF-like features was only

observed for pSS patients developing eMZL (2 out of 3 cases), suggesting highly abundant clonotypes with RF-like features may indeed have potential as a marker for eMZL development (Figure 3A). In circulation, RF-like clonotypes were observed in both cases and controls, albeit at low abundance in comparison to the tissue biopsies (Figure 3B). Interestingly, persistent presence of the same RF-like clonotype as the dominant clonotype (and matching the eMZL clonotype) was a feature restricted to pSS patients developing eMZL, while controls presented with transient elevation of RF-like clonotypes only (Figure 3B). The utility of detection of clonotypes with RF-like features in the peripheral blood for the early detection of eMZL remains unclear, although it would be of interest to explore if an aberrant RF-like clonotype identified in tissue biopsy might be monitored through the peripheral blood as a less invasive alternative.

Intraclonal diversification is limited in the peripheral blood compared to the intraclonal diversification in the eMZL clonotype in the diagnostic biopsy and SHM levels are variable in the peripheral blood over time to diagnosis

Intraclonal diversification as a consequence of high levels of somatic hypermutation is a well described feature of eMZL lymphomas in pSS (6). In the diagnostic tissue biopsies we indeed observed extensive intraclonal diversification (Figures 4B, D). Interestingly, we observed very limited intraclonal diversification within the eMZL clonotype in the peripheral blood (Figures 4A, C). SHM levels in the peripheral blood increased over time to diagnosis of eMZL, while SHM levels in the peripheral blood shortly before diagnosis and at diagnosis matched SHM levels in the diagnostic biopsy, suggesting ongoing SHM during eMZL development (Supplementary Table 3).

Discussion

In this study, we sequenced the IG gene repertoire of tissue biopsies and/or the peripheral blood of eight pSS patients who developed eMZL and nine matched pSS controls. Our results indicate eMZL clonotypes are already present in both tissue biopsies and the peripheral blood prior to eMZL diagnosis. In pSS patients, dendritic cells, T-cells and B-cells react strongly to auto-antigens such as the ribonucleoprotein particles Ro/SS-A and La/SS-B expressed by the epithelium of the exocrine glands. They produce high levels of cytokines and chemokines, resulting in chronic inflammation of the exocrine glands and loss of physiological function (25, 26). The salivary gland epithelium plays a central role in this local autoimmune process by actively stimulating immune cells to accumulate, activate and differentiate (27). The inflammatory microenvironment and activated immune cells then create a vicious cycle by activating the epithelial cells and promoting epithelial cell survival, resulting in a perpetual maintenance and escalation of pSS-associated auto-immune responses (27). A particularly important factor in the context of eMZL development is that epithelial cells may be directly involved in B-cell activation. Salivary gland epithelial cells of pSS patients have been shown to produce BAFF mRNA upon Type 1 IFN stimulation (28). BAFF is essential in promoting B lymphocyte activation and survival and has also been associated with autoimmune B-cell activation (27, 29). B-cells in pSS can be found in focal mononuclear infiltrates and between the ductal epithelial cells in the salivary glands (30, 31). Interestingly, the B-cells found in the focal mononuclear infiltrates express markers associated with memory B-cells and plasma cells, while the B-cells in the lympho-epithelial lesions between the ductal epithelial cells express markers of chronic activation and proliferation in absence of classical memory and plasma cell markers (30, 31). This population of chronically expanding B-cells has been described as the putative source of the eMZLs associated with pSS (32, 33).

One of the most impactful and most studied pathways in pSS is the Type 1 interferon (IFN) pathway (34). The IFN receptor IFNAR is expressed on nearly all cells in the body and its downstream mediators affect the transcription of up to 10 percent of all human genes (35, 36). A remarkable and consistent upregulation of IFN stimulated genes (commonly referred to as the IFN signature) has been reported in pSS (37, 38). A higher expression of IFN stimulated genes is associated with a more severe disease phenotype characterized by B-cell hyperactivity, manifesting through a higher

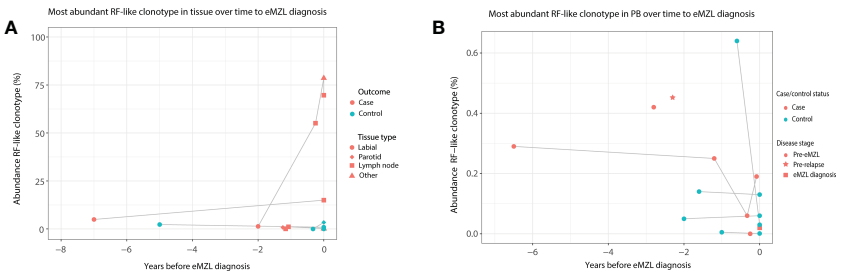


FIGURE 3
Abundance of RF-like clonotypes in cases and controls in tissue and peripheral blood of pSS patients. **(A)** The abundance (% of total reads in the IG gene repertoire) of the most abundant RF-like clonotype in pSS patients is shown over time until eMZL diagnosis in tissue biopsies in comparison to matched pSS controls. **(B)** The abundance (% of total reads in the IG gene repertoire) of the most abundant RF-like clonotype in pSS patients is shown over time until eMZL diagnosis in the peripheral blood in comparison to matched pSS controls.

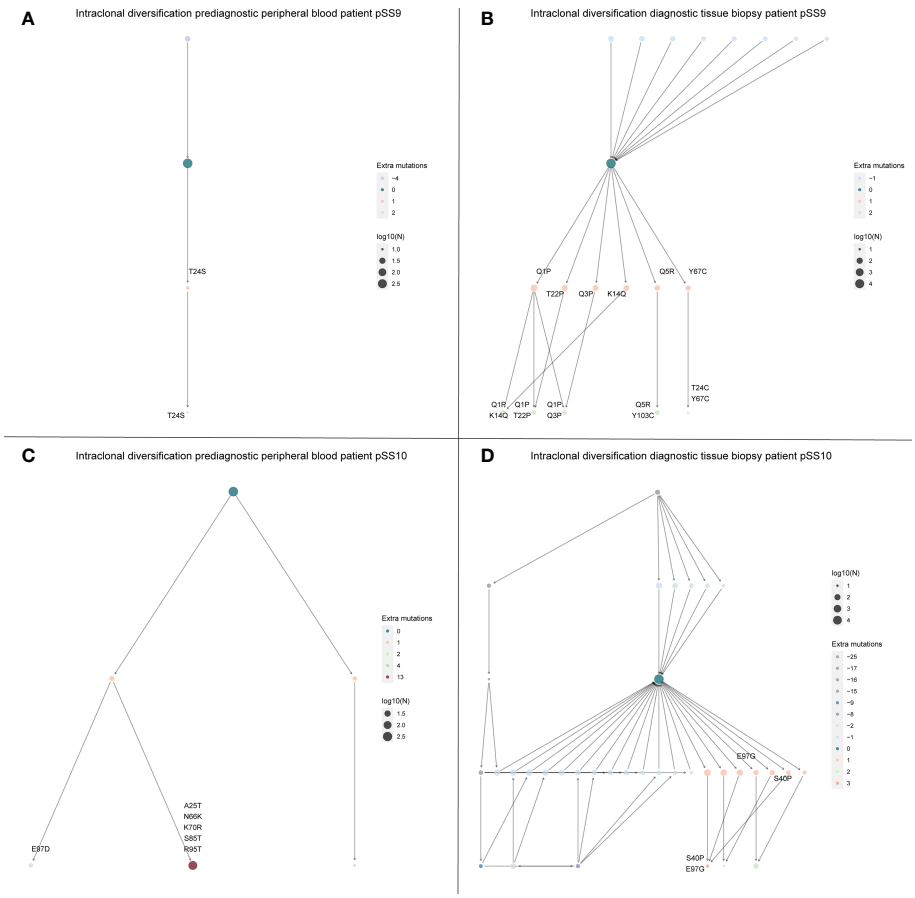


FIGURE 4
Intraclonal diversification of the eMZL clonotype is strongly reduced in the peripheral blood *vs.* the diagnostic tissue material. The intraclonal diversification (ID) of the eMZL clonotype in pSS patients is shown for two patients with prediagnostic PBMC samples available. For these two patients a fresh frozen tissue biopsy was available at eMZL diagnosis. Subclones with reduced mutational load are indicated as precursors, while subclones with additional mutations are indicated as diversification within the eMZL clonotype. **(A)** ID pattern in patient pSS9 in the peripheral blood prior to eMZL diagnosis **(B)** ID pattern in the diagnostic tissue biopsy of patient pSS9. **(C)** ID pattern in patient pSS10 in the peripheral blood prior to eMZL diagnosis. **(D)** ID pattern in patient pSS10 in the diagnostic tissue biopsy of patient pSS10. Log10(N) refers to the number of sequencing reads assigned to each subclone and is used to scale the size of each subclone.

focus score, autoantibody seropositivity, hypergammaglobulinemia, complement consumption, reduced saliva secretion and an increased risk of lymphoma development (39).

The susceptibility of RF clones for lymphomagenesis compared to anti-Ro/La clones has been attributed to differences in affinity maturation. The increased incidence of lymphomagenesis among RF clones likely results from differences in the structure of the recognized autoantigens. Compared to anti-Ro/La clones, RF clones express a restricted public set of immunoglobulin variable regions and an IgM constant region, resulting in extensive intraclonal diversification, which may over time result in formation of cryoglobulins and monoclonal lymphoproliferation (6, 8). In the current study, we observed evidence for this intraclonal diversification in the diagnostic tissue biopsies, but not in the eMZL clonotype in the peripheral blood. Additionally, the degree of intraclonal diversification in the eMZL clonotype in the peripheral blood was stable over time to diagnosis (data not shown). This observation would suggest that intraclonal diversification may hold value as a marker for eMZL development in the context of tissue biopsies but not in the peripheral blood, although this finding should be validated in a larger study.

In the context of this chronic auto-immune inflammatory process, it is perhaps not surprising that malignant B-cell clonotypes are already detectable prior to eMZL diagnosis. It remains unclear at what point after the emergence of these clonotypes progression to eMZL occurs and what drivers are involved in progression. To answer this question it will be essential to ascertain which (genetic) driver events lead to the transformation from auto-immune RF-stereotyped B cells to eMZL and which role the local environment of the parotid gland epithelium plays in this process. Further study of the features of the prediagnostic IG gene repertoire may be an effective method of providing clarification on the dynamics of the expansion during eMZL development, but paired experiments evaluating the local environment will likely be required to achieve new insights. Additionally, in the event of the identification of a clonal population in the initial salivary gland biopsy during pSS diagnosis it may be beneficial to monitor the putative eMZL clonotype in the peripheral blood instead, reducing the burden of

additional biopsies required for monitoring. Another factor to consider is whether there is a difference in the sensitivity of eMZL detection through immunogenetics when sequencing DNA from labial salivary gland biopsies *vs.* parotid gland biopsies.

The detection of a Kde clonotype alone without an IGH rearrangement in the tissue biopsy of one pSS patient (pSS3) highlights the benefit of immunogenetic sequencing beyond the IGH gene repertoire, as in this instance the malignant clonotype would have been missed. Potentially, the extensive degree of somatic hypermutation observed in eMZL has impeded amplification of the IGH rearrangement in this patient (24). The IGK-Kde is a IGK rearrangement in which recombination occurs of a recombination signal sequence 24 kb downstream of IGKC gene with a IGKV gene or with the intron between the IGKJ and IGKC genes. In lymphoma, this rearrangement is well recognized to be used as a clonal marker (23, 24).

RFs present with stereotypic combinations of IGHV and IGHJ genes with shared CDR3 sequence motifs (32, 40, 41). Interestingly, the stereotypy seen in RFs is shared with parotid eMZL developed by pSS patients (7, 32, 40). These eMZLs express one of a specific set of stereotyped combinations: IGHV4-59/IGHJ2, IGHV4-59/IGHJ5, IGHV1-69/IGHJ4, IGHV3-7/IGHJ3 (40). These IGH rearrangements are usually accompanied by a IGK light chain rearrangement involving the IGKV3-15 or IGKV3-20 gene (40). In the current study, five out of eight pSS patients developed eMZL expressing IGHV1-69/IGHJ4 or IGHV4-59/IGHJ5 rearrangements, supporting their putative value as prediagnostic markers for eMZL. Not all eMZL in pSS patients present with RF-like features, as described previously. Additionally, we observed a clonotype with RF-like features (IGHV3-7/IGHJ3) in one control (pSS13) at a ratio of 2.8 (Supplementary Table 3). This clonotype could still be detected 1.6 years later, though it was no longer the clonotype with the highest abundance in the repertoire. Another control (pSS17) presented with a clonotype with RF-like features (IGHV1-69/IGHJ4) at a ratio of 3.5. This clonotype had significantly diminished 7 months later (6-fold reduction in abundance). Therefore, a combination of factors should be considered before characterizing a clonotype as potentially (pre) malignant. These include RF-like stereotypic features, consistent elevation of a clonotype over time and the magnitude of the increase in abundance and/or ratio. Characterization of a clonotype as potentially malignant based on measurement in the circulation alone appears prone to false positives. An interesting potential application of our findings would be to perform routine immunogenetic sequencing of the parotid or labial gland biopsy taken at pSS diagnosis. In the event that a (RF-like) clonotype is detected at an abnormal abundance, this clonotype could then be monitored through the peripheral blood, rather than through repeated biopsies, significantly reducing the monitoring burden for the patient, while putatively detecting eMZL at an earlier stage. Integration of immunogenetic sequencing based risk factors with previously reported risk factors for lymphoma in pSS such as parotid enlargement, cryoglobulinemia, leukopenia, detection of rheumatoid factor autoantibodies and low complement levels may further improve risk stratification (4). In our cohort, pSS patients developing eMZL were characterized by a combination of intermittent parotid enlargement, cryoglobulinemic vasculitis, high focal score (≥4) (13), anti-Ro, anti-La and low complement at time of sampling (seven years to three months before eMZL diagnosis), while most of these risk factors (except anti-Ro) were not reported in the controls (Supplementary Table 4).

The main limitation of the current pilot study is its sample size, preventing further investigation into heterogeneous features of both the lymphomas and included pSS patients. The distinct prediagnostic dynamics of the eMZL clonotypes observed in both the tissue biopsies and the peripheral blood indicate a larger cohort is needed to fully understand the implications of the observed prediagnostic eMZL clonotypes, both for our understanding of lymphomagenesis in pSS patients and to evaluate the potential of these clonotypes for the early detection of eMZL.

In conclusion, eMZL clonotypes can be detected prior to eMZL diagnosis in both tissue biopsies and in the peripheral blood of pSS patients. The context-dependent abnormalities we observed in the IGH gene repertoire of blood and tissues in pSS patients have potential implications for early detection of lymphoma risk. While the absolute abundance of the observed eMZL clonotypes in tissues and peripheral blood was not sufficient to readily distinguish them from clonotypic expansions in controls, features such as RF-stereotypy, ongoing SHM and a persistently increased abundance relative to the other clonotypes in the IGH gene repertoire have the potential to aid in risk stratification and warrant further investigation. Furthermore, our findings highlight the potential of pSS-associated RF clones to transform into eMZL, confirming previous reports. In the age of precision oncology, there is a need for high resolution data and insights on targeted patient groups. While our results suggest potential merit for early detection of eMZL development through immunogenetic sequencing, expansion of the cohort size will be an essential step in validating the intriguing findings from the current pilot study.

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: https://www.ncbi.nlm.nih.gov/, GSE221412.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Review Committee Erasmus MC and Central Committee on Research Involving Human Subjects, Netherlands. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

PMK and JR performed the experiments. PMK and PJTAG analyzed the data. PMK, EH, MJW, CGH-M, PLAD, ZB, JR, KMH, PJTAG, MAV, RMT and AWL interpreted results. PMK and AQL wrote the manuscript. EH, MJW, CGH-M, PLAD, ZB, JR, KMH, PJTAG, MAV, RMT critically reviewed and edited the manuscript. MAV, RMT and AWL designed and supervised the study. All authors contributed to the article and approved the submitted version.

Funding

TRANSCAN / Dutch Cancer Society grant (179; NOVEL consortium).

Acknowledgments

The authors gratefully acknowledge Dr. King H. Lam and Peggy Atmodimedjo of the pathology department of the Erasmus MC for sequencing services and sharing of tissue samples.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1130686/full#supplementary-material

References

1. Risselada AP, Looije MF, Kruize AA, Bijlsma JW, van Roon JA. The role of ectopic germinal centers in the immunopathology of primary sjogren's syndrome: A systematic review. *Semin Arthritis Rheumatol* (2013) 42(4):368–76. doi: 10.1016/j.semarthrit.2012.07.003

2. Espitia-Thibault A, Masseau A, Neel A, Espitia O, Toquet C, Mussini J-M, et al. Sjogren's syndrome-associated myositis with germinal centre-like structures. *Autoimmun Rev* (2017) 16(2):154–8. doi: 10.1016/j.autrev.2016.12.006

3. Martin-Nares E, Hernandez-Molina G. Novel autoantibodies in sjogren's syndrome: A comprehensive review. *Autoimmun Rev* (2019) 18(2):192–8. doi: 10.1016/j.autrev.2018.09.003

4. Retamozo S, Brito-Zeron P, Ramos-Casals M. Prognostic markers of lymphoma development in primary sjogren syndrome. *Lupus* (2019) 28(8):923–36. doi: 10.1177/0961203319857132

5. Baimpa E, Dahabreh IJ, Voulgarelis M, Moutsopoulos HM. Hematologic manifestations and predictors of lymphoma development in primary sjogren syndrome: Clinical and pathophysiologic aspects. *Med (Baltimore)* (2009) 88(5):284–93. doi: 10.1097/MD.0b013e3181b76ab5

6. Broeren MGA, Wang JJ, Balzaretti G, Groenen PJTA, van Schaik BDC, Chataway T, et al. Proteogenomic analysis of the autoreactive b cell repertoire in blood and tissues of patients with sjogren's syndrome. *Ann Rheum Dis* (2022) 81(5):644–52. doi: 10.1136/annrheumdis-2021-221604

7. Bende RJ, Aarts WM, Riedl RG, de Jong D, Pals ST, van Noesel CJM. Among b cell non-hodgkin's lymphomas, MALT lymphomas express a unique antibody repertoire with frequent rheumatoid factor reactivity. *J Exp Med* (2005) 201(8):1229–41. doi: 10.1084/jem.20050068

8. Singh M, Jackson KJL, Wang JJ, Schofield P, Field MA, Koppstein D, et al. Lymphoma driver mutations in the pathogenic evolution of an iconic human autoantibody. *Cell* (2020) 180(5):878–894 e819. doi: 10.1016/j.cell.2020.01.029

9. Salomonsson S, Jonsson MV, Skarstein K, Brokstad KA, Hjelmström P, Wahren-Herlenius M, et al. Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with sjogren's syndrome. *Arthritis Rheumatol* (2003) 48(11):3187–201. doi: 10.1002/art.11311

10. Theander E, Vasaitis L, Baecklund E, Nordmark G, Warfvinge G, Liedholm R, et al. Lymphoid organisation in labial salivary gland biopsies is a possible predictor for the development of malignant lymphoma in primary sjogren's syndrome. *Ann Rheum Diseases* (2011) 70(8):1363–8. doi: 10.1136/ard.2010.144782

11. Brito-Zeron P, Ramos-Casals M, Bove A, Sentis J, Font J. Predicting adverse outcomes in primary sjogren's syndrome: Identification of prognostic factors. *Rheumatol (Oxford)* (2007) 46(8):1359–62. doi: 10.1093/rheumatology/kem079

12. Chiu YH, Chung CH, Lin KT, Lin C-S, Chen J-H, Chen H-C, et al. Predictable biomarkers of developing lymphoma in patients with sjogren syndrome: A nationwide population-based cohort study. *Oncotarget* (2017) 8(30):50098–108. doi: 10.18632/oncotarget.15100

13. Chatzis L, Goules AV, Pezoulas V, Baldini C, Gandolfo S, Skopouli FN, et al. A biomarker for lymphoma development in sjogren's syndrome: Salivary gland focus score. *J Autoimmun* (2021) 121:102648. doi: 10.1016/j.jaut.2021.102648

14. Tzioufas AG, Boumba DS, Skopouli FN, Moutsopoulos HM. Mixed monoclonal cryoglobulinemia and monoclonal rheumatoid factor cross-reactive idiotypes as predictive factors for the development of lymphoma in primary sjogren's syndrome. *Arthritis Rheum* (1996) 39(5):767–72. doi: 10.1002/art.1780390508

15. Quartuccio L, Isola M, Baldini C, Priori R, Bocci EB, Carubbi F, et al. Biomarkers of lymphoma in sjogren's syndrome and evaluation of the lymphoma risk in prelymphomatous conditions: Results of a multicenter study. *J Autoimmun* (2014) 51:75–80. doi: 10.1016/j.jaut.2013.10.002

16. Kolijn PM, Hosnijeh FS, Spath F, Hengeveld PJ, Agathangelidis A, Saleh M, et al. High-risk subtypes of chronic lymphocytic leukemia are detectable as early as 16 years prior to diagnosis. *Blood* (2022) 139(10):1557–63. doi: 10.1182/blood.2021012890

17. Davi F, Langerak AW, de Septenville AL, Kolijn PM, Hengeveld PJ, Chatzidimitriou A, et al. Immunoglobulin gene analysis in chronic lymphocytic leukemia in the era of next generation sequencing. *Leukemia* (2020) 34(10):2545–51. doi: 10.1038/s41375-020-0923-9

18. van Bladel DAG, van der Last-Kempkes JLM, Scheijen B, Groenen P, EuroClonality C. Next-generation sequencing-based clonality detection of immunoglobulin gene rearrangements in b-cell lymphoma. *Methods Mol Biol* (2022) 2453:7–42. doi: 10.1007/978-1-0716-2115-8_2

19. Bystry V, Reigl T, Krejci A, Demko M, Hanakova B, Grioni A, et al. ARResT/Interrogate: an interactive immunoprofiler for IG/TR NGS data. *Bioinformatics* (2017) 33(3):435–7. doi: 10.1093/bioinformatics/btw634

20. van den Brand M, Rijntjes J, Mobs M, Steinhilber J, van der Klift MY, Heezen KC, et al. Next-generation sequencing-based clonality assessment of ig gene rearrangements: A multicenter validation study by EuroClonality-NGS. *J Mol Diagn* (2021) 23(9):1105–15. doi: 10.1016/j.jmoldx.2021.06.005

21. Zaragoza-Infante L, Junet V, Pechlivanis N, Fragkouli S-C, Amprachamian S, Koletsa T, et al. IgIDivA: immunoglobulin intraclonal diversification analysis. *Briefings Bioinf* (2022) 23(5). doi: 10.1093/bib/bbac349

22. Kotouza MT, 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

23. van Dongen JJ, Langerak AW, Bruggemann M, Evans PAS, Hummel M, Lavender FL. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia* (2003) 17(12):2257–317. doi: 10.1038/sj.leu.2403202

24. Evans PAS, Pott C, Groenen PJTA, Salles G, Davi F, Berger F. Significantly improved PCR-based clonality testing in b-cell malignancies by use of multiple

immunoglobulin gene targets. report of the BIOMED-2 concerted action BHM4-CT98-3936. *Leukemia* (2007) 21(2):207–14. doi: 10.1038/sj.leu.2404479

25. Tzioufas AG, Kapsogeorgou EK, Moutsopoulos HM. Pathogenesis of sjogren's syndrome: What we know and what we should learn. *J Autoimmun* (2012) 39(1-2):4–8. doi: 10.1016/j.jaut.2012.01.002

26. Brito-Zeron P, Baldini C, Bootsma H, Bowman SJ, Jonsson R, Mariette X, et al. Sjogren syndrome. *Nat Rev Dis Primers* (2016) 2:16047. doi: 10.1038/nrdp.2016.47

27. Hillen MR, Ververs FA, Kruize AA, Van Roon JA. Dendritic cells, T-cells and epithelial cells: a crucial interplay in immunopathology of primary sjogren's syndrome. *Expert Rev Clin Immunol* (2014) 10(4):521–31. doi: 10.1586/1744666X.2014.878650

28. Ittah M, Miceli-Richard C, Gottenberg JE, Lavie F, Lazure T, Ba N. B cell-activating factor of the tumor necrosis factor family (BAFF) is expressed under stimulation by interferon in salivary gland epithelial cells in primary sjogren's syndrome. *Arthritis Res Ther* (2006) 8(2):R51. doi: 10.1186/ar1912

29. Ittah M, Miceli-Richard C, Lebon P, Pallier C, Lepajolec C, Mariette X. Induction of b cell-activating factor by viral infection is a general phenomenon, but the types of viruses and mechanisms depend on cell type. *J Innate Immunity* (2011) 3(2):200–7. doi: 10.1159/000321194

30. Verstappen GM, Ice JA, Bootsma H, Pringle S, Haacke WA, de Lange K. Gene expression profiling of epithelium-associated FcRL4(+) b cells in primary sjogren's syndrome reveals a pathogenic signature. *J Autoimmun* (2020) 109:102439. doi: 10.1016/j.jaut.2020.102439

31. Haacke EA, Bootsma H, Spijkervet FKL, Visser A, Vissink A, Kluin PM, et al. FcRL4(+) b-cells in salivary glands of primary sjogren's syndrome patients. *J Autoimmun* (2017) 81:90–8. doi: 10.1016/j.jaut.2017.03.012

32. Visser A, Verstappen GM, van der Vegt B, Vissink A, Bende RJ, Bootsma H, et al. Repertoire analysis of b-cells located in striated ducts of salivary glands of patients with sjogren's syndrome. *Front Immunol* (2020) 11:1486. doi: 10.3389/fimmu.2020.01486

33. Verstappen GM, Pringle S, Bootsma H, Kroese FGM. Epithelial-immune cell interplay in primary sjogren syndrome salivary gland pathogenesis. *Nat Rev Rheumatol* (2021) 17(6):333–48. doi: 10.1038/s41584-021-00605-2

34. Gottenberg JE, Cagnard N, Lucchesi C, Letourneur F, Mistou S, Lazure T, et al. Activation of IFN pathways and plasmacytoid dendritic cell recruitment in target organs of primary sjogren's syndrome. *Proc Natl Acad Sci United States America* (2006) 103(13):5242–2. doi: 10.1073/pnas.0510837103

35. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nat Rev Immunol* (2014) 14(1):36–49. doi: 10.1038/nri3581

36. Schoggins JW. Interferon-stimulated genes: What do they all do? *Annu Rev Virol* (2019) 6:567–84. doi: 10.1146/annurev-virology-092818-015756

37. Hillen MR, Pandit A, Blokland SLM, Hartgring SAY, Bekker CPJ, van der Heijden EHM, et al. Plasmacytoid DCs from patients with sjogren's syndrome are transcriptionally primed for enhanced pro-inflammatory cytokine production. *Front Immunol* (2019) 10. doi: 10.1136/annrheumdis-2019-eular.7019

38. Wildenberg ME, van Helden-Meeuwsen CG, van de Merwe JP, Drexhage HA, Versnel MA. Systemic increase in type I interferon activity in sjogren's syndrome: a putative role for plasmacytoid dendritic cells. *Eur J Immunol* (2008) 38(7):2024–33. doi: 10.1002/eji.200738008

39. Brkic Z, Maria NI, van Helden-Meeuwsen CG, van de Merwe JP, van Daele PL, Dalm VA, et al. Prevalence of interferon type I signature in CD14 monocytes of patients with sjogren's syndrome and association with disease activity and BAFF gene expression. *Ann Rheum Dis* (2013) 72(5):728–35. doi: 10.1136/annrheumdis-2012-201381

40. Bende RJ, Janssen J, Beentjes A, Wormhoudt TAM, Wagner K, Haacke EA, et al. Salivary gland mucosa-associated lymphoid tissue-type lymphoma from sjogren's syndrome patients in the majority express rheumatoid factors affinity-selected for IgG. *Arthritis Rheumatol* (2020) 72(8):1330–40. doi: 10.1002/art.41263

41. Falkenburg WJJ, von Richthofen HJ, Rispens T. On the origin of rheumatoid factors: Insights from analyses of variable region sequences. *Semin Arthritis Rheumatol* (2019) 48(4):603–10. doi: 10.1016/j.semarthrit.2018.06.006

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 06 April 2023
DOI 10.3389/fonc.2023.1143811

OPEN ACCESS

EDITED BY
Richard Chahwan,
University of Zurich, Switzerland

REVIEWED BY
Nicholas Chiorazzi,
Feinstein Institute for Medical Research, United States
Elisa Ten Hacken,
Dana–Farber Cancer Institute, United States
Jonathan C. Strefford,
University of Southampton, United Kingdom

*CORRESPONDENCE
Cecilia Österholm
cecilia.osterholm.corbascio@ki.se

†These authors share senior authorship

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

RECEIVED 13 January 2023
ACCEPTED 22 March 2023
PUBLISHED 06 April 2023

CITATION
Oder B, Chatzidimitriou A, Langerak AW, Rosenquist R and Österholm C (2023) Recent revelations and future directions using single-cell technologies in chronic lymphocytic leukemia.
Front. Oncol. 13:1143811.
doi: 10.3389/fonc.2023.1143811

Recent revelations and future directions using single-cell technologies in chronic lymphocytic leukemia

Blaž Oder[1], Anastasia Chatzidimitriou[1,2], Anton W. Langerak[3], Richard Rosenquist[1,4†] and Cecilia Österholm[1*†]

[1]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, [2]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, [3]Department of Immunology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands, [4]Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden

Chronic lymphocytic leukemia (CLL) is a clinically and biologically heterogeneous disease with varying outcomes. In the last decade, the application of next-generation sequencing technologies has allowed extensive mapping of disease-specific genomic, epigenomic, immunogenetic, and transcriptomic signatures linked to CLL pathogenesis. These technologies have improved our understanding of the impact of tumor heterogeneity and evolution on disease outcome, although they have mostly been performed on bulk preparations of nucleic acids. As a further development, new technologies have emerged in recent years that allow high-resolution mapping at the single-cell level. These include single-cell RNA sequencing for assessment of the transcriptome, both of leukemic and non-malignant cells in the tumor microenvironment; immunogenetic profiling of B and T cell receptor rearrangements; single-cell sequencing methods for investigation of methylation and chromatin accessibility across the genome; and targeted single-cell DNA sequencing for analysis of copy-number alterations and single nucleotide variants. In addition, concomitant profiling of cellular subpopulations, based on protein expression, can also be obtained by various antibody-based approaches. In this review, we discuss different single-cell sequencing technologies and how they have been applied so far to study CLL onset and progression, also in response to treatment. This latter aspect is particularly relevant considering that we are moving away from chemoimmunotherapy to targeted therapies, with a potentially distinct impact on clonal dynamics. We also discuss new possibilities, such as integrative multi-omics analysis, as well as inherent limitations of the different single-cell technologies, from sample preparation to data interpretation using available bioinformatic pipelines. Finally, we discuss future directions in this rapidly evolving field.

KEYWORDS

single-cell sequencing, genomics, epigenomics, transcriptomics, immunogenetics, tumor microenvironment, chronic lymphocytic leukaemia

10.3389/fonc.2023.1143811

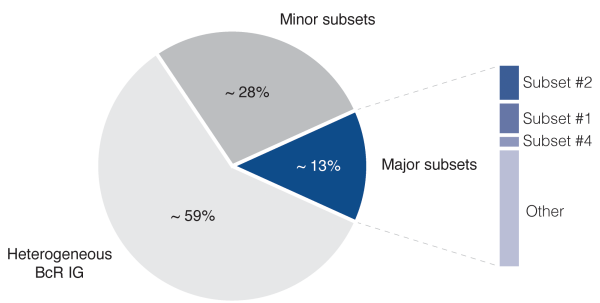
1 Introduction

Chronic lymphocytic leukemia (CLL) is characterized by an expansion of malignant $CD5^+/CD23^+$ B cells, often detected in the peripheral blood of asymptomatic patients (1). The median age at diagnosis is 72 years and men are afflicted more frequently than women. The disease course can range from indolent with a nearly normal life expectancy to aggressive with a poor response to treatment. Although well-established clinical staging systems (2, 3) remain instrumental for determining disease burden, they fail to identify early-stage patients prone to developing aggressive disease. Instead, molecular biomarkers that provide prognostic and/or predictive information have successively been identified. These include i) the immunoglobulin heavy variable (IGHV) gene somatic hypermutation (SHM) status, which divides patients into a poor-prognostic group with unmutated IGHV genes (U-CLL) or a favorable-prognostic group with mutated IGHV genes (M-CLL) (4, 5), and ii) the presence (or absence) of certain genomic lesions, such as deletions of 13q (35-45%), 11q (10-20%), and 17p (5-7%), and trisomy 12 (10-15%), as well as *TP53* mutations (1, 6, 7). These molecular tests are performed for all patients prior to the start of first-line treatment and at subsequent lines of treatment (except the IGHV gene SHM status that is stable throughout the disease course) (1, 8).

The easy access to tumor material from peripheral blood allows for advanced molecular studies of disease progress, from early, pre-cancerous monoclonal B cell lymphocytosis (MBL), to advanced stages of CLL, including Richter's transformation (RT). With the introduction of next-generation sequencing (NGS) technologies more than 10 years ago, the genomic landscape of the different stages of CLL was rapidly uncovered. Today, recurrent genomic alterations have been described in >2000 genes, of which >200 genes have been identified as putative 'drivers'. Of these, >25 genes are associated with clinically aggressive disease, including *ATM, BIRC3, EGR2, NFKBIE, NOTCH1*, *SF3B1*, and *TP53*, among others (9–14). Moreover, CLL is characterized by the expression of almost identical or 'stereotyped' B cell receptor immunoglobulins (BcR IGs) in more than 40% of patients (15). Notably, patients carrying stereotyped BcR IGs can be grouped into distinct subsets that display more similar molecular profiles and clinical outcomes than non-subset CLL patients (16, 17). For instance, patients in subset #1 (utilize IGHV1/5/7 clan I genes, U-CLL) and #2 (IGHV3-21/IGLV3-21, mixed SHM status) respond poorly to chemoimmunotherapy and have a dismal outcome, whereas subset #4 patients (IGHV4-34/IGKV2-30, M-CLL) show indolent disease courses and are rarely in need of treatment (Figure 1) (17, 18). Intriguingly, there is a striking enrichment of specific gene alterations in certain stereotyped subsets, for instance, *SF3B1* mutations in subset #2 and *TP53/NFKBIE/NOTCH1* aberrations in subset #1 (19). This biased acquisition of molecular lesions underscores the importance of both cell-intrinsic and cell-extrinsic factors in CLL pathobiology. In fact, besides the gradual accumulation of genomic lesions, the CLL clone is also dependent on active BcR signaling and interactions with the tumor microenvironment (TME) to promote clonal expansion. However, the TME plays a complex role in CLL pathobiology, where its constituents (including macrophages and their derivatives, mesenchymal stromal cells, and additional lymphocytes), participate in tumor-stimulating, reciprocal signaling, but also suppress anti-tumor immune surveillance mediated primarily by T cells (20–25).

Inhibition of BcR signaling or intrinsic apoptotic pathways by contemporary targeted therapies has shown significant clinical efficacy and prolonged progression-free and overall survival in poor-risk patients with CLL carrying *TP53* aberrations or unmutated IGHV genes (26–29). Nevertheless, the evolution and selection of therapy-resistant subclones may occur during and/or after targeted treatment and lead to disease relapse (Figure 2) (30). Notably, even minor alterations at the subclonal level (e.g., *BTK* and *PLCG2* mutations) are sufficient to drive treatment resistance (Figure 2) (31, 32). Deciphering the clonal evolution of CLL cells under treatment-induced selection pressure is thus critical for a better understanding of resistance mechanisms and the identification of additional predictive biomarkers.

The advent of single-cell sequencing technologies has enabled unprecedented dissection of the intraclonal molecular landscape, further linking genotypes and phenotypes to specific CLL cell subpopulations. In this review, we outline how various single-cell sequencing approaches can be used to unravel intratumoral heterogeneity and track clonal evolution in distinct phases of the disease (Figure 2), but also the contribution of the TME. We highlight the limitations of single-cell technologies and discuss



	Subset #1	Subset #2	Subset #4
% of total CLL	~ 2.2%	~ 2.5%	~ 0.9%
IGHV IGHD IGHJ	IGHV1/5/7 IGHD6-19 IGHJ4	IGHV3-21 Not identified IGHJ6	IGHV4-34 IGHD5-18 IGHJ6
IGKV/IGLV	IGKV1(D)-39	IGLV3-21	IGKV2-30
IGHV status	U-CLL	U-CLL/M-CLL	M-CLL
Recurrent genetic aberrations	*TP53, NFKBIE, NOTCH1*	*SF3B1, ATM*	del(13q)
Clinical course	Aggressive	Aggressive	Indolent

FIGURE 1
Clinicobiological profiles of major CLL subsets #1, #2, and #4. BcR IG, B cell receptor immunoglobulin; U-CLL, CLL with unmutated IGHV genes; M-CLL, CLL with mutated IGHV genes.

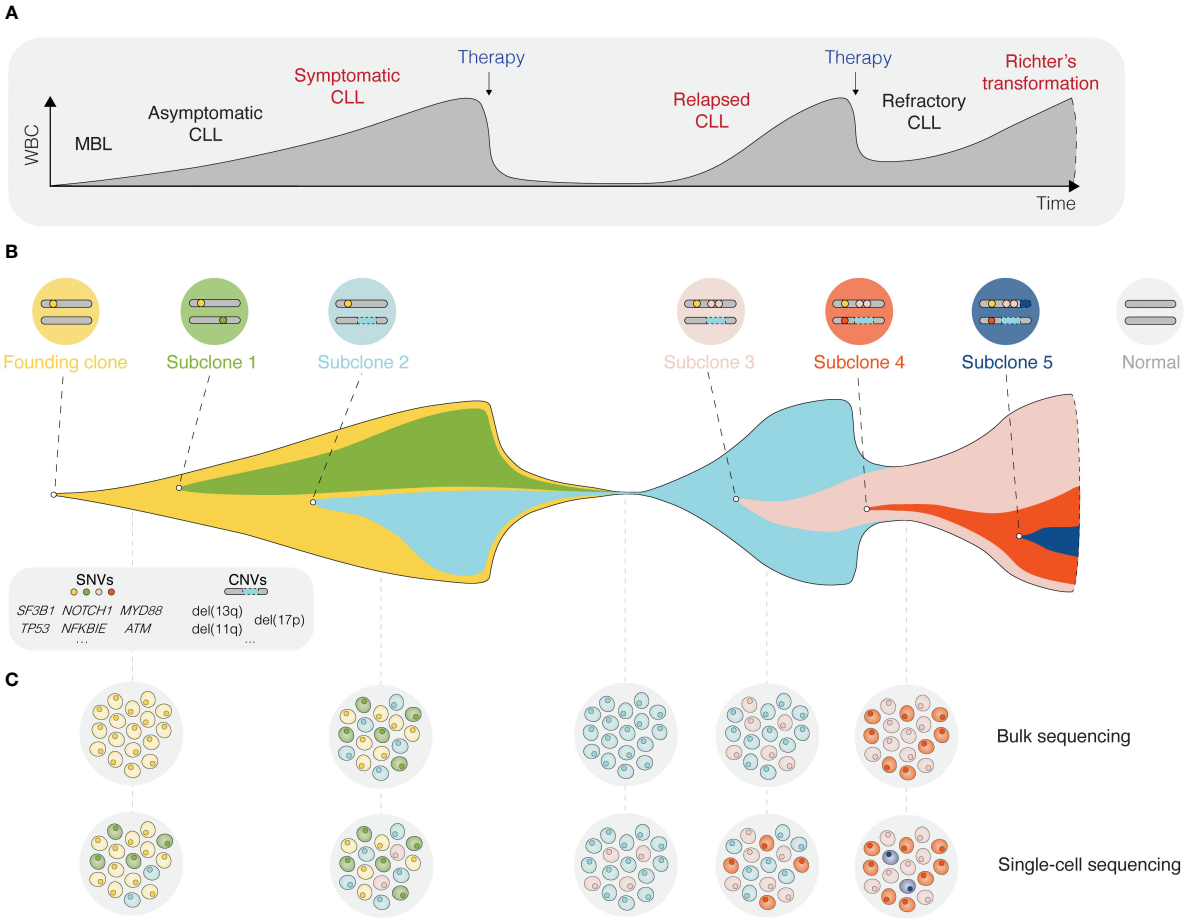


FIGURE 2
Single-cell sequencing deciphers cellular heterogeneity. **(A)** Schematic presentation of the different stages for a typical CLL patient, from the precursor condition monoclonal B cell lymphocytosis (MBL) to relapsed/refractory CLL, and in some cases Richter's transformation. **(B)** The fish plot depicts tumor evolution in which multiple subclones exist within a leukemic cell population. While some clones are eradicated by therapy, others accumulate mutations that confer a clonal advantage. **(C)** The depth of bulk sequencing primarily allows detection of dominant subclones only, whereas single-cell analyses have the potential to detect all subclones, including minor ones as soon as they arise. WBC, white blood count; MBL, monoclonal B cell lymphocytosis; SNV, single nucleotide variant; CNV, copy-number variant.

new directions, such as spatial omics and integrated, multi-omics single-cell analysis.

2 Deciphering clonal heterogeneity and evolution in CLL

As mentioned, CLL is notable for significant genetic diversification and clonal evolution, both during disease progression and upon therapeutic interventions (10, 33, 34). The occurrence of clonal heterogeneity became evident in the 1980-90s, initially by applying traditional cytogenetic techniques, such as chromosome banding analysis and fluorescence *in situ* hybridization (FISH) (35–37). With the development of new technologies, including array- and NGS-based approaches (38, 39), the resolution of detection has increased continuously, enabling to follow clonal dynamics albeit still at a 'bulk' level (40). Frequently used NGS methods include whole-genome sequencing (WGS), whole-exome sequencing (WES), RNA-sequencing (RNA-seq), and analyses of the epigenome, such as methylation profiling and assay for transposase-accessible chromatin with sequencing (ATAC-seq). Using high-resolution genomic technologies, it has been possible to discern early versus late molecular events during CLL pathogenesis (Figure 2). Consequently, we now know that a few genomic aberrations represent early clonal events (e.g., trisomy 12, *MYD88* mutations, and del(13q)), whereas most alterations are present at the subclonal level (10, 12, 41). These technologies have been instrumental in identifying important CLL-specific genomic, epigenomic, and transcriptomic features linked to key dysregulated signaling pathways and cellular processes and have also enabled a multi-layered integrative portrayal of CLL and the discovery of novel subgroups. For instance, in three recent studies, novel patient clusters with distinct clinicobiological features and outcomes were identified using multi-omics approaches, including proteogenomics (12, 14, 42). Additionally, targeted deep-sequencing has become a powerful tool for further in-depth molecular characterization of CLL, allowing for the discovery of previously undetected, smaller-

sized lesions that occur at low frequencies. As a concrete example, minor *TP53*-mutated subclones, undetected by Sanger sequencing but identified by ultra-deep NGS, have been shown to influence clinical outcomes negatively, at least in patients treated with chemoimmunotherapy (43, 44). Furthermore, emerging treatment resistance related to targeted therapy, such as *BTK, PLCG2*, and *BCL2* mutations, can be detected by deep-sequencing or droplet digital PCR (ddPCR) of hotspot positions (45–47).

While these new technologies have advanced the field in terms of deciphering clonal dynamics and treatment resistance in CLL, there are evident limitations when analyzing bulk nucleic acids and proteins from a heterogeneous leukemic sample. Instead, single-cell technologies have opened new possibilities for in-depth studies of clonal diversity in malignant diseases, including CLL.

2.1 Single-cell sequencing technologies

Single-cell sequencing (SCS) technologies, particularly when employed in combination and on longitudinal sets of samples, enable the dissection of subclonal composition and evolutionary dynamics, both in the context of disease progression and response to treatment (48). Moreover, they hold the potential to reveal druggable signaling pathways and mechanisms contributing to treatment resistance. In the following sections, we introduce various SCS technologies and their utility across different modalities (the latter summarized in Figure 3) and discuss their contribution to the understanding of CLL pathogenesis and potential clinical implications. We also provide a timeline of SCS from a technology development aspect, which includes milestones of CLL single-cell research (Figure 4).

2.1.1 Single-cell transcriptomics

One of the first SCS applications was single-cell RNA sequencing (scRNA-seq), which, owing to its ability to assess transcriptomes of individual cells, marked a paradigm shift in cancer research (49–51). The initial scRNA-seq methods, such as Smart-Seq, were low-throughput and relied on sorting single cells into multi-well plates and sequencing of full-length cDNA libraries obtained through whole-transcriptome amplification (WTA) with oligo(dT) priming and template switching (Figure 4) (52, 53). The introduction of unique molecular identifiers (UMIs) facilitated more reliable and absolute quantitation of mRNA molecules with nearly eliminated amplification bias (54). Later, several microwell- and droplet-based scRNA-seq methods based on 3′ cDNA libraries (and less often 5′ cDNA libraries, which are designed to be combined with profiling of cell surface proteins and/or immune repertoire) were developed, such as Seq-Well (55), Drop-seq (56), inDrop (57), and Chromium (Figure 4) (58). These platforms overcame previous limitations of only sequencing tens to hundreds of cells by enabling sequencing of thousands of cells at a time as well as unlocking the potential to capture transcriptomes from heterogeneous cell populations more accurately, but at less comprehensive coverage. For a more detailed overview of scRNA-seq methods, we refer to several extensive review articles (49–51, 59).

In CLL, scRNA-seq has helped to resolve the transcriptomic changes and alternative splicing effects of *SF3B1* mutations, a common subclonal event associated with clinically aggressive disease (Figure 4) (41). In a study by Wang et al., single cells carrying an *SF3B1* mutation possessed significant changes associated with multiple cellular functions, including apoptosis (upregulation of *BIRC3*, *BCL2*, and *KLH21*), DNA damage and cell cycle (increase of *KLF8*, *ATM*, *CDKN2A*, and *CCND1*), telomere maintenance (upregulation of *TERC* and *TERT*), and NOTCH signaling (downregulation of *DTX1* and altered splicing of *DVL2*) (60). Through this study, scRNA-seq was demonstrated to be applicable for in-depth investigation of subclonal events, whose effects could be missed when assessed by 'bulk' sequencing approaches. In a previous study, we used MASC-seq, a method based on single-cell microarray capture of mRNA, to show differences in transcriptional expression patterns among CLL cells from patients classified into distinct stereotyped subsets (i.e., subsets #1, #2, and #4) (Figure 4). Using this approach, we observed major and minor clusters of CLL cells with unique expression signatures in each case (61). Another comprehensive and high-resolution single-cell multi-omics study, in which combined scRNA-seq and 'bulk' ATAC-seq were performed to address ibrutinib treatment response, discovered a tightly regulated ibrutinib-induced signaling program

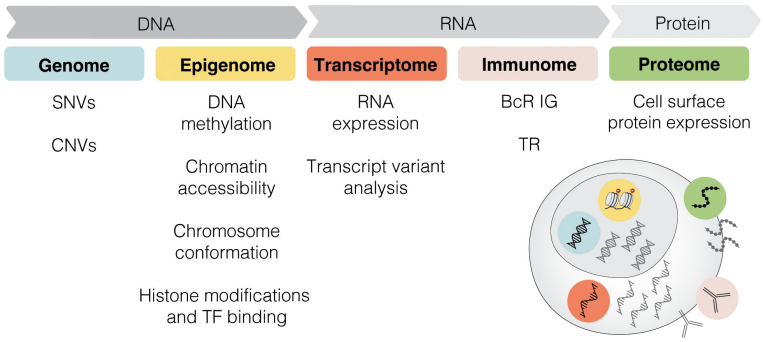


FIGURE 3
Multi-modal profiling. SCS methods can detect alterations across multiple modalities, including genome, epigenome, transcriptome, immunome, and proteome. SNV, single nucleotide variant; CNV, copy-number variant; TF, transcription factor; BcR IG, B cell receptor immunoglobulin; TR, T cell receptor.

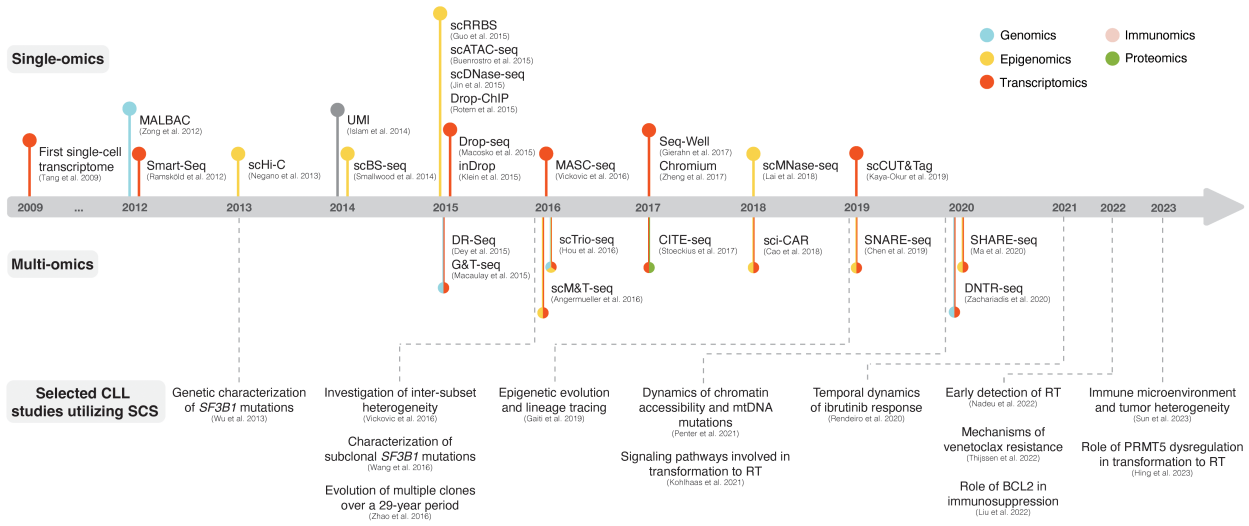


FIGURE 4
The advent of SCS technologies with selected examples of CLL studies. The respective panels above and below the timeline showcase various single-omics and multi-omics technologies that have been developed over the last decade. Selected CLL studies where SCS was employed are exemplified in the lower part of the figure.

of CLL cells (Figure 4). Initially, reduced expression of *BTK*, *CD52* (a CLL disease activity marker), and *CD27* (a memory B cell marker) and diminished chromatin accessibility at NF-κB binding sites were observed, followed by a rapid decrease in the activity of B cell lineage-defining transcription factors (EBF1, FOXM1, IRF4, PAX5, and PU.1), loss of CLL cell identity, and acquisition of a quiescent-like gene signature (upregulation of *CXCR4*, *ZFP36L2*, and *HMGB2)* (62).

2.1.2 Single-cell genomics

In contrast to scRNA-seq approaches, the development of single-cell DNA sequencing (scDNA-seq) methods has proven to be more difficult since a single cell contains only two copies of genomic DNA. Multiple methods for uniform whole-genome amplification (WGA), such as multiple annealing and looping-based amplification cycles (MALBAC), have been developed and enabled single-cell WGS (scWGS) although at a modest throughput (63) (Figure 4). To facilitate charting of the most prevalent gene mutations during clonal evolution, also in the context of CLL (64), high-throughput targeted DNA sequencing using disease-specific gene panels has become available for single-cell analysis. Furthermore, several approaches have been developed for performing simultaneous scRNA-seq and scDNA-seq on the same cell, such as DR-Seq (65) and G&T-seq (66), both of which are based on whole cell lysis and subsequent separation of poly-adenylated RNA from genomic DNA, and direct nuclear tagmentation and RNA-seq (DNTR-seq) (67), in which nucleus and cytosol are physically separated beforehand (Figure 4).

Early efforts to use scDNA-seq in CLL research focused on evaluating *SF3B1* mutations, which had long been assumed to be heterozygous in CLL (68), owing to their typical allelic burden of <50% (Figure 4). Using targeted scDNA-seq, single CLL cells indeed demonstrated heterozygous genotypes, however, a novel subpopulation with homozygous *SF3B1* mutant genotype was discovered, supporting a subclonal evolutionary pathway of *SF3B1* mutations in CLL (69). This study illustrates the applicability of scDNA-seq in detecting subclonal populations not possible to unmask by 'bulk' approaches. Another study, in which scWGS was combined with scRNA-seq and performed on a longitudinal set of samples collected from one patient over a 29-year disease course, demonstrated the power of single-cell approaches for reconstructing cancer evolution based on CNVs and changes in gene expression (Figure 4). Clonal selection in response to treatment was manifested by the disappearance of certain populations and the emergence of a clone with novel CNVs, whereas disease progression was reflected by dynamic transcriptome changes, including upregulation of transcription factors involved in stem cell and cell cycle regulation (*KLF4*, *KLF6*, and *CDKN1A)*, MYD88 signaling (*FOS*, *JUN*, and *NFKBIA*) and downstream of BcR signaling (*REL*, *CDKN1A*, and *NFKBIA*) (70). In a recent study, scDNA-seq of 32 genes, scRNA-seq, and high-throughput immunogenetic analysis were performed on longitudinal samples in patients developing RT, revealing that micro subclones could be identified already at CLL diagnosis up to 19 years before transformation (Figure 4) (64).

2.1.3 Single-cell epigenomics

With the advent of scDNA-seq, SCS approaches for capturing epigenomic alterations have flourished as well and today allow the assessment of methylation dynamics and programs for transcriptional regulation. These technologies were adapted from methods originally applied on bulk nucleic acids and vary based on the epigenomic modality being assayed. Frequently used methods include i) investigation of methylome by single-cell DNA methylation sequencing (scDNAme-seq) approaches that entail bisulfite conversion of genomic DNA, such as scRRBS (71) and scBS-seq (72), ii) analysis of chromatin accessibility by single-cell ATAC-seq (scATAC-seq) (73), single-cell DNase sequencing

(scDNase-seq) (74), and single-cell micrococcal nuclease sequencing (scMNase-seq) (75), iii) exploration of the spatial genome organization and chromatin interactions using chromosome conformation capture, such as single-cell Hi-C (76), and iv) interrogation of histone modifications and transcription factor binding by single-cell chromatin immunoprecipitation followed by sequencing (scChIP-seq), such as Drop-ChIP (77), and single-cell cleavage under targets and tagmentation (scCUT&Tag) (Figure 4) (78–80). Furthermore, multi-omics approaches have been developed for combined capturing of methylome and transcriptome, such as scM&T-seq (81), chromatin accessibility and transcriptome, such as SNARE-seq (82) and SHARE-seq (83), and even combinations of the methylome, genome, and transcriptome, such as scTrio-seq (Figure 4) (84).

While epigenetic studies on bulk material have provided important clues as to the cellular origin of CLL and also identified subtypes with distinct DNA methylation profiles and outcome (85, 86), single-cell epigenomics paves the way for investigation of epigenetic programming and transcriptional regulatory networks in evolving clones. A recent study, in which scDNAme-seq was combined with scRNA-seq, showed that an increased proliferative capacity of CLL cells was reflected in consistently increased epimutation rates with minimal cell-to-cell variability in contrast to healthy B cells (Figure 4) (87). Furthermore, the authors demonstrated that mapping of epimutations can be used as a means for subclone lineage reconstruction and tracing, consistent with previous reports (85, 86). CLL cells with elevated epimutation rates exhibited higher gene expression heterogeneity (also known as transcriptional entropy), consistent with transcriptional dysregulation. Additionally, the investigators demonstrated the enrichment of low epimutation rates in gene promoters for binding motifs of transcription factors with established roles in CLL progression (*NFKB1* and *MYBL1*), and enhancers in proximity to genes implicated in lymphoproliferation (*NOTCH1*, *NFATC1*, and *FOXC1*) and key CLL signaling pathways (e.g, Wnt and MAPK) (87). Another study addressing temporal clonal dynamics employed mitochondrial scATAC-seq and scRNA-seq and revealed that naturally occurring mutations in mitochondrial DNA (mtDNA) could be utilized as biomarkers to distinguish between CLL cell subpopulations with distinct functional states (Figure 4) (88). The presence of mtDNA mutations closely mirrored the disease history and reflected the acquisition of CNVs as well as changes in chromatin accessibility and gene expression, allowing for tracking of existing clones and assessing the emergence of divergent subclones with varying fitness over time, particularly in response to therapy (88).

2.1.4 Single-cell proteomics

The advancement of proteogenomics has allowed for the study of relationships between genetic/transcriptional features and protein expression at 'bulk' level, also in CLL (42, 89, 90). While various single-cell platforms are now commercially available for sequencing nucleic acids, whole-proteome analysis at the single-cell level is still under development. Protein-related difficulties include the wide range of post-translational modifications and the inability of peptides to be amplified. As a result, current efforts are primarily focused on increasingg the signal-to-noise ratio by reducing sample processing volumes and ion contamination (91). Until standardized mass spectrometry-based single-cell proteomics is available, the field is dominated by other techniques. Cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) (92), which uses an antibody-oligonucleotide conjugate-based approach, enables multiplex quantitative profiling of cell surface proteins and has recently been applied in phenotyping of CLL cells for investigating mechanisms of venetoclax resistance, based on short- and long-read scRNA-seq (Figure 4). This study demonstrated a high plasticity of CLL cells in their ability to evade apoptosis upon venetoclax treatment, likely through NF-κB-induced upregulation of the pro-survival protein MCL1 (93).

2.1.5 Single-cell immunomics

Alterations of immunogenetic features constitute a central aspect of clonal evolution in CLL. Using low-throughput sequencing approaches, we have previously demonstrated the occurrence of intraclonal diversification based on SHM patterns of the clonotypic IGH/IGK/IGL gene rearrangements, particularly in subset #4 patients (94, 95). Today, standardized protocols for NGS-based IGH/IGK/IGL sequencing have been developed and enable an in-depth analysis of the clonal composition and dynamics over time (96, 97). In fact, a recent study focusing on stereotyped subset #2 and #169 (a satellite subset to subset #2) demonstrated shared SHM patterns in both subsets at either clonal or subclonal level, reflecting ongoing intraclonal diversification compatible with a branched evolution (98). The recently developed strategy to combine immunomics (99) with global transcriptomics in individual cells, allows for B cell and/or T cell clonotypes to be linked to gene expression signatures (100). Collectively, these technologies facilitate in-depth analysis of immunogenetic features (e.g., ongoing SHM and class-switching) at single-cell level and the discovery of clone-specific phenotypes, which may in turn expedite the identification of therapeutic targets and resistance markers.

2.2 The role of the microenvironment

Major sites for CLL cell propagation are the primary and secondary lymphoid tissues, such as bone marrow (BM) and lymph nodes (LN). The TME, populated also by non-tumor leukocytes and mesenchymal stromal cells, provides stimulatory and anti-apoptotic signals to the malignant clones (21, 101–103) and interplay with tumor cell-intrinsic factors to promote resistance (104). Within the TME, the T cell population is of particular interest, due to its tumor-restricting abilities. In parallel to the biased IG gene repertoire in CLL in general and in patients with stereotyped BcR IGs in particular, the T cell receptor (TR) repertoire is oligoclonal and skewed in terms of the TR beta (TRBV) gene usage, as previously demonstrated by both low-

throughput sequencing analysis and NGS (105–107). Furthermore, similar to other malignancies, T cell exhaustion sustained by continuous antigen exposure and manifested as altered chemokine secretion, reversed ratios of $CD4^+/CD8^+$ cells, altered $CD4^+$ cell helper function, and diminished $CD8^+$ cell cytotoxicity is a common feature of CLL (22, 23, 108, 109).

Concurrent single-cell profiling of TR gene rearrangements and global transcriptomes aids in the investigation of T cell clonality while also elucidating phenotypes and immuno-responsiveness (99, 110). To map transcriptional profiles of T cells in patients with CLL, scRNA-seq was applied on pre-sorted T cells with or without *BCL2* expression. Increased BCL2 levels were suggested to be a marker of T cell dysfunction, and treatment with venetoclax, a BCL2 inhibitor, was able to restore functional T cell immunity by removing BCL2-positive T cells (111). In another study addressing the role of IL-10 receptor signaling in $CD8^+$ T cell exhaustion based on the Eμ-TCL1 mice model, scRNA-seq aided in identifying transcriptional profiles of $CD8^+$ T cells associated with different surface levels of PD1, indicative of an exhausted versus immuno-responsive phenotype (112).

While the abundance of CLL cells in the peripheral blood allows for easy access to tumor material, peripheral blood mononuclear cells (PBMCs) do not represent the active, proliferating tumor population to a large extent. By performing 'bulk' and scRNA-seq on CLL cells isolated from LNs as well as peripheral blood, Sun et al. determined the fraction of activated cells in the LNs to a few percent of the total CLL cell count in the LN compartment. This corroborates previous estimations from *in vivo* labeling of CLL cells using deuterated water (113). A portion of the proliferating cells was also shown to proceed in a unidirectional fashion with mitosis followed by activation and subsequently by quiescence (114). Another study utilizing scRNA-seq on spleen- and LN-derived CLL cells from Eμ-$TCL1^{Akt\text{-}C}$ mice in a model of RT pointed to the importance of sustained Akt signaling for maintaining a pro-proliferative and anti-apoptotic microenvironment through aberrant NOTCH1 activation (115). scRNA-seq was also instrumental in identifying the epigenetic modifier PRMT5 as a potential mediator of RT, in patient tissues as well as in an experimental model employing Eμ-PRMT5/TCL1 mice (116).

While SCS on extracted cells give valuable insight regarding the different disease-driving compartments, attaining information based on spatial context necessitates a preserved histological architecture. The critical role of the stroma and the geographic location of different stromal cell populations in relation to tumor cells (i.e., tumor/stroma boundaries) have been illustrated by spatial multi-omics in solid tumors (117). This pertains not only to direct cell/cell communication but also the composition dynamics of the extracellular matrix and the complex interplay with its producers. Over the last years, the technology has shifted from an initial oligo(dT)-based strategy relying on high-quality, fresh-frozen tissue (118), to gene-specific capture probes with rapidly increasing transcriptome coverage, which makes the currently available platforms for spatial transcriptomics and proteomics suitable for utilization of archival, paraffin-embedded and cryopreserved material containing RNA of compromised integrity (119, 120). In the case of CLL, increased access to detailed information about properties of rapidly dividing tumor cells and accessory cells within the TME will be instrumental to develop new strategies to effectively target the site of birth for CLL cells. Additionally, it will allow for a more meticulous exploration of mechanisms leading to loss of homing and lymphocytosis upon treatment with BTK inhibitors.

3 Discussion

So far, SCS technologies have provided a wide range of options for identifying molecular features at the transcriptomic, genomic, and epigenomic level. In the case of CLL, these methods facilitate the construction of a comprehensive and detailed map of clonal heterogeneity and evolution over time and in response to, above all, targeted treatment (Figure 5). An illustrative example presented above is the ability to identify subclones implicated in RT, which may be present at the time of CLL diagnosis and for up to two decades before clinical manifestation (64). Other important contributions empowered by SCS and which will be valuable for elucidating additional mechanisms behind drug resistance and therapy-induced biological adaptation to BTK and BCL2 inhibitors, include lineage tree reconstruction and identification of key signaling pathways and early markers of progression. Using approaches like this, SCS will help us to better understand which patients are more likely to experience an aggressive disease course, regardless of disease burden at the time of diagnosis. It will also allow us to decipher the subclonal composition of less well studied subgroups of CLL, such as subsets expressing stereotyped BcR IGs, associated with distinct outcomes (Figure 1) (17). Here, an advantage pertains to the ability to perform concomitant single-cell analysis of the transcriptome and expressed IG genes. Additionally, scRNA-seq enables the generation of transcriptional profiles from other accessory cells obtained from PBMCs and lymphoid tissues without the bias implied by prior selection/sorting. Combined with high-resolution spatial omics technologies, this allows for a detailed characterization of the TME with emphasis on pro-proliferative and immunoregulatory properties, and will further aid to identify mechanisms of resistance to contemporary therapies. Some limitations and challenges should be considered when designing studies, preparing samples, and analyzing and interpreting data.

Theoretically, SCS has the potential to detect and in detail investigate minor clones and accessory cells with rare genotypes and/or phenotypes. This, however, necessitates the sequencing of a considerable number of cells, using a sufficient sequencing depth. Therefore, under current circumstances, SCS is less suitable for finding very small subclones or for detection of measurable residual disease due to the current high costs of performing these assays as well as the high resolution and robust output obtained with other established and clinically validated protocols (ddPCR and ultra-deep NGS) on bulk nucleic acids (121–123). Nonetheless, with the anticipated decrease in the cost of sequencing in the coming years, this will enable analyses of a greater number of cells.

Reduced sample viability of cryopreserved cells, due to biological variation or extrinsic factors, may also present a problem as it impacts data quality and reproducibility. A recent study by Massoni-Badosa et al. found that extended storage of

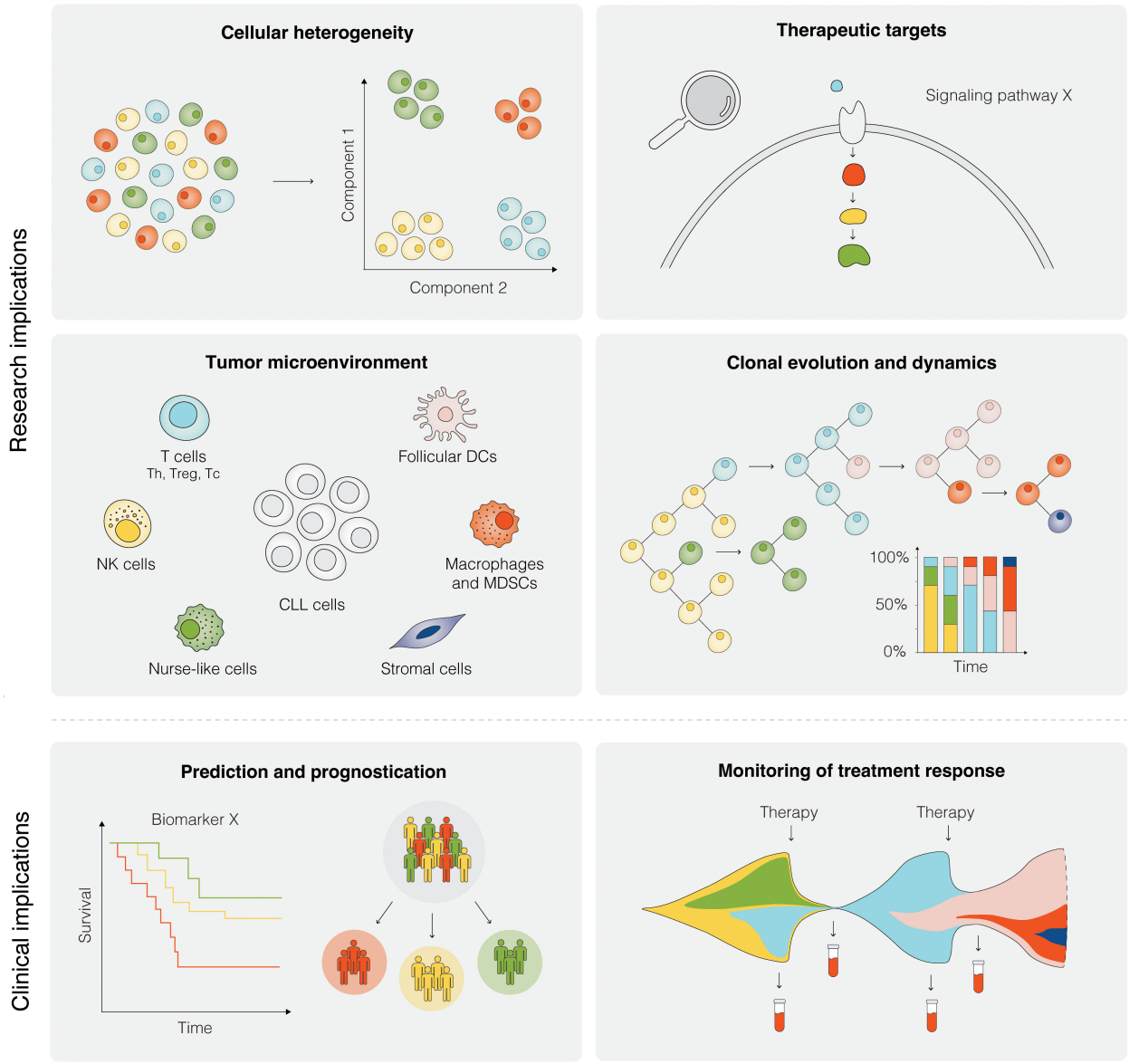


FIGURE 5
Research and clinical implications of SCS in CLL. Research implications of SCS include studies of cellular heterogeneity, identification of potential therapeutic targets, investigation of the TME, and analyses of clonal evolution and dynamics, among others. Clinical implications of SCS are still remote, but could include the prediction of therapeutic intervention, prognostication of clinical outcome, and therapy monitoring.

PBMCs prior to sample preparation and scRNA-seq had a significant impact on gene expression profiles of PBMCs from healthy subjects and CLL patients, even though RNA integrity was preserved during longer storage times. The effect was most noticeable in global gene expression and, to a lesser extent, open chromatin patterns, as measured by scATAC-seq (124). Because simultaneous sampling and the use of freshly harvested samples are generally not possible, the importance of standardized protocols across studies and collaborating centers cannot be overstated (125). Although the issue of initial PBMC storage and processing remains, the recently emerging possibility of using paraformaldehyde-fixed cells for scRNA-seq avoids the challenge of maintaining high cell viability through cryopreservation and transportation (126, 127). Furthermore, if multi-omics is used, different applications may necessitate different sample preparation methods, which in turn requires careful coordination.

SCS technologies have led to a data revolution in CLL research, which inevitably brings challenges also when analyzing and interpreting such data. A major issue usually concerns the lack of appropriate references for the identification of different cell types from heterogeneous samples, although this is not aggravated in CLL where the majority of cells in PBMC samples are leukemic. While several annotation tools based on the expression of cell-type-specific markers have been developed for scRNA-seq, these may not be reliable for the discovery of rare, uncharacterized cell populations or small leukemic subpopulations in heterogeneous samples (128–130). Since such tools are not yet available for most other SCS platforms, cell type assignment is often performed

manually using clustering and dimensionality reduction methods, which limits the reproducibility of the results. False cell type assignment can thus have an impact on downstream analyses such as differential gene expression and lineage tree reconstruction. Although potentially challenging, validation by independent methodologies, such as PCR-based methods, perturbation experiments and flow cytometry, is therefore necessary to ascertain the accuracy and reliability of obtained results. Sparse transcriptomic profiles present another frequent challenge when analyzing scRNA-seq data, and depends not only on the initial relative mRNA abundance but also on technical constraints related to amplification bias, libraries with uneven coverage, and sequencing depth (130–132). Therefore, cell-to-cell variability within the same population must be considered, and absent gene expression should be interpreted cautiously. As there is evidence of aberrant RNA splicing induced by *SF3B1* mutations, as well as an increasing indication of the involvement of long non-coding RNAs and microRNAs in the CLL pathogenesis (13, 133–135), investigating these effects at single-cell resolution has become of considerable interest. However, such analyses of alternative splicing and non-coding RNAs are limited because most scRNA-seq platforms rely on 3' and 5' libraries, which represent merely 3´or 5´parts of transcripts. Despite this, 3´and 5´-based scRNA-seq is advantageous due to a reduced technical noise compared to WTA, and to cost-effectiveness as it requires less sequencing depth to obtain sufficient coverage of gene expression. Nonetheless, for the aforementioned analyses single-cell long-read RNA-seq approaches are gaining momentum and have already demonstrated higher proportions of novel transcripts in CLL (136).

The primary difficulty of scDNA-seq concerns WGA, which can introduce amplification errors and bias towards imbalanced proportions of alleles or even drop out of variant alleles, resulting in unreliable variant detection that consequently hinders characterization of intertumoral heterogeneity and reconstruction of evolutionary history (130). To address uneven genome coverage and challenging variant calling, targeted scDNA-seq, in which only regions of interest are selectively amplified, is gaining popularity and is now commercially available also for many hematological malignancies, including CLL (137, 138). Similarly, analyses of other SCS data may be biased due to inadequate sequencing coverage and depth.

To allow for more systematic and comprehensive studies of CLL pathobiology, in particular clinically aggressive subgroups, integration of various omics is necessary. A new subgroup of patients with aggressive disease (20% of patients) was recently identified by using proteogenomics at 'bulk' level, which could not be identified by genomic analyses alone, emphasizing the value of superimposing and integrating different layers of information (42). Additionally, extended proteomics that accommodates analyses of diverse post-translational modifications, such as phosphorylation and glycosylation, will likely contribute to refining signatures of aggressive and/or therapy-resistant clones, when performed at the single-cell level.

Despite the recent development of rigorous statistical and computational frameworks, such as multi-omics factor analysis (MOFA) (139, 140), omics integration remains a challenge (128, 141). As multi-omics approaches that allow for simultaneous capture of different omics in the same cells are on the rise, data integration may become easier, however, other challenges, such as accounting for dependencies among the measurement types, may emerge. A related data analysis issue concerns comparisons across samples and multiple batches, for which multiple bioinformatics tools have been developed as well (142–144).

Since SCS is still a rapidly evolving field, no methods and bioinformatics pipelines regarded as 'gold standard' currently exist for data analysis, leaving researchers to rely on various options, and base their selection on availability, price, labor intensity, method complexity, and expertise in bioinformatics. Considering that each method has its advantages and disadvantages, the 'right' approach should be carefully selected based on the desired application.

4 Conclusions

The SCS methods described above have extended the range of possibilities for identifying novel signatures and recurrent markers of clonal evolution and treatment resistance in CLL and have also enabled a detailed deciphering of molecular events in anatomical sites that constitute the epicenters of disease progression, more particularly the LNs (Figure 5). While the technologies are under constant development, the current clinical utility of SCS methods *per se* is still in its infancy. Potential future clinical applications include assessment of clonal and microenvironmental composition prior to and during targeted therapy as well as monitoring of treatment response (Figure 5). Nonetheless, through the possibility to capture the genome, epigenome, transcriptome, immunome, and to a limited extent also the proteome in individual cells, SCS signifies a paradigm shift in CLL research. This detailed dissection of the disease at the cellular level will have implications for patient stratification and management in terms of diagnostics, prognostics and tailoring of treatment. As discussed here, SCS poses several challenges; thus, the different aspects of a study, including study design, sample preparation, data analysis, and interpretation should be considered using an integrated approach.

Author contributions

BO designed the review layout, wrote the paper and designed the figures. AC and AL wrote the paper. RR and CÖ designed the review layout and wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was in part supported by from the Swedish Cancer Society (19 0425 Pj 01 H), the Swedish Research Council (2020-01750), the Knut and Alice Wallenberg Foundation (2016.0373), Karolinska Institutet, Karolinska University Hospital, Region Stockholm (ALF/FoUI-962423), and Radiumhemmets Forskningsfonder (194133); 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.

Conflict of interest

RR honoraria/advisory board: Abbvie, AstraZeneca, Janssen, Illumina and Roche. AL honoraria: Janssen, Roche-Genentech.

The remaining 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

References

1. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner 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

2. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood* (1975) 46(2):219–34. doi: 10.1182/blood.V46.2.219.219

3. Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* (1981) 48(1):198–206. doi: 10.1002/1097-0142(19810701)48:1<198::AID-CNCR2820480131>3.0.CO;2-V

4. 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

5. 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):1848–54. doi: 10.1182/blood.V94.6.1848

6. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* (2000) 343(26):1910–6. doi: 10.1056/NEJM200012283432602

7. Zenz T, Eichhorst B, Busch R, Denzel T, Habe S, Winkler D, et al. TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol* (2010) 28(29):4473–9. doi: 10.1200/JCO.2009.27.8762

8. 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

9. Mansouri L, Sutton LA, Ljungstrom V, Bondza S, Arngarden L, Bhoi S, et al. Functional loss of IkappaBepsilon leads to NF-kappaB deregulation in aggressive chronic lymphocytic leukemia. *J Exp Med* (2015) 212(6):833–43. doi: 10.1084/jem.20142009

10. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature* (2015) 526(7574):525–30. doi: 10.1038/nature15395

11. Puente XS, Bea S, Valdes-Mas R, Villamor N, Gutierrez-Abril J, Martin-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* (2015) 526(7574):519–24. doi: 10.1038/nature14666

12. Knisbacher BA, Lin Z, Hahn CK, Nadeu F, Duran-Ferrer M, Stevenson KE, et al. Molecular map of chronic lymphocytic leukemia and its impact on outcome. *Nat Genet* (2022) 54(11):1664–1674. doi: 10.1038/s41588-022-01140-w

13. Mansouri L, Thorvaldsdottir B, Sutton LA, Karakatsoulis G, Meggendorfer M, Parker H, et al. Different prognostic impact of recurrent gene mutations in chronic lymphocytic leukemia depending on IGHV gene somatic hypermutation status: a study by ERIC in HARMONY. *Leukemia* (2022) 37(2):339–347. doi: 10.1038/s41375-022-01802-y

14. Robbe P, Ridout KE, Vavoulis DV, Dreau H, Kinnersley B, Denny N, et al. Whole-genome sequencing of chronic lymphocytic leukemia identifies subgroups with distinct biological and clinical features. *Nat Genet* (2022) 54(11):1675–89. doi: 10.1038/s41588-022-01211-y

15. 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. doi: 10.1182/blood.2020007039

16. 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

17. 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

18. Jaramillo S, Agathangelidis A, Schneider C, Bahlo J, Robrecht S, Tausch E, et al. Prognostic impact of prevalent chronic lymphocytic leukemia stereotyped subsets: analysis within prospective clinical trials of the German CLL study group (GCLLSG). *Haematologica* (2020) 105(11):2598–607. doi: 10.3324/haematol.2019.231027

19. Strefford JC, Sutton LA, Baliakas P, Agathangelidis A, Malcikova J, Plevova K, et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset 2. *Leukemia* (2013) 27(11):2196–9. doi: 10.1038/leu.2013.98

20. Burger JA, Tsukada N, Burger M, Zvaifler NJ, Dell'Aquila M, Kipps TJ. Blood-derived nurse-like cells protect chronic lymphocytic leukemia b cells from spontaneous apoptosis through stromal cell-derived factor-1. *Blood* (2000) 96(8):2655–63. doi: 10.1182/blood.V96.8.2655

21. Fiorcari S, Maffei R, Atene CG, Potenza L, Luppi M, Marasca R. Nurse-like cells and chronic lymphocytic leukemia b cells: A mutualistic crosstalk inside tissue microenvironments. *Cells* (2021) 10(2):217. doi: 10.3390/cells10020217

22. Ramsay AG, Johnson AJ, Lee AM, Gorgun G, Le Dieu R, Blum W, et al. Chronic lymphocytic leukemia T cells show impaired immunological synapse formation that can be reversed with an immunomodulating drug. *J Clin Invest* (2008) 118(7):2427–37. doi: 10.1172/JCI35017

23. Palma M, Gentilcore G, Heimersson K, Mozaffari F, Nasman-Glaser B, Young E, et al. T Cells in chronic lymphocytic leukemia display dysregulated expression of immune checkpoints and activation markers. *Haematologica* (2017) 102(3):562–72. doi: 10.3324/haematol.2016.151100

24. Dumontet E, Mancini SJC, Tarte K. Bone marrow lymphoid niche adaptation to mature b cell neoplasms. *Front Immunol* (2021) 12:784691. doi: 10.3389/fimmu.2021.784691

25. Lagneaux L, Delforge A, Bron D, De Bruyn C, Stryckmans P. Chronic lymphocytic leukemic b cells but not normal b cells are rescued from apoptosis by contact with normal bone marrow stromal cells. *Blood* (1998) 91(7):2387–96. doi: 10.1182/blood.V91.7.2387

26. Byrd JC, Furman RR, Coutre SE, Flinn IW, Burger JA, Blum KA, et al. Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N Engl J Med* (2013) 369(1):32–42. doi: 10.1056/NEJMoa1215637

27. Byrd JC, Harrington B, O'Brien S, Jones JA, Schuh A, Devereux S, et al. Acalabrutinib (ACP-196) in relapsed chronic lymphocytic leukemia. *N Engl J Med* (2016) 374(4):323–32. doi: 10.1056/NEJMoa1509981

28. Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P, et al. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N Engl J Med* (2014) 370(11):997–1007. doi: 10.1056/NEJMoa1315226

29. Roberts AW, Davids MS, Pagel JM, Kahl BS, Puvvada SD, Gerecitano JF, et al. Targeting BCL2 with venetoclax in relapsed chronic lymphocytic leukemia. *N Engl J Med* (2016) 374(4):311–22. doi: 10.1056/NEJMoa1513257

30. Byrd JC, Furman RR, Coutre SE, Burger JA, Blum KA, Coleman M, et al. Three-year follow-up of treatment-naive and previously treated patients with CLL and SLL receiving single-agent ibrutinib. *Blood* (2015) 125(16):2497–506. doi: 10.1182/blood-2014-10-606038

31. Ahn IE, Underbayev C, Albitar A, Herman SE, Tian X, Maric I, et al. Clonal evolution leading to ibrutinib resistance in chronic lymphocytic leukemia. *Blood* (2017) 129(11):1469–79. doi: 10.1182/blood-2016-06-719294

32. Woyach JA, Ruppert AS, Guinn D, Lehman A, Blachly JS, Lozanski A, et al. BTK (C481S)-mediated resistance to ibrutinib in chronic lymphocytic leukemia. *J Clin Oncol* (2017) 35(13):1437–43. doi: 10.1200/JCO.2016.70.2282

33. Burger JA, Landau DA, Taylor-Weiner A, Bozic I, Zhang H, Sarosiek K, et al. Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nat Commun* (2016) 7:11589. doi: 10.1038/ncomms11589

34. Gruber M, Bozic I, Leshchiner I, Livitz D, Stevenson K, Rassenti L, et al. Growth dynamics in naturally progressing chronic lymphocytic leukaemia. *Nature* (2019) 570 (7762):474–9. doi: 10.1038/s41586-019-1252-x

35. Juliusson G, Gahrton G. Abnormal/normal metaphase ratio and prognosis in chronic b-lymphocytic leukemia. *Cancer Genet Cytogenet* (1985) 18(4):307–13. doi: 10.1016/0165-4608(85)90152-9

36. Juliusson G, Oscier DG, Fitchett M, Ross FM, Stockdill G, Mackie MJ, et al. Prognostic subgroups in b-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. *N Engl J Med* (1990) 323(11):720–4. doi: 10.1056/NEJM199009133231105

37. Peterson LC, Lindquist LL, Church S, Kay NE. Frequent clonal abnormalities of chromosome band 13q14 in b-cell chronic lymphocytic leukemia: Multiple clones, subclones, and nonclonal alterations in 82 midwestern patients. *Genes Chromosomes Cancer* (1992) 4(4):273–80. doi: 10.1002/gcc.2870040402

38. Schwaenen C, Nessling M, Wessendorf S, Salvi T, Wrobel G, Radlwimmer B, et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: Development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci U S A.* (2004) 101(4):1039–44. doi: 10.1073/pnas.0304717101

39. Grubor V, Krasnitz A, Troge JE, Meth JL, Lakshmi B, Kendall JT, et al. Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood* (2009) 113 (6):1294–303. doi: 10.1182/blood-2008-05-158865

40. Gunnarsson R, Mansouri L, Isaksson A, Goransson H, Cahill N, Jansson M, et al. Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia. *Haematologica* (2011) 96(8):1161–9. doi: 10.3324/haematol.2010.039768

41. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* (2013) 152(4):714–26. doi: 10.1016/j.cell.2013.01.019

42. Herbst SA, Vesterlund M, Helmboldt AJ, Jafari R, Siavelis I, Stahl M, et al. Proteogenomics refines the molecular classification of chronic lymphocytic leukemia. *Nat Commun* (2022) 13(1):6226. doi: 10.1038/s41467-022-33385-8

43. Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Fama R, et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* (2014) 123(14):2139–47. doi: 10.1182/blood-2013-11-539726

44. Malcikova J, Pavlova S, Kunt Vonkova B, Radova L, Plevova K, Kotaskova J, et al. Low-burden TP53 mutations in CLL: Clinical impact and clonal evolution within the context of different treatment options. *Blood* (2021) 138(25):2670–85. doi: 10.1182/blood.2020009530

45. Mansouri L, Thorvaldsdottir B, Laidou S, Stamatopoulos K, Rosenquist R. Precision diagnostics in lymphomas - recent developments and future directions. *Semin Cancer Biol* (2022) 84:170–83. doi: 10.1016/j.semcancer.2021.10.007

46. Blombery P, Anderson MA, Gong JN, Thijssen R, Birkinshaw RW, Thompson ER, et al. Acquisition of the recurrent Gly101Val mutation in BCL2 confers resistance to venetoclax in patients with progressive chronic lymphocytic leukemia. *Cancer Discovery* (2019) 9(3):342–53. doi: 10.1158/2159-8290.CD-18-1119

47. Woyach JA, Furman RR, Liu TM, Ozer HG, Zapatka M, Ruppert AS, et al. Resistance mechanisms for the bruton's tyrosine kinase inhibitor ibrutinib. *N Engl J Med* (2014) 370(24):2286–94. doi: 10.1056/NEJMoa1400029

48. Nagler A, Wu CJ. The end of the beginning: application of single-cell sequencing to chronic lymphocytic leukemia. *Blood* (2023) 141(4):369–79. doi: 10.1182/blood.2021014669

49. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. *Mol Cell* (2015) 58(4):598–609. doi: 10.1016/j.molcel.2015.05.005

50. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc* (2018) 13(4):599–604. doi: 10.1038/nprot.2017.149

51. Kashima Y, Sakamoto Y, Kaneko K, Seki M, Suzuki Y, Suzuki A. Single-cell sequencing techniques from individual to multiomics analyses. *Exp Mol Med* (2020) 52 (9):1419–27. doi: 10.1038/s12276-020-00499-2

52. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods* (2009) 6(5):377–82. doi: 10.1038/nmeth.1315

53. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* (2012) 30(8):777–82. doi: 10.1038/nbt.2282

54. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* (2014) 11(2):163–6. doi: 10.1038/nmeth.2772

55. Gierahn TM, Wadsworth MH2nd, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* (2017) 14(4):395–8. doi: 10.1038/nmeth.4179

56. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* (2015) 161(5):1202–14. doi: 10.1016/j.cell.2015.05.002

57. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* (2015) 161(5):1187–201. doi: 10.1016/j.cell.2015.04.044

58. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* (2017) 8:14049. doi: 10.1038/ncomms14049

59. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: Current state of the science. *Nat Rev Genet* (2016) 17(3):175–88. doi: 10.1038/nrg.2015.16

60. Wang L, Brooks AN, Fan J, Wan Y, Gambe R, Li S, et al. Transcriptomic characterization of SF3B1 mutation reveals its pleiotropic effects in chronic lymphocytic leukemia. *Cancer Cell* (2016) 30(5):750–63. doi: 10.1016/j.ccell.2016.10.005

61. Vickovic S, Stahl PL, Salmen F, Giatrellis S, Westholm JO, Mollbrink A, et al. Massive and parallel expression profiling using microarrayed single-cell sequencing. *Nat Commun* (2016) 7:13182. doi: 10.1038/ncomms13182

62. Rendeiro AF, Krausgruber T, Fortelny N, Zhao F, Penz T, Farlik M, et al. Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nat Commun* (2020) 11(1):577. doi: 10.1038/s41467-019-14081-6

63. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* (2012) 338 (6114):1622–6. doi: 10.1126/science.1229164

64. Nadeu F, Royo R, Massoni-Badosa R, Playa-Albinyana H, Garcia-Torre B, Duran-Ferrer M, et al. Detection of early seeding of Richter transformation in chronic lymphocytic leukemia. *Nat Med* (2022) 28(8):1662–71. doi: 10.1038/s41591-022-01927-8

65. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol* (2015) 33 (3):285–9. doi: 10.1038/nbt.3129

66. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* (2015) 12 (6):519–22. doi: 10.1038/nmeth.3370

67. Zachariadis V, Cheng H, Andrews N, Enge M. A highly scalable method for joint whole-genome sequencing and gene-expression profiling of single cells. *Mol Cell* (2020) 80(3):541–53 e5. doi: 10.1016/j.molcel.2020.09.025

68. Hahn CN, Scott HS. Spliceosome mutations in hematopoietic malignancies. *Nat Genet* (2011) 44(1):9–10. doi: 10.1038/ng.1045

69. Wu X, Tschumper RC, Jelinek DF. Genetic characterization of SF3B1 mutations in single chronic lymphocytic leukemia cells. *Leukemia* (2013) 27(11):2264–7. doi: 10.1038/leu.2013.155

70. Zhao Z, Goldin L, Liu S, Wu L, Zhou W, Lou H, et al. Evolution of multiple cell clones over a 29-year period of a CLL patient. *Nat Commun* (2016) 7:13765. doi: 10.1038/ncomms13765

71. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res* (2013) 23(12):2126–35. doi: 10.1101/gr.161679.113

72. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* (2014) 11(8):817–20. doi: 10.1038/nmeth.3035

73. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* (2015) 523(7561):486–90. doi: 10.1038/nature14590

74. Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* (2015) 528 (7580):142–6. doi: 10.1038/nature15740

75. Lai B, Gao W, Cui K, Xie W, Tang Q, Jin W, et al. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* (2018) 562(7726):281–5. doi: 10.1038/s41586-018-0567-3

76. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-c reveals cell-to-cell variability in chromosome structure. *Nature* (2013) 502 (7469):59–64. doi: 10.1038/nature12593

77. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* (2015) 33(11):1165–72. doi: 10.1038/nbt.3383

78. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* (2019) 10(1):1930. doi: 10.1038/s41467-019-09982-5

79. Wu SJ, Furlan SN, Mihalas AB, Kaya-Okur HS, Feroze AH, Emerson SN, et al. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol* (2021) 39(7):819–24. doi: 10.1038/s41587-021-00865-z

80. Bartosovic M, Kabbe M, Castelo-Branco G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* (2021) 39(7):825–35. doi: 10.1038/s41587-021-00869-9

81. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* (2016) 13(3):229–32. doi: 10.1038/nmeth.3728

82. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol* (2019) 37(12):1452–7. doi: 10.1038/s41587-019-0290-0

83. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* (2020) 183 (4):1103–16 e20. doi: 10.1016/j.cell.2020.09.056

84. Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* (2016) 26(3):304–19. doi: 10.1038/cr.2016.23

85. Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet* (2012) 44(11):1236–42. doi: 10.1038/ng.2443

86. Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, et al. DNA Methylation dynamics during b cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet* (2016) 48(3):253–64. doi: 10.1038/ng.3488

87. Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, Schulman RC, et al. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* (2019) 569(7757):576–80. doi: 10.1038/s41586-019-1198-z

88. Penter L, Gohil SH, Lareau C, Ludwig LS, Parry EM, Huang T, et al. Longitudinal single-cell dynamics of chromatin accessibility and mitochondrial mutations in chronic lymphocytic leukemia mirror disease history. *Cancer Discovery* (2021) 11(12):3048–63. doi: 10.1158/2159-8290.CD-21-0276

89. Geiger T, Cox J, Mann M. Proteomic changes resulting from gene copy number variations in cancer cells. *PloS Genet* (2010) 6(9):e1001090. doi: 10.1371/journal.pgen.1001090

90. Branca RM, Orre LM, Johansson HJ, Granholm V, Huss M, Perez-Bercoff A, et al. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods* (2014) 11(1):59–62. doi: 10.1038/nmeth.2732

91. Petrosius V, Schoof EM. Recent advances in the field of single-cell proteomics. *Transl Oncol* (2022) 27:101556. doi: 10.1016/j.tranon.2022.101556

92. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* (2017) 14(9):865–8. doi: 10.1038/nmeth.4380

93. Thijssen R, Tian L, Anderson MA, Flensburg C, Jarratt A, Garnham AL, et al. Single-cell multiomics reveal the scale of multilayered adaptations enabling CLL relapse during venetoclax therapy. *Blood* (2022) 140(20):2127–41. doi: 10.1182/blood.2022016040

94. Kostareli E, Sutton LA, Hadzidimitriou A, Darzentas N, Kouvatsi A, Tsaftaris A, et al. Intraclonal diversification of immunoglobulin light chains in a subset of chronic lymphocytic leukemia alludes to antigen-driven clonal evolution. *Leukemia* (2010) 24 (7):1317–24. doi: 10.1038/leu.2010.90

95. Sutton LA, Kostareli E, Hadzidimitriou A, Darzentas N, Tsaftaris A, Anagnostopoulos A, et al. Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen. *Blood* (2009) 114(20):4460–8. doi: 10.1182/blood-2009-05-221309

96. Scheijen B, Meijers RWJ, Rijntjes J, van der Klift MY, Mobs M, Steinhilber J, et al. Next-generation sequencing of immunoglobulin gene rearrangements for clonality assessment: A technical feasibility study by EuroClonality-NGS. *Leukemia* (2019) 33(9):2227–40. doi: 10.1038/s41375-019-0508-7

97. Davi F, Langerak AW, de Septenville AL, Kolijn PM, Hengeveld PJ, Chatzidimitriou A, et al. Immunoglobulin gene analysis in chronic lymphocytic leukemia in the era of next generation sequencing. *Leukemia* (2020) 34(10):2545–51. doi: 10.1038/s41375-020-0923-9

98. Gemenetzi K, Psomopoulos F, Carriles AA, Gounari M, Minici C, Plevova K, et al. Higher-order immunoglobulin repertoire restrictions in CLL: the illustrative case of stereotyped subsets 2 and 169. *Blood* (2021) 137(14):1895–904. doi: 10.1182/blood.2020005216

99. Gupta N, Lindeman I, Reinhardt S, Mariotti-Ferrandiz E, Mujangi-Ebeka K, Martins-Taylor K, et al. Single-cell analysis and tracking of antigen-specific T cells: Integrating paired chain AIRR-seq and transcriptome sequencing: A method by the AIRR community. *Methods Mol Biol* (2022) 2453:379–421. doi: 10.1007/978-1-0716-2115-8_20

100. Rejeski K, Wu Z, Blumenberg V, Kunz WG, Muller S, Kajigaya S, et al. Oligoclonal T-cell expansion in a patient with bone marrow failure after CD19 CAR-T therapy for Richter-transformed DLBCL. *Blood* (2022) 140(20):2175–9. doi: 10.1182/blood.2022017015

101. ten Hacken E, Burger JA. Molecular pathways: targeting the microenvironment in chronic lymphocytic leukemia–focus on the b-cell receptor. *Clin Cancer Res* (2014) 20(3):548–56. doi: 10.1158/1078-0432.CCR-13-0226

102. Herishanu Y, Perez-Galan P, Liu D, Biancotto A, Pittaluga S, Vire B, et al. The lymph node microenvironment promotes b-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood* (2011) 117(2):563–74. doi: 10.1182/blood-2010-05-284984

103. Apollonio B, Ioannou N, Papazoglou D, Ramsay AG. Understanding the immune-stroma microenvironment in b cell malignancies for effective immunotherapy. *Front Oncol* (2021) 11:626818. doi: 10.3389/fonc.2021.626818

104. Bruch PM, Giles HA, Kolb C, Herbst SA, Becirovic T, Roider T, et al. Drug-microenvironment perturbations reveal resistance mechanisms and prognostic subgroups in CLL. *Mol Syst Biol* (2022) 18(8):e10855. doi: 10.15252/msb.202110855

105. Vardi A, Agathangelidis A, Stalika E, Karypidou M, Siorenta A, Anagnostopoulos A, et al. Antigen selection shapes the T-cell repertoire in chronic lymphocytic leukemia. *Clin Cancer Res* (2016) 22(1):167–74. doi: 10.1158/1078-0432.CCR-14-3017

106. Vardi A, Vlachonikola E, Karypidou M, Stalika E, Bikos V, Gemenetzi K, et al. Restrictions in the T-cell repertoire of chronic lymphocytic leukemia: High-throughput immunoprofiling supports selection by shared antigenic elements. *Leukemia* (2017) 31 (7):1555–61. doi: 10.1038/leu.2016.362

107. Vlachonikola E, Stamatopoulos K, Chatzidimitriou A. T Cells in chronic lymphocytic leukemia: A two-edged sword. *Front Immunol* (2020) 11:612244. doi: 10.3389/fimmu.2020.612244

108. Kiaii S, Choudhury A, Mozaffari F, Kimby E, Osterborg A, Mellstedt H. Signaling molecules and cytokine production in T cells of patients with b-cell chronic lymphocytic leukemia (B-CLL): Comparison of indolent and progressive disease. *Med Oncol* (2005) 22(3):291–302. doi: 10.1385/MO:22:3:291

109. Roessner PM, Seiffert M. T-Cells in chronic lymphocytic leukemia: Guardians or drivers of disease? *Leukemia* (2020) 34(8):2012–24. doi: 10.1038/s41375-020-0873-2

110. Daniel B, Yost KE, Hsiung S, Sandor K, Xia Y, Qi Y, et al. Divergent clonal differentiation trajectories of T cell exhaustion. *Nat Immunol* (2022) 23(11):1614–27. doi: 10.1038/s41590-022-01337-5

111. Liu L, Cheng X, Yang H, Lian S, Jiang Y, Liang J, et al. BCL-2 expression promotes immunosuppression in chronic lymphocytic leukemia by enhancing regulatory T cell differentiation and cytotoxic T cell exhaustion. *Mol Cancer* (2022) 21(1):59. doi: 10.1186/s12943-022-01516-w

112. Hanna BS, Llao-Cid L, Iskar M, Roessner PM, Klett LC, Wong JKL, et al. Interleukin-10 receptor signaling promotes the maintenance of a PD-1(int) TCF-1(+) CD8(+) T cell population that sustains anti-tumor immunity. *Immunity* (2021) 54 (12):2825–41 e10. doi: 10.1016/j.immuni.2021.11.004

113. Herndon TM, Chen SS, Saba NS, Valdez J, Emson C, Gatmaitan M, et al. Direct *in vivo* evidence for increased proliferation of CLL cells in lymph nodes compared to bone marrow and peripheral blood. *Leukemia* (2017) 31(6):1340–7. doi: 10.1038/leu.2017.11

114. Sun C, Chen YC, Martinez AZ, Baptista MJ, Pittaluga S, Liu D, et al. The immune microenvironment shapes transcriptional and genetic heterogeneity in chronic lymphocytic leukemia. *Blood Adv* (2022) 7(1):145–158. doi: 10.1182/bloodadvances.2021006941

115. Kohlhaas V, Blakemore SJ, Al-Maarri M, Nickel N, Pal M, Roth A, et al. Active akt signaling triggers CLL toward Richter transformation *via* overactivation of Notch1. *Blood* (2021) 137(5):646–60. doi: 10.1182/blood.2020005734

116. Hing ZA, Walker JS, Whipp EC, Brinton L, Cannon M, Zhang P, et al. Dysregulation of PRMT5 in chronic lymphocytic leukemia promotes progression with high risk of richter's transformation. *Nat Commun* (2023) 14(1):97. doi: 10.1038/s41467-022-35778-1

117. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* (2021) 53 (9):1334–47. doi: 10.1038/s41588-021-00911-1

118. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* (2016) 353(6294):78–82. doi: 10.1126/science.aaf2403

119. Stewart BJ, Fergie M, Young M, Jones C, Sachdeva A, Blain AE, et al. Spatial and molecular profiling of the mononuclear phagocyte network in classic Hodgkin lymphoma. *Blood* (2023). doi: 10.1182/blood.2022015575

120. Mirzazadeh R, Andrusivova Z, Larsson L, Newton PT, Galicia LA, Abalo XM, et al. Spatially resolved transcriptomic profiling of degraded and challenging fresh frozen samples. *Nat Commun* (2023) 14(1):509. doi: 10.1038/s41467-023-36071-5

121. Logan AC, Zhang B, Narasimhan B, Carlton V, Zheng J, Moorhead M, et al. Minimal residual disease quantification using consensus primers and high-throughput IGH sequencing predicts post-transplant relapse in chronic lymphocytic leukemia. *Leukemia* (2013) 27(8):1659–65. doi: 10.1038/leu.2013.52

122. Hengeveld PJ, Levin MD, Kolijn PM, Langerak AW. Reading the b-cell receptor immunome in chronic lymphocytic leukemia: Revelations and applications. *Exp Hematol* (2021) 93:14–24. doi: 10.1016/j.exphem.2020.09.194

123. Hengeveld PJ, van der Klift MY, Kolijn PM, Davi F, Kavelaars FG, de Jonge E, et al. Detecting measurable residual disease beyond 10-4 by an IGHV leader-based NGS approach improves prognostic stratification in CLL. *Blood* (2023) 141(5):519–28. doi: 10.1182/blood.2022017411

124. Massoni-Badosa R, Iacono G, Moutinho C, Kulis M, Palau N, Marchese D, et al. Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol* (2020) 21(1):112. doi: 10.1186/s13059-020-02032-0

125. Hanamsagar R, Reizis T, Chamberlain M, Marcus R, Nestle FO, de Rinaldis E, et al. An optimized workflow for single-cell transcriptomics and repertoire profiling of purified lymphocytes from clinical samples. *Sci Rep* (2020) 10(1):2219. doi: 10.1038/s41598-020-58939-y

126. Chen J, Cheung F, Shi R, Zhou H, Lu W, Consortium CHI. PBMC fixation and processing for chromium single-cell RNA sequencing. *J Transl Med* (2018) 16(1):198. doi: 10.1186/s12967-018-1578-4

127. Phan HV, van Gent M, Drayman N, Basu A, Gack MU, Tay S. High-throughput RNA sequencing of paraformaldehyde-fixed single cells. *Nat Commun* (2021) 12(1):5636. doi: 10.1038/s41467-021-25871-2

128. Yuan GC, Cai L, Elowitz M, Enver T, Fan G, Guo G, et al. Challenges and emerging directions in single-cell analysis. *Genome Biol* (2017) 18(1):84. doi: 10.1186/s13059-017-1218-y

129. Suva ML, Tirosh I. Single-cell RNA sequencing in cancer: Lessons learned and emerging challenges. *Mol Cell* (2019) 75(1):7–12. doi: 10.1016/j.molcel.2019.05.003

130. Lahnemann D, Koster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. *Genome Biol* (2020) 21(1):31. doi: 10.1186/s13059-020-1926-6

131. Vallejos CA, Risso D, Scialdone A, Dudoit S, Marioni JC. Normalizing single-cell RNA sequencing data: Challenges and opportunities. *Nat Methods* (2017) 14 (6):565–71. doi: 10.1038/nmeth.4292

132. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat Rev Genet* (2019) 20(5):273–82. doi: 10.1038/s41576-018-0088-9

133. Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun* (2020) 11(1):1438. doi: 10.1038/s41467-020-15171-6

134. Bhat AA, Younes SN, Raza SS, Zarif L, Nisar S, Ahmed I, et al. Role of non-coding RNA networks in leukemia progression, metastasis and drug resistance. *Mol Cancer* (2020) 19(1):57. doi: 10.1186/s12943-020-01175-9

135. Balatti V, Acunzo M, Pekarky Y, Croce CM. Novel mechanisms of regulation of miRNAs in CLL. *Trends Cancer* (2016) 2(3):134–43. doi: 10.1016/j.trecan.2016.02.005

136. Tian L, Jabbari JS, Thijssen R, Gouil Q, Amarasinghe SL, Voogd O, et al. Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. *Genome Biol* (2021) 22(1):310. doi: 10.1186/s13059-021-02525-6

137. Alberti-Servera L, Demeyer S, Govaerts I, Swings T, De Bie J, Gielen O, et al. Single-cell DNA amplicon sequencing reveals clonal heterogeneity and evolution in T-cell acute lymphoblastic leukemia. *Blood* (2021) 137(6):801–11. doi: 10.1182/blood.2020006996

138. Ruff DW, Dhingra DM, Thompson K, Marin JA, Ooi AT. High-throughput multimodal single-cell targeted DNA and surface protein analysis using the mission bio tapestri platform. *Methods Mol Biol* (2022) 2386:171–88. doi: 10.1007/978-1-0716-1771-7_12

139. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-omics factor analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* (2018) 14(6):e8124. doi: 10.15252/msb.20178124

140. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* (2020) 21(1):111. doi: 10.1186/s13059-020-02015-1

141. Argelaguet R, Cuomo ASE, Stegle O, Marioni JC. Computational principles and challenges in single-cell data integration. *Nat Biotechnol* (2021) 39(10):1202–15. doi: 10.1038/s41587-021-00895-7

142. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM3rd, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177(7):1888–902 e21. doi: 10.1016/j.cell.2019.05.031

143. Hao Y, Hao S, Andersen-Nissen E, Mauck WM3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* (2021) 184(13):3573–87 e29. doi: 10.1016/j.cell.2021.04.048

144. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* (2019) 16(12):1289–96. doi: 10.1038/s41592-019-0619-0

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 18 April 2023
DOI 10.3389/fimmu.2023.1125017

OPEN ACCESS

EDITED BY
Christian Kosan,
Friedrich Schiller University Jena, Germany

REVIEWED BY
Marvyn T. Koning,
Leiden University Medical Center (LUMC), Netherlands
Tomasz Szczepanski,
Medical University of Silesia, Poland

*CORRESPONDENCE
Nikos Darzentas
nikos.darzentas@gmail.com

†These authors have contributed equally to this work and share first authorship

‡These authors share senior authorship

SPECIALTY SECTION
This article was submitted to B Cell Biology, a section of the journal Frontiers in Immunology

RECEIVED 28 December 2022
ACCEPTED 27 March 2023
PUBLISHED 18 April 2023

CITATION
Darzentas F, Szczepanowski M, Kotrová M, Hartmann A, Beder T, Gökbuget N, Schwartz S, Bastian L, Baldus CD, Pál K, Darzentas N and Brüggemann M (2023) Insights into IGH clonal evolution in BCP-ALL: frequency, mechanisms, associations, and diagnostic implications. *Front. Immunol.* 14:1125017. doi: 10.3389/fimmu.2023.1125017

Insights into IGH clonal evolution in BCP-ALL: frequency, mechanisms, associations, and diagnostic implications

Franziska Darzentas[1†], Monika Szczepanowski[1†], Michaela Kotrová[1], Alina Hartmann[1,2,3], Thomas Beder[1,2], Nicola Gökbuget[4], Stefan Schwartz[5,6], Lorenz Bastian[1,2,3], Claudia Dorothea Baldus[1,2,3], Karol Pál[7], Nikos Darzentas[1‡*] and Monika Brüggemann[1,2,3‡]

[1]Medical Department II, Hematology and Oncology, University Hospital Schleswig-Holstein, Kiel, Germany, [2]University Cancer Center Schleswig-Holstein (UCCSH), University Hospital Schleswig-Holstein, Kiel, Germany, [3]Clinical Research Unit "CATCH-ALL" (KFO 5010/1), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Bonn, Germany, [4]Department of Medicine II, Hematology/Oncology, Goethe University Hospital, Frankfurt/M, Germany, [5]Department of Hematology, Oncology and Tumor Immunology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany, [6]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, [7]Central European Institute of Technology, Masaryk University, Brno, Czechia

Introduction: The malignant transformation leading to a maturation arrest in B-cell precursor acute lymphoblastic leukemia (BCP-ALL) occurs early in B-cell development, in a pro-B or pre-B cell, when somatic recombination of variable (V), diversity (D), and joining (J) segment immunoglobulin (IG) genes and the B-cell rescue mechanism of V_H replacement might be ongoing or fully active, driving clonal evolution. In this study of newly diagnosed BCP-ALL, we sought to understand the mechanistic details of oligoclonal composition of the leukemia at diagnosis, clonal evolution during follow-up, and clonal distribution in different hematopoietic compartments.

Methods: Utilizing high-throughput sequencing assays and bespoke bioinformatics we identified BCP-ALL-derived clonally-related IGH sequences by their shared 'DNJ-stem'.

Results: We introduce the concept of 'marker DNJ-stem' to cover the entirety of, even lowly abundant, clonally-related family members. In a cohort of 280 adult patients with BCP-ALL, IGH clonal evolution at diagnosis was identified in one-third of patients. The phenomenon was linked to contemporaneous recombinant and editing activity driven by aberrant ongoing D_H/V_H-DJ_H recombination and V_H replacement, and we share insights and examples for both. Furthermore, in a subset of 167 patients with molecular subtype allocation, high prevalence and high degree of clonal evolution driven by ongoing D_H/V_H-DJ_H recombination were associated with the presence of *KMT2A* gene rearrangements, while V_H replacements occurred more frequently in Ph-like and DUX4 BCP-ALL. Analysis of 46 matched diagnostic bone marrow and

peripheral blood samples showed a comparable clonal and clonotypic distribution in both hematopoietic compartments, but the clonotypic composition markedly changed in longitudinal follow-up analysis in select cases. Thus, finally, we present cases where the specific dynamics of clonal evolution have implications for both the initial marker identification and the MRD monitoring in follow-up samples.

Discussion: Consequently, we suggest to follow the marker DNJ-stem (capturing all family members) rather than specific clonotypes as the MRD target, as well as to follow both VDJ_H and DJ_H family members since their respective kinetics are not always parallel. Our study further highlights the intricacy, importance, and present and future challenges of IGH clonal evolution in BCP-ALL.

KEYWORDS

acute lymphoblastic leukemia, clonal evolution, minimal residual disease, DNJ-stem, V_H replacement, IGH rearrangements, high-throughput sequencing

1 Introduction

The immense diversity of the antibody repertoire in humans is physiologically conveyed by mechanistic editing and assembly of immunoglobulin (IG) genes, particularly in the processes of somatic recombination of variable (V), diversity (D), and joining (J) segment genes. In the early developmental stage of a pro-B cell, one D_H and one J_H gene segment of the IG heavy (H) chain locus become joint, eventually followed by V_H to DJ_H recombination. Consequently, the somatic VDJ_H recombination yields a uniquely rearranged IGH DNA sequence of the VDJ_H joining region in each B lymphocyte. The joining is an imprecise process, in which germline segment ends are cleaved by the RAG1/2 recombinase activity and joined upon random addition of "non-templated" nucleotides (N nucleotides) into the junctions between segments. The resulting DNA sequence variations add to the overall diversity of antigen receptors and constitute a unique "fingerprint" target for the design of leukemia-clone specific MRD markers. They also offer a valuable tool to decipher which mechanism of somatic recombination was active and to gather clonally related clonotypes (1–4).

Malignant transformation is preceded by a premalignant state, in which genetic aberrations accumulate over time until the driver lesion finally complements the transformation process. Expanded malignant cells harbor the successively acquired genetic aberrations allowing for backtracking the mutational trajectories and (sub) clonal architecture of the malignant population. In the case of malignantly transformed immune cells such as B and T lymphocytes, these carry the unique VDJ rearrangement of the preleukemic cell of origin. Importantly, the VDJ recombination is a stepwise and temporarily tightly regulated process, thus, the completeness of the rearrangements basically depends on the developmental state of the malignantly transformed cell of origin.

The malignant transformation in B-cell precursor acute lymphoblastic leukemia (BCP-ALL) occurs at an early stage in the B-cell development, hitting precursor B cells in the stage of a pro-B or a pre-B cell. Therefore, the process of somatic VDJ recombination might still be active in the malignant B cell, driving IGH clonal evolution either through recombination of incomplete DJ_H rearrangements with multiple V_H segments (occasionally with D_H segments as D_H-D_H tandem fusion rearrangements) or by the process of V_H replacement (5–11).

The aforementioned processes are mediated by the RAG1/2 recombinase activity, which recognizes the canonical recombination signal sequences (RSS) during V_H-DJ_H recombination as well as cryptic RSS during D_H-DJ_H recombination and V_H replacement (10, 12, 13). In the latter process, the cryptic RSS of the original V_H segment and the conventional RSS of the incoming V_H segment are processed to replace the original V_H segment by the incoming with the original DJ_H assembly kept. V_H replacement, which can go on for multiple rounds, may leave a remnant of up to five base pairs of the original V_H gene, which can be further diversified by exonuclease activity or N nucleotide addition. V_H replacement also occurs in healthy humans in bone marrow immature B cells to rescue B cells with nonfunctional or autoreactive receptors (14–16).

IGH clonal evolution in BCP-ALL might pose a serious diagnostic challenge for IGH-based MRD assessment upon further diversification of leukemic clonotypes during long-term follow-up. MRD monitoring utilizing clone-specific sequences of the uniquely rearranged and edited gene segments is widely used to assess therapy response after induction and consolidation therapy, which has crucial prognostic importance in BCP-ALL patients (17–23). High-throughput sequencing (HTS)-based marker identification and MRD monitoring offers the benefit of providing additional information on background repertoire including minor accompanying clones, which may reflect clonal evolution [reviewed in (24)].

In this study of BCP-ALL diagnostic and follow-up bone marrow (BM) and peripheral blood (PB) materials utilizing HTS techniques to backtrack rearranged leukemic IGH sequences and

their trajectories, we sought to understand the mechanistic details of leukemic clonal composition at diagnosis, clonal evolution during follow-up, and clonal distribution in different body or hematopoietic compartments. We link diversified IGH clonal family members and discriminate between the mechanisms driving clonal evolution. We correlate these results with immunophenotypes obtained by multiparameter flow cytometry and with transcriptome-based molecular subtypes. We report on the frequency of the phenomenon of clonal evolution, provide examples of clonal sustainability under immunotherapy and finally discuss the implications of clonal evolution in BCP-ALL with regard to initial diagnostics and MRD monitoring.

2 Methods

2.1 Patients and clinical samples

A total of 465 diagnostic and selected MRD-positive follow-up samples from 280 patients aged 18-55 years with BCP-ALL were investigated within the frame of a research project associated with the German Multicenter Adult Acute Lymphoblastic Leukemia (GMALL) 08/2013 therapy optimization trial (EudraCT-No.: 2013-003466-13). Initial diagnostic samples were subjected to IGH amplicon-based HTS, multiparameter flow cytometry, and transcriptome sequencing (RNA-Seq) analyses. An overview of patient characteristics is detailed in Supplementary Table 1. From all 280 patients either diagnostic BM (n=217) or PB (n=63) samples were available (Figure 1, cohort #1). Concurrent diagnostic PB and BM samples were available for pairwise comparisons in 46/280 cases (Figure 1, cohort #2). In addition, a total of n=139 MRD-positive early or late follow-up and relapse samples of 63 patients were selected. Clonal IGH evolution was monitored during early therapy phases, in particular during the cyclophosphamide/dexamethasone-containing prephase (Figure 1, cohort #3, 66 paired samples from 33 patients) and after Induction I including one dose of rituximab (Figure 1, cohort #4, 82 paired samples from 41 patients). Ultimately, we studied 149 diagnostic and follow-up samples of 34 patients to track the kinetics of clonal evolution over time in longitudinal case reports (Figure 1, cohort #5). Main results and details on IGH sequences within the cohorts #1-5 are summarized in Supplementary Table 4. All patients provided informed consent. All research described herein was approved by the Frankfurt Research Ethics Board (188/15F) and performed in accordance with the Declaration of Helsinki.

2.2 Routine diagnostics

Initial diagnoses were established by standard routine diagnostics. Initial immunophenotype, *BCR::ABL1* and *KMT2A* rearrangement status, initial molecular IG/TR markers, and MRD at follow-up were obtained by GMALL trials central diagnostic reference laboratories (Berlin & Kiel, Germany).

2.3 Transcriptome analysis

RNA was extracted according to standard procedures recommended by the manufacturer (Trizol, Life Technologies, Carlsbad, CA). Library preps, transcriptome sequencing, and molecular subtype calling were performed as described previously (26). Cases with intermediate or divergent gene expression profiles could not be assigned to any existing molecular profile category and were thus described as “other”.

2.4 High-throughput sequencing of the IGH locus

We employed the EuroClonality-NGS assay and the IGH-VJ-FR1 and IGH-DJ primer sets to sequence diagnostic and follow-up samples. In general, 100 ng of DNA, extracted according to standard procedures recommended by the manufacturer

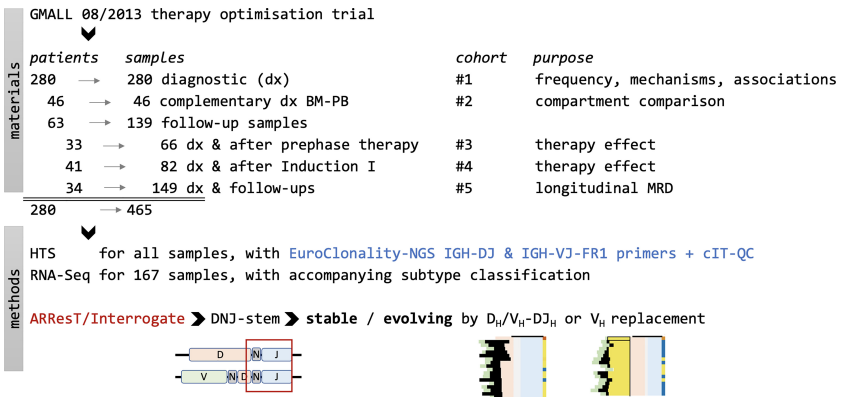


FIGURE 1
Study overview. A total of 465 samples from 280 patients with BCP-ALL enrolled in the GMALL 08/2013 therapy optimization trial were analyzed in five different cohorts to decipher frequency, mechanisms and kinetics of the leukemic clonal evolution. The bioinformatic analyses on IGH HTS data were carried out on DNJ-stem sequences identified by ARResT/Interrogate (25). Transcriptome sequencing and molecular subtype calling were performed as described by Bastian et al. (26) HTS, high-throughput sequencing; RNA-Seq, RNA next- generation sequencing.

(AllPrep, Qiagen, Hilden, Germany), was used for the analysis of the diagnostic samples. In the longitudinal analysis of diagnostic and follow-up samples, we used 500 ng DNA to ensure adequate sensitivity and direct comparability of results. The EuroClonality-NGS central in-tube quality/quantification control (cIT-QC) was spiked into most diagnostic samples and all follow-up samples as a library-specific quality control and for the normalization of abundance from reads to cells (25, 27, 28). (www.euroclonality.org/protocols). Samples were sequenced on a MiSeq (Illumina, San Diego, CA, USA) with 2x250bp reads (sequencing depth analysis in Supplementary Material S1).

2.5 IGH sequence analysis, detection, and characterization of clonal evolution

IGH sequences were analyzed with ARResT/Interrogate (25). Usable reads, the denominator for percentage abundance calculations, were defined as sample reads with identified junctions after the exclusion of cIT-QC reads, in other words, reads with patient-only IGH rearrangements. For better readability, especially in figures, we will refer to IGH segment rearrangements using name acronyms omitting 'IGH' from gene names, e.g., V_H instead of IGHV or DJ_H instead of IGHD-IGHJ.

For IGH clonal evolution assessment, we isolated the 'DNJ-stem' of nucleotide junctions - a subsequence of the junction that consists of the last maximum of 3 D_H nucleotides (or if none are identifiable in a VJ_H rearrangement, of the N region), any N nucleotides between D_H and J_H, and all J_H nucleotides (Figure 2). The DNJ-stem definition is adapted to the understanding that the corresponding junction subsequence is the one remaining generally stable after multiple D_H/V_H to DJ_H (D_H/V_H-DJ_H) recombination and V_H replacement events, and it is therefore used as the link, a shared denominator, between potential family members – whether those clonotypes should indeed be considered related and whether the DNJ-stem should be considered evolving was thoroughly examined by deploying further rules.

We initially removed - as potential noise - clonotypes highly similar to the most abundant DNJ-stem family member, as well as of abundance below 3 reads or below 0.01% of usable reads. We then considered the number of clonotypes removed above (a high number may indicate an overall noisy DNJ-stem), the absolute and relative (to the sample total) number of remaining DNJ-stem family members, the number of their different 5' genes and junction lengths as a sign of expected junctional variability, as well as the DNJ-stem's N region complexity to increase confidence in its specificity. Combined consideration of all these data points led to a consensus call on whether a DNJ-stem was evolving or stable. Please see Supplementary Material S2 for the detailed pseudocode of this algorithm.

V_H replacement was identified by considering (i) the general stability of the D_H region's 5'-site and its preceding N1 region's 3'-site and (ii) the presence of a sequence remnant (fingerprint) of the replaced V_H in the new junction. Evolving DNJ-stems with the absence of aforementioned signs were categorized as D_H/V_H-DJ_H recombinations.

If a DNJ-stem was detected in both the IGH-VJ and IGH-DJ library, the respective DNJ -stem was termed 'rooted', indicating that the underlying root DJ_H rearrangement was also detectable.

A marker DNJ-stem was defined if any of the following conditions were fulfilled: (i) abundance as a sum of all DNJ-stem family members ≥ 5% of usable reads or ≥ 1% of cells (normalized abundance); (ii) presence of clonal evolution ('evolving') at any abundance. We recovered the locus IGHV and IGHD gene order from IMGT (29).

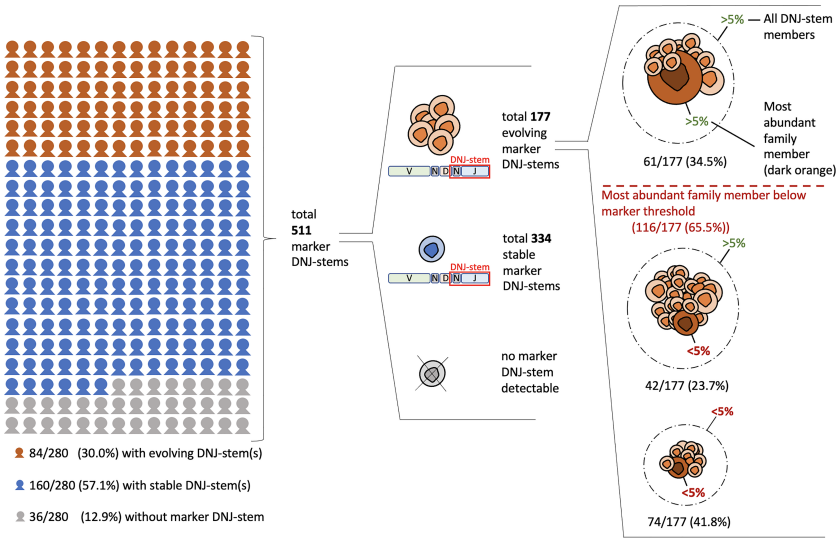


FIGURE 2
Clonal evolution and clonotypic abundances in diagnostic BCP-ALL samples (cohort #1). The figure is displaying the prevalence of clonal evolution in reference to all 280 patients and in reference to the total of 511 detected marker DNJ-stems. Additionally, the evolving DNJ-stems are itemized by their most abundant family member abundance and total DNJ-stem abundance relating to the 5% marker threshold.

Statistical analyses were performed in R version 4.1 using Fisher's exact test with default settings, meaning the significance threshold was set at 0.05.

3 Results

3.1 Clonal evolution is frequent in BCP-ALL

We studied DNJ-stems and the respective VDJ_H and DJ_H rearrangements in the diagnostic cohort of 280 BCP-ALL patients (Figure 1, cohort #1). DNJ-stems with only DJ_H rearrangements were further analyzed only if evolving or otherwise noteworthy.

Clonal evolution, i.e. at least one evolving DNJ-stem, was found in 84/280 (30%) patients. Stable (non-evolving) DNJ-stems were found in 160/280 (57.1%) patients and 36/280 (12.9%) patients had no detectable IGH rearrangements to constitute a marker DNJ-stem – the latter being a rate comparable to previous studies (30, 31). In summary, clonal evolution was prevalent in our cohort affecting one-third of the cases (Figure 2).

In total, 511 marker DNJ-stems (i.e., of ≥ 5% usable reads or evolving at any abundance) were identified in 244/280 (87.1%) patients (range 1-11, average 2.1 per patient). In 116 (65.5%) out of the overall detected 177 evolving DNJ-stems, their most abundant DNJ-stem family member stayed below the generally accepted marker threshold of 5% usable reads and therefore may have been ignored on its own. In other words, applying the standard marker screening approach (by following single clonotypes and not DNJ-stems and using the conventional 5% threshold), 65.5% of markers would not have been reported. By screening for DNJ-stem instead of clonotypes, but still applying the 5% threshold, two-thirds (74) of those 65.5% of markers would still not have been reported. A visual breakdown of the DNJ-stems and family member distribution is depicted in Figure 2.

3.2 Leukemic clonal evolution is driven by ongoing D_H/V_H-DJ_H recombination and V_H replacements

We studied alignments of junction nucleotide sequences of evolving DNJ-stem families from the diagnostic BCP-ALL samples (Figure 1, cohort #1) and observed signs of the two main mechanisms driving IGH clonal evolution, V_H to DJ_H (V_H-DJ_H) recombinations and V_H replacements. Additionally, we identified a variant of somatic recombination, in which D_H to DJ_H (D_H-DJ_H) recombination was evident (10, 11). Both mechanisms, D_H/V_H-DJ_H recombination and V_H replacement, leave distinctive changes in the recombined nucleotide sequences, by which they can be specifically discriminated (Figure 3).

Ongoing D_H/V_H-DJ_H recombination was observed in 78% of evolving DNJ-stems. This mechanism results in sequence trimming during the rearrangement of the D_H/V_H to the existing DJ_H, which in addition to the random editing of N nucleotides results in sequence variability outside the DNJ-stem. This is exemplified in Figure 4A, case #61 and Figure 4B, lower part, case #126. In the illustration, the top row is the 'mother' clonotype for the family of clonotypes to be produced by parallel D_H-DJ_H and V_H-DJ_H recombination (case #61) or solely by V_H-DJ_H recombination (case #126).

The vast majority of D_H/V_H-DJ_H cases featured only V_H-DJ_H recombination-driven clonal evolution, but there were notable exceptions. In particular, of 398 DJ_H marker DNJ-stems identified in 144/280 (51.4%) patients, 21/398 (5.2%) DNJ-stems were evolving in 15 patients - in 12/15 (80.0%) of these patients this ongoing D_H-DJ_H recombination accompanied the V_H-DJ_H recombination, both driving clonal evolution (Figures 4A, B). The mechanism of V_H replacement was not active in those cases. Of note, the only two patients with D_H-DJ_H recombination without accompanying V_H-DJ_H recombination were assigned to the rare molecular subtype CDX2/UBTF, a novel BCP-ALL subgroup with the need of intensified treatment that was identified in 7% of our cohort (26) (see below).

Ongoing V_H replacement was observed in 22% of evolving DNJ-stems. T his mechanism is defined by its most distinct signs: (i) the general stability of the 5′-D_H region site and its preceding 3′-N1 region site; (ii) the actual nucleotide remnants of the replaced V_H in the new junction. There were very few exceptions to the requirement of the incoming V_H to be downstream from the mother clonotype segment: 1.6 upstream segments on average (range 1-5) in lowly abundant clonotypes, in 13/39 DNJ-stems evolving by V_H replacement, representing 1.9% of all family members. These could be erroneous gene annotations facilitated by sequence errors, false positive DNJ-stem members, or perhaps true members having originated previously. A case with ongoing V_H replacement is detailed in Figure 4B (upper part, case #126), where the top row is the 'mother' clonotype for the family members to be produced by replacement of its V_H segment. The resulting family members underneath show no further deletions in the D_H segment, because any sequence loss was buffered by the N1 region and remnants of the replaced V_H (highlighted in yellow) and now residing inside the new N1 regions.

Overall, and in contrast to V_H replacement that most commonly takes place in the presence of one or more highly abundant mother clonotypes, ongoing V_H-DJ_H recombination results in bursts of numerous lowly abundant family members with the role of the mother clonotype taken by the DJ_H root. This translates to an average abundance of the most abundant family member of 51% for V_H replacement and 7% for V_H-DJ_H recombination in the IGH-VJ library of our diagnostic cohort.

3.3 Clonal evolution associates with BCP-ALL immunophenotypes and molecular subtypes

We selected one DNJ-stem from each patient or two DNJ-stems from each of two patients with concomitant V_H-DJ_H recombination and V_H replacement - a total of 53 V_H-DJ_H-evolving stems and 33 V_H replacement-evolving stems - focusing on evolving DNJ-stems with numerous family members, to highlight specific characteristics related to both mechanisms driving clonal evolution. For each DNJ-

10.3389/fimmu.2023.1125017

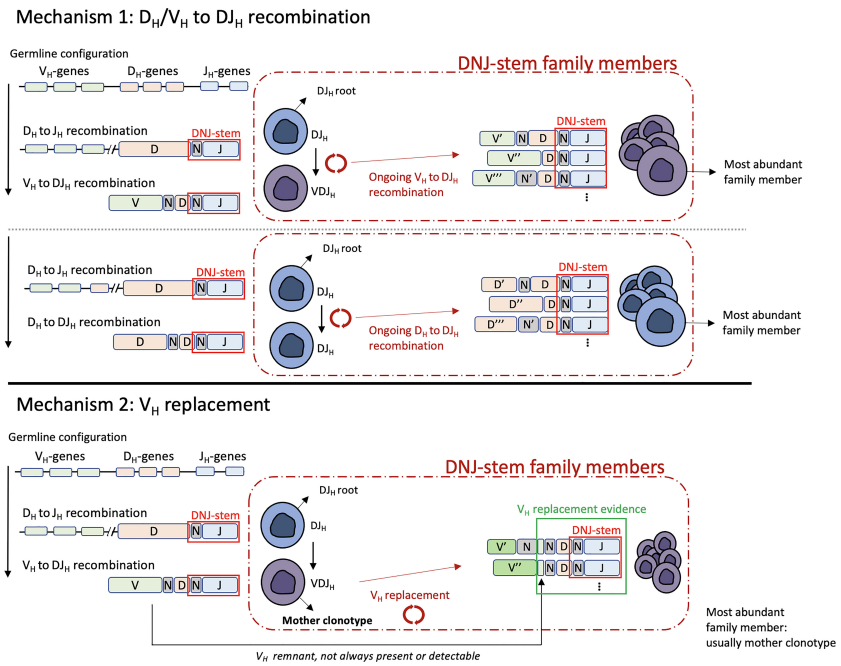


FIGURE 3
Schematic visualization of the terminology used in this study and the two mechanisms driving clonal evolution: D_H/V_H to DJ_H recombination (top) and V_H replacement (bottom). All DNJ-stem family members, including the DJ_H root, the VJ_H mother clonotype (V_H replacement) carry the same DNJ sequence (red box). Evidence of V_H replacement (see main text) is not always present or detectable due to further recombination-related nucleotide sequence trimming (green box).

stem, abundances and IGH evolution classifications such as evolving or stable, DJ_H root detectable or not, D_H/V_H-DJ_H recombination or V_H replacement were annotated along with its locus-ordered V_H gene profile and basic patient metadata. The results are summarized in an overview tabular heatmap (Figure 5).

Of the 244 diagnostic samples of cohort #1 with identified marker DNJ-stems, 41 (16.8%) were classified as pro-B ALL and 203 (83.2%) as pre-B/c-ALL by multiparameter flow cytometry.

We detected significantly higher rates of clonal evolution in patients with a pro-B ALL compared to those with a pre-B/c-ALL immunophenotype (61.0% vs. 29.1%, respectively, p=0.0008), suggesting that the maturation arrest at earlier B-cell development stages might increase the frequency of ongoing IGH recombination.

Other noteworthy observations are that (i) V_H-DJ_H overwhelmingly featured 'early', J-proximal V_H genes, mainly V_H6-1 (42/53 (79.2%) DNJ-stems and most family members) - in strong contrast, V_H6-1 featured only twice in V_H replacement cases, in which the 'late', J-distal V_H3-74 segment was the most frequent; (ii) the DJ_H root was present in the majority of V_H-DJ_H DNJ-stems but in only one V_H replacement DNJ-stem; (iii) D_H-DJ_H-driven clonal evolution was observable in V_H-DJ_H cases but not in V_H replacement cases; (iv) the two mechanisms, V_H-DJ_H recombination and V_H replacement, were almost mutually exclusive (Supplementary Tables 2, 3).

The BCP-ALL molecular subtypes were assigned based on RNA-Seq data available for 167/280 (59.6%) cases and then correlated with the IGH clonal evolution patterns. Overall, 155/167 cases could be assigned to a previously described molecular profile category, while 12 cases showed other gene expression profiles and thus were summarized as "other". Notably, the aforementioned higher rates of clonal evolution in pro-B ALL were almost exclusively linked to the KMT2A subtype (94.4% evolving), while other molecular subtypes with a pro-B ALL immunophenotype were mostly stable, particularly CDX2/UBTF (cases with *UBTF::ATXN7L3* gene fusion, only 22.2% evolving and those only in the IGH-DJ library) and ZNF384 (only 30% evolving) (Figure 6A). Further, the KMT2A subtype showed significantly higher rates of ongoing V_H-DJ_H (16/17 (94.1%) evolving patients, p=3e^{-11}), prevalently initiating the recombination with 'early' J-proximal V_H genes, mostly V_H6-1 (Figure 5). The preferential usage of 'early' J-proximal V_H genes was also seen in the molecular subtypes PAX5 P80R and PAX5alt. Although (and in contrast to the KMT2A subtype) most DNJ-stems in those subtypes were stable (30/40, 75%), the evolving DNJ-stems (10/40, 25%) clearly showed a prevalent (as in 7/10) usage of 'early' J-proximal V_H segment genes including V_H6-1, V_H1-2, V_H1-3 and V_H4-4. Notably, the three DNJ-stems using more J-distal V_H segments belonged to a single patient with concurrent V_H replacement and V_H-DJ_H recombination activity (Supplementary Table 2). While the KMT2A subtype was strongly associated with V_H-DJ_H recombination as the prevalent clonal evolution mechanism, the molecular subtypes Ph-like and DUX4 featured no V_H-DJ_H but were instead associated with V_H replacement (7/37 and 4/13, p=0.00009 and p=0.067, respectively) (Figure 6A).

The D_H-DJ_H evolution was a less frequent observation and only detectable in 11 of 155 (7.1%) patients with molecular subtype assignment, with the low number of patients impeding significant associations with molecular subtypes. Despite the fact that the KMT2A subtype not only showed the highest rate of V_H-DJ_H evolution, but also significantly higher rates of D_H-DJ_H evolution

10.3389/fimmu.2023.1125017

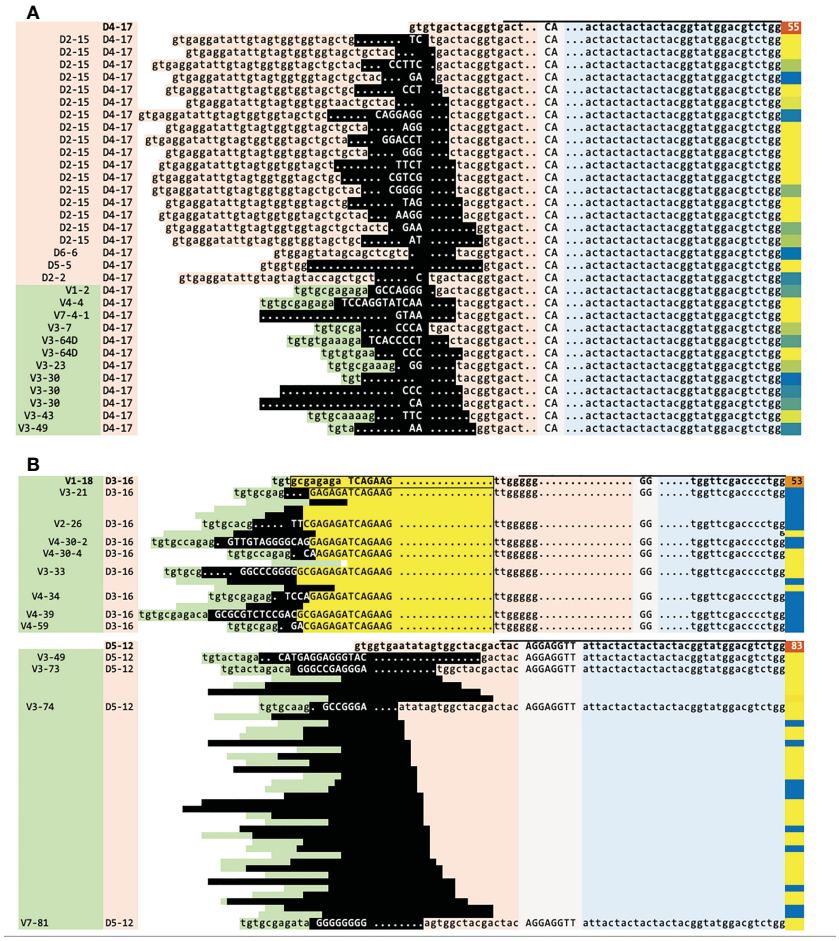


FIGURE 4
Sequence analysis of two cases illustrating the IGH clonal evolution mechanisms. **(A)** Case #61, ongoing D_H-DJ_H and V_H-DJ_H recombination on the same highly abundant DJ_H root (55% of reads) (top row, in bold). **(B)** Case #126, one of two unique cases with DNJ-stems featuring both mechanisms, V_H replacement and V_H-DJ_H recombination (V_H replacement: top, V_H-DJ_H recombination: bottom). The mother clonotype of the V_H replacement and the DJ_H root of the V_H-DJ_H recombination are each shown on the top row and in bold. The DNJ-stem family members evolving by V_H-DJ_H use 'late' D-distal V_H genes, which is uncommon for this mechanism (main text, Figure 5). Legend: green = V_H, orange = D_H, blue = J_H, gray/uppercase = N regions, lowercase = germline, dots = deleted nucleotides, black = region affected by the recombination or replacement, yellow = region kept from the mother clonotype (on which the region is boxed), blue/yellow-to-orange (last column) = clonotype abundance in % reads, horizontal bar on top of first sequence = DNJ-stem. Genes are indented in the 1st column to indicate locus order. A number of rows in **(B)** are 'squeezed' and their V_H and D_H assignment and nucleotide sequences are hidden for space and clarity.

(27.8% evolving, p=0.0001), we found that the presence of D_H-DJ_H and V_H-DJ_H was not fully concordant (Figure 6B). Strikingly, two patients of the rare CDX2/UBTF subtype showed only D_H-DJ evolution (p=0.04), whereas there were no signs of V_H-DJ_H evolution. Overall, the evolving D_H-DJ_H rearrangements showed a remarkable preference for the usage of D_H2-2 as the incoming D_H (14/15 (93.3%) of DNJ-stems) (Supplementary Table 3), which was also observed by others but in the context of normal rearrangements in blood B cells from healthy donors (32).

3.4 Overlap between BM and PB is high for marker DNJ-stems but not for their family members

We sought to understand the overlap of DNJ-stems and their family members between diagnostic BM and PB of the 46 patients with both diagnostic materials available (Figure 1, cohort #2). We identified 39 patients with paired BM and PB DNA samples and with marker DNJ-stem(s) in at least one of these two compartments. The total number of DNJ-stems was 93, with 66 DNJ-stems in both compartments as marker DNJ-stems (see Methods for definition), 21 DNJ-stems also in both compartments but appearing in one compartment outside the marker DNJ-stem rules (and almost always at very low abundance), and the remaining 6 appearing only in BM compartment (Figure 7). This translated into 37/39 patients with at least one of the 87 DNJ-stems in both materials, or 36/39 when we demanded marker DNJ-stems in both materials, i.e., considering only the 66 DNJ-stems; all other cases had the DNJ-stem in BM only. Of the 66 DNJ-stems in both the BM and PB compartments of the 36 patients, all but two had concordant clonal evolution status between BM and PB, with the two exceptional cases evolving in BM and stable in PB. In 32/36 patients we saw a complete marker DNJ-stem overlap between BM and PB.

10.3389/fimmu.2023.1125017

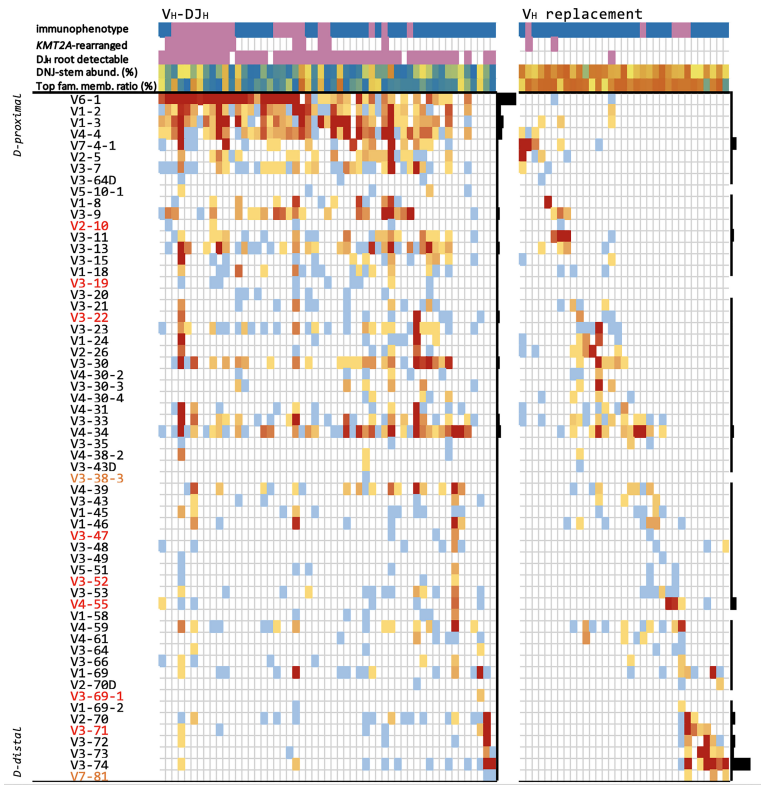


FIGURE 5
Evolving DNJ-stems in the diagnostic cohort. Each column represents one representative evolving DNJ-stem in one patient, ordered by the locus order of the most frequent gene in that column. V_H genes are sorted by their proximity to the D_H genes (proximal top to distal bottom), respectively; red gene names: pseudogenes, yellow: ORFs. Top rows: 1) immunophenotype: pro-B ALL purple, common-/pre-B ALL blue; 2) KMT2A molecular subtype status: KMT2Ar purple, non-KMT2Ar no color; 3) DJ_H root with same DNJ-stem detectable (purple); 4) DNJ-stem abundance (% reads), blue to orange (0.2% to 97%); 5) most abundant family member clonotype to summed DNJ-stem abundance ratio, blue to orange (0.01 to 0.99). Heatmap cells are colored by number of DNJ-stem family members featuring each gene, blue to red (1 to 418); black bar plots on the right of each mechanism group visualize each gene's relative frequency in that group.

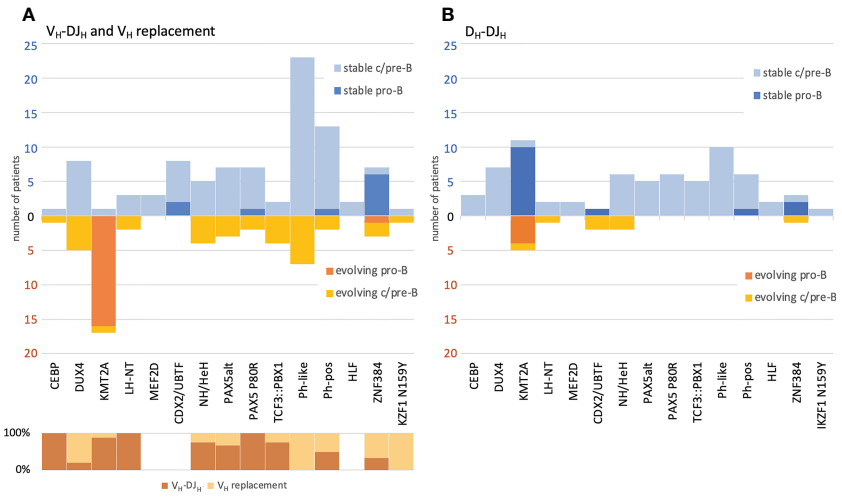


FIGURE 6
Correlation of BCP-ALL immunophenotype and molecular subtypes versus IGH clonal evolution and its respective mechanisms. VH replacement and VH-DJH recombination **(A)** and DH-DJH recombination **(B)**. The numbers and classifications are derived from the one-stem-per-case data depicted also in Figure 5. All cases have been labelled by (i) immunophenotype and (ii) their respective molecular subtype as assigned by the transcriptome analysis. Evolving DNJ-stems (warm colors) are shown below the X-axis line and stable DNJ-stems (cold colors) are shown above the X-axis line. For the evolving DNJ-stems, the respective mechanism driving clonal evolution in % of cases has been depicted in the lower left part of the schematic in **(A)**.

When we compared the overlap of ~3300 family members of 26 DNJ-stems that were evolving both in BM and PB, we found the overlap to be only 8.7%, with the rest split almost equally between BM and PB (Figure 7). Evolving DNJ-stem abundance in BM was on average 18.7% reads (range 0.6-89.9%), while the average abundance of family members was 0.16% reads (range 0.0008-71.8%). In PB the evolving DNJ-stem abundance was on average 12.9% reads (0.2-76.8%) and the average abundance of family members 0.11% (range 0.0002-73.8%). The fact that the minimum abundance of family members is below the 0.01% reads threshold used for family members when calling clonal evolution, is due to searching for members across materials at any abundance to maximize the sensitivity and information content of the overlap analysis.

3.5 Clonal evolution over time and after therapy – implications for MRD diagnostics

In order to understand the longitudinal dynamics of evolving clonotypes during and after treatment, we compared therapy-naive BCP-ALL samples (day 0) and matched samples from different follow-up time points (Figure 1, cohort #3-5): (i) day +6 samples after the 5-day cyclophosphamide/dexamethasone-containing prephase, no rituximab (cohort #3); (ii) day +22 samples after Induction I including one dose of rituximab (cohort #4); and (iii) later time points including additional doses of rituximab and refractory disease/relapse-related secondary immunotherapies, e.g. blinatumomab (cohort #5).

3.5.1 The clonotypic spectrum does not change much during the very early phase of therapy

Cohort #3 encompassed 33 BCP-ALL patients with a total of 66 matched peripheral blood (PB) samples taken at diagnosis and at day +6 of the prephase. In 3/33 (9.1%) diagnostic PB samples, no marker DNJ-stem was identified. In one case, the stable marker DNJ-stem was undetectable after prephase therapy. More than two-thirds of the cases were either concordantly stable (19/33, 57.6%) or concordantly evolving (4/33, 12.1%) across day 0 and day +6. Four cases (12.1%) were evolving at day 0 and stable at day +6, with the loss of family members after prephase therapy probably due to the general reduction of tumor burden. Two cases showed clonal evolution after prephase therapy but not at diagnosis, which could be rather attributed to the overall reduction of benign mature B-cell numbers in the post-prephase samples and the resulting higher sensitivity of HTS. In the majority of concordantly evolving DNJ-stems (4/6), the highest abundant family member at day 0 remained the highest abundant family member at day +6. The overlap of evolving clonotypes between the two time points was on average 29.6%.

We further quantified the MRD at day +6 using either the DNJ-stem, or only the most abundant family member as MRD marker and compared the MRD levels. The DNJ-stem-based MRD was on average 3.5-times higher (range 1.1 to 9.7-times) than the most abundant family member-based MRD. In 2/8 evolving DNJ-stems, the discrepancy between DNJ-stem MRD and most abundant family member MRD equaled a log10 difference.

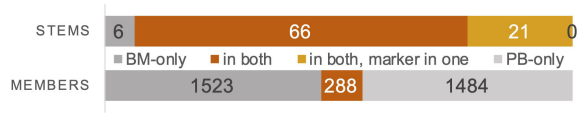


FIGURE 7
Overlap of marker DNJ-stems and their family members detectable in BM and PB. Total numbers of marker DNJ-stems (top) and DNJ-stem family members (bottom) that were found in only one of the compartments are illustrated in grey. Those that were overlapping between the compartments are illustrated in orange. In case of marker DNJ-stems, we distinguished between concordant marker status (dark orange) and detectable in both, but only marker status in one (light orange).

3.5.2 Single dose rituximab and early Induction therapy unlikely to cause a clonotypic selection

Cohort #4 encompassed 41 BCP-ALL patients with a total of 82 matched bone marrow (BM) samples taken at diagnosis and at day +22 after Induction I with 1 dose of rituximab administered after the prephase. In 6/41 (14.6%) of initial diagnostic samples, no DNJ-stem was identified. The evolving marker DNJ-stem in one case and a stable marker DNJ-stem in another two cases were undetectable after Induction I. In accordance with cohort #1, nearly two-thirds of patients were either concordantly stable (20/41, 48.8%) or concordantly evolving (7/41, 17.1%), with 5 patients evolving at day 0 and stable after Induction I. However, no cases were stable at diagnosis but evolving after Induction I. The average overlap of evolving clonotypes was 28.5%.

3.5.3 Clonotypes display more dynamic changes in longitudinal analysis

Cohort #5 encompassed 34 BCP-ALL patients with a total of 149 matched PB and/or BM samples taken at diagnosis and at multiple follow-up time points per patient, i.e. at the end of prephase at day +6, after Induction I at day +22, at later time points with refractory disease or/and at relapse during chemotherapy including 4 or 8 doses of rituximab or at relapse after additional immunotherapy with blinatumomab or inotuzumab ozogamicin. In particular, the clonotypic composition of initial diagnostic samples was compared to 97 matched samples obtained during these aforementioned divergent MRD-positive follow-up time points. In order to evaluate in how many cases the conventional MRD detection (which relies on the tracking of the one or two most abundant family members) would have been hampered by the vanishing of the most abundant family member, we analyzed clonal evolution dynamics over time in the 34 patients of cohort #5. At diagnosis 14 of 34 patients (41.2%) had stable DNJ-stems (no clonal evolution), 4 (11.8%) had no IGH marker and 16 (46.4%) were evolving. In those 16 patients with clonal evolution, we identified 45 evolving DNJ-stems in total (Supplementary Table 4). Noteworthy, 17/45 (37.8%) of evolving DNJ-stems including all family members became untraceable in the

follow-up samples. In detail, 15/17 evolving DNJ-stems were distributed among two patients, in whom the MRD remained only traceable through the *KMT2A* fusion-derived MRD marker assay, underscoring the importance of this alternative approach of MRD monitoring in *KMT2A*-rearranged B-ALL. The remaining 2/17 disappearing evolving DNJ-stems were both related to one patient and presented the most abundant clonotypes at diagnosis but became eradicated during the follow-up. In 7/45 (15.6%) of evolving DNJ-stems, the most abundant family member stayed most abundant across all follow-up time points. In 2/45 (4.4%) evolving DNJ-stems the most abundant family member was traceable at all the time points, but did not stay the most abundant family member at all time points, thus not reflecting the most correct MRD level. In the majority of evolving DNJ-stems (19/45, 42.2%), the most abundant family member disappeared over time, while other family members of the same DNJ-stem remained present and would be available to constitute the MRD signal in the DNJ-stem based MRD approach. The median abundance of the most abundant family member in those 19 evolving DNJ-stems was 0.06% (0.002 -37.7%) cells compared to 20% (0.67 - 87.6%) cells in the 7 evolving DNJ-stems, in which the most abundant family member remained most abundant at later time points. The non-persistence of the most abundant family member was most likely caused by its low abundance in follow-up in combination with a sampling bias. In one case (Supplementary Material S3A, Ph-neg#27), a highly abundant family member (37.7% cells) disappeared in the course of therapy, while only the second most abundant family member persisted. Taken together, substantial changes in the abundance and composition of the initially most abundant family members conventionally chosen as molecular MRD markers were observed in 37/45 (82%) of the evolving DNJ-stems and accordingly in 9/16 (56%) patients with evolving DNJ-stems, in whom the correct assignment of the MRD level could be adversely affected, i.e. underestimated or failing detection. Critically, in 2/34 (5.9%) of patients, the DNJ-stem with the vanishing most abundant family member was the only IGH target for MRD detection.

In the Supplementary Material S3A-D we present four examples of Philadelphia chromosome-negative (Ph-neg) cases with interesting dynamics of family members belonging to the same DNJ-stem, all presenting different kinetics over time, in which the conventional marker identification (5% reads threshold) or the conventional MRD quantification (clonotype based) would have led to impaired MRD detection in follow-up samples.

Of note, in all cases, the mechanism of clonal evolution in all follow-up samples remained the same compared to the respective diagnostic sample. In consideration of the cohorts #3-5, we conclude that applying the following rules could improve MRD monitoring in follow-up samples: (i) consider the DNJ-stem and not specific family members as MRD basis marker, also because new family members of initially stable DNJ-stems could emerge during the course of therapy; (ii) if an evolving DNJ-stem in the IGH-VJ library is also found in the IGH-DJ library sample (and thus rooted), one should also follow the DNJ-stem in the IGH-DJ library for MRD monitoring even if it does not fulfill marker criteria in the diagnostic sample; (iii) use DNJ-stem abundance as the sum of all family members to avoid underestimation of the extent of MRD.

4 Discussion

Clonal evolution in the IGH locus and its significance and implications in BCP-ALL patients have all been well documented (5, 33–35). Our research was conducted on a well-studied cohort in order to develop tools for the identification and characterization of this important phenomenon. We provide refined observations with a more detailed insight into clonal evolution processes during BCP-ALL therapy.

We report that IGH clonal evolution is prevalent in BCP-ALL, conservatively identified in one-third of the cases. Clonal evolution has been shown to be a malignancy-related phenomenon of B cells (8). The approach presented in this manuscript to identify IGH clonal evolution and the resulting IGH oligoclonality is not based on the simple detection of more than two dominant clonal IGH rearrangements, which could be present even without oligoclonality in the case of biallelic rearrangements or trisomy of the IGH locus (e.g., in hyperdiploid ALL) (36, 37). Instead, we introduce the concept of 'marker DNJ-stem' to cover the entire, clonally-related family of evolving clonotypes. We identified 40% of evolving marker DNJ-stems at levels below conventional clonality thresholds (<5% reads) in 38/280 (14%) patients presenting clonal evolution. In such cases, conventional abundance thresholds produce misleading results and should therefore not be applied. In other words, evolving DNJ-stems at any level should be assumed to be BCP-ALL-related and therefore considered as a marker.

We further investigated whether the degree of clonal evolution and the composition of DNJ-stem family members differs between BM and PB. Comparing the diagnostic sample pairs of 46 patients, we found that the high overlap of marker DNJ-stems and their evolution status indicates that PB is a comparable and reliable source for the determination of the IGH oligoclonality status. The V_H usage of DNJ-stem family members did not show a tendency towards more mature clonotypes in PB compared to BM. In line with our previous study, which showed that MRD levels tend to be lower in PB than in BM (38), 22% of marker DNJ-stems were exclusively found in BM, where DNJ-stems and clonotypes are generally more abundant and evolving, indicating that not all malignant family members access the bloodstream in sufficient numbers to be detected. As expected, the combination of the low abundance of evolved family member clonotypes and sampling biases resulted in their low overlap between the two materials - but of note, their total number was not higher in BM compared to PB.

Abandoning the conventional 5% reads clonality threshold for evolving DNJ-stems as suggested above entails the risk of reporting unspecific, malignancy-unrelated markers which can result in impaired MRD detection. Thus, it should be discussed if the DNJ-stem is unique and specific enough to be a reliable MRD target. The insertion of short palindromic P nucleotides during VDJ_H recombination has been reported in around ten percent of IGH rearrangements (39). These nonrandom sequences might reduce the specificity of the DNJ-stem for MRD detection,

especially in cases with short or completely absent N2 nucleotides. Furthermore, it is not clear if the application of DNJ-stem-based MRD detection is of benefit in mature lymphoid malignancies. We, therefore, tested the sensitivity and specificity of this approach using IGH sequences in internal cohorts of two distinct mature B-lymphoid malignancies, i.e. ~100 chronic lymphocytic leukemia (CLL) and ~100 multiple myeloma (MM) cases. While we detected no clonal evolution in the CLL cohort, three DNJ-stems in as many cases in the MM cohort were reported. All three were 'borderline' calls with few family members and low variability of genes and junction lengths, calls that in the context of MM could be dismissed. Additionally, we screened our BCP-ALL cohort for marker DNJ-stems that appear in more than one patient at diagnosis. Only one of the 511 marker DNJ-stems detected in the IGH-VJ library was shared between two patients. Overall, we would advise the application of bio/medical context during the interpretation of results, and to examine DNJ-stem sequences for specificity within each cohort and against a large reference cohort.

The study of the junction alignments of the family members of evolving DNJ-stems allowed us to identify the apparent mechanism involved, D_H/V_H to DJ_H recombination or V_H replacement, and thus provide another perspective into our data and results (Figures 3-5).

D_H-DJ_H segment fusions were first hypothesized in 1982 (40), experimentally verified in 1989 and thereby considered a rare but physiologically legitimate activity adding another combinatorial dimension to the antibody repertoire diversity (11). Since D_H-DJ_H recombination clearly violates the 12-23 rule applying in the process of V(D)J recombination, concerns were raised about a possible aberrant nature of this phenomenon. Further research led to the discovery of hitherto uncharacterized cryptic RSS in the D_H genes, confirming the physiological nature of the occasional D_H-DJ_H recombination (10).

Taken together, our data undoubtedly show that all of the described ongoing recombinatory processes, the D_H/V_H to DJ_H recombination and V_H replacement, drive the clonal evolution in BCP-ALL patients. Moreover, we demonstrate that these mechanisms are almost mutually exclusive and separate patients into clearly delineated groups on different levels:

(i) the D_H/V_H-DJ_H group was characterized by an overrepresentation of *KMT2A* and *PAX5* aberrations, the presence of a DJ_H root, the preferential usage of D_H-proximal V_H genes and the presence of low-abundant DNJ-stem family members, as well as accompanying D_H-DJ_H evolution in *KMT2A* rearranged cases;

(ii) the V_H replacement group was (less strongly) associated with the DUX4 and Ph-like molecular subtypes, as well as with D-distal V_H genes used by DNJ-stem family members, one or more of which playing the role of abundant mother clonotypes.

The two exceptions to the observation that D_H-DJ_H evolution cases associated with V_H-DJ_H evolution, were two cases belonging to the newly described CDX2/UBTF molecular subtype, which otherwise presented no IGH evolution despite the predominance of a pro-B immunophenotype (26).

Our results point towards D_H/V_H-DJ_H recombination as an ontogenically earlier-stage and V_H replacement as an ontogenically later-stage associated mechanism. In most V_H-DJ_H cases, the DNJ-stem was detectable in the IGH-DJ library, suggesting that V_H-DJ_H is a highly active and ongoing process, resulting in a burst of immature rearrangements. Generally, the biased usage of D_H/V_H genes, particularly V_H6-1 in V_H-DJ_H recombination and D_H2-2 in D_H-DJ_H recombination, could be related to the distinctive accessibility of different parts of the V_H and D_H clusters to the VDJ_H recombination machinery as postulated by Tsakou and colleagues (41) and previously shown in mice (42).

The preferential use of V_H6-1 could at least be related to hampered PAX5 activity. Wild-type PAX5 was reported to cooperate with interleukin receptor 7 to confer locus contraction and subsequent V_H-DJ_H recombination of distal V_H segments, whereas it was not required for the preceding D_H to J_H recombination (43, 44). Consequently, at least the biased usage of V_H6-1 could be related to the inaccessibility of distal V_H segments due to absent or hampered activity of PAX5 and/or related factors.

Our results, which clearly demonstrate a correlation of the molecular subtype with different frequencies and even different mechanisms of clonal evolution, entail clinical consequences. Our selected longitudinal case reports show that MRD quantification, in the form in which it is currently routinely performed, can be affected by the degree of clonal evolution and at worst might lead to false negative MRD results. Thus, the reliability of IGH-based MRD quantification varies depending on the molecular subtype. In particular, patients with the KMT2A subtype, of whom more than 94% are evolving, should be monitored with caution.

Besides the high rates of clonal evolution in *KMT2A*-rearranged BCP-ALL, increasing the susceptibility to errors, studies have shown that IG/TR rearrangements in infant *KMT2A*-rearranged ALL are more likely oligoclonal or even completely absent, which makes any IG/TR based MRD-monitoring less convenient. It has therefore been recommended to preferably employ the *KMT2A* gene fusion for more stable and reliable MRD detection (45, 46).

On the basis of our findings, we suggest the following in order to improve IG/TR-based MRD detection in BCP-ALL: i) follow the DNJ-stem (covering all DNJ-stem family members) and not specific family members; ii) follow the DNJ-stem using both IGH-VJ and IGH-DJ assays; iii) express MRD as the DNJ-stem abundance (the sum of all DNJ-stem family members) in order to avoid underestimation of the MRD level in samples with high degree of clonal evolution; iv) be aware that the classical and still widely adopted real-time quantitative PCR (RQ-PCR) often uses primers and probes designed to track only specific clonal rearrangements (unlike HTS), thus potentially missing family members that are either already present at diagnosis or arise during the course of the disease. Crucially, further testing of the above-suggested approach on large well-defined cohorts is necessary to prove its prognostic strength before it can be routinely applied.

Wrapping up, we will outline a number of future perspectives. First, we herein only studied clonal evolution in the IGH locus, but a study on clonal evolution of T-cell receptors might allow deeper insights - particularly as the TRD locus has shown similar recombinase activity as the IGH locus in patients with V_H replacement (47). Second, a recent study has shown that BCP-

ALL harboring *TP53* alterations are associated with immature DJ_H rearrangements (48). Assuming that *TP53* alterations occur already in the pre-leukemic cell compartment, one should study if these cases show clonal evolution rates similar to *KMT2A*-rearranged cases. Finally, it was recently reported that in childhood BCP-ALL, clonal evolution is not only relevant for reliable MRD detection but also directly correlated to the clinical outcome (33). A similarinvestigation into this critical association of outcome and clonal evolution in adult BCP-ALL would require expanded prospective studies.

Data availability statement

The data presented in the study are deposited in the European Nucleotide Archive (ENA) at EMBL-EBI, accession number PRJEB59052 (https://www.ebi.ac.uk/ena/browser/view/PRJEB59052).

Ethics statement

The studies involving human participants were reviewed and approved by Frankfurt Research Ethics Board (188/15F). The patients/participants provided their written informed consent to participate in this study.

Author contributions

MB, ND, MK and MS designed the research. NG supervised the clinical trial. SS and MS performed immunophenotypic analyses. MK, ND and FD processed, analyzed and interpreted high-throughput sequencing data. ND and KP worked on ARResT/Interrogate. LB, AH and TB analyzed and interpreted RNA-Seq data. FD, ND and MS performed statistical analyses. MB, ND, MS and CB supervised the project. FD, ND, MS, MK and MB drafted the first version of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Deutsche José Carreras Leukämie-Stiftung (grants DJCLS R 15/11 and DJCLS 06R/2019) to MS and MB as well as by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 444949889 (KFO 5010/1 Clinical Research Unit 'CATCH ALL') to LB, AH, MB, and CDB.

Acknowledgments

The authors thank the Hematology Laboratory Kiel (HLK) staff for sample, IGH HTS and RNA-Seq processing. The authors are indebted to the GMALL Trial Center (R. Reutzel, C. Fuchs) and participating hospitals for patients´ recruitment, care and logistics.

Conflict of interest

MB received grants and personal fees from Amgen advisory board, speakers bureau, travel support, and personal fees from Becton Dickinson, Janssen, Pfizer speakers bureau and Jazz travel support, all outside the submitted work. MS and MK received a personal fee from Amgen speakers bureau, outside the submitted work. SS has received personal fees advisory board from AMGEN, Gilead Sciences, Pfizer, SERB SAS, honoraria speaker bureau from Akademie für Infektionsmedizin e.V., AMGEN, AVIR Pharma, CSi Hamburg GmbH, Gilead, Labor28, Novartis, Persberg Group GmbH/DGIM e.V., Pfizer, Vivantes GmbH, research grants from Protherics Medicines Development Ltd, and travel support from Gilead Sciences, Novartis, outside the submitted work.

The remaining 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1125017/full#supplementary-material

References

1. Alt FW, Yancopoulos GD, Blackwell TK, Wood C, Thomas E, Boss M, et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J* (1984) 3:1209–19. doi: 10.1002/j.1460-2075.1984.tb01955.x

2. Alt FW, Oltz EM, Young F, Gorman J, Taccioli G, Chen J. VDJ recombination. *Immunol Today* (1992) 13:306–14. doi: 10.1016/0167-5699(92)90043-7

3. Alt FW, Baltimore D. Joining of immunoglobulin heavy chain gene segments: Implications from a chromosome with evidence of three d-J(H) fusions. *Proc Natl Acad Sci USA* (1982) 79:4118–22. doi: 10.1073/pnas.79.13.4118

4. Tonegawa S. Somatic generation of antibody diversity. *Nature* (1983) 302:575–81. doi: 10.1038/302575a0

5. Beishuizen A, Verhoeven M, van Wering E, Hahlen K, Hooijkaas H, van Dongen J. Analysis of ig and T-cell receptor genes in 40 childhood acute lymphoblastic leukemias at diagnosis and subsequent relapse: Implications for the detection of minimal residual disease by polymerase chain reaction analysis. *Blood* (1994) 83:2238–47. doi: 10.1182/blood.V83.8.2238.2238

6. Beishuizen A, Hählen K, Hagemeijer A, Verhoeven MA, Hooijkaas H, Adriaansen HJ, et al. Multiple rearranged immunoglobulin genes in childhood acute lymphoblastic leukemia of precursor b-cell. *Leukemia* (1991) 5(8):657–67.

7. Choi Y, Greenberg SJ, Du TL, Ward PM, Overturf PM, Brecher ML, et al. Clonal evolution in b-lineage acute lymphoblastic leukemia by contemporaneous VH-VH gene replacements and VH-DJH gene rearrangements. *Blood* (1996) 87:2506–12. doi: 10.1182/blood.V87.6.2506.bloodjournal8762506

8. Gawad C, Pepin F, Carlton VEH, Klinger M, Logan AC, Miklos DB, et al. Massive evolution of the immunoglobulin heavy chain locus in children with b precursor acute lymphoblastic leukemia. *Blood* (2012) 120:4407–17. doi: 10.1182/blood-2012-05-429811

9. Theunissen PMJ, van Zessen D, Stubbs AP, Faham M, Zwaan CM, van Dongen JJM, et al. Antigen receptor sequencing of paired bone marrow samples shows homogeneous distribution of acute lymphoblastic leukemia subclones. *Haematologica* (2017) 102:1869–77. doi: 10.3324/haematol.2017.171454

10. Safonova Y, Pevzner PA. V(DD)J recombination is an important and evolutionarily conserved mechanism for generating antibodies with unusually long CDR3s. *Genome Res* (2020) 30:1547–58. doi: 10.1101/gr.259598.119

11. Meek KD, Hasemann CA, Capra JD. Novel rearrangements at the immunoglobulin d locus. inversions and fusions add to IgH somatic diversity. *J Exp Med* (1989) 170:39–57. doi: 10.1084/jem.170.1.39

12. Reth M, Gehrmann P, Petrac E, Wiese P. A novel VH to VHDJH joining mechanism in heavy-chain-negative (null) pre-b cells results in heavy-chain production. *Nature* (1986) 322:840–2. doi: 10.1038/322840a0

13. Kleinfield R, Hardy RR, Tarlinton D, Dangl J, Herzenberg LA, Weigert M. Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a ly 1+ b-cell lymphoma. *Nature* (1986) 322:843–6. doi: 10.1038/322843a0

14. Steenbergen EJ, Verhagen OJHM, van Leeuwen EF, von dem Borne AEGK, van der Schoot CE. Distinct ongoing ig heavy chain rearrangement processes in childhood b- precursor acute lymphoblastic leukemia. *Blood* (1993) 82:581–9. doi: 10.1182/blood.V82.2.581.581

15. Zhang Z. VH replacement in mice and humans. *Trends Immunol* (2007) 28:132–7. doi: 10.1016/j.it.2007.01.003

16. Zhang Z, Zemlin M, Wang YH, Munfus D, Huye LE, Findley HW, et al. Contribution of VH gene replacement to the primary b cell repertoire. *Immunity* (2003) 19:21–31. doi: 10.1016/S1074-7613(03)00170-5

17. Bassan R, Brüggemann M, Radcliffe HS, Hartfield E, Kreuzbauer G, Wetten S. A systematic literature review and meta-analysis of minimal residual disease as a prognostic indicator in adult b-cell acute lymphoblastic leukemia. *Haematologica* (2019) 104:2028–39. doi: 10.3324/haematol.2018.201053

18. Bassan R, Spinelli O, Oldani E, Intermesoli T, Tosi M, Peruta B, et al. Improved risk classification for risk-specific therapy based on the molecular study of minimal residual disease (MRD) in adult acute lymphoblastic leukemia (ALL). *Blood* (2009) 113:4153–62. doi: 10.1182/blood-2008-11-185132

19. Beldjord K, Chevret S, Asnafi V, Huguet F, Boulland ML, Leguay T, et al. Oncogenetics and minimal residual disease are independent outcome predictors in adult patients with acute lymphoblastic leukemia. *Blood* (2014) 123:3739–49. doi: 10.1182/blood-2014-01-547695

20. Berry DA, Zhou S, Higley H, Mukundan L, Fu S, Reaman GH, et al. Association of minimal residual disease with clinical outcome in pediatric and adult acute lymphoblastic leukemia: A meta-analysis. *JAMA Oncol* (2017) 3(7):e170580. doi: 10.1001/jamaoncol.2017.0580

21. Gökbuget N, Kneba M, Raff T, Trautmann H, Bartram CR, Arnold R, et al. Adult patients with acute lymphoblastic leukemia and molecular failure display a poor prognosis and are candidates for stem cell transplantation and targeted therapies. *Blood* (2012) 120:1868–76. doi: 10.1182/blood-2011-09-377713

22. Reiter A, Schrappe M, Ludwig WD, Hiddemann W, Sauter S, Henze G, et al. Chemotherapy in 998 unselected childhood acute lymphoblastic leukemia patients. results and conclusions of the multicenter trial ALL-BFM 86. *Blood* (1994) 84:3122–33. doi: 10.1182/blood.V84.9.3122.3122

23. Ribera JM, Oriol A, Morgades M, Montesinos P, Sarrà J, González-Campos J, et al. Treatment of high-risk Philadelphia chromosome-negative acute lymphoblastic leukemia in adolescents and adults according to early cytologic response and minimal residual disease after consolidation assessed by flow cytometry: Final results of the PETHEMA ALL-AR-03 trial. *J Clin Oncol* (2014) 32:1595–604. doi: 10.1200/JCO.2013.52.2425

24. Kotrova M, Darzentas N, Pott C, Baldus CD, Brüggemann M. Immune gene rearrangements: Unique signatures for tracing physiological lymphocytes and leukemic cells. *Genes (Basel)* (2021) 12:979. doi: 10.3390/genes12070979

25. Bystry V, Reigl T, Krejci A, Demko M, Hanakova B, Grioni A, et al. ARResT/Interrogate: An interactive immunoprofiler for IG/TR NGS data. *Bioinformatics* (2017) 33:435–7. doi: 10.1093/bioinformatics/btw634

26. Bastian L, Hartmann AM, Beder T, Hänzelmann S, Kässens J, Bultmann M, et al. UBTF::ATXN7L3 gene fusion defines novel b cell precursor ALL subtype with CDX2 expression and need for intensified treatment. *Leukemia* (2022) 36(6):1676–80. doi: 10.1038/s41375-022-01557-6

27. Brüggemann M, Kotrová M, Knecht H, Bartram J, Boudjogrha M, Bystry V, et al. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* (2019) 33:2241–53. doi: 10.1038/s41375-019-0496-7

28. Knecht H, Reigl T, Kotrová M, Appelt F, Stewart P, Bystry V, et al. Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. *Leukemia* (2019) 33:2254–65. doi: 10.1038/s41375-019-0499-4

29. Lefranc MP. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res* (2001) 29:207–9. doi: 10.1093/nar/29.1.207

30. Kotrova M, Knecht H, Herrmann D, Schwarz M, Olsen K, Trautmann H, et al. The IG/TR next generation marker screening developed within euroclonality-NGS consortium is successful in 94% of acute lymphoblastic leukemia samples. *Blood* (2018) 132:2830. doi: 10.1182/blood-2018-99-112828

31. Szczepański T, Willemse MJ, Brinkhof B, Van Wering ER, van der Burg M, Van Dongen JJM. Comparative analysis of ig and TCR gene rearrangements at diagnosis and at relapse of childhood precursor-B–ALL provides improved strategies for selection of stable PCR targets for monitoring of minimal residual disease. *Blood* (2002) 99:2315–23. doi: 10.1182/blood.V99.7.2315

32. Hansen TØ, Lange AB, Barington T. Sterile DJ h rearrangements reveal that distance between gene segments on the human ig h chain locus influences their ability to rearrange. *J Immunol* (2015) 194:973–82. doi: 10.4049/jimmunol.1401443

33. Levy G, Kicinski M, van der Straeten J, Uyttebroeck A, Ferster A, de Moerloose B, et al. Immunoglobulin heavy chain high-throughput sequencing in pediatric b-precursor acute lymphoblastic leukemia: Is the clonality of the disease at diagnosis related to its prognosis? *Front Pediatr* (2022) 10:874771. doi: 10.3389/fped.2022.874771

34. Rosenquist R, Thunberg U, Li AH, Forestier E, Lönnerholm G, Lindh J, et al. Clonal evolution as judged by immunoglobulin heavy chain gene rearrangements in relapsing precursor-b acute lymphoblastic leukemia. *Eur J Haematol* (1999) 63:171–9. doi: 10.1111/j.1600-0609.1999.tb01765.x

35. Theunissen PMJ, de Bie M, van Zessen D, de Haas V, Stubbs AP, van der Velden VHJ. Next-generation antigen receptor sequencing of paired diagnosis and relapse samples of b-cell acute lymphoblastic leukemia: Clonal evolution and implications for minimal residual disease target selection. *Leuk Res* (2019) 76:98–104. doi: 10.1016/j.leukres.2018.10.009

36. Fries C, Lee LW, Devidas M, Dai Y, Rabin KR, Gupta S, et al. Prognostic impact of pretreatment immunoglobulin clonal composition in pediatric b-lymphoblastic leukemia. *Haematologica* (2023) 108:900–4. doi: 10.3324/haematol.2022.281146

37. Szczepanski T, Willemse M, Wering EV, Weerden JFV, Kamps W, Dongen JV. Precursor-B-ALL with d-H-J(H) gene rearrangements have an immature immunogenotype with a high frequency of oligoclonality and hyperdiploidy of chromosome 14. *Leukemia* (2001) 15:1415–23. doi: 10.1038/sj.leu.2402206

38. Kotrova M, Volland A, Kehden B, Trautmann H, Ritgen M, Wäsch R, et al. Comparison of minimal residual disease levels in bone marrow and peripheral blood in adult acute lymphoblastic leukemia. *Leukemia* (2019) 34:1154–7. doi: 10.1038/s41375-019-0599-1

39. Jackson KJL, Gaeta B, Sewell W, Collins AM. Exonuclease activity and p nucleotide addition in the generation of the expressed immunoglobulin repertoire. *BMC Immunol* (2004) 5:19. doi: 10.1186/1471-2172-5-19

40. Kurosawa Y, Tonegawa S. Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. *J Exp Med* (1982) 155:201–18. doi: 10.1084/jem.155.1.201

41. Tsakou E, Agathagelidis A, Boudjoghra M, Raff T, Dagklis A, Chatzouli M, et al. Partial versus productive immunoglobulin heavylocus rearrangements in chronic lymphocytic leukemia: Implications for b-cell receptor stereotypy. *Mol Med* (2012) 18:138–45. doi: 10.2119/molmed.2011.00216

42. Sen R, Oltz E. Genetic and epigenetic regulation of IgH gene assembly. *Curr Opin Immunol* (2006) 18:237–42. doi: 10.1016/j.coi.2006.03.008

43. Hesslein DGT, Pflugh DL, Chowdhury D, Bothwell ALM, Sen R, Schatz DG. Pax5 is required for recombination of transcribed, acetylated, 5′ IgH V gene segments. *Genes Dev* (2003) 17:37–42. doi: 10.1101/gad.1031403

44. Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* (2004) 18:411–22. doi: 10.1101/gad.291504

45. Darzentas F, Szczepanowski M, Kotrová M, Kelm M, Hartmann A, Beder T, et al. IGH rearrangement evolution in adult KMT2A-rearranged b-cell precursor ALL: implications for cell-of-origin and MRD monitoring. *Hemasphere* (2022) 7(1):e820. doi: 10.1097/HS9.0000000000000820

46. Stutterheim J, van der Sluis IM, de Lorenzo P, Alten J, Ancliffe P, Attarbaschi A, et al. Clinical implications of minimal residual disease detection in infants with KMT2A-rearranged acute lymphoblastic leukemia treated on the interfant-06 protocol. *J Clin Oncol* (2021) 39:652–62. doi: 10.1200/JCO.20.02333

47. Steenbergen EJ, Verhagen OJHM, van den Berg H, van Leeuwen EF, Behrendt H, Slater RR, et al. Rearrangement status of the malignant cell determines type of secondary IgH rearrangement (V-replacement or V to DJ joining) in childhood b precursor acute lymphoblastic leukemia. *Leukemia* (1997) 11:1258–65. doi: 10.1038/sj.leu.2400720

48. Salmoiraghi S, Cavagna R, Montalvo MLG, Ubiali G, Tosi M, Peruta B, et al. Immature immunoglobulin gene rearrangements are recurrent in b precursor adult acute lymphoblastic leukemia carrying TP53 molecular alterations. *Genes (Basel)* (2020) 11:960. doi: 10.3390/genes11090960

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact