LYMPHOID NEOPLASIA

Higher-order immunoglobulin repertoire restrictions in CLL: the illustrative case of stereotyped subsets 2 and 169

Katerina Gemenetzi,[1,2] Fotis Psomopoulos,[1,3] Alejandra A. Carriles,[4] Maria Gounari,[1] Claudia Minici,[4] Karla Plevova,[5,6] Lesley-Ann Sutton,[3] Maria Tsagiopoulou,[1] Panagiotis Baliakas,[7] Kostas Pasentsis,[1] Achilles Anagnostopoulos,[8] Raphael Sandaltzopoulos,[2] Richard Rosenquist,[3,9] Frederic Davi,[10] Sarka Pospisilova,[5,6] Paolo Ghia,[11,12] Kostas Stamatopoulos,[1,3] Massimo Degano,[4,*] and Anastasia Chatzidimitriou[1,3,*]

[1]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece; [2]Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece; [3]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; [4]Biocrystallography Unit, Division of Immunology, Transplantation and Infectious Diseases, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Ospedale San Raffaele, Milan, Italy; [5]Hematology and Oncology, Department of Internal Medicine, University Hospital Brno, Brno, Czech Republic; [6]Faculty of Medicine, Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czech Republic; [7]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; [8]Hematopoietic Cell Transplantation Unit, Department of Hematology, G. Papanikolaou Hospital, Thessaloniki, Greece; [9]Clinical Genetics, Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden; [10]Department of Hematology, Hôpital Pitié-Salpêtrière and Sorbonne University, Paris, France; [11]Medical Oncology, Università Vita-Salute San Raffaele, Milan, Italy; and [12]Strategic Research Program on CLL, B Cell Neoplasia Unit, Division of Experimental Oncology, IRCCS Ospedale San Raffaele, Milan, Italy

KEY POINTS

- CLL stereotyped subsets 2 and 169 display shared B-cell receptor structural features, supporting shared antigen selection in their ontogeny.

- CLL stereotyped subsets 2 and 169 provide a striking example of higher order restrictions of the immunoglobulin gene repertoire in CLL.

Chronic lymphocytic leukemia (CLL) major stereotyped subset 2 (IGHV3-21/IGLV3-21, ~2.5% of all cases of CLL) is an aggressive disease variant, irrespective of the somatic hypermutation (SHM) status of the clonotypic IGHV gene. Minor stereotyped subset 169 (IGHV3-48/IGLV3-21, ~0.2% of all cases of CLL) is related to subset 2, as it displays a highly similar variable antigen-binding site. We further explored this relationship through next-generation sequencing and crystallographic analysis of the clonotypic B-cell receptor immunoglobulin. Branching evolution of the predominant clonotype through intraclonal diversification in the context of ongoing SHM was evident in both heavy and light chain genes of both subsets. Molecular similarities between the 2 subsets were highlighted by the finding of shared SHMs within both the heavy and light chain genes in all analyzed cases at either the clonal or subclonal level. Particularly noteworthy in this respect was a ubiquitous SHM at the linker region between the variable and the constant domain of the IGLV3-21 light chains, previously reported as critical for immunoglobulin homotypic interactions underlying cell-autonomous signaling capacity. Notably, crystallographic analysis revealed that the IGLV3-21–bearing CLL subset 169 immunoglobulin retains the same geometry and contact residues for the homotypic intermolecular interaction observed in subset 2, including the SHM at the linker region, and, from a molecular standpoint, belong to a common structural mode of autologous recognition. Collectively, our findings document that stereotyped subsets 2 and 169 are very closely related, displaying shared immunoglobulin features that can be explained only in the context of shared functional selection. (*Blood*. 2021;137(14):1895-1904)

Introduction

The existence of stereotyped B-cell receptor immunoglobulin (BcR IG) in chronic lymphocytic leukemia (CLL) strongly implicates antigen selection in disease ontogeny.[1-11] Stereotyped cases are classified in multiple subsets of variable size with distinct BcR IG configuration.[10] Certain major subsets have emerged as distinct clinical subgroups, exemplified by subset 2, the largest in CLL, accounting for ~2.5% to 3% of all patients and ~5.5% of patients requiring treatment.[10,12,13]

The particular BcR IG of subset 2 is composed of heavy and light chains encoded by the IGHV3-21 and the IGLV3-21 genes,

respectively. The clonotypic IGHV3-21 genes bear a variable somatic hypermutation (SHM) load, with most cases classified as mutated (M-CLL), but some cases are unmutated (U-CLL).[6,7,10,12,14] Subset 2 BcR IGs display additional distinctive immunogenetic features, including conservation of certain positions in the variable heavy (VH) and variable light (VL) complementarity determining region-3 (CDR3) and recurrent SHMs.[15,16] Moreover, they are capable of self-association leading to cell-autonomous signaling that is critically dependent on the substitution of glycine (G) to arginine (R) introduced by SHM at the λ VL-CL linker region in all subset 2 cases analyzed thus far.[7,17] From a clinical perspective, it is noteworthy that, independent of SHM status, subset 2 cases

have a particularly dismal clinical outcome,[12,13,18] similar to that of patients with *TP53* aberrations, despite very rarely harboring such aberrations.[12,18-24]

We have demonstrated that stereotyped subset 169, a minor CLL subset (~0.2% of all CLL), bears striking immunogenetic similarities to subset 2. More specifically, subset 169 carries: (1) clonotypic heavy chains encoded by the IGHV3-48 gene, which is closely similar to the IGHV3-21 gene; (2) highly similar VH CDR3 motif; (3) clonotypic light chains encoded by the IGLV3-21 gene; (4) both M-CLL and U-CLL cases.[10]

To obtain comprehensive insight into the ontogenetic relationship and evolution of CLL subsets 2 and 169, we performed next-generation sequencing (NGS) of the respective BcR IG complemented by crystallographic analysis of the BcR Fab fragments. Our findings cement the structural relatedness of the BcR IG of these subsets, providing further evidence for the role of shared antigen selection throughout their natural history.

Materials and methods

Patient group

The patient group comprised 44 patients diagnosed with CLL, according to the International Workshop on CLL/National Cancer Institute guidelines.[25] Thirty-one of 44 CLL cases expressed stereotyped BcR IGs assigned to subset 2 (M-CLL, n = 19; U-CLL, n = 12), 7 to subset 169 (M-CLL, n = 4; U-CLL, n = 3), whereas the remaining 6 CLL cases (M-CLL, n = 4; U-CLL, n = 2) were not assigned to stereotyped subsets and were selected for comparisons on the basis of using the IGLV3-21 gene (supplemental Table 1, available on the *Blood* Web site). The study was approved by the local Ethics Review Committees of the participating institutions and was conducted in accordance with the Declaration of Helsinki.

NGS of IGHV-IGHD-IGHJ and IGLV-IGLJ gene rearrangements

Total cellular RNA was isolated from peripheral blood mononuclear cells. From 1 case assigned to subset 2, we isolated genomic DNA for comparative studies.

Amplification of the IGHV-IGHD-IGHJ and IGLV-IGLJ gene rearrangements was performed by Platinum Taq DNA Polymerase High Fidelity (ThermoFisher Scientific, Waltham, MA) using IG subgroup-specific leader primers as well as IGHJ and IGLC primers for the heavy and λ light chains, respectively, as described.[17] PCR products were gel purified (Qiagen, Hilden, Germany), and library preparation was performed according to the manufacturer's instructions (NEB Next Ultra II DNA Library Prep Kit for Illumina; NEB, Ipswich, MA). Paired-end NGS was performed with the MiSeq Reagent Kit v3 (2 × 300 bp) on the MiSeq Benchtop Sequencer (Illumina).

To assess possible biases related to substrate usage (genomic DNA [gDNA] vs complementary DNA [cDNA]) and different laboratory personnel, we included 2 controls: (1) sample replicates (ie, 1 subset 2 case was analyzed twice starting from either genomic DNA or complementary DNA), and (2) user replicates (ie, 2 libraries from the same subset 2 case prepared by different laboratory members).

Detailed information about the high-throughput sequencing approach is provided in the supplemental Methods.

Bioinformatics analysis and definitions used in this study

Base-calling, quality control, adapter trimming, and demultiplexing were performed with the Illumina signal-processing software. Downstream analysis was performed by the TRIP bioinformatics analytical toolbox.[26,27] Only productive, in-frame gene rearrangements were evaluated. SHM characteristics were evaluated as described.[7,16,28,29]

Clonotypes were defined as IGHV-IGHD-IGHJ or IGLV-IGLJ gene rearrangements with identical IGHV/IGLV gene and VH/VL CDR3 amino acid sequences within a sample. Clonotypes with different amino acid substitutions within the sequence of the rearranged IGHV/IGLV gene excluding the VH/VL CDR3 region were defined as subclones. They were considered to be expanded (shaping clusters) when they included at least 2 sequences, otherwise referred to as a "singleton." The most expanded clonotype within a sample was deemed the dominant one. Analysis of clonotype branching introduced by SHM was performed using the TRIP tool to determine the 20 most frequent clonotypes of each sample.[27]

"Convergent recombination" is the term used for describing the phenomenon of different nucleotide sequences degenerately encoding the same VH/VL CDR3 amino acid sequence.[30-33]

Figure 1 provides a schematic of the definitions used in this manuscript for describing immunogenetic variation at different levels.

A detailed account of the bioinformatics approach is given in supplemental Methods and supplemental Table 2.
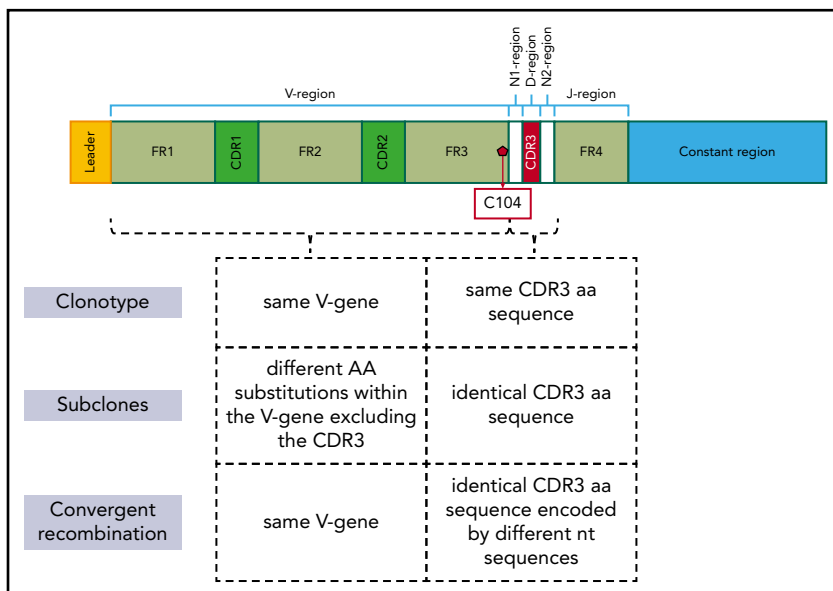
Crystallographic studies

The subset 169 case P6540 BcR Fab fragment was expressed in recombinant form in HEK293 cells, as previously described,[17] then purified and crystallized. Its crystal structure was determined and refined to 3.4 Å resolution (supplemental Table 3).

Full experimental details are presented in the supplemental Methods, also regarding the comparison of the subset 169 case P6540 fragment vs the subset 2 case P11475 BcR Fab fragment, the latter previously reported by our group.[17]

Statistical analysis and data visualization tools

Descriptive statistics for qualitative parameters included counts and frequency distributions. For quantitative variables, statistical measures included mean, median, and minimum/maximum values. Analysis of variance was used to evaluate the mean differences in clonotypes, subclones, and convergent recombination between sample categories. Post hoc analysis was performed to assess the multiple pairwise comparisons, using the Bonferroni correction and the homogeneity of possible subgroups identified by Duncan's test. For the statistics regarding the shared clonotypes between different sample categories, Pearson's χ^2 test with Yates's continuity correction was used. For all comparisons, a significance level of $P = .05$ was set.

GEMENETZI et al

Figure 1. Definitions of immunogenetic variation at different levels.

Data visualization was performed in the R environment and with the open-source data visualization framework RawGraphs (https://rawgraphs.io/). Figures for the crystallographic assays were generated with Pymol (http://www.pymol.org).

Results

Basic information about the NGS run

Overall, 72 PCR amplicons were analyzed, including 39 IGHV-IGHD-IGHJ gene rearrangements (subset 2; n = 32 of which 2 represented duplicates; subset 169, n = 7); and 33 IGLV-IGLJ gene rearrangements (subset 2, n = 21; subset 169, n = 6; nonsubset, n = 6) (supplemental Table 4). Of the aforementioned 39 IGHV-IGHD-IGHJ gene rearrangements, 38 were amplified using cDNA and only 1 by gDNA for the control experiments; detailed information about the results of the control experiments (duplicates) is provided in the supplemental Results.

Overall, 12 925 971 productive gene rearrangement sequences were obtained, corresponding to 4 692 402 IGHV-IGHD-IGHJ (median, 126 681 per sample) and 8 233 569 IGLV-IGLJ gene rearrangement sequences (median, 230 769 per sample) (supplemental Tables 5 and 6).

Clonal composition

We identified 1932 and 3147 unique IGHV-IGHD-IGHJ and IGLV-IGLJ BcR IG clonotypes, respectively. Of these, 1267 (65.6%) heavy chain and 1360 (43.2%) light chain clonotypes were expanded, whereas the remainder were singletons.

In subset 2, each analyzed IGHV-IGHD-IGHJ gene rearrangement carried a median of 81 distinct clonotypes (range, 17-152); the dominant clonotype had a median frequency of 97.7% (range, 66.5%-99.0%). IGLV-IGLJ gene rearrangements carried a higher number of distinct clonotypes per sample compared with their partner heavy chains (median, 216; range, 120-306); the dominant clonotype had a median frequency of 97.2% (range, 73.7%-97.8%) (Figures 2 and 3; supplemental Tables 5 and 6).

In subset 169, each analyzed IGHV-IGHD-IGHJ gene rearrangement carried a median of 54 clonotypes (range, 40-68); the dominant clonotype had a median frequency of 99.1% (range, 97.9%-99.2%). Subset 169 IGLV-IGLJ gene rearrangements carried a higher number of distinct clonotypes per sample compared with their partner heavy chains (median, 144; range, 110-205); the dominant clonotype had a median frequency of 98.3% (range, 97.2%-98.6%; Figures 2 and 3, supplemental Tables 5 and 6).

IGLV3-21–expressing CLL cases that did not belong to either subset 2 or 169 carried a median of 119 distinct clonotypes within their clonotypic IGLV-IGLJ gene rearrangement (range, 100-150); the dominant clonotype had a median frequency of 98.7% (range, 97.2%-98.9%; Figures 2 and 3; supplemental Tables 5 and 6).

Statistical analysis of the clonality patterns revealed that the IGLV-IGLJ gene rearrangements of subset 2 were significantly ($P < .05$) more diverse from all other data sets, whereas those of subset 169 and the non-subset IGLV3-21 cases did not differ from each other. The clonality patterns of IGHV-IGHD-IGHJ gene rearrangements of subsets 2 and 169 did not differ either, but differed significantly from all light chain categories ($P < .05$; supplemental Table 7).

Clonotype sharing in subsets 2 and 169

Heavy chains In subset 2, 415 distinct clonotypes were shared by 2 or more cases (415 of 1618 unique clonotypes; 25.6%). Eight of 415 clonotypes were shared by at least 10 (range 10-19) of the 30 analyzed patients; these were dominant in 1 to 5 patients and of lower frequency in the remaining cases (Figure 4A; supplemental Tables 8A and 9). In subset 169, 55 clonotypes were shared by 2 to 3 of the 7 analyzed cases (55 of 314 unique clonotypes, 17.5%); most shared clonotypes were of low frequency. Two samples carried the same dominant clonotype (98.8%-99.1%), which was also present in 1 more sample, albeit at a low frequency (supplemental Tables 8B and 10). The difference between the 2 subsets regarding the number of shared clonotypes was statistically significant ($P = .0172$); however,
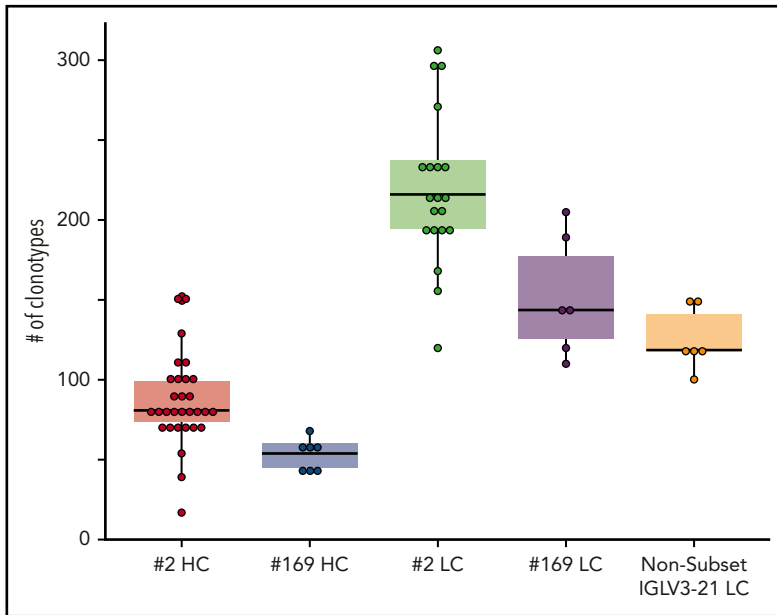
Figure 2. Box plots representing the clonality of all sample categories based on the distinct clonotypes identified per sample. Statistical analysis of the clonality patterns revealed that the IGLV-IGLJ gene rearrangements of subset 2 were significantly ($P < .05$) more diverse than those in all other data sets, whereas those of subset 169 and the nonsubset IGLV3-21 cases did not differ from each other and could form a homogeneous group. The IGHV-IGHD-IGHJ gene rearrangements of subsets 2 and 169 did not differ from each other either, but significantly differed from all LC categories ($P < .05$). For detailed results of the statistical analysis see supplemental Table 7. HC, heavy chain; LC, light chain.

definitive conclusions cannot be drawn, considering that only 7 subset 169 cases were analyzed vs 30 subset 2 cases.

A cross-subset comparison revealed no common VH CDR3 clonotypes between the 2 stereotyped subsets, thus supporting the idea that the finding of shared clonotypes between cases within the same subset reflects a true biological phenomenon rather than contamination during the experimental procedure.

Light chains Overall, 735 clonotypes were shared between all IGLV-IGLJ rearrangements with a 12-aa-long CDR3 (subset 2,

n = 21; subset 169, n = 6; nonsubset, n = 3) vs only 9 that were shared between nonsubset IGLV3-21–expressing cases with an 11-aa-long CDR3 (n = 3; $P < .001$; however, as for the heavy chains, definitive conclusions cannot be drawn, given the great difference in the number of cases analyzed; 30 vs 3). Seventeen of 735 shared clonotypes were highly prevalent, being detected in 20 to 30 samples: 2 of these, QVWDSSSDHPWV and QVWDSGSDHPWV, were not only the most frequently used clonotypes within the entire data set, with relative frequencies of 35.3% and 28.3%, respectively, but also predominated in 12 and 10 samples, respectively (Figure 4B; full lists of the IGLV3-21
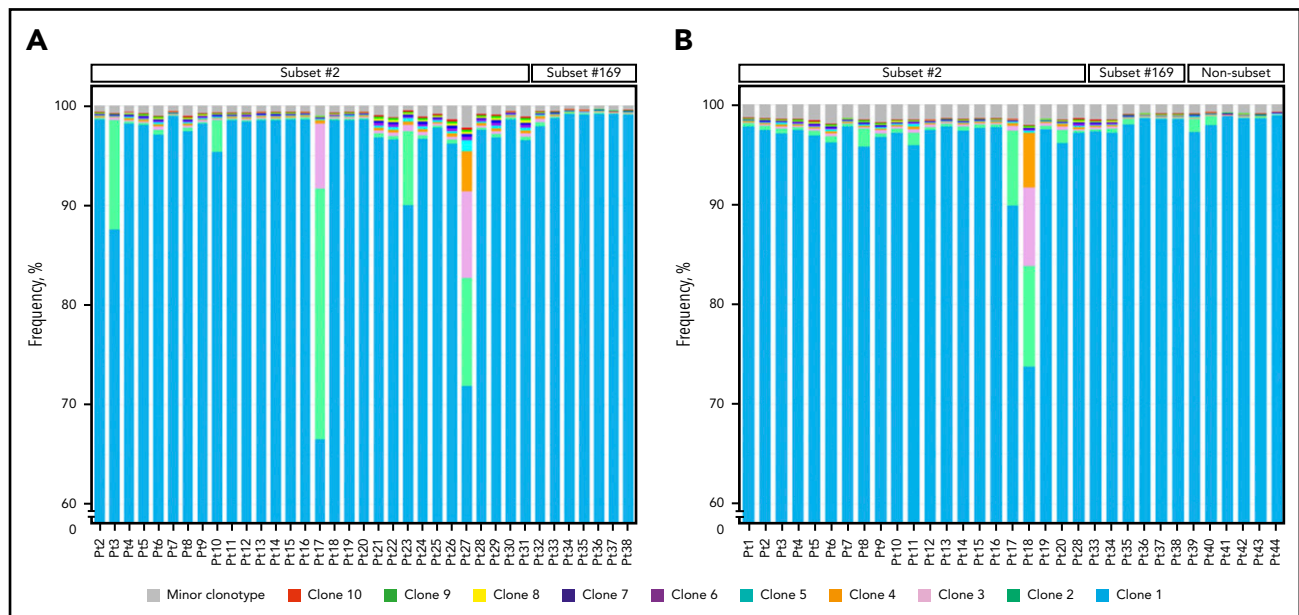

Figure 3. The clonal composition of all heavy chain and IGLV3-21 samples of the cohort is presented. (A) Clustered bar plots depicting the 10 most expanded clonotypes of all heavy chain samples of the present cohort in descending frequency. The relative frequency of each clonotype is computed based on the reads corresponding to the particular clonotype in relation to the total number of reads per sample. The color code represents the ranking of major clonotypes based on their relative frequency. The scale of the graph starts from 60% frequency so that the less expanded clonotypes can be discernible. (B) Clustered bar plots depicting the 10 most expanded clonotypes of all IGLV3-21 samples of the present cohort in descending frequency, as in panel A. Pt, patient.
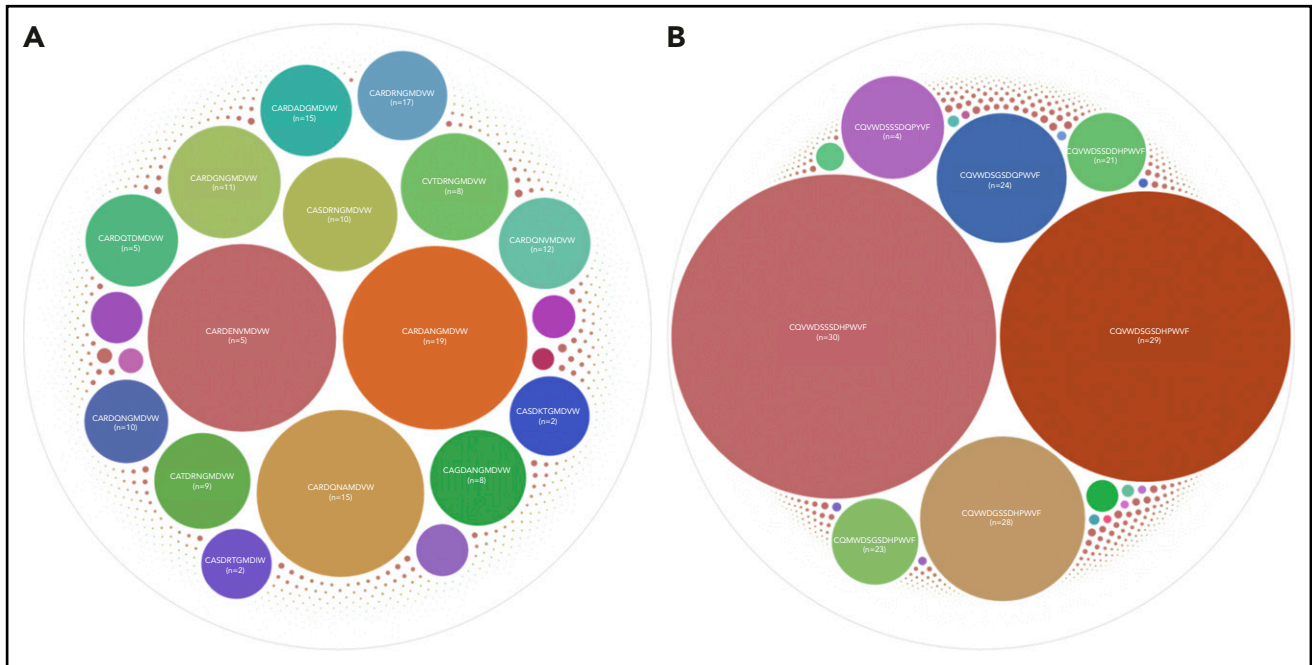
Figure 4. Inter-individual clonotype sharing. (A) Subset 2 heavy chains. All VH CDR3 detected in the subset 2 cohort are depicted by a bubble or dot. The size of the bubbles reflects the relative frequency of each clonotype based on the total number of reads encoding that particular clonotype across all samples. The VH CDR3 of some of the most expanded clonotypes within the entire cohort are stated in the corresponding bubble, and the number of patients sharing that particular clonotype is indicated in the brackets. Most of the expanded clonotypes are shared among different patients (the full list of the entire heavy chain subset 2 clonotypes is given in supplemental Table 8A; for the shared clonotypes, see supplemental Table 9) (B) IGLV3-21 light chains. The entire sequence data set of IGLV3-21 gene rearrangements with 12-aa-long VL CDR3 cohort is depicted. The size of the bubbles reflects the relative frequency of each clonotype based on the total number of reads encoding that particular clonotype across all samples. Some of the expanded clonotypes of the cohort are labeled, and the number of patients sharing that particular clonotype is indicated in the brackets. Most of the expanded clonotypes are shared among different patients (the full list of the entire LC IGLV3-21 with 12-aa-long clonotypes is given in supplemental Table 11A; for the shared clonotypes, see supplemental Table 12).

clonotypes are given in supplemental Table 11; shared clonotypes are given in supplemental Tables 12 and 13).

Higher intraclonal diversification in the light vs the heavy chain gene rearrangements of subsets 2 and 169

In subset 2, the dominant IGHV-IGHD-IGHJ clonotype displayed a median of 3662 subclones per sample (range, 191-11 041), whereas in subset 169, the dominant IGHV-IGHD-IGHJ clonotype had a median of 1803 subclones per sample (range, 1482-6116; $P = .479$; supplemental Tables 5 and 7). In both subsets, intraclonal diversity was also observed for the dominant IGLV3-21 clonotype of each case, with medians of 7 015 subclones per sample (range, 1 946-11 866) for subset 2 and 5 101 subclones per sample (range, 4 237-10 381) for subset 169 ($P = .99$). Nonsubset IGLV3-21 gene rearrangements displayed significantly ($P < .05$) more restricted intraclonal diversification for the dominant clonotype (median: 4163; range, 3404-4620) compared with subset 2 but did not differ in that respect from subset 169 ($P = .99$; supplemental Tables 6 and 7).

Comparison of the heavy chain vs the light chain gene rearrangements within subsets 2 and 169 revealed more pronounced intraclonal diversification in the latter ($P < .05$ for subset 2; $P = .065$ for subset 169; supplemental Table 7).

Also relevant was the analysis of the IGHV-IGHD-IGHJ and IGLV-IGLJ gene junctions at the nucleotide level, which revealed the presence of convergent recombination[30-33] in all sample

categories that was more pronounced in subsets 2 and 169 vs nonsubset/IGLV3-21–expressing cases, although the difference did not reach statistical significance (supplemental Figure 1; supplemental Tables 5 and 7).

Truly unmutated sequences

Subset 2 IGHV-IGHD-IGHJ gene rearrangement sequences were found to carry subclones with truly unmutated IGHV genes (ie, displaying 100% germline identity). In detail, 8 of 30 IGHV-IGHD-IGHJ gene rearrangements from subset 2 cases carried such truly unmutated subclones. Only patient 12 carried mostly truly unmutated reads (95 746 of 124 689; 76.8%), a finding in line with a germline identity of 99.65% by Sanger sequencing (U-CLL; supplemental Table 1). The remaining 7 cases carried too few truly unmutated sequences (median, 2; range, 1-60). No truly unmutated IGHV-IGHD-IGHJ gene rearrangement sequences were detected in subset 169.

Turning to the IGLV-IGLJ gene rearrangements, 11 of 21 subset 2 cases had sequences with 100% germline identity. Patient 18 was particularly noteworthy in this respect, considering that, despite displaying a germline identity of 96.86% by Sanger sequencing (supplemental Table 1), he was found to carry 19 946 of 38 036 (52.4%) truly unmutated sequences; the remaining 10 cases carried few truly unmutated sequences (median, 2.5; range, 1-44). Light chain gene rearrangements from subset 169 carried even fewer (range, 1-3) truly unmutated sequences, whereas only 1 nonsubset case, with a germline identity of

99.28% by Sanger sequencing, carried 36 truly unmutated sequences.

Shared recurrent somatic hypermutations in subsets 2 and 169

Heavy chain gene rearrangements A recurrent 3-nucleotide (AGT) deletion was detected in the VH CDR2 of 28 of 30 examined subset 2 cases, leading to the removal of 1 of 5 consecutive serine residues encoded by the IGHV3-21 germline sequence. This change, previously identified by Sanger sequencing as a recurrent SHM in subset 2 (albeit with an incidence of ∼25%),[7,15] was clonal in 10 of 30 cases (median frequency, 99.4%; range, 91.8%-99.9%) with a median of 4019 subclones per sample (range, 208-8088) and subclonal in 18 of 20 of the remainder (median frequency, 0.1%, range, 0.0004%-3.1%) with a median of 36 subclones per sample (range, 1-429). The deletion was completely absent in only 2 of 30 IGHV-IGHD-IGHJ samples (supplemental Table 14A).

The 5-serine stretch is also present in the germline VH CDR2 of the IGHV3-48 gene used in the BcR IG of subset 169. Notably, an identical 3-nucleotide deletion (AGT) in the VH CDR2 was found as a recurrent change in all examined subset 169 cases as well, albeit at subclonal level (median frequency, 1.6%; range, 0.004%-4.4%), with a median of 157 subclones per sample (range, 6-286; supplemental Table 14A).

Light chains The G-to-R (R110) substitution at the VL-CL linker, previously reported as a ubiquitous SHM in subset 2 (Sanger analysis) and shown to be critical for immunoglobulin self-association leading to cell autonomous signaling in this subset,[17] was a clonal event in all subset 2 cases (median frequency, 96.7%; range, 68.9%-98.1%; Figure 5). The same amino acid substitution was detected in all subset 169 cases as a clonal event with a median frequency of 98.6% (range, 96.8%-98.9%). R110 was also present in 4 of 6 nonsubset IGLV3-21 cases as a clonal event (median frequency, 98.5%; range, 96.7%-98.8%); the remaining 2 samples carried the G-to-R substitution at subclonal level (median frequency, 0.01%; range, 0.009%-0.01%; supplemental Table 14B). The presence of the G-to-R substitution was also validated by flow cytometry using an anti-IGLV3-21/R110 antibody and revealing high expression in the 4 cases with clonal G-to-R substitution and low expression in the remaining 2 cases carrying this change at a low-subclonal level.[34]

Overall similar results were also obtained for the serine (S)-to-G substitution at position 6 within the VL CDR3 which was detected in all examined subset 2 and subset 169 cases at either clonal or subclonal level (supplemental Table 14C).

For both heavy and light chain gene rearrangements, the frequency of mutations at randomly chosen positions (supplemental Tables 14D and 14E) was very different from those reported earlier in this section, further supporting the concept that the findings regarding the recurrent SHMs represent a true biological fact rather than a sequencing artifact.

Highly similar BcR homotypic interactions in subsets 2 and 169

To assess whether the BcR IG from subsets 2 and 169 may share molecular features that parallel their primary structure similarity, we determined the crystal structure of the Fab fragment from the subset 169 P6540 clinical case, which confirmed that subsets 2 and 169 share a similar fold and a conserved modality of intermolecular interaction.

The VH CDR3 loop that defines the CLL subset folded differently in the subset 2 and 169 Fabs, with the latter assuming a more open conformation (Figure 6A-B). These local changes caused a different relative orientation of the VH and VL domains, resulting in a variation of 12° in the domains' pairing angle (supplemental Table 15). These structural differences affected the shape and chemical characteristics of the antigen-binding site of the 2 BcRs, which, despite maintaining a flat, undulating appearance, displayed a distinct distribution of charged and apolar residues (supplemental Figure 2). The VH CDR3 and pairing orientation between the variable domains apparently did not alter the stability of the heavy-light chain pairing, as judged from the similar buried surface areas (supplemental Table 15).

Two molecules of the subset 169 receptors interacted homotypically in the crystals through contacts dominated by the IGLV3-21 light chain. The VL CDR2 loop of the combining site was in contact with a conformational epitope that includes the R110 residue (Figure 6C-D). The intermolecular interactions closely resembled what was observed for subset 2 P11475 BcR,[17] and confirmed that the IGLV3-21 LC–mediated contacts were retained despite the different, albeit homologous, IGHV gene usage in the 2 receptors and distinct VH CDR3 sequences (supplemental Figure 2). Thus, the IGLV3-21-bearing CLL subset 169 BcR retained the same geometry and contact residues for the homotypic intermolecular interaction observed for subset 2[17] and, from a molecular standpoint, belonged to a common structural mode of autologous recognition.

Discussion

CLL stereotyped subset 169 is a minor subset bearing BcR IG with striking similarities to CLL stereotyped subset 2,[10] alluding to shared antigen selection processes in their pathogenesis. We undertook the present study to explore this hypothesis through a comprehensive analysis of the BcR IG, the critical mediator of antigen interactions in the natural history of CLL clones.

Starting from the immunogenetic profiles determined by NGS, the observed patterns of intraclonal diversification within the immunoglobulin genes were similar in both subsets, revealing a dominant subclone for both the heavy and the light chain clonotypic gene rearrangements. The branching of the clone, reflected in the IG gene amino acid sequences, was relatively extensive in both heavy and light chains, albeit consistently more pronounced in the latter, alluding to their relevance in antigen recognition.[16] Further supportive evidence was provided by cross-subset comparisons revealing common clonotypes that were more prevalent for VL CDR3 than for VH CDR3. Admittedly, however, this latter finding may be attributable to the inherently more limited variability of the VL CDR3 compared with the VH CDR3. That cautionary note notwithstanding, our present findings are in line with published evidence regarding the crucial role of the light chains in shaping the structure and, by extension, the function of the BcR IG in CLL.[16,35]

In subset 2, recurrent SHMs within the IGLV3-21–encoded light chains are linked to the structural properties of the clonotypic
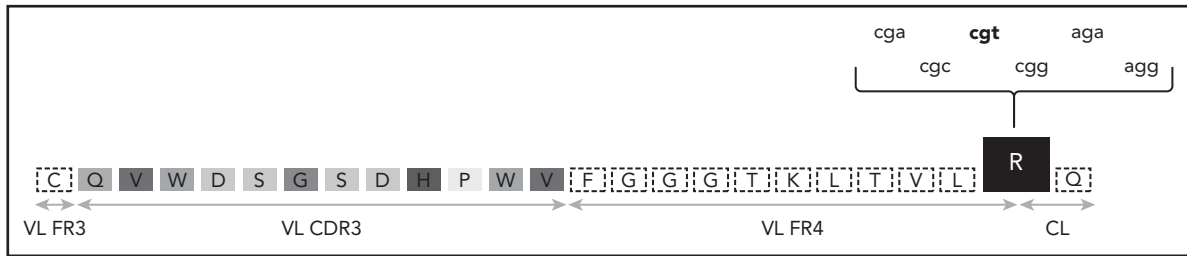
Figure 5. The G-to-R substitution at the VL-CL linker region is being degenerately encoded by different nucleotide sequences. One of the most frequent VL CDR3 of subsets 2 and 169 is depicted. The amino acids of the VL CDR3 region are enclosed in boxes of different shades of gray. The R residue at the VL-CL linker region is in a black box and can be encoded by 6 different nucleotide triplets. The prevailing triplet is set in bold. CL, constant light; FR, framework region.

BcR IG that allow homotypic BcR IG interactions.[17] In more detail, a crucial contribution to the epitope implicated in self-recognition is given by an R residue introduced by SHM at the VL-CL linker position 110. This residue was present in all subset 2 BcR IGs analyzed thus far, and its reversion to the germline-encoded G abrogated cell autonomous signaling.[17] Within this context, it is highly relevant that the G-to-R substitution was detected as a clonal event in all subset 2 and 169 IGLV3-21

rearrangements analyzed in this study, strongly indicating shared functional constraint. This conclusion is strongly supported by our crystallographic studies, which revealed homotypic interactions in subset 169 as well, that, similar to those reported for subset 2,[17] are dominated by the IGLV3-21 light chain. Of note, the R residue at the VL-CL linker region position 110, a ubiquitous SHM in both subsets, was found to play a critical role in these interactions, in subset 169, as well.
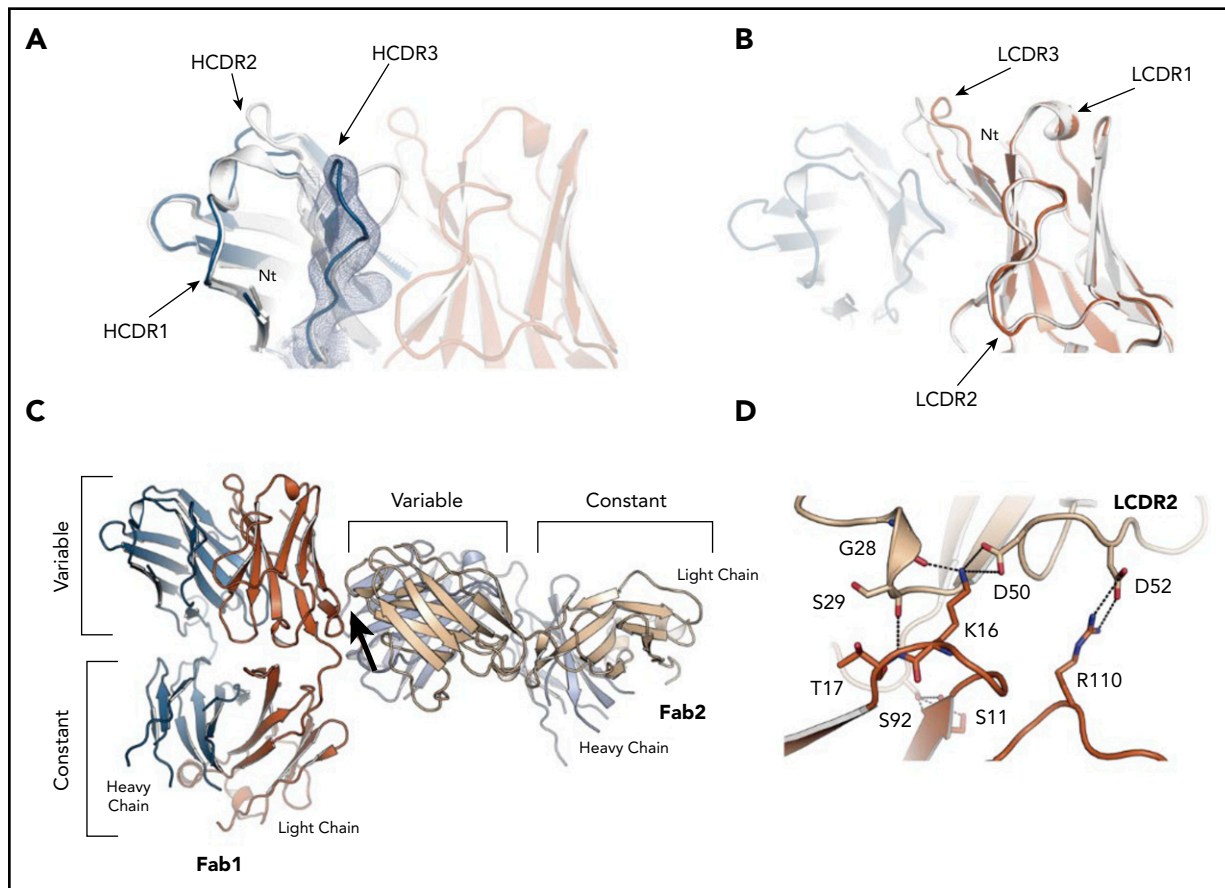

Figure 6. Structure of the subset 169 P6540 case BcR Fab fragment and its homotypic interactions. (A) Superposition of the heavy chain variable domains of P6540 and the subset 2 P11475, showing a different, more open conformation for the VH CDR3 of the P6540 Fab. A cartoon of the P6540 Fab is shown, with heavy chain (blue), light chain (orange), and CDRs labeled accordingly. The subset 2 P11475 Fab is also shown in a cartoon in white. The 2mFo-DFc electron density map is shown from the VH CDR3 loop (blue) and at a contour level of 1.0σ. (B) Superposition of the light chain variable domains of subset 169 P6540 and subset 2 P11475, showing a minor difference in the VL CDR3 conformation caused by a hinged motion of the loop. (C) The homotypic interactions between subset 169 P6540 BCR Fab molecules are revealed by the mutual orientation of 2 interacting Fab molecules in the crystal. Heavy and light chains are shown in the cartoon and are blue and orange, respectively. The interacting region is marked with an arrow. (D) Close-up of the interface between the 2 receptor molecules. The residues involved in the interaction are labeled and shown as sticks. Contacts are dominated by the IGLV3-21 light chain residues through the VL CDR2. All residues shown are also conserved and involved in the homotypic interactions between subset 2 P11475 receptor molecules leading to cell-autonomous signaling.[17] Figures were generated with Pymol (http://www.pymol.org).

The presence of G-to-R mutation in the nonsubset IGLV3-21 cases analyzed at both clonal and subclonal level further extends previous findings about the seminal role of light chains in antigen selection in CLL, while also raising the intriguing hypothesis of light chain dominance in a substantial fraction of IGLV3-21[+] CLL that should be formally proven by future research.[16,35,36] In this context, it is also worth mentioning the enrichment for IGLV3-21 expression in a particular epigenetic subgroup of CLL, the "intermediate-programmed CLL epitype,"[37,38] highlighting the distinctive nature of CLL cases expressing this particular germline specificity. Further support for this notion is offered by published evidence that IGLV3-21 CLL displays a distinct gene expression profile and a poor clinical course regardless of IGHV gene usage, SHM status and classic cytogenetic abnormalities.[39]

Our study provides additional evidence for shared antigen selection in the natural history of subsets 2 and 169 through the examination of the SHM patterns in the heavy chain gene rearrangements of the subsets. More particularly, a recurrent 3-nucleotide (AGT) deletion was detected in the VH CDR2 of 28 of 30 subset 2 cases (10 as a clonal and 18 as a subclonal event) and in all subset 169 cases (all subclonal), leading to the removal of 1 of 5 consecutive serine residues encoded by the germline sequence of both the IGHV3-21 and the IGHV3-48 genes. This finding is noteworthy, given the overall low incidence of deletions and insertions introduced by SHM within productive IG gene rearrangements in normal or pathologic B-cell repertoires (with the notable exception of HIV-1 broadly reactive neutralizing antibodies).[15,40] We have reported that the VH CDR2 deletion in subset 2 could be accommodated without significantly affecting the local structure,[15] although, admittedly, formal proof of the role of this deletion is still pending. That notwithstanding, the very high incidence of this deletion in both subsets, reported here for the first time to our knowledge, argues for shared antigenic drive leaving an identical SHM imprint on the respective heavy chains encoded by 2 phylogenetically related IGHV genes.

An unexpected finding of our study concerns the detection of (mostly) rare subclones lacking any SHM within the IGHV/IGLV genes, even in cases with a considerable SHM load. This rather enigmatic result poses several questions regarding the precise pathogenesis of subsets 2 and 169, alluding to more complex immune trajectories than previously thought, while also prompting research that may reveal whether it is subset specific or a more general phenomenon in CLL.

From a clinical perspective, accumulating evidence suggests that BcR IG stereotypy refines risk stratification in CLL, in certain cases superseding the binary classification into U-CLL and M-CLL.[12,13,41] Perhaps the strongest support of this claim is offered by subset 2, characterized by a very aggressive clinical course, even though most patients within this subset are classified as M-CLL.[13] Of note, a recent meta-analysis of sequential prospective chemoimmunotherapy trials by the German CLL Study Group concluded that subset 2 membership is an independent adverse-prognostic indicator irrespective of IGHV mutational status and should be proposed for risk stratification of patients.[42] On these grounds, the immunogenetic similarities between subsets 2 and 169 argue for their coclustering, with obvious implications for dissecting the heterogeneity of CLL and implementing tailored management approaches. This argument

is supported by published evidence of an overall similar background of genomic aberrations in subsets 2 and 169 (eg, both displaying significant enrichment for SF3B1 mutations compared with generic CLL cohorts)[20,21] and by our present finding of similar time to first treatment and overall survival ($P = .98$ and $P = .23$, respectively) when comparing subsets 2 and 169 (233 vs 14 subset 169 cases with available data from our consortium).

In summary, the immunogenetic and structural profiles reported herein cement the relatedness of subsets 2 and 169, highlighting restricted features at both clonal and subclonal level that can be explained only in the context of shared functional selection. The structural similarities of homotypic interactions identified for subsets 2 and 169 support that the same BcR IG archetype may be shared by cases using different IGHV genes albeit with common ancestry, as in the case of IGHV3-21 and IGHV3-48.

Authorship

Contribution: K.G. performed the experiments, analyzed the data, and wrote the manuscript; F.P. and M.T. assisted in the bioinformatics analysis and with data visualization; A.A.C., M.G., and C.M. performed the crystallographic experiments; K. Pasentsis assisted with the experiments; K. Plevova, L.-A.S., A.A., R.R., F.D., and S.P. provided materials and patient samples; P.B. provided the data; R.S. assisted in the research and interpretation of results; P.G., K.S., M.D., and A.C. designed and supervised the research and wrote the manuscript; and all authors provided final approval of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: K.G., 0000-0002-1331-3925; F.P., 0000-0002-0222-4273; A.A.C., 0000-0003-0481-2551; C.M., 0000-0002-8350-484X; K. Plevova, 0000-0002-6148-8877; M.T., 0000-0002-1653-0327; A.A., 0000-0003-4384-9031; F.D., 0000-0002-3764-063X; S.P., 0000-0001-7136-2680; P.G., 0000-0003-3750-7342.

Correspondence: Anastasia Chatzidimitriou, Institute of Applied Biosciences, Centre for Research and Technology Hellas,

6th km Charilaou-Thermi Rd, 57001 Thermi, Thessaloniki, Greece; e-mail: achatzidimitriou@certh.gr.

Footnotes

*M.D. and A.C. contributed equally to the study as senior authors.

REFERENCES

1. Messmer BT, Albesiano E, Efremov DG, et al. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J Exp Med.* 2004;200(4):519-525.

2. Ghiotto F, Fais F, Valetto A, et al. Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *J Clin Invest.* 2004;113(7):1008-1016.

3. Tobin G, Thunberg U, Karlsson 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):2879-2885.

4. Widhopf GF II, Rassenti LZ, Toy TL, Gribben JG, Wierda WG, Kipps TJ. Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. *Blood.* 2004;104(8):2499-2504.

5. Ghia P, Stamatopoulos K, Belessi C, et al. Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: the lesson of the IGHV3-21 gene. *Blood.* 2005;105(4): 1678-1685.

6. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood.* 2007;109(1):259-270.

7. Murray F, Darzentas N, Hadzidimitriou A, 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):1524-1533.

8. Bomben R, Dal Bo M, Capello D, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. *Br J Haematol.* 2009;144(4):492-506.

9. Darzentas N, Hadzidimitriou A, Murray F, et al. A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. *Leukemia.* 2010;24(1): 125-132.

10. Agathangelidis A, Darzentas N, Hadzidimitriou A, 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):4467-4475.

11. Stamatopoulos K, Agathangelidis A, Rosenquist R, Ghia P. Antigen receptor stereotypy in chronic lymphocytic leukemia. *Leukemia.* 2017;31(2):282-291.

12. Baliakas P, Hadzidimitriou A, Sutton LA, et al. Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. *Lancet Haematol.* 2014;1(2):e74-e84.

13. Baliakas P, Agathangelidis A, Hadzidimitriou A, et al. Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. *Blood.* 2015;125(5):856-859.

14. Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood.* 2003;101(12): 4952-4957.

15. Belessi CJ, Davi FB, Stamatopoulos KE, et al. IGHV gene insertions and deletions in chronic lymphocytic leukemia: "CLL-biased" deletions in a subset of cases with stereotyped receptors. *Eur J Immunol.* 2006;36(7): 1963-1974.

16. Hadzidimitriou A, Darzentas N, Murray F, et al. Evidence for the significant role of immunoglobulin light chains in antigen recognition and selection in chronic lymphocytic leukemia. *Blood.* 2009;113(2):403-411.

17. Minici C, Gounari M, Übelhart R, et al. Distinct homotypic B-cell receptor interactions shape the outcome of chronic lymphocytic leukaemia. *Nat Commun.* 2017;8(1):15746.

18. Baliakas P, Mattsson M, Hadzidimitriou A, et al. No improvement in long-term survival over time for chronic lymphocytic leukemia patients in stereotyped subsets #1 and #2 treated with chemo(immuno)therapy. *Haematologica.* 2018;103(4):e158-e161.

19. Marincevic M, Cahill N, Gunnarsson R, et al. High-density screening reveals a different spectrum of genomic aberrations in chronic lymphocytic leukemia patients with "stereotyped" IGHV3-21 and IGHV4-34 B-cell receptors. *Haematologica.* 2010;95(9): 1519-1525.

20. Strefford JC, Sutton LA, Baliakas P, 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): 2196-2199.

21. Rossi D, Spina V, Bomben R, et al. Association between molecular lesions and specific B-cell receptor subsets in chronic lymphocytic leukemia. *Blood.* 2013;121(24):4902-4905.

22. Malcikova J, Stalika E, Davis Z, et al. The frequency of TP53 gene defects differs between chronic lymphocytic leukaemia subgroups harbouring distinct antigen receptors. *Br J Haematol.* 2014;166(4):621-625.

23. Sutton LA, Young E, Baliakas P, et al; ERIC, the European Research Initiative on CLL. Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica.* 2016;101(8):959-967.

24. Navrkalova V, Young E, Baliakas P, et al. ATM mutations in major stereotyped subsets of chronic lymphocytic leukemia: enrichment in subset #2 is associated with markedly short telomeres. *Haematologica.* 2016;101(9): e369-e373.

25. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood.* 2018; 131(25):2745-2760.

26. Vardi A, Vlachonikola E, Karypidou M, 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): 1555-1561.

27. Kotouza MT, Gemenetzi K, Galigalidou C, et al; Hellenic Precision Medicine Network in Oncology. TRIP - T cell receptor/immunoglobulin profiler. *BMC Bioinformatics.* 2020; 21(1):422.

28. Sutton LA, Kostareli E, Hadzidimitriou 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):4460-4468.

29. Kostareli E, Sutton LA, Hadzidimitriou 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):1317-1324.

30. Venturi V, Kedzierska K, Price DA, et al. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc Natl Acad Sci USA.* 2006; 103(49):18691-18696.

31. Madi A, Shifrut E, Reich-Zeliger S, et al. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res.* 2014;24(10):1603-1612.

32. Venturi V, Price DA, Douek DC, Davenport MP. The molecular basis for public T-cell responses? *Nat Rev Immunol.* 2008;8(3): 231-238.

33. Roy A, Bystry V, Bohn G, et al. High resolution IgH repertoire analysis reveals fetal liver as the likely origin of life-long, innate B lymphopoiesis in humans. *Clin Immunol.* 2017;183:8-16.

34. Maity PC, Bilal M, Koning MT, et al. *IGLV3-21*01* is an inherited risk factor for CLL through the acquisition of a single-point mutation enabling autonomous BCR signaling. *Proc Natl Acad Sci USA.* 2020;117(8):4320-4327.

35. Stamatopoulos K, Belessi C, Hadzidimitriou A, et al. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood.* 2005;106(10):3575-3583.

36. Belessi C, Stamatopoulos K, Hadzidimitriou A, et al. Analysis of Expressed and Non-Expressed IGK Locus Rearrangements in Chronic Lymphocytic Leukemia. *Mol. Med.* 2005;11:52-58.

37. Giacopelli B, Zhao Q, Ruppert AS, et al. Developmental subtypes assessed by DNA methylation-iPLEX forecast the natural history of chronic lymphocytic leukemia. *Blood.* 2019;134(8):688-698.

38. Wojdacz TK, Amarasinghe HE, Kadalayil L, et al. Clinical significance of DNA methylation in chronic lymphocytic leukemia patients: results from 3 UK clinical trials. *Blood Adv.* 2019;3(16):2474-2481.

39. Stamatopoulos B, Smith T, Crompot E, et al. The Light Chain IgLV3-21 Defines a New Poor Prognostic Subgroup in Chronic Lymphocytic Leukemia: Results of a Multicenter Study. *Clin Cancer Res.* 2018;24(20):5048-5057.

40. Kepler TB, Liao H-X, Alam SM, et al. Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe.* 2014;16(3):304-313.

41. Baliakas P, Moysiadis T, Hadzidimitriou A, et al; European Research Initiative on CLL (ERIC). Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica.* 2019;104(2):360-369.

42. Jaramillo S, Agathangelidis A, Schneider C, et al. Prognostic impact of prevalent chronic lymphocytic leukemia stereotyped subsets: analysis within prospective clinical trials of the German CLL Study Group (GCLLSG) [published online ahead of print 26 December 2019]. *Haematologica.* 2019; doi:10.3324/haematol.2019.231027.

GEMENETZI et al