CHRONIC LYMPHOCYTIC LEUKEMIA WITH MUTATED IGHV4-34 RECEPTORS: SHARED AND DISTINCT IMMUNOGENETIC FEATURES AND CLINICAL OUTCOMES

Aliki Xochelli[1,2], Panagiotis Baliakas[2], Ioannis Kavakiotis[3], Andreas Agathangelidis[1,4], Lesley-Ann Sutton[2,5], Eva Minga[1], Stavroula Ntoufa[1], Eugen Tausch[6], Xiao-Jie Yan[7], Tait Shanafelt[8], Karla Plevova[9], Myriam Boudjogra[10], Davide Rossi[11], Zadie Davis[12], Alba Navarro[13], Yorick Sandberg[14], Fie Juhl Vojdeman[15], Lydia Scarfo[4], Niki Stavroyianni[16], Andrey Sudarikov[17], Silvio Veronese[18], Tatiana Tzenou[19], Teodora Karan-Djurasevic[20], Mark Catherwood[21], Dirk Kienle[6], Maria Chatzouli[22], Monica Facco[23], Jasmin Bahlo[24], Christiane Pott[25], Lone Bredo Pedersen[15], Larry Mansouri[2,5], Karin E. Smedby[26], Charles C Chu[7], Véronique Giudicelli[27], Marie-Paule Lefranc[27], Panagiotis Panagiotidis[19], Gunnar Juliusson[28], Achilles Anagnostopoulos[16], Ioannis Vlahavas[3], Darko Antic[29,30], Livio Trentin[23], Marco Montillo[18], Carsten Niemann[15], Hartmut Döhner[6], Anton W Langerak[14], Sarka Pospisilova[9], Michael Hallek[24], Elias Campo[13], Nicholas Chiorazzi[7], Nikos Maglaveras[31], David Oscier[12], Gianluca Gaidano[11], Diane F. Jelinek[32], Stephan Stilgenbauer[6], Ioanna Chouvarda[31], Nikos Darzentas[9], Chrysoula Belessi[22], Frederic Davi[10], Anastasia Hadzidimitriou[1,2], Richard Rosenquist[2,5], Paolo Ghia[4,33], Kostas Stamatopoulos[1,2,16]

1. Institute of Applied Biosciences, CERTH, Thessaloniki, Greece
2. Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
3. Department of informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
4. Division of Experimental Oncology and Department of Onco-Hematology, IRCCS San Raffaele Scientific Institute, Milan, Italy
5. Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
6. Department of Internal Medicine III, Ulm University, Ulm, Germany
7. Feinstein Institute for Medical Research, Northwell Health Manhasset, New York, United States
8. Department of Hematology, Department of Medicine, Mayo Clinic, Rochester, Minnesota, United States
9. Central European Institute of Technology, Masaryk University, Brno, Czech Republic

10. Department of Hematology, Hopital Pitie-Salpetriere and University Pierre et Marie Curie, Paris, France
11. Division of Haematology, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy
12. Department of Haematology, Royal Bournemouth Hospital, Bournemouth, UK
13. Institut d'investigacions biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic, University of Barcelona, Barcelona, Spain
14. Department of Immunology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands
15. Department of Hematology, Rigshospitalet, Copenhagen, Denmark
16. Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece
17. National Research Center for Hematology, Moscow, Russian Federation
18. Molecular Pathology Unit and Haematology Department, Niguarda Cancer Center, Niguarda Ca' Granda Hospital, Milan, Italy
19. First Department of Propaedeutic Medicine, University of Athens, Athens, Greece
20. Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia
21. Department of Haemato-Oncology, Belfast City Hospital, Belfast, UK
22. Hematology Department, Nikea General Hospital, Piraeus, Greece
23. Department of Medicine, Hematology and Clinical Immunology Branch, Padua University School of Medicine, Italy
24. Department I of Internal Medicine and Center of Integrated Oncology, University Hospital Cologne, Cologne, Germany
25. Second Medical Department, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany
26. Department of Medicine Solna, Clinical Epidemiology Unit, Karolinska Institutet, Stockholm, Sweden
27. IMGT®, the international ImMunoGeneTics information system®, Université de Montpellier, LIGM, Institut de Génétique Humaine IGH, UPR CNRS 1142, Montpellier, France
28. Lund University and Hospital Department of Hematology, Lund Stem Cell Center, Lund, Sweden

29. Clinic for hematology, Clinical Center, Belgrade, Serbia
30. Medical faculty, University of Belgrade, Belgrade, Serbia
31. Laboratory of Medical Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
32. Department of Immunology, Mayo Clinic, Rochester, Minnesota, United States
33. Università Vita-Salute San Raffaele and IRCCS Istituto Scientifico San Raffaele, Milan, Italy

Running title: IGHV4-34 stereotyped CLL subsets

Keywords: CLL, stereotypy, IGHV4-34, prognostication, clinicobiological associations

Financial support

This work was supported in part by H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe", by the EU; H2020 No. 692298 project "MEDGENET, Medical Genomics and Epigenomics Network" by the EU; the Swedish Cancer Society, the Swedish Research Council, Uppsala University, Uppsala University Hospital, Lion's Cancer Research Foundation, and Selander's Foundation, Uppsala; Associazione Italiana per la Ricerca sul Cancro (AIRC; #IG15189 and Special Program Molecular Clinical Oncology – 5 per mille #9965 and 10007), Milano, Italy and Ricerca Finalizzata 2010 (RF-2010-2318823) and 2011 (RF-2011-02349712) – Ministero della Salute, Roma, Italy; MEYS CZ project NPUII - CEITEC 2020 (LQ1601).

Conflict-of-interest disclosure: A.H. and K.S. received research support from Novartis SA and Janssen Pharmaceuticals. PG received research support from Celgene, Gilead, GSK, Roche, CN received research support from Novo Nordisk Foundation and Abbvie.

Corresponding author:

Anastasia Hadzidimitriou

Institute of Applied Biosciences

Center for Research and Technology Hellas

57001 Thermi, Thessaloniki, Greece

e-mail: anastasiaxtz@gmail.com

Phone: +30 2310 498474 | Fax: +30 2310 498270

STATEMENT OF TRANSLATIONAL RELEVANCE

IGHV4-34 is the most frequently utilized IGHV gene in chronic lymphocytic leukemia (CLL) cases expressing B cell receptor immunoglobulin (BcR IG) with somatically hypermutated IGHV genes (M-CLL). Amongst IGHV4-34 M-CLL cases, different subsets exist, each defined by a distinctive stereotyped BcR IG. Here, we explored whether the similarity between cases belonging to the same subset extends from immunogenetic features to other biological features and also clinical outcome. We report that IGHV4-34 M-CLL stereotyped subsets have distinct clinicobiological profiles and highlight subset #4 and #16 cases, both expressing IgG-switched BcR IG with similar somatic hypermutation patterns, as particularly indolent. Overall, these findings indicate that identifying IG sequence relationships has implications for clinicobiological research aimed at dissecting the heterogeneity of CLL and, eventually, improving clinical decision-making through the implemention of tailored patient management strategies.

ABSTRACT

PURPOSE: We sought to investigate if B cell receptor immunoglobulin (BcR IG) stereotypy is associated with particular clinicobiological features amongst chronic lymphocytic leukemia (CLL) patients expressing mutated BcR IG (M-CLL) encoded by the IGHV4-34 gene, and also ascertain whether these associations could refine prognostication.

EXPERIMENTAL DESIGN: In a series of 19,907 CLL cases with available immunogenetic information, we identified 339 IGHV4-34 expressing cases assigned to one of the four largest stereotyped M-CLL subsets, namely subsets #4, #16, #29 and #201, and investigated in detail their clinicobiological characteristics and disease outcomes.

RESULTS: We identified shared and subset-specific patterns of somatic hypermutation (SHM) amongst patients assigned to these subsets. The greatest similarity was observed between subsets #4 and #16, both including IgG-switched cases (IgG-CLL). In contrast, the least similarity was detected between subsets #16 and #201, the latter concerning IgM/D-expressing CLLs. Significant differences between subsets also involved disease stage at diagnosis and the presence of specific genomic aberrations. IgG subsets #4 and #16 emerged as particularly indolent with a significantly ($p<0.05$) longer time-to-first-treatment (TTFT) (median TTFT: not yet reached) compared to the IgM/D subsets #29 and #201 (median TTFT: 11 and 12 years, respectively).

CONCLUSIONS: Our findings support the notion that BcR IG stereotypy further refines prognostication in CLL, superseding the immunogenetic distinction based solely on SHM load. Additionally, the observed distinct genetic aberration landscapes and clinical heterogeneity suggests that not all M-CLL cases are equal, prompting further research into the underlying biological background with the ultimate aim of tailored patient management.

INTRODUCTION

The human IGHV4-34 gene has attracted great interest due to its inherent ability to encode autoreactive antibodies (Abs)(1). B cells expressing B cell receptor immunoglobulin (BcR IG) utilizing the IGHV4-34 gene are expanded following infections by microbial pathogens [including particular lymphotropic viruses e.g. cytomegalovirus (CMV) and the Epstein-Barr virus (EBV), and bacteria e.g *Mycoplasma pneumoniae*] as well as certain autoimmune disorders (e.g. systemic lupus erythematosous, SLE)(2). Especially for SLE, IGHV4-34 Abs represent a major fraction of the total serum Abs and their IgG-switched counterparts have been associated with increased disease activity and progression(3). In contrast, IgG-switched IGHV4-34 Abs are underrepresented in the serum of healthy adults(4), strongly indicating that B cells expressing IGHV4-34 BcR IG are normally under close scrutiny in order to avoid the consequences of unwanted autoreactivity.

Chronic lymphocytic leukemia (CLL) is a malignancy of mature B cells, the most frequent adult hematologic malignancy(5). Ample evidence suggests that the development and evolution of CLL are critically dependent upon microenvironmental drive: key players in this process are the immune receptors, especially the B cell receptor (BcR)(5). Strong support to this notion has been provided by the immunogenetic analysis of the clonotypic BcR IG which revealed (i) distinct outcomes for patients with differential imprint of somatic hypermutation (SHM) within the clonotypic immunoglobulin heavy variable (IGHV) genes, since patients bearing a substantial SHM load ('mutated' CLL, M-CLL) follow considerably more indolent clinical courses compared to those with no or limited SHM ('unmutated' CLL, U-CLL)(6,7); and, (ii) pronounced IG gene repertoire skewing(8,9), culminating in the existence of (quasi)identical, alias stereotyped, BcR IG in a remarkable one-third of all patients(9-11).

The IGHV4-34 gene ranks among the most frequent IG genes (~9%) in the BcR IG repertoire of CLL(8,9). Interestingly, it is utilized at even higher frequency (~12%) amongst M-CLL(12) and peaks at a remarkable 44% amongst M-CLL cases of the rare IgG variant(13). Moreover, a significant proportion (~30%) of IGHV4-34 M-CLL cases are assigned to different stereotyped subsets[9-11]. The best studied is subset #4, the largest stereotyped subset within M-CLL(9) (~2% of all M-CLL) which has also emerged as a prototype for indolent disease, likely due to the paucity of adverse genomic aberrations combined with attenuated signaling through the BcR IG(14-16).

Immunogenetically, subset #4 BcR IGs are quite distinctive as they display long and positively charged variable heavy complementarity determining region 3 (VH CDR3)(9), reminiscent of pathogenic anti-DNA auto-Abs(17,18).

Previous studies have reported distinctive SHM patterns in IGHV4-34 CLL, especially in subset #4 cases(12,19), which resemble edited autoreactive Abs(20). These observations mainly pertain to the introduction of negatively charged residues in either heavy or light or even both immunoglobulin chains(12,19). However, certain positions that constitute binding motifs for the N-acetyllactosamine (NAL) carbohydrate epitope, namely residue W7 and the AVY motif (codons 24-26 in framework region 1, FR1-IMGT) (21) remain unmutated in the vast majority of IGHV4-34 BcR in CLL(12,22-24); in principle retaining the ability to engage in (super)antigenic interactions with NAL-containing epitopes in both self and exogenous antigens(21).

Prompted by the unique biological make-up and distinctive clinical behavior of subset #4, analogous studies have been performed for other IGHV4-34-utilizing subsets; however, definitive conclusions could not be drawn due to small patient numbers(25). This limitation is not unexpected when dealing with stereotyped subsets due to the fact that even the largest subset, subset #2 (IGHV3-21/IGLV3-21), accounts for only ~3% of all CLL cases, clearly indicating that for meaningful conclusions to be reached, large patient series are imperative(9). Here, taking advantage of a cohort of ~20,000 CLL patients consolidated in the context of a multi-institutional collaboration, we performed a systematic analysis of IGHV4-34 M-CLL with a major focus on the SHM profiles, clinicobiological characteristics and prognosis of patients assigned to the 4 largest stereotyped subsets, namely subsets #4, #16, #29 and #201.

MATERIALS AND METHODS

Patients

Overall, 19,907 CLL patients with available immunogenetic data (sequences deposited in the IMGT/CLL-DB (http://www.imgt.org/CLLDBInterface/) from collaborating institutions in Europe and the United States were included in the study. All patients were diagnosed following IWCLL criteria(26). The collected clinicobiological information concerned IGHV4-34 M-CLL cases. The study was approved by local Ethics Review Committee.

FISH analysis

Interphase FISH analysis was performed using probes for the cytogenetic abnormalities included in the Döhner hierarchical model(27); namely del(17)(p13), del(11)(q23), del(13)(q14) and trisomy 12. Cell preparations were counterstained with 4,6-diamidino-phenyl-indole (DAPI) and a minimum of 200 interphase nuclei were examined(28).

CD38 expression

CD38 expression was assessed using flow-cytometry and 30% was used as a threshold to indicate positivity(6,29,30).

Analysis of gene mutations

Mutational screening for *NOTCH1, TP53 and SF3B1* genes was performed as previously described(31). In brief, PCR amplification and Sanger sequencing was performed for the following exons: 4-9 of the *TP53* gene, and 14–16 of the *SF3B1* gene. For *NOTCH1,* exon 34 or the specific mutation hotspot (del7544-45/p.P2514Rfs*4) was analyzed.

PCR amplification and sequence analysis of IGHV-IGHD-IGHJ rearrangements

PCR amplification of IGHV-IGHD-IGHJ gene rearrangements was performed on either genomic DNA (gDNA) or complementary DNA (cDNA), as previously described(9,12,32,33). PCR amplicons were subjected to direct sequencing on both strands. IGHV-IGHD-IGHJ sequence data were analyzed using the IMGT® databases and the IMGT/V-QUEST tool (http://www.imgt.org)(34,35). Only productive rearrangements were included in downstream analyses. Output data extracted

and used concerned IGH gene repertoires, VH CDR3 length and amino acid (AA) composition as well as nucleotide/amino acid changes introduced by SHM.

Silent (S) and replacement (R) mutations were calculated per FR and CDR. Considering that VH regions differ in length, absolute mutation counts were normalized by the actual nucleotide length of the corresponding VH region as previously described(12). VH FR1 sequences were not included in such comparisons as in 698 cases the primers used for the amplification of IGHV4-34 gene rearrangements were located in the VH FR1, thereby this part had to be excluded in order to avoid ambiguity.

To identify novel N-glycosylation (N-glyc) sites introduced by SHM, all IGHV4-34 rearranged sequences with less than 100% germline identity (GI) were analysed by the NETGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/).

Assignment to stereotyped subsets

Assignment to stereotyped subsets was performed using established bioinformatics methods as previously described(9,36). In brief, VH CDR3 sequences clustered together if sharing: (i) initially, at least 50% AA identity and 70% similarity regarding AA physicochemical properties, (ii) phylogenetically related IGHV genes, (iii) identical VH CDR3 lengths and (iv) identical offsets of shared VH CDR3 motifs. Each resulting subset was then described in a Bayesian model respecting the same underlying clustering criteria, enabling new sequences to be statistically assigned to the subsets(36)

Comparisons of amino acid changes introduced by somatic hypermutation in IGHV4-34 stereotyped subsets

Evaluation of the AA changes introduced by SHM was based on both qualitative and quantitative comparisons of the AA composition at each codon in the IGHV4-34-encoded portion of the VH domain spanning from VH CDR1 down to VH FR3 following a purpose-designed bioinformatics pipeline that was based on the Euclidean Distance method(37,38).

Each subset was considered as a collection of 77 vectors, one vector per AA position (codon) from VH CDR1 to VH FR3 (i=27..104; n=77). Each vector xi contained 22 attributes, which included the percentages across the whole subset of the 20 AA, gaps and stop codons for that specific position.

The difference (distance) between two subsets (X and Y) in AA distribution at the i^{th} position position was expressed as the Euclidean distance between the vectors xi and yi. This distance was calculated as follows:

$$positionDistance_i(X,Y) = \sqrt{(x_i - y_i)^2}$$

All distance metrics were adjusted and compared to the maximum difference identified.

The total distance between two subsets X and Y was employed to express the cumulative difference in AA distribution, and was calculated according to the following equation:

$$dist(X,Y) = \sum_{i=27}^{104} positionDistance_i(X,Y)$$

Statistical analysis

Differences in frequencies were evaluated using descriptive statistics. Contingency tables depicted the common distribution of pairs of categorical variables, and associations between them were assessed by the Chi-square or the Fisher's exact test for independence. TTFT was evaluated from the date of diagnosis until the date of initial treatment or the date of last follow-up for untreated cases. Survival curves were constructed using the Kaplan-Meier method, and the log-rank test was used to determine differences between survival proportions. All tests were two sided and significance was defined as a p value less than 0.05. All statistical analyses were performed using Statistica Software 10.0 (Stat Soft Inc., Tulsa, OK).

RESULTS

Overview of the immunogenetic features of IGHV4-34 CLL

A total of 20331 productive IGHV-IGHD-IGHJ rearrangements were obtained from 19,907 CLL patients; 424 cases (2.1%) carried two productive rearrangements, in line with previous reports(33). Of these 20331 IGHV-IGHD-IGHJ productive rearrangements, 1790 (8.8%) expressed the IGHV4-34 gene: 1420/1790 (79.3%) IGHV4-34 gene rearrangements had a GI <98% and, thus, represented M-CLL, whereas the remainder (370/1790, 20.7%) had a GI≥98% and were considered as unmutated (U-CLL). Within IGHV4-34 expressing U-CLL, 114 cases (6.4% of all IGHV4-34 CLL) exhibited some impact of SHM activity (98%<GI<100%), whereas the remaining 256 cases (14.3% of all IGHV4-34 CLL) carried truly unmutated sequences (GI=100%).

The VH CDR3 characteristics differed between IGHV4-34 rearrangements with distinct SHM status. More specifically: (i) U-CLL IGHV4-34 cases had significantly longer VH CDR3 (median: 21 AA for U-CLL vs 17 AA for M-CLL, p<0.001); (ii) the IGHD3-3 and IGHD2-2 genes predominated amongst U-CLL, whereas the IGHD2-15 and IGHD3-22 genes were the most frequent amongst M-CLL cases; and, (iii) M-CLL IGHV4-34 rearrangements displayed equal frequencies of the IGHJ4 and IGHJ6 genes [534/1420 (37.6%) and 510/1420 (35.9%) respectively), whereas U-CLL rearrangements showed a clear bias to the IGHJ6 gene (236/370, 63.8%), with fewer cases (75/370, 20.3%) utilizing the IGHJ4 gene. That notwithstanding, interesting exceptions became apparent when cases were grouped in different stereotyped subsets, particularly for M-CLL (see below).

Following the approach described in methods(9,39), 546/1790 (30.5%) cases were assigned to stereotyped subsets, 423 cases (23.6%) belonging to M-CLL and the remaining 123 cases (6.9%) to U-CLL (Table 1)(Supplemental Figure 1). Hereafter, we focused our attention on those IGHV4-34 M-CLL subsets comprising 50 or more cases (the largest IGHV4-34 M-CLL subsets observed), namely: subsets #4 (n=185), subset #16 (n=51), subset #29 (n=50), and subset #201 (n=53).

Differential imprints of SHM on IGHV4-34 stereotyped subsets

Differences in the distribution of SHM were observed amongst the various IGHV4-34 stereotyped subsets. In particular, subsets #16, #29 and #201 had lower R/S mutation ratios within the VH CDR1 compared to the VH CDR2, while the opposite was evidenced in subset #4 (statistical comparisons were performed based on the number of R mutations; p=0.03 for R

mutations within CDR1). Within the FRs, subsets #4 and #16 had overall similar R/S mutation ratios within VH FR2 and VH FR3, whereas subsets #29 and #201 had higher R/S mutation ratios in VH FR3 compared to VH FR2 (Figure 1). More frequent targeting of AID/APOBEC hotspot(40)/(41) motifs was identified in the CDR1 over the CDR2 (53% versus 32%, p<0.001) of subset #201 cases, thus contrasting subsets #4, #16 and #29 that followed the opposite pattern.

In a proportion (698/1790 cases, 39%) of the IGHV4-34 M-CLL cases under study, the clonotypic IGHV-IGHD-IGHJ gene rearrangement was PCR-amplified using VH FR1 primers, hence the VH FR1 could not be analysed completely. Therefore, in order to avoid confounding effects and/or possible biases, when performing comparisons between IGHV4-34 cases, we focused our attention on codons 27-104 within the VH domain (from CDR1-IMGT to FR3-IMGT) and assessed the sequence distance/similarity between subsets and the corresponding IGHV4-34 germline sequence based on a pairwise qualitative and quantitative comparison of the respective amino acid composition. The minimum distance calculated, and hence the greatest similarity, was observed between subsets #4 and #16, both being IgG-switched cases (IgG-CLL), which is notable given the overall rarity of IgG-CLL(42,43). In contrast, the maximum distance, implying the least similarity, was detected between subsets #16 and #201, the latter representing IgM/D-CLL (Figure 2).

Extreme variations between subsets were noted in codons spanning the entire VH domain, highlighting a subset-biased distribution of SHM (results summarized in Table 2). In more detail, we observed that almost all (155/156, 99.4%) IGHV4-34 M-CLL subset cases with available VH FR1 sequence data retained the germline-encoded W at codon 7, critically involved in the creation of the N-acetyl-lactosamine binding motif; however, differences were noted between subsets regarding the incidence of SHM in the other residues of this motif, namely AVY at codons 24-26 (ranging from 0% for subset #16 to 32.5% for subset #201 (p=0.0006). An additional example concerned the germline-encoded N-glycosylation (N-glyc) Asn-His-Ser motif at codons VH CDR2 57-59 which was abrogated by SHM significantly less frequently in subset #16 (9/51 cases, 17.6%) compared to subsets #4 (72/185, 39%), #29 (23/50, 46%) and #201 (23/53, 43.4%) (p=0.016). Prompted by this novel finding and also considering the emerging role of SHM-induced N-glyc motifs as a mechanism to modulate antibody avidity, potentially alleviating autoreactivity(44,45), we analyzed the VH FR1 - VH FR3 part of the VH domain of the stereotyped IGHV4-34 subsets for the presence of additional N-glyc motifs. A remarkable enrichment for N-glyc motifs generated by SHM was identified in the VH FR3 of subset #201,

particularly codons 66-68 (37.7% versus 0-3.9% in the remaining subsets, p<0.001); codons 67-69 (22.6% versus 0% in the remaining subsets, p<0.001); and, codons 77-79 (13.2% versus 0% in the remaining subsets, p<0.001).

In keeping with previous observations(12), the stereotyped IGHV4-34 M-CLL subsets displayed recurrent replacement AA changes, albeit often with markedly different frequencies within each subset (Table 2 and Supplemental figure 2A). Illustrative examples of the most striking statistically significant differences (p<0.0001) include the following:

(i) codon 28 in VH CDR1 was heavily targeted for a recurrent change from glycine to glutamic or aspartic acid (G28>E/D) in subsets #4 (68.6%) and #16 (84.4%), thus sharply contrasting subsets #29 (6.9%) and #201 (6.5%);

(ii) codon 36 in VH CDR1 carried a recurrent G36>E/D change in 72.9% of subset #29 cases thus contrasting all other subsets, especially #16 that showed no AA change at this codon in any case examined;

(iii) codon 40 in VH FR2 displayed a conservative serine to threonine change (S40>T) in 40.5% and 57% of subset #4 and #16 cases, respectively, in contrast to subsets #29 and #201 where the vast majority of cases (0%-1.8%, respectively) remained in germline configuration;

(iv) codon 45 in VH FR2 exhibited a proline to serine change (P45>S) in 37.8%, 34% and 22.4% of subset #4, #16 and #29 cases, that was seen in almost none of subset #201 cases (frequency, 1.8%)

(v) codon 55 in VH FR2 carried a glutamic acid to glutamine change (E55>Q) in 47% of subset #16 cases versus only 3.7% of subset #4 cases and no subset #29 or subset #201 case;

(vi) codon 64 in VH CDR2 that was targeted for a recurrent serine to isoleucine change (S64>I) in 48% of subset #29 cases, thus contrasting all other subsets (0-9.4% frequency of the S64>I change).

Clinicobiological associations

In order to explore whether the distinct immunogenetic profiles identified here were associated with different biological and clinical characteristics, we assessed the clinicobiological characteristics of 275 IGHV4-34 stereotyped subset cases (Table 3 and Supplemental figure 2B). Significant differences were observed between subsets regarding i) disease burden at diagnosis; ii) CD38 expression; iii) frequency of del(13q); and iv) *TP53* abnormalities (deletion of chromosome 17p and/or *TP53* mutations, *TP53*abn). In more detail, although the great majority

of all IGHV4-34 stereotyped subset cases were diagnosed at Binet stage A, percentages ranged from >90% in IgG subsets #4 and #16 to 83% in subset #201 and 74% in subset #29 (p=0.029). However, when comparing with the remaining M-CLL no statistically significant differences were observed (p=0.067). CD38 positivity ranged from extremely low (1%) in subset #4 to 10% in subset #201 (p=0.013).

Notably, the large IGHV4-34 M-CLL subsets under study (#4, #16, #29 and #201) mainly concerned younger patients (<70 years old) in contrast to the remaining M-CLL (p=0.009) and even more interestingly they had a higher prevelance of patients <55 years old (p=0.03 for all comparisons).

Regarding genetic aberrations, del(13q) was identified as a sole aberration in 76% of subset #29, 55.8% of subset #201 and 44.7% and 34% of subsets #4 and #16 respectively (p=0.01 for all comparisons). Interestingly, 4/29 (14%) cases in subset #29 carried *TP53*abn, thus significantly contrasting subsets #4 (3.6%), #16 (0%) and #201 (3%) (p<0.05 for all comparisons); such aberrations were also less frequent in the remaining IGHV4-34 M-CLL (6.7%), however the difference from subset #29 did not reach statistical significance (p=0.14). No differences between subsets were identified regarding other genetic aberrations that were either absent (*NOTCH1* exon 34 mutations) or infrequent (*SF3B1* mutations, trisomy 12, del(11q): all with frequencies ranging from 0% to 9%) (Table 3).

TTFT was analyzed for 204 cases. IgG subsets #4 and #16 had significantly (p=0.046) longer TTFT (median not yet reached at 8.8 years) compared to the IgM/D subsets #29 and #201 (median: 11 and 12 years, respectively) (Figure 3). Even more interestingly, subsets #4 and #16 also displayed significantly (p=0.023) longer TTFT compared to non-subset IGHV4-34 M-CLL cases (median TTFT: 12.6 years) (Supplemental figure Figure 3A) or M-CLL cases utilizing other IGHV genes (median TTFT: 11.9 years, p=0.00038) (Supplemental Figure 3B). The differences in TTFT held even after excluding cases carrying *TP53*abn (p value for comparisons to non-subset IGHV4-34 M-CLL cases and M-CLL cases utilizing other IGHV genes is 0,031 and 0,0003 respectively).

We also assessed clinical implications of stereotypy within the broader context of M-CLL. Evaluated parameters included advanced clinical stage (Binet B/C), male gender, CD38 positivity (30% cut-off), cytogenetic aberrations of the Döhner model [del(13q), +12, del(11q) and del(17p)] and subset #4 membership; less populated subsets such as #16, #29 and #201 were not evaluated as the small group size would not allow reaching powerful statistical conclusions. On univariable analysis (n=2335; Supplemental Table 1), membership in stereotyped subset #4

was significantly (p=0.0001) associated with longer TTFT, whereas advanced clinical stage, CD38 positivity, trisomy 12, del(11q) and del(17p) predicted shorter TTFT (p<0.05 for all comparisons). On multivariable analysis (n=1107), only advanced clinical stage, CD38 positivity and del(11q) retained statistical significance (p<0.001), while subset #4 membership had borderline significance with a p-value of 0.08, predicting for a 37% lower risk for treatment administration (Supplemental Table 1).

DISCUSSION

The SHM status of the clonotypic IGHV genes is a robust prognosticator in CLL(6,7). However, recent evidence suggests that immunogenetic features in addition to SHM, particularly BcR stereotypy, can also be clinically relevant(25) and define distinct subgroups with different prognosis even amongst cases with similar SHM status. Regarding the latter, studies by us and others have highlighted subset #4 as remarkably indolent even when compared with M-CLL cases harboring isolated del(13q)(25,46,47), traditionally considered as the most favorable-prognostic subgroup of CLL patients(27). However, no firm conclusions could be drawn for other, less populated M-CLL subsets, mainly due to the small number of patients analysed.

Here, taking advantage of a very large dataset of IG gene rearrangement sequences from cases with CLL from our multi-institutional consortium deposited in the IMGT/CLL-DB (http://www.imgt.org/CLLDBInterface/query), we reappraised the immunogenetic features and clinicobiological profiles of M-CLL stereotyped subsets, focusing on cases utilizing the IGHV4-34 gene, the most frequent IGHV gene amongst M-CLL. We report that IGHV4-34 expressing M-CLL stereotyped subsets are characterized by 'public' as well subset-biased SHM patterns that allude to both shared and distinct immunopathogenetic processes, while also supporting a functional purpose for the observed AA changes introduced by SHM.

More specifically, all subset #16 cases and the majority (>80%) of subset #4 and #29 cases have retained an intact carbohydrate NAL-binding motif, in principle allowing for superantigenic interactions with NAL-containing self and exogenous antigens. This is in contrast to subset #201 cases, where disruption of the AVY motif due to SHM was identified in 32.5% of cases, which is higher than the frequency observed in all IGHV4-34 M-CLL cases evaluated. Subset #201 is also noteworthy owing to the high frequency of novel N-glycosylation motifs created by SHM within the VH FR3 (30/53 cases, 56%). These findings along with several other examples of subset-biased distribution of SHM throughout the VH domain allude to particular antigen exposure histories and/or immune responses. However additional research is needed in order to clarify the functional purpose for each and every one of these AA changes (Table 2).

In an attempt to obtain insight into the potential functional and/or clinical relevance of the SHM characteristics observed on IGHV4-34 M-CLL stereotyped subsets, we developed a novel bioinformatics approach for assessing the overall similarity of their primary sequences as shaped by SHM. This approach, enabled us to identify high overall similarity between subsets #4 and

#16, both IgG-switched CLL, which differed significantly from subsets #29 and #201, both expressing the IgM/D isotype. Taking into consideration the overall rarity of IgG-expressing CLL(42,43), this observation is unlikely to be due to serendipity alone, but instead may be considered as further evidence in support of selective forces driving the ontogeny and, perhaps, evolution of different stereotyped subsets(10,48,49).

From a clinical perspective, our study confirms and significantly extends previous reports regarding the indolent clinical behavior of subset #4, while also offering firm evidence that subset #16 is another particularly indolent variant of CLL. Hence, the overall immunogenetic similarity between these two subsets is also reflected in their clinical behavior and outcome. Of note, preliminary investigations into the signaling capacity of subset #16 based on the effects of TLR stimulation and stimulation through the BcR using anti-IgG strongly allude to a signaling profile akin to subset #4, for which we recently reported an anergic phenotype(16), indicating that IgG-switched stereotyped subsets of IGHV4-34 CLL cluster together and are distint from the IgM/D variant.

Additional noteworthy observations concern the exceptionally high frequency (73%) of del(13q) as a sole aberration in subset #29, combined with a 14% frequency of *TP53*abn which is unusually high for M-CLL, where such aberrations are rather scarce (<3%) at diagnosis(50). Admittedly, further analysis with next generation sequencing techniques is needed in order to confirm and possibly highlight this subset-biased distribution in *TP53* aberrations. That said, these findings further highlight a more complex disease model for this particular stereotyped subset, potentially implying that an intricate interplay of cell-intrinsic and cell-extrinsic factors are involved in disease development and progression.

In conclusion, we document different spectra of SHM and AA changes between stereotyped IGHV4-34 CLL subsets. The finding of subset-biased, recurrent AA changes at certain codons indicates that the respective progenitor cells may have responded in a specific manner to the selecting antigen(s), despite expressing the same IGHV gene, indicating a functional purpose for these modifications. Moreover, the finding of differing outcomes for different stereotyped subsets with similar SHM status i.e. M-CLL reinforces our previous claim that integration of subset membership into well established prognostic models such as the hierarchical Döhner model can further refine prognostication in CLL(10,25).

ACKNOWLEDGMENTS

A.X., P.B. and A.A. performed research and wrote the paper. I.K., LA.S., I.C. performed research. E.M. was responsible for clinicobiological data registration. A.H. was responsible for the clinicobiological database and performed research. S.N. performed experiments. E.T., XJ.Y., T.S., K.P., M.B., D.R., Z.D., A.N., Y.S., F.J.V., L.S., N.S., A.S., S.V., T.T., T.K.D., M.C., D.K., M.C., M.F., J.B., C.P., L.B.P., L.M., K.S., C.C.C., V.G., M.P.L., P.P., G.J., A.A., I.V., D.A., L.T., M.M., C.N., H.D., A.W.L., S.P., M.H., E.C., N.C., N.M., D.O., G.G., D.F.J., S.S., N.D., C.B., and F.D. provided data. R.R. P.G. and K.S. designed the study and wrote the paper which was approved by all the authors.

The authors wish to thank Stavroula Smerla, Eva Koravou, Evangelia Mouchtaropoulou and Diane Hatzioannou for their technical support with data assessment and definitions.

REFERENCES

1. Bhat NM, Bieber MM, Chapman CJ, Stevenson FK, Teng NN. Human antilipid A monoclonal antibodies bind to human B cells and the i antigen on cord red blood cells. J Immunol **1993**;151(9):5011-21.

2. Mockridge CI, Rahman A, Buchan S, Hamblin T, Isenberg DA, Stevenson FK*, et al.* Common patterns of B cell perturbation and expanded V4-34 immunoglobulin gene usage in autoimmunity and infection. Autoimmunity **2004**;37(1):9-15.

3. van Vollenhoven RF, Bieber MM, Powell MJ, Gupta PK, Bhat NM, Richards KL*, et al.* VH4-34 encoded antibodies in systemic lupus erythematosus: a specific diagnostic marker that correlates with clinical disease characteristics. The Journal of rheumatology **1999**;26(8):1727-33.

4. Pugh-Bernard AE, Silverman GJ, Cappione AJ, Villano ME, Ryan DH, Insel RA*, et al.* Regulation of inherently autoreactive VH4-34 B cells in the maintenance of human B cell tolerance. The Journal of clinical investigation **2001**;108(7):1061-70 doi 10.1172/JCI12462.

5. Kipps TJ, Stevenson FK, Wu CJ, Croce CM, Packham G, Wierda WG*, et al.* Chronic lymphocytic leukaemia. Nature reviews Disease primers **2017**;3:17008 doi 10.1038/nrdp.2017.8.

6. 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.

7. 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.

8. 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. The Journal of clinical investigation **1998**;102(8):1515-25 doi 10.1172/JCI3009.

9. 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.

10. 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.

11. Messmer BT, Albesiano E, Efremov DG, Ghiotto F, Allen SL, Kolitz J*, et al.* Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. The Journal of experimental medicine **2004**;200(4):519-25 doi 10.1084/jem.20040544.

12. 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 doi 10.1182/blood-2007-07-099564.

13. 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. Clinical cancer research : an official journal of the American Association for Cancer Research **2016**;22(1):167-74 doi 10.1158/1078-0432.CCR-14-3017.

14. 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**;18:1281-91 doi 10.2119/molmed.2011.00480.

15. Marincevic M, Cahill N, Gunnarsson R, Isaksson A, Mansouri M, Goransson H*, 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):1519-25 doi 10.3324/haematol.2009.021014.

16. Ntoufa S, Papakonstantinou N, Apollonio B, Gounari M, Galigalidou C, Fonte E*, et al.* B Cell Anergy Modulated by TLR1/2 and the miR-17 approximately 92 Cluster Underlies the Indolent Clinical Course of Chronic Lymphocytic Leukemia Stereotyped Subset #4. J Immunol **2016**;196(10):4410-7 doi 10.4049/jimmunol.1502297.

17. Krishnan MR, Jou NT, Marion TN. Correlation between the amino acid position of arginine in VH-CDR3 and specificity for native DNA among autoimmune antibodies. J Immunol **1996**;157(6):2430-9.

18. Jang YJ, Stollar BD. Anti-DNA antibodies: aspects of structure and pathogenicity. Cellular and molecular life sciences : CMLS **2003**;60(2):309-20.

19. Hadzidimitriou A, Darzentas N, Murray F, Smilevska T, Arvaniti E, Tresoldi C*, et al.* Evidence for the significant role of immunoglobulin light chains in antigen recognition and

selection in chronic lymphocytic leukemia. Blood **2009**;113(2):403-11 doi 10.1182/blood-2008-07-166868.

20. Li H, Jiang Y, Prak EL, Radic M, Weigert M. Editors and editing of anti-DNA receptors. Immunity **2001**;15(6):947-57.

21. Potter KN, Hobby P, Klijn S, Stevenson FK, Sutton BJ. Evidence for involvement of a hydrophobic patch in framework region 1 of human V4-34-encoded Igs in recognition of the red blood cell I antigen. J Immunol **2002**;169(7):3777-82.

22. 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.

23. Sutton LA, Ljungstrom V, Mansouri L, Young E, Cortese D, Navrkalova V*, et al.* Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. Haematologica **2015**;100(3):370-6 doi 10.3324/haematol.2014.109777.

24. 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.

25. Baliakas P, Hadzidimitriou A, Sutton LA, Minga E, Agathangelidis A, Nichellati M*, et al.* Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicenter study. The Lancet Haematology **2014**;1(2):e74-e85.

26. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H*, et al.* Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood **2008**;111(12):5446-56 doi 10.1182/blood-2007-06-093906.

27. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L*, et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. The New England journal of medicine **2000**;343(26):1910-6 doi 10.1056/NEJM200012283432602.

28. Baliakas P, Iskas M, Gardiner A, Davis Z, Plevova K, Nguyen-Khac F*, et al.* Chromosomal translocations and karyotype complexity in chronic lymphocytic leukemia: a systematic

reappraisal of classic cytogenetic data. American journal of hematology **2014**;89(3):249-55 doi 10.1002/ajh.23618.

29. Thunberg U, Johnson A, Roos G, Thorn I, Tobin G, Sallstrom J*, et al.* CD38 expression is a poor predictor for VH gene mutational status and prognosis in chronic lymphocytic leukemia. Blood **2001**;97(6):1892-4.

30. Hamblin TJ, Orchard JA, Ibbotson RE, Davis Z, Thomas PW, Stevenson FK*, et al.* CD38 expression and immunoglobulin variable region mutations are independent prognostic variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease. Blood **2002**;99(3):1023-9.

31. Baliakas P, Hadzidimitriou A, Sutton LA, Rossi D, Minga E, Villamor N*, et al.* Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia **2015**;29(2):329-36 doi 10.1038/leu.2014.196.

32. 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.

33. Langerak AW, Davi F, Ghia P, Hadzidimitriou A, Murray F, Potter KN*, et al.* Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. Leukemia **2011**;25(6):979-84 doi 10.1038/leu.2011.49.

34. 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 research **2008**;36(Web Server issue):W503-8 doi 10.1093/nar/gkn316.

35. Lefranc MP. Immunoglobulin and T Cell Receptor Genes: IMGT((R)) and the Birth and Rise of Immunoinformatics. Frontiers in immunology **2014**;5:22 doi 10.3389/fimmu.2014.00022.

36. Bystry V, Agathangelidis A, Bikos V, Sutton LA, Baliakas P, Hadzidimitriou A*, et al.* ARResT/AssignSubsets: a novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy. Bioinformatics **2015**;31(23):3844-6 doi 10.1093/bioinformatics/btv456.

37. Chakrabarti S, Bryant SH, Panchenko AR. Functional specificity lies within the properties and evolutionary changes of amino acids. Journal of molecular biology **2007**;373(3):801-10 doi 10.1016/j.jmb.2007.08.036.

38. Gilbert PB, Wu C, Jobes DV. Genome scanning tests for comparing amino acid sequences between groups. Biometrics **2008**;64(1):198-207 doi 10.1111/j.1541-0420.2007.00845.x.

39. 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.

40. Rogozin IB, Pavlov YI. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. Mutation research **2003**;544(1):65-85.

41. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. J Immunol **2004**;172(6):3382-4.

42. Vardi A, Agathangelidis A, Sutton LA, Chatzouli M, Scarfo L, Mansouri L*, et al.* IgG-switched CLL has a distinct immunogenetic signature from the common MD variant: ontogenetic implications. Clinical cancer research : an official journal of the American Association for Cancer Research **2014**;20(2):323-30 doi 10.1158/1078-0432.CCR-13-1993.

43. 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. Clinical cancer research : an official journal of the American Association for Cancer Research **2006**;12(6):1672-9 doi 10.1158/1078-0432.CCR-05-2164.

44. 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. Proceedings of the National Academy of Sciences of the United States of America **2014**;111(25):E2567-75 doi 10.1073/pnas.1406974111.

45. Garces F, Hyun Lee J, De Val N, Torrents de la Pena A, Kong L, Purchades C*, et al.* Affinity Maturation of a Potent Family of HIV Antibodies is primarily focused on accomodating or Avoidding Glycans. Immunity **2015**;43:1053-63.

46. 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.

47. Bomben R, Dal Bo M, Capello D, Forconi F, Maffei R, Laurenti L*, et al.* Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. British journal of haematology **2009**;144(4):492-506 doi 10.1111/j.1365-2141.2008.07469.x.

48. 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 research **2014**;74(16):4211-6 doi 10.1158/0008-5472.CAN-14-0630.

49. Sutton LA, Agathangelidis A, Belessi C, Darzentas N, Davi F, Ghia P*, et al.* Antigen selection in B-cell lymphomas--tracing the evidence. Seminars in cancer biology **2013**;23(6):399-409 doi 10.1016/j.semcancer.2013.07.006.

50. Best OG, Gardiner AC, Davis ZA, Tracy I, Ibbotson RE, Majid A*, et al.* A subset of Binet stage A CLL patients with TP53 abnormalities and mutated IGHV genes have stable disease. Leukemia **2009**;23(1):212-4 doi 10.1038/leu.2008.260.

Tables

Table 1. Summary of the immunogenetic features of all IGHV4-34 subset cases from the present cohort. Subsets investigated in-depth are highlighted in cyan.

Subset	n	IGHD	IGHJ	Mutated/ unmutated	median IGHV gene germline identity%	VH CDR3-IMGT length
#4	185	IGHD5-18	IGHJ6	Mutated	93.3	20
#16	51	IGHD2-15	IGHJ6	Mutated	93.8	24
#29	50	IGHD6-19	IGHJ3	Mutated	93	14
#201	53	unassigned	IGHJ3	Mutated	92.6	17
#N4-34-1	5	unassigned	IGHJ4	Mutated	92.69	10
#N4-34-10	2	unassigned	IGHJ1	Mutated	96.31	16
#N4-34-11	3	IGHD3-10	IGHJ2	Mutated	96.14	16
#N4-34-12	5	IGHD2-2	IGHJ4	Mutated	94.04	16
#N4-34-13	2	IGHD2-21	IGHJ4	Mutated	92.81	16
#N4-34-14	5	IGHD6-6	IGHJ6	Mutated	93.68	17
#192	3	IGHD5-12	IGHJ4	Mutated	93.68	12
#198	5	IGHD6-19	IGHJ4	Mutated	94.12	9
#N4-34-2	3	unassigned	IGHJ4	Mutated	92.28	10
#N4-34-3	3	IGHD2-15	IGHJ4	Mutated	95.09	13
#N4-34-5	2	IGHD2-15	IGHJ4	Mutated	89.62	13
#N4-34-6	4	IGHD4-23	IGHJ4	Mutated	94.91	13
#N4-34-8	2	IGHD5-18	IGHJ4	Mutated	94.03	14
#N4-34-9	3	IGHD6-19	IGHJ4	Mutated	94.56	14
#N4-34-GF	3	unassigned	IGHJ4	Mutated	90.88	13
#N4-34-WE	2	IGHD3-3	IGHJ4	Mutated	92.76	13
#N4-34X	7	IGHD3-22	IGHJ4	Mutated	94.57	15
#N4-34-7	5	IGHD2-2	IGHJ4	Mutated	93.9	14
#11	20	IGHD3-10	IGHJ4	Mutated	94.11	15
#125	7	IGHD3-3	IGHJ6	Unmutated	100	25
#129	3	IGHD2-2	IGHJ4	Unmutated	100	21
#130	16	IGHD3-3	IGHJ6	Unmutated	100	23
#205	17	IGHD6-19	IGHJ6	Unmutated	100	17
#207	2	unassigned	IGHJ2	Unmutated	100	20
#64D	25	IGHD2-2	IGHJ6	Unmutated	100	21
#N4-34-15	3	IGHD6-6	IGHJ6	Unmutated	100	19
#N4-34-16	4	IGHD2-2	IGHJ6	Unmutated	100	20
#N4-34-17	4	IGHD2-2	IGHJ6	Unmutated	100	20
#N4-34-20	7	IGHD2-2	IGHJ6	Unmutated	100	23
#N4-34-21	11	IGHD3-3	IGHJ6	Unmutated	100	23
#N4-34-22	4	IGHD3-3	IGHJ6	Unmutated	100	24
#N4-34-23	14	IGHD2-2	IGHJ6	Unmutated	100	26
#N4-34-19	2	IGHD2-21	IGHJ6	Unmutated	100	21
#N4-34-18	4	IGHD2-2	IGHJ6	Unmutated	98.39	20

Table 2. Summary of Somatic Hypermutation characteristics in major IGHV4-34 stereotyped subsets and and the remaining IGHV4-34 M-CLL cases.

	Subset #4	Subset #16	Subset #29	Subset #201	p-value (across subsets)	remaining IGHV4-34 M-CLL	p-value (across subsets and the remaining IGHV4-34 M-CLL)
SHM at codon 7 (FR1-IMGT)	1/87 (1.14%)	0/19 (0%)	0/27 (0%)	0/23 (0%)	p=non significant	2/557 (0.35%)	p=non significant
AVY motif disruption (Codons 24-26; FR1-IMGT)	17/149 (11.4%)	0/39 (0%)	6/35 (17.1%)	13/40 (32.5%)	p=0.0002	108/816 (13.23%)	p=0.0006
SHM at codon 28 (CDR1-IMGT)	115/166 (69.3%)	38/45 (84.4%)	5/43 (11.6%)	3/46 (6.5%)	p<0.0001	115/945 (12.1%)	p<0.0001
Codon 28: G28->D/E	114/166 (68.6%)	38/45 (84.4%)	3/43 (6.9%)	3/46 (6.5%)	p<0.0001	77/945 (8.1%)	p<0.0001
SHM at codon 36 (CDR1-IMGT)	42/184 (22.8%)	0/48 (0%)	38/48 (79%)	7/52 (13.4%)	p<0.0001	380/1024 (37.1%)	p<0.0001
Codon 36: G36->D/E	39/184 (21.1%)	0/48 (0%)	35/48 (72.9%)	5/52 (9.6%)	p<0.0001	242/1024 (23.6%)	p<0.0001
SHM at codon 40 (FR2-IMGT)	97/185 (52.4%)	28/49 (57%)	0/49 (0%)	3/53 (5.6%)	p<0.0001	456/1038 (43.9%)	p<0.0001
Codon 40: S40->T	75/185 (40.5%)	28/49 (57%)	0/49 (0%)	1/53 (1.8%)	p<0.0001	303/1038 (29.1)	p<0.0001
SHM at codon 45 (FR2-IMGT)	103/185 (55.7%)	23/50 (46%)	22/49 (44.9%)	3/53 (5.6%)	p<0.0001	319/1045 (30.5%)	p<0.0001
Codon 45: P45->S	70/185 (37.8%)	17/50 (34%)	11/49 (22.4%)	1/53 (1.8%)	p<0.0001	231/1045 (22.1%)	p<0.0001
SHM at codon 55 (FR2-IMGT)	20/185 (10.8%)	26/51 (50.9%)	0/50 (0%)	0/53 (0%)	p<0.0001	80/1057 (7.5%)	p<0.0001
Codon 55: E55->Q	7/185 (3.7%)	24/51 (47%)	0/50 (0%)	0/53 (0%)	p<0.0001	20/1057 (1.9%)	p<0.0001
Disruption of CDR2 N-glycosylation motif	72/185 (39%)	9/51 (17.6%)	23/50 (46%)	23/53 (43.4%)	p=0.11	433/1057 (41%)	p=0.016
Recurrent disruption of both AVY and N-glyc motifs	13/149 (8.7%)	0/51 (0%)	2/35 (5.7%)	5/53 (9.4%)	p=non significant	39/815 (4.8%)	p=non significant
SHM at codon 64 (CDR2-IMGT)	76/185 (41%)	25/51 (49%)	50/50 (100%)	29/53 (54.7%)	p<0.0001	405/1057 (38.3%)	p<0.0001
Codon 64: S64->I	9/185 (4.8%)	0/51 (0%)	24/50 (48%)	5/53 (9.4%)	p<0.0001	40/1057 (3.7%)	p<0.0001
Creation of novel N-glycosylation motifs in FR3 by SHM	4/185 (2.16%)	2/51 (3.9%)	1/50 (2%)	23/53 (43.3%)	p<0.0001	55/1057 (5.2%)	p<0.0001

Table 3. Summary of the clinicobiological characteristics in different IGHV4-34 M-cases evaluated in the present study

	#4|n=150	#16|n=44	#29|n=39	#201|n=42	p-value (across subsets)	Remaining IGHV4-34 M-CLL|n=354	p-value (across subsets and the remaining IGHV4-34 M-CLL)
Male	76/150 (51%)	21/44 (48%)	18/39 (46%)	22/42 (52%)	p=0.93	218/354 (61.6%)	p=0.057
Age at diagnosis (median)	57 (37-95)	56 (37-85)	58.5 (36-77)	57 (41-100)		66 (36-92)	
<55	58/136 (43%)	17/41 (41%)	14/37 (38%)	15/42 (36%)	p=0.85	101/354 (28.5%)	p=0.03
50<n<70	59/136 (43.4%)	20/41(48.8%)	18/37 (48.6%)	21/42 (50%)	p=0.83	164/354 (46.3%)	p=0.92
>70	19/136 (14%)	4/41 (10%)	5/37 (14%)	6/42 (14%)	p=0.9	89/354 (25.1%)	p=0.009
Clinical Stage (Binet)							
A	116/127 (91%)	37/40 (93%)	26/35 (74%)	30/36 (83%)	p=0.029	228/265 (86%)	p=0.067
B	8/127 (6%)	2/40 (5%)	6/35 (17%)	4/36 (11%)	p=0.16	22/265 (8.3%)	p=0.2
C	3/127 (2%)	1/40 (2%)	3/35 (9%)	2/36 (6%)	p=0.33	16/265 (6%)	p=0.4
High CD38 expression[a]	1/95 (1%)	1/26 (4%)	1/21 (5%)	3/29 (10%)	p=0.12	34/190 (18%)	p=0.0003
High ZAP-70 expression	4/37 (10.8%)	0/9 (0%)	0/10 (0%)	1/10 (10%)	p=0.53	21/97 (21.6%)	p=0.13
del(13q)[b]	43/96 (44.7%)	9/26 (34.6%)	19/25 (76%)	19/34 (55.8%)	p=0.01	71/150 (47.3%)	p=0.027
Trisomy 12	3/102 (3%)	1/28 (4%)	0/27 (0%)	3/37 (8%)	p=0.34	25/157 (16%)	p=0.001
del(11q)[c]	3/104 (3%)	1/28 (4%)	1/27 (4%)	0/38 (0%)	p=0.72	6/171 (3.5%)	p=0.84
***TP53* abnormality**[d]	4/112 (3.6%)	0/31 (0%)	4/29 (14%)	1/37 (3%)	p=0.04	11/178 (6.2%)	p=0.11
del(17p)	3/103 (2.9%)	0/28 (0%)	3/27 (11.11%)	1/37 (3%)		11/178 (6.2%)	
***TP53* mutation**	2/48 (4.1 %)	0/11(0%)	1/18 (5.5%)	1/15 (7%)		N/A	
SF3B1	1/46 (2%)	1/11 (9%)	0/17 (0%)	0/18 (0%)	p=0.35	N/A	N/A
NOTCH1	0/57 (0%)	0/14 (0%)	0/18 (0%)	0/18 (0%)		N/A	N/A
Other malignancy	13/65 (20%)[e]	4/20 (20%)[f]	6/25 (24%)[g]	1/15 (7%)[h]	p=0.58	N/A	N/A

a: cut-off >30%, b: deletion of chromosome 13q, c: deletion of chromosome 11q, d: deletion of chromosome 17p and/or TP53 mutation, e: of the 13 patients diagnosed with a second malignancy, only 4 received treatment; 3/4 were diagnosed with a second malignancy prior to treatment initiation; only 1/13 developed a hematological malignancy; f: of the 4 patients diagnosed with a second malignancy, only one received treatment after the diagnosis of the second malignancy; none of these patients developed a hematological malignancy; g: of the 6 patients diagnosed with a second malignancy, only 3 received treatment; 2/3 were diagnosed with a second malignancy prior to treatment initiation; none of these patients developed a hematological malignancy; h: Only one patient was diagnosed with a second malignancy (non-hematological), following treatment.

FIGURE LEGENDS

Figure 1. Subset-biased distribution of replacement/silent (R/S) mutation ratios in stereotyped IGHV4-34 CLL. IGHV4-34 M-CLL stereotyped subsets displayed an asymmetric distribution of R/S mutations within the different VH subregions.

Figure 2. Sequence similarity between subsets and the corresponding IGHV4-34 germline sequence (G4-34). The results are presented as a heatmap. Color gradient sets to the lowest value (light red)-least identity-to the highest (dark red) – greatest identity (explained also in the color bar above). The greatest similarity was calculated for subsets #4 and #16, whereas the least similar subsets were subsets #16 and #201

Figure 3. CLL with mutated IGHV4-34 receptors: shared and distinct clinical outcomes. IgG subsets #4 and #16 had significantly (p=0.046) longer TTFT (median not yet reached) compared to the IgM/D subsets #29 and #201 (median: 11 and 12 years, respectively).

Figure 1

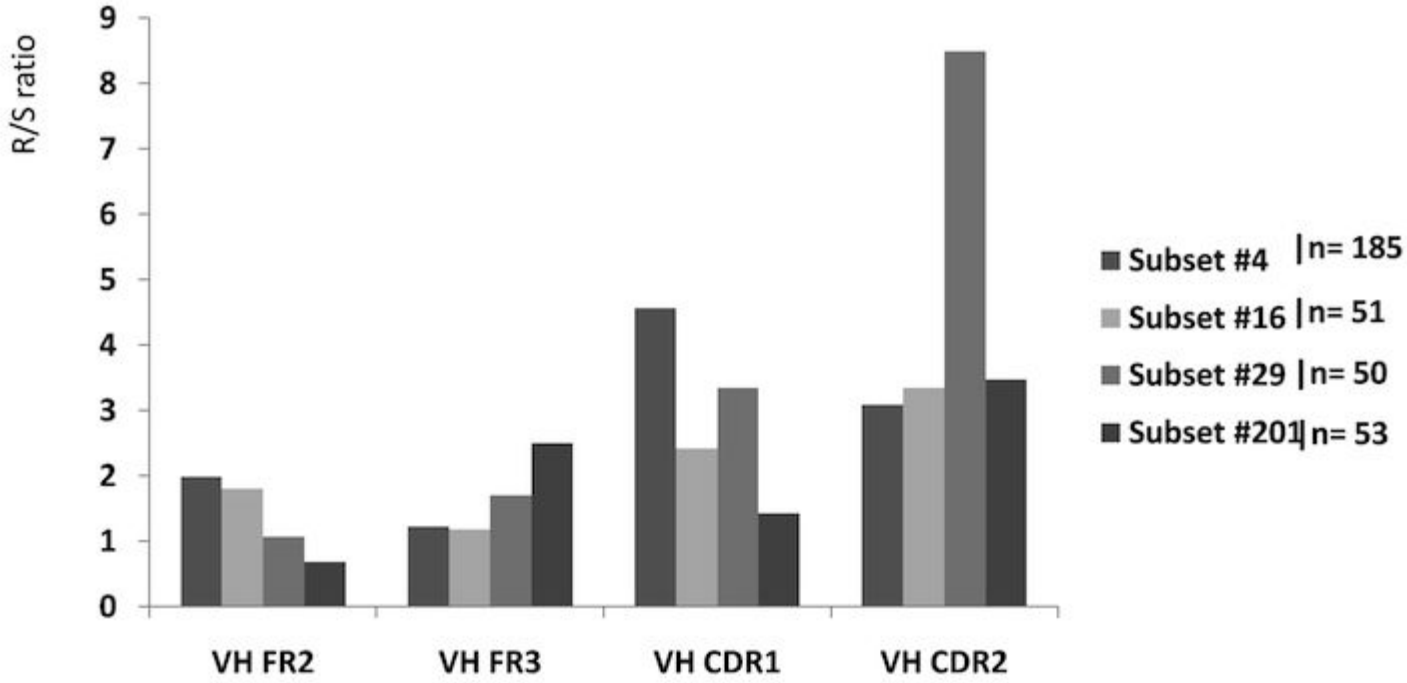


Figure 2

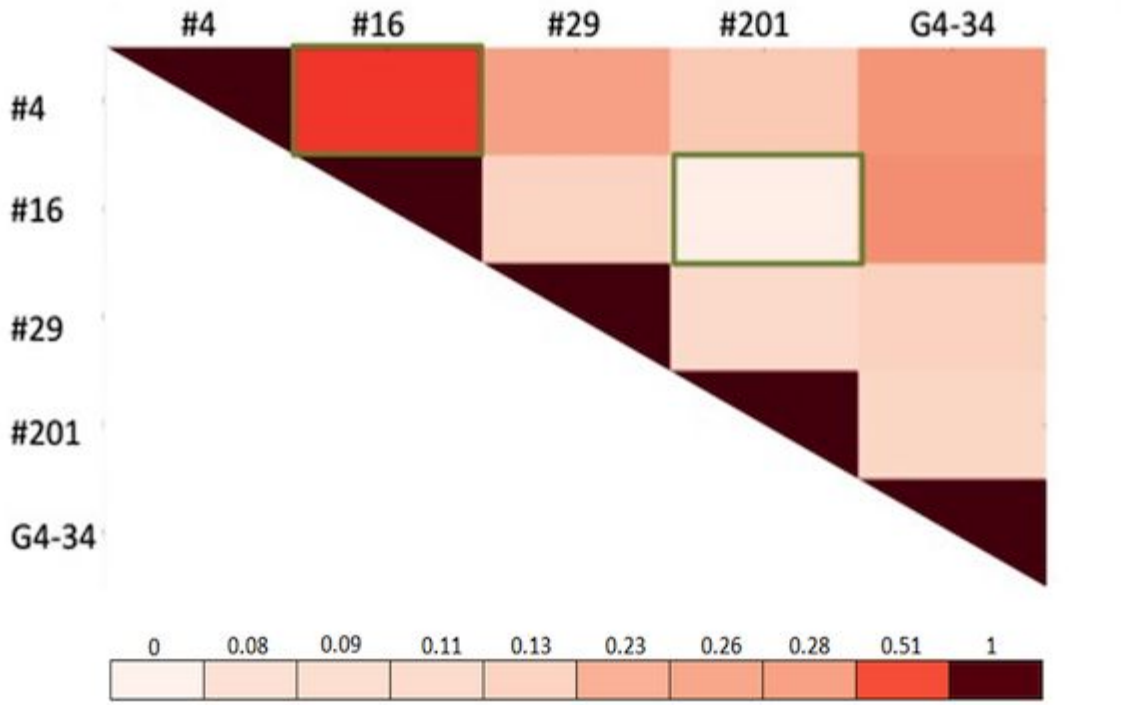
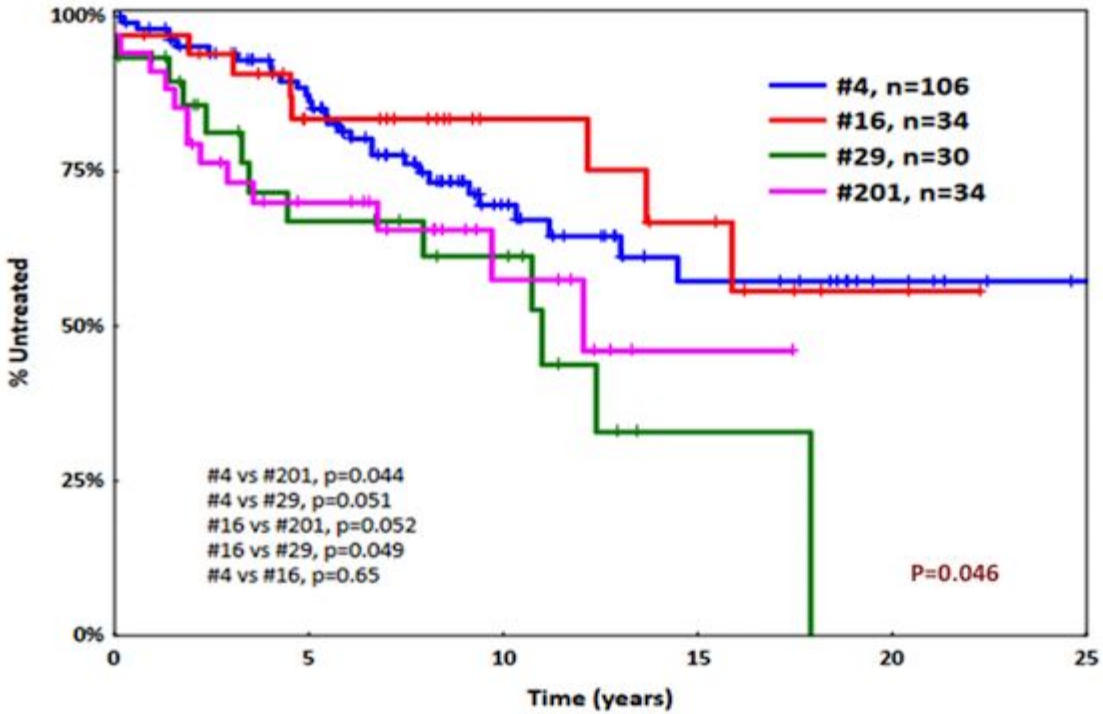


Figure 3



AACR American Association for Cancer Research

Clinical Cancer Research

CHRONIC LYMPHOCYTIC LEUKEMIA WITH MUTATED IGHV4-34 RECEPTORS: SHARED AND DISTINCT IMMUNOGENETIC FEATURES AND CLINICAL OUTCOMES

Aliki Xochelli, Panagiotis Baliakas, Ioannis Kavakiotis, et al.

Clin Cancer Res Published OnlineFirst May 23, 2017.