

An Immunogenetic Signature of Ongoing Antigen Interactions in Splenic Marginal Zone Lymphoma Expressing IGHV1-2*04 Receptors

Vasilis Bikos^{1,2}, Maria Karypidou^{1,3}, Evangelia Stalika³, Panagiotis Baliakas^{1,4}, Aliki Xochelli^{3,4}, Lesley-Ann Sutton⁴, George Papadopoulos⁵, Andreas Agathangelidis⁶, Evdoxia Papadopoulou⁵, Zadie Davis⁷, Patricia Algara⁸, George Kanellis⁹, Alexandra Traverse-Glehen¹⁰, Manuela Mollejo⁸, Achilles Anagnostopoulos¹, Maurilio Ponzoni¹¹, David Gonzalez¹², Sarka Pospisilova², Estella Matutes¹², Miguel Angel Piris¹³, Theodora Papadaki⁹, Paolo Ghia⁶, Richard Rosenquist⁴, David Oscier⁷, Nikos Darzentas², Dimitrios Tzovaras⁵, Chrysoula Belessi¹⁴, Anastasia Hadzidimitriou^{3,4}, and Kostas Stamatopoulos^{1,3,4}

Abstract

Purpose: Prompted by the extensive biases in the immunoglobulin (IG) gene repertoire of splenic marginal-zone lymphoma (SMZL), supporting antigen selection in SMZL ontogeny, we sought to investigate whether antigen involvement is also relevant post-transformation.

Experimental Design: We conducted a large-scale subcloning study of the IG rearrangements of 40 SMZL cases aimed at assessing intraclonal diversification (ID) due to ongoing somatic hypermutation (SHM).

Results: ID was identified in 17 of 21 (81%) rearrangements using the immunoglobulin heavy variable (IGHV)1-2*04 gene versus 8 of 19 (40%) rearrangements utilizing other IGHV genes

($P = 0.001$). ID was also evident in most analyzed IG light chain gene rearrangements, albeit was more limited compared with IG heavy chains. Identical sequence changes were shared by subclones from different patients utilizing the IGHV1-2*04 gene, confirming restricted ongoing SHM profiles. Non-IGHV1-2*04 cases displayed both a lower number of ongoing SHMs and a lack of shared mutations (per group of cases utilizing the same IGHV gene).

Conclusions: These findings support ongoing antigen involvement in a sizable portion of SMZL and further argue that IGHV1-2*04 SMZL may represent a distinct molecular subtype of the disease. *Clin Cancer Res*; 22(8); 2032–40. ©2015 AACR.

¹Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece. ²Central European Institute of Technology, Masaryk University, Brno, Czech Republic. ³Institute of Applied Biosciences, CERTH, Thessaloniki, Greece. ⁴Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ⁵Information Technologies Institute, CERTH, Thessaloniki, Greece. ⁶Division of Experimental Oncology and Department of Onco-Hematology, Università Vita-Salute San Raffaele and Istituto Scientifico San Raffaele, Milan, Italy. ⁷Department of Haematology, Royal Bournemouth Hospital, Bournemouth, United Kingdom. ⁸Department of Pathology, Hospital Virgen de la Salud, Toledo, Spain. ⁹Hematopathology Department, Evangelismos Hospital, Athens, Greece. ¹⁰Department of Pathology and Hematology, Hospices Civils de Lyon, Université Lyon, Lyon, France. ¹¹Pathology Unit, San Raffaele Scientific Institute, Milan, Italy. ¹²Section of Haemato-Oncology, Institute of Cancer Research, London, United Kingdom. ¹³Hospital Universitario Marques de Valdecilla, Santander, Cantabria, Spain. ¹⁴Hematology Department, Nikea General Hospital, Pireaus, Greece.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

V. Bikos and M. Karypidou contributed equally to this article.

Corresponding Author: Kostas Stamatopoulos, Institute of Applied Biosciences, Center for Research and Technology Hellas, 57001 Thessaloniki, Greece. Phone: 3023-1049-8271; Fax: 3023-1049-8270; E-mail: kostas.stamatopoulos@gmail.com

doi: 10.1158/1078-0432.CCR-15-1170

©2015 American Association for Cancer Research.

Introduction

Splenic marginal-zone lymphoma (SMZL) is recognized as a distinct clinical and pathologic entity in the 2008 WHO classification of Tumors of the Hematopoietic and Lymphoid tissues (1). Studies over the last decade have shed light on SMZL pathophysiology revealing that both cell-intrinsic defects (genomic and epigenomic aberrations, dysregulated miRNAs etc.; refs. 2–13) as well as cell-extrinsic mechanisms contribute to disease development and evolution. Regarding the latter, seminal immunogenetic findings regarding the distinctive features of the B-cell receptor immunoglobulin (BcR IG) repertoire of SMZL (14–18) underscore the role of functional selection, suggesting that antigens and/or superantigens are involved in disease ontogeny.

Focusing on IG heavy chains, at striking odds with serendipity, 20% to 30% of SMZL cases express BcR IGs utilizing a single polymorphic variant of the immunoglobulin heavy variable (IGHV)1-2 gene, namely IGHV1-2*04 (18), which is noteworthy given that this allele is much less frequent in other B-cell lymphomas (19, 20). Additional molecular imprints of antigen selection are evident in the heavy complementarity determining region 3 (VH CDR3) of IGHV1-2*04 receptors which display biased usage of the immunoglobulin heavy

Translational Relevance

Almost one-third of cases with splenic marginal zone lymphoma (SMZL) express B-cell receptor immunoglobulin (BcRIG) encoded by the IGHV1-2*04 gene, suggesting antigen selection in disease ontogeny. Emerging evidence suggests that IGHV1-2*04 cases comprise a distinct SMZL subtype associated with specific cytogenetic aberrations, gene mutations, and an overall worse outcome. Prompted by these observations, we explored whether antigen involvement is relevant throughout the disease course of SMZL through investigating intraclonal diversification (ID) within the IG gene rearrangements expressed by the malignant cells. We report that (i) ID was more frequent by far in IGHV1-2*04 cases versus cases expressing BcR IG utilizing other IGHV genes; and, (ii) somatic hypermutation (SHM) characteristics were effectively distinguishing IGHV1-2*04 from other rearrangements. Overall, these findings support ongoing antigen interactions throughout the natural history of IGHV1-2*04 SMZL and have implications for optimal patient management, indicating BcR inhibition as a reasonable therapeutic option at least for IGHV1-2*04 cases.

diversity (IGHD)3-3 and 3-10 genes, leading to the creation of antigen binding sites with restricted sequence features. Furthermore, the great majority (>85%) of the clonotypic IGHV genes in SMZL show evidence of somatic hypermutation (SHM) with a distribution and molecular characteristics typical of receptors that have undergone selection by antigen (18). A similar overall picture has emerged from the analysis of the clonotypic IG light chains (21), revealing restrictions in the immunoglobulin kappa variable (IGKV) and immunoglobulin lambda variable (IGLV) gene repertoire, biased associations of the IGHV1-2*04 with the IGKV3-20, IGKV1-8, and IGLV2-14 genes, and, distinctive SHM patterns, cumulatively alluding to functional constraints.

Although these results convincingly implicate antigen involvement in the development of SMZL, the question remains as to whether antigen is relevant not only in selecting SMZL progenitors but also post-transformation, thus exerting an ongoing selective pressure on the malignant cells and, by extension, shaping their behavior. Preliminary work from our group indicated a continued influence of antigen on four SMZL clones expressing IGHV1-2*04 receptors reflected by intraclonal diversification (ID) within the rearranged IG heavy variable genes (18).

Here, we systematically explored the occurrence of ID in the IG heavy genes of 40 patients with SMZL, including 21 cases expressing IGHV1-2*04 receptors. In order to obtain a more comprehensive view, we also profiled ID in the partner IG light chains in 12 cases. Our findings argue for ongoing antigen activation in SMZL leading to clonal evolution, at least for a substantial proportion of cases. The extensive and distinctive ID observed in IGHV1-2*04 cases compared with cases utilizing other IGHV genes further supports the hypothesis that the former might constitute a distinct subgroup, with important implications for both SMZL subclassification and future clinical and biologic research.

Patients and Methods

Study group

Forty patients with a diagnosis of SMZL from collaborating institutions in Greece, the United Kingdom, Spain, Italy, and France were included in the study. The study group included 24 males and 16 females with a median age of 71 years (range, 49–81). For splenectomized patients ($n = 22$), the diagnosis was based on spleen biopsy histopathologic findings, following the 2008 WHO Classification criteria (1). For non-splenectomized patients ($n = 18$), the diagnosis was based on the Splenic B-cell Lymphoma Group guidelines (22). Intraclonal diversification results within the IGHV genes of 4 patients have previously been reported (18). The study was conducted in accordance with the Declaration of Helsinki and approved by the local Ethics Committees.

PCR amplification and subcloning of IG rearrangements

Immunoglobulin heavy variable-diversity-joining (IGHV-IGHD-IGHJ) gene rearrangements and immunoglobulin kappa or lambda variable-joining (IGK/LV-IGK/LJ) gene rearrangements were PCR amplified as previously reported (18, 21) using: (i) for heavy chain gene rearrangements: IGHV leader primers or consensus primers for the heavy chain variable framework region 1 (VH FR1) and IGHJ-specific primers; and, (ii) for light chain gene rearrangements: consensus primers for the VK FR1/VL FR1 and IGKJ/IGLJ-specific primers (Supplementary Tables S1A–S1C). The starting material was extracted from peripheral blood (23 cases) or spleen specimens (17 cases); genomic DNA was used in 16 cases, whereas for the remaining 24 cases, we used cDNA extracted from total cellular RNA. No treatment other than splenectomy had been administered in any of the analyzed cases before sample collection. All reactions were performed using the high-fidelity Accuprime Pfx polymerase (Invitrogen, Life technologies), a thermostable, proofreading enzyme with a very low error rate ($2.9\text{--}3.7 \times 10^{-6}$ /base).

Purified PCR products were ligated into the pCR2.1 vector (Invitrogen Life technologies) and the vectors were transformed into *E. Coli/TOP10F* competent bacteria (Invitrogen Life technologies). A median of 21 colonies per amplicon (range, 6–72) were randomly chosen and sequenced using M13 primers (Supplementary Tables S1A and S1C).

Sequence data analysis, interpretation, and definitions

Sequences were analyzed using IMGT databases (23, 24) and the IMGT/V-QUEST tool (25, 26) as well as a suite of in-house, purpose-built algorithms and tools. Codons and amino acid (AA) positions are described according to the IMGT unique numbering of the V domain. For sequences obtained with FR1 consensus primers, nucleotide changes were evaluated from codon 15 in FR1-IMGT to the end of CDR3-IMGT. For the subgroup of cases amplified using leader primers, the entire VH domain was evaluated (from FR1-IMGT codon 1 to the end of CDR3-IMGT).

Each nucleotide mutation in every sequence was recorded, as was the change or preservation of the corresponding amino acid (AA), identified as replacement (R) or silent (S), respectively. AA changes were characterized as "conservative" or "non-conservative" following standardized biochemical criteria. Sequences were also analyzed for targeting to the tetranucleotide (4-NTP) motifs RGYW/WRCY (R = A/G, Y = C/T, and W = A/T) as well as the dinucleotide (2-NTP) motifs WA and TW (27).

Intraclonal diversification was assessed through examination of sequence variation in the V domain, following definitions previously proposed by our group (28). In brief, all "non-ubiquitous" sequence changes from the germline were evaluated and characterized as follows: (i) unconfirmed mutation (UCM): a mutation observed in only one subcloned sequence from the same specimen (unique); and (ii) confirmed mutation (CM): a mutation observed more than once among subcloned sequences from the same specimen (partially shared). Amino acid changes resulting from UCMs or CMs are designated by the abbreviations UAA or CAA, respectively. We also defined as trademark mutations those present in only certain subclones of a given case, serving as the defining feature for distinct "branches" (subclones) of the clonal population.

To account for the fact that a mutation is more likely to occur in a FR rather than a CDR simply due to its greater length, each mutation was "weighted," or normalized, as reported by our group (28). In particular, normalized mutation frequencies (NMF) were calculated according to the following formula: $\Sigma (CM+UCM)/[\text{number of subclones}] \times [\text{region length}]$.

Pattern discovery within the VH CDR3

In order to identify common motifs within the VH CDR3, in IGHV1-2*04 rearrangements, we first extracted the VH CDR3 region sequences from all the respective subclones and pairwise compared them in order to end up with one representative for each group of identical VH CDR3 regions. This resulted in a total of 51 unique sequences. Because the VH CDR3 sequences differed in their length (range 18–27 AAs), gaps were introduced in shorter rearrangements to facilitate their alignment. The alignment was performed using default settings within the Clustal Omega (29) command line interface tool. Aligned sequences were then processed by applying an adaptive dictionary algorithm (30). In brief, this algorithm operates by first scanning each possible AA combination within a sequence and then checking to see whether a particular combination already exists. An additional computation was performed whereby the position of the AA was taken into consideration, thus enabling the identification of common patterns shared between the VH CDR3 sequences (30–32).

Results

IG gene repertoires and mutational status

Heavy chains. Overall, 830 subcloned IGHV-IGHD-IGHJ rearrangement sequences from 40 patients with SMZL were evaluated for intraclonal diversification. For 3 cases, samples from two different timepoints were investigated. The study was intentionally biased to cases with clonotypic BcRs utilizing the IGHV1-2*04 gene; indeed, of the 40 analyzed cases, 21 (53%) expressed this gene, whereas the remaining 19 cases utilized one of 8 different IGHV genes. Sequences were categorized according to IGHV gene mutational status following definitions recently reported by our group (18, 20). In particular, 2 of 40 (5%) were classified as "truly unmutated" (100% germline identity, GI); 26 of 40 (65%) were classified as "borderline mutated" (97%–99.9% GI); and finally, 12 of 40 (30%) were "significantly mutated" (<97% GI). Of note, 19 of 21 IGHV1-2*04 cases were assigned to the "borderline mutated" group with only a single case being assigned to each of the other two mutational groups. Detailed information about the IG heavy chain gene repertoire and SHM status is provided in Supplementary Table S1A.

Light chains. Intraclonal diversification within IG light chain genes was also analyzed for 12 of 40 cases (Supplementary Tables S1B and S1C). Overall, 357 subcloned IGKV-IGKJ ($n = 263$) and IGLV-IGLJ ($n = 94$) rearrangement sequences were evaluated. In two cases, samples from different timepoints were investigated.

Eight of the 12 cases (67%) analyzed for IGK/IGL gene rearrangements were classified as "borderline mutated" while the remaining were "significantly mutated". Nine of 12 (75%) cases analyzed for both IG heavy and light chain genes displayed concordant SHM, that is, both the IGH and IGK/IGL rearrangements were assigned to the same SHM category. For 2 of 12 cases (17%), both expressing IGHV1-2*04, the partner kappa light chains were more somatically hypermutated; finally, the opposite scenario was observed in one case, that is, a heavier SHM load within the IGH gene rearrangement compared with the light chain gene rearrangement.

Intraclonal diversification analysis at the nucleotide level

Heavy chains. Alignment of the subcloned IGHV-IGHD-IGHJ rearrangement sequences revealed that: (i) 4 of 43 studied samples (9.3%) carried identical subclones (no ID); (ii) 12 of 43 (27.9%) carried only unconfirmed mutations (UCMs, mutations in single subclones; unconfirmed ID, UID); and, finally, (iii) 27 of 43 (62.8%) carried confirmed mutations (CMs, identical mutations in at least two subclones; confirmed ID, CID). CID was significantly more frequent among rearrangements utilizing the IGHV1-2*04 versus other IGHV genes [19/23 (83%) vs. 8/20 samples (40%), χ^2 test, $P = 0.001$]. This was also evident when comparing the NMF values for the group of IGHV1-2*04 rearrangements versus all other rearrangements, showing that the introduction of point mutations was significantly more frequent among the former group (Mann–Whitney U test, $P = 0.003$; Fig. 1 and Supplementary Table S2A).

All observed CMs and UCMs ($n = 570$) were single base changes, no insertions and/or deletions were identified. Overall, 152 unique CMs and 418 unique UCMs were identified, of which 395 were transitions whereas 175 were transversions: thus, transitions predominated in keeping with a canonical SHM process (33). A significantly higher CM/UCM ratio was observed among cases utilizing the IGHV1-2*04 (119/269, 0.44) versus other genes, for example, IGHV3-23 (4/27, CM/UCM ratio = 0.15; χ^2 test, $P = 0.036$) and, especially, IGHV4-34 (2/52, CM/UCM ratio = 0.04; χ^2 test, $P = 0.00003$; Supplementary Table S2B).

Moreover, we examined the distribution of CMs and UCMs across FR1/FR2/FR3 and CDR1/CDR2 and noted that the CDRs were significantly more targeted than the FRs in IGHV1-2*04 rearrangements, especially when only CMs were considered, whereas the opposite was observed for IGHV4-34 gene rearrangements. That said, IGHV1-2*04 rearrangements displayed ID (CMs and UCMs) in all VH sub-regions (Supplementary Table S2C). The VH CDR3 was assessed separately and found to display limited ID, essentially restricted to IGHV1-2*04 rearrangements. In particular, we found that 14 of 16 (87.5%) CMs and 50 of 79 (63%) UCMs concerned IGHV1-2*04 cases. Overall, CMs within the VH CDR3 were identified in 8 of 21 IGHV1-2*04 cases versus 1 of 19 cases utilizing other IGHV genes (χ^2 test, $P = 0.02$; Supplementary Table S2D).

Finally, analysis for SHM targeting to the RGYW/WRCY and WA/TW hotspot motifs along the IGHV gene region revealed that only a minor fraction of the identified CMs and UCMs was located within such motifs and this was particularly prominent in

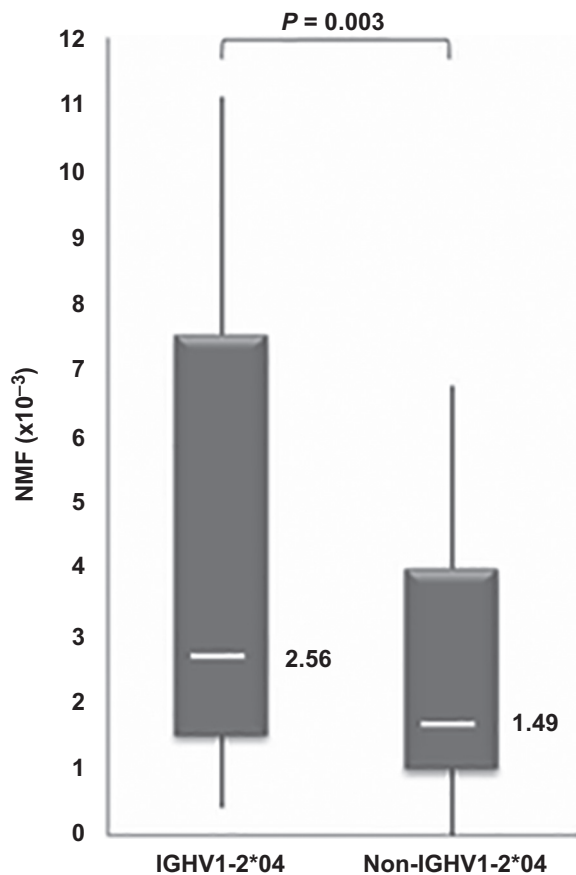


Figure 1. Box plot representation of the normalized mutational factor (NMF) values. IGHV1-2*04 rearrangements ($n = 23$ samples from 21 cases) versus all other rearrangements ($n = 20$ samples from 19 cases). The introduction of point mutations was significantly more frequent among the IGHV1-2*04 group in comparison with rearrangements that utilize all other IGHV genes.

IGHV1-2*04 subcloned sequences (18/77 CMs and 35/164 UCMs; all V-regions except VH CDR3; Supplementary Table S2E).

Light chains. Only 1 of 15 (6.7%) subcloned rearrangements exhibited no ID; 5 of 15 (33.3%) rearrangements carried UID, and, finally, 9 of 15 (60%) rearrangements were found to carry CMs and hence had CID (Supplementary Table S3A). Similar to the IGH rearrangements, only point mutations were observed amongst the subcloned IGK/IGL rearrangement sequences. Overall, 25 CMs and 44 UCMs were identified with transitions predominating over transversions (19:6 CMs and 34:10 UCMs), in agreement with a canonical SHM process (ref. 33; Supplementary Table S3B).

Analysis of the distribution of point mutations per gene group was meaningful only for IGKV1-8 and IGKV4-1 rearrangements in which a sufficient number of mutations were identified: for these groups, higher NMF values were observed within the CDRs (Supplementary Table S3C). As evidenced in the VH CDR3, limited ID occurred within the VK or VL CDR3 (8 CMs and 5 UCMs; Supplementary Table S3D). Only a minor fraction of the observed CMs and UCMs was located within WA/TW and RGYW/WRCY hotspot motifs (Supplementary Table S3E).

Pairs of IGH and IGK or IGL gene rearrangements. The extent of ID was generally concordant between heavy and light chains (Supplementary Table S3F). Taking only CMs into consideration, 9 of 12 (75%) cases displayed confirmed ID in either the heavy or the light chain or both. In the case of IGHV1-2*04 cases, with paired heavy/light data, all heavy chain rearrangements displayed CID and were coupled with a light chain rearrangement displaying CID in 2/3 occasions. Moreover, one could notice the concurrent transition of the heavy/light rearrangement from UID to CID state in the IGHV1-2*04/IGKV1-8 case where the rearrangements from two distinct timepoints were subjected to analysis.

Intraclonal diversification analysis at the amino acid level

Seventy-four of 124 CMs (59.6%) identified in the subclones of IGH gene rearrangements resulted in a novel amino acid (AA) change; a similar ratio was observed for UCMs (219/363, 60.3%; Supplementary Tables S2B and S4A). The IGHV1-2*04 subcloned rearrangements displayed enrichment for confirmed AA changes (CAA) within the VH FR2, clearly distinguishing them from all other rearrangements (Supplementary Table S4B). Amongst IG light gene rearrangements, 14 of 25 CMs (56%) and 34 of 43 UCMs (79%) resulted in an AA change (Supplementary Table S3B); AA changes were more frequent within IGKV1-8 and IGKV4-1 gene rearrangements; however, the majority were unconfirmed changes (UAA) rather than CAA.

Recurrent amino acid replacements in IGH gene rearrangements

Identical CAAs were shared by subcloned sequences of different rearrangements utilizing the same IGHV gene (Supplementary Table S5). In fact, 16 identical CAAs were documented in at least two different rearrangements and hence characterized as "recurrent"; notably, all but one (15/16, 94%) recurrent CAAs were identified amongst IGHV1-2*04 gene rearrangements, alluding to restriction within the ongoing SHM process (Fig. 2). The most striking example concerns codon 39 in VH FR2 which was targeted for a methionine to isoleucine (M>I) CAA in 5 of 21 IGHV1-2*04 rearrangements. The M>I change was identified as ubiquitous (i.e., present in all subcloned sequences) in an additional 14 IGHV1-2*04 rearrangements; hence, overall, 19 of 21 (90.5%) IGHV1-2*04 SMZL cases carried an identical AA change introduced by SHM either as a clonal or subclonal event. A similar finding occurred at codon 87 in VH FR3, where 8 of 21 (38.1%) IGHV1-2*04 rearrangements carried the same alanine to valine (A>V) change introduced by SHM at either a clonal ($n = 5$) or subclonal level ($n = 3$). This pattern of clonal/subclonal mutational events resulted in groups of subclones from different patients sharing a similar SHM profile at certain positions throughout the VH domain. An example of such cases is depicted in the sequence network figure (Fig. 3).

Identification of common VH CDR3 motifs within IGHV1-2*04 expressing SMZL

Of the 503 IGHV1-2*04 rearrangement sequences included in our cohort, we excluded sequences with identical VH CDR3s from the alignment and performed pattern analysis for the remaining rearrangements ($n = 51$). The most frequent patterns concerned dipeptide motifs. To be more precise, the valine-valine dipeptide present at positions 17-18 within the aligned VH CDR3 sequences occurred in 28 of 51 (55%), while the glycine-valine dipeptide located at positions 16-17 were present in 27 of 51 (53%) of

Downloaded from <http://aacrjournals.org/clinccancerres/article-pdf/22/8/2035/2037003/2032.pdf> by guest on 20 July 2024

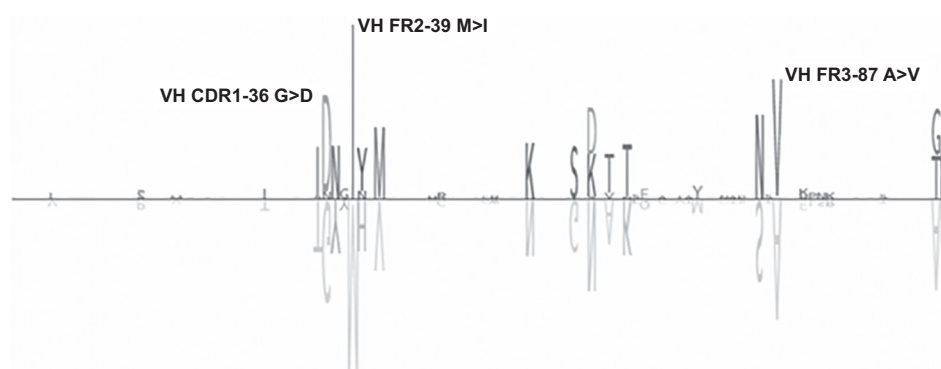


Figure 2.

Recurrent amino acid replacements introduced by ongoing SHM in IGHV1-2*04 gene rearrangements ($n = 504$). In this subfamily logo, each VH position is displayed as a stack of upright amino acid (AA) symbols. The height of each single letter AA symbol is directly proportional to the relative frequency of that AA at a given position in the alignment. The germline AAs of the IGHV1-2*04 allele are shown upside down: the height of the inverted germline AA symbol equates to the sum of the height of the upright amino acids. Positions with an equal distribution are left blank. The logo was created with TeXshade, an alignment shading software completely written in TEX/LATEX (51), through MikTeX 2.9, an up-to-date implementation of TeX/LaTeX for Windows.

sequences. Less prevalent motifs included the glycine-arginine motif at positions 21-22 (20/51; 39%) and the threonine-isoleucine motif at positions 13-14 (18/51; 35%).

When patterns were extended to include three amino acids, the most frequent tripeptide motif was found to be a phenylalanine-glycine-valine tripeptide (25/51; 49%) followed by a glycine-valine-valine tripeptide (19/51; 37%). Both of these tripeptides stemmed from the aforementioned "VV" and "GV" dipeptide motifs. Thus, our search for common motifs within the VH CDR3 revealed glycine at position 16 and valine at position 17 in our aligned VH CDR3 sequences to be the "parent" amino acids of the most common patterns within IGHV1-2*04 SMZL cases (Supplementary Fig. S1).

Subclonal architecture and temporal dynamics in IGHV1-2*04 rearrangements

The pronounced ID within IGHV1-2*04 rearrangements was also evidenced by the presence of distinct "clusters" of subcloned sequences with "cluster-specific" SHM profiles. Analysis of the SHM patterns in such clusters revealed their common ancestry, but also indicated an early "branching" of the SMZL clone into distinct subclones, perhaps able to evolve along separate pathways through the stepwise accumulation of SHMs. This phenomenon was evidenced in 8 of 21 IGHV1-2*04 rearrangements, whereas it was not observed in rearrangements utilizing other IGHV genes. Within these latter cases, the introduced SHMs were mostly ubiquitous (i.e., present in all subcloned sequences).

The overtime dynamics of ID within the clonotypic IGHs in SMZL was assessed by longitudinal analysis of three cases in samples collected over a period spanning from 2 to 10 years; two cases utilized the IGHV1-2*04 gene, whereas the third case expressed the IGHV3-23 gene. Ongoing SHM was observed within both IGHV1-2*04 rearrangements but not in the IGHV3-23 case nor in the partner light chain of any of these three cases. Of particular interest was the pattern observed in one of the two IGHV1-2*04 rearrangements in which the subcloned sequences from the first timepoint branched into two subgroups each characterized by a "trademark" mutation. This early branching was maintained in the subcloned sequences from the second time-point; however,

one of the two groups of subcloned sequences was further branched into three smaller groups by two additional mutations acting as the branching nodes, indicating evolution via a distinct pathway (Fig. 4).

Discussion

Immunogenetic and functional evidence supports a role for antigen in the natural history of SMZL. However, the timing and duration of antigen interactions and their relevance for the evolution of the disease remain elusive. Imprints of ID in the IGH gene rearrangements have been previously reported and interpreted in the context of ongoing SHM (34–37). However these studies concerned very few cases, thus hindering definitive conclusions.

In order to address the question as to whether antigen is implicated only in the process of selecting SMZL progenitors or if it is also relevant post-transformation, we conducted a comprehensive subcloning analysis in a series of 40 SMZL cases, by far the largest ever studied for this purpose. Our study was intentionally biased toward cases utilizing the IGHV1-2*04, not only due to its overexpression in SMZL (18), but also due to increasing evidence suggesting that IGHV1-2*04 gene usage defines a particular subtype of SMZL displaying a distinctive clinicobiologic profile with high frequency of both deletion of chromosome 7(3) and *KLF2* mutations (12, 13) as well as inferior outcome compared with non-IGHV1-2*04 cases (38). That said, cases utilizing other IGHV genes were also included in the study, thus offering a comprehensive view of ID across various immunogenetic subgroups of SMZL.

We confirmed ID within the majority of samples and gene groups analyzed, thus supporting the notion that the SHM mechanism remains active post-transformation, further diversifying the clonotypic IGH receptors. Importantly, ID was significantly more pronounced in cases carrying IGHV1-2*04 receptors, thus further supporting our previous hypothesis that such cases constitute a unique molecular subgroup of SMZL. Admittedly, the extent of ID differed between IGHV1-2*04 cases, as reflected in different NMF values, which might be related to the fact that these cases differ

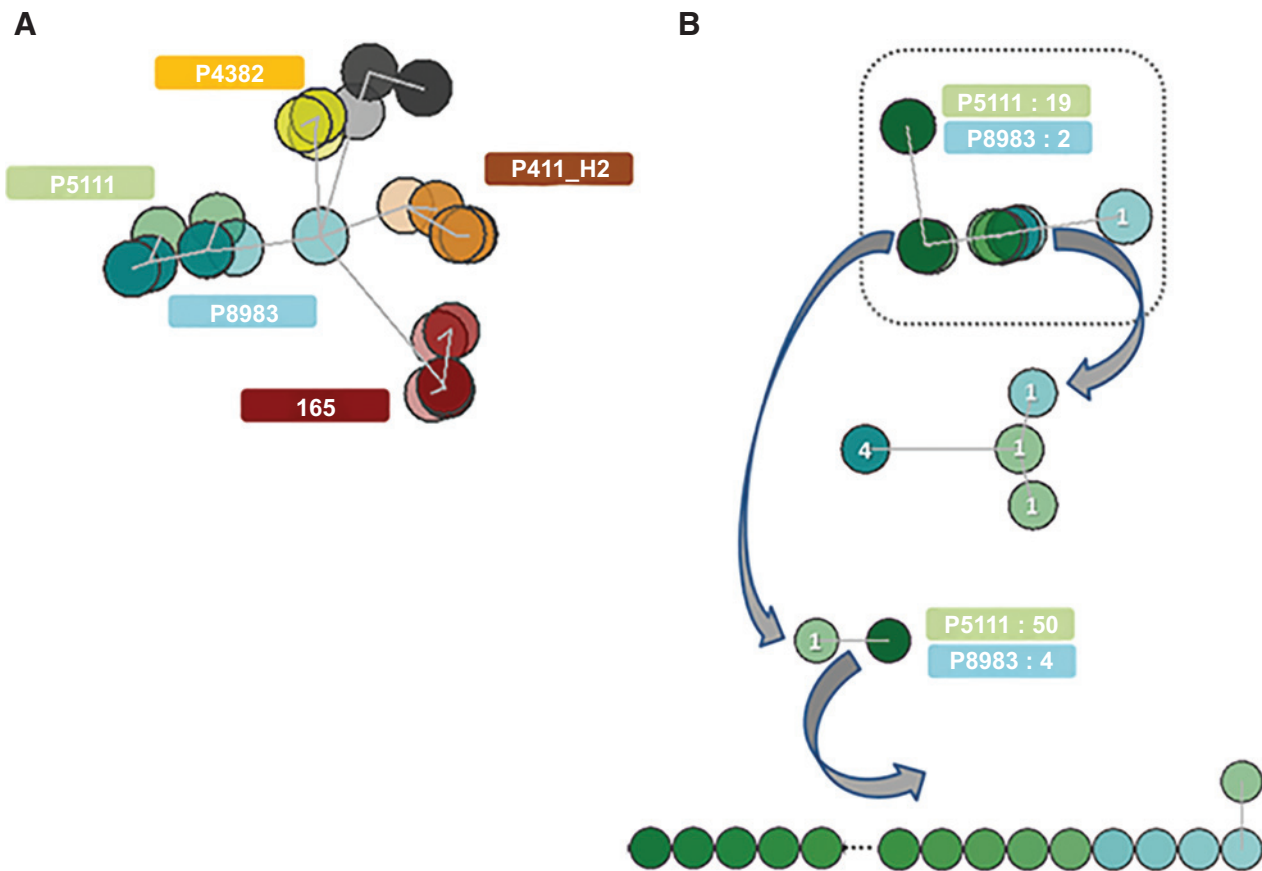


Figure 3. Sequence relationship network for four cases utilizing the IGHV1-2*04 allele. The network was created using an in-house, purpose-built algorithm based on the calculation of distances between aligned antigen receptor sequences considering the variability in their amino acid content at each position (Damerau-Levenshtein method). Spheres represent distinct subclone sequences and are color coded to match the patient tag. Different shades of the same color indicate the presence of more than one sequences of the same case. Figure 3A illustrates cluster formation following analysis of the IGHV-IGHD-IGHJ subcloned amino acid sequences from six cases ($n = 152$). Four distinct clusters were observed; subclones from two cases appear to develop a unique mutational profile, while two cores were created by clonal sequences from two cases, P5111/P8983 and P4382/LS32, respectively. A single subclone from the P5111/P8983 stands out and appears as connecting the four clusters. Figure 3B provides a more detailed view of the latter core. The core is framed by dotted lines and each node is then further dissected. Subclone sequences from each of the two patients appear to be present at each node. Tangent spheres at the bottom of Fig. 3B represent identical subclone sequences.

regarding both their VH CDR3 and partner IG light chains. These parameters are critical for the overall shape and antigen affinity of any IG and have been found to affect the responses to ongoing antigenic stimulation, leading to distinct SHM profiles (28, 39).

Affinity maturation through SHM is evolutionarily optimized so that the germline sequences of the IG genes should be able to enhance their stability and viability in the presence of random mutations (40). In other words, SHM topology is influenced not only by functional constraints, but also by the existence of mutable microsequences inside the variable domain, which affect both the introduction of SHM and the operation of repair mechanisms (41). IGHV1-2 polymorphic variants are a prime example since their primary sequence renders them highly mutable (42), perhaps providing an explanation for the increased occurrence of ID in IGHV1-2*04 rearrangements. Combined with the remarkable bias toward one of its polymorphic variants in the SMZL repertoire, namely allele *04, this could imply selection by a set of common epitopes and further diversification through SHM.

Along these lines, distinctive molecular features of ID were observed within the IGHV1-2*04 clonotypic receptors: (i) the highest ratio of CM/UCM compared with all other IGHV gene subgroups studied; (ii) G-to-C transversion mutations exclusively located within the FRs, which is notable given that these transversions are usually silent or lead to the introduction of highly conservative AA substitutions (33) as in this case; (iii) localization of CMs outside established hotspot motifs (27) either in the CDRs or the FRs. Taken together, this molecular profile is highly suggestive of selection driven by specific antigen(s) imposing functional or structural constraints.

In SMZL, IGHV1-2*04 gene rearrangements are characterized by biased usage of certain IGHD genes and long VH CDR3s enriched in hydrophobic residues. Regarding this latter feature, either a valine-valine (V/V) or valine-isoleucine (V/I) dipeptide motif, encoded by either the IGHD3-3 or the IGHD3-10 gene, was present at the tip of the VH CDR3 in 11/14 IGHV1-2*04 rearrangements (Supplementary Fig. S1) and remained unaltered in all subcloned sequences despite ID within the VH CDR3, again

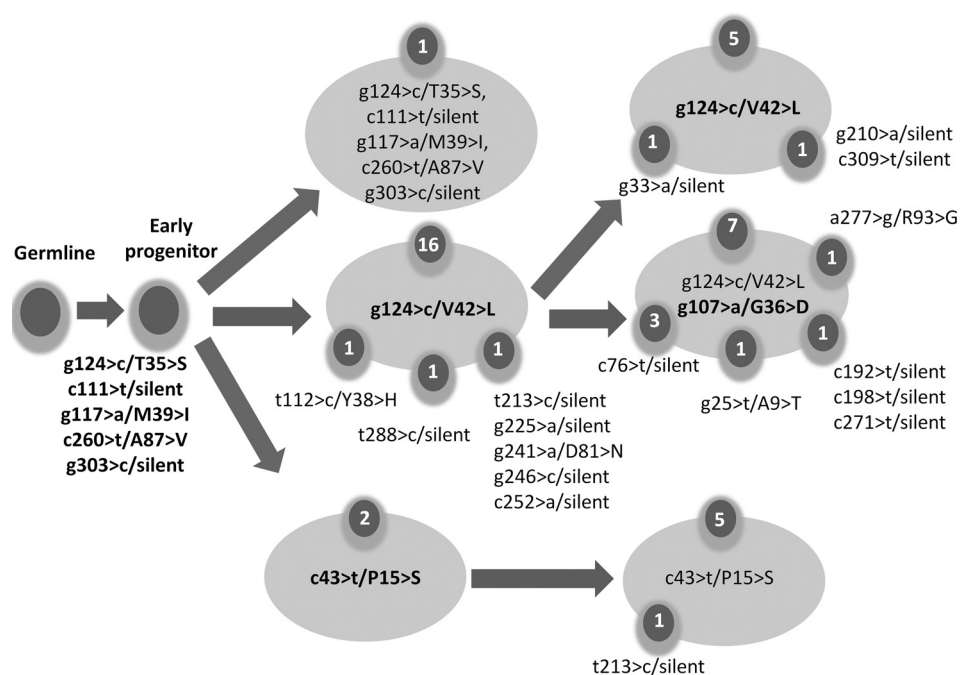


Figure 4. Dynamic patterns of overtime intracлонаl diversification in an IGHV1-2*04 case. Clonal evolution is evidenced since the subcloned sequences of the same case analyzed at different timepoints possess the same ubiquitous mutations and are divided into distinct subgroups based on the presence/lack of "trademark" mutations. Members of each group may also have additional mutations, mainly UCMs.

strongly indicative of selection. This latter finding, that is pronounced hydrophobicity, is also reminiscent of poly/auto-reactive antibodies and is in keeping with recent reports regarding the polyreactivity of IGHV1-2*04 monoclonal antibodies from patients with SMZL (43). Along these lines, especially noteworthy was the observed high prevalence of the conservative replacement of Valine (V) to Leucine/Isoleucine (L/I) at position VH FR2-39, alluding to functional constraint.

How should one interpret these findings? An important point to consider is that despite the VH CDR3 being the principal determinant of specificity, at least within the primary repertoire, VH CDR3 diversity itself is not sufficient to realize the full potential of antibody diversity (44, 45). Furthermore, unconventional antigens, such as B-cell superantigens, may be recognized through critical residues outside the VH CDR3, also located within the FRs (46–48). From this perspective, the molecular features of ID within the IGHV1-2*04 clonotypic receptors of patients with SMZL seem to go along with the concept of superantigenic interactions within this subgroup of patients, while also underscoring the importance of selecting this particular germline specificity in the SMZL clonogenic progenitor cell repertoire. This argument is further supported by the finding of a lower level of ID in the IG light chain genes of cases with paired IG heavy/light gene analysis, perhaps reflecting IG heavy chain dominance (49, 50).

In order to obtain more insight into the temporal dynamics of SHM in general and the ID process in particular, we performed longitudinal analysis of three SMZL cases, of which two expressed IGHV1-2*04 antigen receptors. In these latter cases, ID resulted in a "branching" of the subcloned sequences which was evidenced even at the first timepoint. Additional SHM was observed in each major branch at the second timepoint, raising the possibility that different subclones may evolve across related yet distinct immune pathways.

In conclusion, our present immunogenetic findings seem to justify the claim that IGHV1-2*04 cases constitute a distinct molecular subtype of SMZL, possibly deriving from different progenitor cells compared with non-IGHV1-2*04 cases and/or also having distinct antigen exposure histories.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: D. Tzovaras, A. Hadzidimitriou, K. Stamatopoulos
Development of methodology: M. Karypidou, E. Stalika, L.-A. Sutton, G. Papadopoulos, E. Papadopoulou, N. Darzentas, D. Tzovaras
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): P. Baliakas, Z. Davis, P. Algara, G. Kanellis, A. Traverse-Glehen, A. Anagnostopoulos, M. Ponzoni, S. Pospisilova, E. Matutes, M.A. Piris, P. Chia, D. Oscier, C. Belessi
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): V. Bikos, P. Baliakas, A. Xochelli, L.-A. Sutton, G. Papadopoulos, A. Agathangelidis, E. Papadopoulou, D. Gonzalez, R. Rosenquist, D. Tzovaras
Writing, review, and/or revision of the manuscript: V. Bikos, P. Baliakas, A. Xochelli, L.-A. Sutton, G. Papadopoulos, E. Papadopoulou, A. Anagnostopoulos, M. Ponzoni, S. Pospisilova, M.A. Piris, T. Papadaki, P. Chia, R. Rosenquist, D. Oscier, K. Stamatopoulos
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Karypidou, A. Agathangelidis, Z. Davis, D. Tzovaras
Study supervision: A. Hadzidimitriou, K. Stamatopoulos
Other (provided samples): M. Mollejo
Other (contributed to diagnosis of SMZL): E. Matutes

Grant Support

This work was supported in part by the ENosAI project (code 09SYN-13-880) cofunded by the EU and the General Secretariat for Research and Technology of Greece; the KRIPIS action, funded by the General Secretariat for Research and Technology of Greece; the EU Seventh Framework Programme under the "Capacities" specific programme; H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe", by the EU; the Swedish Cancer

Society, the Swedish Research Council, the Lion's Cancer Research Foundation, and Selander's Foundation, Uppsala; Associazione Italiana per la Ricerca sul Cancro AIRC (Investigator Grant and Special Program Molecular Clinical Oncology – 5 per mille #9965), Milano, Italy and Ricerca Finalizzata 2010 – Ministero della Salute, Roma. V. Bikos was supported by the EU Seventh Framework Programme under the "Capacities" specific program (Contract No. 286154 – SYLICA). M. Karypidou is recipient of a scholarship from the Hellenic Association for Molecular Cancer Research. V. Bikos, S. Pospisilova, and N. Darzentas were supported by the EU Seventh Framework Programme under the "Capacities"

specific program research and projects CEITEC MU (CZ.1.05/1.1.00/02.0068) and SuPREMMe (CZ.1.07/2.3.00/20.0045).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 18, 2015; revised October 13, 2015; accepted November 19, 2015; published OnlineFirst December 8, 2015.

References

1. Swerdlow S, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al. (eds). WHO Classification of Tumors of Haemopoietic and Lymphoid Tissues. Lyon, France: IARC Press; 2008.
2. Novara F, Arcaini L, Merli M, Passamonti F, Zibellini S, Rizzi S, et al. High-resolution genome-wide array comparative genomic hybridization in splenic marginal zone B-cell lymphoma. *Hum Pathol* 2009;40:1628–37.
3. Salido M, Baro C, Oscier D, Stamatopoulos K, Dierlamm J, Matutes E, et al. Cytogenetic aberrations and their prognostic value in a series of 330 splenic marginal zone B-cell lymphomas: a multicenter study of the Splenic B-Cell Lymphoma Group. *Blood* 2010;116:1479–88.
4. Rossi D, Deaglio S, Dominguez-Sola D, Rasi S, Vaisitti T, Agostinelli C, et al. Alteration of BIRC3 and multiple other NF-kappaB pathway genes in splenic marginal zone lymphoma. *Blood* 2011;118:4930–4.
5. Rossi D, Trifonov V, Fangazio M, Brusca A, Rasi S, Spina V, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med* 2012;209:1537–51.
6. Fresquet V, Robles EF, Parker A, Martinez-Useros J, Mena M, Malumbres R, et al. High-throughput sequencing analysis of the chromosome 7q32 deletion reveals IRF5 as a potential tumour suppressor in splenic marginal-zone lymphoma. *Br J Haematol* 2012;158:712–26.
7. Watkins AJ, Hamoudi RA, Zeng N, Yan Q, Huang Y, Liu H, et al. An integrated genomic and expression analysis of 7q deletion in splenic marginal zone lymphoma. *PLoS ONE* 2012;7:e44997.
8. Yan Q, Huang Y, Watkins AJ, Kocalkowski S, Zeng N, Hamoudi RA, et al. BCR and TLR signaling pathways are recurrently targeted by genetic changes in splenic marginal zone lymphomas. *Haematologica* 2012;97:595–8.
9. Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG, Chiang MY, et al. Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *J Exp Med* 2012;209:1553–65.
10. Martinez N, Almaraz C, Vaque JP, Varela I, Derdak S, Beltran S, et al. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. *Leukemia* 2014;28:1334–40.
11. Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, Forster J, et al. Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. *PLoS ONE* 2013;8:e83244.
12. Arribas AJ, Rinaldi A, Mensah AA, Kwee I, Cascione L, Robles EF, et al. DNA methylation profiling identifies two splenic marginal zone lymphoma subgroups with different clinical and genetic features. *Blood* 2015;125:1922–31.
13. Clipsas A, Wang M, de Leval L, Ashton-Key M, Wotherspoon A, Vassiliou G, et al. KLF2 mutation is the most frequent somatic change in splenic marginal zone lymphoma and identifies a subset with distinct genotype. *Leukemia* 2015;29:1177–85.
14. Arcaini L, Zibellini S, Passamonti F, Rattotti S, Lucioni M, Invernizzi R, et al. Splenic marginal zone lymphoma: clinical clustering of immunoglobulin heavy chain repertoires. *Blood Cells Mol Dis* 2009;42:286–91.
15. Hockley SL, Giannouli S, Morilla A, Wotherspoon A, Morgan GJ, Matutes E, et al. Insight into the molecular pathogenesis of hairy cell leukaemia, hairy cell leukaemia variant and splenic marginal zone lymphoma, provided by the analysis of their IGH rearrangements and somatic hypermutation patterns. *Br J Haematol* 2010;148:666–9.
16. Traverse-Glehen A, Davi F, BenSimon E, Callet-Bauchu E, Felman P, Basseggio L, et al. Analysis of VH genes in marginal zone lymphoma reveals marked heterogeneity between splenic and nodal tumors and suggests the existence of clonal selection. *Haematologica* 2005;90:470–8.
17. Zibellini S, Capello D, Forconi F, Marcatili P, Rossi D, Rattotti S, et al. Stereotyped patterns of B-cell receptor in splenic marginal zone lymphoma. *Haematologica* 2010;95:1792–6.
18. Bikos V, Darzentas N, Hadzidimitriou A, Davis Z, Hockley S, Traverse-Glehen A, et al. Over 30% of patients with splenic marginal zone lymphoma express the same immunoglobulin heavy variable gene: ontogenetic implications. *Leukemia* 2012;26:1638–46.
19. 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;119:4467–75.
20. Hadzidimitriou A, Agathangelidis A, Darzentas N, Murray F, Delfau-Larue MH, Pedersen LB, et al. Is there a role for antigen selection in mantle cell lymphoma? Immunogenetic support from a series of 807 cases. *Blood* 2011;118:3088–95.
21. Bikos V, Stalika E, Baliakas P, Darzentas N, Davis Z, Traverse-Glehen A, et al. Selection of antigen receptors in splenic marginal-zone lymphoma: further support from the analysis of the immunoglobulin light-chain gene repertoire. *Leukemia* 2012;26:2567–9.
22. Matutes E, Oscier D, Montalban C, Berger F, Callet-Bauchu E, Dogan A, et al. Splenic marginal zone lymphoma proposals for a revision of diagnostic, staging and therapeutic criteria. *Leukemia* 2008;22:487–95.
23. Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res* 2015;43:D413–22.
24. Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, et al. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res* 2006;34:D781–4.
25. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* 2008;36:W503–8.
26. Giudicelli V, Brochet X, Lefranc MP. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. *Cold Spring Harb Protoc* 2011;2011:695–715.
27. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta* 1992;1171:11–8.
28. Sutton LA, Kostareli E, Hadzidimitriou A, Darzentas N, Tsiftaris 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;114:4460–8.
29. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
30. Jakobsson M. Compression of character strings by an adaptive dictionary. *BIT* 1985;25:593–603.
31. Ziv J, Lempel A. A universal algorithm for sequential data compression. *Inf Theory* 1977;23:337–43.
32. Rodeh M, Pratt VR, Even S. Linear algorithm for data compression via string matching. *J ACM* 1981;28:16–24.
33. Zheng NY, Wilson K, Jared M, Wilson PC. Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation. *The J Exp Med* 2005;201:1467–78.
34. Dunn-Walters DK, Boursier L, Spencer J, Isaacson PG. Analysis of immunoglobulin genes in splenic marginal zone lymphoma suggests ongoing mutation. *Hum Pathol* 1998;29:585–93.

35. Zhu D, Orchard J, Oscier DG, Wright DH, Stevenson FK. V(H) gene analysis of splenic marginal zone lymphomas reveals diversity in mutational status and initiation of somatic mutation *in vivo*. *Blood* 2002;100:2659–61.
36. Tierens A, Delabie J, Malecka A, Wang J, Gruszka-Westwood A, Catovsky D, et al. Splenic marginal zone lymphoma with villous lymphocytes shows on-going immunoglobulin gene mutations. *Am J Pathol* 2003;162:681–9.
37. Tracey L, Aggarwal M, Garcia-Cosio M, Villuendas R, Algara P, Sanchez-Beato M, et al. Somatic hypermutation signature in B-cell low-grade lymphomas. *Haematologica* 2008;93:1186–94.
38. Rose-Zerilli MJ, Parry M, Ljungstrom V, Gibson J, Wang J, Walewska R, et al. Deep-sequencing reveals the molecular landscape of splenic marginal zone lymphoma: biological and clinical implications. *Blood* 2014;124:76.
39. 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;24:1317–24.
40. Hershberg U, Shlomchik MJ. Differences in potential for amino acid change after mutation reveals distinct strategies for kappa and lambda light-chain variation. *Proc Natl Acad Sci U S A* 2006;103:15963–8.
41. Cohen RM, Kleinstein SH, Louzoun Y. Somatic hypermutation targeting is influenced by location within the immunoglobulin V region. *Mol Immunol* 2011;48:1477–83.
42. Shapiro GS, Aviszus K, Murphy J, Wysocki LJ. Evolution of Ig DNA sequence to target specific base positions within codons for somatic hypermutation. *J Immunol* 2002;168:2302–6.
43. Warsame AA, Aasheim HC, Nustad K, Troen G, Tierens A, Wang V, et al. Splenic marginal zone lymphoma with VH1-02 gene rearrangement expresses poly- and self-reactive antibodies with similar reactivity. *Blood* 2011;118:3331–9.
44. Davies DR, Cohen GH. Interactions of protein antigens with antibodies. *Proc Natl Acad Sci U S A* 1996;93:7–12.
45. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 2000;13:37–45.
46. Silverman GJ, Goodyear CS. A model B-cell superantigen and the immunobiology of B lymphocytes. *Clin Immunol* 2002;102:117–34.
47. Silverman GJ, Goodyear CS. Confounding B-cell defences: lessons from a staphylococcal superantigen. *Nat Rev Immunol* 2006;6:465–75.
48. Ekiert DC, Bhabha G, Elsliger MA, Friesen RH, Jongeneelen M, Throsby M, et al. Antibody recognition of a highly conserved influenza virus epitope. *Science* 2009;324:246–51.
49. Jang YJ, Stollar BD. Anti-DNA antibodies: aspects of structure and pathogenicity. *Cell Mol Life Sci* 2003;60:309–20.
50. Li H, Jiang Y, Prak EL, Radic M, Weigert M. Editors and editing of anti-DNA receptors. *Immunity* 2001;15:947–57.
51. Beitz E. TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics* 2000;16:135–9.