

# blood

2009 114: 4460-4468  
Prepublished online Aug 27, 2009;  
doi:10.1182/blood-2009-05-221309

## **Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen**

Lesley-Ann Sutton, Efterpi Kostareli, Anastasia Hadzidimitriou, Nikos Darzentas, Athanasios Tsaftaris, Achilles Anagnostopoulos, Richard Rosenquist and Kostas Stamatopoulos

---

Updated information and services can be found at:  
<http://bloodjournal.hematologylibrary.org/cgi/content/full/114/20/4460>

Articles on similar topics may be found in the following *Blood* collections:  
[Lymphoid Neoplasia](#) (520 articles)

---

Information about reproducing this article in parts or in its entirety may be found online at:  
[http://bloodjournal.hematologylibrary.org/misc/rights.dtl#repub\\_requests](http://bloodjournal.hematologylibrary.org/misc/rights.dtl#repub_requests)

Information about ordering reprints may be found online at:  
<http://bloodjournal.hematologylibrary.org/misc/rights.dtl#reprints>

Information about subscriptions and ASH membership may be found online at:  
<http://bloodjournal.hematologylibrary.org/subscriptions/index.dtl>



# Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen

\*Lesley-Ann Sutton,<sup>1</sup> \*Efterpi Kostareli,<sup>2,3</sup> Anastasia Hadzidimitriou,<sup>4</sup> Nikos Darzentas,<sup>4</sup> Athanasios Tsaftaris,<sup>4</sup> Achilles Anagnostopoulos,<sup>3</sup> Richard Rosenquist,<sup>1</sup> and Kostas Stamatopoulos<sup>3</sup>

<sup>1</sup>Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden; <sup>2</sup>Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece; <sup>3</sup>Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece; and <sup>4</sup>Institute of Agrobiotechnology, Centre for Research and Technology, Thessaloniki, Greece

Several studies indicate that the development of chronic lymphocytic leukemia (CLL) may be influenced by antigen recognition through the clonotypic B-cell receptors (BCRs). However, it is still unclear whether antigen involvement is restricted to the malignant transformation phase or whether the putative antigen(s) may continuously trigger the CLL clone and affect not only the progenitor cell but also the leukemic cells themselves. To address this issue, we conducted a large-scale

subcloning study of rearranged immunoglobulin heavy variable (IGHV) genes of diverse mutational status from 71 CLL cases (total, 1496 subcloned sequences), belonging to both the common IgM/IgD variant and the rare IgG-positive variant. Although most cases showed no or low levels of intraclonal diversification (ID), we report intense ID in the IGHV genes of selected cases, especially a subgroup of 13 IgG-switched cases expressing stereotyped, mutated IGHV4-34 rearrangements

(subset 4). We demonstrate that the ID evident in subset 4 cases cannot be attributed to IGHV4-34 usage, IGHV gene-mutated status, class-switch recombination, or BCR stereotypy in general; rather, it represents a unique phenomenon strongly correlated with the distinctive BCR of subset 4. In such cases, the observed ID patterns may imply a stereotyped response to an active, ongoing interaction with antigen(s). (Blood. 2009; 114:4460-4468)

## Introduction

Over the last decade, immunophenotypic and gene expression profiling studies have provided conclusive evidence that chronic lymphocytic leukemia (CLL) malignant B cells are antigen-experienced.<sup>1-3</sup> A strong sequence-based argumentation supporting the role of antigen in CLL was the finding that somatic hypermutations (SHMs) are present in the immunoglobulin heavy variable (IGHV) genes expressed by the leukemic B cells in more than 50% of CLL cases.<sup>4,5</sup> The presence or absence of mutations was shown to have important clinical implications and is now widely used to segregate CLL cases into 2 clinical entities displaying significantly different outcomes.<sup>6,7</sup>

Further evidence for the role of antigen in leukemogenesis is the remarkably restricted IGHV gene repertoire of CLL, with just a handful of genes predominating (eg, *IGHV1-69*, *IGHV4-34*, *IGHV3-7*, *IGHV3-23*, and *IGHV3-21*).<sup>5,8,9</sup> From a sequence perspective, however, the most striking argument for the active role of antigen in CLL is the existence of subsets of geographically distant and unrelated cases with quasi-identical, "stereotyped" B cell receptors (BCRs).<sup>10-18</sup> Interestingly, cases expressing certain stereotyped BCRs may also share other biologic and clinical features. For instance, the *IGHV3-21/IGLV3-21* subset (known as subset 2)<sup>17</sup> has been associated with a poor prognosis irrespective of IGHV gene mutational status.<sup>16,17,19</sup> In contrast, the *IGHV4-34/IGKV2-30* subset (known as subset 4)<sup>17</sup> appears to be associated with an indolent

course of the disease.<sup>17</sup> Furthermore, cases belonging to this subset are relatively young<sup>17</sup> and uniformly express IgG-switched IGs.<sup>17,18</sup>

We recently demonstrated the existence of more than 100 different subsets with stereotyped heavy-chain complementarity-determining region 3 (HCDR3) sequences in a series of 1939 patients with CLL.<sup>18</sup> A comprehensive analysis of SHMs in the same series revealed "CLL-biased" SHM patterns typified by recurrent hypermutations throughout the entire IGHV region in subsets of cases with stereotyped HCDR3s, in particular those expressing stereotyped *IGHV3-21/IGLV3-21* IGs (subset 2) and *IGHV4-34/IGKV2-30* IGs (subset 4).<sup>18</sup>

Although the study of IG genes in CLL has strongly implicated an antigen-driven model of CLL development, an as yet unanswered question concerns the B-cell developmental stage when leukemic transformation occurred.<sup>20,21</sup> Furthermore, it is still unclear whether antigen involvement is restricted to the malignant transformation phase or whether the putative antigen(s) may continuously trigger the CLL clone and affect not only the progenitor cell but also the leukemic cells themselves.<sup>20,21</sup>

Evidence from several other types of B-cell lymphoproliferative disorders suggests that valuable insight into these issues may be gleaned from the study of intraclonal diversification (ID) within IG genes through ongoing mutational activity. Evidence of ID pointing to ongoing SHM has been observed in several types of B-cell malignancies,<sup>22-28</sup> thereby indicating that the neoplastic cells

Submitted May 10, 2009; accepted August 20, 2009. Prepublished online as *Blood* First Edition paper, August 27, 2009; DOI 10.1182/blood-2009-05-221309.

\*L.-A.S. and E.K. contributed equally to this study as first authors.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2009 by The American Society of Hematology

interact continuously with their cognate antigen(s). On the contrary, multiple myeloma is conspicuous for a complete lack of ID within IG genes, supporting the notion that malignant transformation probably occurs at a post-germinal center (GC) stage of B-cell development.<sup>29,30</sup>

The available data on ID within IG genes in CLL are limited and conflicting. Several studies demonstrate some level of ID (ranging from limited to pronounced) leading to clonal evolution,<sup>31-36</sup> with only one study reporting a complete lack of ID.<sup>37</sup> We conducted the present study to address these discrepancies and systematically explore the issue as to whether CLL cells continue to acquire hypermutations after leukemic transformation. To this end, we followed a very stringent methodology and conducted a comprehensive analysis of ID in IGHV genes of diverse mutational status from 71 patients with CLL, belonging to the common IgM/IgD variant and the rare IgG-positive variant. Although the majority of cases showed no or low levels of ID, an intense and functionally driven ID process was evident in selected cases, especially those belonging to subset 4, who express highly distinctive, stereotyped *IGHV4-34/IGKV2-30* BCRs.<sup>17,18</sup> In such cases, the observed ID patterns among sets of subcloned IG sequences attest to the very precise targeting of the SHM process and may be considered as evidence for a “stereotyped response” to an active, ongoing interaction with antigen(s).

## Methods

### Patient group

Seventy-one patients with CLL from collaborating institutions in Scandinavia and Greece were included in the study. The main criteria for case selection were IGHV gene usage and mutational status. Thus, our study is intentionally biased for cases using the *IGHV3-21* and *IGHV4-34* genes (25 of 71 and 28 of 71 cases, respectively). All cases were immunophenotyped as described previously<sup>10,15,17,18</sup> and met the recently revised diagnostic criteria of the National Cancer Institute Working Group.<sup>38</sup> Written informed consent was obtained according to the Declaration of Helsinki, and the study was approved by the local ethics review committee of each institution.

### Polymerase chain reaction amplification of rearranged IGH genes

Polymerase chain reaction (PCR) amplification was performed on either genomic DNA (gDNA) or complementary DNA (cDNA), extracted mainly from blood and bone marrow (63 cases), but also from spleen and lymph nodes in 1 and 7 cases, respectively. Reverse transcriptase (RT)-PCR or gDNA-PCR amplification of IGHV-IGHD-IGHJ rearrangements was performed using consensus primers for the HFR1 and IGHJ genes or the appropriate sense IGHV leader primer and the antisense IGHC primer (supplemental Table 1, available on the *Blood* website; see the Supplemental Materials link at the top of the online article), as previously described.<sup>10,15,17,18</sup> All amplification reactions were run using the high-fidelity Accuprime Pfx polymerase (Invitrogen). Purified PCR amplicons were subjected to direct sequencing, and data obtained were analyzed using the ImMunoGeneTics (IMGT) database and tools.<sup>39,40</sup>

### Subcloning

PCR amplification products were gel-purified with the QIAGEN DNA purification columns (QIAGEN), ligated into the pCR2.1 vector (Invitrogen), and subsequently transformed into *Escherichia coli*/TOP10F' competent bacteria (Invitrogen). A range of 14 to 35 colonies per case (median, 21 colonies) were chosen randomly and sequenced using the -20 universal primer or M13 primers.

### Sequence data analysis

The sequences obtained from subcloning were analyzed with the IMGT V-QUEST<sup>40</sup> and ClustalW/EMBL<sup>41</sup> tools. To avoid misidentification of mutations when HFR1 consensus primers were used in the amplification reactions, nucleotide changes in the obtained sequences were evaluated from codon 24 in IMGT-HFR1 to the end of HCDR3. For the subgroup of cases amplified using leader primers and a constant region primer, the entire V region was evaluated (from IMGT HFR1 codon 1 down to the end of HFR4). The following information was extracted: (1) IGHV gene usage, percentage of identity to germline, and HCDR3 length; (2) SHM characteristics: 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 “nonconservative” following standardized biochemical criteria, as previously described.<sup>42</sup> To account for the fact that a mutation is more likely to occur in an HFR than an HCDR simply because of its greater length, each mutation was “weighted,” or normalized, as recently reported by our group<sup>18</sup>; (3) hotspot targeting: mutated sequences were also analyzed for targeting to the tetranucleotide (4-NTP) motifs RGYW/WRCY (R = A/G, Y = C/T, and W = A/T) and DGYW/WRCH (D = A/G/T, H = T/C/A) as well as the dinucleotide (2-NTP) motifs WA and TW.<sup>43,44</sup>

### Definitions

Intraclonal heterogeneity in sets of subcloned sequences obtained from the same sample was assessed by examination of the sequence variation in the V domain. All “nonubiquitous” sequence changes from the germline were evaluated and further characterized as follows: (1) unconfirmed mutation (UCM), a mutation observed in only one subcloned sequence from the same specimen (“unique”); and (2) confirmed mutation (CM), a mutation observed more than once among subcloned sequences from the same specimen (“partially shared”). AA changes resulting from UCMs or CMs are designated by the abbreviations UAA or CAA, respectively.

To compare mutation counts between the different rearrangements included in the analysis, mutations were normalized to both the nucleotide length and the number of subcloned sequences for each rearrangement. Thus, the normalized mutation frequencies (NMFs) were calculated according to the following formula:  $\sum (CM + UCM)/\text{number of subcloned sequences} \times \text{sequence length}$ ; for instance, if 9 mutations (confirmed and/or unconfirmed) were identified in a set of 20 subcloned sequences of a certain rearrangement over a region of 300 nucleotides, the NMF for this rearrangement would be:  $9/(20 \times 300) = 1.5 \times 10^{-3}$ .

### Evolutionary history of sets of subcloned sequences

The evolutionary history of the sets of subcloned sequences was inferred using the maximum parsimony method.<sup>45</sup> The maximum parsimony trees were obtained using the close-neighbor-interchange algorithm with search level 2 in which the initial trees were obtained with the random addition of sequences (10 replicates).<sup>45</sup> Branch lengths were calculated using the average pathway method<sup>45</sup> and are in the units of the number of changes over the whole sequence. They are shown next to the branches. The tree is drawn to scale. All positions containing gaps and missing data were eliminated from the dataset (Complete Deletion option). Phylogenetic analyses were conducted in MEGA4.<sup>45</sup>

### Statistical analysis

Descriptive statistics for discrete parameters included counts and frequency distributions. For quantitative variables, statistical measures included means, medians, SD, and ranges. Significance of bivariate relationships between factors was assessed with  $\chi^2$  and Fisher exact tests. For all comparisons, a significance level of  $P = .05$  was set, and all statistical analyses were performed with the Statistical Package SPSS, Version 12.0 (SPSS Inc).

**Table 1. Intracлонаl diversification analysis of IGHV-D-J rearrangements from cases in subset 4 with stereotyped IGHV4-34/IGKV2-30 BCRs**

Case no.	Identity, %	CM	UCM	NMF	C-AA	U-AA
Swe56	95.1	27	11	$3.05 \times 10^{-3}$	9	7
Swe181	94.3	4	1	$0.94 \times 10^{-3}$	1	0
Swe193	93.1	6	4	$1.65 \times 10^{-3}$	3	4
P0103	95.9	32	3	$2.89 \times 10^{-3}$	11	2
P0907	93.2	6	15	$1.68 \times 10^{-3}$	3	10
P1422	91.9	52	12	$6.27 \times 10^{-3}$	22	7
P1939	94.5	3	9	$0.91 \times 10^{-3}$	1	5
P2451	91.7	39	27	$4.99 \times 10^{-3}$	14	15
P2920	93.5	24	19	$4.38 \times 10^{-3}$	16	14
P3020	90	1	4	$0.83 \times 10^{-3}$	1	4
P3551	93.3	52	3	$7.28 \times 10^{-3}$	25	3
P3916	91.2	3	3	$0.93 \times 10^{-3}$	1	1
P6520	94.4	1	8	$1.08 \times 10^{-3}$	1	6

All cases shown had confirmed intracлонаl diversification (CID).

CM indicates confirmed mutation; UCM, unconfirmed mutation; NMF, normalized mutation frequency; C-AA, confirmed amino acid change; and U-AA, unconfirmed amino acid change.

## Results

### IGHV repertoire and mutational status

Productive IGHV-IGHD-IGHJ rearrangements from 71 CLL cases were included in this analysis. Detailed information on IGHV, IGHD, and IGHJ gene repertoires is provided in supplemental Table 1. Overall, 7 different IGHV genes were used: *IGHV1-69*, 6 of 71 cases; *IGHV3-21*, 25 of 71 cases; *IGHV3-23*, 5 of 71 cases; *IGHV3-30*, one case; *IGHV3-7*, 2 of 71 cases; *IGHV4-34*, 28 of 71 cases; and *IGHV4-39*, 4 of 71 cases. Sequences were categorized according to IGHV gene mutational status following definitions recently reported by our group<sup>18</sup>: (1) “truly unmutated” (100% germline identity): 16 of 71 sequences (22.5%); (2) “minimally mutated” (99.0%-99.9% identity): 6 of 71 sequences (8.5%); (3) “borderline mutated” (98.0%-98.9% identity): 5 of 71 sequences (7%); and (4) “mutated” (< 98% identity): 44 of 71 sequences (62%).

Forty-six of 71 cases expressed stereotyped BCRs and were assigned to 11 different subsets, as previously described (supplemental Table 1). More specifically, 16 cases expressed *IGHV3-21/IGLV3-21* BCRs (subset 2),<sup>17,18</sup> 13 of 71 cases expressed *IGHV4-34/IGKV2-30* BCRs (subset 4),<sup>17,18</sup> 4 of 71 cases expressed *IGHV4-39/IGKV1(D)-39* BCRs (subset 8),<sup>17,18</sup> whereas 5 of 71 cases expressed *IGHV4-34/IGKV3-20* (subset 16).<sup>17,18</sup> Notably, subset 4 and subset 16 cases were uniformly mutated, whereas subset 2 cases exhibited significant heterogeneity with regard to mutational load (supplemental Table 1); finally, all subset 8 cases had 100% germline identity. The heavy chain isotype was the same among members of a subset (for cases with available data). All subsets expressed IgMD, except for subsets 4 (*IGHV4-34* gene), 8 (*IGHV4-39* gene), and 16 (*IGHV4-34* gene), which included IgG-expressing cases, as previously reported.<sup>13,17,18</sup>

### Intracлонаl diversification analysis at the nucleotide level

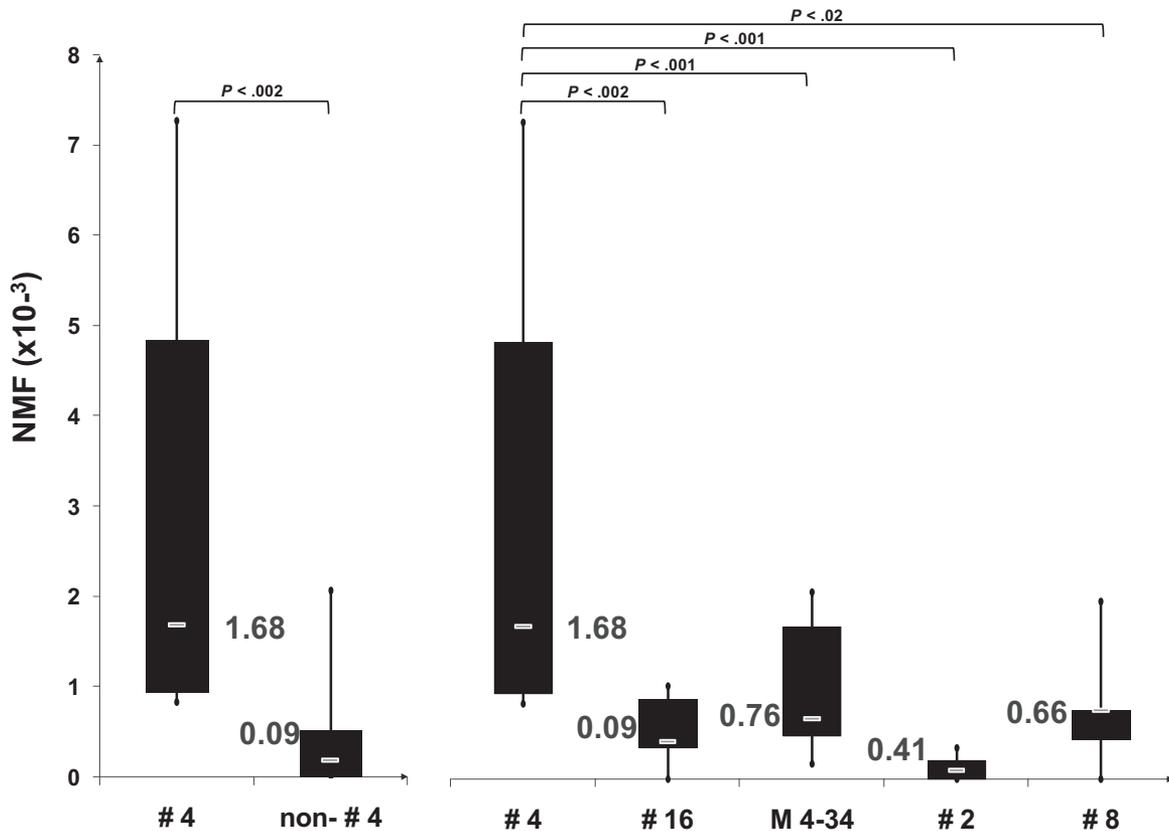
Following the definitions detailed in “Definitions,” cases were allocated to one of 3 categories: (1) no intracлонаl diversity: 20 cases with identical sets of subcloned sequences; (2) unconfirmed intracлонаl diversity (UID): 23 cases exhibiting only UCMs in certain positions of the V domain; and (3) confirmed intracлонаl diversity (CID): 28 cases exhibiting at least one CM among subcloned sequences (supplemental Table 2). Based on direct sequencing analysis, 23 of 28 cases (82%) carrying CMs belonged to the “mutated” category (< 98% germline identity).

Interesting biases with regard to the IGHV gene repertoire were observed among the 3 categories defined by ID status. In particular, 20 of 28 (71.4%) cases exhibiting CID concerned *IGHV4-34*-expressing cases, whereas the corresponding percentages for *IGHV4-34* usage by cases belonging to the UID category and non-ID category were 26% (6 of 23) and 10% (2 of 20), respectively ( $\chi^2$  test:  $P < .001$ ). Notably, 12 of 16 truly unmutated rearrangements included in the present study were found to exhibit some degree of ID (CID, 4 of 16 cases [25%]; UID, 8 of 16 cases [50%]).

Among 28 cases exhibiting CMs (supplemental Table 2), 13 (46.4%) belonged to subset 4 with stereotyped *IGHV4-34/IGKV2-30* BCRs (Table 1). Furthermore, a much more pronounced impact of ID was noted among subset 4 rearrangements versus all other rearrangements included in the analysis, as evidenced by significantly higher NMFs ( $t$  test:  $P < .002$ ), regardless of IGHV gene usage and mutation status, BCR stereotypy, or heavy chain isotype (Figure 1; supplemental Table 2). Interestingly, the high number of CMs (and UCMs) observed in several subset 4 cases (P0103, P1422, P2451, P2920, P3551, and Swe56) could be attributed to the existence of distinct “clusters” of subcloned sequences with “cluster-specific” mutational profiles. All such clusters exhibited certain shared mutations and closely similar, if not identical, HCDR3s (including identical HCDR3 length), pointing to the emergence from a common ancestor (Figure 2; supplemental Figure 1).

With the exception of 2 confirmed 3-bp deletions, all observed CMs and UCMs concerned single-base changes and resulted in both silent (S) and replacement (R) mutations. Overall, 212 silent CMs/UCMs were identified, compared with 288 replacement CMs/UCMs. Nucleotide substitution analysis revealed that transitions predominated over transversions both for CMs and UCMs (175 vs 123 and 158 vs 44, respectively), in keeping with a canonical SHM process.<sup>46</sup> The analysis for targeting of SHM hotspots over HFR1-3 revealed that only 141 of 345 (41%) mutations occurred within a hotspot motif.

Comparison of subset 4 *IGHV4-34* sequences to all other sequences revealed significant differences with regard to the distribution of CMs/UCMs. The most striking difference concerned the fact that *IGHV4-34* rearrangements of subset 4 were highly targeted for UCMs/CMs within HCDR3. Furthermore, a general shift of mutations to the 3' part of the V-D-J region, in particular HCDR3-HFR4, was noted among subset 4 rearrangements (supplemental Figure 2).



**Figure 1. Box-plot analysis of NMF values.** Comparison of NMF values among subset 4 rearrangements (subset 4, *IGHV4-34/IGKV2-30*)<sup>17,18</sup> versus (1) all other rearrangements included in the analysis (nonsubset 4); (2) subset 16 rearrangements (subset 16, *IGHV4-34/IGKV3-20*)<sup>17,18</sup>; *IGHV4-34* rearrangements from cases expressing the common IgM/D isotype (M 4-34); subset 2 rearrangements (subset 2, *IGHV3-21/IGLV3-21*)<sup>17,18</sup>; and subset 8 rearrangements (subset 8, *IGHV4-39/IGKV1(D)-39*).<sup>17,18</sup>

#### Intraclonal diversification of *IGHV-D-J* rearrangements at the AA level

A total of 273 novel AA changes and 2 single-AA deletions were identified in 47 of 71 evaluated cases (66.2%). Following definitions presented in “Definitions,” 134 of 273 AA changes (49%) were considered as unconfirmed (UAA), whereas the remainder (139 of 273; 51%) were detected in at least 2 subcloned sequences from the same patient and, therefore, were considered as confirmed (CAA; supplemental Table 2). In addition, 2 cases using the *IGHV3-21* gene were observed to carry a confirmed serine deletion at IMGT/HCDR2 codon 59 (supplemental Table 3).

CAA changes were detected across the entire *IGHV* gene among 27 cases using the *IGHV3-21*, *IGHV3-23*, *IGHV4-34*, or *IGHV4-39* genes (supplemental Tables 3-6); 13 of these cases belonged to subset 4 with stereotyped *IGHV4-34/IGKV2-30* BCRs. Intriguingly, analysis of the distribution of the so-called UAA changes provided further evidence for the very precise targeting of mutations introduced as part of the ID process. Thus: (1) 26 UAA changes were identified at the same codon among cases using the same *IGHV* gene (“confirmed by another case”); and (2) some AA changes identified in single subcloned sequences from one case were shared by all subcloned sequences of another case using the same *IGHV* gene, especially when expressing homologous, “stereotyped” HCDR3s (Figure 3).

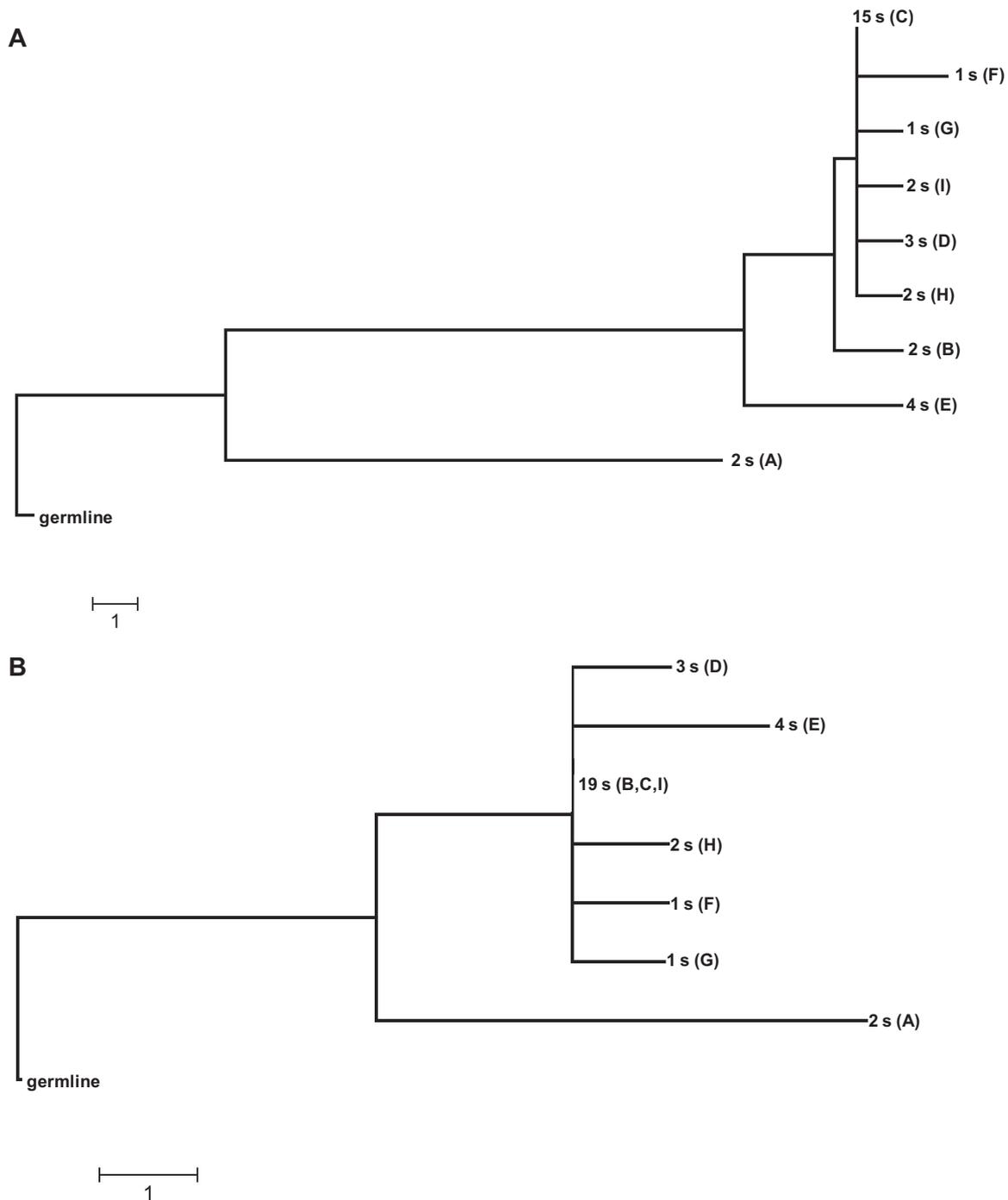
#### AA changes in subcloned *IGHV-D-J* sequences of subset 4 cases

The frequency of AA changes among subcloned sequences was significantly higher among subset 4 cases than all other cases

analyzed (supplemental Tables 2-6; supplemental Figure 3). Of note, however, several changes identified in sets of subcloned sequences from subset 4 cases were rather conservative in terms of AA physicochemical properties (Figure 4). Novel changes were often “stereotyped,” in the sense that different cases could share recurrent AA changes among subcloned sequences; therefore, the “options” for certain positions were remarkably restricted (Figure 4). A comprehensive list of AA changes in sets of subcloned sequences from *IGHV4-34* rearrangements of subset 4 cases is given in supplemental Table 4.

A remarkable impact of ID was evident in the HCDR3 sequences of subset 4 cases, with 11 of the 20 HCDR3 codons showing CAA changes. Importantly, for several cases, the AA diversity observed intraclonally was found to reflect the overall heterogeneity observed at subset level. In particular, AAs appearing in some subcloned sequences of one case represented the predominant residues at the same position in another case (Figure 4; supplemental Table 4). For instance, the only permitted options among subcloned sequences at position 105 (following the IMGT numbering for the V domain) were A or V, with “intra-subset” variation also reflected at intraclonal level. An unprecedented high level of ID was observed in the HFR4 of subset 4 rearrangements, with 3 HFR4 positions (codons 120, 123, and 125) identified as ID “hotspots” (supplemental Table 4).

Taking into account the high number of CMs and UCMs among subset 4 *IGHV4-34* rearrangements, it is remarkable that certain critical positions of the VH domain were essentially left unaltered by the ID process. For instance, the W residue at codon 7 in the HFR1 of *IGHV4-34*, which plays a critical role in the recognition of



**Figure 2. Clonal evolution trees.** Clonal evolution trees constructed by MEGA algorithms and tools (maximum parsimony, bootstrap tree) for case P0103 from subset 4. (A) Nucleotide tree. (B) AA tree. s indicates subcloned sequence.

the I/i NAL epitope,<sup>47</sup> was preserved in 322 of 326 subcloned sequences from 13 evaluated subset 4 cases. A similar case of sequence preservation was also observed in HCDR1. As previously shown by our group,<sup>18</sup> 2 glycine (G) residues at codons 28 and 36 of the HCDR1 of subset 4 cases were very frequently targeted for “subset-biased” substitution by acidic residues (aspartate, D, or glutamate, E). This hallmark of SHM activity for subset 4 cases was not affected by ongoing mutational activity, as all observed mutations at codon 28 involved only G-D or G-E substitutions. Indeed, the only CAA change at this position concerned the introduction of D or E residues and, in one case, substitution of D for E (supplemental Table 4). Finally, the distinctive and “subset-

biased” couplet of RR or KR dipeptides at the IGHD-IGHJ junction of subset 4 IGHV4-34 rearrangements<sup>17,18</sup> was not targeted by SHM, and the only sign of mutational activity was an unconfirmed R-K replacement (Figure 4).

## Discussion

Molecular analysis of the BCR in CLL has revealed biases in IG gene repertoire suggesting a potential role for antigenic stimulation in CLL ontogeny.<sup>8,9,21</sup> This notion has been further supported by the discovery of subsets of CLL cases carrying stereotyped BCRs.<sup>10-18</sup>

Case	n	Subset #4							
		28:G	59:S	82:T	84:K	85:N	90:K	105:A	128:S
P0103	32	G2, E30	-	-	K31, E1	N30, S2	-	-	-
P0907	33	D33	S32, R1	-	-	-	-	A32, V1	-
P1422	27	E21, D6	-	-	K26, R1	-	K21, Q6	A21, V6	S26, K1
P1939	35	G1, E34	-	T34, A1	-	-	N35	V35	S34, K1
P2451	35	G1, D33, E1	-	T34, A1	-	N34, S1	K2, Q33	A2, V33	-
P2920	26	-	S25, R1	-	-	-	K25, R1	-	-
P3020	16	-	F16	-	-	-	-	-	-
P3551	20	G8, E12	-	-	-	-	T11, R8, A1	-	-
P3916	17	D17	-	A17	-	-	-	-	-
P6520	22	-	-	-	-	-	R22	-	-
Swe181	14	E14	-	-	-	-	R14	-	-
Swe193	16	-	-	-	-	K16	T16	-	-
Swe56	33	E33	S31, H2	-	K32, R1	N32, D1	-	-	S31, K2

**Figure 3. Unconfirmed mutations at various positions throughout the V domain of subcloned sequences from subset 4 cases illustrating the “confirmed by another case” phenomenon.** In the column headers, the number refers to the IMGT codon numbers, whereas the germline AA at each codon is indicated using the single-letter AA code. Within each column, the number of subcloned sequences containing a certain AA is indicated after the single-letter AA code. Unconfirmed mutations, which may be considered as “confirmed by another case,” are highlighted in bold. Dashes indicate that all subcloned sequences carried the germline AA. For example, the single K to R mutation at position IMGT/HFR3-90 in case P2920 is confirmed by 8 subcloned sequences in P3551 as well as K to R mutations in all subclones of cases P6520 and Swe181.

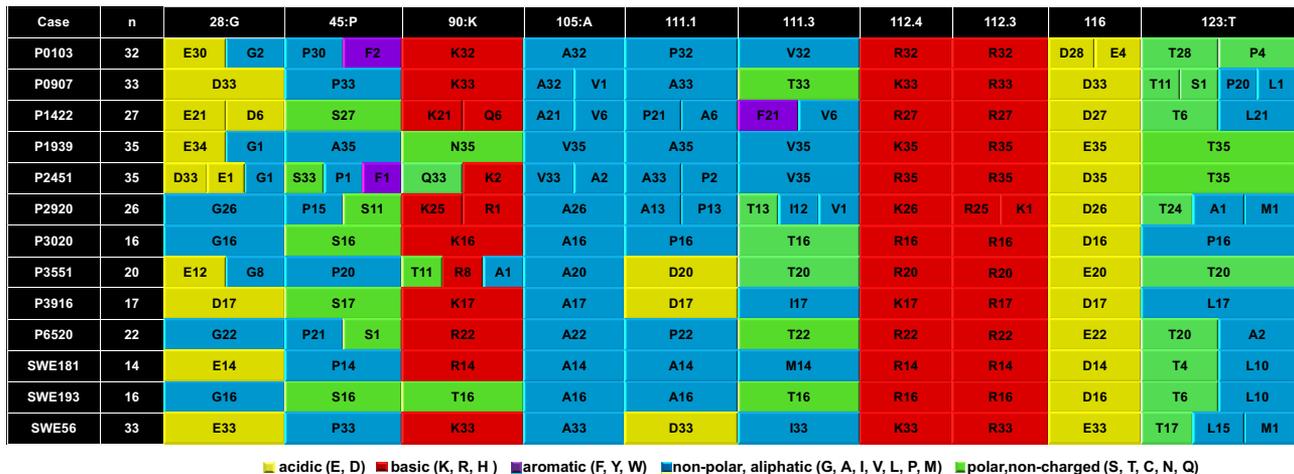
As we have recently shown, a recurrent theme in subsets carrying mutated stereotyped BCRs is the presence of stereotyped, “CLL-” and “subset-biased” mutation patterns, alluding to a very precise, functionally driven response to a restricted set of antigenic stimuli.<sup>18,48</sup> Therefore, it can be concluded that SHM studies help to better understand and delineate critical issues in the history and biology of clonal malignant CLL cells. That notwithstanding, several important issues remain unresolved, especially with regard to the location, duration, and nature of the interactions between the selecting antigens and the CLL precursors as well as the differentiation status of the precursor cells themselves.

As previously shown for other B-cell malignancies, the study of ID of the clonotypic IG genes may provide important information that helps to address the aforementioned issues. More specifically, in some entities, such as follicular lymphoma, sporadic Burkitt lymphoma, splenic marginal zone lymphoma, and mucosa-associated lymphoid tissue lymphoma, the analysis of SHM reveals ID, indicating that the neoplastic cells further diversify their IG genes through ongoing mutational activity,<sup>22-28</sup> either through conventional SHM in GCs or unconventional SHM outside GCs. On the contrary, molecular analyses of IG genes in multiple

myeloma showed intraclonal homogeneity suggesting that the myeloma clone derived from a post-GC memory B cell.<sup>29,30</sup>

The studies devoted to clarifying the possible occurrence and hence implications of ID in CLL have reached conflicting results.<sup>31-37</sup> These discrepancies may be attributed, at least in part, to differences in the sensitivity of various methods used for detection of mutations,<sup>31-37</sup> usage of low-fidelity *Taq* polymerase,<sup>31-37</sup> number of analyzed subcloned sequences,<sup>34</sup> sample size, or case selection. For instance, most studies analyzed relatively few patients<sup>31-33,35,37</sup> or even a single case.<sup>36</sup> Furthermore, an underrepresentation of certain CLL subsets with distinctive SHM patterns, namely, cases using the *IGHV3-21* and *IGHV4-34* genes (especially in subsets 2 or 4), was observed throughout all previous studies. Finally, the lack of a standard criterion for ID evaluation may also lead to ambiguity among previously reported data.

Taking into consideration the aforementioned discrepancies, we conducted the present large-scale subcloning study following a strict methodologic approach to readdress ID in CLL. Our study included 71 carefully selected CLL cases, thereby enabling us to explore the presence of ID not only at cohort level, but also in subgroups defined by BCR stereotypy, IGHV gene mutational



**Figure 4. Recurrent (“stereotyped”) AA changes among sets of subcloned sequences from subset 4 cases.** Recurrent AA changes among subcloned sequences were observed in certain codons throughout the VH domain (IMGT unique codon numbering); such changes could be considered as “stereotyped,” in the sense that different cases could share the same AA in several subcloned sequences. AAs are represented by a single-letter code. The number of subcloned sequences containing a certain AA is indicated after the single-letter code for each AA.

status, and heavy chain isotype. In addition, a high-fidelity polymerase was used for all analyses, thus minimizing the incidence of errors. Finally, a large number of subcloned sequences per case (median, 21; range, 14-35) were evaluated.

At cohort level, 28% (20 of 71) of cases carried sets of identical subcloned sequences. An additional 32% (23 of 71) of cases were characterized by the presence of mutations in single subcloned sequences and, following the strict definitions adopted in this study, could not be formally assigned to the intraclonally diversified category. What is important to mention, however, is that almost 40% (28 of 71) of cases carried intraclonally diversified IGHV-IGHD-IGHJ genes with CM among subclones. CMs were identified in cases assigned to all mutational categories, even in the “truly unmutated” category (ie, 100% identity to the germline). Altogether, our results strongly support the notion that the SHM mechanism remains operative in CLL and suggest that antigen stimulation may be a promoting factor not only in the development of CLL clones but also in their evolution. However, the fact that ID was generally insubstantial in leukemic clones with unmutated IGHV genes that usually follow more aggressive clinical courses may be considered as evidence that their expansion is antigen-independent or, more probably, that their unmutated status confers a clonal advantage by permitting very efficient stimulation by a variety of self-antigens that would be compromised or even abrogated on alteration of the original BCR structure.

The molecular characteristics of mutations occurring in the context of ID are compatible with the idea that the underlying causal mechanisms may generally obey the same rules as the canonical SHM.<sup>46,49</sup> More specifically, transitions were more common than transversions, in keeping with previous reports.<sup>46,49</sup> However, the observed distribution patterns of CMs/UCMs differed from those reported as typical for mutations introduced by canonical SHM, leading to R/S ratios not characteristic of selection by conventional, T-dependent antigens.<sup>50</sup>

The breakdown of the group of cases exhibiting ID revealed a striking bias for usage of the *IGHV4-34* gene, especially in stereotyped *IGHV4-34/IGKV2-30* BCRs of subset 4.<sup>18</sup> It is also worth underscoring the fact that the majority of mutations introduced within the context of the ID process were identified in subcloned sequences of *IGHV4-34* rearrangements from subset 4 cases. Therefore, one could reasonably argue that in discussing the molecular characteristics of ID in CLL we are essentially describing the operation of ongoing mutational activity in subset 4.

In keeping with our previous reports,<sup>17,18</sup> subset 4 cases included in the present study had a young median age at diagnosis (51 years; range, 37-69 years) and were diagnosed in early clinical stages; they were uniformly CD38-negative and IgG-switched; finally, 5 of 11 cases with available data carried del(13q) as the sole cytogenetic abnormality (supplemental Table 7). As recently shown by our group,<sup>18</sup> at a molecular level the distinctive features of this subset extend to SHM patterns in the clonotypic IG genes, including: (1) “noncanonical” distribution of R mutations, with low R/S mutation ratios in HCDRs (especially HCDR2); (2) complete lack of mutations, leading to preservation of the germline conformation, at the *IGHV4-34*-specific motif, which mediates superantigenic-like interactions with self-elements and exogenous elements<sup>47</sup>; and (3) recurrent, “stereotyped” hypermutations.<sup>18</sup> On these grounds, we proposed that the SHM characteristics in subset 4 are indicative of a particular mode of interaction with and selection by distinct (super)antigenic elements.<sup>18</sup>

The results reported here reveal that the “subset-biased” impact of SHM extends to very distinctive patterns of ID. In particular,

certain residues of the stereotyped *IGHV4-34* rearrangements of subset 4 cases were identified as “hotspots” for ID; these “hotspots” were scattered throughout the V domain, including, somewhat unexpectedly, HFR4. Interestingly, the observed AA changes at these “hotspots” were generally restricted (“stereotyped”) and, with a few notable exceptions, conservative. This suggests that the permissible “options” at these positions are also restricted, perhaps by strong functional constraints for preservation of critical physicochemical properties.

From a different perspective, further evidence that SHM in subset 4 is functionally driven and also very precisely targeted is offered by our finding that certain positions across the entire VH domain remained essentially unaltered, despite intense ID in nearby positions. Perhaps the most illustrative example is provided by the couplet of basic AAs (RR, KR) located at the *IGHD-IGHJ* junctions of subset 4 *IGHV4-34* rearrangements,<sup>18</sup> which endow the corresponding HCDR3s with a positive charge. Prompted by the analogy to HCDR3 sequence motifs of *IGHV4-34* antibodies against both apoptotic cells<sup>51,52</sup> and DNA,<sup>53-55</sup> we have previously postulated that the progenitors of CLL cases assigned to subset 4 may have originated as cells with reactivity against DNA or apoptotic bodies.<sup>17,18</sup> Along these lines, it is worth mentioning that the single mutation observed at this dipeptide among 326 subcloned sequences from subset 4 rearrangements concerned a substitution of K for R. This cannot be attributed to an overall lack of ID within HCDR3, given that other HCDR3 positions were frequently targeted for CMs or UCMs.

Interestingly, this single K-to-R mutation, though “unconfirmed,” following the definitions adopted in our study, was “nonrandom”: a simple inspection of the aligned subcloned sequences justifies considering this mutation as “confirmed by another case.” Based on the findings reported in “Intraclonal diversification of *IGHV-D-J* rearrangements at the AA level” and also depicted graphically in Figure 3, the same argument may hold true for several other mutations detected in single subcloned sequences. Therefore, although the practice of subdividing mutations into CM or UCM categories could be considered as a “safeguard” for reliability, finding mutations “unconfirmed at case level,” which are “confirmed at subset level” raises the intriguing possibility that at least some UCMs should perhaps be counted as signs of true ID.

The strikingly different impact of ID in subset 4 cases versus all other cases analyzed in this study could be thought to reflect an inherent mutability of the *IGHV4-34* gene, or, alternatively, their mutated status or IgG-switched phenotype. However, on the evidence presented here, we argue that these findings cannot be attributed to *IGHV4-34* usage, IGHV gene-mutated status, class-switch recombination, or BCR stereotypy in general. Indeed, this is evident by comparison of stereotyped subset 4 *IGHV4-34* rearrangements to: (1) nonsubset 4 *IGHV4-34* rearrangements of the common IgMD variant; (2) stereotyped *IGHV4-34* rearrangements of IgG-switched cases in subset 16<sup>18</sup>; and (3) both mutated and unmutated rearrangements using other IGHV genes. Therefore, the ID patterns observed in subset 4 represent a unique phenomenon strongly correlated with the distinctive *IGHV4-34/IGKV2-30* BCR archetype expressed by subset 4 cases and allude to “stereotyped” interactions with the cognate antigen(s) during both the preleukemic phase but also posttransformation.

The intense ID activity in subset 4 is also evidenced by the identification of distinct “clusters” of subcloned sequences with “cluster-specific” mutational profiles. Analysis of the SHM patterns in such clusters reveals their common ancestry yet indicates

an early “branching” of the leukemic clone into distinct subclones, perhaps able to evolve along similar, although separate, pathways. Taking into account the polyreactivity of the clonotypic BCRs in CLL<sup>56-59</sup> as well as our recent report that persistent activation by Epstein-Barr virus and cytomegalovirus may be specifically implicated in the history of subset 4 leukemic clones,<sup>60</sup> one might consider this “branching” as evidence for special, selective pressures occurring in parallel in distinct subclones and thereby fine-tuning their BCR affinities.

In conclusion, our study convincingly demonstrates that the SHM mechanism may operate continuously in certain subsets of patients with CLL, especially patients expressing stereotyped IGHV4-34 rearrangements typical of subset 4. However, it is still difficult to reach definitive conclusions with regard to the duration of exposure to and stimulation by antigen as well as the functional impact of antigen on CLL evolution. That notwithstanding, the results reported here suggest a role for persistent antigenic stimulation rather than clonal expansion promoted by nonspecific stimuli, at least for subsets of CLL cases.

## Acknowledgments

The authors thank Prof Marie-Paule Lefranc and Dr Veronique Giudicelli, Laboratoire d’Immunogenetique Moleculaire, Universite Montpellier II, Montpellier, France, for their enormous support and help with the large-scale immunoglobulin sequence analysis throughout this project; Jesper Jurlander, Karin Karlsson, Mats

Merup, Lyda Osorio, Göran Roos, Christer Sundström, and Juhani Vilpo for providing samples and associated data concerning Swedish, Danish, and Finnish CLL patients; and Andreas Agathage-lidis, Vasilis Bikos, Maria Gounari, Ulf Thunberg, and Gerard Tobin for the sequence analysis.

This work was supported by the Swedish Cancer Society, the Swedish Medical Research Council, the Medical Faculty of Uppsala University, Uppsala University Hospital, and the Lion’s Cancer Research Foundation, Uppsala, Sweden; and the BioSapiens Network of Excellence (contract no. LSHG-CT-2003-503265) and the General Secretariat for Research and Technology of Greece (Program INA-GENOME). E.K. is a recipient of a fellowship from the Propontis Foundation, Athens, Greece.

## Authorship

Contribution: L.-A.S. and E.K. performed research, analyzed data, and wrote the paper; A.H. and N.D. performed research; A.A. provided samples and associated data; A.T. supervised research; and R.R. and K.S. designed and supervised the research and wrote the paper.

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

Correspondence: Kostas Stamatopoulos, Hematology Department and HCT Unit, G. Papanicolaou Hospital, 57010 Thessaloniki, Greece; e-mail: stavstam@otenet.gr.

## References

- Rosenwald A, Alizadeh AA, Widhopf G, et al. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med*. 2001;194(11):1639-1647.
- Klein U, Tu Y, Stolovitzky GA, et al. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med*. 2001;194(11):1625-1638.
- Damle RN, Ghiotto F, Valetto A, et al. B-cell chronic lymphocytic leukemia cells express a surface membrane phenotype of activated, antigen-experienced B lymphocytes. *Blood*. 2002;99(11):4087-4093.
- Schroeder HW Jr, Dighiero G. The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire. *Immunol Today*. 1994;15(6):288-294.
- Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest*. 1998;102(8):1515-1525.
- Damle R, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999;94(6):1840-1847.
- Hamblin T, Davis Z, Gardiner A, et al. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999;94(6):1848-1854.
- Chiorazzi N, Ferrarini M. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Annu Rev Immunol*. 2003;21:841-894.
- Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *N Engl J Med*. 2005;352(8):804-815.
- Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood*. 2003;101(12):4952-4957.
- Ghiotto F, Fais F, Valetto A, et al. Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *J Clin Invest*. 2004;113(7):1008-1016.
- Widhopf GF 2nd, Rassenti LZ, Toy TL, Gribben JG, Wierda WG, Kipps TJ. Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. *Blood*. 2004;104(8):2499-2504.
- Messmer BT, Albesiano E, Efremov DG, et al. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J Exp Med*. 2004;200(4):519-525.
- Tobin G, Thunberg U, Karlsson K, et al. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood*. 2004;104(9):2879-2885.
- Ghia P, Stamatopoulos K, Belessi C, et al. Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: the lesson of the IGHV3-21 gene. *Blood*. 2005;105(4):1678-1685.
- Thorselius M, Krober A, Murray F, et al. Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-utilizing chronic lymphocytic leukemia independent of geographical origin and mutational status. *Blood*. 2006;107(7):2889-2894.
- Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. *Blood*. 2007;109(1):259-270.
- Murray F, Darzentas N, Hadzidimitriou A, et al. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. *Blood*. 2008;111(3):1524-1533.
- Tobin G, Thunberg U, Johnson A, et al. Somatic mutation of Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood*. 2002;99(6):2262-2264.
- Caligaris-Cappio F, Ghia P. The normal counterpart to the chronic lymphocytic leukemia B cell. *Best Pract Res Clin Haematol*. 2007;20(30):385-397.
- Ghia P, Chiorazzi N, Stamatopoulos K. Microenvironmental influences in chronic lymphocytic leukaemia: the role of antigen stimulation. *J Intern Med*. 2008;264(6):549-562.
- Chapman CJ, Mockridge CI, Rowe M, Rickinson AB, Stevenson FK. Analysis of VH genes used by neoplastic B cells in endemic Burkitt’s lymphoma shows somatic hypermutation and intraclonal heterogeneity. *Blood*. 1995;85(8):2176-2181.
- Chapman CJ, Zhou JX, Gregory C, Rickinson AB, Stevenson FK. VH and VL gene analysis in sporadic Burkitt’s lymphoma shows somatic hypermutation, intraclonal heterogeneity, and a role for antigen selection. *Blood*. 1996;88(9):3562-3568.
- Ottensmeier CH, Thompsett AR, Zhu D, Wilkins BS, Sweetenham JW, Stevenson FK. Analysis of VH genes in follicular and diffuse lymphoma shows ongoing somatic mutation and multiple isotype transcripts in early disease with changes during disease progression. *Blood*. 1998;91(11):4292-4299.
- Thiede C, Alpen B, Morgner A, et al. Ongoing somatic mutations and clonal expansions after cure of *Helicobacter pylori* infection in gastric mucosa-associated lymphoid tissue B-cell lymphoma. *J Clin Oncol*. 1998;16(12):3822-3831.
- Aarts WM, Willemze R, Bende RJ, Meijer CJ, Pals ST, van Noesel CJ. VH gene analysis of primary cutaneous B-cell lymphomas: evidence for ongoing somatic hypermutation and isotype switching. *Blood*. 1998;92(10):3857-3864.

27. Lossos IS, Alizadeh AA, Eisen MB, et al. Ongoing immunoglobulin somatic mutation in germinal center B cell-like but not in activated B cell-like diffuse large cell lymphomas. *Proc Natl Acad Sci U S A*. 2000;97(18):10209-10213.
28. Zhu D, Orchard J, Oscier DG, Wright DG, 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(7):2659-2661.
29. Bakkus MH, Heirman C, Van Riet I, Van Camp B, Thielemans K. Evidence that multiple myeloma Ig heavy chain VDJ genes contain somatic mutations but show no intraclonal variation. *Blood*. 1992;80(9):2326-2335.
30. Sahota SS, Leo R, Hamblin TJ, Stevenson FK. Myeloma VL and VH sequences reveal a complementary imprint of antigen selection in tumor cells. *Blood*. 1997;89(1):219-226.
31. Hashimoto S, Dono M, Wakai M, et al. Somatic diversification and selection of immunoglobulin heavy and light chain variable region genes in IgG+ CD5+ chronic lymphocytic leukemia B cells. *J Exp Med*. 1995;181(4):1507-1517.
32. Dono M, Hashimoto S, Fais F, et al. Evidence for progenitors of chronic lymphocytic leukemia B cells that undergo intraclonal differentiation and diversification. *Blood*. 1996;87(4):1586-1594.
33. Gurrieri C, McGuire P, Zan H, et al. Chronic lymphocytic leukemia B cells can undergo somatic hypermutation and intraclonal immunoglobulin V(H)DJ(H) gene diversification. *J Exp Med*. 2002;196(5):629-639.
34. Degan M, Bomben R, Dal Bo M, et al. Analysis of IgVH gene mutations in B cell chronic lymphocytic leukaemia according to antigen-driven selection identifies subgroups with different prognosis and usage of the canonical somatic hypermutation machinery. *Br J Haematol*. 2004;126(1):29-42.
35. Volkheimer AD, Weinberg JB, Beasley BE, et al. Progressive immunoglobulin gene mutations in chronic lymphocytic leukemia: evidence for antigen-driven intraclonal diversification. *Blood*. 2007;109(4):1559-1567.
36. Bagnara D, Callea V, Stelitano C, et al. IgV gene intraclonal diversification and clonal evolution in B-cell chronic lymphocytic leukaemia. *Br J Haematol*. 2006;133(1):50-58.
37. Schettino EW, Cerutti A, Chiorazzi N, Casali P. Lack of intraclonal diversification in Ig heavy and light chain V region genes expressed by CD5+IgM+ chronic lymphocytic leukemia B cells: a multiple time point analysis. *J Immunol*. 1998;160(2):820-830.
38. Hallek M, Cheson BD, Catovsky D, et al. International Workshop on Chronic Lymphocytic Leukemia. 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):5446-5456.
39. Lefranc MP, Giudicelli V, Ginestoux C, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res*. 2009;37(Database issue):D1006-D1012.
40. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res*. 2008;36(Web Server issue):W503-W508.
41. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. 2002;Chapter 2:Unit 2.3.
42. Pommie C, Levadoux S, Sabatier R, Lefranc G, Lefranc MP. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J Mol Recognit*. 2004;17(1):17-32.
43. Rogozin IB, Pavlov YI. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res*. 2003;544(1):65-85.
44. 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):3382-3384.
45. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24(8):1596-1599.
46. Zheng NY, Wilson K, Jared M, Wilson PC. Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation. *J Exp Med*. 2005;201(9):1467-1478.
47. 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):3777-3782.
48. Hadzidimitriou A, Darzentas N, Murray F, et al. Evidence for the significant role of immunoglobulin light chains in antigen recognition and selection in chronic lymphocytic leukemia. *Blood*. 2009;113(2):403-411.
49. Teng G, Papavasiliou FN. Immunoglobulin somatic hypermutation. *Annu Rev Genet*. 2007;41:107-120.
50. William J, Euler C, Christensen S, Shlomchik MJ. Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. *Science*. 2002;297(5589):2066-2070.
51. Cocca BA, Seal SN, D'Agnillo P, et al. Structural basis for autoantibody recognition of phosphatidyserine-beta 2 glycoprotein I and apoptotic cells. *Proc Natl Acad Sci U S A*. 2001;98(24):13826-13831.
52. Lorenz HM, Herrmann M, Winkler T, Gaipal U, Kalden JR. Role of apoptosis in autoimmunity. *Apoptosis*. 2000;5(5):443-449.
53. Jang YJ, Stollar BD. Anti-DNA antibodies: aspects of structure and pathogenicity. *Cell Mol Life Sci*. 2003;60(2):309-320.
54. Li Z, Schettino EW, Padlan EA, Ikematsu H, Casali P. Structure-function analysis of a lupus anti-DNA autoantibody: central role of the heavy chain complementarity determining region 3 Arg in binding of double- and single-stranded DNA. *Eur J Immunol*. 2000;30(7):2015-2026.
55. 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):2430-2439.
56. Herve M, Xu K, Ng YS, et al. Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. *J Clin Invest*. 2005;115(6):1636-1643.
57. Lanemo Myhrinder A, Hellqvist E, Sidorova E, et al. A new perspective: molecular motifs on oxidized LDL, apoptotic cells, and bacteria are targets for chronic lymphocytic leukemia antibodies. *Blood*. 2008;111(7):3838-3848.
58. Catera R, Silverman GJ, Hatzi K, et al. Chronic lymphocytic leukemia cells recognize conserved epitopes associated with apoptosis and oxidation. *Mol Med*. 2008;14(11):665-674.
59. Chu CC, Catera R, Hatzi K, et al. Chronic lymphocytic leukemia antibodies with a common stereotypic rearrangement recognize nonmuscle myosin heavy chain IIA. *Blood*. 2008;112(13):5122-5129.
60. Kostareli E, Hadzidimitriou A, Stavroyianni N, et al. Molecular evidence for EBV and CMV persistence in a subset of patients with chronic lymphocytic leukemia expressing stereotyped IGHV4-34 B-cell receptors. *Leukemia*. 2009;23(5):919-924.