Available online at www.sciencedirect.com
ScienceDirect
Leukemia Research 31 (2007) 1247–1252
Leukemia Research
www.elsevier.com/locate/leukres

Use of IGHJ and IGHD gene mutations in analysis of immunoglobulin sequences for the prognosis of chronic lymphocytic leukemia

Cindy E.H. Lee [a], Katherine J.L. Jackson [a], William A. Sewell [b], Andrew M. Collins [a,*]

[a] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Kensington, Sydney, NSW 2052, Australia
[b] Institute of Laboratory Medicine, St. Vincent's Hospital Sydney and St. Vincent's Clinical School, University of New South Wales, Australia

Received 21 September 2006; received in revised form 21 September 2006; accepted 19 October 2006
Available online 13 December 2006

Abstract

The level of somatic point mutation in immunoglobulin genes is an important prognostic indicator for patients with chronic lymphocytic leukemia (CLL). Mutation analysis presently focuses solely upon the heavy chain IGHV gene, however mutation is a stochastic process that also targets IGHD and IGHJ genes. Here, we evaluate the completeness and reliability of the reported IGHJ gene repertoire, and demonstrate the likely consequences of the inclusion of IGHD and IGHJ mutations in CLL analysis, using a dataset of 607 sequences. Inclusion of these mutations would lead to the re-classification of many sequences, which should significantly improve the prognostic value of mutation analysis.

Keywords: Chronic lymphocytic leukemia; Somatic mutation; Immunoglobulin; IGHJ; Repertoire development

1. Introduction

CLL is the commonest form of leukemia in Western countries. In most individuals, the disease remains stable for many years, whereas in others, there is rapid disease progression. In patients who present with isolated lymphocytosis, there are no simple clinical criteria to determine whether they are likely to have stable or progressive disease, but the extent of somatic point mutation in immunoglobulin genes from the leukemic cells is recognized to be an important prognostic indicator for the disease [1,2]. Patients whose tumour cell immunoglobulin heavy chain genes show relatively high levels of somatic mutation have a significantly increased mean survival time compared with patients with fewer mutations. The clinical estimation of the extent of somatic point mutations in immunoglobulin genes is presently based upon an examination of the heavy chain IGHV genes, and 'mutated' sequences are usually defined as those sequences in which mutations affect 2% or more of the nucleotides in the IGHV gene. The somatic point mutation process is not, however, confined to the IGHV genes.

Functional immunoglobulin genes are generated, during the early development of B cells, by the recombination of a number of short gene segments. The immunoglobulin heavy chains are the products of approximately 46 unique IGHV genes, 23 IGHD genes and six IGHJ genes. Early in the development of each B cell, one of each kind of gene is randomly chosen from the sets of available genes, and during an immune response, all three genes in an assembled VDJ gene may accumulate somatic point mutations. The determination of mutation levels in an assembled heavy chain gene requires the VDJ gene sequence to be aligned against the sets of germline IGHV, IGHD and IGHJ genes. Perhaps because of perceived difficulties in the identification of IGHD and IGHJ genes within a rearranged VDJ gene, analysis of CLL mutations has always been confined to the IGHV gene, despite the fact that virtually all studies generate longer VDJ sequences.

Although IGHV genes are highly polymorphic [3], they are approximately 300 nucleotides in length, making it

* Corresponding author. Tel.: +61 2 9385 3441; fax: +61 2 9385 1483.
E-mail address: a.collins@unsw.edu.au (A.M. Collins).

0145-2126/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.leukres.2006.10.013

relatively straightforward to identify them within a heavy chain VDJ gene. In contrast, the identification of IGHD genes is made extremely difficult by their short lengths. The germline genes range in length from just 11 to 37 nucleotides, and as a consequence of exonuclease activity, the expressed IGHD genes are even shorter. We have recently developed a hidden Markov model-based alignment program (iHHMune-align) that can reliably identify IGHD genes in rearranged heavy chain genes, and have confirmed the relative completeness of the reported IGHD repertoire [4]. It should therefore be possible to consider point mutations of IGHD genes in CLL analysis. This should improve the definition of CLL patient groups, for the IGHD gene represents a major part of the heavy chain third complementarity determining region (CDR3), which is generally considered to be critical to antigen binding [5], and which contains a relatively high number of mutational hotspots [6].

The 5′ ends of IGHJ genes also contribute to the CDR3 regions, and frequently accumulate mutations. However, if mutations of the IGHJ genes are to be incorporated into measures of the extent of mutation of CLL immunoglobulin genes, the reliability of alignments to IGHJ genes must also be established. IGHJ genes are of intermediate length (approximately 50 nucleotides), and little polymorphism of IGHJ genes has been reported. Together the six IGHJ genes include just six reported allelic variants [3]. Although the identification of IGHJ gene segments in heavy chain gene rearrangements may therefore seem straightforward, we are aware of no study that has systematically considered the IGHJ repertoire with a view to identifying problems with such alignments.

In this study, we describe analyses of the frequencies of utilization of the different genes, and of the apparent levels of mutation in both the IGHJ genes and their associated IGHV genes, which allows conclusions to be drawn regarding both the accuracy and the completeness of the reported germline repertoire. We conclude that the reported IGHJ gene repertoire is essentially complete, and that misidentifications of IGHJ genes are likely to be relatively rare. Using the iHHMune-align program, the three immunoglobulin heavy chain genes can therefore be aligned with confidence, and IGHV, IGHD and IGHJ mutations could therefore be included in the analysis of CLL sequences. To gauge the consequences of such a change, we performed a re-analysis of published CLL sequence data. This showed that 59% of the sequences with between 4 and 10 IGHV gene mutations – which could either be classified as ‘mutated’ or ‘unmutated’ according to contending reports of appropriate cut-off values – had one or more mutations in their associated IGHD or IGHJ genes. Six of these sequences had five or more additional mutations in their IGHD and IGHJ genes. This therefore suggests that IGHD and IGHJ gene mutations should be included in the enumeration of immunoglobulin gene mutations for the prognosis of CLL.

2. Materials and methods

2.1. Compilation of databases

Rearranged cDNA sequences were collected from the EMBL database [7], and 5294 sequences remained after the exclusion of disease-related sequences. IGHV, IGHD, and IGHJ segments in each sequence were identified using the iHMMune-align program, an alignment tool based around a hidden Markov model of the rearranged variable region [4]. Sequences were only included in the analysis if the ends of the IGHJ alignments were at least three nucleotides downstream of the critical nucleotides that define the different, known allelic variants. An exception to this rule was made in the case of IGHJ4*03 where the most 5′ nucleotide of the reported sequence differs from the other IGHJ4 alleles. Inclusion of alignments to IGHJ4*03 required the sequence to terminate at least three nucleotides downstream of the second critical nucleotide that distinguishes IGHJ4*03 from the other IGHJ4 alleles. Where clonally-related sequences were found, on the basis of shared IGHV, IGHD and IGHJ genes, N nucleotides and junction length, only the least mutated sequences were retained. Duplicate, incomplete and ambiguous sequences were also removed from the database. Sequences having five or fewer mutations in the IGHV gene segment were defined as the LowMut database, while the remaining sequences made up the HighMut database.

The immunoglobulin sequences from both the LowMut and HighMut datasets were analyzed for the frequency of IGHJ gene usage. The extent of apparent somatic point mutations in each IGHJ gene and in their associated IGHV genes were then determined. Mismatches that could have arisen during PCR amplification, as a result of the use of degenerate primers, were identified by searches of literature identified in the sequence annotations, and by alignments of suspect sequences against multiple IGHJ germline sequences. Primer-mediated mismatches were identified where such 3′ mismatches perfectly matched an alternative IGHJ gene. Such mismatches were excluded from further analysis.

An additional set of 607 CLL heavy chain immunoglobulin gene sequences was retrieved from the EMBL sequence database [7]. Sequences were aligned against the germline IGHV, IGHD and IGHJ gene repertoires using the iHHMune-align program, and all nucleotide mismatches were identified as mutations. IGHD alignments were accepted where alignments included at least eight consecutive nucleotides, or included one mismatch in at least 10 nucleotides, or two mismatches in at least 12 nucleotides, or three mismatches in at least 14 nucleotides, or four mismatches in at least 16 nucleotides [4]. No sequence alignment included more than four IGHD mutations.

2.2. Identification of putative unreported IGHJ gene polymorphisms

The frequency of mismatches at each nucleotide of each IGHJ gene was calculated. If mismatches of different kinds

were seen at any position, this was taken as *prima facie* evidence of somatic point mutation. Where only one kind of apparent substitution was seen in multiple sequences, the sequence was considered to represent a putative polymorphism. To assess the completeness of the reported repertoire, the frequency distribution of different numbers of mutations of the IGHJ genes were determined, as well as the levels of mutation in their associated IGHV genes. Sets of sequences that aligned to a particular IGHJ gene or allele were analyzed by chi-squared test to see if they carried significantly more mutations than other IGHJ genes.

3. Results

3630 near full-length sequences were obtained from public immunoglobulin sequence databases which successfully aligned against the germline IGHJ gene repertoire. The IGHV genes associated with these IGHJ genes were determined, and 1308 sequences were identified in which the IGHV genes had five or fewer mutations. This restricted dataset was defined as the LowMut database. An additional 2296 more mutated sequences made up the HighMut database. The frequencies with which the different genes and alleles of the germline repertoire were utilized in the databases is presented as Table 1. 88.5% of the alignments in the LowMut dataset utilized just four IGHJ genes: IGHJ3*02 (14.3%), IGHJ4*02 (43.2%), IGHJ5*02 (10.4%) and IGHJ6*02 (20.6%). In contrast, the IGHJ3*01, IGHJ5*01 and IGHJ6*01 alleles together accounted for just 0.8% of the total LowMut alignments, while the two alternative IGHJ4 alleles accounted for a further 1.4% of alignments.

A comparison of the frequency distributions of the genes and alleles in the LowMut and HighMut databases, showed a number of significant differences. IGHJ1*01 ($p < 0.05$) and IGHJ3*01 ($p < 0.001$) were over-represented in the HighMut database, when compared with the LowMut database. This suggests that many of these HighMut alignments could be in error. Such alignment errors could easily result from the presence of mutations, because of similarities between IGHJ1, IGHJ4 and IGHJ5 and between IGHJ3*01 and IGHJ3*02.

Table 1
Frequency of IGHJ gene usage amongst the 1308 sequences of the LowMut database and the 2296 sequences of the HighMut database

J gene	Alleles	LowMut dataset		HighMut dataset	
		No.	%	No.	%
IGHJ1	01	16	1.2	53	2.3
IGHJ2	01	47	3.6	58	2.5
IGHJ3	01	4	0.3	47	2.0
	02	188	14.4	263	11.5
IGHJ4	01	13	1.0	31	1.4
	02	565	43.2	1031	44.9
	03	5	0.4	15	0.7
IGHJ5	01	1	0.1	2	0.1
	02	137	10.5	240	10.5
IGHJ6	01	5	0.4	8	0.3
	02	271	20.7	458	19.9
	03	56	4.3	90	3.9

Ten IGHJ6*02 sequences were identified that included a common 'mutation' of the third codon of the conserved WGXG amino acid motif that defines the 5′ end of the FR4 region of the IGHJ genes. It was concluded that these sequences aligned to a new putative allele IGHJ6*p04 (attactactactacggtatggacgtctggggcAaagggaccacggtcaccgtctcctca). No other candidate polymorphisms were identified on the basis of shared mutations, although both IGHJ3*02 and IGHJ6*03 included very small but significantly increased numbers of more mutated sequences ($p < 0.05$ and $p < 0.01$, respectively) (Table 2). These sequences were accepted as mutated sequences, rather than alignments to unreported polymorphisms because of a lack of common nucleotide mismatches, and because of the high numbers of mutations in the IGHV genes that were associated with these sequences. The IGHJ3*02 sequences with two or more IGHJ mutations were associated with IGHV genes having an average 2.5 mutations. The IGHJ6*03 sequences with one or more IGHJ mutations were associated with IGHV genes having an average 2.9 mutations. In each case, this level of IGHV mutation is significantly higher than the level of IGHV mutation in the other IGHJ3*02 and IGHJ6*03 alignments, suggesting that these are truly more mutated IGHJ sequences, rather than unreported polymorphisms.

In order to determine the implications of the inclusion of IGHD and IGHJ gene mutations on the categorization of CLL immunoglobulin sequences, we re-analyzed 607 published CLL sequences, scoring mutations in the full length VDJ gene rearrangements (Table 3). The inclusion of IGHD and IGHJ mutations in the total mutation load had major consequences in some cases. Overall, 51% of sequences included IGHD and/or IGHJ mutations. Forty-nine of 83 sequences (59%) with between 4 and 10 IGHV mutations, whose categorisation as 'mutated' or 'unmutated' varies according to the clinical definitions used, had IGHD and/or IGHJ mutations. Twenty-eight of the sequences (34%) had two or more mutations, and as many as four IGHD mutations and five IGHJ mutations were seen. Many of the sequences which would be considered 'unmutated' on the basis of IGHV mutations alone, would therefore be included in the 'mutated' prognostic category if IGHD and IGHJ mutations were included in the analysis.

4. Discussion

In 1999, two papers demonstrated an association between the presence or absence of substantial numbers of mutations in patients' immunoglobulin genes, and prognosis in B-CLL (1, 2). An association was also identified between CD38 expression and mutation number (1), but subsequent studies demonstrated that many cases are discordant for these two parameters and that CD38 is an independent prognostic

Table 2
Frequency of apparent mutations in IGHV and IGHJ gene segments amongst sequences of the LowMut database, according to their IGHJ gene alignment

IGHJ gene	Allele	Mean IGHV mutations	Mean IGHJ mutations	No. of sequences with 0, 1, 2 and ≥3 IGHJ mutations			
				0	1	2	≥3
IGHJ1	01	1.4	0.3	13	2	1	0
IGHJ2	01	1.7	0.4	38	8	0	1
IGHJ3	01	1.3	0.8	2	1	1	0
	02	1.7	0.3	152	28	2	6
IGHJ4	01	2.1	0.1	12	1	0	0
	02	1.7	0.2	486	64	13	2
	03	2.2	0.8	3	0	2	0
IGHJ5	01	1	0	1	0	0	0
	02	1.2	0.2	113	18	5	1
IGHJ6	01	1.2	0.2	4	1	0	0
	02	1.5	0.3	212 (212)[a]	44 (35)[a]	12 (11)[a]	3
	03	2.2	0.5	38	11	6	1
	p04	1.3	0.1	9[b]	1[b]	0	0
Total				1067	180	42	19

[a] Values in parenthesis represent results after removal of sequences that align to the putative polymorphism IGHJ6*p04.
[b] Values are not included in the total sequence counts, as they are also recorded as IGHJ6*02 sequences.

Table 3
Frequency of IGHD and IGHJ mutations in 607 CLL sequences with varying levels of point mutation in their IGHV

IGHV mutations	Combined IGHD and IGHJ mutations							
	0	1	2	3	4	≥5	Total	% IGHD/IGHJ mutations
≤3	202	21	2	1	0	0	226	10.6
4	7	4	1	1	0	0	13	46.2
5	6	1	2	3	0	0	12	50.0
6	3	2	1	0	0	0	6	50.0
7	6	3	1	1	1	1	13	53.9
8	4	3	1	2	0	0	10	60.0
9	4	5	1	4	1	2	17	76.5
10	4	3	2	0	0	3	12	66.7
≥11	61	67	62	42	26	40	298	76.2

marker [8]. Further studies showed an association between immunoglobulin gene (IGHV) mutational status and expression levels of ZAP-70 [9], a tyrosine kinase that has been found to be a strong, independent factor for predicting prognosis and requirement for therapy. There are a number of technical issues with ZAP-70 assessment that have limited its introduction as a prognostic factor, which together lead to striking interlaboratory variability in its measurement. For this reason, the IGHV mutational status remains a very important parameter [10], and a BIOMED-2 collaborative project has recently reported a standard protocol for the sequencing of immunoglobulin genes [11], in the expectation that the enumeration of mutations will become a part of routine B-CLL investigations. What has not yet been standardised is the analysis of the gene sequences, and the subsequent categorisation of patients.

The BIOMED-2 protocol amplifies full length CLL sequences because of the use of reverse primers designed to anneal to the 3′ ends of the IGHJ genes. This study was therefore conducted to first determine whether or not IGHJ gene alignments can be determined with sufficient accuracy for IGHJ gene mutations to be included in investigations of CLL. Such accuracy particularly requires the available repertoire to have been accurately and completely reported, and a detailed evaluation of the reported repertoire was therefore a major focus of this study.

The frequency with which each reported gene is represented in the expressed repertoire was determined by an analysis of large databases of both highly mutated and relatively unmutated VDJ gene rearrangements. The frequencies with which the different genes were utilized was broadly similar to that previously described from smaller studies [12,13], and ranged from 1.2% for IGHJ1 to 44.6% for IGHJ4. This variable utilization of the different genes may be a consequence of variations in the IGHJ gene recombination signal sequences (RSS) that are required for gene rearrangement, for the hierarchy of utilization of the different genes corresponds well with the similarity of the different RSS to the ‘consensus’ RSS sequence [14].

In contrast with previous studies, this study was large enough to allow us to accurately determine the frequency with which different IGHJ alleles are represented in the

expressed repertoire. The IGHJ3*01, IGHJ4*01, IGHJ5*01 and IGHJ6*01 alleles were all used far less frequently than alternative alleles. Together these alleles accounted for just 23 of the 1308 alignments in the LowMut database. Alignments to IGHJ3*01 ($n = 4$), IGHJ5*01 ($n = 1$) and IGHJ6*01 ($n = 5$) were very rare indeed. It is curious that these three rare alleles were all reported as part of the first description of the IGHJ locus [15]. When a large study of rearranged sequences subsequently found very few alignments to these alleles, genomic sequences were amplified from 39 individuals, but the alleles were not found [16]. Together they have been described as part of a rare haplotype [16], but until confirmation of the alleles is provided from genomic studies, the very existence of these alleles should be questioned.

The IGHJ3*01, IGHJ4*01 and IGHJ6*01 alleles each differ from the more common IGHJ3*02, IGHJ4*02 and IGHJ6*02 alleles by a single nucleotide. IGHJ5*01 differs from IGHJ5*02 by two nucleotides. It is therefore relatively easy for mismatches arising by mutation to lead to the misidentification of a utilized allele, and therefore to errors in the enumeration of mutations. This appears to be a particular risk for the sequences utilizing the IGHJ3 gene. The IGHJ3*01 allele was seen in 0.4% of sequences in the LowMut database, but in 2.0% of sequences in the HighMut database. The IGHJ3*01 alignments represents 2.1% of all IGHJ3 alignements in the LowMut database, but 15.2% of the IGHJ3 alignments in the HighMut database. The critical nucleotide that distinguishes these alleles appears to be a highly mutable nucleotide, and the data can be explained if many of these alignments to IGHJ3*01 are the result of somatic point mutations of IGHJ3*02-utilizing sequences.

It has been suggested that additional errors in the enumeration of mutations in CLL sequences could arise from alignment errors resulting from the use of an incomplete database of germline sequences [17]. We therefore evaluated the completeness of the reported IGHJ repertoire as part of this study. Only a single putative polymorphism was identified. The IGHJ6*p04 allele was identified in 3.5% of sequences that first aligned to the IGHJ6*02 gene. The IGHV genes that are associated with the IGHJ6*p04 sequences had few mutations (mean 1.3), and the IGHJ6*p04 sequences themselves had very few mismatches to the germline IGHJ6*02 sequence other than the mismatch that defines this candidate allele. This gave us considerable confidence in designating this sequence as a new allele. In fact, it has recently been identified as such by genomic screening. It has been reported to be carried by 2% of the Danish population, and has been designated as IGHJ6*04 [18]. This report confirms the validity of the bioinformatics approach used in the present study, and so if other unreported IGHJ gene polymorphisms remain to be discovered, they must be very rare or be largely confined to populations that are not well represented in sequence databases.

The completeness of the reported repertoire means that errors in the enumeration of mutation levels in IGHJ genes are only likely to result from misalignment of sequences against the germline repertoire. Similarities between IGHJ genes, particularly similarities within the FR4 regions of the genes, means that errors in the identification of IGHJ genes are likely be relatively common in sets of highly mutated sequences, particularly when there has been substantial removal of 5′ nucleotides by exonuclease activity. In CLL mutation analysis, errors could be mitigated if mutational counts in very rare alleles were accepted to potentially be in error by one or two ‘mutations’.

The results of this study suggest that it would be prudent to include IGHJ gene sequences in analyses of mutations in immunoglobulin genes of patients with chronic lymphocytic leukemia. In order to determine the implications of such a change for the clinical designation of sequence type, we analyzed the frequencies of IGHD and IGHJ mutations in a dataset of CLL sequences. Although most sequences included IGHD and/or IGHJ mutations, the consequences of this for the categorisation of CLL sequences depends upon the mutation cut-off point that defines the two prognostic categories.

Various definitions of the cut-off levels of mutation have been used in different studies of CLL mutations. It was originally believed that poor prognosis was associated with completely unmutated sequences, however sequences with very high homology were defined as ‘unmutated’. While Hamblin and colleagues defined ‘unmutated’ sequences as having ≥98% identity to a germline sequence [2], Damle and colleagues defined the cut-off as >98% identity [1]. These cut-offs were conceived as a way of discriminating between truly mutated sequences and sequences with a small number of apparent mutations resulting from alignments against an incomplete database of germline genes. The cut-offs were defined to account for variability between unreported alleles and the reported repertoire, based upon variability that had been noted between different reported immunoglobulin alleles [19]. It is now known that the ‘unmutated’ sequences include a large proportion of sequences that have truly accumulated small numbers of mutations in their IGHV genes [20,21], and it has been proposed that the progressive form of the disease is associated with sequences having 97% or greater identity [22], or even 95% identity with the germline [23]. Allowing for rounding, at one extreme the Damle group defines sequences with less than five mutations to be associated with progressive disease [1], while at the other extreme, sequences with as many as 16 mutations may be indicative of progressive disease [23].

Now that it is accepted that the mutational load should be properly determined, a more careful definition of the cut-off level of mutation is appropriate. If this is done, we believe it should identify a number of mutations, rather than a percentage of mutations. The only justification for the use of a percentage is to assist in the analysis of partial sequences, but analysis of partial sequences should be discouraged. The incorporation of IGHD and IGHJ genes into CLL mutation analysis, and the definition of the most predictive cut-off for such full length VDJ gene rearrangements can only increase

the prognostic power of immunoglobulin gene mutation analysis for the management of CLL.

Acknowledgements

This work was supported in part by a grant from the National Health and Medical Research Council.

Contributions: CEHL and AMC were responsible for the sequence analysis, KJLJ was responsible for the development and management of databases, while WAS and AMC were responsible for the overall design of the project.

References

[1] Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood 1999;94:1840–7.

[2] Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukaemia. Blood 1999;94:1848–54.

[3] Lefranc MP. Nomenclature of the human immunoglobulin heavy (IGH) genes. Exp Clin Immunogenet 2001;18:100–16.

[4] Lee CEH, Gaëta B, Malming HR, Bain ME, Sewell WA, Collins AM. Reconsidering the human immunoglobulin heavy chain locus. I. An evaluation of the expressed human IGHD gene repertoire. Immunogenetics 2006;57:917–25.

[5] Guddat LW, Shan L, Broomell C, Ramsland PA, Fan Z, Anchin JM, et al. The three-dimensional structure of a complex of a murine Fab (NC10. 14) with a potent sweetener (NC174): an illustration of structural diversity in antigen recognition by immunoglobulins. J Mol Biol 2000;302:853–72.

[6] Dunn-Walters DK, Spencer J. Strong intrinsic biases towards mutation and conservation of bases in human IgVH genes during somatic hypermutation prevent statistical analysis of antigen selection. Immunology 1998;95:339–45.

[7] Kanz C, Aldebert P, Althorpe N, Baker W, Baldwin A, Bates K, et al. The EMBL nucleotide sequence database. Nucleic Acids Res 2005;33:1.

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

[9] Hamblin TJ. Predicting progression-ZAP-70 in CLL New. Eng J Med 2004;351:856–7.

[10] Tobin G, Rosenquist R. Prognostic usage of V-H gene mutation status and its surrogate markers and the role of antigen selection in chronic lymphocytic leukaemia. Med Oncol 2005;22:217–28.

[11] van Dongen JJM, Langerak AW, Bruggemann M, Evans PAS, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003;17:2257–317.

[12] Wasserman R, Ito Y, Galili N, Yamada M, Reichard BA, Shane S, et al. The pattern of joining (JH) gene usage in the human IgH chain is established predominantly at the B precursor cell stage. J Immunol 1992;149:511–6.

[13] Yamada M, Wasserman R, Reichard BA, Shane S, Caton AJ, Rovera G. Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. J Exp Med 1991;173:395–407.

[14] Ramsden DA, Baetz K, Wu GE. Conservation of sequence in recombination signal sequence spacers. Nucleic Acids Res 1994;22:1785–96.

[15] Ravetch JV, Siebenlist U, Korsmeyer S, Waldmann T, Leder P. Structure of the human immunoglobulin μ locus: characterization of embryonic and rearranged J and D genes. Cell 1981;27:583–91.

[16] Mattila PS, Schugk J, Wu H, Makela O. Extensive allelic sequence variation in the J region of the human immunoglobulin heavy chain gene locus. Eur J Immunol 1995;25:2578–82.

[17] Pekova S, Baran-Marszak F, Schwarz J, Matoska V. Mutated or non-mutated? Which database to choose when determining the IgVH hypermutation status in chronic lymphocytic leukemia? Haematologica 2006;91:ELT01.

[18] Ohm-Laursen L, Larsen SR, Barington T. Identification of two new alleles, IGHV3-23*04 and IGHJ6*04, and the complete sequence of the IGHV3-h pseudogene in the human immunoglobulin locus and their prevalences in Danish Caucasians. Immunogenetics 2005;57:621–7.

[19] Matsuda F, Shin EK, Nagaoka H, Matsumura R, Haino M, Fukita Y, et al. Structure and physical map of 64 variable segments in the 3'0.8-megabase region of the human immunoglobulin heavy-chain locus. Nat Genet 1993;3:88–94.

[20] Davis ZA, Orchard JA, Corcoran MM, Oscier DG. Divergence from the germ-line sequence in unmutated chronic lymphocytic leukemia is due to somatic mutation rather than polymorphisms. Blood 2003;102:3075.

[21] Forconi F, Sahota SS, Lauria F, Stevenson FK. Revisiting the definition of somatic mutational status in B-cell tumors: does 98% homology mean that a V(H)-gene is unmutated? Leukemia 2004;18:882–3.

[22] Krober A, Seiler T, Benner A, Bullinger L, Bruckle E, Lichter P, et al. V-H mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukaemia. Blood 2002;100:1410–6.

[23] Lin K, Sherrington PD, Dennis M, Matrai Z, Cawley JC, Pettitt AR. Relationship between p53 dysfunction, CD38 expression, and IgV(H) mutation in chronic lymphocytic leukaemia. Blood 2002;100:1404–9.