Somatic hypermutation signature in B-cell low-grade lymphomas

Lorraine Tracey,[1] Mohit Aggarwal,[1] Mónica García-Cosio,[2] Raquel Villuendas,[1] Patrocinio Algara,[3] Margarita Sánchez-Beato,[1] Abel Sánchez-Aguilera,[1] Juan F. García,[4] Antonia Rodríguez,[5] Francisca I. Camacho,[6] Nerea Martínez,[1] Elena Ruiz-Ballesteros,[3] Manuela Mollejo,[3] and Miguel Á. Piris[1]

[1]Molecular Pathology Programme, Centro Nacional de Investigaciones Oncológicas, Madrid; [2]Pathology Department, Hospital Ramon y Cajal, Madrid; [3]Genetics and Pathology Departments, Hospital Virgen de la Salud, Toledo; [4]Pathology Department, MD Anderson International Spain, Madrid; [5]Hematology Department, Hospital Doce de Octubre, Madrid and [6]Pathology Department, Hospital Universitario de Getafe, Madrid, Spain

ABSTRACT

Background
Immunoglobulin gene somatic hypermutation is a biologically relevant and clinically useful prognostic factor in different types of low-grade B-cell lymphomas, including chronic lymphocytic leukemia, mantle cell lymphoma and splenic marginal zone lymphoma.

Design and Methods
With the aim of identifying surrogate markers of somatic hypermutation, a combined investigation of IgV_H mutational status and expression profiles of 93 samples from patients with small B-cell lymphoma was performed.

Results
The analysis identified an somatic hypermutation signature of genes involved in the regulation of gene transcription, DNA repair and replication, and chromosome maintenance. Eight of these genes were subjected to protein analysis using tissue microarrays, for a set of 118 cases. We found a clear link between RAD51C and CDK7 protein expression and somatic hypermutation status, in that positive expression of either marker was significantly associated with a mutated status ($p<0.003$). We also found that positive expression of TFDP1 and POLA was significantly associated with ongoing somatic hypermutation ($p<0.001$). To assess the potential clinical applicability of these somatic hypermutation markers, we studied a series of cases of mantle cell lymphoma included in a tissue microarray. The expression of RCC1 and CDK7, separately and together, was found to be significantly associated with longer overall survival.

Conclusions
An somatic hypermutation signature has been identified for different types of small B-cell lymphoma. This has a potential mechanistic and diagnostic value.

Key words: lymphomas, somatic hypermutation, profiling.

Citation: Tracey L, Aggarwal M, García-Cosio M, Villuendas R, Algara P, Sánchez-Beato M, Sánchez-Aguilera A, García JF, Rodríguez A, Camacho FI, Martínez N, Ruiz-Ballesteros E, Mollejo M, and Piris MA. Somatic hypermutation signature in B-cell low-grade lymphomas. Haematologica 2008; 93:1186-1194. doi: 10.3324/haematol.12999

The online version of this article contains a supplemental appendix.

Acknowledgments: we are indebted to Isabel Fernandez, Mercedes Navarrete, Mar Lopez, Paloma de la Cueva, Giovanna Roncador and Cristina Romero of the CNIO for technical assistance. Special thanks are extended to the Tumor Bank Network of the CNIO for providing all frozen and paraffin-embedded tissue used in this study, and to all the hospitals that collaborated in this work.

Funding: this work was supported by grants from the Ministerio de Ciencia y Tecnología (SAF2004-04286, SAF2005-00221), Fundacion La Caixa, Ministerio de Sanidad y Consumo (FIS PI051623) and Consejería de Sanidad de la Junta de Comunidades de Castilla-La Mancha (#02031), Spain. LT received support from grants from the CNIO and the Higher Education Authority of Ireland through the Department of Haematology/Institute of Molecular Medicine, St James Hospital, Dublin, Ireland. MA is supported by European Commission grants from the Marie Curie Fellowship. AR was the recipient of a grant from the Fondo de Investigaciones Sanitarias (FIS), Ministerio de Sanidad y Consumo, Spain. ERB received support from the Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha, Spain.

Manuscript received March 3, 2008. Revised version arrived on April 17, 2008. Manuscript accepted April 23, 2008.

Correspondence:
Miguel A. Piris, Centro Nacional de Investigaciones Oncológicas, Melchor Fernández Almagro, 3, Madrid 28029, Spain.
E-mail: mapiris@cnio.es

Introduction

A diverse antibody repertoire is critical in order to maintain immune system capability. This diversity is developed and maintained by virtue of a variety of genetic rearrangements and changes that occur in B cells, including V(D)J recombination, immunoglobulin (Ig) class switching and somatic hypermutation (SHM). Initially, the antibody repertoire is produced through recombination of the V, D and J exons of the Ig gene during B-cell differentiation. On completion of this process the mature B cells migrate to the secondary lymphoid organs where antigen is encountered. Stimulation by antigen is, in turn, responsible for further diversification of the antibody repertoire by inducing class-switch recombination and SHM, a process by which point mutations are introduced into the variable regions of the heavy and light Ig chains.[1]

In humans, SHM of Ig genes occurs at rates of 10^{-3} to 10^{-4} mutations per base pair per generation, which is up to six orders of magnitude greater than the spontaneous mutation rate of most other genes.[2,3] The majority of changes are single base substitutions (most commonly transitions) in the Ig variable region, starting 150-200 base pairs downstream of the promoter and continuing about 1.5 kb downstream.[4,5] Although mutations occur throughout the rearranged variable regions, RGYW and WRCY (R: A or G; Y: T or C; W: A or T) motifs are preferential targets (hotspots) for SHM. Higher-order structures or differences in local sequences may also play a role in targeting SHM to specific bases, as not all potential hotspots in the same region are affected by SHM.[6,7] Likewise, the vast majority of genes are not affected by SHM although similar hotspots may be present. For example the Ig constant region gene, which does not exhibit SHM, is found only a few kilobases from the variable region genes.[5]

Activation-induced cytidine deaminase (AID) is expressed by activated B cells in the germinal centers of peripheral lymphoid organs, and is critical to the process of SHM. However, the mechanisms of SHM are not fully understood. For example, the molecules and processes responsible for DNA unwinding, which allow AID access to the DNA template, and the signaling pathways involved in the process are unknown. High AID expression is insufficient to explain SHM in chronic lymphocytic leukemia (CLL), since strong AID mRNA expression is associated with unmutated *IgV$_H$* gene status.[8]

Besides the biological relevance of SHM, there is a broad base of evidence to suggest that *IgV$_H$* SHM is a clinically significant phenomenon in various types of B-cell lymphoma, as demonstrated by studies reporting that cases of hypermutated CLL[9] and splenic marginal zone lymphoma (SMZL) have a better prognosis,[10] while cases of hypermutated mantle cell lymphoma (MCL) display specific clinicopathological features (leukemic course, longer survival).[11-14]

There were two aims of the current study. First, we wished to identify a set of markers that could be analyzed by immunohistochemical assays, since this could have a potential clinical value for analyzing SHM status. Secondly, we wanted to understand the mechanisms and markers of SHM more thoroughly, since genes associated with high SHM status could help to elucidate the mechanisms of SHM and the co-factors involved with AID. These mechanisms and markers are currently insufficiently characterized, except, to some extent, in the case of AID, whose role was mentioned above, and ZAP70, which is associated with unmutated *IgV$_H$* genes in CLL.[15] However, this finding is confined to CLL: no association between ZAP70 and SHM has been observed in cases of MCL.[16]

We used expression profiling in 93 samples from patients with B-cell lymphoma to identify genes that could serve as markers of high SHM levels, a characteristic that has prognostic significance. The study focuses on small B-cell lymphomas in which the presence of Ig somatic hypermutation has been reported to be a biological variable of clinical significance, such as CLL, MCL and SMZL, as noted above.[9-14] Furthermore, expression profiling was used in patients' samples to identify genes that may play a role in the process of SHM.

Finally, the association of several markers with overall survival was assessed.

Design and Methods

Case selection

All cases included in this study were selected from the medical records of member hospitals of the Spanish Tumor Bank Network. All paraffin-embedded and frozen tissue samples were collected through the protocols of the Tumor Bank of the *Centro Nacional de Investigaciones Oncológicas* (CNIO). Tissue distribution and analysis was performed under the supervision of the Hospitals' ethical committees. All samples were centrally reviewed by a panel of pathologists and diagnosed using uniform criteria based on clinical, histological, immunophenotypic and molecular characteristics. A total of 93 cases of small B-cell lymphoma, comprising 24 SMZL, 33 MCL and 36 CLL, were available for expression profiling. Controls consisting of five reactive lymph nodes, five normal spleen samples and mantle zone B cells from a tonsillectomy specimen were also included. Protein expression of significant genes in the germinal center was checked in six reactive lymph nodes. Immunohistochemical validation was carried out in 49 MCL, 34 SMZL and 35 CLL cases. Thirty-six percent of the cases analyzed by immunohistochemistry were unique to the validation analyses and were not included in the microarray analysis. Sixty MCL cases included in a tissue microarray were used for analysis of the correlation of marker expression with overall survival. These cases were almost uniquely used for the survival analysis with overlaps of only 10% with the microarray analysis and of 30% with the marker validation analysis.

Analysis of IgV$_H$ somatic hypermutation and ongoing mutations in patients' samples

DNA was extracted from frozen tissue blocks using proteinase-K and purified by phenol-chloroform extraction. Rearranged *IgV$_H$* genes were amplified by using a

semi-nested polymerase chain reaction (PCR) method, as described previously.[10,13,17] The first round of the PCR was performed using a mixture of six framework region 1 (FR1) V_H family-specific primers and two consensus primers for the J_H gene. The second round of PCR was performed in six separate reactions with one of the six V_H FR1 primers and J_H internal primers. Direct sequencing was performed in an ABI PRISM 310 or 3700 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) from both strands, using the same primers as in the amplification. Mutations were identified by comparison with the germ-line sequence (Ig BLAST and V BASE sequence directory). Cases with a high mutational load were defined as those in which the homology between the amplified sequence and the original sequence was less than 98%, while those cases with less than 2% variation with respect to the original sequence were considered to have a low mutational status. In all cases, analysis of IgV_H SHM was performed in duplicate.

Ongoing mutation was analyzed in a subset of 31 patients with more than 2% SHM, comprising 14 CLL, 4 MCL and 13 SMZL cases. The V_H fragments were amplified as outlined above and the PCR products were purified and cloned into the pCR2.1-TOPO vector (Invitrogen, CA, USA). Several colonies were sequenced from each case. The sequences obtained were compared with that of the wild type and the number of stable and ongoing mutations was determined. A case was considered to have ongoing mutation if at least two of its colonies had a mutation that was not present in the other colonies. For evaluation of intraclonal heterogeneity, a mutation was considered to be confirmed if it was observed more than once in V_H gene molecular clones from the same tumor specimen. Only confirmed mutations were considered as evidence of intraclonal heterogeneity. Unconfirmed mutations, a substitution mutation observed in only one of the V_H gene molecular clones from the same tumor specimen, were disregarded because they could have been caused by a Taq polymerase error.

RNA isolation, cDNA microarray target preparation and hybridization

Total RNA was extracted from frozen tumor samples using the Trizol reagent (Invitrogen) and RNA was purified and treated with RNase-free DNase I using the RNeasy kit (Qiagen Inc., Valencia, CA, USA). Next, 1-5 μg of target RNA was amplified using T-7 *in vitro* transcription[18] and 2.5 μg of amplified RNA were directly labeled with cyanine 5-conjugated dUTP or cyanine 3-conjugated dUTP (Amersham, Uppsala, Sweden). The reference sample used was 2.5 μg of amplified RNA from the Universal Human Reference RNA (Stratagene, La Jolla, CA, USA).

Microarray studies were carried out using the CNIO OncoChip and labeling and hybridizations were performed as previously described.[18-20] The cloned sequences of all the genes included in the OncoChip and the reproducibility of the expression data (measured by quantitative PCR) of multiple genes have been verified.[18,20-23] Scanning and image analysis were performed using a Scanarray 5000 XL (GSI Lumonics, Kanata, Ontario, Canada) and GenePix Pro Software (Axon Instruments Inc., Union City, CA, USA), respectively.

Data analysis and normalization of microarray data

Raw microarray data were processed as previously described.[19,20,22] Data from CLL and SMZL cases were normalized against the average expression of each gene from five reactive lymph node and five normal spleen samples, respectively. Normalization was carried out only for those genes for which at least 50% of the data were available in the control samples. All other genes were excluded from analysis. Data from MCL samples were normalized against data from tonsillar mantle cells purified using magnetic beads.[24] Raw and normalized data of the genes studied are presented in the *Online Supplementary Appendix*.

Statistical analyses

To identify the genes of importance in distinguishing cases with a high or low IgV_H SHM load, Welch's t-statistic, which does not assume equal variances,[25] was calculated, and false discovery rates were obtained from tests of 100,000 permutations. Genes were considered significant if they were differentially expressed between patients with high and low SHM burdens to a false discovery rate < 0.20.

The biological functions of genes were assigned using the Gene Ontology (www.geneontology.org) and Genecards[26,27] (*http://bioinformatics.weizmann.ac.il/cards*) databases. Gene pathways were analyzed with the help of Gene Set Enrichment Analysis (GSEA) version 2.0.1. The gene-set database included Biocarta pathways,[28] clusters of functionally related, coregulated genes identified by our unsupervised clustering and molecular signatures defined for lymphoma in the Staudt molecular signature database (*http://lymphochip.nih.gov/signaturedb*).[29]

Fisher's exact test was used to investigate associations between SHM marker expression and SHM status and with ongoing mutation status in all tumor types. The statistical significance of relationships between Ki67 and SHM markers was evaluated using Pearson's χ^2 test. Differences were considered statistically significant for p values <0.05. Overall survival was analyzed in patients with different levels of expression of SHM markers using the Mantel-Cox test. Univariate receiver operation curves (ROC) were analyzed to identify the specificity of marker genes. All statistical analyses were done using SPSS version 13.0 (SPSS Inc., Chicago, IL, USA).

Immunohistochemistry

Antibodies against CDK7, DP1 (TFDP1), HMGB2, POLA, PRIM1, RAD51C, RCC1 (CHC1) and RFC4 were used to determine protein expression employing a previously described protocol.[24,30] Suppliers and dilutions of antibodies for immunohistochemistry were as follows: Ki67 at 1:50 (MIB1 antibody from DAKO, Glostrup, Denmark), CDK7 at 1:225, HMGB2 at 1:10, RCC1 (CHC1) at 1:150, RFC4 at 1:150, RAD51C at 1:30 and POLA at 1:100 (Santa Cruz Biotechnology, Santa Cruz, CA, USA) and DP1 (TFDP1) at 1:60, and DNA primase (PRIM1) at 1:35 (NeoMarkers, Lab Vision, Fremont, CA, USA). To analyze SHM status and marker association

with ongoing mutation, samples were evaluated semi-quantitatively by a panel of pathologists. Given that the 60 cases of MCL used for survival analysis were included in a tissue microarray amenable to automatic quantification, the expression of the markers studied was ranked as strong, weak or negative, as determined by the ARIOL semi-automated computerized training system (*http://www.aicorp.com/products/02path.htm*). The cellular location of each marker was determined using information from published studies and defined to the ARIOL training system. On the basis of this, the system was trained by an expert team of technicians and pathologists to classify the staining on tissue microarray slides as strong, weak or negative for each marker selection. The automated selection was then verified to ensure the robustness of results. Each cellular location was counted as a single point and, depending on its intensity, was assigned to the strong, weak or negative category. The operation is similar to flow cytometry in tissue sections, in which quantitative measurements (e.g., number of cells in a core from a patient's sample) and qualitative measurements (e.g., intensity of staining) are made. Cores with low quality staining were considered not evaluable.

Results

Genes associated with somatic hypermutation in clinical samples

Expression analysis of 93 cases of small B-cell lymphoma identified genes significantly associated with a high level of SHM (patients had been classified as having either a low or a high mutational status according to whether they had ≤2% or > 2% SHM, respectively; false discovery rate <0.20). The top 58 genes found to be upregulated in cases with a high mutational status are involved in cell cycling and DNA replication (*PRIM1, RFC4, HOXB7, CCNA2, AIM2, BUB3, PCNA, CDK2, JUNB, MCM3, CHEK1, RFC3, POLA, WEE1*), DNA repair (*CDK7, RAD51C, POLS, HMG2, RAD54B, FRAP1, PRKDC*), chromosome condensation (*CHC1*) and transcription regulation (*TFDP1, POLA, MED6, TCEA2, E2F5*). As a validation step, we found the same genes in the top ranks using GSEA (Table 1).

Expression of somatic hypermutation markers in germinal center cells

SHM is an active process known to take place in the germinal center and so we would expect genes associated with SHM to be strongly expressed there.[17] Such expression would confirm these genes as potential markers of SHM. We selected eight genes involved in DNA repair and replication and transcription regulation. Their expression was studied in six normal lymph node samples as a way of confirming whether the SHM markers observed here in small B-cell lymphomas were simultaneously expressed by B cells in the germinal center. The markers analyzed were *DP1* (*TFDP1*), *HMGB2, POLA, PRIM1, RAD51C, RCC1* (*CHC1*) and *RFC4*. All were strongly expressed in germinal centers from reactive lymph node and tonsil (Figure 1). As a validation step, the genes overexpressed for high SHM were significantly enriched in the germinal center signature in the MSig Database.[29]

Correlation of expression of DNA repair and replication markers with somatic hypermutation status

A semi-quantitative analysis was performed by a panel of pathologists on the eight SHM markers selected in the 118 cases of CLL, MCL and SMZL. The conventional 2% cut-off between SHM-positive and SHM-negative cases was used. Two markers were significantly overexpressed in mutated cases (*CDK7*, p=0.04; *RAD51C*, p=0.04), irrespective of the diagnosis (Figure 1). Expression of either *CDK7* or *RAD51C* was significantly associated with a high SHM status (p<0.003) (Table 2). Specificities for *CDK7* and *RAD51C* were 86.67% and 81.36% respectively, so we may conclude that both genes were highly specific and therefore qualified as potential SHM markers.

A relationship was found also between some of the SHM markers analyzed, especially *TFDP1, CDK7, RCC* and *PRIM1,* and Ki67 expression (Table 2B).

Correlation of expression of DNA repair and replication markers with ongoing somatic hypermutation

The relationship between the expression level of the selected SHM markers and the presence of ongoing SHM was analyzed (Figure 1). Ongoing mutations were not observed in 13 CLL cases, but were present in 2/4 MCL and 8/12 SMZL cases (Table 3A). Correlation analysis between the markers of SHM and ongoing

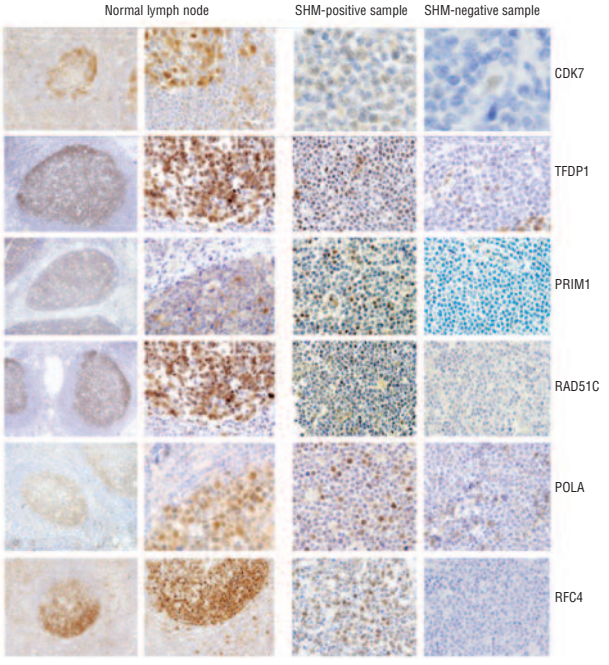

Figure 1. Immunohistochemical staining of markers significantly associated with SHM status, ongoing SHM and prognosis and their expression in normal lymph node.

Table 1. Genes expressed in cases of mutated small B-cell lymphoma.

Gene	*GSEA score*	*FDR t-test*	*Description of functions*	*Pathways*
RFC4	-2.80057	0.109046	Accessory protein required for elongation of primed DNA templates by DNA polymerase delta and DNA polymerase epsilon	DR
HOXB7	-2.54487	0.108268	Sequence-specific transcription factor involved in cell proliferation and differentiation	TF, DR
HCLS1	-2.77987	0.107942	Cytoskeleton organization in B cells and leukemic B cells	
CCNA2	-2.7495	0.107942	Promotes both cell cycle G1/S and G2/M transitions	cell cycle
GARS	-2.75294	0.107743	Encodes glycyl-tRNA synthetase, one of the aminoacyl-tRNA synthetases that charge tRNA with their cognate amino acids	transcription
POLS	-2.7621	0.1073	DNA polymerase that is probably involved in DNA repair. Also, the encoded protein may be required for sister chromatid adhesion	DNA repair
AIM2	-2.83162	0.1073	Plays a putative role in tumorigenic reversion and may control cell proliferation. Interferon-gamma induces expression of AIM2	cell cycle
TFAM	-2.78792	0.104791	Mitochondrial transcription factor that is a key activator of mitochondrial transcription and a participant in mitochondrial genome replication	TF, DR
TFDP1	-2.79054	0.104497	Transcription factor, DNA replication, Cell cycle regulation	TF, DR, cell cycle
BUB3	-2.87171	0.099578	Encodes a protein involved in spindle checkpoint function	DR
POLR2G	-2.87838	0.099578	Encodes the seventh-largest subunit of RNA polymerase II	transcription
HDAC8	-3.11226	0.087322	Histone acetylation/deacetylation alters chromosome structure and affects transcription factor access to DNA	transcription
RINZF	-3.01756	0.087627	Zinc finger protein	TF
MLL3	-2.7755	0.084392	Encodes a nuclear protein possessing histone methylation activity and is involved in transcriptional coactivation	transcription
PCNA	-3.09846	0.082301	A cofactor of DNA polymerase delta, helps increase the processivity of leading strand synthesis during DNA replication. Ubiquitinated in response to DNA damage and is involved in the RAD6-dependent DNA repair pathway	DR, DNA repair
CBFB	-3.01497	0.082301	Transcription factor that master-regulates a host of genes specific to hematopoiesis	DR
CDK2	-2.93673	0.082301	A catalytic subunit of the cyclin-dependent protein kinase complex, whose activity is restricted to the G1-S phase, and is essential for cell cycle G1/S phase transition	cell cycle
TOB2	-3.1206	0.082301	Family of antiproliferative proteins involved in the regulation of cell-cycle progression	cell cycle
JUNB	-1.22195	0.081626	Binds directly to retinoblastoma protein; interacts with BRCA1 and may have a role in the regulation of cell proliferation and differentiation	cell cycle
TNFRSF8	-3.48039	0.082301	Receptor expressed by activated T- and B-cells. It leads to the activation of NF-κB and is a positive regulator of apoptosis	apoptosis
BTF3	-3.08539	0.081626	Forms a stable complex with RNA polymerase IIB and is required for transcriptional initiation	transcription
DEK	-3.16899	0.081626	Binds to cruciform and superhelical DNA and induces positive supercoils into closed circular DNA and is also involved in splice site selection during mRNA processing	transcription
MCM3	-3.09084	0.081626	The hexameric protein complex formed by MCM proteins is a key component of the prereplication complex	DR
RRM1	-3.09764	0.081557	Ribonucleoside-diphosphate reductase, essential for the production of deoxyribonucleotides before DNA synthesis in S phase of dividing cells.	DR
KRAS2	-3.11601	0.081557	Oncogene	DR. DNA repair
TCEB1	-3.11433	0.081557	Activates elongation by RNA polymerase II by suppressing transient pausing of the polymerase at many sites within transcription units	transcription
MKI67	-3.17159	0.075623	Ki67	cell cycle
FGFR3	-3.20483	0.075357	Extracellular portion of the protein interacts with fibroblast growth factors, setting in motion a cascade of downstream signals, ultimately influencing mitogenesis and differentiation	cell cycle
SRI	-3.27728	0.073197	May play a role in tumor cell metastasis	
PRKRIR	-3.26681	0.073197	Interferon-inducible double-stranded RNA-dependent inhibitor, repressor of p58	
HMG2	-3.32223	0.072405	DNA end-joining processes of DNA double-strand breaks repair and V(D)J recombination	DNA repair
RAD54B	-3.09276	0.07151	Implicated in homologous recombination and repair of DNA. Homozygous mutations of this gene were observed in primary lymphoma	DNA repair
BTG1	-3.29686	0.070316	Involved in a t(8;12)(q24;q22) in CLL. Negatively regulates cell proliferation.	cell cycle
CHEK1	-3.37342	0.069019	Prevents abrogation of the G2 checkpoint in p53+/+ cells	cell cycle
FRAP1	-3.44359	0.069019	mTOR; mediates cellular responses to stresses such as DNA damage and nutrient deprivation	DNA repair
RFC3	-3.71964	0.064778	Elongation of primed DNA templates by DNA polymerase delta and DNA polymerase epsilon	DR
PSA	-3.49318	0.061038	Kallikreins are implicated in carcinogenesis and some have potential as novel cancer and other disease biomarkers.	
RAD51C	-3.41514	0.057975	DNA repair, homologous recombination repair of double-strand DNA breaks, DNA replication	DR, DNA repair
SIRT1	-3.47165	0.052774	Regulates epigenetic gene silencing and suppresses recombination of rDNA	DR
PTPN1	-3.4374	0.048666	Regulates a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation	DR
PAXIP1L	-3.59922	0.047629	Plays a critical role in maintaining genome stability, condensation of chromatin and progression through mitosis	DR
POLA	-3.64993	0.047085	Polymerase, DNA synthesis	DR
HDAC1	-3.52541	0.047085	Interacts with retinoblastoma tumor-suppressor protein and this complex is a key element in the control of cell proliferation and differentiation	cell cycle
UBE2D2	-3.52188	0.046836	Ubiquitin-conjugating enzyme E2D 2	
FBXO5	-3.84689	0.046836	Mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex	cell cycle
TMPO	-3.64424	0.046776	Consistent with its regulation by E2F, LAP2alpha overexpression in primary tumors is correlated with tumour-proliferation rate	DR
CDK7	-3.60075	0.045508	Cell-cycle regulation, Positive regulation of cell growth, DNA repair, transcription initiation	DNA repair
RBM12	-3.76853	0.031672	Protein that contains several RNA-binding motifs, potential transmembrane domains, and proline-rich regions	transcription
PPP1R8	-3.08753	0.033225	Cleaves specific sites in A+U-rich regions of RNA	transcription
CARD12	-4.03519	0.030722	Positive regulator of apoptosis	apoptosis
WEE1	-3.98373	0.030043	Coordinates the transition between DNA replication and mitosis by protecting the nucleus from activated CDC2 kinase	DR
PRIM1	-4.0843	0.029198	DNA primase, DNA replication, formation of Okazaki fragments	DR
MFAP1	-4.14503	0.021039	E2F1 interacts with MFAP1 chromatin	TF
PRKDC	-4.23498	0.021039	XRCC7, DNA non-homologous end joining, double-strand-break repair, V(D)J recombination	DNA repair
E2F5	-4.42853	0.017798	E2F family plays a crucial role in the control of cell cycle and action of tumor suppressor proteins	TF
TCEA1	-4.75182	0.010258	Transcription elongation factor A	TF
PLAGL2	-4.84224	0.009459	Leukemic oncogenes, zinc finger protein that recognizes DNA and RNA	TF
MED6	-5.29143	0.002932	RNA polymerase II transcription-factor activity	TF

GSEA: Gene-set enrichment analysis; FDR: false discovery rate; DR: DNA replication; TF: transcription factor.

mutation demonstrated that *TFDP1* and *POLA* were significantly associated with ongoing mutation (p=0.038 and p=0.014, respectively) and that expression of both markers was highly correlated with a status of ongoing mutation (p<0.001) (Table 3B).

Association of DNA replication marker expression with overall survival in mantle cell lymphoma

Two markers, *RCC1* and *CDK7*, were significantly associated with improved overall survival. The combined expression of *RCC1* and *CDK7* yielded an even more significantly longer overall survival in SHM (p=0.005) (Figure 2). Ki67 expression was not significantly associated with outcome in this series.

Discussion

The gene expression analysis presented here of almost 100 cases of small B-cell lymphoma, representing sub-types in which SHM is of clinical and prognostic significance, identified a large number of genes that may be surrogate markers of the SHM process. As might be expected, these genes are selectively expressed in reactive germinal center cells, the specific cell subset in which SHM takes place.[17] Subsequent analysis of selected markers involved in DNA replication and repair revealed some of protein markers to be significantly associated with SHM status and ongoing SHM. We obtained a molecular signature, comprising genes involved in the cell cycle, regulation of transcription, and DNA repair and replication, which mainly recognizes mechanistic aspects of the process in the low-grade lymphoma types. It did not include genes whose expression is only associated with SHM in specific tumor types, such as ZAP70[16] or lipoprotein lipase (LPL) in CLL.[15] Studies of ZAP70 in a wide spectrum of cases of B-cell lymphoma have shown that the relationship between its expression and SHM is confined to CLL; there is no association between ZAP70 expression and SHM in MCL.[16] Our results from this series confirm these observations, since ZAP70 was only associated with SHM in cases of CLL (*data not shown*).

Conceptually, SHM must consist of at least two main steps: (i) unwinding of DNA during replication or transcription to allow access to the SHM machinery; (ii) mutation of the DNA by a polymerase during replication or by a DNA repair enzyme. A large proportion of the genes identified as SHM markers or as being involved in the SHM mechanism (Tables 1 and 2, respectively) play a role in DNA repair, replication, transcription and cell cycling. Most of these genes had not previously been associated with SHM and so represent novel findings. The composition of the SHM signature identified here seems to confirm that SHM is indeed associated with a characteristic set of changes in the cell machinery in charge of cell cycling, transcription and DNA repair and replication, when considering different B-cell lymphoma types together. This contrasts with the findings when considering only CLL, in which SHM has been mainly found to be associated with changes in B-cell receptor signaling genes, such as ZAP70 or others.[31-33] Although theoretically, malignant transformation could have taken place after silencing of the hypermutation process in a post-germinal center B cell, these findings seem to show that in fact hypermutated tumors, when considered as a whole, preserve a distinctive signature. An additional finding supporting this observation is the increased expression of some of the genes composing this SHM signature (*TFDP1* and *POLA*) in cases with ongoing SHM, in which the SHM machinery is still active. Nevertheless, the limited number of cases analyzed for ongoing SHM impedes generalization regarding the value of the genes associated with ongoing SHM in conditions such as CLL, in which ongoing SHM is more infrequent.

DNA replication appears to play a crucial role in SHM given that a large number of critical genes involved in the process are overexpressed in cases with high SHM status. These genes include *PRIM1*, *POLA, RFC4, CDK7, TFDP1* and *RCC1*. POLA is a replication polymerase in a complex with DNA primase (PRIM1). It is expressed in hypermutating cells but not in resting B cells,[34] as our study of ongoing mutation confirms (Table 3B). PRIM1, which is responsible for synthesizing the small RNA primers during discontinuous DNA replication, is significantly associated with SHM in microarray analysis and tends to show higher protein expression in cases with > 2% SHM (Table 2). CHC1 (RCC1) is involved in regulating the onset of chromosome con-

Table 2A. Protein marker expression association with somatic hypermutation status in cases of small B-cell lymphoma.

Marker	Cases analyzed	High SHM status (> 2% SHM) Positive	High SHM status (> 2% SHM) Negative	Low SHM status (≤2% SHM) Positive	Low SHM status (≤2% SHM) Negative	p
CDK7	103	13	30	8	52	0.0477
RCC1 (CHC1)	108	8	37	6	57	NS
DP1 (TFDP1)	96	22	19	23	32	NS
PRIM 1	104	24	20	24	36	NS
HMG2	100	34	8	42	16	NS
POLA	97	23	18	39	17	NS
RAD51C	101	16	26	11	48	0.0401
RFC4	97	23	19	30	25	NS

Marker	Cases analyzed	1+ markers positive	Negative	1+ markers positive	Negative	p
CDK7 and RAD51C	101	21	21	12	47	0.0025

NS: not significant. Missing values were considered not valuable.

Table 2B. Relation between Ki67 and somatic hypermutation markers.

	Ki67	CDK7	RCC1 (CHC1)	PRIM1	RAD51c	TFDP1	HMG2	RFC4	POLA
Pearson	1	0.322	0.320	0.368	-0.005	0.405	0.037	0.039	-0.057
Significance		0.007	0.007	0.001	0.962	0.000	0.758	0.751	0.640
Number	68	68	68	68	68	68	68	68	68

Table 3A. Analysis of ongoing somatic hypermutation in small B-cell lymphomas.

Diagnosis	*Wild type VH*	*Clones evaluated*	*Stable mutations*	*Additional mutations*	*Confirmed additional mutations*	*Single unconfirmed mutations*	*Intraclonal variation (additional mutations/total clones)*	*Ongoing mutations*
CLL01	VH1-46	11	14	0				No
CLL02	VH4-31	7	6	0				No
CLL04	VH3-21	26	6	0				No
CLL05	VH3-21	10	7	0				No
CLL18	VH2-5	8	11	1		1	0.124	No
CLL34	VH3-35	5	8	0				No
CLL34	VH4-59	9	5	4		4	0.44	No
CLL36	VH3-9	7	6	1		1	0.142	No
CLL19	VH6-1	10	9	0				No
CLL20	VH4-34	7	11	1		1	0.142	No
CLL22	VH5-51	7	12	0				No
CLL27	VH3-9	15	15	6		6	0.4	No
CLL30	VH3-7	11	20	0				No
MCL02	VH3-7	8	6	9	2	5	1,125	Yes
MCL10	VH3-73	10	7	0				No
MCL15	VH3-23	22	8	11	4	2	0.5	Yes
MCL19	VH3-7	10	5	1		1	0.1	No
SMZL06	VH1-2	9	5	6	2		0.66	Yes
SMZL06	VH3-11	8	12	9	1	5	1,125	Yes
SMZL11	VH1-2	8	9	16	4	1	2	Yes
SMZL16	VH3-30	19	4	35	7	1	1.84	Yes
SMZL20	VH3-9	17	5	19	4		1.1	Yes
SMZL21	VH1-2	12	9	22	6	2	1.83	Yes
SMZL22	VH4-59	9	10	19	3	5	2.1	Yes
SMZL22	VH5-51	5	4	4		4	0.8	No
SMZL29	VH1-2	10	5	0				No
SMZL01	VH3-23	14	16	1		1	0.07	No
SMZL17	VH4-34	8	16	0				No
SMZL33	VH1-2	10	14	21	3	3	2.1	Yes

densation in the S phase. Replication factor C (RFC4) is a DNA polymerase accessory protein whose function is to elongate primed DNA templates through the action of DNA polymerase. It acts as a *clamp loader* to enable the polymerase to bind to DNA.

CDK7 and *RAD51C* showed significantly greater expression in cases with higher SHM, and both genes were significantly associated with SHM in validation studies using protein expression. Furthermore nodal marginal zone lymphomas with a high SHM burden also showed increased *CDK7* mRNA expression (*data not shown*). Both genes can activate p53, which leads to cell-cycle arrest during which DNA repair can progress. CDK7, a cyclin-dependent kinase, is a mediator of cell-cycle progression through activation by binding to cyclins. Additionally, CDK7 is a component of the transcription factor TFIIH, which helps control transcription by RNA polymerase II and possesses DNA repair and helicase activities. RAD51C is involved in the homologous recombination repair pathway of double-stranded DNA breaks that arise during DNA replication or are induced by DNA-damaging agents, and in meiotic recombination. Its overexpression may contribute to SHM by repairing the DNA after mutations have been induced, perhaps during DNA replication, and by allowing the cells to progress through the cell cycle. Finally, in terms of DNA damage repair, microarray analysis also revealed overexpression of PRKDC in IgV_H mutated cases. PRKDC is a serine/threonine-protein kinase involved in DNA non-homologous end joining, which is required for double-strand-break repair and V(D)J recombination. PRKDC is also involved in the modulation of transcription. However, the role of PRKDC is unclear, as SHM is known to be able to proceed essentially unaffected by deficient DNA-PK activity.[35]

Interestingly, these findings show that some specific components of protein complexes involved in DNA

Table 3B. Protein marker expression association with ongoing somatic hypermutation in cases of small B-cell lymphoma.

		Ongoing SHM		*Stable SHM*		
Marker	*Cases analyzed*	*Positive*	*Negative*	*Positive*	*Negative*	*p*
CDK7	21	0	6	6	9	NS
RCC1 (CHC1)	21	2	4	3	12	NS
TFDP1 (DP1)	20	5	0	6	9	0.038
PRIM 1	21	6	0	8	7	NS
HMG2	21	6	0	13	2	NS
POLA	20	6	0	5	9	0.014
RAD51C	20	3	2	7	8	NS
RFC4	20	4	1	8	7	NS

Marker	Cases analyzed	2 positive markers	1+ negative markers	2 markers positive	1+ negative markers	*p*
TFDP1 and POLA	21	6	0	2	13	0.0005

NS: not significant.

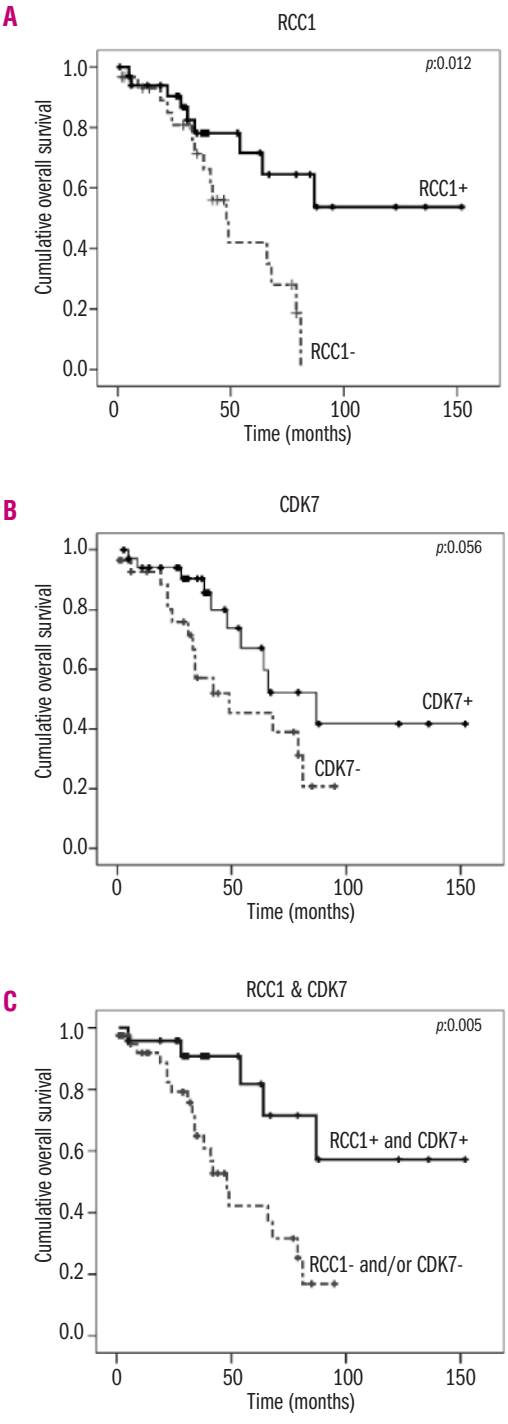

Figure 2. Association between SHM markers and overall survival in MCL. (A) RCC1 expression is significantly associated with longer overall survival in MCL (B) CDK7 expression is significantly associated with longer overall survival in MCL (C) Combined expression of RCC1 and CDK7 is highly significantly associated with longer overall survival in MCL (overall survival at 60 months: 72% for double positive RCC1 + CDK7 *vs.* 36% for all the other combinations).

repair and transcription control, such as CDK7 and RAD51C, appear to be selectively increased in association with SHM. The observations require additional confirmation, but do raise an interesting hypothesis to be explored in functional assays.

B-cell translocation gene 1 (*BTG1*) is a negative regulator of cell proliferation whose expression was increased in mutated cases. The t(8;12)(q24;q22) *BTG1/MYC* translocation has been identified in CLL. TFDP1 has been associated in this study with ongoing mutation. It is a component of the E2F/DP transcription factor complex and a regulator of a number of genes whose products are involved in cell-cycle regulation and DNA replication. Its expression can be induced by p53 and it is also responsible for indirect activation of p53 through its role in the cell cycle. HMG2 may also be involved in the final ligation step in DNA end-joining processes of DNA double-strand-break repair, and V(D)J recombination. p53-dependent DNA-repair may also play a role in the upregulation of a variety of transcription factors that are associated with high SHM status in patients' samples.

This SHM signature does not include AID, an enzyme required for SHM but largely absent in multiple lymphoproliferative processes with increased SHM, as shown by Pasqualucci and co-workers.[36]

Finally, a potential clinical application of these findings was explored, since some of the markers identified here as being associated with SHM are significantly associated with improved overall survival in patients, with MCL, underlining the fact that SHM is a prognostic factor in the major types of low-grade B-cell lymphoma, including MCL.[14] Taking into consideration all these findings, studies can be developed to examine the role of the specific genes detected here in the generation of IgV_H SHM in small B-cell lymphomas, and to analyze their potential prognostic utility in larger series of standardized patients.

Authorship and Disclosures

LT: experimental procedures, contribution to and/or analysis of microarray data, contribution of essential new reagents, manuscript preparation; MA: experimental procedures, contribution to and/or analysis of microarray data, contribution of essential new reagents, manuscript preparation; MG-C: immunohistochemistry evaluation; RV: data analysis and interpretation, contribution to and/or analysis of microarray data; PA: experimental procedures, data analysis and interpretation; AS-A: experimental procedures; JFG: immunohistochemistry evaluation; AR, FIC, NM, ER-B, MM: contribution to and/or analysis of microarray data; MAP: immunohistochemistry evaluation, contribution to and/or analysis of microarray data, contribution of essential new reagents, manuscript preparation.

The authors reported no potential conflicts of interest.

References

1. Frieder D, Larijani M, Tang E, Parsa JY, Basit W, Martin A. Antibody diversification: mutational mechanisms and oncogenesis. Immunol Res 2006;35: 75-88.
2. Rajewsky K, Forster I, Cumano A. Evolutionary and somatic selection of the antibody repertoire in the mouse. Science 1987;238:1088-94.
3. Wang L, Jackson WC, Steinbach PA, Tsien RY. Evolution of new nonantibody proteins via iterative somatic hypermutation. Proc Natl Acad Sci USA 2004;101:16745-9.
4. Rada C, Milstein C. The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. Embo J 2001;20:4570-6.
5. Muller-Hermelink HK, Greiner A. Molecular analysis of human immunoglobulin heavy chain variable genes (IgVH) in normal and malignant B cells. Am J Pathol 1998;153:1341-6.
6. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim Biophys Acta 1992;1171:11-8.
7. Kinoshita K, Honjo T. Linking class-switch recombination with somatic hypermutation. Nat Rev Mol Cell Biol 2001;2:493-503.
8. Heintel D, Kroemer E, Kienle D, Schwarzinger I, Gleiss A, Schwarzmeier J, et al. High expression of activation-induced cytidine deaminase (AID) mRNA is associated with unmutated IGVH gene status and unfavourable cytogenetic aberrations in patients with chronic lymphocytic leukaemia. Leukemia 2004; 18:756-62.
9. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 1999;94:1848-54.
10. Algara P, Mateo MS, Sanchez-Beato M, Mollejo M, Navas IC, Romero L, et al. Analysis of the IgV(H) somatic mutations in splenic marginal zone lymphoma defines a group of unmutated cases with frequent 7q deletion and adverse clinical course. Blood 2002;99:1299-304.
11. Bertoni F, Conconi A, Cogliatti SB, Schmitz SF, Ghielmini M, Cerny T, et al. Immunoglobulin heavy chain genes somatic hypermutations and chromosome 11q22-23 deletion in classic mantle cell lymphoma: a study of the Swiss Group for Clinical Cancer Research. Br J Haematol 2004; 124:289-98.
12. Cogliatti SB, Bertoni F, Zimmermann DR, Henz S, Diss TC, Ghielmini M, et al. IgV H mutations in blastoid mantle cell lymphoma characterize a subgroup with a tendency to more favourable clinical outcome. J Pathol 2005;206:320-7.
13. Camacho FI, Algara P, Rodriguez A, Ruiz-Ballesteros E, Mollejo M, Martinez N, et al. Molecular heterogeneity in MCL defined by the use of specific VH genes and the frequency of somatic mutations. Blood 2003; 101:4042-6.
14. Lai R, Lefresne SV, Franko B, Hui D, Mirza I, Mansoor A, et al. Immunoglobulin VH somatic hypermutation in mantle cell lymphoma: mutated genotype correlates with better clinical outcome. Mod Pathol 2006;19:1498-505.
15. Wiestner A, Rosenwald A, Barry TS, Wright G, Davis RE, Henrickson SE, et al. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. Blood 2003;101:4944-51.
16. Carreras J, Villamor N, Colomo L, Moreno C, Ramón y Cajal S, Crespo M, et al. Immunohistochemical analysis of ZAP-70 expression in B-cell lymphoid neoplasms. J Pathol 2005;205:507-13.
17. Kuppers R, Zhao M, Hansmann ML, Rajewsky K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. Embo J 1993;12:4955-67.
18. Tracey L, Villuendas R, Ortiz P, Dopazo A, Spiteri I, Lombardia L, et al. Identification of genes involved in resistance to interferon-alpha in cutaneous T-cell lymphoma. Am J Pathol 2002;161:1825-37.
19. Tracey L, Villuendas R, Dotor AM, Spiteri I, Ortiz P, Garcia JF, et al. Mycosis fungoides shows concurrent deregulation of multiple genes involved in the TNF signaling pathway: an expression profile study. Blood 2003;102:1042-50.
20. Tracey L, Spiteri I, Ortiz P, Lawler M, Piris MA, Villuendas R. Transcriptional response of T cells to IFN-α: changes induced in IFN-α sensitive and resistant cutaneous T cell lymphoma. J Interf Cyt Res 2004;24: 185-95.
21. Martinez-Delgado B, Melendez B, Cuadros M, Alvarez J, Castrillo JM, Ruiz De La Parte A, et al. Expression profiling of T-cell lymphomas differentiates peripheral and lymphoblastic lymphomas and defines survival related genes. Clin Cancer Res 2004; 10:4971-82.
22. Tracey L, Perez-Rosado A, Artiga MJ, Camacho FI, Rodriguez A, Martinez N, et al. Expression of the NF-κB targets BCL2 and BIRC5/survivin characterizes small B-cell and aggressive B-cell lymphomas, respectively. J Pathol 2005;206:123-34.
23. Ruiz-Ballesteros E, Mollejo M, Rodriguez A, Camacho FI, Algara P, Martinez N, et al. Splenic marginal zone lymphoma. Proposal of new diagnostic and prognostic markers identified after tissue and cDNA microarray analysis. Blood 2005; 106: 1831-8.
24. Martinez N, Camacho FI, Algara P, Rodriguez A, Dopazo A, Ruiz-Ballesteros E, et al. The molecular signature of mantle cell lymphoma reveals multiple signals favoring cell survival. Cancer Res 2003;63:8226-32.
25. Miller RG. Beyond Anova. New York, Chapman & Hall. 1997.
26. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics 1998;14:656-64.
27. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. Trends Genet 1997; 13:163.
28. Wei G, Twomey D, Lamb J, Schlis K, Agarwal J, Stam RW, et al. Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. Cancer Cell 2006;10:331-42.
29. Shaffer AL, Rosenwald A, Hurt EM, Giltnane JM, Lam LT, Pickeral OK, et al. Signatures of the immune response. Immunity 2001;15:375-85.
30. Rodríguez A, Martínez N, Camacho FI, Ruiz-Ballesteros E, Algara P, García JF, et al. Variability in the degree of expression of phosphorylated IkappaBalpha in chronic lymphocytic leukemia cases with nodal involvement. Clin Cancer Res 2004;10:6796-806.
31. Kienle DL, Korz C, Hosch B, Benner A, Mertens D, Habermann A, et al. Evidence for distinct pathomechanisms in genetic subgroups of chronic lymphocytic leukemia revealed by quantitative expression analysis of cell cycle, activation, and apoptosis-associated genes. J Clin Oncol 2005; 23:3780-92.
32. Rodriguez A, Villuendas R, Yanez L, Gomez ME, Diaz R, Pollan M, et al. Molecular heterogeneity in chronic lymphocytic leukemia is dependent on BCR signaling: clinical correlation. Leukemia 2007;21:1984-91.
33. Kienle D, Benner A, Krober A, Winkler D, Mertens D, Buhler A, et al. Distinct gene expression patterns in chronic lymphocytic leukemia defined by usage of specific VH genes. Blood 2006;107:2090-3.
34. Winter DB, Gearhart PJ. Altered spectra of hypermutation in DNA repair-deficient mice. Philos Trans R Soc Lond B Biol Sci 2001;356:5-11.
35. Bemark M, Sale JE, Kim HJ, Berek C, Cosgrove RA, Neuberger MS. Somatic hypermutation in the absence of DNA-dependent protein kinase catalytic subunit (DNA-PK(cs)) or recombination-activating gene (RAG)1 activity. J Exp Med 2000;192: 1509-14.
36. Pasqualucci L, Guglielmino R, Houldsworth J, Mohr J, Aoufouchi S, Polakiewicz R, et al. Expression of the AID protein in normal and neoplastic B cells. Blood 2004;104:3318-25.