Improved classification of leukemic B-cell lymphoproliferative disorders using a transcriptional and genetic classifier

B-cell chronic lymphoproliferative disorders (B-CLPD) encompass a group of hematologic tumors that often present with leukemic involvement.[1] Their heterogeneity and the lack of relatively specific diagnostic markers for most of these diseases make their diagnosis challenging, especially in cases that only have blood involvement or when histology is not available. With the currently used immunophenotypic and molecular markers, around 10% of B-CLPD cases remain unclassifiable and are categorized as B-CLPD, not otherwise specified (B-CLPD, NOS). Few recurrent gene mutations and chromosomal abnormalities have been documented in some entities: *BRAF* and *MYD88* mutations in hairy cell leukemia (HCL) and lymphoplasmacytic lymphoma (LPL), respectively,[2,3] in addition to the recurrent 7q31-q32 deletion in splenic marginal zone lymphoma (SMZL).[1] However, none of them are diagnostic hallmarks of any particular entity. Gene expression profiling studies have recognized specific signatures that identify most common hematological neoplasms.[4,5] Based on these results we postulated that the analysis of the gene expression profiling (GEP) of a large series of leukemic B-CLPD could identify specific signatures for each leukemic disease entity. These signatures could be useful for the classification of cases with undetermined diagnosis (B-CLPD, NOS). In this study, we have investigated the GEP of a large series of leukemic lymphoid neoplasms and identified specific gene signatures for most entities that were validated in an independent cohort. We have also derived and validated a simplified quantitative polymerase chain reaction (qPCR)-based 8-gene assay that reliably recognized these entities and could assist in the diagnosis in routine practice, particularly in atypical cases and B-CLPD, NOS.

We initially studied the global GEP (Affymetrix U133+2.0 arrays, GEO GSE79196) of 159 well-defined leukemic B-cell neoplasms with highly purified tumor cells (mean 96%, range 79-100%; training series) (*Online Supplementary Figure S1*). The diagnosis was established in the peripheral blood (PB) based on cytology, immunophenotype, and, in some cases, cytogenetics according to the World Health Organization (WHO) classification.[1] These cases included 54 chronic lymphocytic leukemia (CLL), 30 conventional and 24 leukemic non-nodal mantle cell lymphoma (cMCL and nnMCL, respectively), 12 follicular lymphoma (FL), 4 HCL, 4 HCL variant (HCLv), 4 LPL, 23 SMZL, and 4 splenic diffuse red pulp lymphomas (SDRPL). Subsequent biopsies obtained in 70 cases, including all SMZL and SDRPL were reviewed (*Online Supplementary Table S1*). We investigated 30 additional cases considered leukemic B-CLPD, NOS because they did not have the diagnostic features characteristic of a specific disease entity, lacked spleen histology in splenomegalic disorders, and/or had unusual features for a given entity such as CLL with *BCL2* rearrangements. All these B-CLPD, NOS also had high

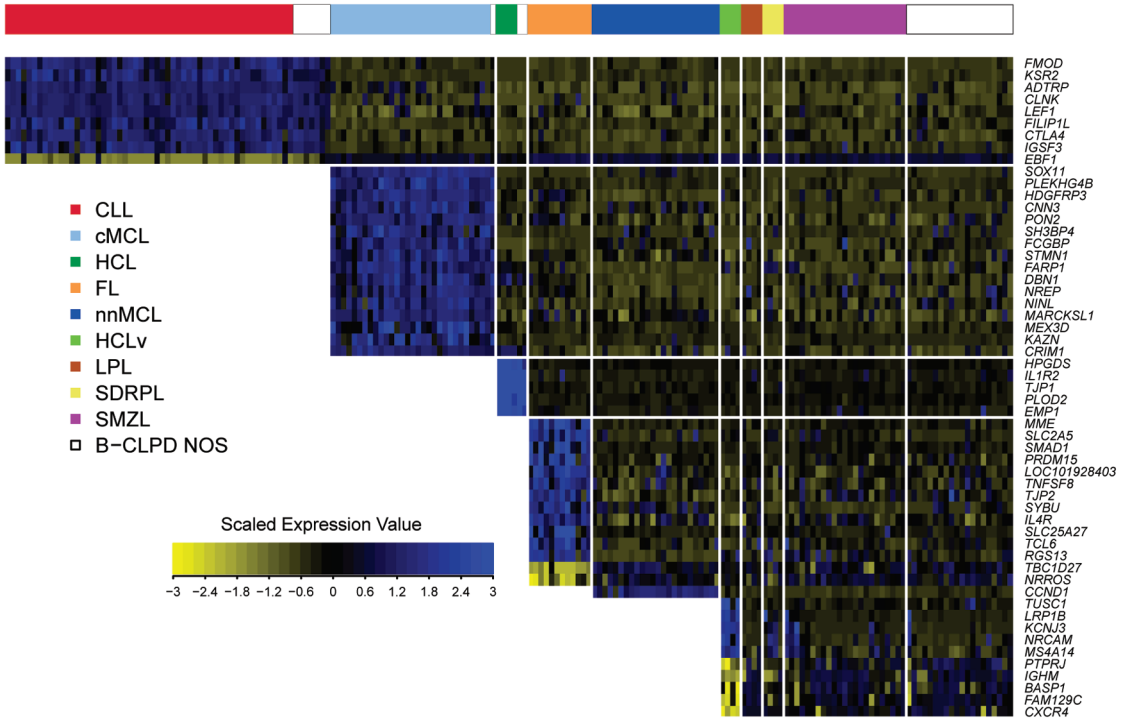

Figure 1. Molecular diagnostic signatures of 159 leukemic B-cell chronic lymphoproliferative disorders (B-CLPD) and 30 B-CLPD, not otherwise specified (NOS). Heatmap representing the genes identified at each step of the GEP55. Each column represents a B-CLPD patient and each row represents a gene. Most genes showed differential high expression (blue), whereas only 1, 2 and 5 genes showed low expression for chronic lymphocytic leukemia (CLL), follicular lymphoma (FL), and hairy cell leukemia variant (HCLv), respectively. In the right part of each entity the B-CLPD, NOS (white) are classified according to their expression. For each gene only relevant samples of the multistep approach are shown. cMCL: conventional mantle cell lymphoma; HCL; hairy cell leukemia; nnMCL: non-nodal mantle cell lymphoma; LPL: lymphoplasmacytic lymphoma; SDRPL; splenic diffuse red pulp lymphoma; SMZL: splenic marginal zone lymphoma.

tumor cell content (mean 96%, range 94-100%). The study was approved by the Institutional Review Board of the Hospital Clinic of Barcelona.

The GEP of this series was first analyzed by an unsupervised hierarchical clustering approach. The cases were distributed into 2 main branches: 1 exclusively comprised CLL, whereas the other included MCL, FL, HCL, and HCLv, which were grouped together according to their respective diagnosis. The 30 B-CLPD, NOS cases were distributed among the different well-defined entities, suggesting that they may correspond to these respective diseases (*Online Supplementary Figure S2*). To identify the specific gene expression signatures of the different entities, we developed a 6-step GEP predictor model using 55 genes (GEP55; Figures 1 and 2A, *Online Supplementary Methods*, and *Online Supplementary Table S2*). Nine genes were specific for CLL and separated the 54 CLL from the other disorders. Subsequently, 30 cMCL were identified using 16 genes, then HCL was separated using 5 genes (*Online Supplementary Figure S3*), and the following step classified 12 FL based on the expression of 14 genes. Once cMCL and HCL were classified, *CCND1* expression identified the 24 nnMCL. Finally, the lack of *CXCR4* expression, together with the deregulation of another 9 genes, specifically identified the 4 HCLv. We were unable to determine distinct signatures for the 31 LPL, SMZL, and SDRPL, thus these 3 diseases were considered as a miscellaneous group. This latter finding correlates with the overlapping cytology and immunophenotype of these diseases.[6] In summary, the GEP55 model discriminated most entities with high sensitivity (77-100%) and specificity (89-100%), except for the miscellaneous group (*Online Supplementary Table S2*).

We next applied the GEP55 model to the subset of 30 B-CLPD, NOS. Ten (33%) were classified by GEP as 7 CLL, 1 cMCL, and 2 HCL (Figure 3A,B and *Online Supplementary Table S3*). To determine whether additional genetic and molecular studies could improve the classification of the B-CLPD, NOS cases we performed the mutational analysis of *BRAF, MAP2K1, MYD88, NOTCH1, NOTCH2, SF3B1*, and *TP53* by Sanger sequencing[7] and investigated chromosomal alterations by cytogenetics and/or Affymetrix SNP 6.0 microarrays.[8] Combining the GEP with this molecular information, 10/20 of the remaining B-CLPD, NOS were classified: 6 SMZL based on the presence of *NOTCH2* mutations, 2 SMZL based on concurrent trisomy 3 and 12 and absence of paraprotein, and 2 LPL based on *MYD88* mutations (*Online Supplementary Figure S4*). Although 10/30 (33%) B-CLPD, NOS remained unclassifiable, we could rule out

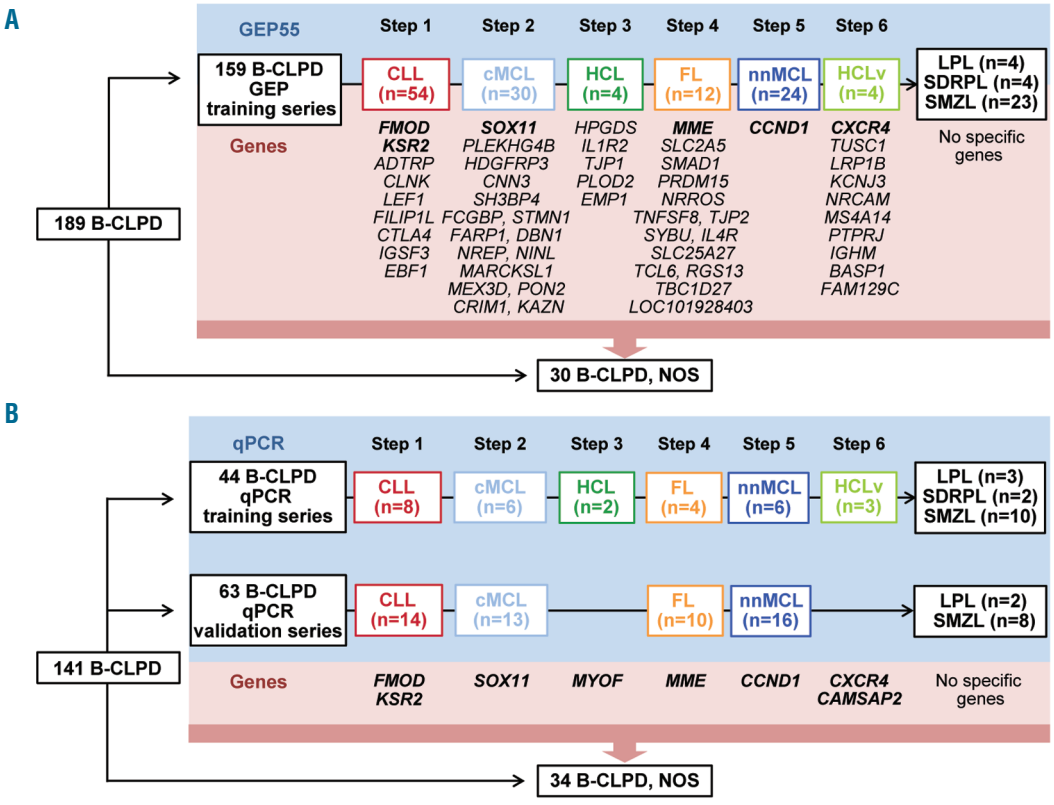

Figure 2. Schematic representation of the predictor models. (A) Box flow chart of the gene expression profiling GEP55. The successive steps and the number of cases per entity used to build the model are shown in boxes. In the lower part, the genes used for the classification in each step are indicated. In bold, the genes that were later selected for quantitative polymerase chain reaction (qPCR) analysis. This GEP55 was also used to classify 30 B-cell chronic lymphoproliferative disorders, not otherwise specified (B-CLPD, NOS) cases. (B) Box flow chart of the qPCR model. In the upper part, the successive steps and the number of cases from the qPCR training and independent validation series. In the lower part, the number of cases of each entity used to build the qPCR model and the genes used at each step. This qPCR model was also used to classify 34 B-CLPD cases. CLL: chronic lymphocytic leukemia; cMCL: conventional mantle cell lymphoma; HCL; hairy cell leukemia; FL: follicular lymphoma; nnMCL: non-nodal mantle cell lymphoma; HCLv; hairy cell leukemia variant. LPL: lymphoplasmacytic lymphoma; SDRPL; splenic diffuse red pulp lymphoma; SMZL: splenic marginal zone lymphoma.

the diagnosis of the well-defined entities (CLL, FL, cMCL, nnMCL, HCL, and HCLv; Figure 3B). Overall, the combination of GEP55 and additional genetic data helped to accurately classify 20/30 (67%) B-CLPD, NOS.

To develop a simple qPCR assay that could be used in routine diagnosis we reanalyzed the GEP of the 159 cases. We selected a subset of genes representative of each disease and measured their expression by qPCR (Fluidigm BioMark 48.48 Dynamic Array, Fluidigm®) in 44 samples previously examined by microarrays (Figure 2B, *Online Supplementary Methods, Online Supplementary Table S4*, and *Online Supplementary Figures S1* and *S5*). Subsequently, we selected 8 genes that corresponded to the minimal number that reliably recognized the different entities, except the miscellaneous group. The final qPCR 8-gene step-wise predictor included *FMOD* and *KSR2* for CLL, *SOX11* for cMCL, *MYOF* for HCL, *MME* for FL, *CCND1* for nnMCL, and *CXCR4* and *CAMSAP2* for HCLv (*Online Supplementary Table S5* and *Online Supplementary Figure S6*). Next, we applied this 8-gene predictor model to an independent validation series of 63 leukemic B-CLPD (mean tumor cell content 81%, range 60-99.9%). We confirmed the classification of CLL, cMCL, FL, and nnMCL in all cases (confidence interval = (94.3%-100%); Figure 2B and *Online Supplementary Figure S1*). As expected, LPL and SMZL remained in the miscellaneous group. We did not observe differences in the classification based on the tumor cell content (60-80% *vs.* 81-99.7%). Moreover, we performed serial dilution experiments (100%-70%-40%-20%) in 8 samples (2 CLL, 2 cMCL, 1 HCL, 1 FL, 1 nnMCL, and 1 HCLv). All entities were correctly identified with 70% tumor cells, and CLL, cMCL, and nnMCL were correctly classified even with 40% tumor cells. The 8-gene predictor model was then applied to 34 additional B-CLPD, NOS (mean 81% tumor cells, range 60-99.9%) and it classified 19/34 (56%) as 14 CLL, 1 FL, 3 HCLv, and 1 nnMCL. Additional molecular/genetic features helped to classify 4 additional cases in the miscellaneous group as LPL based on *MYD88* mutations (Figure 3C and *Online Supplementary Table S6*). Finally, 11/34 (32%) cases remained unclassified. In summary, this simple 8-gene qPCR-model together with molecular analysis accurately classified 100% of the well-defined entities in addition to 68% B-CLPD, NOS.

Taken together, these results show that the GEP55 and the simplified 8-gene qPCR models contribute to the pre-

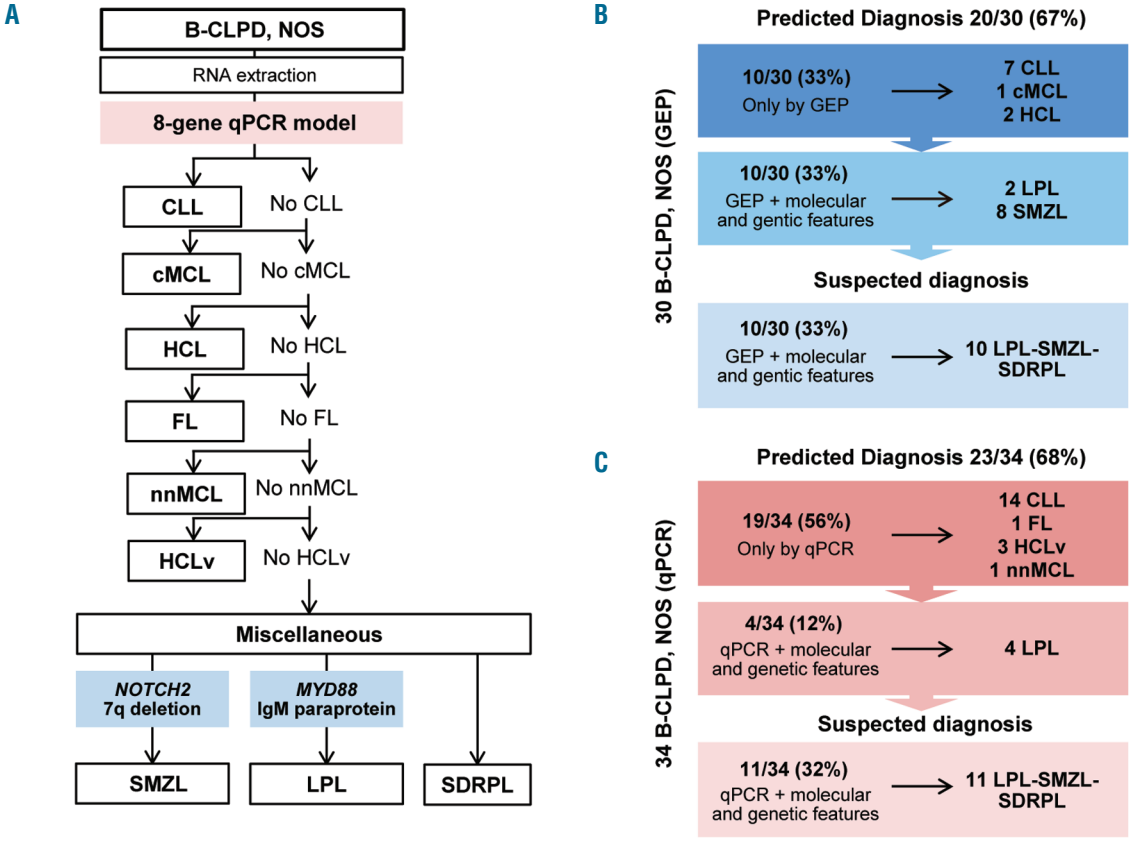

Figure 3. Classification flow charts of the B-cell chronic lymphoproliferative disorders, not otherwise specified (B-CLPD, NOS) patients from the gene expression profiling (GEP) and quantitative polymerase chain reaction (qPCR) series. (A) Flow diagram of the steps of B-CLPD classification, based on gene expression and other molecular/genetic data such as 7q deletion and/or *NOTCH2* mutation (for SMZL) and *MYD88* mutation for LPL classification. (B) Classification of 30 B-CLPD, NOS cases of the GEP series: i) 10 cases were classified by GEP only; 7 as CLL (including 3 with *BCL2* rearrangements) and 1 as cMCL (lacking t(11;14)(q13;q32) but positive for the IGL/*CCND2* rearrangement) and 2 cases in which a definitive diagnosis could not be established in the initial sample (due to the low number of circulating neoplastic cells) were classified as HCL by GEP alone, ii) 10 cases by combining GEP with molecular/genetic features; 2 as LPL (*MYD88* mutations), and 8 as SMZL (*NOTCH2* mutations and 7q deletion in 6 cases and 3q and chromosome 12 gains in the remaining 2 cases), and iii) 10 cases remained in the miscellaneous group of LPL, SMZL, and SDRPL. (C) Classification of 34 B-CLPD, NOS cases using the 8-gene qPCR model: i) 19 cases were classified by qPCR only; 14 as CLL (9 with t(14;19)(q32;q13) involving IGH and *BCL3* loci), 1 as FL (with add(14)(q32) but without t(14;18)(q32;q21) by fluorescence *in situ* hybridization, 3 HCLv, and 1 nnMCL, ii) 4 cases classified as LPL by combining qPCR with molecular features (*MYD88* mutations), and iii) 11 cases remained in the miscellaneous group of LPL, SMZL, and SDRPL. RNA; ribonucleic acid; CLL: chronic lymphocytic leukemia; cMCL: conventional mantle cell lymphoma; HCL; hairy cell leukemia; FL: follicular lymphoma; nnMCL: non-nodal mantle cell lymphoma; HCLv; hairy cell leukemia variant; LPL: lymphoplasmacytic lymphoma; SDRPL; splenic diffuse red pulp lymphoma; SMZL: splenic marginal zone lymphoma.

cise diagnosis of B-CLPD, particularly in cases with atypical/inconclusive phenotypes or unusual features (B-CLPD, NOS) with the exception of the miscellaneous group. Interestingly, atypical cases such as CLL with atypical phenotypes or carrying the t(14;18)(q32;q21) or t(14;19)(q32;q13), cyclin D1-negative cMCL and leukemic FL without the t(14;18)(q32;q21) were clearly assigned to their specific categories. Moreover, both expression assays could assign 45% of the B-CLPD, NOS to a specific disease based solely on gene expression data. The addition of a small subset of genetic/molecular studies (i.e., *NOTCH2* mutation and 7q deletion in SMZL, and *MYD88* mutation in LPL) refined the diagnosis of the miscellaneous group, increasing the total number of classified cases by up to 67%. However, as *MYD88* mutations with or without an immunoglobulin M (IgM) band are also detected in CLL[9] and *NOTCH2* mutations in FL[10] and MCL,[8] a definitive diagnosis without the information of gene expression would have been difficult. Thus, the contribution of GEP or qPCR assays allowed us to exclude all these diseases, and instead allocate these cases into the miscellaneous group (LPL, SMZL, SDRPL).

Overall, our GEP analysis expanded upon previous reports, highlighting that most leukemic B-CLPD have specific expression profiles, even if they present with atypical features. The most distinct was CLL (high *FMOD*, *LEF1*, and *KSR2*),[11] followed by cMCL (high *SOX11*, *DBN1*, and *HDGFRP*), nnMCL (high *CCND1*),[12,13] FL (high *MME*), and HCL (high *IL1R2*, *ANXA1*, and *MYOF*).[14] Of interest was our finding of a distinct gene signature for HCLv (low *CXCR4* and high *CAMSAP2*), a disease that still is a provisional entity in the WHO classification,[1,15] however, further studies are needed to confirm this finding due to the small number of HCLv analyzed.

In summary, we describe a "ready to use" 8-gene qPCR prediction assay which can be helpful in the classification of leukemic B-CLPD in routine practice, particularly in cases with atypical or non-specific features. The combination of this assay with additional molecular and genetic studies may improve the diagnosis of these entities.

Alba Navarro,[1,2] Guillem Clot,[1,2*] Alejandra Martínez-Trillos,[1,2] Magda Pinyol,[2,3] Pedro Jares,[1,2] Blanca González-Farré,[1,2] Daniel Martínez,[1,2] Nicola Trim,[4] Verónica Fernández,[1] Neus Villamor,[1,2] Dolors Colomer,[1,2] Dolors Costa,[1,2] Itziar Salaverria,[1,2], David Martín-Garcia,[1,2] Wendy Erber,[5] Cristina López,[6,7] Sandrine Jayne,[8] Reiner Siebert,[6,7] Martin J. S. Dyer,[8] Adrian Wiestner,[9] Wyndham H. Wilson,[10] Marta Aymerich,[1,2] Armando López-Guillermo,[1,2] Àlex Sánchez,[11,12] Elías Campo,[1,2] Estella Matutes[2] and Sílvia Beà[1,2]*

[*]AN and GC contributed equally to this work

[1]Institut d'Investigacions Biomèdiques August Pi i Sunyer, Hospital Clínic, Universitat de Barcelona, Spain; [2]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Spain; [3]Genomics Unit, IDIBAPS, Barcelona, Spain; [4]West Midlands Regional Genetics Laboratory, Birmingham, UK; [5]School of Pathology and Laboratory Medicine, The University of Western Australia, Crawley, WA, Australia; [6]Institute of Human Genetics, University Kiel, Germany; [7]Institute of Human Genetics, University Hospital of Ulm, Germany; [8]Ernest and Helen Scott Haematological Research Institute, Department of Biochemistry, University of Leicester, UK; [9]National Heart, Lung, and Blood Institute, Bethesda, MD, USA; [10]Lymphoid Malignancies Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA; [11]Department of Genetics Microbiology and Statistics, University of Barcelona, Spain and [12]Statistic and Bioinformatics Unit, Vall d'Hebron Research Institute, Barcelona, Spain

Acknowledgements: the authors would like to thank the Hematopathology Collection for sample procurement and to the Genomics core facility of the Institut d'Investigacions Biomèdiques August Pi i Sunyer for the technical help. We thank Miriam Prieto for excellent technical assistance, Beatriz Bellosillo for MYD88 real-time allele-specific PCR determination and the members of the cytogenetics and molecular cytogenetics laboratories of Kiel. The genotyping service was carried out at CGEN (PTI3/0001, ISCIII-SGEFI/FEDER).

Funding: this work was supported by research funding from Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III PI14/00571 (SB), Fundació La Marató de TV3 TV3-Cancer/2013410 (SB), Ministerio de Economía y Competitividad, Grant No. SAF2015-64885-R (EC), MTM2015-64465-C2-1-R (AS), Generalitat de Catalunya Suport Grups de Recerca 2013-SGR-378 (SB), 2013-SGR-0795 (EC), 2014SGR967 (DC), and 2014-SGR-464 (AS), and the European Regional Development Fund "Una manera de fer Europa", CERCA Programme / Generalitat de Catalunya. EC is an Academia Researcher of the "Institució Catalana de Recerca i Estudis Avançats" of the Generalitat de Catalunya.

Correspondence: sbea@clinic.cat
doi:10.3324/haematol.2016.160374

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Swerdlow S, Campo E, Harris N et, al (Eds.): WHO classification of tumours of haematopoietic and lymphoid tissues. IARC: Lyon, 2008.
2. Tiacci E, Trifonov V, Schiavoni G et al. BRAF mutations in hairy-cell leukemia. N Engl J Med. 2011;364(24):2305-2315.
3. Gachard N, Parrens M, Soubeyran I, et al. IGHV gene features and MYD88 L265P mutation separate the three marginal zone lymphoma entities and Waldenstrom macroglobulinemia/lymphoplasmacytic lymphomas. Leukemia. 2013;27(1):183-189.
4. Orsborne C, Byers R. Impact of gene expression profiling in lymphoma diagnosis and prognosis. Histopathology. 2011;58(1):106-127.
5. Haferlach T, Kohlmann A, Wieczorek L, et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. J Clin Oncol. 2010; 28(15):2529-2537.
6. Dufresne SD, Felgar RE, Sargent RL, et al. Defining the borders of splenic marginal zone lymphoma: a multiparameter study. Hum Pathol. 2010;41(4):540-551.
7. Martinez D, Navarro A, Martinez-Trillos A, et al. NOTCH1, TP53, and MAP2K1 mutations in splenic diffuse red pulp small B-cell lymphoma are associated with progressive disease. Am J Surg Pathol. 2016;40(2):192-201.
8. Bea S, Valdes-Mas R, Navarro A, et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc Natl Acad Sci USA 2013;110(45):18250-18255.
9. Puente XS, Bea S, Valdes-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015;526(7574):519-524.
10. Karube K, Martinez D, Royo C, et al. Recurrent mutations of NOTCH genes in follicular lymphoma identify a distinctive subset of tumours. J Pathol. 2014;234(3):423-430.
11. Cornet E, Debliquis A, Rimelen V, et al. Developing molecular signatures for chronic lymphocytic leukemia. PLoS One. 2015; 10(6):e0128990.
12. Fernandez V, Salamero O, Espinet B, et al. Genomic and gene expression profiling defines indolent forms of mantle cell lymphoma. Cancer Res. 2010;70(4):1408-1418.
13. Royo C, Navarro A, Clot G, et al. Non-nodal type of mantle cell lymphoma is a specific biological and clinical subgroup of the disease. Leukemia. 2012;26(8):1895-1898.
14. Falini B, Tiacci E, Liso A, et al. Simple diagnostic assay for hairy cell leukaemia by immunocytochemical detection of annexin A1 (ANXA1). Lancet. 2004;363(9424):1869-1870.
15. Summers TA, Jaffe ES. Hairy cell leukemia diagnostic criteria and differential diagnosis. Leuk Lymphoma. 2011;52 Suppl 2:6-10.