***Supplementary material and methods***

***miR RT- qPCR analysis***

A total of 664 human miRs were investigated using a RT-looped qPCR performed with the TaqMan Human A+B microRNA fluidic card system (Applied Biosystems, Darmstadt, Germany). Two pools of mutiplex RT reactions were performed that finally allowed the quantitative measurement of the miRs and several additional negative and endogenous controls per sample. Pre-amplification was used to analyze a total of 10ng of total RNA/miR. The fluidic cards were run on the 7900HT Fast Real-Time PCR system using the SDS 2.3 software (Applied Biosystems). Detectable miR expression was considered from Ct values below 35. Individual thresholds and baselines were established for each miR in order to achieve optimal amplification curves for Ct definition, and these parameters were consistently maintained across the whole series. Relative expression was referred to the same calibrator used previously and composed of a pool of purified CD19+CD5+ cells (1). Best endogenous control selection among all that are included in the fluidic cards was performed applying the Genorm algorithm. This analysis identified *RU48* gene amplicon as the most robust candidate. Relative expression for each miR was median-centered and log10 transformed. Finally, we used individual miR RT-looped qPCR assays for validation, in the lymph node series, of the selected miR (miR-34a) with significant survival value in the profiling study. Additionally, *RU48* was also analyzed as endogenous control. These PCR reactions were performed using Step-One qPCR thermocycler (Applied Biosystems), and the relative quantification was performed in reference to the same pool of normal lymph nodes as previously described (1).

***Consensus clustering***

The unsupervised miR expression data was analyzed with the Consensus Clustering (CC) method as implemented in the Consensus Cluster Plus (CCP) package from Bioconductor (2). CC is a previously validated approach to improve the classical hierarchical clustering methods in order to obtain reliability in the number of subgroups present in the dataset and the group memberships of cases (3). Briefly, the algorithm begins by subsampling a proportion of items and a proportion of features from the data matrix. Each subsample is then partitioned into up to *k* groups by hierarchical clustering using Euclidean average metrics. Then, the proportion that two items occupied the same cluster is calculated for each k obtaining a consensus matrix that allows a visual checking of the number of robust clusters to find the 'cleanest' cluster partition where items nearly always either cluster together giving a high consensus (dark blue colour) or do not cluster together giving a low consensus (white). (3). We used 1000 subsamples, each consisting of a random sample of 80% of the total items. Additional tools implemented in the CCP allow the visual identification of robust cluster numbers (using empirical distribution function –CDF-) and samples that show consistent memberships to the robust clusters (item tracking plots) (2). CDF plots display consensus distributions for each *k* to find the *k* at which the distribution reaches an approximate maximum, which indicates a maximum stability. The item tracking plot shows the consensus cluster of items (in columns) at each *k* (in rows). This allows a user to track an item's cluster assignments across different *k*, to identify promiscuous items that are suggestive of weak class membership. Only robust core clusters obtained by CC will be considered for later analyses as previously described (4).

***miR expression supervised analyses***

Initially, differentially expressed miRs were considered between separate categories of (i) *IGHV* mutational status categories (Unmutated -U- and Mutated –M-) according to a previously used homology cutoff (98%) (1), that also showed significant prognostic value in the present leukemic MCL series, and (ii) SOX11 expression categories of high and low expression were defined by a cutoff (6.4 RU of relative expression) estimated using a maximally selected log-rank statistic (Maxstat package from R). For further exploration of the possible influence of cluster membership into the expression of miRs related with *IGHV* status, several subgroups were considered defined by a combination of two factors: robust cluster membership (A, B and C), and *IGHV* mutational status categories. The chosen method for supervised analysis was empirical Bayes moderated t-statistic as implemented in the Bioconductor package limma (http://cran.r-project.org). For statistical correction of multiple comparisons, *P* values were adjusted using the FDR method following Benjamini and Hochberg (5). Adjusted *P*-values less than 0.05 were considered statistically significant. Venn diagrams were generated using the implemented Venn tool in R version 2.14.1 (http://cran.r-project.org). Categorical data were compared using Fischer's exact test (two-sided *P*) and Kruskal-Wallis nonparametric test was used for comparisons of continuous variables.

***Gene expression and related statistical analyses***

mRNA expression was measured by hybridization to Affymetrix HG133Plus 2.0 microarrays (Affymetrix, Santa Clara, CA, US) in 16 MCL cases included in the series analyzed for miR expression (identified in Supplementary Table 3a). The analysis of the scanned images and the determination of the signal value for each probe set of the array

were obtained with GeneChip® Command Console® Software (AGCC) (Affymetrix). Raw data were imported to Expression Console build 1.2.0.20 (Affimetrix) package for normalization using the Robust Multichip Analysis (RMA) algorithm. Probesets were collapsed to single gene names using a previously described method implemented as *collapseRows* function in the WGCNA R package (http://www.r-project.org/) (6). On one hand, some of these data were used to calculate a gene proliferation signature previously defined as prognostic factor in MCL as previously described (1, 7). Correlation Spearman's Rho correlation coefficients and the associated *P*-values between proliferation signature and the individual miR expression data were computed using R 2.14.1. On the other hand, the gene set was more stringently selected for those gene with higher variability (25% with higher standard deviation) and used for searching of correlations between the expression of miRs previously found related to *IGHV* status or SOX11 expression categories and their predicted gene targets by PITA and Targetscan algorithms using the MAGIA webtool (http://gencomp.bio.unipd.it/magia) (8). From the obtained gene lists ranked by the correlation Spearman Rho value, two types of pathway enrichment analyses were used. First, we performed an enrichment pathway analysis using the GSEA desktop application (GSEA, Broad Institute at MIT, Cambridge, MA), and several reference gene set collections including the curated collection of canonical pathways (3276 gene sets), KEGG pathways (186 gene sets) and GO terms (BP) (825 gene sets). This analysis approach has been previously used for miR target analysis (9). Top enriched gene sets were determined on the basis of statistical significance ($P<0.05$) and ranked on the basis of false discovery rate (FDR<0.25) and the normalized enriched score (NES>|1.3|). In addition, to explore further possible enrichments exclusively among the inversely correlated predicted targets of miR-708 and miR-455-5p/3p we used a Gene

Group Functional Profiling method implemented in the g-Profiler webpage that it has been described as an improved method for multiple testing correction on complex functional term analysis with the g:GOSt webtool (included in g:Profiler; http://biit.cs.ut.ee/gprofiler/)(10). This analysis was performed in parallel at two levels of stringency, selecting 75% and 30%, respectively, from the top inversely correlated predicted targets for these miRs.

***miR/mRNA RT- qPCR validation***

For the validations of miR-34a/MYC expression prognostic value in MCL and potential inverse relationship of miR-455-5p/miR-708 with several candidate target genes, RT-qPCRs were performed following the supplier instructions (all reagents from Applied Biosystems). Reverse transcription was performed using TaqMan MicroRNA Reverse Transcription kit and High Capacity RNA-to-cDNA kit. For quantitative PCR TaqMan microRNA assays for miR-34a and RU48 (as endogenous control for miR quantification) were used. Gene expression assays for EZH1: Hs00940463_m1; MLL2: Hs00231606_m1; SETDB1: Hs00180850_m1; HEMK1: Hs00275076_m1; MYC: Hs00905030_m1 and GUS Hs00939627_m1 (as endogenous control for mRNA quantification) were used for mRNA quantification. All RT-PCRs were run in a 7900HT Fast Real-Time PCR system and using the SDS 2.3 software (Applied Biosystems). Ct values were obtained using automatic baseline and threshold. Relative expression calculations were performed using Expression Suite Software v1.0 (Applied Biosystems) in relation to a equimolar mixture of all cases (for miRs) or the Universal Human Reference RNA (Stratagene) (for mRNAs).

***Survival analysis***

Survival analysis was first performed in the test leukemic MCL series considering each individual miR expression as a continuous variable together with previously defined *IGHV* status and SOX11 expression categories using multivariate Cox regression (R survival package) and a high stringent statistical significance ($P<0.01$). An independent series of leukemic (26 cases) and spleen (5 cases) MCL samples were studied for validation of miR-34a prognosis value using an actuarial survival analysis as previously described (11), including Kaplan-Meier/log-rank test analysis. Moreover, 50 lymph node MCL and 38 leukemic MCL samples with available RNA were used for exploration of potential survival interactions of miR-34a and MYC expression levels. Categorical groups of high/low expression of these genes were defined maximizing the number of patients per group, and by using cutoffs with a maximally selected log-rank statistic (Maxstat package) along the studied MCL series. Actuarial survival analysis was performed as the previous ones.

**References**


1. Navarro A, Bea S, Fernandez V, Prieto M, Salaverria I, Jares P et al. MicroRNA expression, chromosomal alterations, and immunoglobulin variable heavy chain hypermutations in Mantle cell lymphomas. Cancer Res 2009;69:7071-8.

2. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26:1572-3.

3. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine learning 2003;52:91-118.

4. Brannon AR, Reddy A, Seiler M, Arreola A, Moore DT, Pruthi RS et al. Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. Genes Cancer 2010;1:152-63.

5. Benjamini Y, Hoechberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B 1995;57:289-300.

6. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR et al. Strategies for aggregating gene expression data: the collapseRows R function. BMC Bioinformatics 2011;12:322.

7. Rosenwald A, Wright G, Wiestner A, Chan WC, Connors JM, Campo E et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell 2003;3:185-97.

8. Sales G, Coppe A, Bisognin A, Biasiolo M, Bortoluzzi S, Romualdi C. MAGIA, a web-based tool for miRNA and Genes Integrated Analysis. Nucleic Acids Res 2010;38:W352-W359.

9. Mestdagh P, Lefever S, Pattyn F, Ridzon D, Fredlund E, Fieuw A et al. The microRNA body map: dissecting microRNA function through integrative genomics. Nucleic Acids Res 2011;39:e136.

10. Reimand J, Arak T, Vilo J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res 2011;39:W307-W315.

11. Hernandez L, Bea S, Pinyol M, Ott G, Katzenberger T, Rosenwald A et al. CDK4 and MDM2 gene alterations mainly occur in highly proliferative and aggressive mantle cell lymphomas with wild-type INK4a/ARF locus. Cancer Res 2005;65:2199-206.