Experimental procedure

Patients and treatments
The study population consisted of a retrospective series of 258 unselected, de novo cases of DLBCL obtained from multiple centers in Spain, Italy and USA.

The whole series of patients was divided into two major groups, a DISCOVERY group (36 patients for whom frozen tissue was available for profiling using array technologies) and a TEST group (240 patients with available FFPE tissue from the diagnostic pathological sample). The TEST group was further divided systematically into two sets of patients, the training group (123 patients) and the validation group (117 patients) on the basis of the moment when the sample was received.

The DISCOVERY series was intended to involve a complete miRnome screening (with the available high-throughput technologies) using RNA extracted from frozen samples, while the TEST series was intended to serve as an “easy to use” quantitative method for rapid clinical translation. 16 patient samples were studied in both series.

The clinical characteristics according to sex, age, stage, extranodal disease, serum LDH levels, ECOG, and IPI score of the different sets of patients are summarized in Table 1.

RNA and DNA extraction
For miRNA hybridization of fresh frozen samples, total RNA was isolated by Trizol reagent (Invitrogen, Carlsbad, CA, USA). RNA quality was checked using total RNA (small fraction chip) with the Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA) and by 1% agarose electrophoresis, following the standard procedure.

For gene expression studies, RNA extracted with Trizol was purified by RNase-free DNase I RNeasy kit (Qiagen Inc., Valencia, CA) following the standard procedure. mRNA quality was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies).

For quantitative PCR of FFPE samples, RNA was extracted as previously described\textsuperscript{1,2} with minor modifications. In brief, three 10-μm tissue sections were deparaffinized by xylene incubation. After centrifugation and an ethanol washing and drying procedure, the pellet was resuspended in the lysis buffer and incubated with proteinase K (Qiagen). RNA was purified by phenol-chloroform extraction followed by isopropanol precipitation.

Microarray procedures: miRNA, mRNA hybridization and normalization
For miRNA hybridization, 100 ng of total RNA were hybridized on an Agilent 8 × 15K Human miRNA one-color microarray for detecting 470 human and 64 viral miRNAs, following the manufacturer’s instructions (Agilent Technologies).\textsuperscript{3} The sample was labeled with Cy3 and cRNA product was hybridized overnight following the manufacturer’s instructions. Between-array median normalization was carried out using Genespring GX9 software to make miRNA expression datasets comparable.
RNA for gene expression profiling was hybridized on a Whole Human Genome Agilent 4 × 44K v1 Oligonucleotide Microarray. Briefly, 500 ng of mRNA were amplified and fluorescence-labeled with Agilent’s Low RNA Input Fluorescence Linear Amplification kit (Agilent Technologies). The sample was labeled with Cy5 and the reference Universal Human Reference RNA (Stratagene, La Jolla, CA, USA) was labeled with Cy3. cRNA product was hybridized overnight following the manufacturer’s instructions.

Washing steps for all platforms were performed using corresponding buffers at recommended temperatures in an ozone-free environment. Scanning was carried out immediately using the Agilent G2565AA Microarray Scanner System (Agilent Technologies) and data were collected with Feature Extraction v9.5 software (Agilent Technologies).

For gene-expression profiling, microarray background subtraction was carried out using the normexp tool available in Bioconductor’s limma package (http://www.bioconductor.org). To normalize the dataset, loess within-array normalization and quartile between-array normalization were performed. MiRNA and gene expression data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number (Super Series GSE21849). http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21849

**Differential miRNA expression analysis in DLBCL subtypes according to the COO and identification of its putative target genes and gene pathways.**

Gene expression data were available from most of the discovery set (29 cases) and were used to classify patients according to the Cell of Origin (COO) using the gene set classifier described by Wright et al. After categorizing the cases as GC type (11 cases) or ABC type (18 cases), their miRNA expression levels were analyzed. Twenty-four miRNAs were identified as being differentially expressed by a factor of at least two-fold among these categories (Table S4). We then performed T-tests (significant for values of p < 0.02) and FDR tests (using the Benjamini-Hochber FDR for multiple-testing correction) to select 8 miRNAs that were differentially expressed between those subtypes with a value of FDR < 0.03 (see Table S4). Genes differentially expressed between GC and ABC subtypes were then filtered using miRBase (miRBase v11.0, MICROCOSM). We selected only those genes that were predicted to be targets of any of the differentially expressed miRNAs. Gene Set Enrichment Analysis (GSEA v2) was performed on this group of genes using BIOCARTEA-lymphoma-curated gene sets as classifiers.

**Identification of candidate gene targets for the miRNAs contained in the predictor models.**

miRBase (miRBase v11.0, MICROCOSM) and TargetScan (TargetScan release 5.1, Whitehead Institute for Biomedical Research) databases were interrogated to identify the genes that might be regulated by the miRNAs related with outcome in this series of DLBCL cases. Details of algorithms used by these bioinformatic tools can be found on their respective web sites (http://www.mirbase.org/ and http://www.targetscan.org/).

In addition to this bioinformatic prediction we examined the correlations of the gene and miRNA expression data from the discovery set. We performed a Pearson correlation test and considered any significant negative correlations to be a main mechanism by which miRNAs modulate
protein expression, i.e., downregulation of mRNA. The complete correlation grid obtained is shown in Table S4.

**Real-time PCR for relative miRNA quantification using RNA from FFPE tissue and identification of a set of miRNAs associated with outcome in DLBCL.**

MiRNA expression profiling of FFPE tissues was conducted by the Applied Biosystems 384-well multiplexed real-time PCR assay using 250 ng of total RNA. RNA from each case was reversed-transcribed in triplicate using a multiplex looped primer pool with the selected miRNA probes together with let7a and let7d. Each completed reaction was loaded onto the plate and real-time PCR was done on the ABI 7900HT Prism. Two miRNAs, with only minor variations in the studied series (let7a and let7d), were used as endogenous miRNAs.\(^\text{14}\) Ct values were exported using SDS software (SDS 2.3), and the data were analyzed with real-time StatMiner software (INTEGROMICSTM; www.integromics.com). An miRNA was considered to be present if the Ct was less than 36 in all three biological replicates; \(-\Delta CT\) values \((\text{Ct value of miRNA of interest} – \text{mean Ct value for let7a & let7d})\) were considered for further statistical analysis. After the previously described procedures, a final set of 14 candidate miRNAs were selected on the basis of their relationship with the COO signature and the outcome (according to their p values in T-tests between CG-ABC and univariate Cox regression analysis by SignS). After Cox regression using backward stepwise selection procedures for a subset of 87 cases from the training group of patients, 9 miRNAs were selected for further quantification in all 240 cases from the test group of FFPE tissue samples. These data were statistically analyzed further.

**Immunohistochemistry and TMA construction**

All cases of DLBCL with available FFPE tissue were histologically reviewed and representative areas were selected. We used a tissue arrayer device (Beecher Instruments, Sun Prairie, WI) to construct tissue microarray (TMA) blocks, according to conventional protocols.\(^\text{15}\) Standard tissue sections were also analyzed in some cases.

Immunohistochemical staining was performed as follows: 2–4-μm–thick paraffin-embedded TMAs were cut onto Dako slides (DAKO), and subsequently de-waxed, rehydrated and subjected to antigen retrieval by heating in 50 mM Tris (Trizma base)-1 mM EDTA (ethylenediaminetetraacetic acid) (Sigma Chemical) (pH 8) or citrate 10 mM pH 6.5 in a pressure cooker for 2 minutes. The slides were cooled and treated with peroxidase-blocking solution (DAKO) for 5 minutes.

Sections were then immunostained with antibodies against Gcet1, MUM1, CD10, bcl6 and FOXP1, counterstained with hematoxylin and mounted. Incubations, either omitting the specific antibody or containing unrelated antibodies, were used as a negative control of the technique.

The antibodies used were CD10 (clone 56C6, dilution 1:10; Novocastra, Newcastle, UK), bcl6 (clone G191E/A8, dilution 1:4, CNIO\(^\text{16}\)), MUM-1/IRF4 (polyclonal Ab, dilution 1:250; Santa Cruz Biotechnology, Santa Cruz, CA), Gcet1 (clone RAM, dilution 1:4, CNIO\(^\text{17}\)) and FOXP1 (clone JC12, dilution 1:100, CNIO). Antigen retrieval and visualization methods are shown in the table below (Table S1)
<table>
<thead>
<tr>
<th>Antibody</th>
<th>Clone</th>
<th>Source</th>
<th>Antigen retrieval/Visualization method</th>
<th>Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD20</td>
<td>L26</td>
<td>DAKO</td>
<td>EDTA 1mM pH 9/FLEX DAKO</td>
<td>1/50</td>
</tr>
<tr>
<td>GCET1</td>
<td>RAM</td>
<td>CNIO, Madrid, Spain</td>
<td>EDTA 1mM pH 8/Novolink (Novocastra)</td>
<td>1/4</td>
</tr>
<tr>
<td>MUM1/IRF4</td>
<td>Polyclonal goat</td>
<td>Santa Cruz Biotechnology, Santa Cruz, CA</td>
<td>Citrate 10 mM pH 6.5/LSAB (DAKO)</td>
<td>1/250</td>
</tr>
<tr>
<td>CD10</td>
<td>56C6</td>
<td>Novocastra, Newcastle, UK</td>
<td>Citrate 10 mM pH 6.5/Novolink (Novocastra)</td>
<td>1/10</td>
</tr>
<tr>
<td>BCL6</td>
<td>GI191E/A8</td>
<td>CNIO, Madrid, Spain</td>
<td>Citrate 10 mM pH 6.5 and Proteinase K/Novolink (Novocastra)</td>
<td>1/4</td>
</tr>
<tr>
<td>FOXP1</td>
<td>JC12</td>
<td>CNIO, Madrid, Spain</td>
<td>Citrate 10 mM pH 6.5/novolink (Novocastra)</td>
<td>1/100</td>
</tr>
</tbody>
</table>

Table S1. Antibody clones used, source, antigen retrieval and visualization method and dilutions used

Immunoreactivity was scored by two independent pathologists (SMM and JADP) and the percentage of tumor-cell staining was estimated by visual inspection and categorized into the appropriate decile. Disagreements were resolved by joint review using a multihread microscope. For each case the most representative core was selected (highest percentage of neoplastic cells). Scoring of GCET1, MUM1, CD10, bcl6, and FOXP1 was based on the cutoffs used by Choi et al.18

General statistical methods

Over-all survival (OS)19 was considered as the time from diagnosis to the date of death from any cause or last contact. Progression-free survival (PFS) was calculated from the time of diagnosis to the date of progression. Cases were censored on the date the patient was last known to be alive or, for patients dying as a result of causes unrelated to lymphoma or treatment, the date of death. The Kaplan-Meier method was used to estimate OS and PFS distributions.20 The log-rank and Breslow tests were used to compare survival distributions.21 The chi-square test was used to assess differences in the proportions of individual prognostic factors between series.

A summary of the clinical characteristics of the whole set of patients used in the study can be found in Table 1. Kaplan-Meier estimates for OS and PFS in the whole test group of R-CHOP–treated DLBCL patients and according to IPI score and GC-ABC phenotyping are shown in Fig. S1.

The two classification systems (GC-ABC and miRNA based) were directly compared with a Chi-square test. The systems were non-overlapping and complementary, thereby establishing different subsets of cases (p < 0.001). Furthermore, after multivariate analysis including IPI, miRNA score and GC-ABC subclassification based on IHQ, only miRNAs and IPI were still significant for both OS and PFS (Table S4, sheet D).
Figure S1. Kaplan-Meier estimates for OS and PFS in the entire test group of R-CHOP–treated DLBCL patients
The estimated 2-year OS was 74.7% ± 3 (A) and the estimated PFS was 67.5 ± 3 (B). IPI stratification predicts both OS (C) and PFS (D) in this series of patients (log-rank test, p < 0.05). Patient stratification after GC-ABC phenotyping according to Choi’s algorithm18 also predicts for both OS and PFS. The estimated 2-year OS for ABC type DLBCL cases was 69.8% ± 4.5, which was significantly worse than for GC type DLCBL patients (81.4% ± 4.3; p < 0.05) (E). Differences were also found for PFS (60.7% ± 4.7 for ABC type vs. 75.6% ± 4.6 for GC type; p < 0.05) (F).
Statistical model building

A) Cox regression analysis

Cox regression was used to derive three independent survival models based on IPI score, miRNA expression (of 9 selected miRNAs) and GC-ABC classification based on immunohistochemistry for both OS and PFS in the training set of patients (123 patients). miRNA-based survival models are summarized in the following survival functions:

For OS:
\[ h(t) = (-0.61681 \times \text{miR221}) + (0.53702 \times \text{miR222}) - (0.3742 \times \text{miR331}) + (0.07177 \times \text{miR451}) + (0.29792 \times \text{miR93}) + (0.21191 \times \text{miR491}) + (0.05561 \times \text{miR28}) - (0.00875 \times \text{miR151}) - (0.05425 \times \text{miR148a}). \]

For PFS:
\[ h(t) = (-0.5795 \times \text{miR221}) + (0.4665 \times \text{miR222}) - (0.3358 \times \text{miR331}) + (0.0636 \times \text{miR451}) + (0.2412 \times \text{miR93}) + (0.1743 \times \text{miR491}) + (0.0617 \times \text{miR28}) + (0.0341 \times \text{miR151}) - (0.0287 \times \text{miR148a}). \]

More complex models using combinations of the three predictor models were generated and the improved model fit was assessed using a chi-square test to determine the significance of the change in log likelihood. Only the combination of miRNAs and IPI score was significantly better than the individual models for both OS and PFS (p < 0.05 for all comparisons), justifying the fitting of a large model with these variables. Model terms were chosen by stepwise backward selection using AUC as the criterion. This yielded definitive combined survival models for OS and PFS that can be summarized in the following survival functions:

For OS:
\[ h(t) = 0.400 \times \text{IPI(1)} + 3.113 \times \text{IPI(2)} + 3.019 \times \text{IPI(3)} + 3.856 \times \text{IPI(4)} + 4.723 \times \text{IPI(5)} - 0.567 \times \text{miR221} + 0.370 \times \text{miR222} - 0.536 \times \text{miR331} + 0.405 \times \text{miR93} + 0.342 \times \text{miR491}. \]

For PFS:
\[ h(t) = -0.661 \times \text{IPI(1)} + 1.810 \times \text{IPI(2)} + 1.695 \times \text{IPI(3)} + 2.376 \times \text{IPI(4)} + 3.074 \times \text{IPI(5)} - 0.560 \times \text{miR221} + 0.315 \times \text{miR222} - 0.465 \times \text{miR331} + 0.276 \times \text{miR93} + 0.242 \times \text{miR491} + 0.222 \times \text{miR151}. \]

Hazard ratios estimated from the models after variable selection are shown in Fig. 4 in the main text.

Assumptions of key models were checked. We used Schoenfeld residuals and the “correlation with time” test of Grambsch and Therneau to assess the proportional hazards assumption. To check the functional form of the covariates we used martingale residuals. Possible outliers and influential observations were assessed with deviance residuals and dfbetas, respectively. These diagnostics were examined in the full model with IPISCORE and miRNA and in the model from variable selection.
**B) Validation of predictive performance**

Models generated in the training series were validated in the validation group using the (integrated) area under the ROC curve, the concordance index and the Brier score.\textsuperscript{26,27,28,29,30,31} When evaluating the Brier score, we used a conditional weighting scheme from a Cox model of the censoring distribution,\textsuperscript{31} including all miRNAs and the IPI score as predictors.

Results are shown in Tables S2 and S3. Fig. S2 shows the estimated Brier score at different times for each of the models studied.

<table>
<thead>
<tr>
<th></th>
<th>C-Index</th>
<th>iAUC</th>
<th>iBSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null (or overall KM)</td>
<td>0.50</td>
<td>0.50</td>
<td>0.19</td>
</tr>
<tr>
<td>IPISCORE</td>
<td>0.64</td>
<td>0.61</td>
<td>0.21</td>
</tr>
<tr>
<td>miRNA</td>
<td>0.63</td>
<td>0.63</td>
<td>0.16</td>
</tr>
<tr>
<td>IPISCORE + miRNA, full</td>
<td>0.63</td>
<td>0.63</td>
<td>0.19</td>
</tr>
<tr>
<td>IPISCORE + miRNA, variable selection</td>
<td>0.63</td>
<td>0.63</td>
<td>0.19</td>
</tr>
</tbody>
</table>

**Table S2. Predictive accuracy PFS**

<table>
<thead>
<tr>
<th></th>
<th>C-Index</th>
<th>iAUC</th>
<th>iBSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null (or overall KM)</td>
<td>0.50</td>
<td>0.50</td>
<td>0.09</td>
</tr>
<tr>
<td>IPISCORE</td>
<td>0.62</td>
<td>0.62</td>
<td>0.10</td>
</tr>
<tr>
<td>miRNA</td>
<td>0.68</td>
<td>0.67</td>
<td>0.08</td>
</tr>
<tr>
<td>IPISCORE + miRNA, full</td>
<td>0.71</td>
<td>0.72</td>
<td>0.10</td>
</tr>
<tr>
<td>IPISCORE + miRNA, variable selection</td>
<td>0.71</td>
<td>0.74</td>
<td>0.11</td>
</tr>
</tbody>
</table>

**Table S3. Predictive accuracy OS**

**Tables S2 and S3. Predictive performance in the validation group of patients**

The predictive performance using the (integrated) area under the ROC curve, the concordance index and the Brier score are shown for the different models generated from the training set. Overall, predictions were more accurate when miRNAs were incorporated into IPI-based models.

Null (or overall KM): represents the expected value either from a null model with no predictive ability (C-index and iAUC) or from a Kaplan-Meier estimate where covariate information was ignored; IPISCORE: Independent survival model based on IPI score; miRNA: Independent survival model based on the expression of the eight selected miRNAs; IPI score + miRNA full: Combined survival model based on the IPI score and the whole set of miRNAs; IPISCORE + miRNA, variable selection: Combined survival model based on IPI score and miRNAs after backwards stepwise variable selection (see text and Fig. 4 for variables in the model).
Figure S2. Estimated prediction error curves according to Brier score at different times for each model
According to this analysis the prediction based on miRNAs has the lowest predictive error and improves the IPI based models. This is particularly true early on when IPI-based models perform poorly.

C) Kaplan-Meier representation of miRNA-based and combined survival model based on IPI score and miRNAs
miRNA-based and combined survival models (functions) based on IPI score and miRNAs after variable selection (see survival functions $h_i(t)$ (OS) and $h_i(t)$ (PFS) were applied to the all patients in the test group and in the separate training and validation sets of patients. The distribution of results for each patient in OS and PFS models was represented after Kaplan-Meier estimation. The log-rank and Breslow tests were used to compare survival distributions (see Figs. 2, 3 and Figs. S3 and S4).
miRNA-based survival scores were calculated for each patient in the training and validation groups (123 and 115 patients with available miRNA expression values, respectively) according to the survival function obtained from the training set. After median stratification, patients from the two groups were classified as low miRNA score (below median) or high miRNA score (above median). Kaplan-Meier estimates were plotted for OS and PFS (log-rank test, p < 0.05 for both OS and PFS in all comparisons, except for PFS in the validation set, p = 0.25).
Figure S4. Kaplan-Meier representation of combined survival model based on IPI score and miRNAs

Survival scores according to both IPI and miRNAs were calculated for each patient in the training and validation groups (123 and 115 patients with available miRNA expression and IPI values, respectively) according to the survival function obtained in the training set of patients. After tercile stratification, Kaplan-Meier estimates were plotted for OS and PFS (log rank test, p < 0.05 for OS and PFS).

Models were built and validated using the R statistical program, specifically with packages survival (T. Therneau), pec (T. Gerds), and survcomp (B. Haibe-Kains and C. Sotiriou and G. Bontempi). Additional analyses were done with SPSS version 15.0.0 (SPSS, Chicago, IL).
REFERENCES


