REGULAR ARTICLE

blood advances

A gene expression assay based on chronic lymphocytic leukemia activation in the microenvironment to predict progression

Pau Abrisqueta,[1,2] Daniel Medina,[2,3] Guillermo Villacampa,[4] Junyan Lu,[5] Miguel Alcoceba,[6] Julia Carabia,[2,3] Joan Boix,[2,3] Barbara Tazón-Vega,[1,2] Gloria Iacoboni,[1,2] Sabela Bobillo,[1,2] Ana Marín-Niebla,[1,2] Marcos González,[6] Thorsten Zenz,[7,*] Marta Crespo,[2,3,*] and Francesc Bosch[1,2,*]

[1]Servei d'Hematologia, Vall d'Hebron Hospital Universitari, Experimental Hematology, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain; [2]Departament de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain; [3]Experimental Hematology, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain; [4]Oncology Data Science (OdysSey) Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain; [5]European Molecular Biology Laboratory (EMBL), Heidelberg, Germany; [6]Department of Hematology, University Hospital of Salamanca (HUS/IBSAL), CIBERONC and Center for Cancer Research-IBMCC (USAL-CSIC), Salamanca, Spain; and [7]Department of Medical Oncology and Hematology, University Hospital and University of Zurich, Zurich, Switzerland

Key Points

- The CLL15 signature, based on the expression of genes associated with microenvironment signaling, predicts early progression in CLL.
- The predictive power of the CLL15 signature is independent of the *IGHV* mutational status and IPS-E CLL score.

Several gene expression profiles with a strong correlation with patient outcomes have been previously described in chronic lymphocytic leukemia (CLL), although their applicability as biomarkers in clinical practice has been particularly limited. Here we describe the training and validation of a gene expression signature for predicting early progression in patients with CLL based on the analysis of 200 genes related to microenvironment signaling on the NanoString platform. In the training cohort (n = 154), the CLL15 assay containing a 15-gene signature was associated with the time to first treatment (TtFT) (hazard ratio [HR], 2.83; 95% CI, 2.17-3.68; $P < .001$). The prognostic value of the CLL15 score (HR, 1.71; 95% CI, 1.15-2.52; $P = .007$) was further confirmed in an external independent validation cohort (n = 112). Notably, the CLL15 score improved the prognostic capacity over *IGHV* mutational status and the International Prognostic Score for asymptomatic early-stage (IPS-E) CLL. In multivariate analysis, the CLL15 score (HR, 1.83; 95% CI, 1.32-2.56; $P < .001$) and the IPS-E CLL (HR, 2.23; 95% CI, 1.59-3.12; $P < .001$) were independently associated with TtFT. The newly developed and validated CLL15 assay successfully translated previous gene signatures such as the microenvironment signaling into a new gene expression–based assay with prognostic implications in CLL.

Introduction

It is well accepted that patients with chronic lymphocytic leukemia (CLL) who are asymptomatic and in an early clinical phase do not require therapy.[1] Nevertheless, cumulative data on the risk of clonal evolution[2-4] renewed interest in early therapeutic intervention in patients at diagnosis who are likely to progress rapidly.[5] Therefore, the identification of these patients at diagnosis has been an intense focus of clinical research in the field of CLL. Prognostication in this setting has classically relied on a myriad of laboratory values, cytogenetic abnormalities, gene mutations, or the mutational status of the *IGHV* genes.[6-10] More recently, the International Prognostic Score for Early-stage CLL (IPS-E) has been

Submitted 7 March 2022; accepted 28 July 2022; prepublished online on *Blood Advances* First Edition 16 August 2022; final version published online 8 November 2022. https://doi.org/10.1182/bloodadvances.2022007508.

*T.Z., M.C., and F.B. contributed equally to this study.

Data are available on request from the corresponding author, Pau Abrisqueta (pabrisqueta@vhio.net).

The full-text version of this article contains a data supplement.

developed employing 3 covariates: unmutated *IGHV*, absolute lymphocyte count $> 15 \times 10^9$/L, and presence of palpable lymph nodes.[11]

Despite this extensive investigation, the accuracy of these models may be improved.[11-16] In addition, the emergence of novel targeted agents has attracted interest in the early treatment of patients at high risk of early progression.[5]

Gene expression profiles and the clinical course of patients with CLL have been correlated in various studies.[7,8,17-26] Unfortunately, biomarkers based on gene expression profiles exhibit several caveats that preclude them from being widely applied in the prognostication of patients with CLL. These include the lack of reproducibility and standardization and the complexity of bioinformatics analysis. Significantly, the prognostic value of clustering methods is limited by the fact that the assignment of an individual may vary when different patients are included in the clustering process, thus impeding the use of these methods in real time. In this regard, the development of new platforms that allow direct and reproducible quantification of gene expression, such as NanoString nCounter, should facilitate the attainment of gene expression biomarkers applicable in clinical settings.[27,28] Among different gene signatures, and because CLL is a malignancy that is particularly dependent on interaction with the microenvironment for survival and proliferation,[25] *IGHV* mutational status signature[7,8,17,18] and genes involved in the activation of malignant cells in the microenvironment, including stimulation of the B-cell receptor (BCR),[24-26,29] are of particular interest. Indeed, this notion is reinforced by the standard use of different small molecules targeting CLL-microenvironment interactions, particularly Bruton's tyrosine kinase inhibitors.[30,31]

Herein, we developed, evaluated, and validated a multigene expression signature using genes associated with the activation of CLL cells in the microenvironment and the *IGHV* mutational status. This assay, based on the NanoString platform, should facilitate its applicability in clinical settings.

Materials and methods

Study design and patient population

The overall design of the process for developing and evaluating a new assay to assess the risk of progression in patients with CLL is shown in supplemental Figure 1. For the training cohort of the study, 156 untreated samples, 119 from the University Hospital Vall d'Hebron and 37 from the University of Salamanca, were used. The assay was validated using 112 samples from an independent cohort of patients from the German Cancer Research Center, Heidelberg, Germany. The details of the validation cohort have been reported elsewhere.[32]

Samples were obtained at diagnosis, whenever possible. For patients who did not have a sample at the time of diagnosis, samples were collected during follow-up but always before the patients received any treatment. Gene expression quantification was performed in blood samples from untreated patients diagnosed with CLL. Peripheral blood mononuclear cells were obtained using Ficoll-Paque Plus density gradient (GE Healthcare, Buckinghamshire, United Kingdom) and subsequently cryopreserved until analysis. Tumor cells were purified using immunomagnetic depletion by EasySep Human B Cell Enrichment Kit (StemCell Technologies), and the final tumor content was assessed by flow cytometry. The estimated median tumor content was 98.3% (range, 80-99.9) in the training cohort and 95.7% (range, 86.8%-99.4%) in the validation cohort.

Written informed consent was obtained from all individuals in accordance with the Declaration of Helsinki. The study was approved by the clinical research ethics committee of the Vall d'Hebron Barcelona Hospital Campus.

Gene expression analysis

Gene expression was quantified in 250 ng of RNA on the NanoString platform (NanoString Technologies, WA) using the "high sensitivity" setting on the nCounter PrepStation and 555 fields of view on the nCounter Digital Analyzer. A total of 178 genes were selected from the literature, including genes related to the activation of CLL cells in the microenvironment,[23-26] genes that were differentially expressed according to the mutational status of *IGHV*,[7,8,17,18] and other genes of prognostic interest in CLL (supplemental Methods, supplemental Table 1). Normalization for RNA loading was performed using the geometric mean of 22 housekeeping genes (supplemental Table 1). The normalized data were $\log_{10}$ transformed. The reference gene selection is further described in the "Data supplement."

Predictive gene expression score

Detailed descriptions of model building and performance assessment are provided in the "Data supplement." In brief, we used the gene expression data from the training cohort to produce a parsimonious predictive model for time to first treatment (TtFT) using a penalized Cox model.[33] To evaluate the global performance of the multivariate Cox model obtained from the selected genes, different diagnostic parameters were calculated and are summarized in the "Data supplement" (supplemental Table 2), including R^2, Brier score, iAUC (a summary measure of the area under the receiver operating characteristic curve calculated for the different times), and Harrell's C-statistic, a generalization of the AUC.[34,35] The graph obtained for the AUC values at the different time points is shown in supplemental Figure 2. For illustrative purposes, we dichotomized the predictive gene expression score in 3 risk groups using the R *partykit* package.

Statistical analysis

The statistical analysis plan was prespecified before the evaluation of the gene expression in the training and validation cohort. The primary end point of the study was TtFT, defined as the time from the date of obtaining the sample to the date of treatment onset. To study the predictive capacity of the gene expression score, we relaxed the linearity assumption using restricted cubic splines by means of *rms* R package (Harrell, F. E. Jr Package "rms" [The Comprehensive R Archive Network, 2016]). Harrell's C-statistic was calculated to compare the discrimination capacities of different models. The analysis of deviance (analysis of variance R function) was used to study whether the inclusion of new factors had a significant improvement in the predictive capacity of the model. Survival curves were estimated using the Kaplan-Meier method to visualize gene expression risk groups and were compared by the log-rank test. Cox proportional-hazard models were used to obtain hazard ratios (HRs) with 95% CIs without dichotomizing continuous factors.[36] To select prognostic variables

Table 1. Patient and disease characteristics of the training cohort

	Total cohort (n = 154), n (%)	CLL15 categories			*P* value
		Low-risk group (n = 85), n (%)	Intermediate-risk group (n = 31), n (%)	High-risk group (n = 38), n (%)	
Male	88 (57.1)	47 (55.3)	16 (51.6)	25 (65.8)	.435
Female	66 (42.9)	38 (44.7)	15 (48.4)	13 (34.2)	
Age, median (range) years	70 (34-91)	72 (34-91)	69 (46-91)	64 (44-85)	.05
Binet stage					<.01
A	116 (76.3)	79 (92.9)	24 (80)	13 (35.1)	
B	27 (17.8)	6 (7.1)	4 (13.3)	17 (45.9)	
C	9 (5.9)	0	2 (6.7)	7 (18.9)	
Missing, n	2	–	1	1	
Lymphocyte cell count, 10^9/L – median (range)	16.8 (3.2-323)	16.8 (3.2-238)	15.6 (4.2-323)	22.1 (7.9-207.4)	.18
Missing, n	38	3	7	28	
β2-microglobulin					<.01
≤3.5 mg/dL	108 (74)	68 (81.9)	24 (85.7)	16 (45.7)	
>3.5 mg/dL	38 (26)	15 (18.1)	4 (14.3)	19 (54.3)	
Missing, n	8	2	3	3	
CLL-IPI					<.01
Low (0-1)	54 (43.9)	38 (56.7)	11 (42.3)	5 (16.7)	
Intermediate (2-3)	32 (20.8)	16 (23.9)	9 (34.6)	7 (23.3)	
High (4-6)	31 (20.1)	12 (17.9)	5 (19.2)	14 (46.7)	
Very high (7-10)	6 (3.9)	1 (1.5)	1 (3.8)	4 (13.3)	
Missing, n	31	18	5	8	
CLL IPS-E					.785
Low (0)	24 (27.3)	16 (25.4)	7 (36.8)	1 (16.7)	
Intermediate (1)	44 (50)	32 (50.8)	9 (47.4)	3 (50)	
High (2-3)	20 (22.7)	15 (23.8)	3 (15.8)	2 (33.3)	
Missing, n	28	16	5	7	
***IGHV* mutational status**					<.01
Mutated	90 (62.5)	58 (74.4)	22 (71)	10 (28.6)	
Unmutated	54 (37.5)	20 (25.6)	9 (29)	25 (71.4)	
Undetermined, n	9	7	–	2	
Missing, n	1	–		1	
ZAP-70					.121
<20%	88 (74.6)	54 (78.3)	20 (80)	14 (58.3)	
≥20%	30 (25.4)	15 (21.7)	5 (20)	10 (41.7)	
Missing, n	36	16	6	14	
CD38					.011
<30%	117 (84.2)	69 (90.8)	24 (85.7)	24 (68.6)	
≥30%	22 (15.8)	7 (9.2)	4 (14.3)	11(31.4)	
Missing, n	15	9	3	3	
FISH analysis					
17 deletion	11 (7.9)	6 (8.1)	3 (10.3)	2 (5.4)	
11q deletion	14 (10)	8 (10.8)	3 (10.3)	3 (8.1)	
13q deletion	77 (55)	38 (51.4)	21 (72.4)	18 (48.6)	
Trisomy 12	26 (18.6)	18 (24.3)	5 (17.2)	3 (8.1)	
Missing, n	14	11	2	1	

P values are for comparisons across the 3 risk groups determined by the CLL15 score.
CLL-IPI, International Prognostic Index for CLL; FISH, fluorescence in situ hybridization.

Table 1 (continued)

	Total cohort (n = 154), n (%)	CLL15 categories			*P* value
		Low-risk group (n = 85), n (%)	Intermediate-risk group (n = 31), n (%)	High-risk group (n = 38), n (%)	
Complex karyotype (≥3 abnormalities)					.877
No	63 (90)	42 (89.4)	14 (93.3)	7 (87.5)	
Yes	7 (10)	5 (10.6)	1 (6.7)	1 (12.5)	
Missing, n	84	38	16	30	
TP53 mut					.009
No	92 (92)	61 (98.4)	20 (83.3)	11 (78.6)	
Yes	8 (8)	1 (1.6)	4 (16.7)	3 (21.4)	
Missing, n	54	23	7	24	
NOTCH1 mut					.351
No	86 (82.7)	49 (79)	23 (92)	14 (82.4)	
Yes	18 (17.3)	13 (21)	2 (8)	3 (17.6)	
Missing, n	50	23	6	21	
SF3B1 mut					.066
No	92 (94.8)	61 (98.4)	21 (91.3)	10 (83.3)	
Yes	5 (5.2)	1 (1.6)	2 (8.7)	2 (16.7)	
Missing, n	57	23	8	26	
MYD88 mut					.395
No	85 (94.6)	60 (96.8)	21 (91.3)	7 (87.5)	
Yes	5 (5.4)	2 (3.2)	2 (8.7)	1 (12.5)	
Missing, n	61	23	8	30	
Median follow-up, months (mo)	43.8	43.6	43.8	45.9	.61

P values are for comparisons across the 3 risk groups determined by the CLL15 score.
CLL-IPI, International Prognostic Index for CLL; FISH, fluorescence in situ hybridization.

with the highest impact in TtFT, we performed a least absolute shrinkage and selection operator regression using package *glmnet* in the R software to build the most parsimonious multivariate model. Imputation of random missing values was carried out via the *mice* R package (supplemental Table 3). The median follow-up was calculated using the reverse Kaplan-Meier method. All analyses were performed using the R statistical software version 3.6.2.

Results

Generation of a prognostic model based on gene expression: the CLL15 assay

The training cohort was comprised of 156 patients with previously untreated CLL. The median age of the series was 66 years (range, 34-90 years), and 57% of the patients were men. In total, 37% of samples were obtained at the time of CLL diagnosis, whereas 63% were obtained during the follow-up of patients before any CLL treatment. The median time from CLL diagnosis to sample collection was 11.9 months (95% CI, 7.1-22.6). The analysis of TtFT was calculated from the date of collecting the sample to the date of treatment onset. The main clinical and biological characteristics of the series are shown in Table 1. Ninety-two cases (59%) were *IGHV* mutated, 54 cases (35%) were *IGHV* unmutated, and 9 cases (6%) were undetermined because of polyclonal, unproductive, or biclonal rearrangement. In 1 case, no *IGHV* mutational data were obtained.

Digital gene expression for 178 genes of interest and 22 housekeeping genes (supplemental Table 1) was determined in 156 samples from the training cohort. Adequate gene expression was obtained in 154 (99%) samples. Two samples (1%) with not enough quality for expression testing were excluded from the analysis.

The expression of 76 genes was significantly associated with TtFT in univariate Cox regression analysis (adjusted *P* value controlling for false discovery rate [FDR] < .05), and 88 with FDR < .1. A total of 46 genes (FDR < .1) met the prespecified inclusion criteria and were selected for further analysis (see "Methods"). Among them, a total of 15 genes (*MYC, ITGA4, CERS6, ZNF471, ZNF667, SEPT10P1, ZAP70, LTK, CCL3, CNR1, EGR2, TNF, IL4R, FGL2, PPBP*) were finally selected for a prognostic model of TtFT using a penalized Cox method. In addition, 15 housekeeping genes were selected based on their low variance across the samples. A final model, named CLL15, to predict TtFT in the training cohort was developed using the expression of the 15 predictive genes normalized with the 15 housekeeping genes (Figure 1). Subsequently, a linear equation comprising log-transformed, normalized gene expression levels of the 15 genes multiplied by their respective regression coefficients was established and calculated for each patient of the training cohort to obtain the CLL15 score. The C-statistic for the model was 0.77. Figure 2A shows the shape of the association between the CLL15 score and TtFT risk after relaxing the linearity assumption for continuous variables. As a

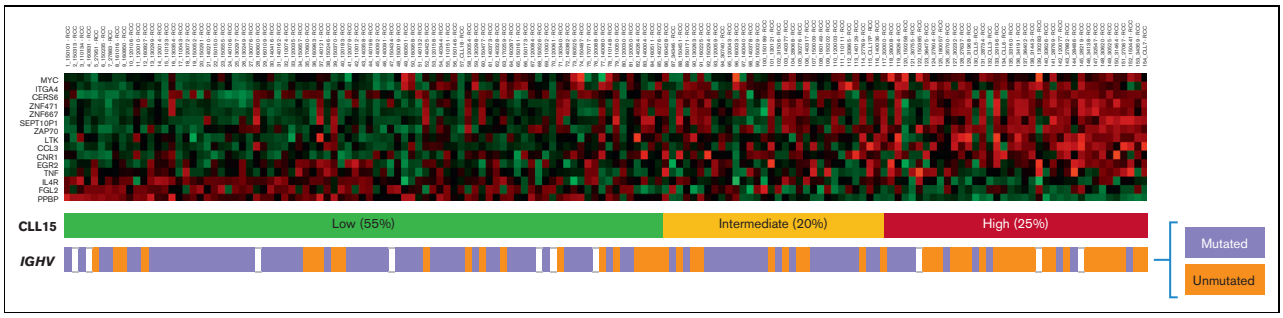


Figure 1. The gene expression–based model to predict TtFT in patients with CLL. A heatmap of the CLL15 assay with 15 informative genes shown as rows and 154 patient samples as columns. The 3 patient groups identified by the assay are shown below the heatmap together with the mutational status of the *IGHV* genes.

continuous variable, the CLL15 assay score was associated with TtFT (HR, 2.83; 95% CI, 2.17-3.68; $P < .001$). To better stratify the risk of progression, the optimal thresholds for defining 3 groups with differentiated outcomes (TtFT) were determined using the R *partykit* package. The low-risk group (score ≤ 2.718, comprising 55% of the cohort) had a 5-year estimated risk of treatment initiation of 30.5%. In the intermediate-risk group (score ≤ 3.535 and >2.718, comprising 20% of the cohort), the 5-year estimated risk of treatment initiation was 57.8% (HR, 2.67; 95% CI, 1.39-5.10; $P = .003$). Finally, in the high-risk group (score > 3.535, comprising 25% of the cohort) the 5-year estimated risk of treatment initiation was 93.4% (HR, 10.9; 95% CI, 6.12-19.3; $P < .001$) (Figure 2B). Notably, the CLL15 score exhibited a similar prognostic capacity in the subgroup of patients with an early clinical stage (n = 116), with a 5-year estimated risk of treatment initiation of 18.2%, 44.8%, and 79.54% in the low-, intermediate-, and high-risk groups, respectively (Figure 2C).

The prognostic value of the CLL15 score is independent of the *IGHV* mutational status and IPS-E CLL

We analyzed the association between the progression risk groups obtained by the CLL15 assay with known biological prognostic factors in CLL, including the most common chromosomal alterations determined by FISH (del17p, del11q, and trisomy 12), the level of protein expression of ZAP-70 and CD38 determined by flow cytometry, the mutations in *TP53*, *NOTCH1*, *SF3B1, and MYD88 genes,* the mutational status of *IGHV*, CLL-IPI, and the IPS-E CLL score.

In the univariate analysis, several factors such as the *SF3B1* mutations, *IGHV* status, the expression of ZAP-70 and CD38 by flow cytometry, clinical stage (RAI and Binet), the CLL-IPI, and the IPS-E score were associated with TtFT (Figure 3). In the final multivariate analysis, the CLL15 score, the IPS-E CLL, and the Binet stage were the only factors that maintained their independent statistical significance (Figure 3).

We subsequently explored the introduction of the mutational status of the *IGHV* (mutated/unmutated) as a variable in an expression model and compared its performance with that of a previous model of only gene expression. The C-statistic for the combined model was 0.79, and the analysis of deviance showed that the addition of *IGHV* status to the gene expression score (and vice versa) provided significant predictive information (analysis of deviance $P < .001$). According to these results, the model combining gene expression with the *IGHV* variable performed better in predicting TtFT than the models of gene expression and *IGHV* by themselves. In the pairwise multivariate Cox models, both variables, *IGHV* mutational status, and the categorized groups of progression risk according to the gene expression model contributed prognostically (Figure 2D; supplemental Table 4).

The inclusion of the CLL15 score also improved the capacity to predict TtFT of the IPS-E score. Figure 4A shows the increment in discrimination capacity in terms of C-statistic when the CLL15 score was included in the model concurrently with the IPS-E score or *IGHV* status. Moreover, in pairwise multivariate Cox models, the CLL-IPI and CLL15 also independently contributed to TtFT in the training cohort, with a C-statistic of 0.73 for the CLL-IPI alone and 0.81 for the combination. However, when the IPS-E score was included, the information on the CLL-IPI did not improve the model (supplemental Table 4). Finally, the (1) CLL15 score, (2) *IGHV* status, and (3) IPS-E score were all independent factors that improved the prediction of TtFT (all analyses of deviance pairwise comparison, $P < .01$) (Figure 4B).

Validation and reproducibility of the CLL15 assay

The CLL15 assay was then validated in cryopreserved samples from 112 patients from an independent cohort from Heidelberg (supplemental Table 5). As a continuous variable, the CLL15 score was significantly associated with TtFT (HR, 1.71; 95% CI, 1.15-2.52; $P = .007$). Figure 5A shows the association between CLL15 score and TtFT risk after relaxing the linearity assumption in the validation cohort. Using the preestablished cut-off in the training cohort, the assay assigned 22 (19.6%) patients to the low-risk group, 42 (37.5%) to the intermediate-risk group, and 48 (42.9%) to the high-risk group. These 3 groups presented differentiated outcomes with a 60-month estimated risk of treatment initiation of 16.5%, 40%, and 58.1% in the low-, intermediate-, and high-risk groups, respectively ($P = .03$ overall log-rank test, Figure 5B). Moreover, as observed in the training cohort, the gene expression information, both as a continuous variable and as a risk group, was an independent prognostic factor in the presence of *IGHV* mutational status (supplemental Table 6). The C-statistic for the *IGHV* mutational status and the gene expression model was 0.6 and 0.63, respectively, whereas the C-statistic for the combined model was 0.67. As observed in the training cohort, 3 risk groups were identified by combining the CLL15 score and the *IGHV* mutational status information (supplemental Figure 3). To

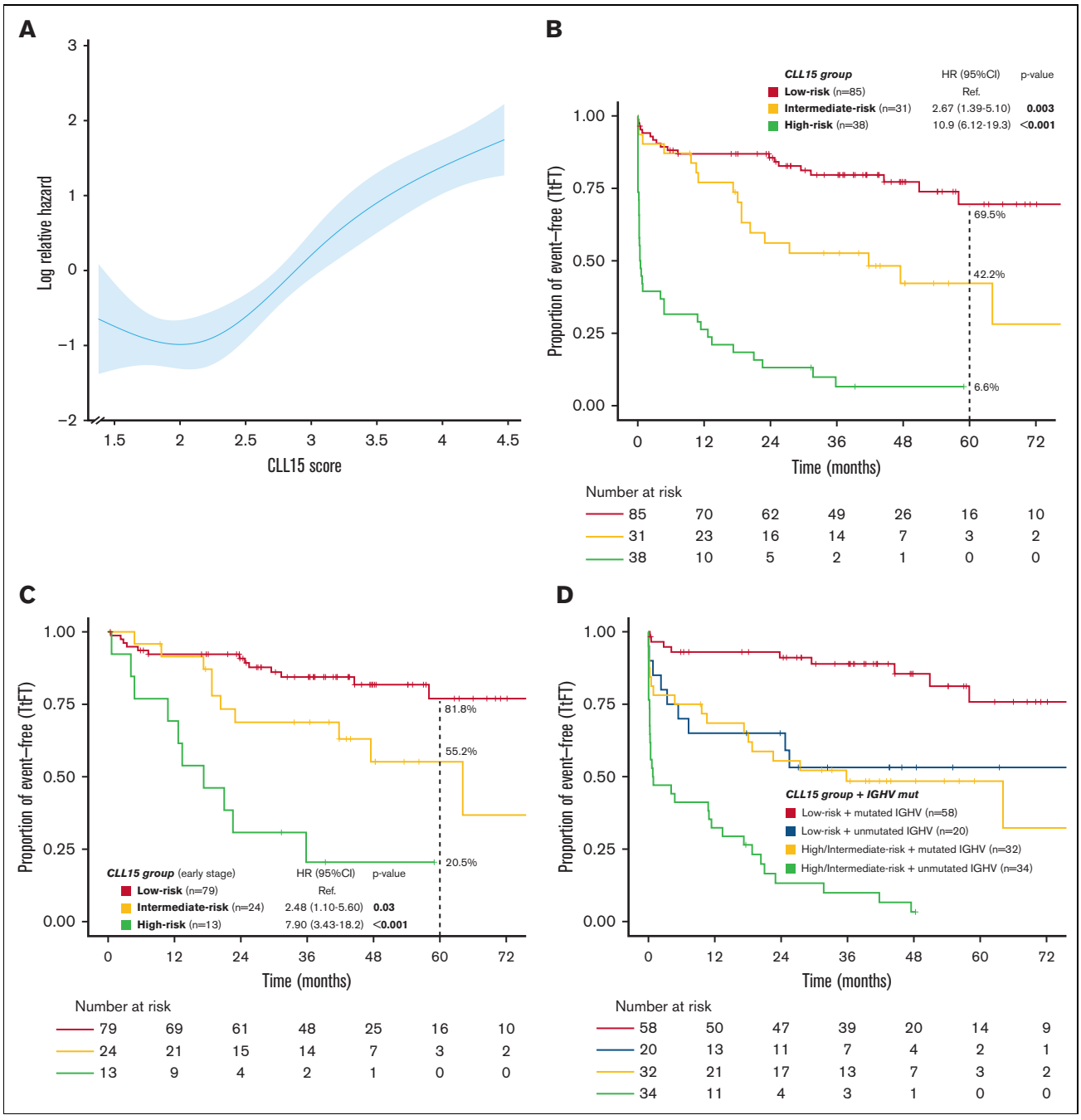


Figure 2. Time to first treatment in the training cohort. (A) Log-relative hazard according to the CLL15 score. (B) Kaplan-Meier curves of the TtFT of the 3 patient groups identified by the CLL15 assay. (C) Kaplan-Meier curves of the TtFT of the 3 patient groups identified by the CLL15 assay in the subgroup of patients with early-stage disease (Binet A 0/I). (D) Kaplan-Meier curves of the TtFT according to CLL15 assay and *IGHV* mutational status.

determine the reproducibility of the CLL15 assay, we selected 9 samples with scores distributed across the assay (low risk, intermediate risk, and high risk). The RNA from each of the samples was run on the CLL15 assay in triplicate, with each run performed on a different NanoString cartridge. The results showed 100% concordance of risk-group assignment across triplicates (supplemental Figure 4), with a standard deviation of 0.073 points.

Discussion

In this study, we translated a gene expression prognostic signature comprising genes involved in the microenvironment activation and *IGHV* mutational status into a test applicable to categorize patients into the differentiated risk of progression and requiring treatment for their CLL. The assay demonstrated the ability to identify patients

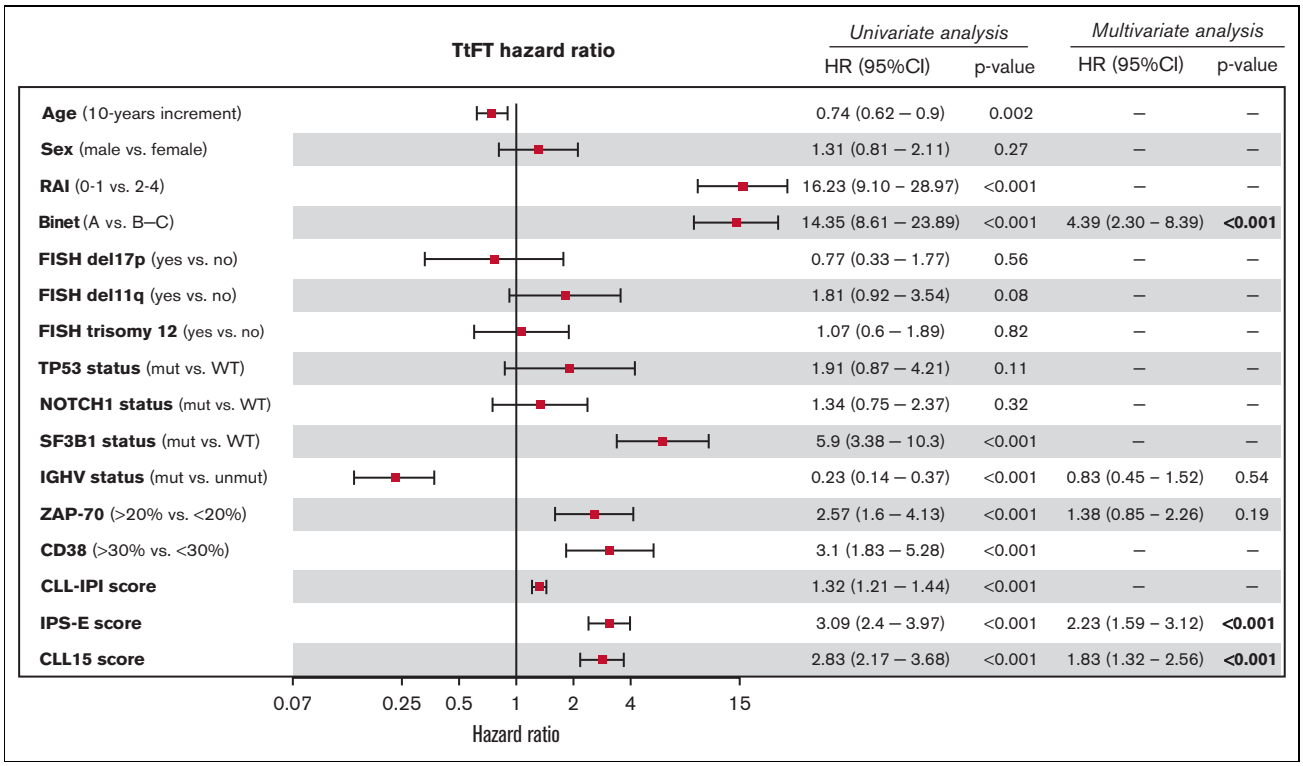
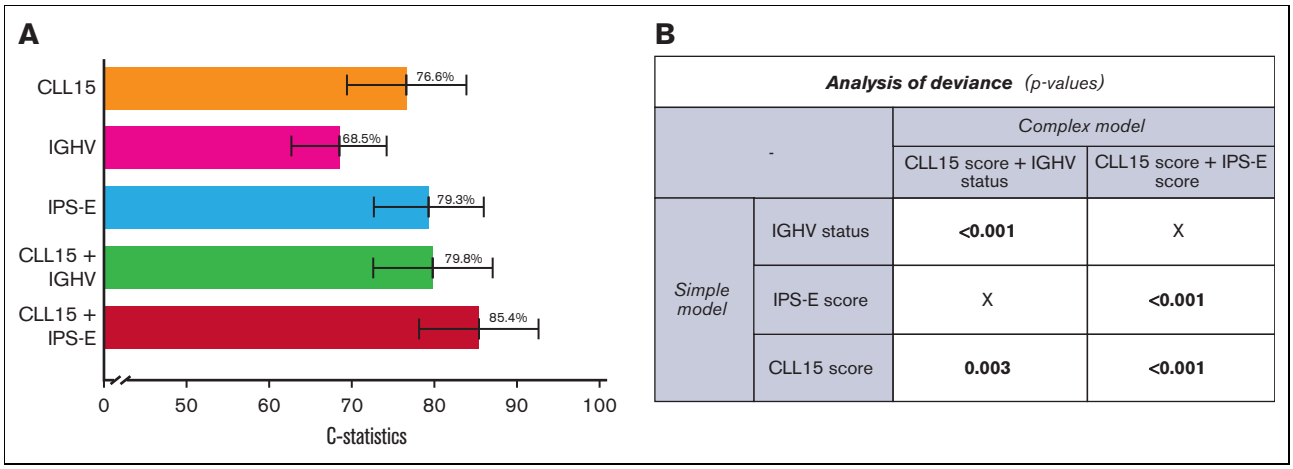


Figure 3. Univariate and multivariate analysis for TtFT according to prognostic factors in CLL.

at a high risk of requiring treatment in a short time or with an extremely stable disease.

Based on the enormous advances in the biology and treatment of CLL, classical staging systems have been complemented by a plethora of new prognostic parameters based on CLL genetics and biology, including gene expression profiles.[14,15,37,38] Despite the fact that gene expression profiles have been strongly correlated with the clinical course of the patients,[7,8,17-26] their translational value in clinical practice has been difficult to implement due to methodological reasons. The recent advent of new platforms such as NanoString nCounter, capable of digital, direct quantification on a real-time basis for individual patients, allows the attainment of gene expression analysis in a clinical setting.[27,28] In this regard, we demonstrated the clinical strength and reproducibility of the CLL15 assay in an independent cohort of previously untreated patients with CLL and its analytical reproducibility by showing a very low variability across repeated measurements.

Several in vitro and in vivo data indicate that CLL is a malignancy highly dependent on microenvironment signals for survival and proliferation, with BCR signaling being the most prominent



B

Analysis of deviance (*p-values*)			
-		*Complex model*	
		CLL15 score + IGHV status	CLL15 score + IPS-E score
Simple model	IGHV status	**<0.001**	X
	IPS-E score	X	**<0.001**
	CLL15 score	**0.003**	**<0.001**

Figure 4. Different models to predict TtFT according to CLL15 score, *IGHV* and IPS-E CLL score. (A) Discrimination capacity in terms of C-statistic according to models including CLL15 score, *IGHV* mutational status, and IPS-E CLL score. (B) Pairwise ANOVA comparisons. ANOVA, analysis of variance.

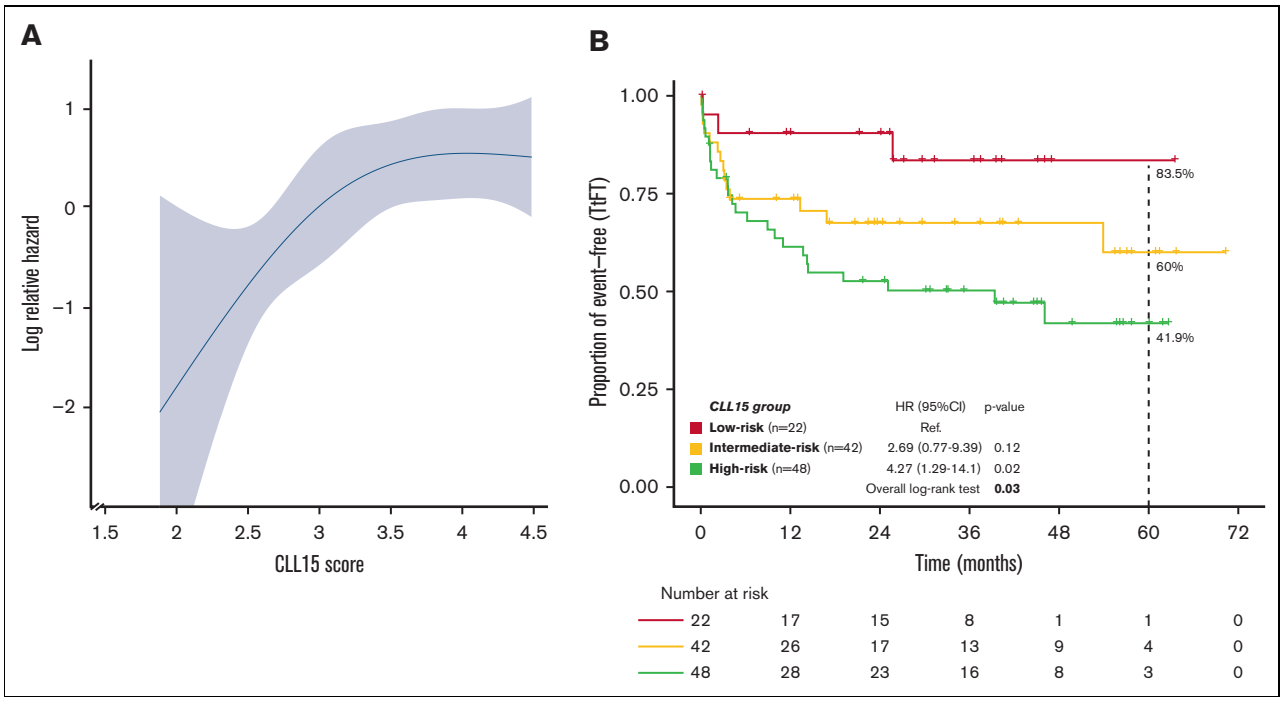


Figure 5. Time to first treatment in the validation cohort. (A) Log-relative hazard according to the CLL15 score in the validation cohort. (B) Kaplan-Meier curves of the TtFT of the 3 patient groups in the validation cohort identified by the CLL15 assay.

pathway activated in CLL cells isolated from lymph nodes.[25] The role of the microenvironment in CLL pathogenesis has been reinforced when molecules targeting CLL-microenvironment interactions have shown unprecedented therapeutic results.[30,31] The CLL15 assay included genes coding for cytokines, chemokines, and cytokines receptors such as *CCL3*, *TNF*, *PPBP,* and *IL4R*; integrins such as *ITGA4*; and transcription regulatory factors such as *MYC* and *EGR2*, which are involved in microenvironment activation in different studies in CLL.[23-26,39-41] In addition, genes previously reported to be differentially expressed according to the *IGHV* mutational status, including *CERS6*, *CNR1*, *FGL2*, *LTK*, *SEPT10P1*, *ZAP70*, *ZNF471*, and *ZNF667,* were also selected in the CLL15 assay.[17,18,24,26,41-44] Notably, the levels of expression of the aforementioned genes could also be regulated in microenvironment activation processes.[18,25,45] Thus, *ZAP-70* expression has been associated with enhanced and prolonged BCR signaling,[46,47] higher responsiveness to chemokines [56-58], and enhanced migration of CLL cells,[48,49] reinforcing the notion that increased *ZAP-70* expression is associated with a more aggressive clinical course of patients with CLL.[37,50,51]

It is worth mentioning that the CLL15 signature kept its predictive value independent of the *IGHV* mutational status, the CLL-IPI, and the IPS-E CLL score. More importantly, the inclusion of the CLL15 score improved the discrimination capacity to predict TtFT when *IGHV* or IPS-E was included in the model, suggesting that the CLL15 signature could complement the prognostic value of these other variables. In addition, the combination of the CLL15 and CLL-IPI provided independent predictive information; however, with the inclusion of the IPS-E score, the information of the CLL-IPI did not contribute prognostically to the model. In this sense, the combination of the IPS-E and the CLL15 assay was highly discriminative for the TtFT, with a C-statistic of 0.85. It appears that combining a more clinical–based score, such as the IPS-E, with a molecular score (CLL15) could increase the accuracy of both models. Unfortunately, the IPS-E score was not available for the validation cohort and this comparison could not be validated in this cohort.

On the other hand, the combination of *IGHV* and CLL15 also improved the predictive capacity of the model. Three clearly different risk groups were identified after combining the CLL15 and *IGHV* status. However, a limited improvement of the C-statistic was observed, and the lower statistical power in the validation cohort did not allow for the validation of all findings.

Currently, one of the moving fields is the possibility of early treatment of patients at early stages that are likely to progress within a short period.[5] The selection of these patients is usually based on standard prognostic scores. The usage of more accurate methods for prognostication, such as the CLL15 score, should allow for better identification of patients with an increased risk of early progression and thus support future trials based on risk-adapted therapeutic intervention.

In conclusion, the biological prognostication of CLL relies on the use of genetic aberrations together with the mutational status of *IGHV*. Unfortunately, the use of gene expression profiles has been difficult owing to its technical difficulties and reproducibility, precluding its use in clinical practice. The use of newer and more reproducible methods to assess gene expression could round off well-established prognostic parameters, appraising the entire biological profile for the prognostication of patients with CLL. The study presented herein successfully translates previously described gene expression signatures with strong prognostic value

into a new gene expression–based assay, the CLL15, applicable in the routine diagnostic setting.

Acknowledgments

The authors thank the patients for their blood donations. The authors would also like to thank the Cellex Foundation for providing research facilities and equipment and the CERCA Programme / Generalitat de Catalunya for institutional support. Part of the statistical analysis was performed in the Statistics and Bioinformatics Unit (UEB) of the Vall Hebron Hospital Research Institute.

This work was supported by research funding from the Asociación Española Contra el Cáncer grant [5U01CA157581-05, ECRIN-M3 - A29370] and in part by the Instituto de Salud Carlos III, Fondo de Investigaciones Sanitarias [PI17/00950, M.C., PI17/00943, F.B, PI18/01392, P.A.], and the Spanish Ministry of Economy and Competitiveness [CIBERONC-CB16/12/00233], the Education Council or Health Council of the Junta de Castilla y León [GRS 2036/A/19], and Gilead Sciences [GLD15/00348]. This work was supported by research funding from the Asociación Española Contra el Cáncer grant [5U01CA157581-05, ECRIN-M3 - A29370] and in part by the Instituto de Salud Carlos III, Fondo de Investigaciones Sanitarias [PI17/00950, M.C., PI17/00943, F.B, PI18/01392, P.A.], and the Spanish Ministry of Economy and Competitiveness [CIBERONC-CB16/12/00233], the Education Council or Health Council of the Junta de Castilla y León [GRS 2036/A/19], Gilead Sciences [GLD15/00348] and Gilead Fellowships [GLD16/00144, GLD18/00047, F.B.], and Fundació la Marató de TV3 [201905-30-31 F.B]. All Spanish funding was cosponsored by the European Union FEDER program "Una manera de hacer Europa". M.C. holds a contract from Ministerio de Ciencia, Innovación y Universidades [RYC-2012-2018].

Authorship

Contribution: P.A., T.Z., M.C., and F.B. designed and supervised this work; P.A., M.C., J.L., T.Z., M.A., M.G., G.I., S.B., and A.M.-N. provided samples; D.M., J.C., J.B., and B.T.-V. performed experiments; P.A., G.V., M.C., and F.B. analyzed and interpreted the data; P.A., G.V., M.C., and F.B. wrote the manuscript; and all authors revised the manuscript.

Conflict-of-interest disclosure: P.A. received honoraria from Janssen, Roche, Celgene, AbbVie, and AstraZeneca. G.V. received research honoraria for speaker activities from MSD and an advisory role from AstraZeneca. M.A. received honoraria for speaker activities from AstraZeneca, an advisory role from Janssen, and nonfinancial support from Janssen, and AbbVie. A.M.-N. received honoraria from Janssen, Roche, Takeda, Gilead, AbbVie, and Celgene for speaker activities and from Janssen, Takeda, Gilead, Kiowa Kirin, AstraZeneca, and Beigene for participating in advisory boards. M.C. received research funding from Janssen, Roche, and AstraZeneca. F.B. received honoraria and research grants from Roche, Celgene, Takeda, AstraZeneca, Novartis, AbbVie, Lilly, Beigene, and Janssen. The remaining authors declare no competing financial interests.

ORCID profiles: D.M., 0000-0001-9421-949X; G.V., 0000-0003-4868-6585; J.L., 0000-0002-9211-0746; G.I., 0000-0003-0805-9288; A.M.-N., 0000-0002-7592-6815; M.G., 0000-0001-6637-1072; M.C., 0000-0002-7732-7808; F.B., 0000-0001-9241-2886.

Correspondence: Pau Abrisqueta, Department of Hematology, University Hospital Vall d'Hebron, Pssg Vall d'Hebron 119-129, 08035, Barcelona, Spain; email: pabrisqueta@vhio.net.

References

1. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131(25):2745-2760.
2. Landau DA, Carter SL, Stojanov P, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013;152(4):714-726.
3. Barrio S, Shanafelt TD, Ojha J, et al. Genomic characterization of high-count MBL cases indicates that early detection of driver mutations and subclonal expansion are predictors of adverse clinical outcome. *Leukemia*. 2017;31(1):170-176.
4. Gruber M, Bozic I, Leshchiner I, et al. Growth dynamics in naturally progressing chronic lymphocytic leukaemia. *Nature*. 2019;570(7762):474-479.
5. Langerbeins P, Bahlo J, Rhein C, et al. The CLL12 trial protocol: a placebo-controlled double-blind phase III study of ibrutinib in the treatment of early-stage chronic lymphocytic leukemia patients with risk of early disease progression. *Future Oncol*. 2015;11(13):1895-1903.
6. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000;343(26):1910-1916.
7. 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(6):1848-1854.
8. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999;94(6):1840-1847.
9. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475(7354):101-105.
10. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365(26):2497-2506.
11. Condoluci A, Terzi di Bergamo L, Langerbeins P, et al. International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia. *Blood*. 2020;135(21):1859-1869.
12. Wierda WG, O'Brien S, Wang X, et al. Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia. *J Clin Oncol*. 2011;29(31):4088-4095.

13. Molica S, Giannarelli D, Gentile M, et al. External validation on a prospective basis of a nomogram for predicting the time to first treatment in patients with chronic lymphocytic leukemia. *Cancer*. 2013;119(6):1177-1185.

14. Rossi D, Rasi S, Spina V, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood*. 2013;121(8):1403-1412.

15. group IC-lw. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol*. 2016;17(6):779-790.

16. Cohen JA, Rossi FM, Zucchetto A, et al. A laboratory-based scoring system predicts early treatment in Rai 0 chronic lymphocytic leukemia. *Haematologica*. 2020;105(6):1613-1620.

17. Klein U, Tu Y, Stolovitzky GA, et al. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med*. 2001;194(11):1625-1638.

18. Rosenwald A, Alizadeh AA, Widhopf G, et al. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med*. 2001;194(11):1639-1647.

19. Haslinger C, Schweifer N, Stilgenbauer S, et al. Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J Clin Oncol*. 2004;22(19):3937-3949.

20. Friedman DR, Weinberg JB, Barry WT, et al. A genomic approach to improve prognosis and predict therapeutic response in chronic lymphocytic leukemia. *Clin Cancer Res*. 2009;15(22):6947-6955.

21. Fernandez V, Jares P, Salaverria I, et al. Gene expression profile and genomic changes in disease progression of early-stage chronic lymphocytic leukemia. *Haematologica*. 2008;93(1):132-136.

22. Morabito F, Cutrona G, Gentile M, et al. Definition of progression risk based on combinations of cellular and molecular markers in patients with Binet stage A chronic lymphocytic leukaemia. *Br J Haematol*. 2009;146(1):44-53.

23. Chuang HY, Rassenti L, Salcedo M, et al. Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. *Blood*. 2012;120(13):2639-2649.

24. Ferreira PG, Jares P, Rico D, et al. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res*. 2014;24(2):212-226.

25. Herishanu Y, Pérez-Galán P, Liu D, et al. The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood*. 2011;117(2):563-574.

26. Guarini A, Chiaretti S, Tavolaro S, et al. BCR ligation induced by IgM stimulation results in gene expression and functional changes only in IgV H unmutated chronic lymphocytic leukemia (CLL) cells. *Blood*. 2008;112(3):782-792.

27. Scott DW, Mottok A, Ennishi D, et al. Prognostic significance of diffuse large B-cell lymphoma cell of origin determined by digital gene expression in formalin-fixed paraffin-embedded tissue biopsies. *J Clin Oncol*. 2015;33(26):2848-2856.

28. Scott DW, Abrisqueta P, Wright GW, et al. New molecular assay for the proliferation signature in Mantle cell lymphoma applicable to formalin-fixed paraffin-embedded biopsies. *J Clin Oncol*. 2017;35(15):1668-1677.

29. Yosifov DY, Idler I, Bhattacharya N, et al. Oxidative stress as candidate therapeutic target to overcome microenvironmental protection of CLL. *Leukemia*. 2020;34(1):115-127.

30. Byrd JC, Furman RR, Coutre SE, et al. Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N Engl J Med*. 2013;369(1):32-42.

31. Burger JA, Tedeschi A, Barr PM, et al. Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia. *N Engl J Med*. 2015;373(25):2425-2437.

32. Dietrich S, Oleś M, Lu J, et al. Drug-perturbation-based stratification of blood cancer. *J Clin Invest*. 2018;128(1):427-445.

33. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J*. 2010;52(1):70-84.

34. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105-1117.

35. Schmid M, Kestler HA, Potapov S. On the validity of time-dependent AUC estimators. *Briefings Bioinf*. 2015;16(1):153-168.

36. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst*. 1994;86(11):829-835.

37. Crespo M, Bosch F, Villamor N, et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N Engl J Med*. 2003;348(18):1764-1775.

38. Cramer P, Hallek M. Prognostic factors in chronic lymphocytic leukemia-what do we need to know? *Nat Rev Clin Oncol*. 2011;8(1):38-47.

39. Mittal AK, Chaturvedi NK, Rai KJ, et al. Chronic lymphocytic leukemia cells in a lymph node microenvironment depict molecular signature associated with an aggressive disease. *Mol Med*. 2014;20:290-301.

40. Zucchetto A, Sonego P, Degan M, et al. Surface-antigen expression profiling (SEP) in B-cell chronic lymphocytic leukemia (B-CLL): identification of markers with prognostic relevance. *J Immunol Methods*. 2005;305(1):20-32.

41. Mansouri L, Gunnarsson R, Sutton LA, et al. Next generation RNA-sequencing in prognostic subsets of chronic lymphocytic leukemia. *Am J Hematol*. 2012;87(7):737-740.

42. Abruzzo LV, Barron LL, Anderson K, et al. Identification and validation of biomarkers of IgV(H) mutation status in chronic lymphocytic leukemia using microfluidics quantitative real-time polymerase chain reaction technology. *J Mol Diagn*. 2007;9(4):546-555.

43. Fält S, Merup M, Tobin G, et al. Distinctive gene expression pattern in VH3-21 utilizing B-cell chronic lymphocytic leukemia. *Blood*. 2005;106(2):681-689.

44. Kienle D, Benner A, Kröber A, et al. Distinct gene expression patterns in chronic lymphocytic leukemia defined by usage of specific VH genes. *Blood*. 2006;107(5):2090-2093.

45. Lanham S, Hamblin T, Oscier D, Ibbotson R, Stevenson F, Packham G. Differential signaling via surface IgM is associated with VH gene mutational status and CD38 expression in chronic lymphocytic leukemia. *Blood*. 2003;101(3):1087-1093.

46. Chen L, Huynh L, Apgar J, et al. ZAP-70 enhances IgM signaling independent of its kinase activity in chronic lymphocytic leukemia. *Blood*. 2008;111(5):2685-2692.

47. Gobessi S, Laurenti L, Longo PG, Sica S, Leone G, Efremov DG. ZAP-70 enhances B-cell-receptor signaling despite absent or inefficient tyrosine kinase activation in chronic lymphocytic leukemia and lymphoma B cells. *Blood*. 2007;109(5):2032-2039.

48. Richardson SJ, Matthews C, Catherwood MA, et al. ZAP-70 expression is associated with enhanced ability to respond to migratory and survival signals in B-cell chronic lymphocytic leukemia (B-CLL). *Blood*. 2006;107(9):3584-3592.

49. Calpe E, Codony C, Baptista MJ, et al. ZAP-70 enhances migration of malignant B lymphocytes toward CCL21 by inducing CCR7 expression via IgM-ERK1/2 activation. *Blood*. 2011;118(16):4401-4410.

50. Rassenti LZ, Jain S, Keating MJ, et al. Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood*. 2008;112(5):1923-1930.

51. Rassenti LZ, Huynh L, Toy TL, et al. ZAP-70 compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. *N Engl J Med*. 2004;351(9):893-901.