blood advances

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodadvances@hematology.org

A gene expression signature based on microenvironment signaling to predict progression in chronic lymphocytic leukemia

Tracking no: ADV-2022-007508R1

Pau Abrisqueta (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Daniel Medina-Gil (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Guillermo Villacampa (Oncology Data Science Group, Vall d'Hebron Institute of Oncology. Barcelona, Spain, Spain) Junyan Lu (European Molecular Biology Laboratory, Germany) Miguel Alcoceba (Department of Hematology, University Hospital of Salamanca (HUS/IBSAL), CIBERONC and Center for Cancer Research-IBMCC (USAL-CSIC), Spain) Julia Carabia (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Joan Boix (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Barbara Tazón-Vega (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Gloria Iacoboni (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Sabela Bobillo (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Ana Marín-Niebla (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Marcos González (Hospital Universitario de Salamanca, Spain) Thorsten Zenz (University of Zurich, Switzerland) Marta Crespo (Departament de Medicina, Universitat Autònoma de Barcelona, Spain) Francesc Bosch (Departament de Medicina, Universitat Autònoma de Barcelona, Spain)

Abstract:
Several gene expression profiles with a strong correlation with patient outcome have been previously described in chronic lymphocytic leukemia (CLL), although their applicability in clinical practice as biomarkers has been particularly limited. Here we describe the training and validation of a gene expression signature for predicting early progression of 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 time to first treatment (TtFT) (HR: 2.83, 95%CI 2.17-3.68; p<0.001). The prognostic value of the CLL15 score (HR 1.71, [95%CI 1.15-2.52]; p=0.007) was further confirmed in an external independent validation cohort (n=112). Of note, the CLL15 score improved the prognostic capacity over the *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<0.001) and the IPS-E CLL (HR: 2.23, 95%CI 1.59-3.12; p<0.001) were independently associated with TtFT. The newly developed and validated CLL15 assay successfully translates previous gene signatures, such as the microenvironment signaling, into a new gene expression-based assay with prognostic implications in CLL.

Conflict of interest: COI declared - see note

COI notes: P.A. has received honoraria from Janssen, Roche, Celgene, Abbie and Astrazeneca. G.V has received research honoraria for speaker activities from MSD and advisory role from Astrazeneca. M.A has received honoraria for speaker activities from Astrazeneca, advisory role from Janssen and non-financial support from Janssen and Abbvie. AMN. has received honoraria from Janssen, Roche, Takeda, Gilead, Abbvie, Celgene for speaker activities and from Janssen, Takeda, Gilead, Kiowa Kirin, Astra-Zeneca and Beigene for participating in advisory boards. M.C. has received research funding from, Janssen, Roche and AstraZeneca. F.B has received honoraria and research grants from Roche, Celgene, Takeda, AstraZeneca, Novartis, AbbVie, Lilly, Beigene, and Janssen. All remaining authors have declared no conflicts of interest.

Preprint server: No;

Author contributions and disclosures: Designed research and supervised the work: PA, TZ, MC and FB. Provided samples: PA, MC, JL, TZ, MA, MG, GI, SB, AMN. Performed experiments: DM, JC, JB, BTV. Analyzed and interpreted data: PA, GV, MC, FB. Wrote the manuscript: PA, GV, MC, FB. Revised the manuscript: all authors.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: emails to the corresponding author

Clinical trial registration information (if any):

Article type: Original article

Title:

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

Authors: 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*] Francesc Bosch[1,2*]

Affiliations:

[1]Servei d'Hematologia, Vall d'Hebron Hospital Universitari, Experimental Hematology, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain

[2]Departament de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

[3]Experimental Hematology, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Barcelona Hospital Campus, C/ Natzaret, 115-117, 08035 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

[7]Department of Medical Oncology and Hematology, University Hospital and University of Zurich, Zurich, Switzerland.

* These authors contributed equally to this work

Running Title: CLL15 assay for time to progression in CLL

Contact the corresponding author for data sharing: pabrisqueta@vhio.net.

Corresponding Author:

Dr. Pau Abrisqueta,
Department of Hematology, University Hospital Vall d'Hebron, Pssg Vall d'Hebron 119-129, 08035, Barcelona, Spain;
Phone: +34 934893000 ext 4895

e-mail: pabrisqueta@vhio.net

Word Count:
Abstract: 191/250
Main Text: 3189/4000

Figures: 5 (main), 4 (supplemental)
Tables: 1 (main), 6 (supplemental)

References: 51/100

Abstract:

Several gene expression profiles with a strong correlation with patient outcome have been previously described in chronic lymphocytic leukemia (CLL), although their applicability in clinical practice as biomarkers has been particularly limited. Here we describe the training and validation of a gene expression signature for predicting early progression of 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 time to first treatment (TtFT) (HR: 2.83, 95%CI 2.17-3.68; p<0.001). The prognostic value of the CLL15 score (HR 1.71, [95%CI 1.15-2.52]; p=0.007) was further confirmed in an external independent validation cohort (n=112). Of note, the CLL15 score improved the prognostic capacity over the *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<0.001) and the IPS-E CLL (HR: 2.23, 95%CI 1.59-3.12; p<0.001) were independently associated with TtFT. The newly developed and validated CLL15 assay successfully translates previous gene signatures, such as the microenvironment signaling, into a new gene expression-based assay with prognostic implications in CLL.

Key Points

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

Introduction

It is well accepted that patients with chronic lymphocytic leukemia (CLL) that are asymptomatic and in early clinical phase do not require therapy.(1) Nevertheless, cumulative data on the risk of clonal evolution (2-4) renewed the interest about an early therapeutic intervention on patients at diagnosis who are likely to progress rapidly.(5) Therefore, the identification at diagnosis of these patients has been an intense focus of clinical research in the field of CLL. Prognostication in this setting has classically relied upon a myriad of laboratory values, cytogenetic abnormalities, gene mutations or the mutational status of the immunoglobulin genes (*IGHV*).(6-10) More recently, the International Prognostic Score for Early-stage CLL [IPS-E] has been developed employing three covariates, unmutated *IGHV*, absolute lymphocyte count > 15 × 10^9/L, and presence of palpable lymph nodes.(11)

In spite of all this extensive investigation, the accuracy of these models may be improved.(11-16) In addition, the emergence of novel targeted agents has reignited interest in early treatment of patients with high risk of early progression.(5)

Gene expression profile and the clinical course of the patients with CLL have been correlated in different studies.(7, 8, 17-26) Regrettably, biomarkers based on gene expression profiles exhibit several caveats that preclude them from being widely applicable in the prognostication of CLL patients. 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 can vary when different patients are included during 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 since CLL is a malignancy particularly dependent on the interaction with their 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 the 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 BTK inhibitors.(30, 31)

Herein, we developed, evaluated and validated a multi-gene expression signature using genes associated with the activation of CLL cells in the microenvironment and the *IGHV* mutational status. This assay, based on a NanoString platfom, 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 the new assay for the risk of progression in CLL patients is shown in Supplemental Fig 1. For the training cohort of the study, 156 untreated samples, 119 from University Hospital Vall d'Hebron and 37 from University of Salamanca were used. The assay was validated using 112 samples from an independent cohort of patients from the German Cancer Research Centre, Heidelberg. Details of the validation cohort have been reported elsewhere.(32) Samples were obtained at diagnosis whenever possible. For patients that did not have a sample at the moment of diagnosis, samples during the follow-up were collected 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 (PBMC) were obtained by Ficoll-Paque Plus (GE Healthcare, Buckinghamshire, United Kingdom) density gradient 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.

A 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.

147

148

149 ***Gene expression analysis***

150 Gene expression was quantified in 250 ng of RNA on the NanoString platform
151 (NanoString Technologies, WA) using the "high sensitivity" setting on the nCounter™
152 PrepStation and 555 fields of view (FOV) on the nCounter™ Digital Analyzer. One
153 hundred seventy-eight genes were selected from the literature, including genes related
154 to the activation of CLL cells in the microenvironment,(23-26) genes differentially
155 expressed according to the mutational status of *IGHV*,(7, 8, 17, 18) and other genes with
156 prognostic interest in CLL (Supplementary methods, Supplemental Table S1).
157 Normalization for RNA loading was performed using the geometric mean of 22
158 housekeeping genes (Supplemental Table S1). The normalized data were log10
159 transformed. Reference gene selection is further described in the Data Supplement.

160

161 ***Predictive gene expression score***

162 Detailed descriptions of model building and model performance assessment are
163 provided in the Data Supplement. In brief, we used the gene expression data from the
164 training cohort to produce a parsimonious predictive model for Time to First Treatment
165 (TtFT) using a penalized Cox model.(33) To evaluate the global performance of the
166 multivariate Cox model obtained from the selected genes, different diagnostic
167 parameters have been calculated and are summarized in the Data supplement (Table
168 S2), including R^2, the Brier score, iAUC (a summary measure of the area under the

receiver operating characteristic [ROC] curve calculated for the different times), and the 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 illustration purposes, we dichotomized the predictive gene expression score in three risk-groups using R *partykit* package.

Statistical Analysis

The statistical analysis plan was pre-specified prior to the evaluation of the gene expression from the training and validation cohort. The primary endpoint 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]). The Harrell's C-statistic was calculated to compare the discrimination capacity of different models. The analysis of deviance (*anova* R function) was used to study if 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 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) In order to select prognostic variables with the highest impact in TtFT, we performed a least absolute shrinkage and selection operator (LASSO) regression using package *glmnet* in R software to build the most parsimonious

multivariate model. Imputation of random missing values was carried out via the *mice* R package (Supplemental Table S3). The median follow-up was calculated using the reverse Kaplan-Meier method. All analyses were performed using 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. Median age of the series was 66 years (range, 34 to 90 years) and 57% were men. A total of 37% of samples were obtained at diagnosis of CLL, while 63% of samples were obtained during the follow-up of the patients prior to any CLL treatment. 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 obtaining 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, whereas 9 cases (6%) were undetermined because of polyclonal, unproductive or biclonal rearrangement. In 1 case no *IGHV* mutational data was obtained.

Digital gene expression for 178 genes of interest and 22 housekeeping genes (Supplement Table S1) was determined in the 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] <0.05), and 88 with a FDR <0.1. A total of 46 genes (FDR <0.1) met the pre-specified 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 also 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). Then, a linear equation comprising the 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 association between the CLL15 score and TtFT risk after relaxing the linearity assumption for continuous variables. As a continuous variable, the CLL15 assay score was associated with TtFT (HR: 2.83, 95%CI 2.17-3.68; p <0.001). To better stratify the risk of progression, 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= 0.003). 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 <0.001) (Figure 2B). Notably, the CLL15 score exhibited similar prognostic capacity in the subgroup of patients with 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 of the International Prognostic Score for Asymptomatic Early-stage (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 of *TP53*, *NOTCH1*, *SF3B1 and MYD88 genes,* the mutational status of *IGHV,* the International Prognostic Index for Chronic Lymphocytic Leukemia (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 Binet stage were the only factors that maintained their independent statistical significance (Figure 3).

We next explored the introduction of the mutational status of the *IGHV* (mutated/ unmutated) as a variable in the expression model, comparing its performance with the 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) provides significant predictive information (analysis of deviance p <0.001). 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, the *IGHV* mutational status and the categorized groups of progression risk according to the gene expression model did contribute prognostically (Figure 2D and Supplementary 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 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 the CLL15 also independently contributed to TTfT in the training cohort, with a C-statistic for the CLL-IPI alone of 0.73, and of 0.81 for the combination. However, when the IPS-E score was included, the information of the CLL-IPI did not improve the model (Supplementary Table 4). Finally, i) CLL15 score, ii) *IGHV* status, and iii) IPS-E score are all independent factors that improve prediction of TtFT (all analysis of deviance pairwise comparison p <0.01) (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 S5). As a continuous

283 variable, the CLL15 score was significantly associated with TtFT (HR 1.71, [95%CI 1.15-
284 2.52]; p= 0.007). Figure 5A shows the association between CLL15 score and TtFT risk
285 after relaxing the linearity assumption in the validation cohort. Using the pre-
286 established cut-off in the training cohort, the assay assigned 22 (19.6%) patients to the
287 low-risk group, 42 (37.5%) to intermediate-risk and 48 (42.9%) to the high-risk group.
288 These three groups presented differentiated outcomes with 60-months estimated risk
289 of treatment initiation of 16.5%, 40% and 58.1% in the low, intermediate, and high-risk
290 groups, respectively (p= 0.03 overall log-rank test, Fig 5B). Moreover, as observed in the
291 training cohort, the gene expression information both as a continuous variable and as a
292 risk-group was an independent prognostic factor in the presence of *IGHV* mutational
293 status (Supplemental Table S6). The C-statistic for the IGHV mutational status and for
294 the gene expression model was 0.6 and 0.63, respectively, while the C-statistic for the
295 combined model was 0.67. As observed in the training cohort, three risk groups were
296 identified after combining the CLL15 score and the *IGHV* mutational status information
297 (Supplemental Figure S3).To determine the reproducibility of the CLL15 assay, we
298 selected 9 samples with scores distributed across the assay (low risk, intermediate, and
299 high risk). The RNA from each of these samples was run on the CLL15 assay in triplicate,
300 with each run performed on a different NanoString cartridge. The results showed 100%
301 concordance of risk group assignment across the triplicates (Supplemental Figure S4),
302 with a standard deviation of 0.073 points.
303

Discussion

In this study we translated a gene expression prognostic signature composed by genes involved in the microenvironment activation and *IGHV* mutational status into a test applicable to categorize patients into differentiated risk of progression and requiring treatment for their CLL. The assay demonstrated the ability to identify patients with a high risk of requiring treatment in a short time or with an extremely stable disease.

On the basis of 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) In spite 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 the NanoString nCounter, capable of a 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 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) Of note, 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 independently 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* were 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 the CLL-IPI had independent predictive information,

but 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 TtFT with a C-statistics of 0.85. It appears that combining a more clinical-based score 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 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 validating all the findings.

Nowadays, 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 of time.(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, biological prognostication in 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 penetration in clinical practice. The employment of newer

and more reproducible methods to assess gene expression could round off well-established prognostic parameters, appraising the entire biological profile in 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.

Authors' Disclosures

P.A. has received honoraria from Janssen, Roche, Celgene, Abbie and Astrazeneca. G.V has received research honoraria for speaker activities from MSD and advisory role from Astrazeneca. M.A has received honoraria for speaker activities from Astrazeneca, advisory role from Janssen and non-financial support from Janssen and Abbvie. AMN. has received honoraria from Janssen, Roche, Takeda, Gilead, Abbvie, Celgene for speaker activities and from Janssen, Takeda, Gilead, Kiowa Kirin, Astra-Zeneca and Beigene for participating in advisory boards. M.C. has received research funding from, Janssen, Roche and AstraZeneca. F.B has received honoraria and research grants from Roche, Celgene, Takeda, AstraZeneca, Novartis, AbbVie, Lilly, Beigene, and Janssen.

All remaining authors have declared no conflicts of interest.

Authors' contributions

Designed research and supervised the work: PA, TZ, MC and FB. Provided samples: PA, MC, JL, TZ, MA, MG, GI, SB, AMN. Performed experiments: DM, JC, JB, BTV. Analyzed and interpreted data: PA, GV, MC, FB. Wrote the manuscript: PA, GV, MC, FB. Revised the manuscript: all authors.

Acknowledgements

The authors thank the patients for their blood donations. Authors also would 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 has been carried out in the Statistics and Bioinformatics Unit (UEB) of the Vall Hebron Hospital Research Institute (VHIR).

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]. All Spanish funding is co-sponsored by the European Union FEDER program “Una manera de hacer Europa”. Gilead Fellowships [GLD16/00144, GLD18/00047, F.B.] and Fundació la Marató de TV3 [201905-30-31 F.B]. M.C. holds a contract from Ministerio de Ciencia, Innovación y Universidades [RYC-2012-2018].

References:

1. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood. 2018;131(25):2745-60.
2. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013;152(4):714-26.
3. Barrio S, Shanafelt TD, Ojha J, Chaffee KG, Secreto C, Kortüm KM, 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):170-6.
4. Gruber M, Bozic I, Leshchiner I, Livitz D, Stevenson K, Rassenti L, et al. Growth dynamics in naturally progressing chronic lymphocytic leukaemia. Nature. 2019;570(7762):474-9.
5. Langerbeins P, Bahlo J, Rhein C, Cramer P, Pflug N, Fischer K, 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):1895-903.
6. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. The New England journal of medicine. 2000;343(26):1910-6.
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):1848-54.
8. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood. 1999;94(6):1840-7.
9. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475(7354):101-5.
10. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med. 2011;365(26):2497-506.
11. Condoluci A, Terzi di Bergamo L, Langerbeins P, Hoechstetter MA, Herling CD, De Paoli L, et al. International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia. Blood. 2020;135(21):1859-69.
12. Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia. J Clin Oncol. 2011;29(31):4088-95.
13. Molica S, Giannarelli D, Gentile M, Cutrona G, Di Renzo N, Di Raimondo F, et al. External validation on a prospective basis of a nomogram for predicting the time

462 to first treatment in patients with chronic lymphocytic leukemia. Cancer.
463 2013;119(6):1177-85.
464 14. Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated
465 mutational and cytogenetic analysis identifies new prognostic subgroups in chronic
466 lymphocytic leukemia. Blood. 2013;121(8):1403-12.
467 15. group IC-Iw. An international prognostic index for patients with chronic
468 lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. Lancet
469 Oncol. 2016;17(6):779-90.
470 16. Cohen JA, Rossi FM, Zucchetto A, Bomben R, Terzi-di-Bergamo L, Rabe KG, et
471 al. A laboratory-based scoring system predicts early treatment in Rai 0 chronic
472 lymphocytic leukemia. Haematologica. 2020;105(6):1613-20.
473 17. Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretti G, Husson H, et al. Gene
474 expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous
475 phenotype related to memory B cells. J Exp Med. 2001;194(11):1625-38.
476 18. Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X, et al. Relation
477 of gene expression phenotype to immunoglobulin mutation genotype in B cell
478 chronic lymphocytic leukemia. J Exp Med. 2001;194(11):1639-47.
479 19. Haslinger C, Schweifer N, Stilgenbauer S, Döhner H, Lichter P, Kraut N, et al.
480 Microarray gene expression profiling of B-cell chronic lymphocytic leukemia
481 subgroups defined by genomic aberrations and VH mutation status. J Clin Oncol.
482 2004;22(19):3937-49.
483 20. Friedman DR, Weinberg JB, Barry WT, Goodman BK, Volkheimer AD, Bond
484 KM, et al. A genomic approach to improve prognosis and predict therapeutic
485 response in chronic lymphocytic leukemia. Clin Cancer Res. 2009;15(22):6947-55.
486 21. Fernandez V, Jares P, Salaverria I, Gine E, Bea S, Aymerich M, et al. Gene
487 expression profile and genomic changes in disease progression of early-stage
488 chronic lymphocytic leukemia. Haematologica. 2008;93(1):132-6.
489 22. Morabito F, Cutrona G, Gentile M, Matis S, Todoerti K, Colombo M, et al.
490 Definition of progression risk based on combinations of cellular and molecular
491 markers in patients with Binet stage A chronic lymphocytic leukaemia. Br J
492 Haematol. 2009;146(1):44-53.
493 23. Chuang HY, Rassenti L, Salcedo M, Licon K, Kohlmann A, Haferlach T, et al.
494 Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways
495 that associate with disease progression. Blood. 2012;120(13):2639-49.
496 24. Ferreira PG, Jares P, Rico D, Gómez-López G, Martínez-Trillos A, Villamor N, et
497 al. Transcriptome characterization by RNA sequencing identifies a major molecular
498 and clinical subdivision in chronic lymphocytic leukemia. Genome Res.
499 2014;24(2):212-26.
500 25. Herishanu Y, Pérez-Galán P, Liu D, Biancotto A, Pittaluga S, Vire B, et al. The
501 lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB
502 activation, and tumor proliferation in chronic lymphocytic leukemia. Blood.
503 2011;117(2):563-74.
504 26. Guarini A, Chiaretti S, Tavolaro S, Maggio R, Peragine N, Citarella F, et al. BCR
505 ligation induced by IgM stimulation results in gene expression and functional
506 changes only in IgV H unmutated chronic lymphocytic leukemia (CLL) cells. Blood.
507 2008;112(3):782-92.

27. Scott DW, Mottok A, Ennishi D, Wright GW, Farinha P, Ben-Neriah S, 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):2848-56.
28. Scott DW, Abrisqueta P, Wright GW, Slack GW, Mottok A, Villa D, et al. New Molecular Assay for the Proliferation Signature in Mantle Cell Lymphoma Applicable to Formalin-Fixed Paraffin-Embedded Biopsies. J Clin Oncol. 2017:JCO2016707901.
29. Yosifov DY, Idler I, Bhattacharya N, Reichenzeller M, Close V, Ezerina D, et al. Oxidative stress as candidate therapeutic target to overcome microenvironmental protection of CLL. Leukemia. 2020;34(1):115-27.
30. Byrd JC, Furman RR, Coutre SE, Flinn IW, Burger JA, Blum KA, 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, Robak T, Owen C, Ghia P, et al. Ibrutinib as Initial Therapy for Patients with Chronic Lymphocytic Leukemia. N Engl J Med. 2015;373(25):2425-37.
32. Dietrich S, Oleś M, Lu J, Sellner L, Anders S, Velten B, et al. Drug-perturbation-based stratification of blood cancer. J Clin Invest. 2018;128(1):427-45.
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):1105-17.
35. Schmid M, Kestler HA, Potapov S. On the validity of time-dependent AUC estimators. Brief Bioinform. 2015;16(1):153-68.
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):829-35.
37. Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. N Engl J Med. 2003;348(18):1764-75.
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, Gilling-Cutucache CE, Nordgren TM, Moragues M, 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, Bomben R, Dal Bo M, Russo S, 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, Ameur A, Hooper SD, Mayrhofer M, et al. Next generation RNA-sequencing in prognostic subsets of chronic lymphocytic leukemia. Am J Hematol. 2012;87(7):737-40.
42. Abruzzo LV, Barron LL, Anderson K, Newman RJ, Wierda WG, O'brien S, 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):546-55.
43. Fält S, Merup M, Tobin G, Thunberg U, Gahrton G, Rosenquist R, et al. Distinctive gene expression pattern in VH3-21 utilizing B-cell chronic lymphocytic leukemia. Blood. 2005;106(2):681-9.
44. Kienle D, Benner A, Kröber A, Winkler D, Mertens D, Bühler A, et al. Distinct gene expression patterns in chronic lymphocytic leukemia defined by usage of specific VH genes. Blood. 2006;107(5):2090-3.
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):1087-93.
46. Chen L, Huynh L, Apgar J, Tang L, Rassenti L, Weiss A, et al. ZAP-70 enhances IgM signaling independent of its kinase activity in chronic lymphocytic leukemia. Blood. 2008;111(5):2685-92.
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):2032-9.
48. Richardson SJ, Matthews C, Catherwood MA, Alexander HD, Carey BS, Farrugia J, 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):3584-92.
49. Calpe E, Codony C, Baptista MJ, Abrisqueta P, Carpio C, Purroy N, 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):4401-10.
50. Rassenti LZ, Jain S, Keating MJ, Wierda WG, Grever MR, Byrd JC, et al. Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. Blood. 2008;112(5):1923-30.
51. Rassenti LZ, Huynh L, Toy TL, Chen L, Keating MJ, Gribben JG, 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.

588

589

	Total Cohort (n=154)	CLL15 categories			*P* value
		Low-Risk Group (n=85)	Intermediate-Risk Group (n=31)	High-Risk Group (n=38)	
Male	88 (57.1%)	47 (55.3%)	16 (51.6%)	25 (65.8%)	0.435
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)	0.05
Binet stage					
A	116 (76.3%)	79 (92.9%)	24 (80%)	13 (35.1%)	<0.01
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	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)	0.18
Missing	38	3	7	28	
β2-microglobulin					<0.01
≤ 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	8	2	3	3	
CLL-IPI					<0.01
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	31	18	5	8	
CLL IPS-E					0.785
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	28	16	5	7	
IGHV mutational status					<0.01
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	9	7	-	2	
Missing	1	-	-	1	
ZAP-70					0.121
<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	36	16	6	14	
CD38					0.011
<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		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		14		11		2		1	
Complex karyotype (≥3 abnormalities)									0.877
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		84		38		16		30	
TP53 mut									0.009
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		54		23		7		24	
NOTCH1 mut									0.351
No		86 (82.7%)		49 (79%)		23 (92%)		14 (82.4%)	
Yes		18 (17.3%)		13 (21%)		2 (8%)		3 (17.6%)	
Missing		50		23		6		21	
SF3B1 mut									0.066
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		57		23		8		26	
MYD88 mut									0.395
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		61		23		8		30	
Median follow up - months		43.8 m		43.6 m		43.8 m		45.9 m	0.61

Table 1: Patient and disease characteristics of the training cohort

Abbreviations: CLL-IPI: International Prognostic Index for Chronic Lymphocytic Leukemia; CLL IPS-E: International Prognostic Score for Asymptomatic Early-stage Chronic Lymphocytic Leukemia; *P* values are for comparisons across the 3 risk groups determined by the CLL15 score.

Figure Legends:

Figure 1. The gene expression-based model to predict Time to First Treatment in patients with chronic lymphocytic leukemia. Heatmap of the CLL15 assay with the 15 informative genes shown as rows and the 154 patient samples as columns. The three patient groups identified by the assay are shown below the heatmap together with the mutational status of the *IGHV* genes.

Figure 2. (A) Log relative hazard according to the CLL15 score. (B) Kaplan-Meier curves of the Time to First Treatment of the three patient groups identified by the CLL15 assay. (C) Kaplan-Meier curves of the Time to First Treatment of the three 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 Time to First Treatment according to CLL15 assay and *IGHV* mutational status.

Figure 3. Univariate and multivariate analysis for Time to First Treatment according to prognostic factors in CLL.

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

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

Figure 1

Figure

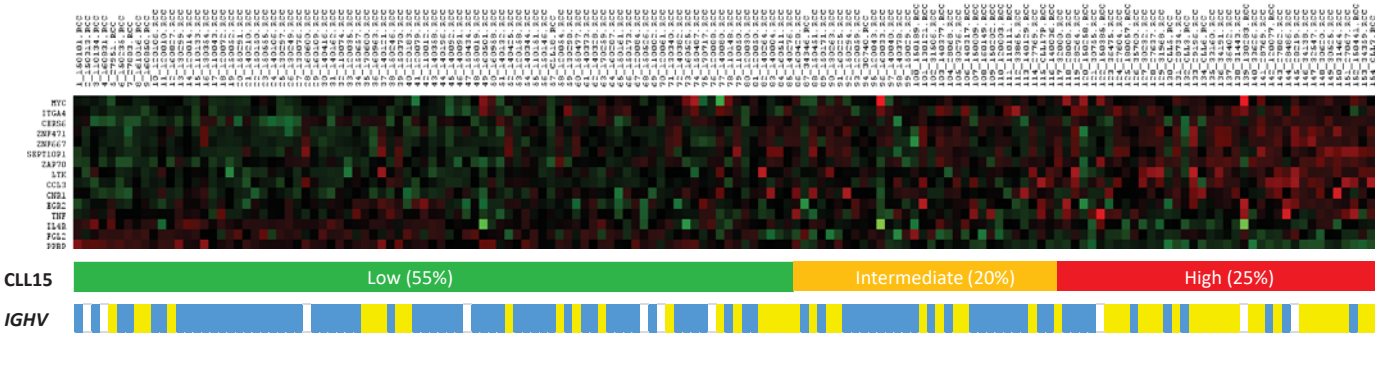
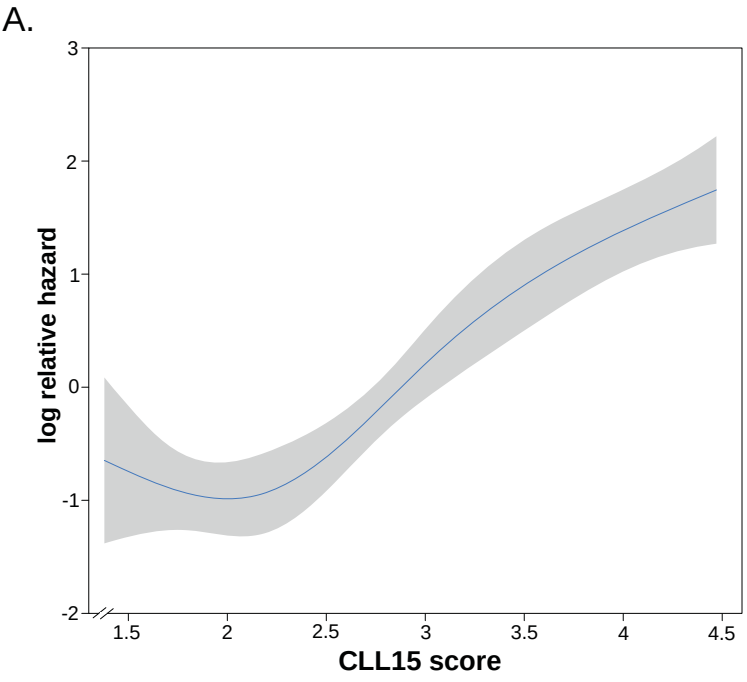
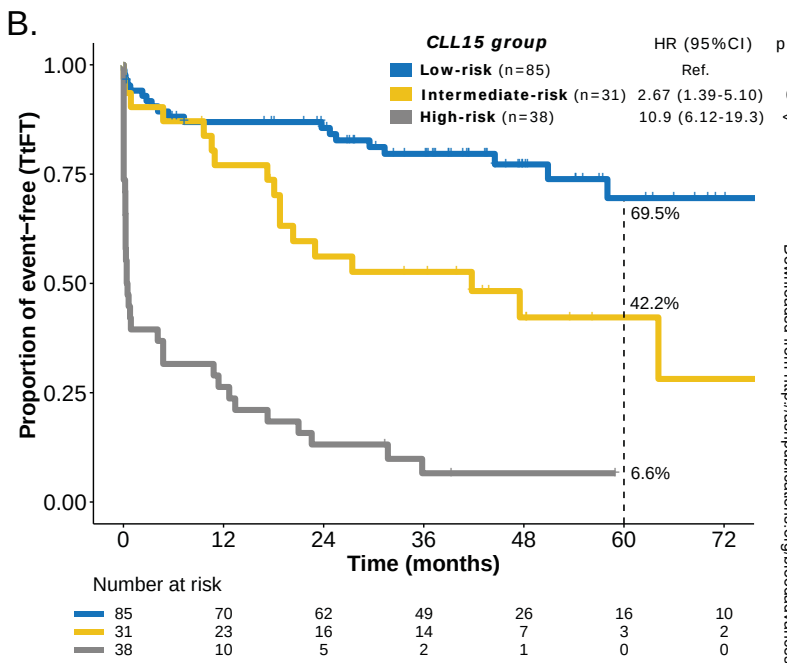
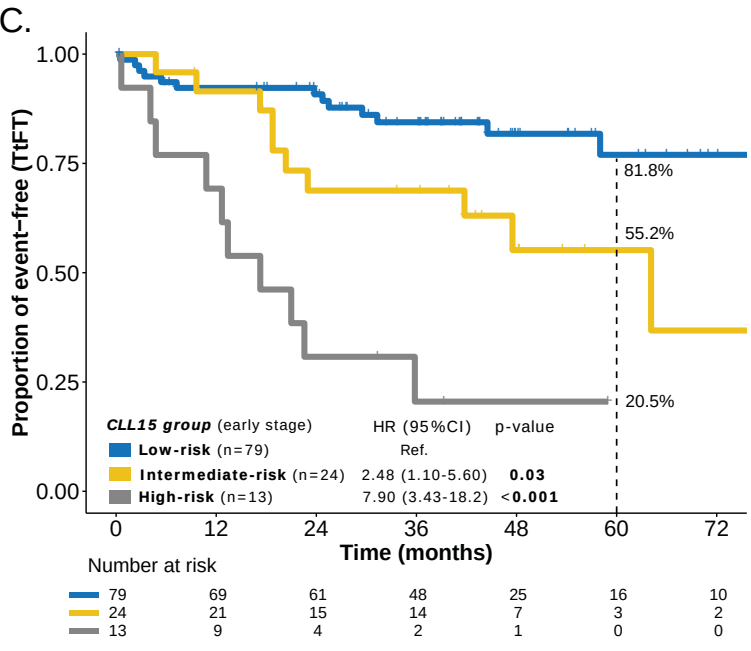


Figure 2







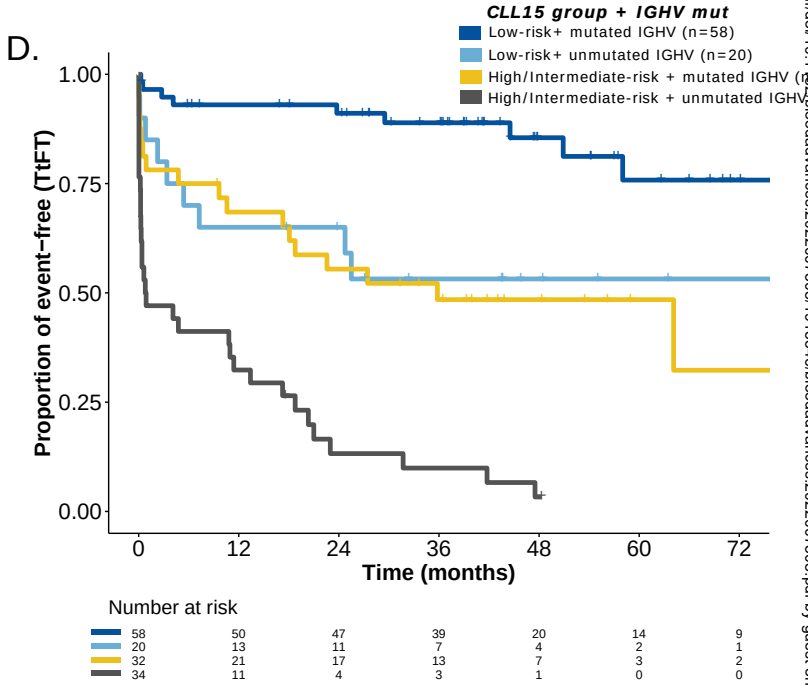


Figure 3

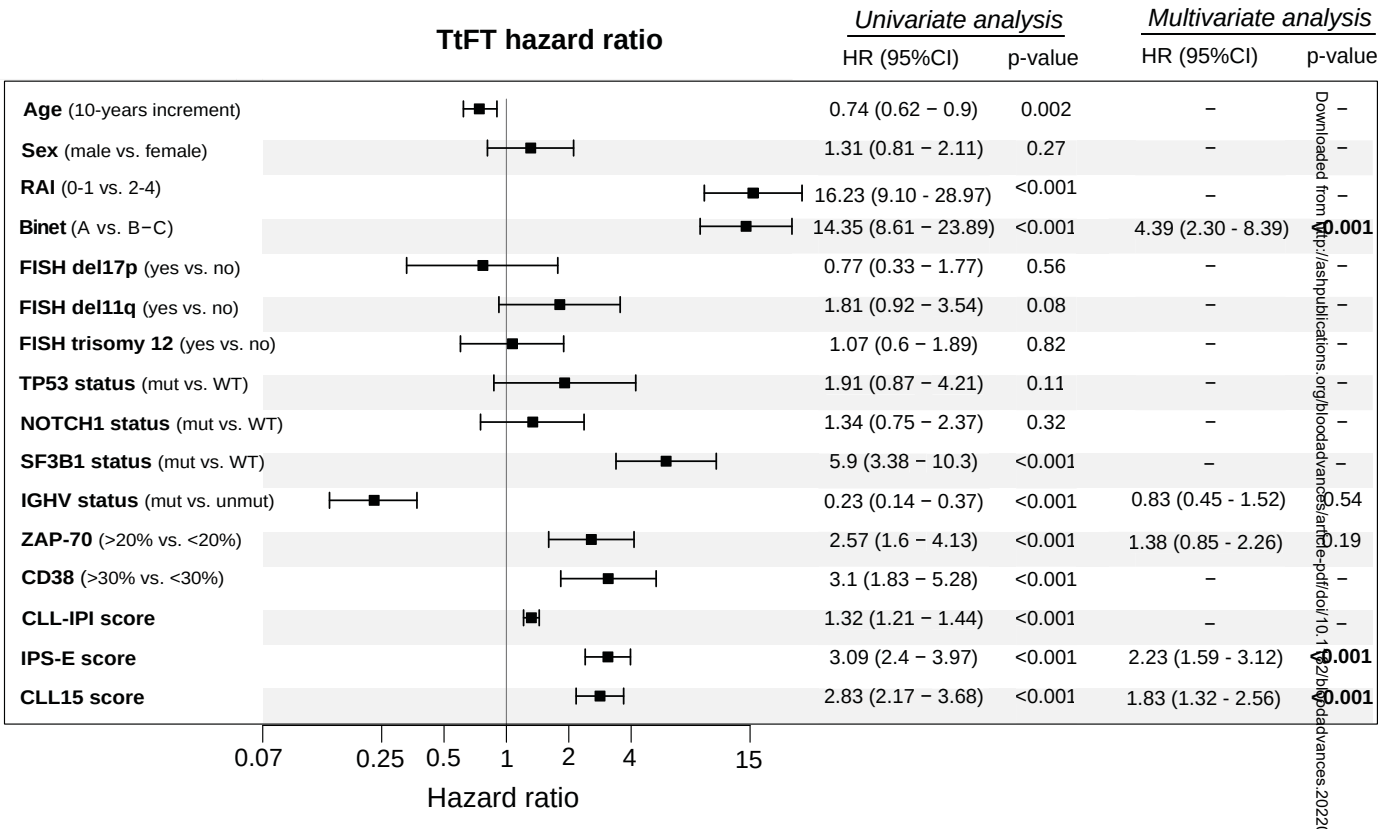
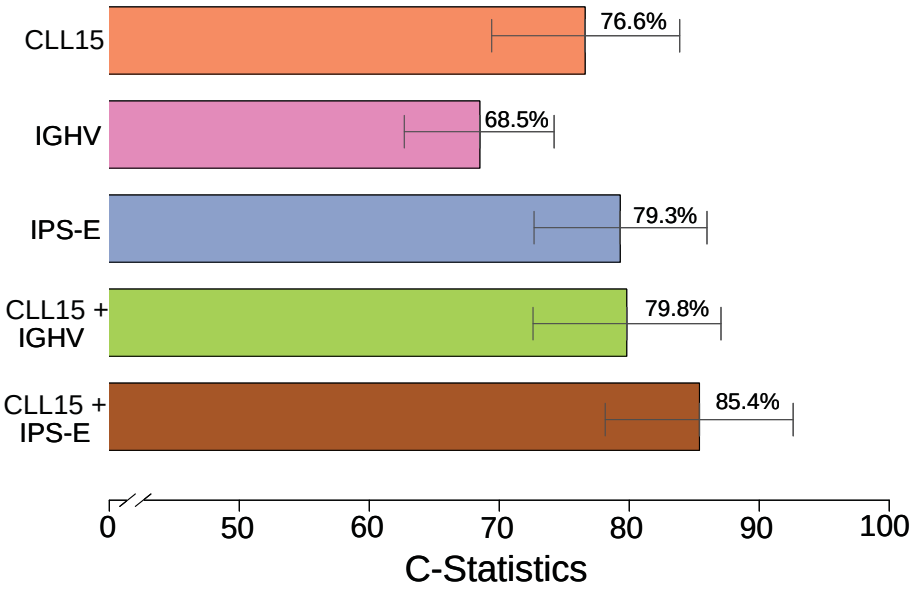


Figure 4

A.



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 5