**The Engineered MIPI (e-MIPI), a Candidate Data-Mining Based Mantle Cell Lymphoma Prognostic Index Developed from the Dataset of the Fondazione Italiana Linfomi (FIL) MCL0208 Phase III Trial**

# The Engineered MIPI (e-MIPI), a Candidate Data-Mining Based Mantle Cell Lymphoma Prognostic Index Developed from the Dataset of the Fondazione Italiana Linfomi (FIL) MCL0208 Phase III Trial

*Zaccaria GM[1], Ferrero S[1], Passera R[2], Evangelista A[3], Loschirico M[1], Dogliotti I[1], Ghislieri M[4], Genuardi E[1], Bomben R[5], Gattei V[5], Ciccone G[3], Tani M[6], Gaidano G[7], Volpetti S[8], Cabras MG[9], Di Renzo N[10], Merli F[11], Vallisa D[12], Spina M[13], Pascarella A[14], Latte G[15], Patti C[16], Pozzato G[17], Fabbri A[18], Cortelazzo S[19], Ladetto M[20]*

*[1]Department of Molecular Biotechnology and Health Sciences, Università di Torino, Turin, Italy; [2]Division of Nuclear Medicine, Università di Torino, Turin, Italy; [3]Clinical Epidemiology, Città della Salute e della Scienza and CPO Piemonte, Torino; [4]Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy; [5]Clinical and Experimental Onco-Hematology Unit, IRCCS, Aviano (PN); [6]U.O.C di Ematologia Ospedale S. Maria delle Croci Ravenna, Italy; [7]Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy; [8]Clinica Ematologica, Azienda Sanitaria Universitaria Integrata di Udine, Udine, Italy; [9]Ematologia e CTMO, Ospedale Businco Cagliari, Italy; [10]Divisione di Ematologia, Polo Oncologico Giovanni Paolo II, Presidio Ospedaliero Vito Fazzi, Lecce, Italy. [11]Clinics and Public Health, "Arcispedale S.Maria Nuova", University of Modena e Reggio Emilia, Hematology Division, Department of Diagnostic Medicine, Reggio Emilia, Italy. [12]Unità Operativa di Ematologia, Dip. Di Oncologia ed Ematologia, Ospedale Guglielmo da Saliceto, Piacenza, Italy. [13]Division of Medical Oncology A, National Cancer Institute, Aviano, Italy. [14]Division of Hematology, Ospedale SS. Giovanni e Paolo, Venezia Mestre, Italy. [15]Unità di Ematologia e Trapianto di Midollo Osseo, Ospedale "San Francesco" Nuoro, Italy. [16]Divisione di Ematologia I Azienda Ospedali riuniti Villa Sofia- Cervello. [17]Department of Medical and Surgical Sciences, University of Trieste, Trieste, Italy. Trieste Italy; [18]UOC ematologia, Azienda Ospedaliera Universitaria Senese, Siena, Italy. [19]Oncology Unit, Humanitas/Gavazzeni Clinic, Bergamo, Italy. [20]Division of Hematology, Azienda Ospedaliera SS Antonio e Biagio e Cesare Arrigo, Alessandria, Italy.*

## BACKGROUND

The amount of clinical and biological data stored within clinical trials is growing exponentially. The highly translational FIL-MCL0208 trial has been used to test a data-ware house (DW) to improve data quality and to discover putative associations [1]. In this study we developed an engineered prognostic model, focusing on easily accessible clinical variables. For this purpose, we exploited hierarchical clustering with the aim of seeking hidden patterns of interest in large datasets. Hence, these tools allowed to develop a novel prognostic model: the engineered MIPI index (e-MIPI).

## OBJECTIVES

Herein we present the first results, on baseline clinical characteristics:

- clustering analysis and definition of a signature of predictive variables.
- construction of the e-MIPI to detect patients' risk of relapse.
- comparison with known prognostic indexes for MCL.
- validation of the signature on independent subset of patients.

## METHODS

**A**

| FEATURES | Sex | Age | BMI | ECOGps | Sym | Bulky | LDH | PLTs | WBC | Hb | ALT | AST | Creatinine | Prot | Alb | Bili | GGT | ALP | B2M | IgG | IgA | IgM | flow BM | flow PB | Ki67 | SOX11 | Hist | BM Inf | IgH Omo | DN | EN | PET |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DATA TYPE | Morphometric | | | Clinical | | | Laboratory | | | | | | | | | | | | | | | | Pathology | | | | | | | Imaging | | |
| UNIT | - | y | Kg/m$^2$ | - | - | cm | mg /dL | 10^9 /L | 10^9 /L | g/dL | IU/L | IU/L | mg/dL | g/dL | g/dL | mg/dL | IU/L | IU/L | mg/dL | g/dL | g/dL | g/dL | % | % | % | - | - | - | % | - | - | - |
| MISSING VALUES | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 5 | 5 | - | 18 | 36 | 19 | 20 | 26 | 51 | 78 | 78 | 77 | 48 | 15 | 29 | 141 | 0 | 0 | 89 | 0 | 0 | 67 |
| NORM**/LOW/ CUT-OFF | - | µ | [18,5 - 24,9] | 0 | A vs B | 5 | 1 | [150 - 450] | [4 - 11] | [11.7 - 18] | [7 - 56] | [10 - 40] | * | [6 - 8.3] | [3.4 - 5.4] | [0.2 - 1.2] | [8 - 65] | [44 - 147] | - | [0.7 - 1.6] | [0.07 - 0.40] | [0.04 - 0.23] | µ | µ | 30 | - | N vs B | Nl vs l | µ | Nl vs l | Nl vs l | Nl vs l |
| Legend | F vs M | L vs H | L vs H | N vs Abn | N vs Abn | N vs Abn | L vs H | N vs Abn | N vs Abn | N vs Abn | N vs Abn | N vs Abn | N vs Abn | N vs Abn | N vs Abn | N vs Abn | N vs Abn | N vs Abn | L vs H | N vs Abn | N vs Abn | N vs Abn | L vs H | L vs H | L vs H | N vs Abn | N vs Abn | N vs Abn | L vs H | N vs Abn | N vs Abn | N vs Abn |

Baseline features extracted from FIL-MCL0208 clinical trial. Features are grouped in morphometric, clinical, laboratory, pathology and Imaging. Each feature is described in UNIT of magnitude, number of MISSING VALUES (MV), NORMALITY RANGE or CUT-OFF levels to dichotomize continues variables.

**A.** Clinical Baseline data of 300 patients enrolled in FIL-MCL0208 clinical trial

**B.** Definition of the training-set without MVs and application of the Unsupervised data-analysis

**C.** Application of clinical outcomes and construction of a novel candidate p-index (eMIPI)

**D.** Definition of a surrogate "signature" of variables

**E.** Validation of the "signature" on a series derived from the FIL-MCL0208 phase III clinical study

**B** **TRAINING-SET DEFINITION.** starting from 300 enrolled subjects, the main goal was to achieve a training-set without MVs. To do that, we assumed not eligible both the variables with high amount of MVs (>13%) and the subjects with MVs among 5% and 13%. At the end of the process we reached a subset of 185 subs X 24 fts.

- **32 dichotomized variables for 300 subjects**
- **8 variables not eligible because >40 (13%) of MVs**
  - B2, IgM, IgG, IgA
  - FlowBM, IgH_Omo, SOX11
  - PET inv
- **5 variables with <15 of MVs (5%) imputed by median of the patients' observations**
  - PLTs, ALT, AST, Creatinine
  - BMInf
- **For the variables with MVs among 5% and 13% we preferred to exclude patients from training-set**
  - 38 MVs of Albumin
  - 29 MVs of Ki67
  - 15 MVs of FlowPB
- **32-8=24 variables for the TRAINING SET on 185 subs free from MVs**
- **4440 total observations**

**C**

**Prognostic Indexes**
- eMIPI
- MIPI Standard
- MIPI Biologic

C index

**Clinical Outcomes**
- PFS
- TTP
- OS

We assumed each cluster to be a different risk of relapse (Low/Int/High) correlating each sub-cohort with PFS, TTP and OS outcomes. Hence, we compared the eMIPI results with the known MCL indexes [2] assessing the C-index as a goodness of prediction.

**D** From the starting 24 features we firstly applied an Exploratory Data Analysis (EDA). Thus, to detect a "signature" of variables to define a surrogate of PFS, TTP and OS a Recursive Feature Elimination was deployed (RFE).

Start

**24 STARTING FEATURES**
- Sex
- Age
- BMI
- ECOGps
- Sym
- Bulky
- LDH
- PLTs
- WBC
- Hb
- AST
- ALT
- Creatinine
- Protein
- Albumin
- Bili
- GGT
- ALP
- flowPB
- Ki67
- BMinf
- SOX11
- Hist
- DN
- EN

**NOT ELIGIBLE FEATURES**
- Sex
- AST
- Protein
- ALP
- Hist
- DN
- EN
- Creatinine
- BMI
- Bili
- Bulky
- Age
- GGT
- ALT
- ECOGps

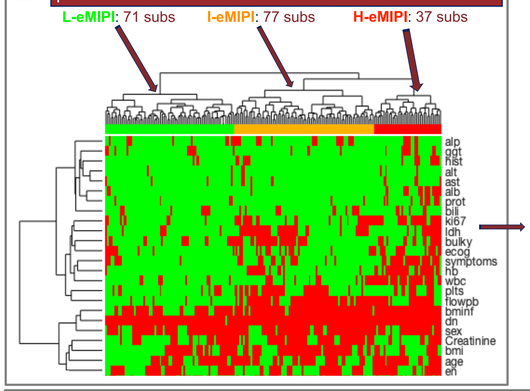
**FINAL «SIGNATURE»**
- flowPB
- Hb
- WBC
- Alb
- BMinf
- LDH
- Ki67
- PLTs
- symptoms

**E** 115 subjects for the validation set → 75/115 subjects with at least a MV → MVs Imputation with K-nn

The validation set involved the missing 115 subjects. To optimize results in terms of clinical outcomes, a K-nearest neighbor [3] has been performed to impute MVs.

## RESULTS

**B** Hierarchical Clustering to find out sub-cohorts of patients who share the same characteristics – 24 ft

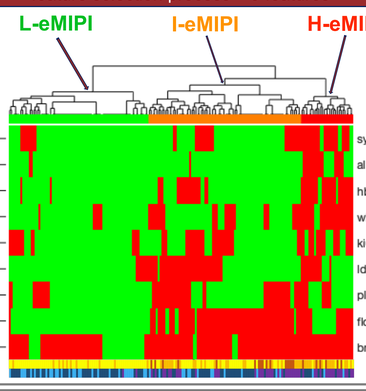



Normal/Low

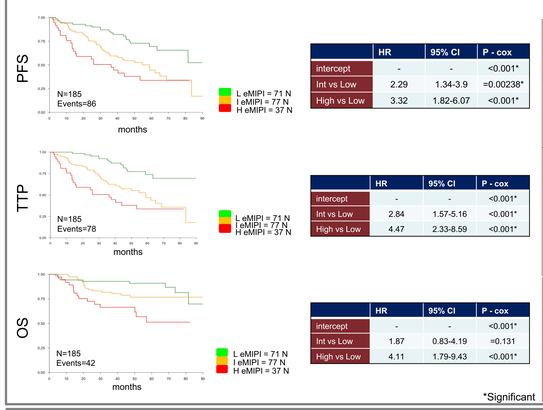
Abnormal/High

| 83% | signature | | |
|---|---|---|---|
| | L | I | H |
| all variables L | 62 | 9 | 0 |
| all variables I | 11 | 65 | 1 |
| all variables H | 2 | 8 | 27 |

**D** Hierarchical Clustering on the "signature" derived from the feature selection process – 9 features




**C** Correlation of the e-MIPI with PFS, TTP, OS



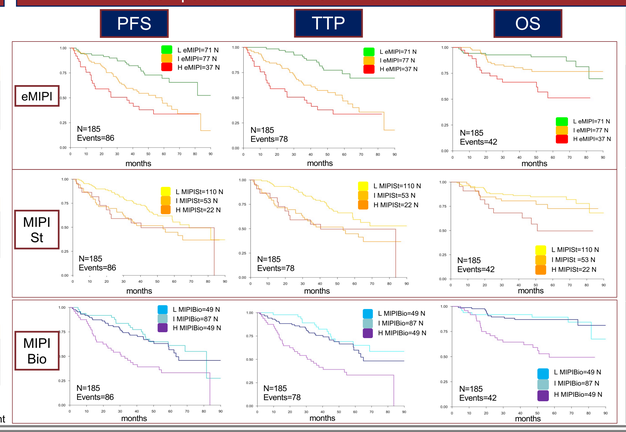

| | HR | 95% CI | P - cox |
|---|---|---|---|
| intercept | - | - | <0.001* |
| Int vs Low | 2.29 | 1.34-3.9 | =0.00238* |
| High vs Low | 3.32 | 1.82-6.07 | <0.001* |

| | HR | 95% CI | P - cox |
|---|---|---|---|
| intercept | - | - | <0.001* |
| Int vs Low | 2.84 | 1.57-5.16 | <0.001* |
| High vs Low | 4.47 | 2.33-8.59 | <0.001* |

| | HR | 95% CI | P - cox |
|---|---|---|---|
| intercept | - | - | <0.001* |
| Int vs Low | 1.87 | 0.83-4.19 | =0.131 |
| High vs Low | 4.11 | 1.79-9.43 | <0.001* |

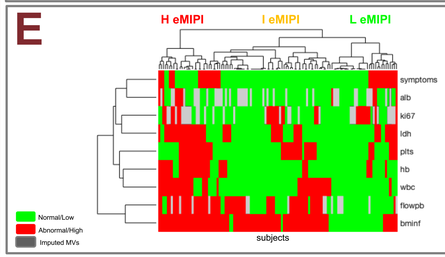
*Significant

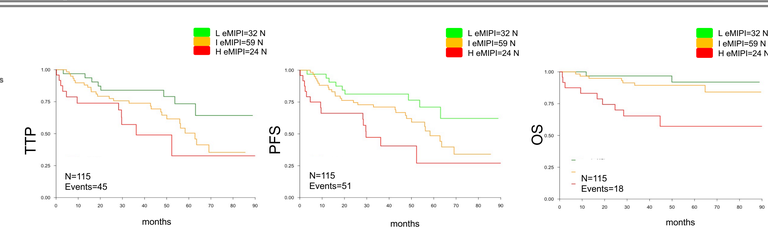
**Comparison between eMIPI and both MIPISt and MIPIBio**




| C-index | PFS | TTP | OS |
|---|---|---|---|
| eMIPI | 0.64 | 0.67 | 0.65 |
| MIPI St | 0.58 | 0.61 | 0.60 |
| MIPI Bio | 0.61 | 0.64 | 0.67 |

**E**




## CONCLUSIONS

e-MIPI is a new first prognostic index derived from hierarchical clustering. Our results indicate that this approach might allow to model engineered prognostic indexes based on comprehensive analysis of large datasets. Even if promising, it needs validation through its application to independent series of MCL patients. Additional efforts aiming at integrating biological variables in the model are ongoing.

*No Disclosures for this study*

## REFERENCES

1. Zaccaria et al. *Blood 130 (Suppl 1) , 2017*
2. Hoster et al., *Blood, 2008*
3. Han et al. *Data-Mining, 2012*

*gizaccar@unito.it*