LEUKEMIA2022

Rome, Hotel NH Collection - Vittorio Veneto

May 5-6, 2022

AIL President: G. Toro
Coordinators: A.M. Carella, S. Amadori

AIL
ASSOCIAZIONE ITALIANA CONTRO LEUCEMIE LINFOMI E MIELOMA

UNDER THE AUSPICES OF:

fondazione GIMEMA onlus
per la promozione e lo sviluppo della ricerca scientifica sulle malattie ematologiche. FRANCO MANDELLI

GITMO

SIES

SIE - Società Italiana di Ematologia

Towards a new prognostication of MDS

A gene-bank statistical learning approach for personalized medicine

Dott. Matteo Bersanelli, PhD.
Humanitas University,
Via Rita Levi Montalcini,
20090 Pieve Emanuele, Milan, Italy

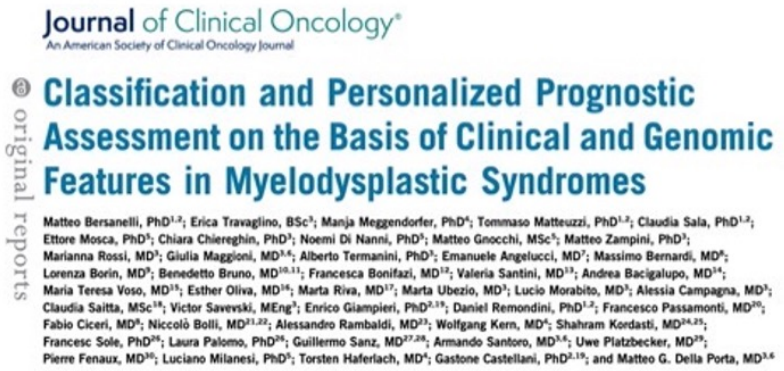
The EUROMDS cohort

This study included **2361 patients** with MDS according to 2016 WHO classification of myeloid neoplasms:

- **learning cohort:** a retrospective cohort of **2043 patients** collected in the context of EuroMDS Consortium (including 21 hematological centers from Italy, Germany, Spain and France)
- **validation cohort:** an independent prospective cohort of **318 patients** diagnosed at Humanitas Research Hospital, Milan, Italy.

Main data points by category:

- General (Age, Sex)
- Clinical (BMB, Hemoglobin, Platelets, Neuthophils, …)
- Cytogenetics alterations
- Genomics (NGS panel of 47 genes)
- Outcome data (OS, LFS)


Journal of Clinical Oncology
An American Society of Clinical Oncology Journal

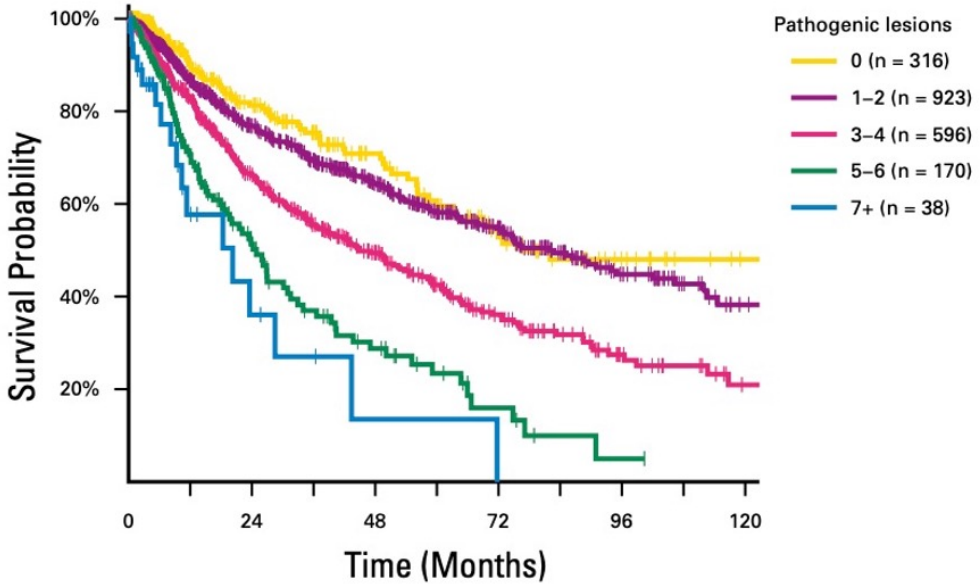
original reports

Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes

Matteo Bersanelli, PhD[1,2]; Erica Travaglino, BSc[3]; Manja Meggendorfer, PhD[4]; Tommaso Matteuzzi, PhD[1,2]; Claudia Sala, PhD[1,2]; Ettore Mosca, PhD[3]; Chiara Chiereghin, PhD[3]; Noemi Di Nanni, PhD[3]; Matteo Gnocchi, MSc[3]; Matteo Zampini, PhD[3]; Marianna Rossi, MD[3]; Giulia Maggioni, MD[3,4]; Alberto Termanini, PhD[3]; Emanuele Angelucci, MD[7]; Massimo Bernardi, MD[8]; Lorenza Borin, MD[9]; Benedetto Bruno, MD[10,11]; Francesca Bonifazi, MD[12]; Valeria Santini, MD[13]; Andrea Bacigalupo, MD[14]; Maria Teresa Voso, MD[15]; Esther Oliva, MD[16]; Marta Riva, MD[17]; Marta Ubezio, MD[3]; Lucio Morabito, MD[3]; Alessia Campagna, MD[3]; Claudia Saitta, MSc[18]; Victor Savevski, MEng[3]; Enrico Giampieri, PhD[2,19]; Daniel Remondini, PhD[1,2]; Francesco Passamonti, MD[20]; Fabio Ciceri, MD[5]; Niccolò Bolli, MD[21,22]; Alessandro Rambaldi, MD[23]; Wolfgang Kern, MD[4]; Shahram Kordasti, MD[24,25]; Francesc Sole, PhD[26]; Laura Palomo, PhD[26]; Guillermo Sanz, MD[27,28]; Armando Santoro, MD[3,6]; Uwe Platzbecker, MD[29]; Pierre Fenaux, MD[30]; Luciano Milanesi, PhD[3]; Torsten Haferlach, MD[4]; Gastone Castellani, PhD[2,19]; and Matteo G. Della Porta, MD[3,6]

The EUROMDS cohort

Survival curves stratify by number of pathogenic lesions.

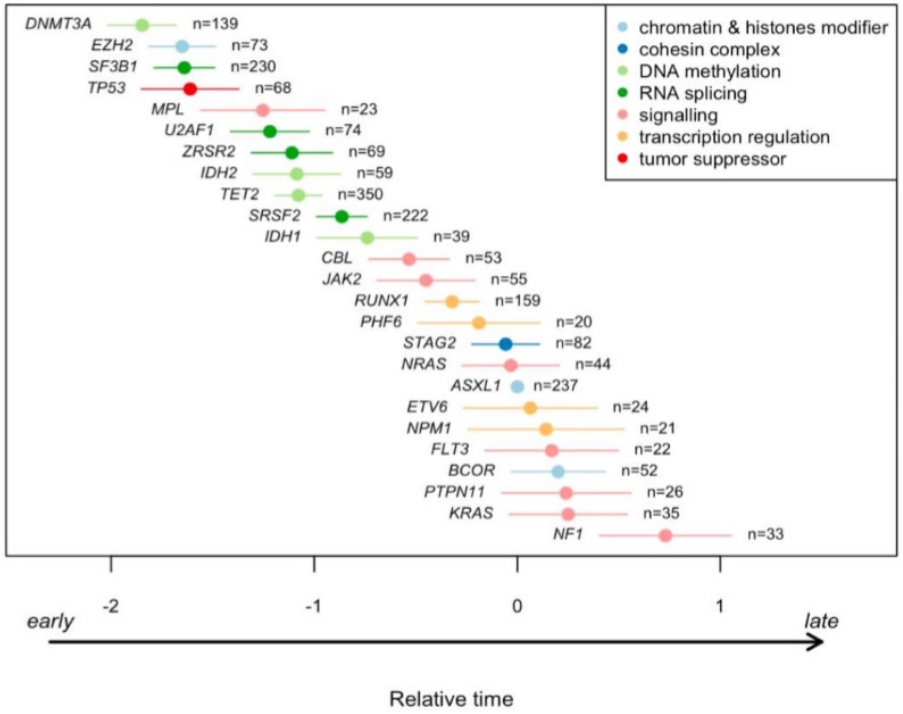


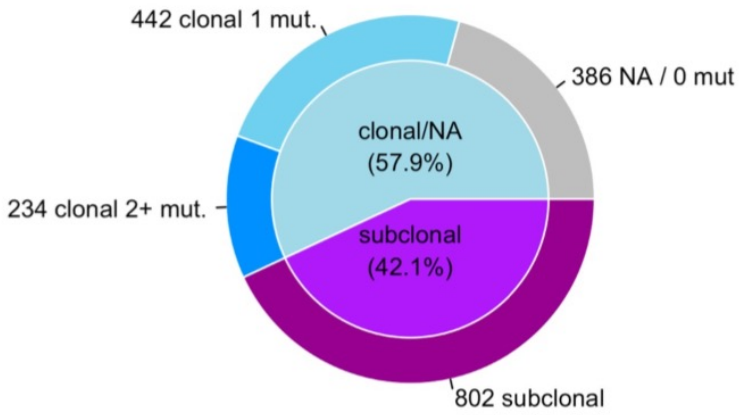
Number at risk

0	192	58	30	18	11	7
1–2	508	157	116	85	37	20
3–4	356	107	70	39	16	8
5–6	125	25	14	5	1	0
7+	33	4	1	0	0	0

Bradley-Terry (BT) model for detection of clonal / subclonal mutations

- **Determination of relative order of mutation acquisition**
- Comparisons are made for **each pair of mutations co-occurring in the same sample**
- for each patient are considered the **proportions of cells carrying each mutation**, the **variant allele fractions** corrected for any **copy number change** at the site of the variant.

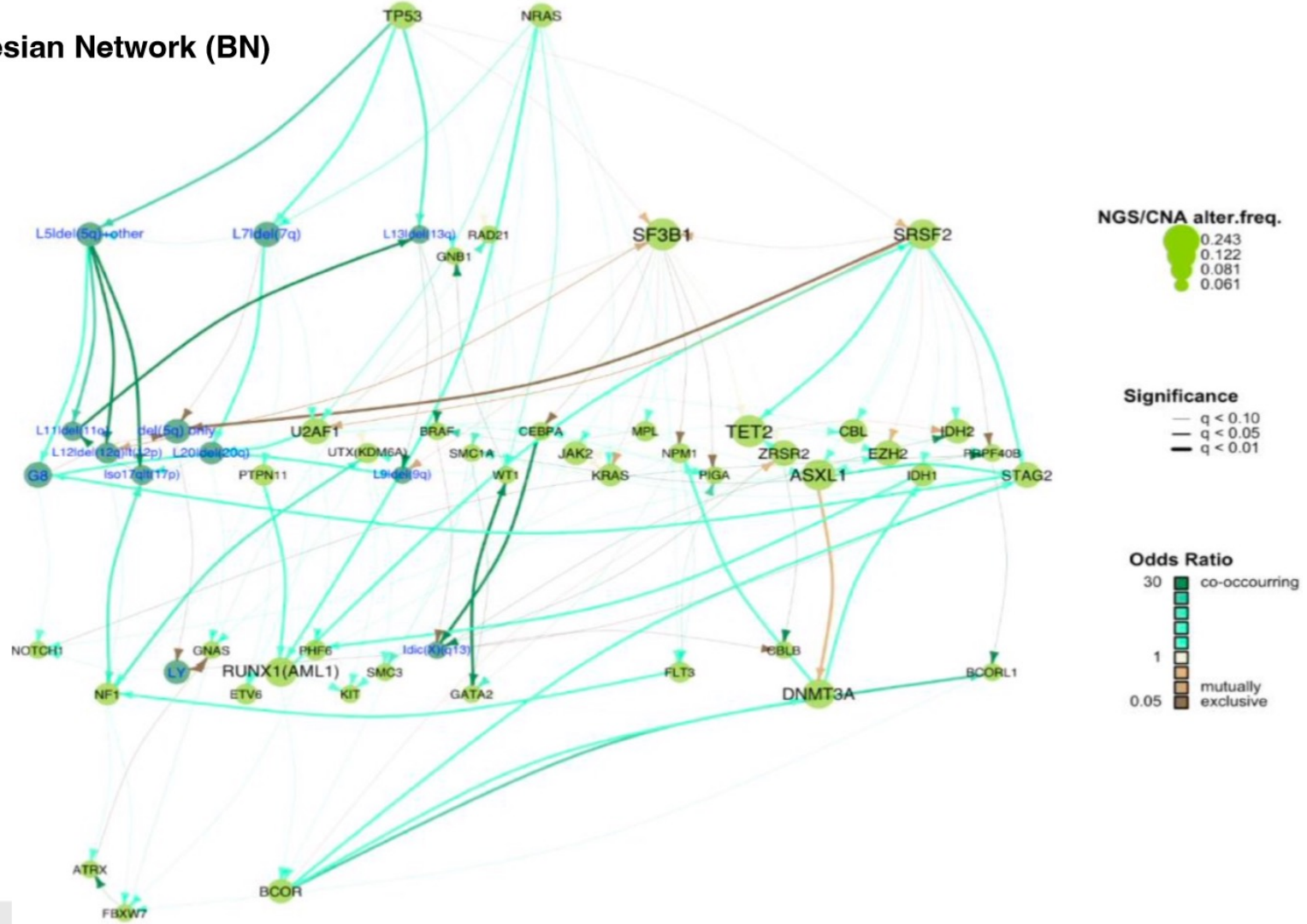




Both clonal and sub-clonal mutations have a significant impact on patient outcome. The different impact of clonal vs subclonal mutations needs to be further investigated.

Mutation causality: MDS Bayesian Network (BN)

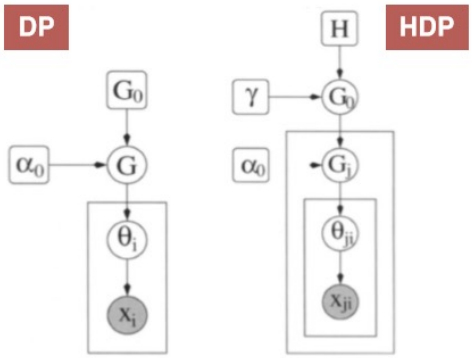
- BN are able to infer the statistical causal link that exists between mutations occurring in patients affected by the same disease.
- It must be interpreted as follows: parent mutations tend to be important in causing or not (statistically) the children mutations.
- Parent mutations tend to be on the top of the network layout, while children mutations tend to be on the bottom, even if with some exceptions.



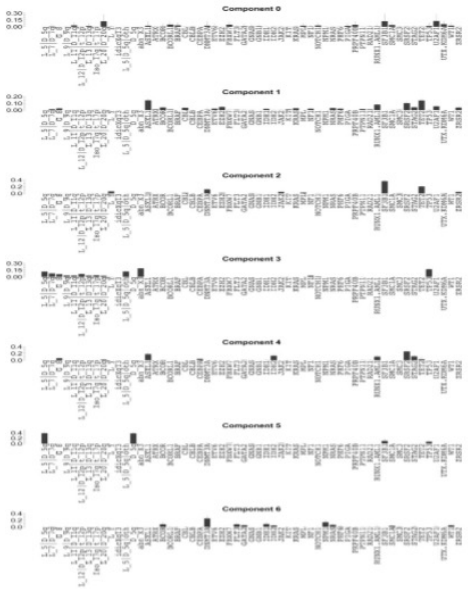
Genomic classification of MDS using Hierarchical Dirichelet Processes

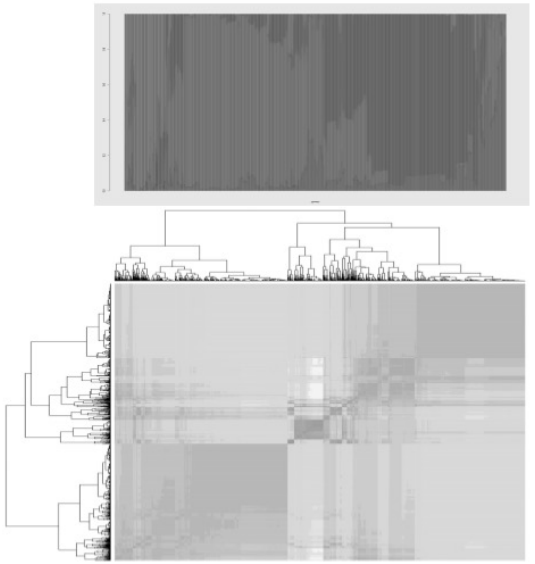
- **Unsupervised non-parametric Bayesian method.**
- Objectives: identify disease-specific molecular subtypes, patient stratification
- Criticalities: heterogeneous data, long tail distribution, binary data (0,1), low signal (2-3 median mutations per patient)

a. HDP modeling of the dataset G. → **b. Extraction of molecular components** → **c. Patients Stratification**



- $\theta \sim DP(Dirichlet(\alpha), \alpha_0)$
- $X \mid \theta, N \sim Multinomial(\theta, N_j)$
- N_j is the number of mutations in sample j. As prior we assume a Dirichlet distribution with parameter $\alpha = (1/n,...,1/n)$.



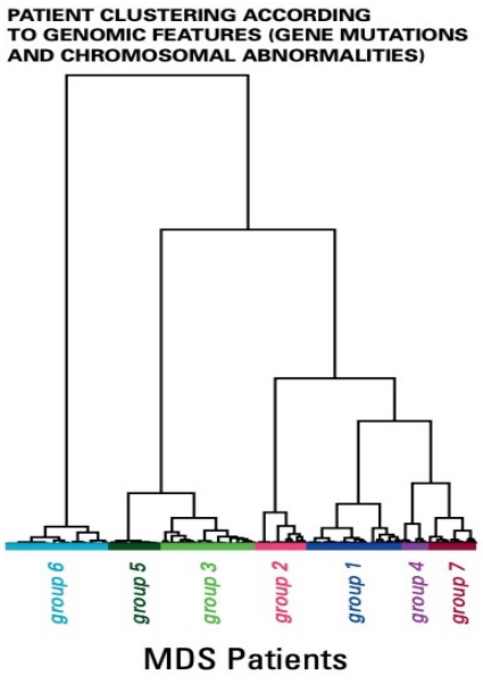


Genomic classification of MDS using Hierarchical Dirichelet Processes

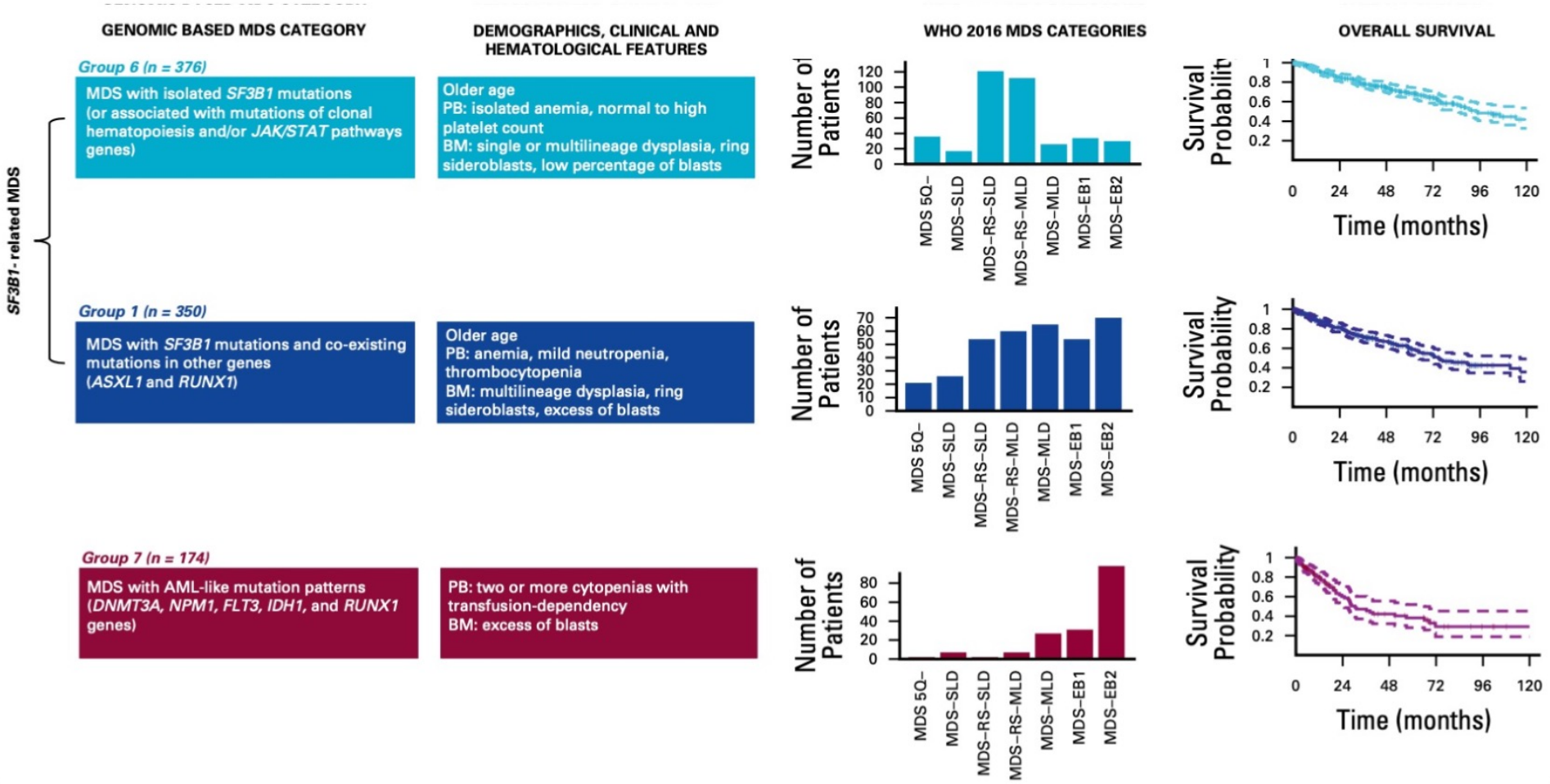
- **Eight genomic groups** were identified using Hierarchical Dirichelet Processing (HDP) for patient stratification, out of six components retrieved in the latent space. HDP is an unsupervised stratification method capable to handle far-from-normal distributed datsets

MDS genomic based group

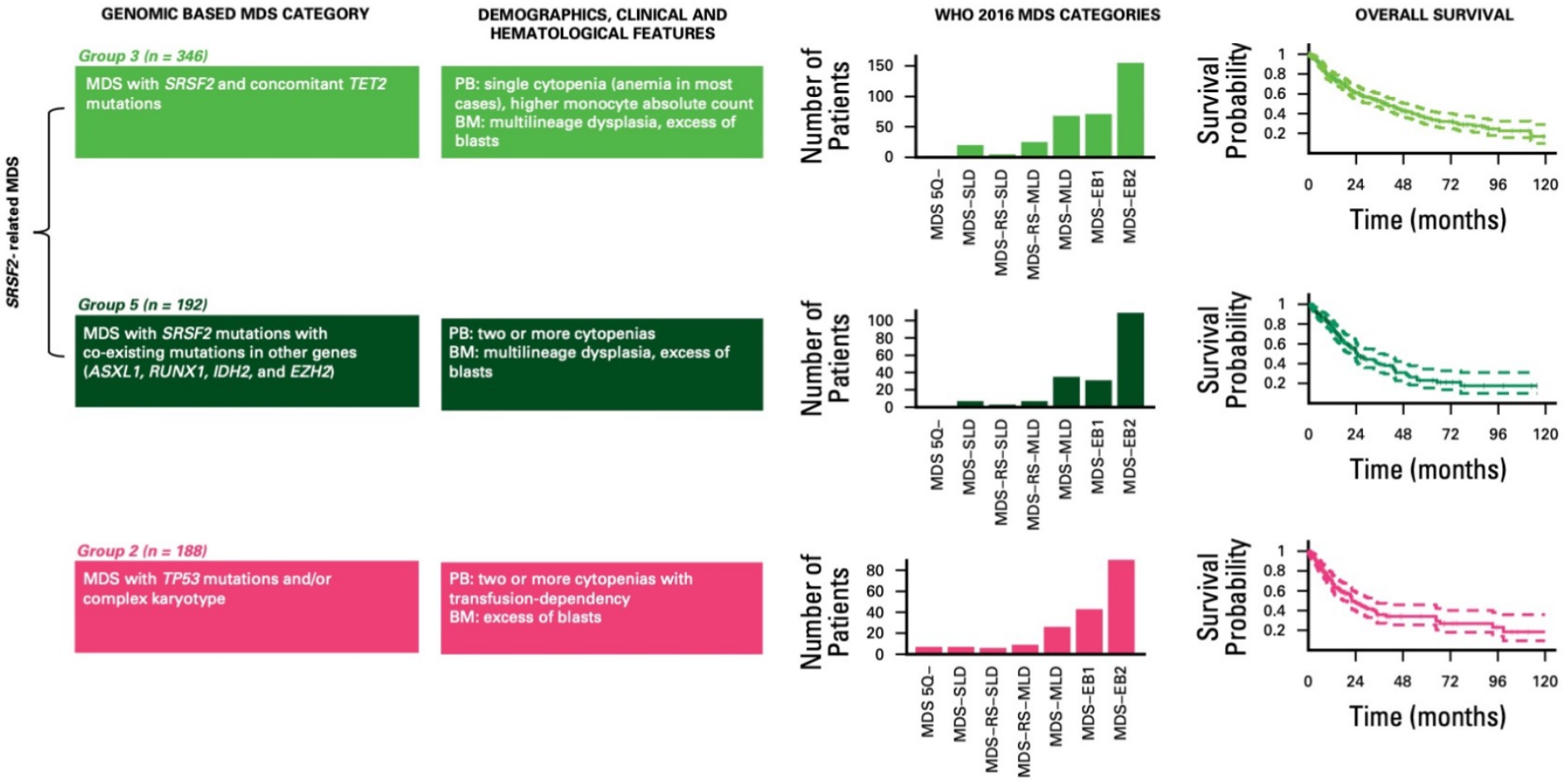
- MDS with isolated SF3B1 mutations (or associated with mutations of clonal hematopoiesis and/or JAK/STAT pathways genes) [Group 6]
- MDS with SF3B1 with co-existing mutations [Group 1]
- MDS with SRSF2 and concomitant TET2 mutations [Group 3]
- MDS with SRSF2 mutations with co-existing mutations [Group 5]
- MDS with U2AF1 mutations associated with deletion of chromosome 20q, isolated del(7q) or chromosome 7 monosomy [Group 4]
- MDS with TP53 mutations and/or complex karyotype [Group 2]
- MDS with AML-like mutations [Group 7]
- MDS without specific genomic profiles [Group 0]



Genomic classification of MDS using Hierarchical Dirichelet Processes

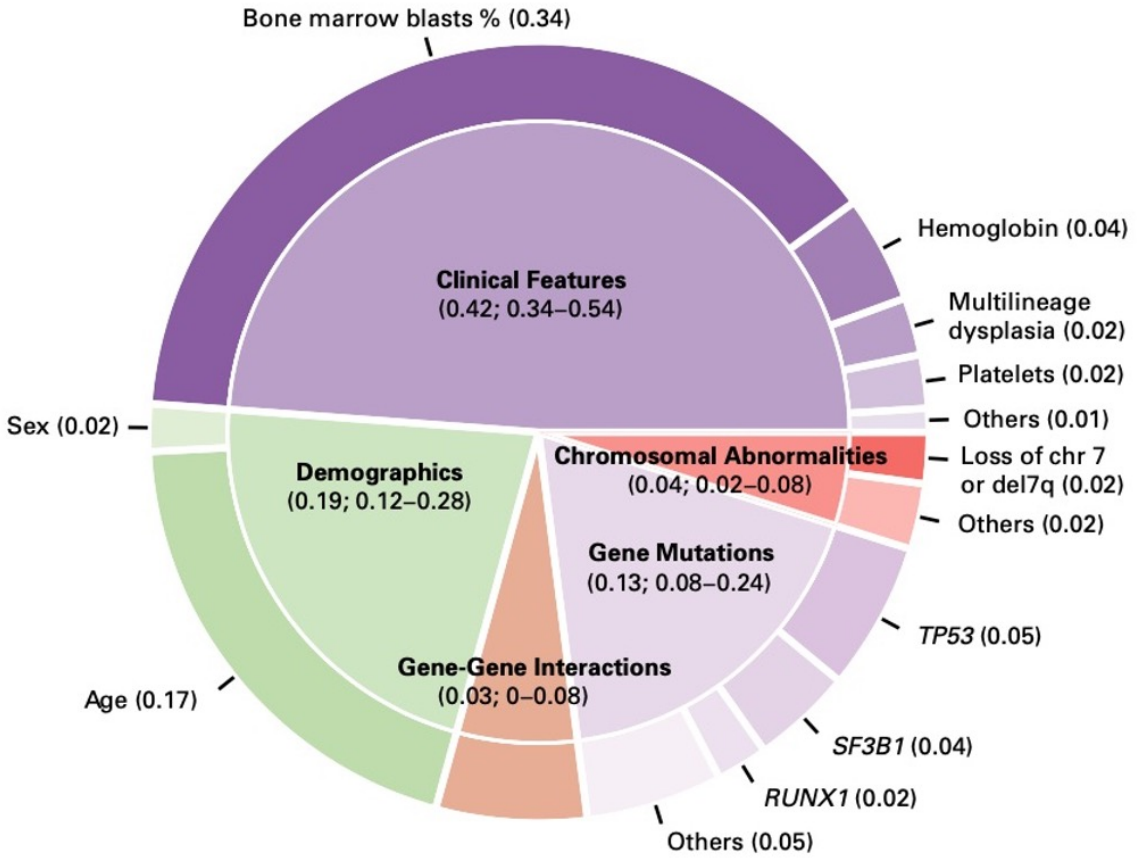


Genomic classification of MDS using Hierarchical Dirichelet Processes



Towards a new prognostication of MDS

- **Random effects COX proportional hazards** model was used for modelling overall survival with the study variables treated as random effects.
- The weight of genomic mutations over prognostic outcome is significantly higher than the weight of chromosomal abnormalities.
- The combined weight of gene mutations, gene-gene interactions and cytogenetic data covers approximately 1/3 of the total.



Towards a new prognostication of MDS

- In terms of concordance score, the model significantly improved the state of the art, at the same time giving the possibility to estimate a personalised outcome.

	Training (66% of EuroMDS Patients)		Test (33% of EuroMDS Patients)	
Statistical Model and Variable Selection	**Concordance**	**SD**	**Concordance**	**SD**
Cytogenetics IPSS-R risk groups	0.576	0.012	0.567	0.016
Age-adjusted IPSS-R risk groups	0.620	0.015	0.659	0.019
Dirichlet processes	0.649	0.014	0.629	0.020
CoxRFX_Clinical + demographics + Dirichelet processes	0.729	0.015	0.713	0.021
CoxRFX_Clinical + demographics + genomics	0.742	0.015	0.709	0.021

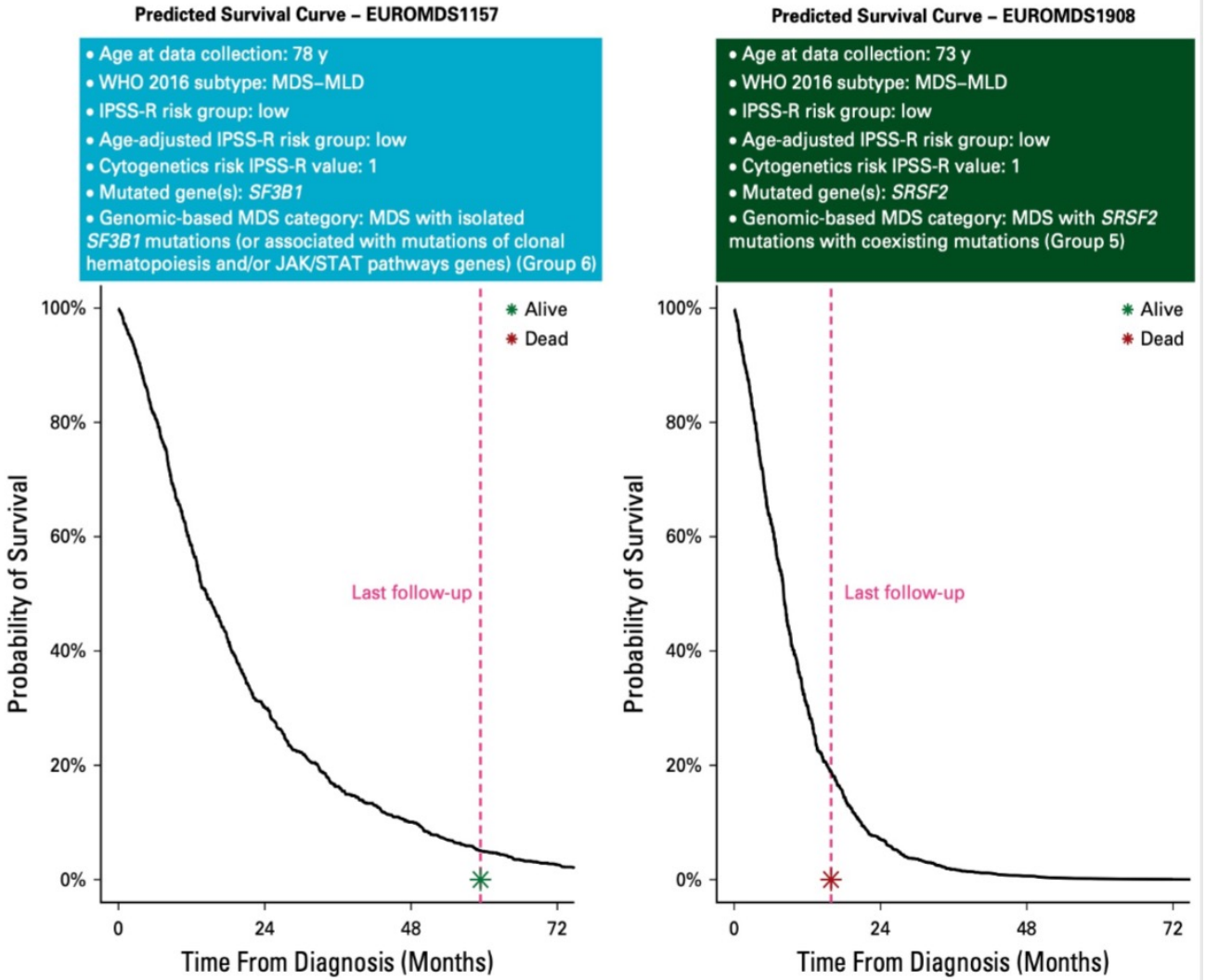
	Training (EuroMDS Cohort)		Validation (Humanitas Cohort)	
Statistical Model and Variable Selection	**Concordance**	**SD**	**Concordance**	**SD**
CoxRFX_Clinical + demographics + Dirichlet processes	0.715	0.012	Not applicable	Not applicable
CoxRFX_Clinical + demographics + genomics	0.737	0.012	0.753	0.037

Towards a new prognostication of MDS

Comparison of predicted survival curves for two real patients with same

- Age range
- IPSS-R classification
- Cytogenetic risk

but different genomic classification show significantly different behavior.



Towards a new prognostication of MDS

In order to reach clinical practice effectively, the results must converge into a score that is simple to understand and compute.

Webserver available at:
https://mds.itb.cnr.it/#/mds/home

- **IPSS-R** Revised international Prognostic Scoring System (IPSS-R)
- **PSS** Sex-informed Prognostic-Scoring-System (Including sex & age at diagnosis)
- **GSS** Sex-informed Genomic Scoring System

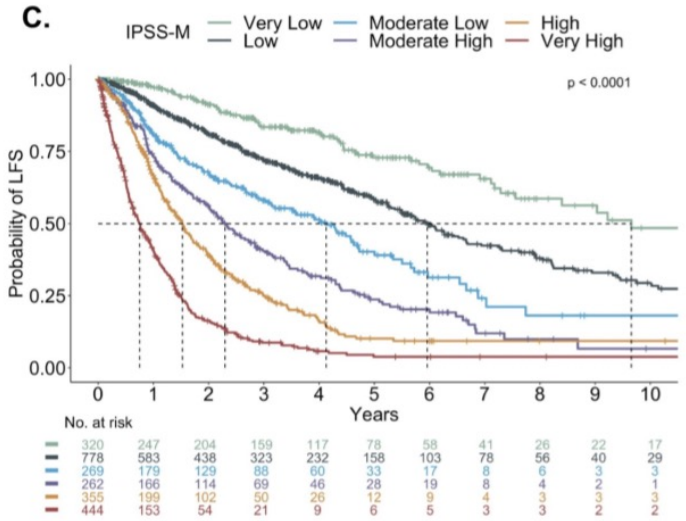
IPSS-M Molecular International Prognostic Scoring System

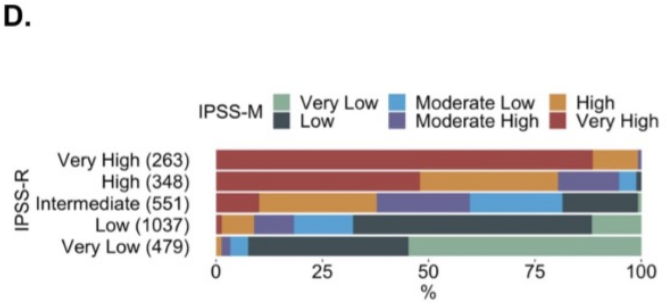
- Based on a panel of several genes and cytogenetic information
- Integrating demographic and clinical data
- The weight of each data entry contribution are retrieved from a survival model
- Additive score (similar to IPSS-R)

International Working Group on Myelodysplastic Syndromes

Molecular International Prognostic Scoring System for Myelodysplastic Syndromes

- Diagnostic MDS samples from **2,957 patients** with less than 20% blasts were profiled for mutations in 156 genes (discovery cohort). The model was validated in an independent cohort of **718 patients**.
- 9,339 driver point mutations or short indels involving 124 genes across the 2,957 patients were characterized.
- At least one gene mutation was characterized in 90% of patients, and 2 or more in 71%.
- The IPSS-M risk score consisted of
  - hemoglobin, platelets and bone marrow blasts
  - IPSS-R cytogenetic category
  - 22 binary features derived from the presence of mutations in 21 predictive genes and one feature representing the number of mutations from a group of 17 additional genes.





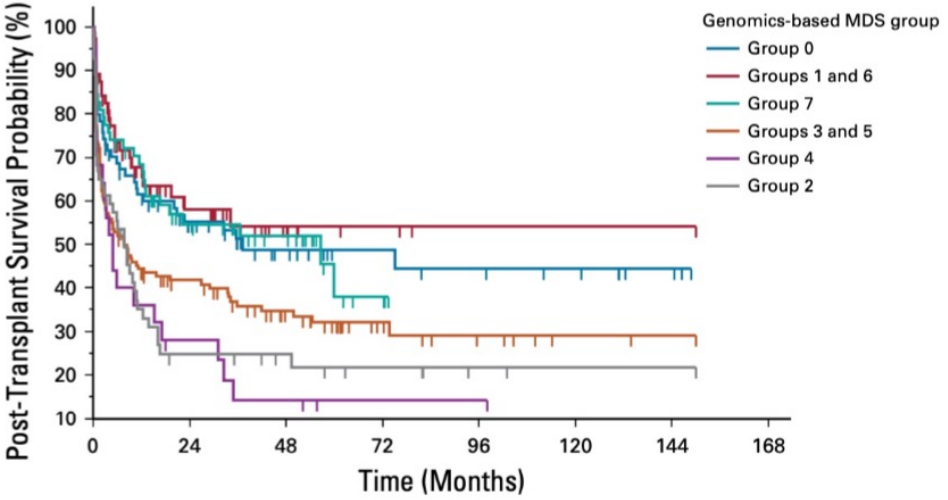
Elsa Bernard et al. 2021 (ASH)

Conclusions

- Both **clonal and sub-clonal mutations** have a **significant impact on patient outcome**. The different impact of clonal vs subclonal mutations needs to be further investigated.
- Performed **genomic classification of MDS** using BN and HDP with clinically interpretable outcome.
- Including NGS data allows to define new **predictive models** that:
  - measure a **high impact of specific genomic profile over prognostic outcome**
  - show **significantly better predictive performances** with respect to traditional scores
  - allows **personalized outcome prediction models** (PSS, GSS)
  - Allows the **definition of simplified molecular scores** (IPSS-M) that aim at entering the clinical practice

Future Work

- **IPSS-M**: Independent validation, robustness, applicability (what is the minimum number of genes to be tested in order to significantly improve prognostication accuracy for the wide majority of patients? / How much does an innovative score lose in predictive accuracy in relation to how much information is not available?)
- **Personalised treatment**, with focus on **transplant policies** using multi-state modelling.
- Integration with **more layers of data** (Multi-omics, Protein-protein interaction networks, Biological pathways, Imaging, Single Cell)
- **Interpretable AI** in order to reduce black-box effects



Number at risk								
Group: Group 0	94	32	17	11	9	7	3	0
Group: Groups 1 and 6	75	21	8	3	1	1	1	0
Group: Group 7	65	24	16	3	0	0	0	0
Group: Groups 3 and 5	163	45	27	12	7	3	2	0
Group: Group 4	31	6	3	1	1	0	0	0
Group: Group 2	62	11	8	5	2	1	1	0

Acknowledgements

C.A.L.R.

CENTER FOR ACCELERATING LEUKEMIA/LYMPHOMA RESEARCH

Artificial Intelligence and real world data analysis to improve patient care and advance medical research in hematology

HUMANITAS HU HUMANITAS UNIVERSITY POLITECNICO MILANO 1863

GEN MED 4ALL

HU HUMANITAS UNIVERSITY

Horizon2020 European Union Funding for Research & Innovation

HUMANITAS RESEARCH HOSPITAL

AI people

- Ettore Mosca
- Elisabetta Sauta
- Saverio D'Amico
- Victor Savevski
- Francesca Ieva
- Gastone Castellani
- Matteo Gnocchi

Clinical team

- Matteo Della Porta
- Erica Travaglino
- Giulia Maggioni
- Alessia Campagna
- Marta Ubezio
- Antonio Russo
- Cristina A Tentori
- Luca Lanino

Thank you for your attention