



The 65th ASH Annual Meeting Abstracts

ONLINE PUBLICATION ONLY

636.MYELODYSPLASTIC SYNDROMES-BASIC AND TRANSLATIONAL

Multi-Omics Factor Analysis (MOFA) Identifies Transposable Element Expression As a Risk Factor and Inflammaging As a Protective Factor in Myelodysplastic Syndromes

Shan Hama, MSc¹, Nogayhan Seymen, MSc¹, Sila Gerlevik, MSc¹, Deniz Ece Kaya, PhD¹, Giorgio Napolitani, PhD², Seishi Ogawa, MD PhD^{3,4}, Ghulam Mufti, DM, FRCPath², Mohammad M. Karimi, PhD²

¹Comprehensive Cancer Centre, School of Cancer and Pharmaceutical Sciences, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom, London, United Kingdom

²King's College London, Comprehensive Cancer Centre, School of Cancer and Pharmaceutical Sciences, London, United Kingdom

³Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto City, Japan

⁴Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan

Introduction

Myelodysplastic syndromes (MDS) are haematological diseases characterised by clonal proliferation due to genetic and epigenetic alterations within haematopoietic stem and progenitor cells. Thus far, the prediction of a patient's overall survival (OS) and event-free survival (EFS) has been dependent predominantly on the degree of peripheral blood cytopenias, bone marrow blast percentage, cytogenetics, and genetic features. However, this may overlook other biological phenotypes, including, for example, the expression of transposable elements (TEs), ageing, and immunology profile. In this study, we present a comprehensive analysis of three data modalities (clinical, genotypic, and transcriptomic) and eight different views derived from these modalities to identify the hidden factors that significantly impact MDS prognosis.

Methods

Three (immunology, ageing, and cellular composition) out of eight views of the data were derived from RNA-seq by applying singscore, a gene set scoring tool developed by Foroutan *et al.*, 2018, on RNA-seq gene expression from bone marrow samples of 94 MDS patients obtained from Shiozawa *et al.*, 2017 (Fig 1.A). Each of these three views was carried forward in the workflow by a number of gene sets. The other five views were clinical numeric data, clinical categorical data, TE expression, genotype data, and known AML/MDS antigens (obtained from TANTIGEN 2.0), expressed in at least 10% of MDS samples in the study. We then employed Multi-Omics Factor Analysis (MOFA), a computational framework developed by Argelaguet *et al.*, 2018, to uncover hidden phenotypic states within a multi-omic dataset. Through MOFA, we generated factors representing specific biological data modalities, establishing connections between features across these different categories. The total variance (R^2) that can be attributed to specific factors is also exhibited by MOFA.

Results

Factors were ranked based on their explained variance. We present the 10 factors with a threshold of variance explained of 2% from the 15 default factors generated by MOFA (Fig 1.A). Factor 1 was one of the factors that linked the most categories of data: immunology, cellular composition, ageing, and TEs. Most notably, Factor 1 was aligned with features of less differentiated cells (more HSC-like and less GMP-like). We observed that samples represented by high Factor 1 scores also had a higher expression of SINE:Alu elements. (Fig 1.A). Moreover, Factor 1 aligns the following immunology features: increased T-helper 1 (Th1) cells, macrophages, and T regulatory cells. The relationship between Th1 and SINE elements is particularly interesting, given the role of Th1 cell activation in host defence against infection that may be modulated by SINE activation. Factor 2 comprised of ageing and clinical categorical data, which showed that less inflammaging may lead to a poorer EFS ($P=0.00023$) (Fig 1.B). Factor 4 linked immunology, cellular decomposition, ageing and clinical categorical data. The following features mainly influenced Factor 4: increased HSC-like cells, decreased CD8+ T cells, decreased immunosenescence and increased inflammatory chemokines. Using Kaplan-Meier plots to predict OS, we observed that patients with more HSC-like cells, fewer leukocytes and T cells, and less immunosenescence and inflammaging were linked to poorer EFS ($P=0.0001$) and OS ($P=0.00017$, data not shown). Furthermore, Factor 9 comprised solely of the TEs category and exhibited a correlation between increased TE expression and poorer EFS ($P=0.0074$) and OS ($P=0.0065$, data not shown) (Fig 1.B).

Conclusions

Our analysis using MOFA on a diverse set of biological data modalities uncovered hidden factors in MDS, providing insights into the complex interplay between various biological layers. Notably, we identified TE expression as a risk factor and inflam-maging as a protective factor in MDS. HSC-like cells and specific immunological profiles were also associated with significantly poor OS and EFS. This study contributes to a deeper understanding of MDS pathogenesis and identifies potential prognostic markers for this disease. It also elucidates the importance of considering the relationships between different pathways, mark-ers, and mutations in predicting patient outcomes, highlighting the efficacy of a comprehensive approach that goes beyond all the scoring systems described thus far for MDS.

Disclosures No relevant conflicts of interest to declare.

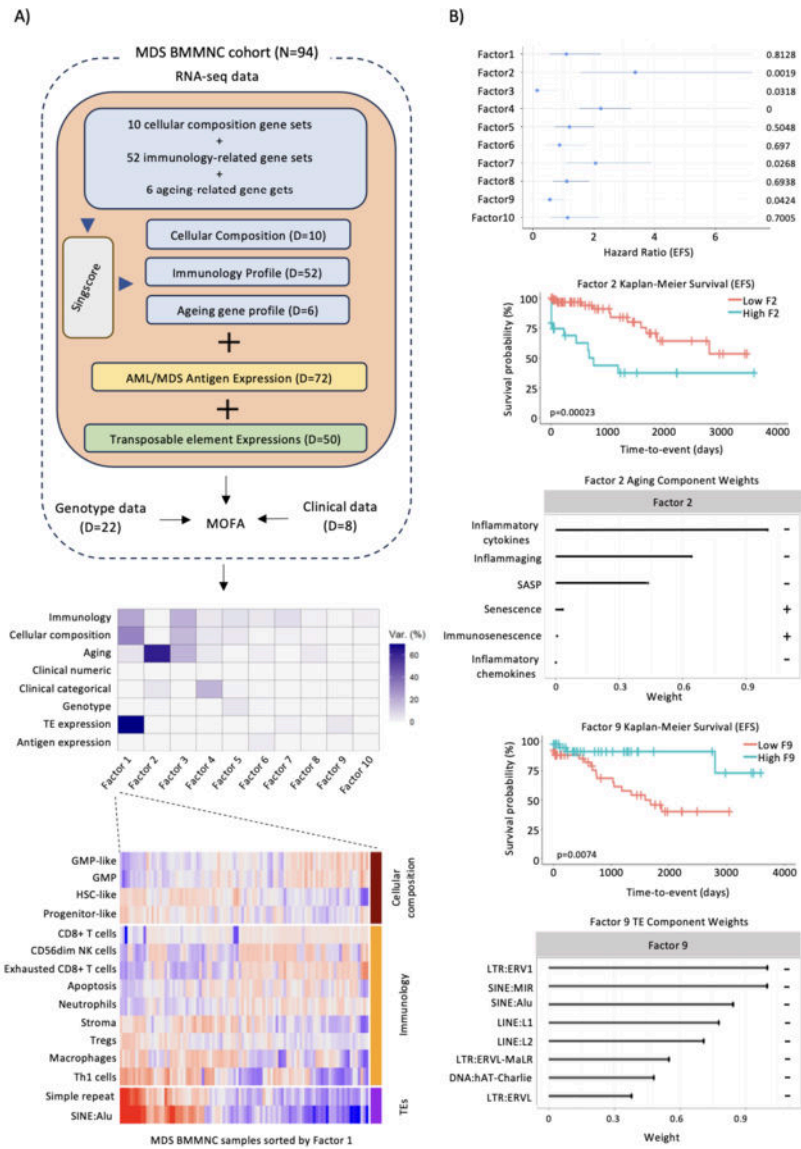


Figure 1: Workflow overview and downstream analyses. A) RNA-seq, genotype and clinical data were obtained from bone marrow mononuclear cell (BMMNC) samples of 94 MDS patients from Shiozawa *et al.*, 2017. Gene sets related to cellular composition, immunology and ageing profiles were processed through singscore to obtain gene signature scores. In total, there were eight views of the data: immunology profile, cellular composition, ageing, clinical numeric, clinical categorical, genotype, TE expression and AML/MDS antigen expression. The data were put through MOFA to identify factors and the variance decomposition by factor. The heatmap shows the characterisation of Factor 1, depicting which features have the greatest influence on this factor per sample. B) Univariate Cox regression was conducted to determine the hazard ratio for time to event associated with each factor. *P*-values are based on the Wald statistic. Error bars depicting 95% confidence interval. Kaplan-Meier plots measuring time to event for individual MOFA factors, with *P*-values created using a log-rank test. Absolute loadings for features with largest absolute weights on Factors 2 and 9 can be seen. Plus or minus symbols on the right depict an increased or decreased load.

Figure 1

<https://doi.org/10.1182/blood-2023-177683>