

Identification of novel myelodysplastic syndromes prognostic subgroups by integration of inflammation, cell-type composition, and immune signatures in the bone marrow

Reviewed Preprint

v2 • August 21, 2024

Revised by authors

Reviewed Preprint

v1 • June 11, 2024

Sila Gerlevik, Nogayhan Seymen, Shan Hama, Warisha Mumtaz, I Richard Thompson, Seyed R Jalili, Deniz E Kaya, Alfredo Iacoangeli, Andrea Pellagatti, Jacqueline Boulwood, Giorgio Napolitani, Ghulam J Mufti , Mohammad M Karimi 

Comprehensive Cancer Centre, School of Cancer & Pharmaceutical Sciences, Faculty of Life Sciences & Medicine, King's College London, London SE5 8AF, United Kingdom • Department of Basic and Clinical Neuroscience, King's College London, London, United Kingdom • Department of Biostatistics and Health Informatics, King's College London, London, United Kingdom • NIHR BRC SLAM NHS Foundation Trust, London, United Kingdom • Perron Institute for Neurological and Translational Science, University of Western Australia Medical School, Perth, WA 6009, Australia • Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Mutational profiles of Myelodysplastic syndromes (MDS) have established that a relatively small number of genetic aberrations, including SF3B1 and SRSF2 spliceosome mutations, lead to specific phenotypes and prognostic subgrouping. We performed a Multi-Omics Factor Analysis (MOFA) on two published MDS cohorts of bone marrow mononuclear cells (BMMNCs) and CD34+ cells with three data modalities (clinical, genotype, and transcriptomics). Seven different views, including immune profile, inflammation/aging, Retrotransposon (RTE) expression, and cell-type composition, were derived from these modalities to identify the latent factors with significant impact on MDS prognosis. SF3B1 was the only mutation among 13 mutations in the BMMNC cohort, indicating a significant association with high inflammation. This trend was also observed to a lesser extent in the CD34+ cohort. Interestingly, the MOFA factor representing the inflammation shows a good prognosis for MDS patients with high inflammation. In contrast, SRSF2 mutant cases show a granulocyte-monocyte progenitor (GMP) pattern and high levels of senescence, immunosenescence, and malignant myeloid cells, consistent with their poor prognosis. Furthermore, MOFA identified RTE expression as a risk factor for MDS. This work elucidates the efficacy of our integrative approach to assess the MDS risk that goes beyond all the scoring systems described thus far for MDS.

eLife assessment

This manuscript uses public datasets of myelodysplastic syndrome (MDS) patients to undertake a multi-omics analysis of clinical, genomic, and transcriptomic datasets.

Useful findings are provided by way of interesting correlations of specific mutations with inflammation and differing clinical outcomes. The evidence is **solid** and interesting, and the manuscript is of substantive value to hematologists and clinical immunologists.

<https://doi.org/10.7554/eLife.97096.2.sa2>

Introduction

Myelodysplastic syndromes (MDS) are haematological diseases characterised by clonal proliferation due to genetic and epigenetic alterations within haematopoietic stem and progenitor cells (1, 2). Thus far, the prediction of a patient's overall survival (OS) and event-free survival (EFS) has been predominantly dependent on the degree of peripheral blood cytopenias, bone marrow (BM) blast percentage, cytogenetics, and genetic features (3–6). This may, however, overlook other biological phenotypes and risk factors in MDS, including inflammation, age and aging characteristics, immune profile, BM cell-type composition, and expression of the noncoding sequences in particular retrotransposable elements (RTEs).

Inflammation can either suppress or promote cancer development (7). Inflammation can be protective against malignancies, including MDS, through innate and adaptive immune responses, enhancing antitumor immunity by promoting the maturation and function of dendritic cells (DCs) and initiating effector T cell responses (8). However, low-level chronic inflammation can result in an immunosuppressive milieu preventing the innate and T-cell antitumor immunity (9). Inflammaging is the process by which an age-related increase in chronic inflammation occurs, but the extent to which this process and other age-related events can impact the overall prognosis in MDS is yet to be uncovered (10, 11).

RTEs are genomic remnants of ancient DNA sequences that comprise a large part of the human non-coding genome and are evolutionarily silenced. RTEs include three classes: long terminal repeat (LTR) RTEs, and long and short interspersed nuclear elements, known as LINEs and SINEs, respectively. As part of the host defence mechanism, RTEs are silenced through DNA methylation and histone modifications in somatic cells (12). Mutations in genes implicated in DNA methylation and histone modifications (DNMT3A, TET2, ASXL1, and IDH1/2) are frequently reported in MDS. It is postulated that global hypomethylation can reactivate RTEs in MDS cases with epigenetic mutations, particularly in DNMT3A mutant cases, as has previously been shown for other cancer types (13, 14). Despite the availability of multiple transcriptomics data for MDS (15–18), a comprehensive assessment of RTE expression and its relationship to genetic variations or prognosis has not been documented.

Splicing factor (SF3B1, SRSF2, and U2AF1) mutations are the most common mutations in MDS (19–22). SF3B1 mutations are associated with the MDS ring sideroblasts type, good prognosis, and low leukemic transformation (20). In contrast, SRSF2 mutations are associated with poor prognosis and are more prevalent in the male sex and older age (23). Recent studies have shown mutations in splicing factors can induce chronic innate immunity and enhance NF-κB signalling in MDS through aberrant splicing of various target genes. (18, 24, 25). These recent mechanistic studies have significantly advanced our understanding of the consequences of

splicing factor mutation in human and model organisms. However, we still lack a systematic approach to integrate splicing factor mutations with other players in the tumour microenvironment, including immune profile and BM cell-type composition.

Recent work has explored the relationship between transcriptional signatures and critical signalling pathways to determine survival prognosis and diagnostic efficacy in MDS patient cohorts (26 [↗](#)). However, no studies to date integrated clinical MDS phenotypes, RTE expression, cell-type composition, and immune and aging gene signatures. This study employs MOFA for a comprehensive analysis of three data modalities (clinical, genotypic, and transcriptomic) and seven different “views” derived from these modalities to identify the factors that may impact MDS prognosis. MOFA could not identify any factor representing splicing factor mutations; hence, we examined our entire feature sets from cell-type composition, immune profile, and inflammation/aging views to identify the features associated with mutations in SF3B1 and SRSF2 genes in MDS cohorts.

Methods

MDS RNA-seq cohorts

We utilised two RNA-seq datasets for MDS (**Supplementary Table 1**). The first was data from bone marrow mononuclear cells (BMMNCs) of 94 MDS patients obtained from the Shiozawa *et al.* study (17 [↗](#)) that is enriched with splicing factor mutations (**Supplementary Table 2**). The second dataset utilised bone marrow CD34+ haematopoietic stem and progenitor cells (HSPCs) data that Pellagatti *et al.* derived from 82 MDS patients (15 [↗](#)), which again focused on MDS cases with splicing factor mutations (**Supplementary Table 2**).

Generating gene signature scores using *singscore*

For the immunology and inflammation/aging gene sets, we performed a meticulous literature review and generated a list of gene sets from previously published articles (**Supplementary Table 3**). The markers for the cellular composition gene sets were taken from van Galen *et al.* (27 [↗](#)). For each of the curated gene sets, instead of looking at individual gene expressions, we used *singscore* (version 1.20.0) (28 [↗](#)), a method that scores gene signatures in a cohort of samples using rank-based statistics on their gene expression profiles. For RNA-seq data, we provided Reads per Million (RPM) normalized expression values to *singscore*. In the case of microarray data, we first compiled the microarray expression matrix for average expression values for all probes overlapping each gene using the *limma* package (29 [↗](#)) in R; then we used the *rankGenes* function from the *singscore* package to rank each gene sample-wise. Eventually, the *multiScore* function was used to calculate signature scores for all gene sets at once.

Multi-Omics Factor Analysis (MOFA)

MOFA was applied on seven views derived from the BMMNC and CD34+ cohorts: immune profile, cell-type composition, inflammation/aging, genotype, RTE expression, clinical numeric, and clinical categorical views, using the *MOFA2* package (version 1.10.0) in R (**Fig. 1a** [↗](#)). Each view consists of non-overlapping features of the same sample set of patients. Since aging and inflammation views share some gene sets, including inflammatory chemokines and cytokines, and we did not want to repeat these features through MOFA analysis, we combined these two views as the inflammation/aging views throughout this study. The views were scaled to have the same unit variance via the *scale_view* option from the MOFA model. The model pruned inactive factors incapable of capturing significant variance within the biological views, generating ten factors with minimum explained variance 2% in at least one biological view. Significant features within each view were determined based on the absolute weight threshold above 0.5 in at least one of the ten identified factors.

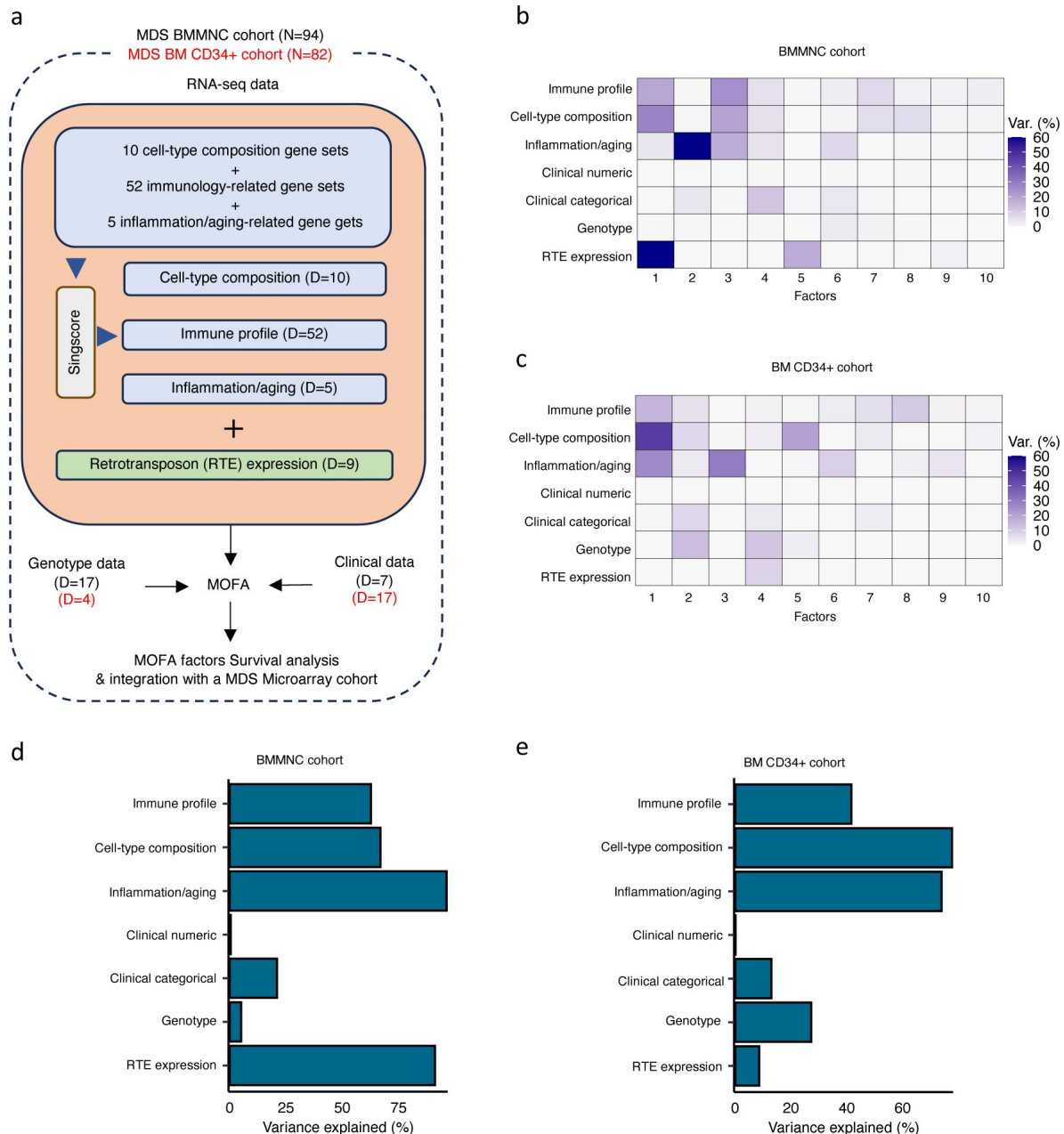


Figure 1.

Schematic of MOFA workflow overview, downstream analyses, and factor determination in the BMMNC and BM CD34+ cohorts.

a, RNA-seq, genotype and clinical data were obtained from bone marrow mononuclear cell (BMMNC) samples of 94 MDS patients from Shiozawa *et al.* and BM CD34+ samples of 82 patients from Pellagatti *et al.* studies. We generated seven views of the data where three of which were derived from RNA-seq data after applying *Singscore*: immune profile, cell-type composition, inflammation/aging. The other four views were clinical numeric and categorical (**Supplementary Table 4**), genotype and retrotransposable element (RTE) expression. The data were put through MOFA to identify latent factors and the variance decomposition by factors. The number of features (dimensions) per view is abbreviated by “D”. **b-c**, The determined factors for the BMMNC and BM CD34+ cohorts and the percentage of explained variance for each view per identified factor were shown. **d-e**, Bar charts depict the total variance explained for each biological data view by all the factors combined in the BMMNC and BM CD34+ cohorts.

Survival analysis

The association between latent factors and survival outcomes was investigated with Cox regression analysis and Kaplan-Meier curves via the R package *survival* (version 3.5-5). Within the BMMNC cohort, overall and event-free survivals were used as separate response variables in the univariate Cox regression, with latent factors employed as predictors. Additionally, in the multivariate Cox regression, age and sex were included as predictors alongside the factors. In the CD34+ cohort, only overall survival was used as the response variable for both regression analyses. Kaplan-Meier plots were constructed for factors exhibiting a significant hazard ratio in the univariate Cox regression, categorizing factor values into three groups: ‘low’ for the 1st quartile, ‘high’ for the 4th quartile, and ‘intermediate’ otherwise. The statistical significance between the high and low factor groups was determined using the log-rank test via the R package *survminer* (version 0.4.9).

Differential Expression and Gene Set Enrichment Analysis (GSEA)

We performed differential expression analysis in high versus low Factor 1 groups using *DESeq2* (30) and generated p-values and statistics for each gene. The genes were sorted based on the “stat” column in *DESeq2* and provided to the GSEA software (31) as an input. GSEA was separately run on cancer hallmark and Reactome gene set databases. The GSEA output was the list of up or down-regulated pathways from each database.

RTE expression

To generate the RTE expression, we mapped the RNA-seq reads to RepeatMasker to extract the reads covering the RTE regions and calculated the Reads Per Million (RPM) scores for each class and family of RTEs. We included nine families from three main RTE classes: 1) CR1, L1 and L2 families from LINE; 2) Alu and MIR from SINE; and 3) ERV1, ERVL, ERVL-MaLR, and ERVK families from LTR.

Results

MOFA identified latent factors linking different views in multimodal MDS data

We applied MOFA to identify latent factors within BMMNC and CD34+ MDS cohorts (Fig. 1a). To run MOFA on these two MDS cohorts, three (immune profile, inflammation/aging profile, and cell-type composition) out of seven views were derived from RNA-seq by applying *singscore* (28) on RNA-seq gene expression. Each of these three views were carried forward in the workflow by a number of gene sets, and per gene set, *singscore* generated relative gene signature scores for all samples within each cohort (Fig. 1a and Supplementary Table 3). The other four views were clinical numeric, clinical categorical, genotype, and RTE expression.

We have provided the data generated for these seven views to MOFA in two separate runs for BMMNC and CD34+ MDS cohorts. For each cohort, we could identify ten factors (minimum explained variance 2% in at least one biological view) from the 15 default factors generated by MOFA (Fig. 1b-c). These factors displayed a relationship with different biological views, with Factor 1 as the most dominant factor, linking immune profile, cell-type composition, and inflammation/aging profile in both cohorts. Further to this, in the BMMNC cohort, a high level of variance for RTE expression was explained by Factor 1 (Fig. 1b). We also observed similar trends in terms of the level of variance explained by the identified factors for the biological views in both cohorts, with the only notable difference being a greater level of variance explained for

genotype data and lower level of variance explained for RTE expression in the BM CD34⁺ cohort versus the BMMNC cohort (**Fig. 1d-e**). Additionally, dividing the patients based on low (1st quartile), intermediate (2nd and 3rd quartile), and high (4th quartile) levels of Factor 1 in both cohorts could successfully stratify patients in the dimensionality reduction plots obtained from applying principal component analysis (PCA) on gene expression data from both cohorts (**Supplementary Fig. 3**).

MOFA also identified the highly weighted features within each factor in the BMMNC and BM CD34⁺ cohorts (**Fig. 2a-b**), in which each feature belongs to a specific biological view. We further characterised Factor 1 as the most dominant factor linking multiple features from different views. We observed that high Factor 1 in the BMMNC cohort represented patients who have cells with stem and progenitor-like characteristics. This includes but is not limited to a positive correlation with features comprising progenitor-like and HSC-like (**Fig. 2c**). In contrast, there is an inverse correlation between GMP/GMP-like and Factor 1 scores (**Supplementary Fig. 1c**). SINE:Alu expression is increased in patients high in Factor 1 within this cohort, representing a group that may have significant levels of genetic instability (**Supplementary Fig. 1a**). Moreover, Factor 1 correlates with the following immunology features: increased T-helper 1 (Th1) cells, and decrease in certain immune cells, especially in neutrophils and exhausted CD8⁺ T cells (**Supplementary Fig. 1b**). The relationship between Th1 and SINE elements is particularly interesting given the role of Th1 cell activation in host defence against inflammation that may be modulated by SINE activation. Furthermore, there is a modest correlation between immunosenescence and exhausted CD8⁺ T cells scores, particularly in patients with higher levels of GMPs, demonstrating the ineffectiveness of the immune system in GMP-dominant patients (**Fig. 2c**).

MOFA analysis of the CD34⁺ MDS Cohort revealed factors associated with different immune signatures. Factor 1 is associated with signatures of differentiated myeloid cells (**Fig. 2d** and **Supplementary Fig. 2c**), Factor 2 with HSPC, and Factor 4 with GMP progenitor cells, suggesting that these factors might correlate with distinct differentiation states of MDS blasts (**Fig. 2b**). Factor 1 is also associated with cytolytic and cytotoxicity score (**Fig. 2d** and **Supplementary Fig. 2b**). Since CD34⁺ cells used for this analysis were purified using anti-CD34 conjugated beads and not FACS sorted, this signal might depend on the presence of T or NK cells contaminations in the CD34⁺ fraction. The trend that we observed for Factor 1 prompted us to conduct a differential expression and gene set enrichment analysis (GSEA) to identify the pathways that are dysregulated in patients having high (4th quartile) versus low (1st quartile) levels of Factor 1. Interestingly, we found upregulation of inflammatory, Interferons, TNFA, JAK-STAT signalling pathways within cancer hallmark gene sets (**Supplementary Fig. 4a**) and upregulation of chemokines and Neutrophils amongst Reactome gene sets (**Supplementary Fig. 4b**) supporting the trend observed by MOFA (**Fig. 2d** and **Supplementary Fig. 2b**). This analysis suggests Factor 1 as an immune-active factor associated with high-level of cytotoxicity and inflammation in the CD34⁺ MDS Cohort.

The Cox regression models identified the high expression of RTEs as a risk factor and inflammation as a protective factor in MDS

We explored the power of the latent factors generated by MOFA as predictors of MDS survival. Univariate and multivariate Cox regression models identified three of the ten factors identified by MOFA were significantly associated with OS or EFS in the BMMNC cohort (**Table 1**). The univariate analysis was conducted to investigate the association of the factors with survival, and the multivariate analysis controlled for sex and age. Our univariate analysis demonstrated that Factors 4 and 9 significantly impact OS (HR: 0.39; P = 0.028 and HR: 0.49; P = 0.012, respectively) and EFS (HR: 0.48; P < 0.001 and HR: 0.59; P = 0.007, respectively) in patients in the BMMNC cohort (**Table 1**). Interestingly, the impact of Factor 4 on EFS became insignificant after controlling for sex and age. Though no significance was associated with OS and Factor 2, EFS in both the univariate

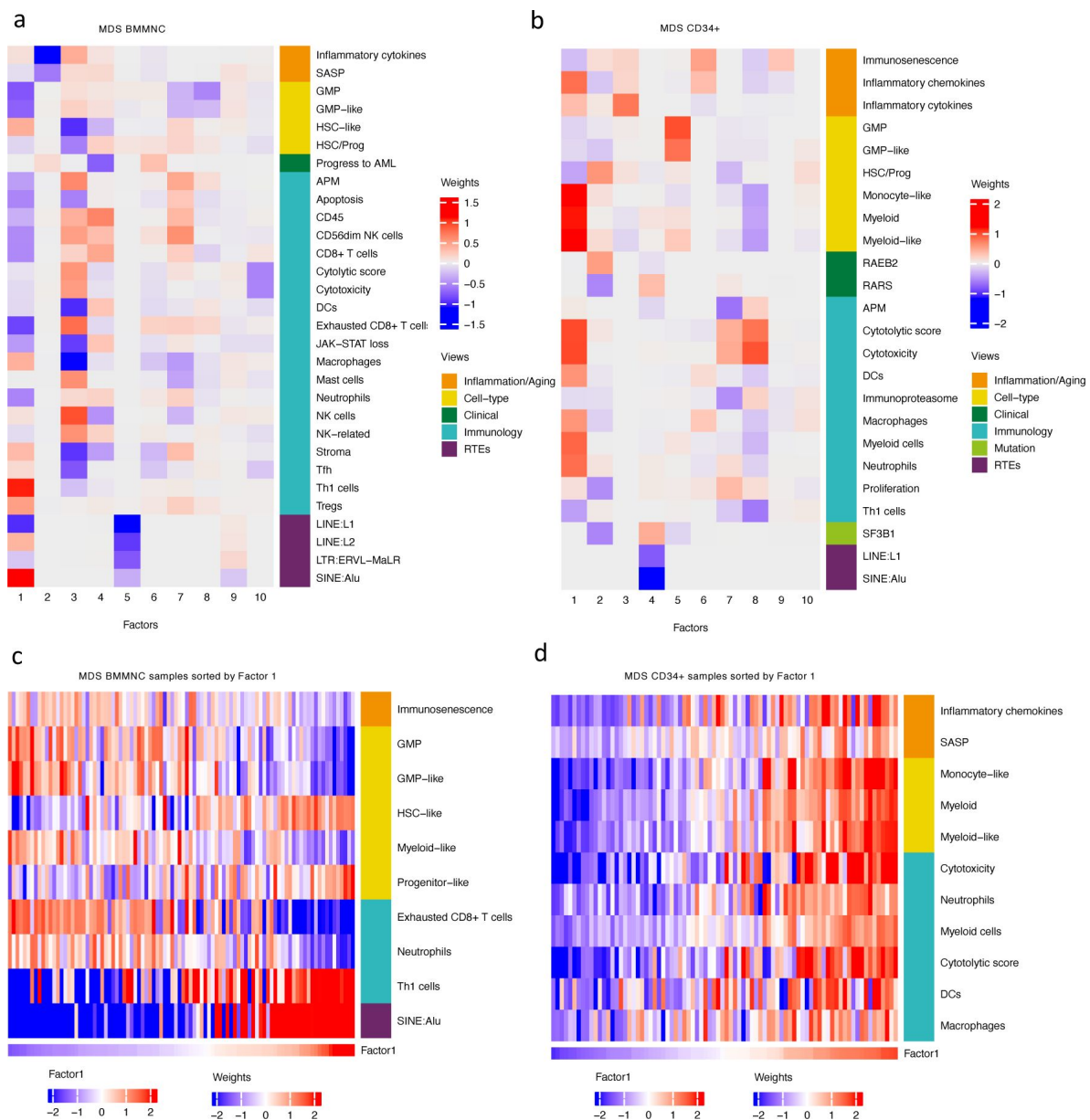


Figure 2.

Breakdown of important features for each factor generated by MOFA.

a-b, The important features with high weights for each biological view per factor were shown for the BMMNC and BM CD34+. Blue represents features with the inverse correlation with the factor, and red shows the positive correlation. **c-d,** Characterisation of Factor 1 in the BMMNC and BM CD34+ cohorts, showing only those features highly influencing Factor 1 for the patients in these cohorts. Patients were sorted by Factor 1 values.

and multivariate analyses exhibited a statistically significant association with this factor (HR: 1.7; $P = 0.033$ and HR: 2.27; $P = 0.033$, respectively) (**Table 1**). The 10 factors identified by MOFA in BM CD34⁺ cohort did not show any significance associated with MDS overall survival (**Supplementary Table 5**).

Dividing the patients based on low (1st quartile), intermediate (2nd and 3rd quartile), and high (4th quartile) levels of Factors 2, 4, and 9 in the BMMNC cohort, we found Factors 4 and 9 exert a protective influence over the prognosis of MDS patients (**Supplementary Fig. 5a** and **Fig. 3a**, respectively), whilst higher levels of Factor 2 predict a poorer prognosis for these same patients (**Fig. 4a**). Upon investigation of the RTE absolute loading of features in Factor 9, patients with low Factor 9 level showed increased LTR:ERV1, SINE:MIR and SINE:Alu demonstrating the high level of RTE expression as a potential risk factor for MDS (**Fig. 3a-b**). In contrast, for Factor 2, some component weights for inflammation/aging including inflammatory cytokines and SASP were increased once the Factor 2 level is decreased suggesting the secretion of cytokines and the down-stream inflammation as a protective factor for MDS (**Fig. 4a-b**).

Factor 4 in the BMMNC cohort shows several data views and displays a more specific phenotype of the patients. This phenotype includes accumulation of immune cells, such as CD56dim NK cells, CD8⁺ T cells, DCs, and neutrophils (**Fig 2.a**). Thus, Factor 4 points to a more immune-active disease. In line with this, there is a decreased rate of the 'progress to AML' phenotype (**Fig 2.a** and **Supplementary Fig 5.b**) in patients with high level of Factor 4. In contrast, for the patients showing low level of Factor 4, we observed that the number of healthy stem and progenitor cells (HSCs) decreased (**Supplementary Fig 5.c**), whereas the number of malignant HSCs (HSC-like) increased (**Supplementary Fig 5.c-d**). Most notably, the BM of these patients have a high content of stroma and a low content of leukocytes, represented by low CD45 marker score (**Supplementary Fig 5.c and e**). Overall, this result suggests that for a subset of MDS patients, represented by the low level of Factor 4, the haematopoiesis is impaired due to the depletion of healthy HSCs, leading to the decrease in the number of leukocytes, an increase of stroma, and features of secondary AML.

BM CD34⁺ cohorts support better prognosis for high-inflamed cases but poor prognosis for IFN-1-induced cases

To better understand the drivers of prognosis for MDS, patients were split into quartiles (high 25%, low 25% and intermediate 50%) for survival analyses depending on their levels for individual features (inflammatory cytokines, inflammatory chemokines, and IFN-I signature) in the two MDS cohorts. In addition, a microarray gene expression dataset obtained from the HSPCs isolated from BM of 183 MDS patients (16) was also included in this analysis. Recall, in the BMMNC cohort, Factor 2 was primarily represented by the inflammation/aging biological view (**Fig. 4b**) and Factor 9 was represented by RTE expression (**Fig. 3b**). To confirm the isolated impact of the aforementioned features on patient survivals in CD34⁺ cohorts, we created survival plots for each of the three cohorts separating patients into quartiles based on high, low and intermediate levels of the features. The only situation in which we observed statistical significance was the high score of inflammatory cytokines in the BMMNC cohort, which was associated with superior OS ($P = 0.019$) (**Fig. 4c**) supporting the strong contribution of this feature to Factor 2 (**Fig. 4b**). For the rest of the analyses, statistical significance was not achieved, though there may be a trend pertaining to higher inflammatory chemokines signature and better prognosis in CD34⁺ cohorts (**Fig. 4c**). Further to this, higher IFN-I signature score was also related to a poorer prognosis in CD34⁺ cohorts (**Fig 3.c**). Strikingly, this pattern was not observed in the BMMNC cohort (**Fig 3.c**), suggesting an alternative mechanism other than IFN-I activation can contribute to poor MDS prognosis due to retroviral activations (**Fig 3.a-b**).

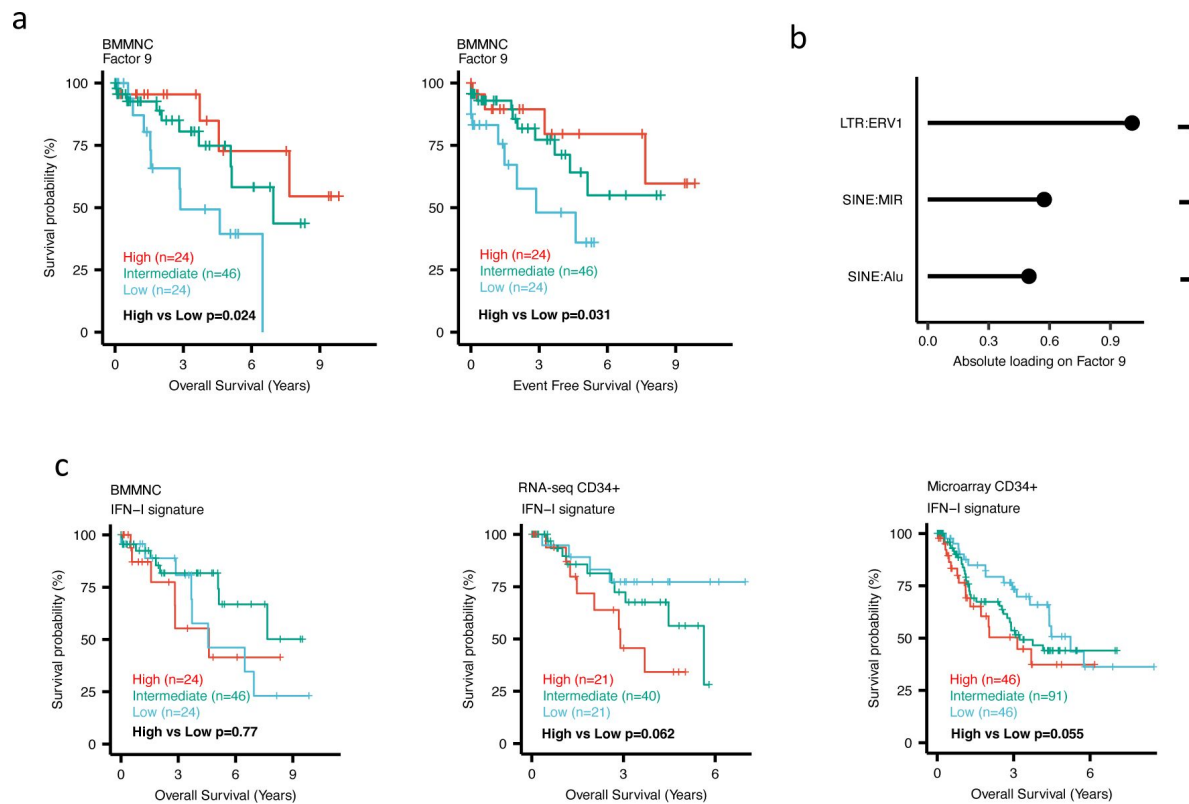


Figure 3.

Impact of Factors 9 and IFN-I levels on MDS prognosis.

a, Kaplan-Meier plots for the BMMNC cohort where patients were split based on low (1st quartile), intermediate (2nd and 3rd quartile), and high (4th quartile) levels of Factor 9. **b**, The absolute loading of the top three features affecting Factor 9 in RTE expression view in the BMMNC cohort. **c**, Kaplan-Meier plots where patients were split into quartiles (high 25%, low 25% and intermediate 50%) for survival analyses depending on their IFN-I signature score levels in the BMMNC, RNA-seq CD34+, and Microarray CD34+ cohorts. All p-values were calculated using log-rank test on overall and event-free survival values of high versus low groups.

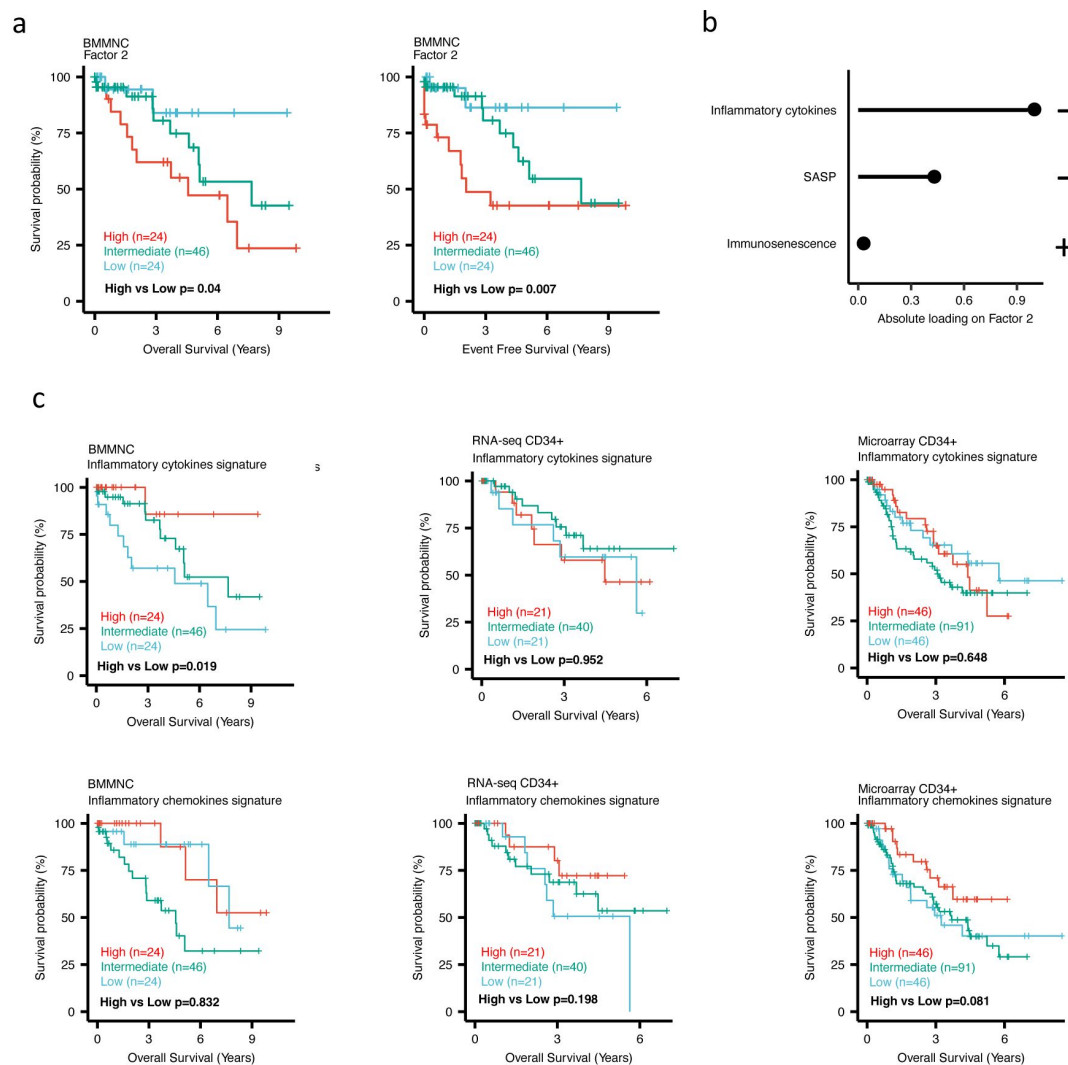


Figure 4.

Impact of Factors 2 and inflammation levels on MDS prognosis.

a, Kaplan-Meier plots for the BMMNC cohort where patients were split based on low (1st quartile), intermediate (2nd and 3rd quartile), and high (4th quartile) levels of Factor 2. **b**, The absolute loading of the top three features affecting Factor 2 in inflammation/aging biological view in the BMMNC cohort. **c**, Kaplan-Meier plots where patients were split into quartiles (high 25%, low 25% and intermediate 50%) for survival analyses depending on their inflammatory cytokines and chemokines levels in the BMMNC, RNA-seq CD34+, and Microarray CD34+ cohorts. All p-values were calculated using log-rank test on overall and event-free survival values of high versus low groups.

SF3B1 mutant cases show low HSPC content and high levels of inflammation in the BMMNC and BM CD34+ cohorts

When investigating the inflammation/aging and cell-type views in both the BMMNC and BM CD34+ cohorts, we identified a significant association between the occurrence of SF3B1 mutation and the high level of inflammatory chemokines (BMMNC $P < 0.01$; BM CD34+ $P < 0.05$) (**Fig. 5a-c**). However, we needed to know whether the co-occurrence of other mutations with SF3B1, as confounding factors, can affect the statistical significance of the association of SF3B1 mutation with inflammatory chemokines. Therefore, we applied a multiple linear regression model to examine the association between various mutations (covariates) and inflammatory chemokines while controlling the effect of other mutations. Strikingly, out of 13 different mutations in the BMMNC cohort, SF3B1 mutation was the only mutation showing a significant association with the inflammatory chemokine level in the multiple linear regression model ($P < 0.0001$) (**Table 2**).

In both cohorts, the SF3B1 mutant pertained to decreased HSC/Prog cells (BMMNC $P < 0.05$, BM CD34+ $P < 0.01$) (**Fig. 5a-b**). Furthermore, the BMMNC cohort saw a statistically significant increase in myeloid, HSC-like, and promo-like scores, along with inflammatory cytokines level ($P < 0.05$) in SF3B1 mutant cases (**Fig. 5a-b**). Applying multiple linear regression models and controlling the effect of other mutations supported the identified correlation between SF3B1 mutation and inflammatory cytokines level ($P < 0.02$) (**Table 2**). We then examined the Immunology view in the BMMNC cohort to see whether we could identify any immune signature that the SF3B1 mutation may influence (**Supplementary Fig. 6a**). We found that SF3B1 mutants tend to have low levels of leukocytes but high levels of stroma and macrophages (**Fig. 5d**). This result supports previous findings by Pollyea et al. (32, 33) demonstrating the generation of inflammatory cytokines due to macrophage activation in human BM samples obtained from SF3B1 mutant patients.

We also divided the SF3B1 mutant cases from the BMMNC cohort into high versus low groups depending on their levels of inflammatory chemokines and cytokines. Although survival plots did not reach statistical significance, the trend shows that SF3B1 mutant cases with higher inflammation tend to survive better than wild-type SF3B1 (**Fig. 5f-g**). This trend indicates that the SF3B1 mutation perhaps only induces macrophage activation to survive better in a subset of mutant cases.

SRSF2 mutant cases show high GMP content and high levels of senescence and immunosenescence

When assessing inflammation/aging and cell-type features for MDS SRSF2 mutant cases in both cohorts, we saw a statistically significant increase in GMP and GMP-like features ($P < 0.05$), which are precursor cells to granulocytes and monocytes (**Fig. 6a-c**). The increased frequency of GMPs is an inherent feature of high risk MDS (34, 35). In the BMMNC cohort, within SRSF2 mutant cases, there was a statistically significant decrease in healthy myeloid features ($P < 0.05$), along with an increase in myeloid-like (malignant myeloid) and senescence/immunosenescence ($P < 0.05$) (**Fig. 6a-b**). Further analyses of the BMMNC cohort indicated that SRSF2 mutants tend to have low levels of T cells and a reduction of T cell activity. The levels of several immune cells were decreased in SRSF2 mutants, including T, Th1, and Treg cells ($P < 0.05$) (**Fig. 6d** and **Supplementary Fig. 6b**). Patients with SRSF2 mutation, related to poorer prognosis in literature, also had a lower cytolytic score ($P < 0.001$) and increased central memory cells ($P < 0.01$) (**Fig. 6d**). Taken together, this result finds relevance to the immunosenescence and the reduction of the T cells and their cytolytic activity in SRSF2 mutant MDS, leading to a severe and high-risk situation for MDS.

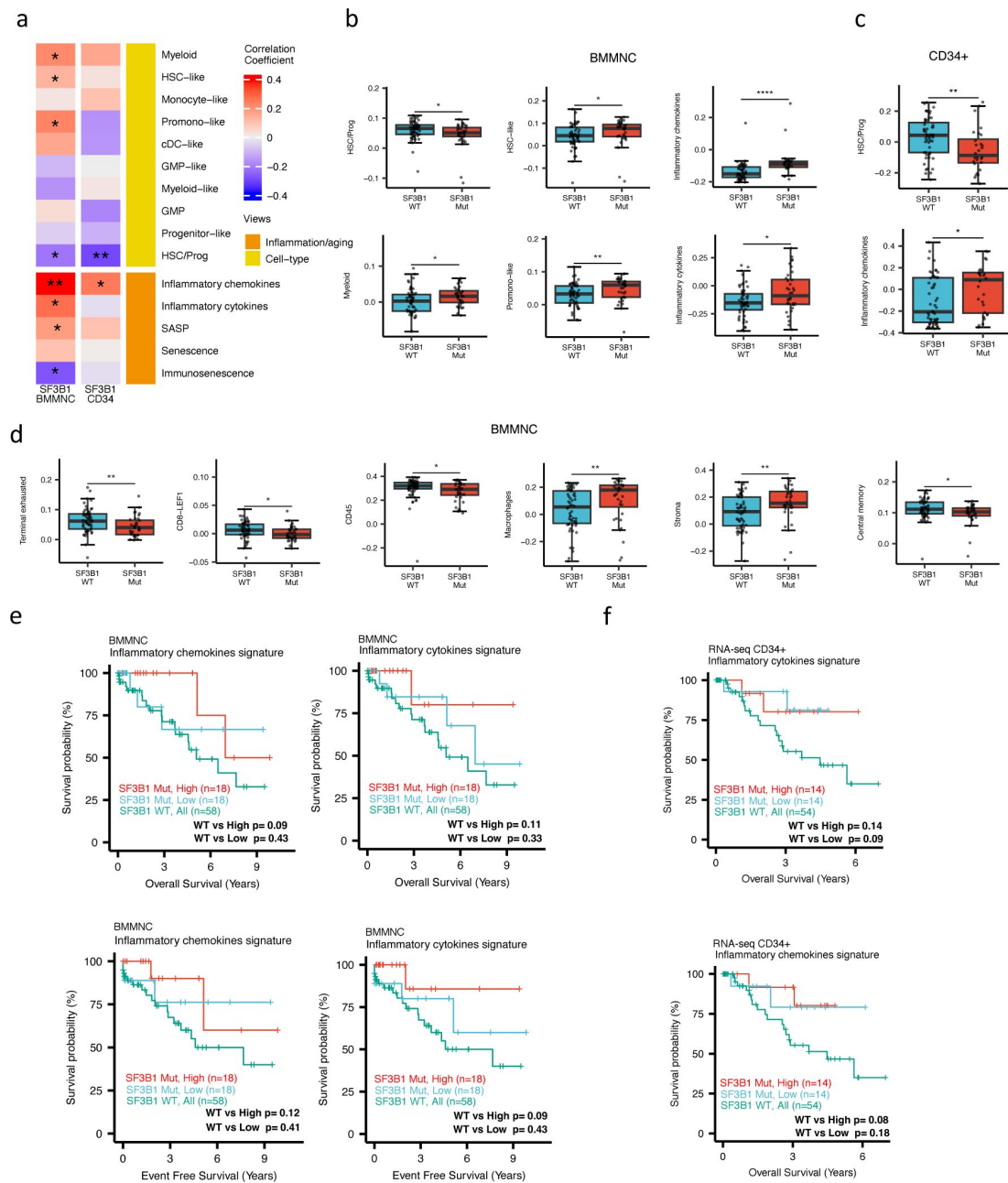


Figure 5.

Characterisation of SF3B1 mutant MDS using gene signatures from multiple biological views.

a, Association of a subset of inflammation/aging and cell-type features with SF3B1 mutation in the BMMNC and BM CD34+ cohorts, with red depicting a positive correlation and blue an inverse correlation with SF3B1 mutation. The significances were calculated with the Wilcoxon rank-sum test, and the significant associations were shown by * ($P < 0.05$), ** ($P < 0.01$) or *** ($P < 0.001$). **b-c**, Boxplots comparing the levels of the significant individual features from the cell-type and inflammation/aging biological views for SF3B1 mutant versus SF3B1 WT cases in the BMMNC and CD34+ cohorts. **d**, Boxplots comparing the levels of the significant individual features from the immune profile biological views for SF3B1 mutant versus SF3B1 WT cases in the BMMNC cohort. **e-f**, Kaplan-Meier plots displaying overall and event-free survivals for SF3B1 mutant cases split by high and low levels of inflammatory cytokines and chemokines versus SF3B1 mutant WT in the BMMNC and CD34+ cohorts. Event-free survival data was only available for the BMMNC cohort. All p-values were calculated using log-rank test on overall and event-free survival values of SF3B1 mutant high and low versus WT groups.

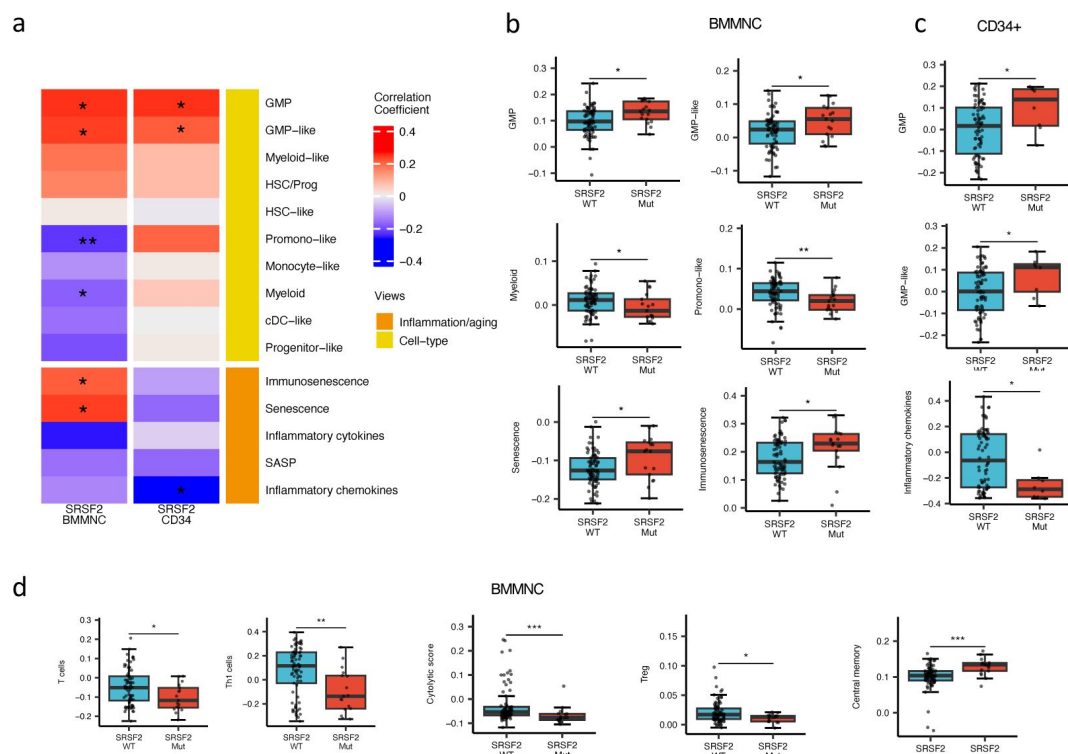


Figure 6.

SRSF2 mutant MDS is characterized by high GMP content and high levels of senescence and immunosenescence.

a, Association of a subset of inflammation/aging and cell-type features with SRSF2 mutation in the BMMNC and BM CD34+ cohorts, with red depicting a positive correlation and blue an inverse correlation with SRSF2 mutation. The significances were calculated with the Wilcoxon rank-sum test, and the significant associations were shown by * ($P < 0.05$), ** ($P < 0.01$) or *** ($P < 0.001$). **b-c**, Boxplots comparing the levels of the significant individual features from the cell-type and inflammation/aging biological views for SRSF2 mutant versus SRSF2 WT cases in the BMMNC and CD34+ cohorts, respectively. **d**, Boxplots comparing the levels of the significant individual features from the immune profile biological views for SRSF2 mutant versus SRSF2 WT cases in the BMMNC cohort.

Discussion

We applied MOFA to seven biological views derived from two MDS patient cohorts. MOFA could identify latent and important phenotypes from multimodal MDS data. Notably, we identified RTE expression as a risk factor and inflammation as a protective factor in MDS. Moreover, we uncovered that Factor 4 correlated with progress to AML, with patients lower in this factor more likely to develop AML. Low level of Factor 4 is related to a phenotype with low frequency of leukocytes; patients also typically had increased malignant HSCs and stroma. Our analysis shows that these patients probably have a defect in haematopoiesis, preventing the production of sufficient blood cells and ultimately allowing for the stroma to invade the whole marrow (36, 37). Although this observation is not novel, it is confirmatory of prior researches and demonstrates the power of our approach for integrating multimodal MDS data.

Literature has clearly defined the relationship between the SF3B1 mutation and long OS and EFS, with a low risk of progression to AML (23, 38, 39). We showed that SF3B1 mutant MDS cases tend to have high levels of inflammation, perhaps due to macrophage activation, and thus confers good prognosis for patients in terms of anti-tumour activity. Therefore, inflammation might help MDS survival. Further division of SF3B1 mutant patients into high and low levels of inflammation shows that the SF3B1 mutants with higher inflammation can generally survive better despite not being statistically significant. Investigating the synergic factors including epigenetic factors remains elusive. Enriching the multi-omics data with new modalities, including epigenomes, CyTOF, and Luminex cytokine/chemokine data in the future might help us understand why SF3B1 mutants show distinct patterns in terms of inflammation and survival.

Our work revealed that SRSF2 mutant MDS cases show a reduction of T cells. The decrease in the ability of patients to accumulate T cells, such as Th1 cells, which play crucial roles in modulating killing of tumour cells may cause worse outcomes for SRSF2 mutant patients (40).

We also observed increased immunosenescence level in SRSF2 mutants. Recent studies have shown that the expression of programmed death-ligand 1 (PD-L1) protein is significantly elevated in senescent cells (41–43). Increased PD-L1 protein levels protect senescent cells from being cleared by cytotoxic immune cells that express the PD-1 checkpoint receptor. In fact, activation of the PD-1 receptor inhibits the cytotoxic capabilities of CD8+ T and NK cells, increasing immunosenescence. Notably, patients with MDS who possess particular somatic mutations, such as those in the TP53, ASXL1, SETBP1, TET2, SRSF2, and RUNX1 genes, have an increased propensity to react favourably to PD-1/PD-L1 inhibitors (44) confirming that many cellular and molecular mechanisms, known to promote cellular senescence, including alteration of splicing machinery, are crucial stimulators of the expression of PD-L1 protein. Interestingly, in our analysis, we also observed a correlation between the senescence gene signature score and the expression of the PD-L1 gene in CD34+ cells (**Supplementary Figure 7**), supporting the previous findings linking PD-L1 gene expression to cellular senescence.

The immunology and ageing features extracted from the MDS transcriptomic data used in our analysis pipeline can enhance the conventional risk-scoring systems for MDS by providing new insights into this disease, particularly in the context of inflammation and ageing. For some patients, the clinical and genetic features may remain relatively the same until follow-up. Still, the transcriptomic features might differ considerably from the baseline diagnosis, affecting the course of treatment.

This study contributed to a deeper understanding of MDS pathogenesis and identified potential prognostic markers for this disease. It also elucidated the importance of considering the relationships between different pathways, markers, and mutations in predicting patient outcomes,

highlighting the efficacy of a comprehensive approach that goes beyond all the scoring systems that have been described thus far for MDS.

Acknowledgements

This study was supported with funding from Bristol Myers Squibb (BMS) company during this project. We thank Dr Sh. Kordasti for critical reading of the manuscript. AP and JB were supported by Blood Cancer UK (grants 13042 and 19004).

Authors' contributions

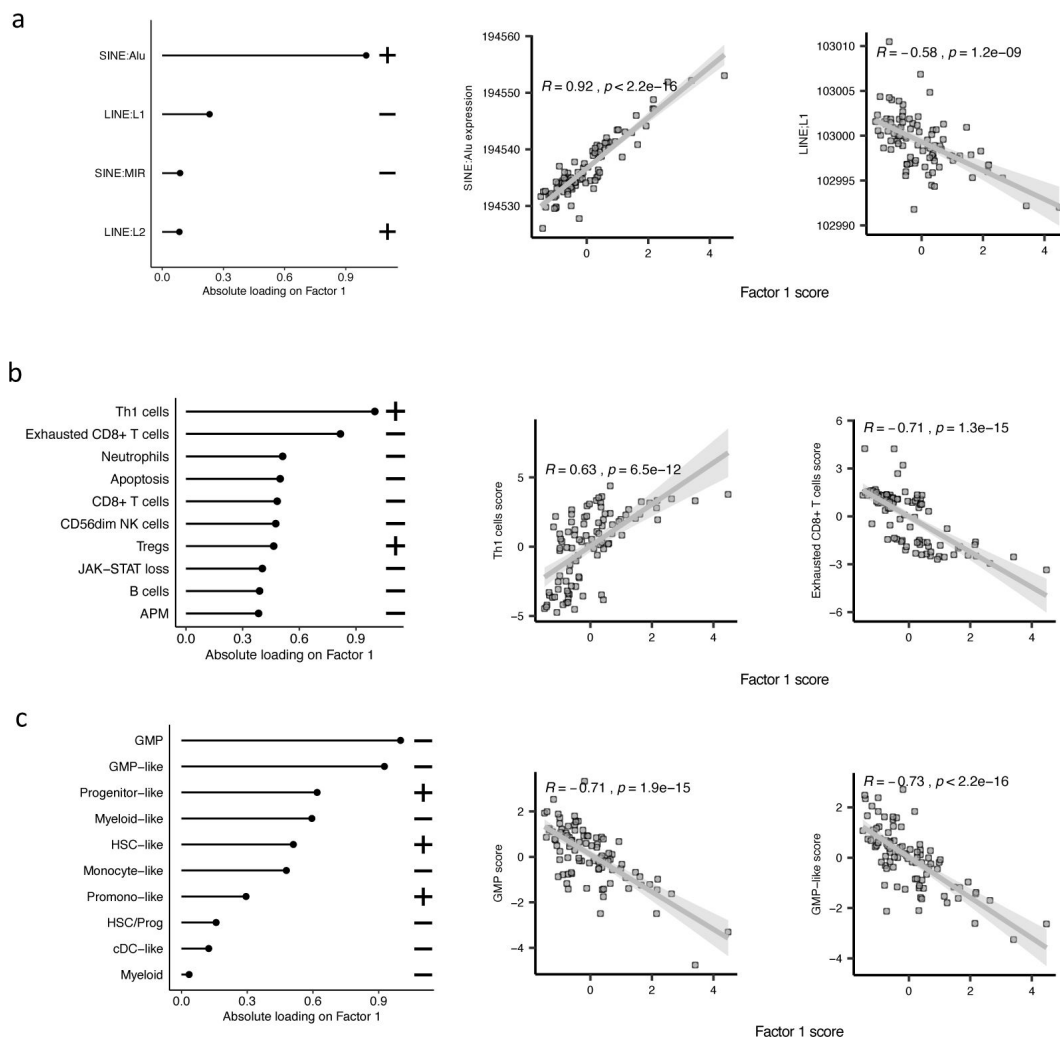
MMK and GJM conceived this analysis. The analysis was mainly performed by SG, and NS with contributions from WM, IRT, and DEK. MMK, SG, and NS designed the graphics. MMK, SG, SH, NS, and SRJ wrote the manuscript with input from all authors. GN, SO, AP, and JB contributed to the interpretation of the analytics. AL provided statistical support during the study. MMK and GJM supervised the work. All authors read and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

Data Availability

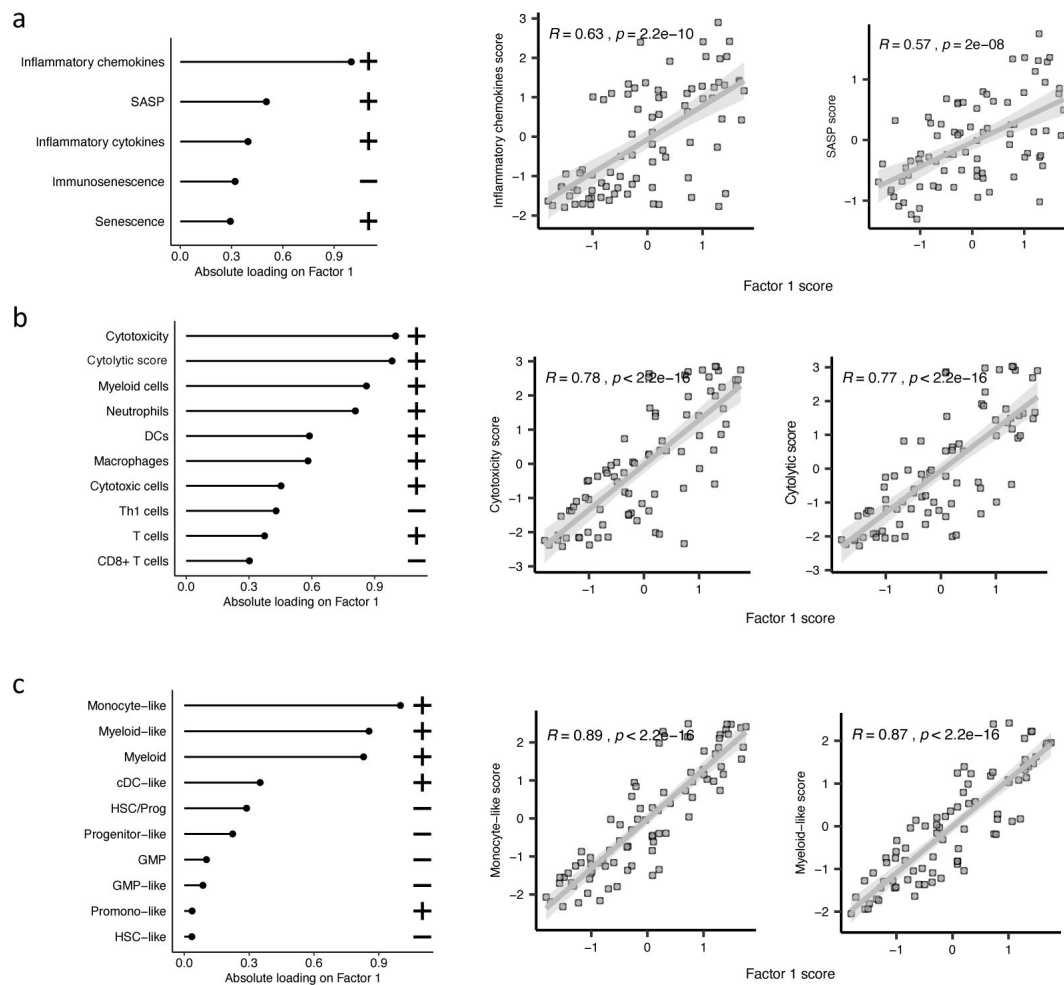
Scripts and data used in this study are available on Github (https://github.com/Karimi-Lab/MDS_MOFA )



Supplementary Figure 1.

The top features influencing Factor 1 form various biological views in the BMMNC cohort.

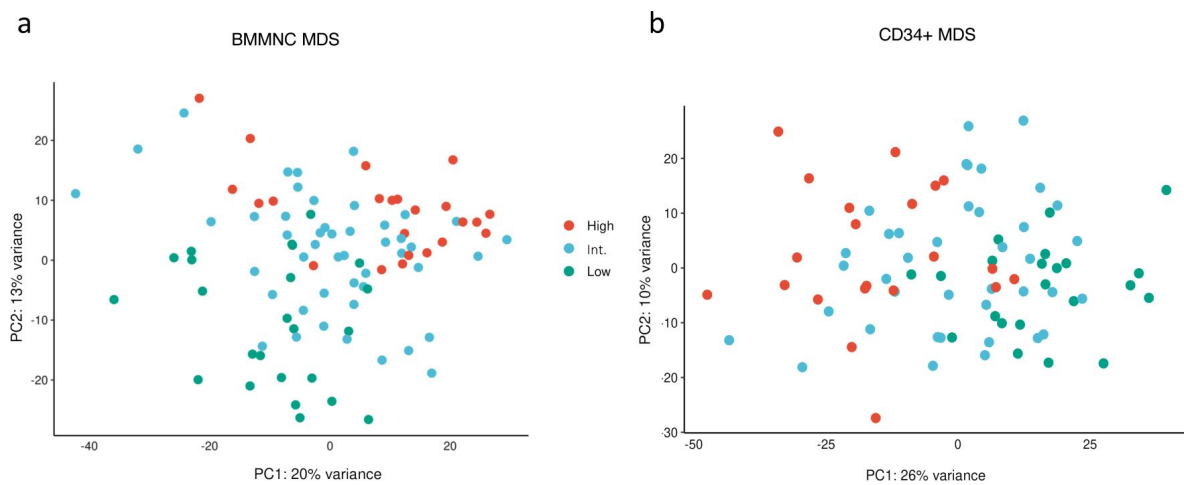
a-c, Absolute loadings of the top features and their positive or inverse correlations with Factor 1 from RTE expression, immune profile, and cell-type biological views, respectively. Pearson correlation coefficients (R) and p -values were displayed on top of the scatter plots.



Supplementary Figure 2.

The top features influencing Factor 1 form various biological views in the CD34+ cohort.

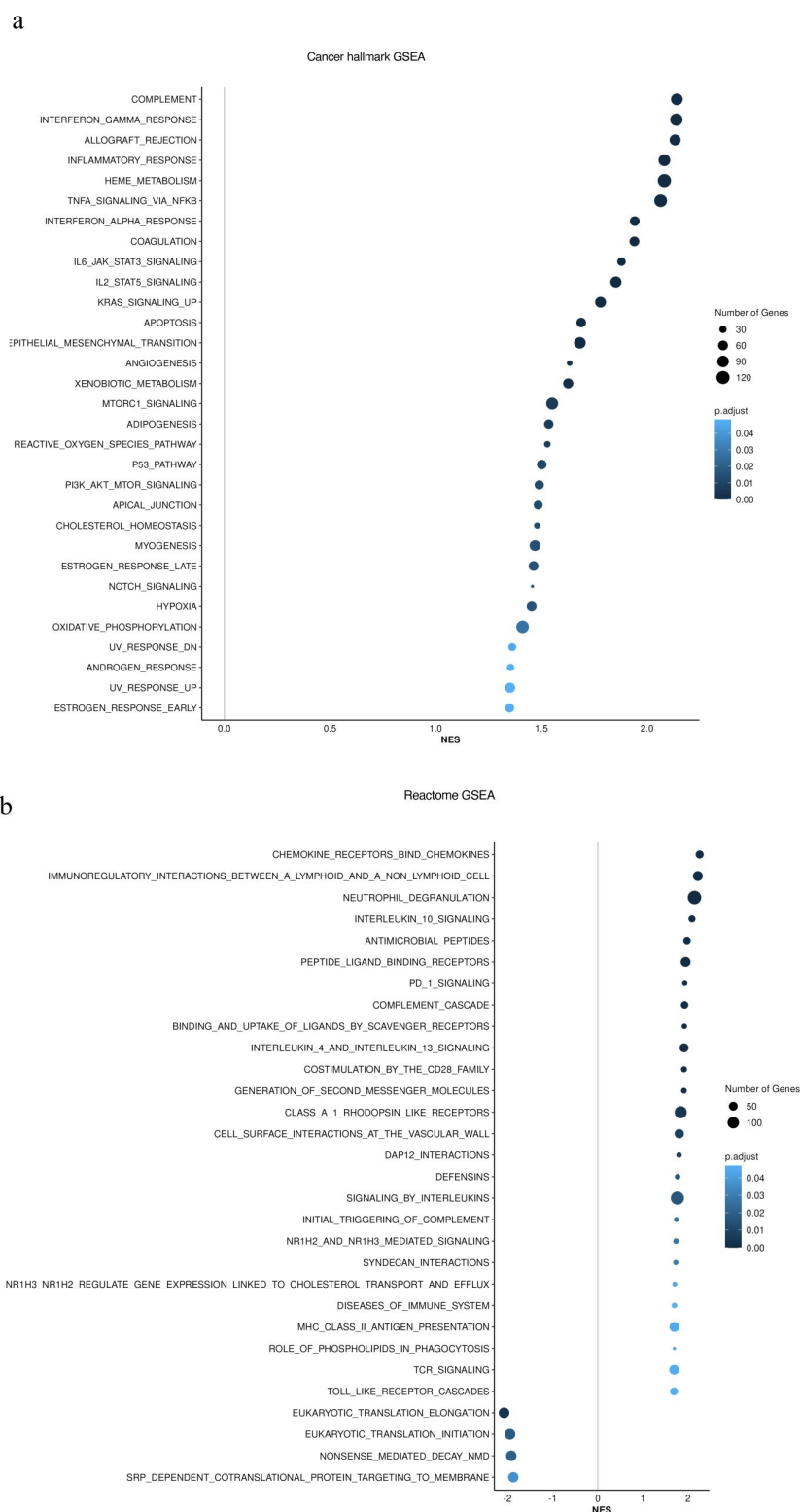
a-c, Absolute loadings of the top features and their positive or inverse correlations with Factor 1 from inflammation/aging, immune profile, and cell-type biological views, respectively. Pearson correlation coefficients (R) and p -values were displayed on top of the scatter plots.



Supplementary Figure 3.

Factor 1 from the BMMNC and CD34+ MDS cohorts stratified the patients in the gene expression PCA plots.

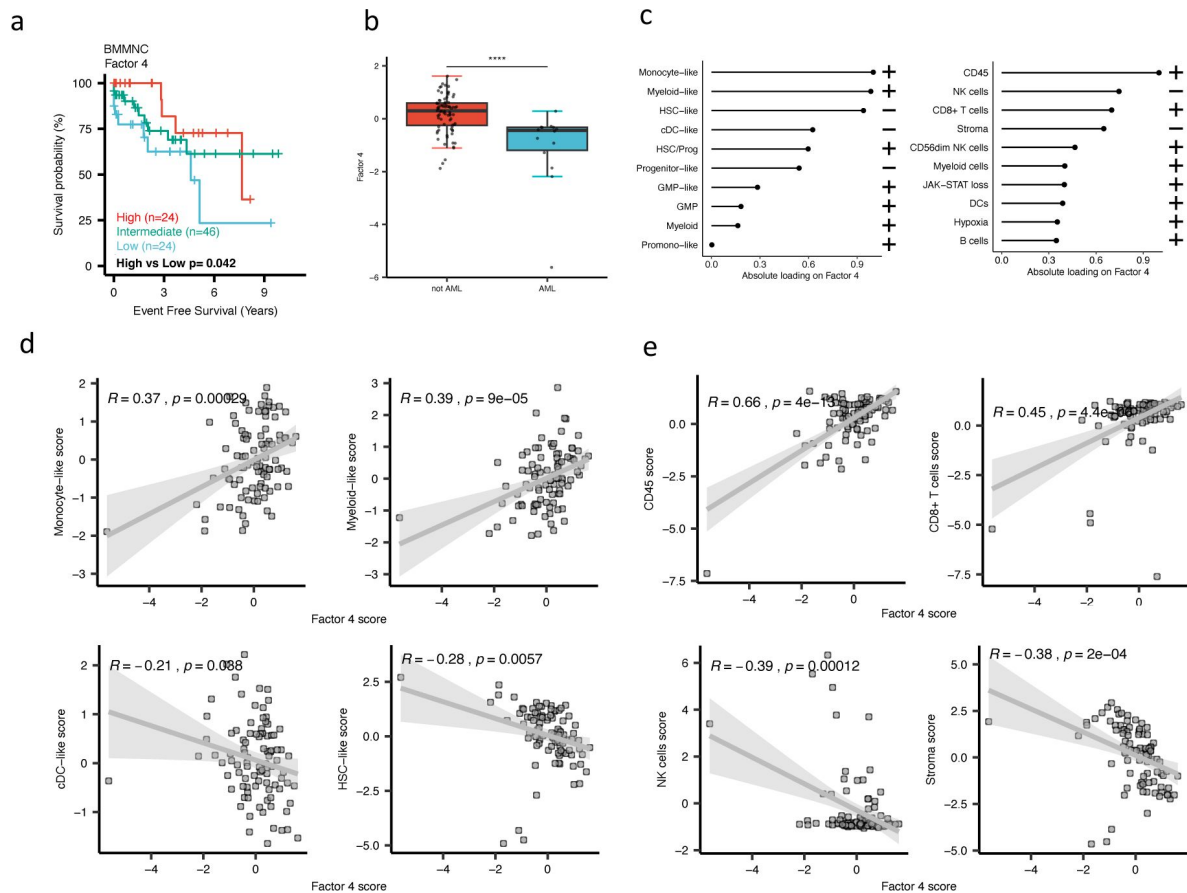
a-b, Principal component analysis on gene expression values of the BMMNC and CD34+ cohorts revealed separate clusters for patients split based on low (1st quartile), intermediate (2nd and 3rd quartile), and high (4th quartile) levels of Factor 1 in both cohorts. In both plots, the intermediate group is in between the high and low groups in the PCA plots.



Supplementary Figure 4.

Gene set enrichment analysis (GSEA) identified upregulation of inflammation related pathways in patients having high (4th quartile) versus low (1st quartile) levels of Factor 1 in the CD34+ cohort.

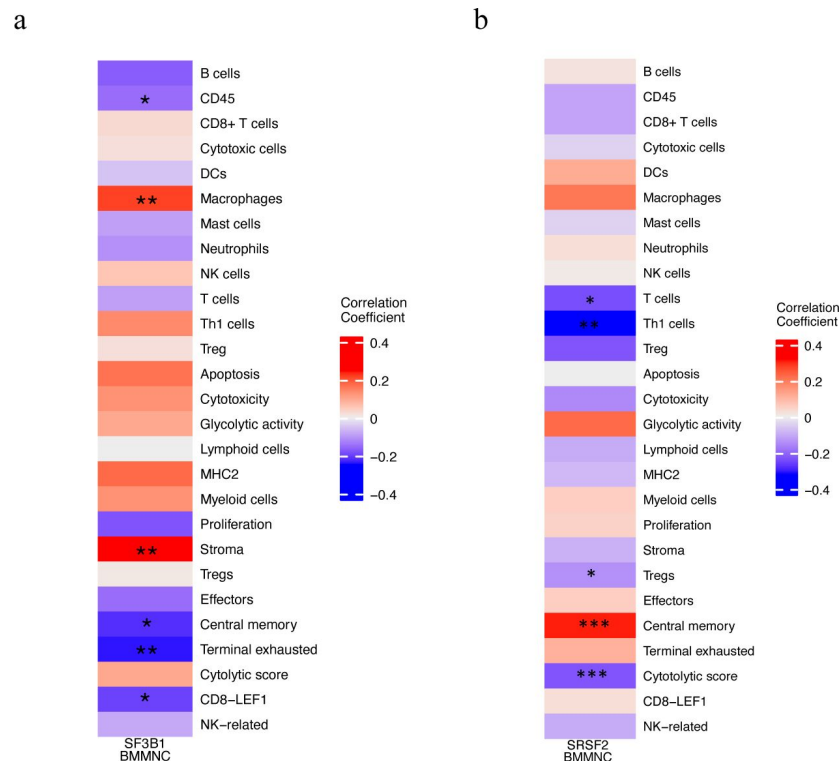
a, Upregulation of inflammatory, Interferons, TNFA, JAK-STAT signalling pathways within cancer hallmark gene sets. **b**, GSEA analysis using Reactome gene sets shows upregulation of gene sets associated with chemokines, Neutrophils deregulation and IL10 signalling.



Supplementary Figure 5.

Characterisation of Factor 4 in the BMMNC cohort using gene signatures from multiple biological views.

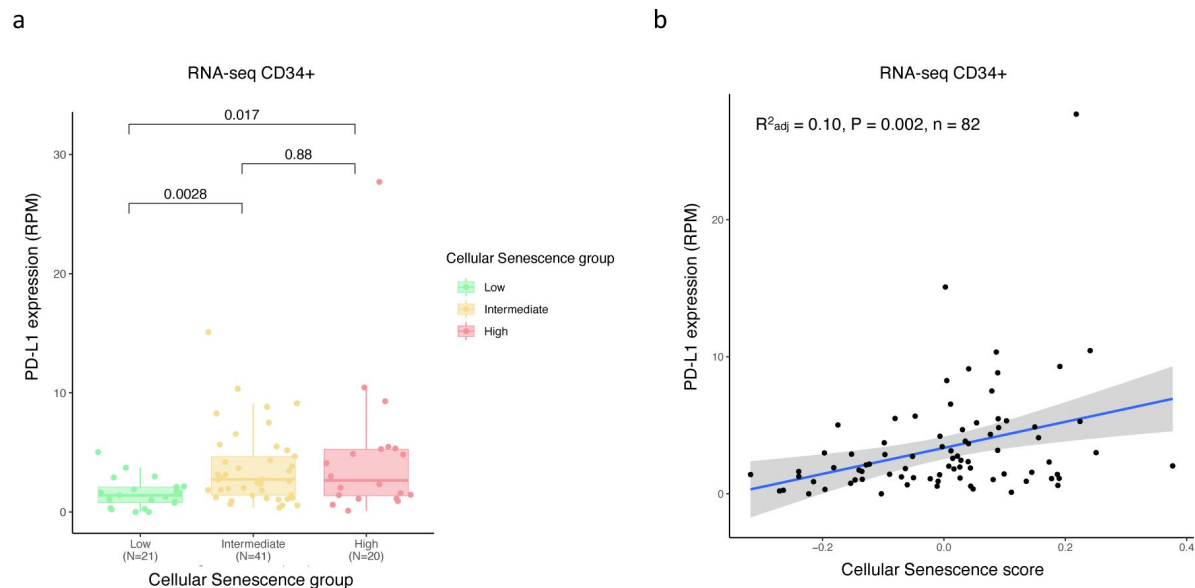
a, Kaplan-Meier plots for the BMMNC cohort where patients were split based on low (1st quartile), intermediate (2nd and 3rd quartile), and high (4th quartile) levels of Factor 4. The p-value was calculated using log-rank test on event-free survival values of high versus low groups. **b**, Boxplot depicting low level of Factor 4 for patients who progressed to AML. The significance was calculated with the Wilcoxon rank-sum test, and the significance was shown by **** (P < 0.0001). **c**, The absolute loading of the top features affecting Factor 4 in cell-type composition and immunology biological views in the BMMNC cohort. **d-e**, Positive or inverse correlations of selected top factors from the cell-type composition and immunology biological views with Factor 4. Pearson correlation coefficients (R) and p-values were displayed on top of the scatter plots.



Supplementary Figure 6.

Correlation of selected immune profile gene sets with SF3B1 and SRSF2 mutations in the MDS BMMNC cohort.

a-b, Correlation of a subset of immune profile features with SF3B1 and SRSF2 mutations in the BMMNC cohort, with red depicting a positive correlation and blue an inverse correlation with these mutations. The significances were calculated with the Wilcoxon rank-sum test, and the significant associations were shown by * ($P < 0.05$), ** ($P < 0.01$) or *** ($P < 0.001$).



Supplementary Figure 7.

Positive correlation between PD-L1 expression and senescence score in BM CD34+ RNA-seq cohort.

a, Boxplot showing higher expression of PD-L1 gene in patient groups with high (4th quartile) and intermediate (2nd and 3rd quartile) vs low (1st quartile) senescence levels. The significances were calculated with the Wilcoxon rank-sum. **b**, Scatter plot displaying a positive correlation between PD-L1 expression and senescence score.

References

1. Sperling AS, Gibson CJ, Ebert BL (2017) **The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia** *Nat Rev Cancer* **17**:5–19
2. Mufti GJ, Bennett JM, Goasguen J, Bain BJ, Baumann I, Brunning R, et al. (2008) **Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts** *Haematologica* **93**:1712–7
3. Zhang Y, Wu J, Qin T, Xu Z, Qu S, Pan L, et al. (2022) **Comparison of the revised 4th (2016) and 5th (2022) editions of the World Health Organization classification of myelodysplastic neoplasms** *Leukemia* **36**:2875–82
4. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. (2016) **The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia** *Blood* **127**:2391–405
5. Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF, et al. (2022) **The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms** *Leukemia* **36**:1703–19
6. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Sole F, et al. (2012) **Revised international prognostic scoring system for myelodysplastic syndromes** *Blood* **120**:2454–65
7. Zhao H, Wu L, Yan G, Chen Y, Zhou M, Wu Y, et al. (2021) **Inflammation and tumor progression: signaling pathways and targeted intervention** *Signal Transduct Target Ther* **6**
8. Ma Y, Adjemian S, Mattarollo SR, Yamazaki T, Aymeric L, Yang H, et al. (2013) **Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells** *Immunity* **38**:729–41
9. McLaughlin M, Patin EC, Pedersen M, Wilkins A, Dillon MT, Melcher AA, et al. (2020) **Inflammatory microenvironment remodelling by tumour cells after radiotherapy** *Nat Rev Cancer* **20**:203–17
10. Leonardi GC, Accardi G, Monastero R, Nicoletti F, Libra M (2018) **Ageing: from inflammation to cancer** *Immun Ageing* **15**
11. Weeks LD, Marinac CR, Redd R, Abel G, Lin A, Agrawal M, et al. (2022) **Age-related diseases of inflammation in myelodysplastic syndrome and chronic myelomonocytic leukemia** *Blood* **139**:1246–50
12. Anwar SL, Wulaningsih W, Lehmann U (2017) **Transposable Elements in Human Cancer: Causes and Consequences of Deregulation** *Int J Mol Sci* **18**
13. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, et al. (2010) **Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer** *PLoS Genet* **6**

14. Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, et al. (2014) **Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis** *Gut* **63**:635–46
15. Pellagatti A, Armstrong RN, Steeples V, Sharma E, Repapi E, Singh S, et al. (2018) **Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations** *Blood* **132**:1225–40
16. Pellagatti A, Cazzola M, Giagounidis A, Perry J, Malcovati L, Della Porta MG, et al. (2010) **Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells** *Leukemia* **24**:756–64
17. Shiozawa Y, Malcovati L, Galli A, Pellagatti A, Karimi M, Sato-Otsubo A, et al. (2017) **Gene expression and risk of leukemic transformation in myelodysplasia** *Blood* **130**:2642–53
18. Choudhary GS, Pellagatti A, Agianian B, Smith MA, Bhagat TD, Gordon-Mitchell S, et al. (2022) **Activation of targetable inflammatory immune signaling is seen in myelodysplastic syndromes with SF3B1 mutations** *Elife* **11**
19. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. (2014) **Landscape of genetic lesions in 944 patients with myelodysplastic syndromes** *Leukemia* **28**:241–7
20. Papaemmanuil E, Cazzola M, Boultonwood J, Malcovati L, Vyas P, Bowen D, et al. (2011) **Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts** *N Engl J Med* **365**:1384–95
21. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al., quiz 99 (2013) **Clinical and biological implications of driver mutations in myelodysplastic syndromes** *Blood* **122**:3616–27
22. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. (2011) **Frequent pathway mutations of splicing machinery in myelodysplasia** *Nature* **478**:64–9
23. Wu L, Song L, Xu L, Chang C, Xu F, Wu D, et al. (2016) **Genetic landscape of recurrent ASXL1, U2AF1, SF3B1, SRSF2, and EZH2 mutations in 304 Chinese patients with myelodysplastic syndromes** *Tumour Biol.* **37**:4633–40
24. Lee SC, North K, Kim E, Jang E, Obeng E, Lu SX, et al. (2018) **Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations** *Cancer Cell* **34**:225–41
25. Smith MA, Choudhary GS, Pellagatti A, Choi K, Bolanos LC, Bhagat TD, et al. (2019) **U2AF1 mutations induce oncogenic IRAK4 isoforms and activate innate immune pathways in myeloid malignancies** *Nat Cell Biol* **21**:640–50
26. Tuerxun N, Wang J, Zhao F, Qin YT, Wang H, Chen R, et al. (2022) **Bioinformatics analysis deciphering the transcriptomic signatures associated with signalling pathways and prognosis in the myelodysplastic syndromes** *Hematology* **27**:214–31
27. van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, et al. (2019) **Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity** *Cell* **176**:1265–81
28. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ (2018) **Single sample scoring of molecular phenotypes** *BMC Bioinformatics* **19**

29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. (2015) **limma powers differential expression analyses for RNA-sequencing and microarray studies** *Nucleic Acids Res* **43**
30. Love MI, Huber W, Anders S (2014) **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2** *Genome Biol* **15**
31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. (2005) **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles** *Proc Natl Acad Sci U S A* **102**:15545–50
32. Pollyea DA, Kim HM, Stevens BM, Lee FF, Harris C, Hedin BR, et al. (2021) **MDS-associated SF3B1 mutations enhance proinflammatory gene expression in patient blast cells** *J Leukoc Biol* **110**:197–205
33. Pollyea DA, Harris C, Rabe JL, Hedin BR, De Arras L, Katz S, et al. (2019) **Myelodysplastic syndrome-associated spliceosome gene mutations enhance innate immune signaling** *Haematologica* **104**:e388–e92
34. Pang WW, Pluvineau JV, Price EA, Sridhar K, Arber DA, Greenberg PL, et al. (2013) **Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes** *Proc Natl Acad Sci U S A* **110**:3011–6
35. Will B, Zhou L, Vogler TO, Ben-Neriah S, Schinke C, Tamari R, et al. (2012) **Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations** *Blood* **120**:2076–86
36. Tripodo C, Sangaletti S, Piccaluga PP, Prakash S, Franco G, Borrello I, et al. (2011) **The bone marrow stroma in hematological neoplasms--a guilty bystander** *Nat Rev Clin Oncol* **8**:456–66
37. Vega F, Medeiros LJ, Lang WH, Mansoor A, Bueso-Ramos C, Jones D (2002) **The stromal composition of malignant lymphoid aggregates in bone marrow: variations in architecture and phenotype in different B-cell tumours** *Br J Haematol* **117**:569–76
38. Migdady Y, Barnard J, Al Ali N, Steensma DP, DeZern A, Roboz G, et al. (2018) **Clinical Outcomes With Ring Sideroblasts and SF3B1 Mutations in Myelodysplastic Syndromes: MDS Clinical Research Consortium Analysis** *Clin Lymphoma Myeloma Leuk* **18**:528–32
39. Tang Y, Miao M, Han S, Qi J, Wang H, Ruan C, et al. (2019) **Prognostic value and clinical feature of SF3B1 mutations in myelodysplastic syndromes: A meta-analysis** *Crit Rev Oncol Hematol* **133**:74–83
40. Knutson KL, Disis ML (2005) **Tumor antigen-specific T helper cells in cancer immunity and immunotherapy** *Cancer Immunol Immunother* **54**:721–8
41. Wang TW, Johmura Y, Suzuki N, Omori S, Migita T, Yamaguchi K, et al. (2022) **Blocking PD-L1-PD-1 improves senescence surveillance and ageing phenotypes** *Nature* **611**:358–64
42. Pippin JW, Kaverina N, Wang Y, Eng DG, Zeng Y, Tran U, et al. (2022) **Upregulated PD-1 signaling antagonizes glomerular health in aged kidneys and disease** *J Clin Invest* **132**
43. Onorati A, Havas AP, Lin B, Rajagopal J, Sen P, Adams PD, et al. (2022) **Upregulation of PD-L1 in Senescence and Aging** *Mol Cell Biol* **42**

44. Chien KS, Kim K, Nogueras-Gonzalez GM, Borthakur G, Naqvi K, Daver NG, et al. (2021) **Phase II study of azacitidine with pembrolizumab in patients with intermediate-1 or higher-risk myelodysplastic syndrome** *Br J Haematol* **195**:378–87

Editors

Reviewing Editor

Satyajit Rath

Indian Institute of Science Education and Research (IISER), Pune, India

Senior Editor

Satyajit Rath

Indian Institute of Science Education and Research (IISER), Pune, India

Reviewer #2 (Public Review):

Summary:

The authors performed a Multi-Omics Factor Analysis (MOFA) on analysis of two published MDS patient cohorts-1 from bone marrow mononuclear cells (BMMNCs) and CD34 cells (ref 17) and another from CD34+ cells (ref 15) –with three data modalities (clinical, genotype, and transcriptomics). Seven different views, including immune profile, inflammation/aging, Retrotransposon (RTE) expression, and cell- type composition, were derived from these modalities to attempt to identify the latent factors with significant impact on MDS prognosis.

SF3B1 was found to be the only mutation among 13 mutations in the BMMNC cohort that indicated a significant association with high inflammation. This trend was also observed to a lesser extent in the CD34+ cohort. The MOFA factor representing inflammation showed a good prognosis for MDS patients with high inflammation. In contrast, SRSF2 mutant cases showed a granulocyte-monocyte progenitor (GMP) pattern and high levels of senescence, immunosenescence, and malignant myeloid cells, consistent with their poor prognosis. Also, MOFA identified RTE expression as a risk factor for MDS. They proposed that this work showed the efficacy of their integrative approach to assess MDS prognostic risk that 'goes beyond all the scoring systems described thus far for MDS'.

<https://doi.org/10.7554/eLife.97096.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public Review):

In their manuscript, Gerlevik et al. performed an integrative analysis of clinical, genetic and transcriptomic data to identify MDS subgroups with distinct outcomes. The study was based on the building of an "immunoscore" and then combined with genotype and clinical data to analyze patient outcomes using multi-omics factor analysis.

Strengths: Integrative analysis of RNA-seq, genotyping and clinical data

Weaknesses: Validation of the bioinformatic pipeline is incomplete

Major comments:

(1) This study considered two RNA-seq data sets publicly available and generated in two distinct laboratories. Are they comparable in terms of RNA-seq technique: polyA versus rRNA depletion, paired-end sequencing, fragment length?

We want to reemphasize that the main point of this study is not to compare the BMMNC with the HSPC cohort. These datasets are not comparable because they were

collected from different cell types, and we should not expect them to be matched. We just analysed them in parallel to check how much HSPCs contribute to the molecular signatures we see in BMMNC samples. However, we agree with the reviewer that similar RNA-seq experimental techniques should be employed to control for confounding factors. Here is the information that we found for HSPC and BMMNC RNA-seq studies:

HSPC RNA-seq cohort: Total RNA was extracted using TRIzol (Thermo Scientific), and Sequencing was performed on an Illumina HiSeq4000 with 100-bp paired-end reads.

BMMNC RNA-seq cohort: The RNA was extracted with TRIzol reagent (Thermo Scientific). RNA-sequencing libraries were prepared from poly(A)-selected RNA and were sequenced using Illumina HiSeq 2000 or 2500 platform with 100-bp paired-end reads.

The only difference between the two cohorts is that one cohort includes total RNAs, whereas the other has polyA-selected RNAs. Since the gene set signatures use the expression of protein-coding genes, which all have polyA tails and are included in total RNA libraries, the analysis will not be affected by total vs. polyA-selected RNA-seq techniques.

(2) Data quality control (figure 1): the authors must show in a graph whether the features (dimensions) of factor 1 were available for each BMMNC and CD34+ samples.

By features of Factor 1, we think the reviewer means the features with high weights for Factor 1 in BMMNC and CD34+ samples. Figure 2c-d clearly illustrates the important features and their associations with Factor 1 for all samples in both cohorts. The samples are the columns of the two heatmaps.

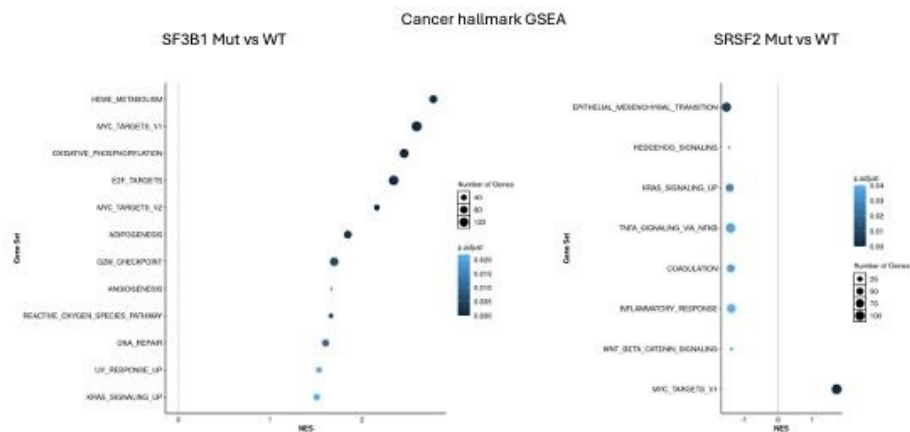
(3) How to validate the importance of "immunoscore"? If GSEA of RNA-seq data was performed in the entire cohort, in the SF3B1-mutated samples or SRSF2-mutated samples (instead of patients having a high versus low level of factor 1 shown in Sup Fig. 4), what would be the ranking of Hallmarks or Reactome inflammatory terms among the others?

Our GSEA analysis was an attempt to validate the importance of our identified factors. As described in the paper, Factor 1 represents a combination of immunology scores (or "immunoscores") in CD34+ cohort. Applying GSEA, we identified upregulation of inflammation related pathways, chemokines, and Neutrophils in patients having high (4th quartile) versus low (1st quartile) levels of Factor 1. Interestingly, sorting patients by Factor 1 resulted in similar pattern based on gene signature scores (Figure 2d).

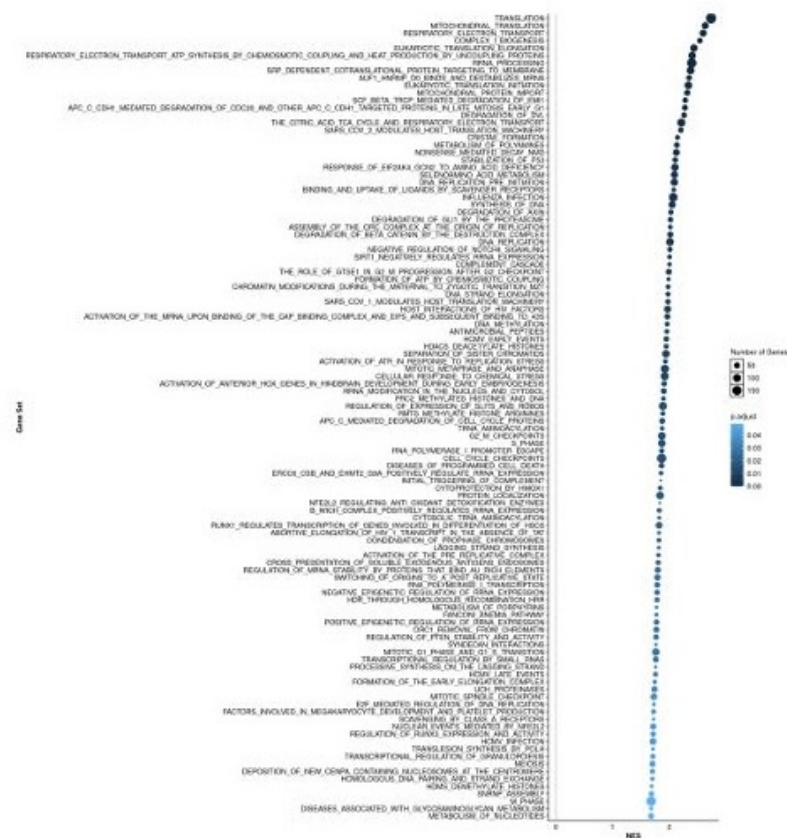
To show that Factor1 generated by MOFA is important and different from known MDS categories such as SF3B1 and SRSF2 mutants, we performed GSEA in SF3B1-mutated vs. SF3B1-WT samples and SRSF2-mutated vs. SRSF2-WT samples in the CD34+ cohort. As shown in Author response image 1, we did not see the upregulation of inflammation and interferon pathways in SF3B1 and SRSF2 mutant MDS.

Author response image 1.

GSEA showed no upregulation of inflammation and interferon pathways for SF3B1 and SRSF2 mutant in CD34+ cohort.



Reactome GSEA - SF3B1 Mut vs WT



Reactome GSEA - SRSF2 Mut vs WT



(4) To decipher cell-type composition of BMMNC and CD34+ samples, the authors used van Galen's data (2019; supplementary table 3). Cell composition is expressed as the proportion of each cell population among the others. Surprisingly, the authors found that the promonocytelike score was increased in SF3B1-mutated samples and not in SRSF2-mutated samples, which are frequently co-mutated with TET2 and associated with a CMML-like phenotype. Is there a risk of bias if bone marrow subpopulations such as megakaryocytic-erythroid progenitors or early erythroid precursors are not considered?

We thank the reviewer for their insightful comment about CMML and the high prevalence of SRSF2 mutation (> 45%) in CMML cases. Using single-cell RNA sequencing and high-parameter flow cytometry, Ferrall-Fairbanks et al. (DOI: 10.1158/2643-3230.BCD-21-0217) recently showed that CMML can be classified into three differentiation trajectories: monocytic, megakaryocyte-erythroid progenitor (MEP), and normal-like. One hallmark of monocytic-biased trajectory was the enrichment of inflammatory granulocyte-macrophage progenitor (GMP)-like cells, which we observed through our analysis for SRSF2 mutants (Figure 6a).

Unfortunately, van Galen's data does not provide any gene set for MEP, and there is no singlecell RNA-seq atlas for MDS to employ to calculate the MEP score. Also, we compared the Promono-like and GMP-like gene sets from van Galen's data, and we could not find any overlap, meaning that Promono-like is not specific enough to capture the signatures coming from the more differentiated progenitors such as GMPs. Therefore, as described in the paper, we focused on GMP-like rather than Promono-like.

(5) Figures 2a and 2b indicated that the nature of retrotransposons identified in BMMNC and CD34+ was different. ERVs were not detected in CD34+ cells. Are ERVs not reactivated in CD34+ cells? Is there a bias in the sequencing or bioinformatic method?

As described above, the two cohorts' sequencing methods, read length, etc., are identical.

CD34+ RNA-seq is total RNA-seq that includes both polyA and non-polyA RTE transcripts.

Therefore, the chance of bias and missing RTE signatures in CD34+ cohort is very low. L1 and Alu, which are shared between the two cohorts, are the two RTE families that are still active and make new insertions in humans. Our interpretation is that ERV activation in BM is associated with immune cells. As shown by Au et al. (DOI: 10.1016/j.ccell.2021.10.001), several ERV loci had expression in purified immune cell subsets in renal cell carcinoma samples, potentially explaining ERV upregulation in tumours responding to treatment as those biopsies had increased tumour infiltration.

(6) What is the impact of factor 1 on survival? Is it different between BMMNC and CD34+ cells considering the distinct composition of factor 1 in CD34+ and BMMNC?

As shown in Table 1, Factor 1 in the BMMNC cohort is associated with overall survival (P-val < 0.05) when we did multivariate analysis but not univariate analysis. We did not observe any association between Factor 1 and event-free survival in the BMMNC cohort. Also, The 10 factors identified by MOFA in BM CD34+ cohort did not show any significance associated with MDS overall survival (Supplementary Table 5).

(7) In Figure 1e, genotype contributed to the variance of in the CD34+ cell analyses more importantly than in the BMMNC. Because the patients are different in the two cohorts, differences in the variance could be explained either by a greater variability of the type of mutations in CD34 or an increased frequency of poor prognosis mutations in CD34+ compared to BMMNC. The genotyping data must be shown.

The genotype has already been reported in Supplementary Table 2. In fact, the number of inspected genes was much higher in the BMMNC cohort (17 genes) compared to the CD34+ cohort (3 genes). Therefore, we have more significant variability of the type of mutations in the BMMNC cohort compared to the CD34+ cohort. For the CD34+ cohort, we only had mutations for three spliceosome genes, where most cases (n=28) were SF3B1 mutants with good prognosis. We think that the result makes sense because the less genetic variability, the more homogenous groups and the more chance that one factor or a group of factors can explain the genetic variance.

(8) Fig. 2a-b: Features with high weight are shown for each factor. For factor 9, features seemed to have a low weight (Fig. 1b and 1c). However, factor 9 was predictive of EFS and OS in the BMMNC cohort. What are the features driving the prognostic value of factor 9?

As shown in Figure 3b, The main features are RTE expression from LTR:ERV1, SINE:MIR, and SINE:Alu family.

(9) The authors also provided microarray analyses of CD34+ cell. It could be interesting to test more broadly the correlation between features identified by RNA-seq or microarrays.

The microarray data did not come with any genetic information or clinical data except survival information. Therefore, we could not apply MOFA on Microarray data. However, we did generate gene signature scores from Microarray data and investigated the relationship between inflammatory chemokines and cytokines, and IFN-I signature scores with MDS survival (Figure 3c and 4c).

(10) The authors should discuss the relevance of immunosenescence features in the context of SRSF2 mutation and extend the discussion to the interest of their pipeline for patient diagnosis and follow up under treatments.

We have added the below text to the discussion:

Recent studies have shown that the expression of programmed death-ligand 1 (PD-L1) protein is significantly elevated in senescent cells (DOIs: 10.1128/mcb.00171-22, 10.1172/JCI156250, 10.1038/s41586-022-05388-4). Increased PD-L1 protein levels protect senescent cells from being cleared by cytotoxic immune cells that express the PD-1 checkpoint receptor. In fact, activation of the PD-1 receptor inhibits the cytotoxic capabilities of CD8 + T and NK cells, increasing immunosenescence.

Notably, patients with MDS who possess particular somatic mutations, such as those in the TP53, ASXL1, SETBP1, TET2, SRSF2, and RUNX1 genes, have an increased propensity to react favourably to PD-1/PD-L1 inhibitors (DOIs: 10.1111/bjh.17689, <https://doi.org/10.1182/blood2020-141100>) confirming that many cellular and molecular mechanisms, known to promote cellular senescence, including alteration of splicing machinery, are crucial stimulators of the expression of PD-L1 protein. Interestingly, in our analysis, we also observed a correlation between the senescence gene signature score and the expression of the PD-L1 gene in CD34+ cells (Supplementary Figure 7), supporting the previous findings linking PD-L1 gene expression to cellular senescence.

The immunology and ageing features extracted from the MDS transcriptomic data used in our analysis pipeline can enhance the conventional risk-scoring systems for MDS by providing new insights into this disease, particularly in the context of inflammation and ageing. For some patients, the clinical and genetic features may remain relatively the same

until follow-up. Still, the transcriptomic features might differ considerably from the baseline diagnosis, affecting the course of treatment.

Reviewer #2 (Public Review):

The authors performed a Multi-Omics Factor Analysis (MOFA) on analysis of two published MDS patient cohorts-1 from bone marrow mononuclear cells (BMMNCs) and CD34 cells (ref 17) and another from CD34+ cells (ref 15) --with three data modalities (clinical, genotype, and transcriptomics). Seven different views, including immune profile, inflammation/aging, Retrotransposon (RTE) expression, and cell-type composition, were derived from these modalities to attempt to identify the latent factors with significant impact on MDS prognosis.

SF3B1 was found to be the only mutation among 13 mutations in the BMMNC cohort that indicated a significant association with high inflammation. This trend was also observed to a lesser extent in the CD34+ cohort. The MOFA factor representing inflammation showed a good prognosis for MDS patients with high inflammation. In contrast, SRSF2 mutant cases showed a granulocyte-monocyte progenitor (GMP) pattern and high levels of senescence, immunosenescence, and malignant myeloid cells, consistent with their poor prognosis. Also, MOFA identified RTE expression as a risk factor for MDS. They proposed that this work showed the efficacy of their integrative approach to assess MDS prognostic risk that 'goes beyond all the scoring systems described thus far for MDS'.

Several issues need clarification and response:

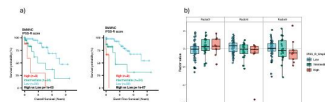
(1) The authors do not provide adequate known clinical and molecular information which demonstrates prognostic risk of their sample cohorts in order to determine whether their data and approach 'goes 'beyond all the scoring systems described thus far for MDS'. For example, what data have the authors that their features provide prognostic data independent of the prior known factors related to prognosis (eg, marrow blasts, mutational, cytogenetic features, ring sideroblasts, IPSS-R, IPSS-M, MDA-SS)?

We agree with the reviewer that we did not generate a new cumulative risk score and compare it with the conventional risk scores for MDS. However, we identified individual MOFA factors, which are risk or protective factors for MDS, based on survival analysis in the BMMNC cohort. One reason that we did not generate our independent, cumulative score and compare it with other scores was that we did not receive any conventional risk score for the BMMNC cohort. However, we had access to all the clinical and genetic variables from the BMMNC cohort (except for three patients) that were required to calculate IPSS-R; hence, we calculated the IPSS-R in our resubmission for the BMMNC cohort. We made three IPSS-R risk categories by combining low and very low as low risk, and high and very high as high risk, and keeping intermediate as intermediate risk. Our survival analysis of these three categories showed a clear match between IPSS-R score and MDS survival (Author response image 2a).

We then investigated the relationship between factors 2, 4, and 9 from MOFA with three IPSS-R risk groups. Integration of IPSS-R risk groups with factor values confirmed the finding in the manuscript that Factors 4 and 9 generally exert a protective influence over the MDS risk, whilst higher levels of Factor 2 predict a high-risk MDS (Author response image 2b). However, we see so many outliers in all three factors, indicating that some patients were assigned to the wrong IPSS-R categories because IPSS-R calculation is based on clinical and genetic variables and does not include the transcriptomics data for coding and non-coding genomic regions.

Author response image 2.

Comparison of IPSS-R risk categories and MOFA risk and protective factors.

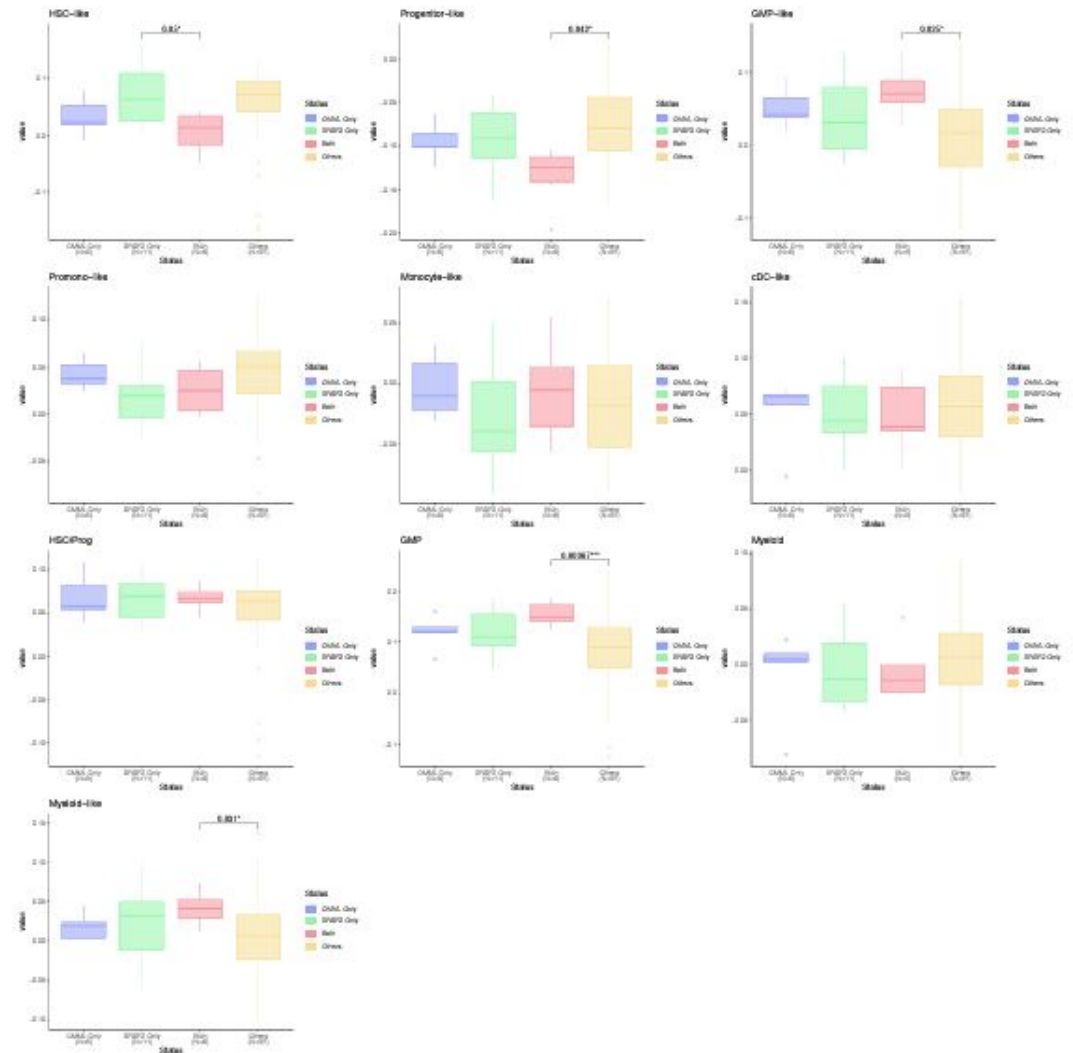


(2) A major issue in analyzing this paper relates to the specific patient composition from whom the samples and data were obtained. The cells from the Shiozawa paper (ref 17) is comprised of a substantial number of CMML patients. Thus, what evidence have the authors that much of the data from the BMMNCs from these patients and mutant SRSF2 related predominantly to their monocytic differentiation state?

We thank the reviewer for the insightful comment about the monocytic differentiation state of CMML and SRSF2 mutant cases. The BMMNC cohort has 11 CMML and 17 SRSF2 mutant cases, of which six are shared between the two groups. We have divided the patients into four groups: CMML only, SRSF2 mutant only, CCML and SRSF2 mutant, and others. We have generated boxplots for all cellular composition gene signature scores for these groups and compared the scores between these groups. As explained above, Ferrall-Fairbanks et al. (DOI: 10.1158/2643-3230.BCD-21-0217) recently showed that CMML can be classified into three differentiation trajectories: monocytic, megakaryocyte-erythroid progenitor (MEP), and normal-like. One hallmark of monocytic-biased trajectory was the enrichment of inflammatory granulocyte-macrophage progenitor (GMP)-like cells, which we observed through our analysis for the CMML cases with SRSF2 mutation (Author response image 3.).

Author response image 3.

Cellular composition gene signature scores for CMML and SRSF2 mutant versus other cases. CMML cases with SRSF2 mutation show a significant higher level of GMP and GMP-like scores compared to other MDS cases.



(3) In addition, as the majority of patients in the Shiozawa paper have ring sideroblasts ($n=59$), thus potentially skewing the data toward consideration mainly of these patients, for whom better outcomes are well known.

We disagree with the reviewer. We used 94 BMMNC samples from Shiozawa's paper, of which 19 cases had Refractory Anemia with Ring Sideroblasts (RARS), 4 cases had Refractory Anemia with Ring Sideroblasts and thrombocytosis (RARS-T), and 5 cases had Refractory cytopenia with multilineage dysplasia and ring sideroblasts (RCMD-RS). In total, we had 28 cases (~30%) with Ring Sideroblasts (RS), which are not large enough to skew the data.

(4) Further, regarding this patient subset, what evidence have the authors that the importance of the SF3B1 mutation was merely related to the preponderance of sideroblastic patients from whom the samples were analyzed?

We had 34 SF3B1 mutant cases, of which 25 had Ring Sideroblasts (RS). The total number of cases with RS in the BMMNC cohort was 28. Therefore, the BMMNC cohort is not an RSdominant cohort, and RS cases did not include all SF3B1 mutants. Furthermore, it was recently shown by Ochi et al. (DOI: 10.1038/s41598-022-18921-2) that RS is a consequence of SF3B1K700E mutation, and it is not a cause to affect the SF3B1 importance.

(5) An Erratum was reported for the Shiozawa paper (Shiozawa Y, Malcovati L, Galli A, et al. Gene expression and risk of leukemic transformation in myelodysplasia. *Blood*. 2018 Aug 23;132(8):869-875. doi: 10.1182/blood-2018-07-863134) that resulted from a coding error in the construction of the logistic regression model for subgroup prediction based on the gene expression profiles of BMMNCs. This coding error was identified after the publication of the article. The authors should indicate the effect this error may have had on the data they now report.

Thank you for bringing this important issue to our attention. The error resulted from a mistake in the construction of the logistic regression model for subgroup prediction based on the gene expression profiles of BMMNCs. However, this issue does not affect our result because we analysed the expression data from scratch and generated our own gene signature scores. Also, the error has no impact on the genetics and clinical information that we received from the authors.

(6) What information have the authors as to whether the dicing RTE findings were not predominantly related to the differentiation state of the cell population analyzed (ie higher in BM MNCs vs CD34, Fig 1)? What control data have the authors regarding these values from normal (non-malignant) cell populations?

As described above, L1 and Alu, the two RTE families shared between the two cohorts, are still active and make new insertions in humans (Figure 2.a-b). Our interpretation is that ERV activation in BM is associated with immune cells. This interpretation is further supported by the findings of Au et al. (DOI: 10.1016/j.ccell.2021.10.001), where several ERV loci had expression in purified immune cell subsets in renal cell carcinoma samples.

Unfortunately, none of these two cohorts had normal (non-malignant) cell populations. We think that the MOFA unbiased way of modelling the heterogeneity is sufficient to capture the RTE derepressed phenotype of a subset of MDS cases compared to others, and we do not need normal cases to further support the finding.

(7) The statement in the Discussion regarding the effects of SRSF2 mutation is speculative and should be avoided. Many other somatic gene mutations have known stronger effects on prognosis for MDS.

One aim of this study is to identify specific immune signatures associated with SRSF2 and SF3B1 mutations, which are highly prevalent in MDS. Although other mutations, such as TP53, may have a stronger correlation with poor survival, numerous studies have demonstrated a clear link between SRSF2 mutations and poor prognosis.

<https://doi.org/10.7554/eLife.97096.2.sa0>