
Figures and figure supplements

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

Sila Gerlevik, Nogayhan Seymen and Shan Hama *et al.*

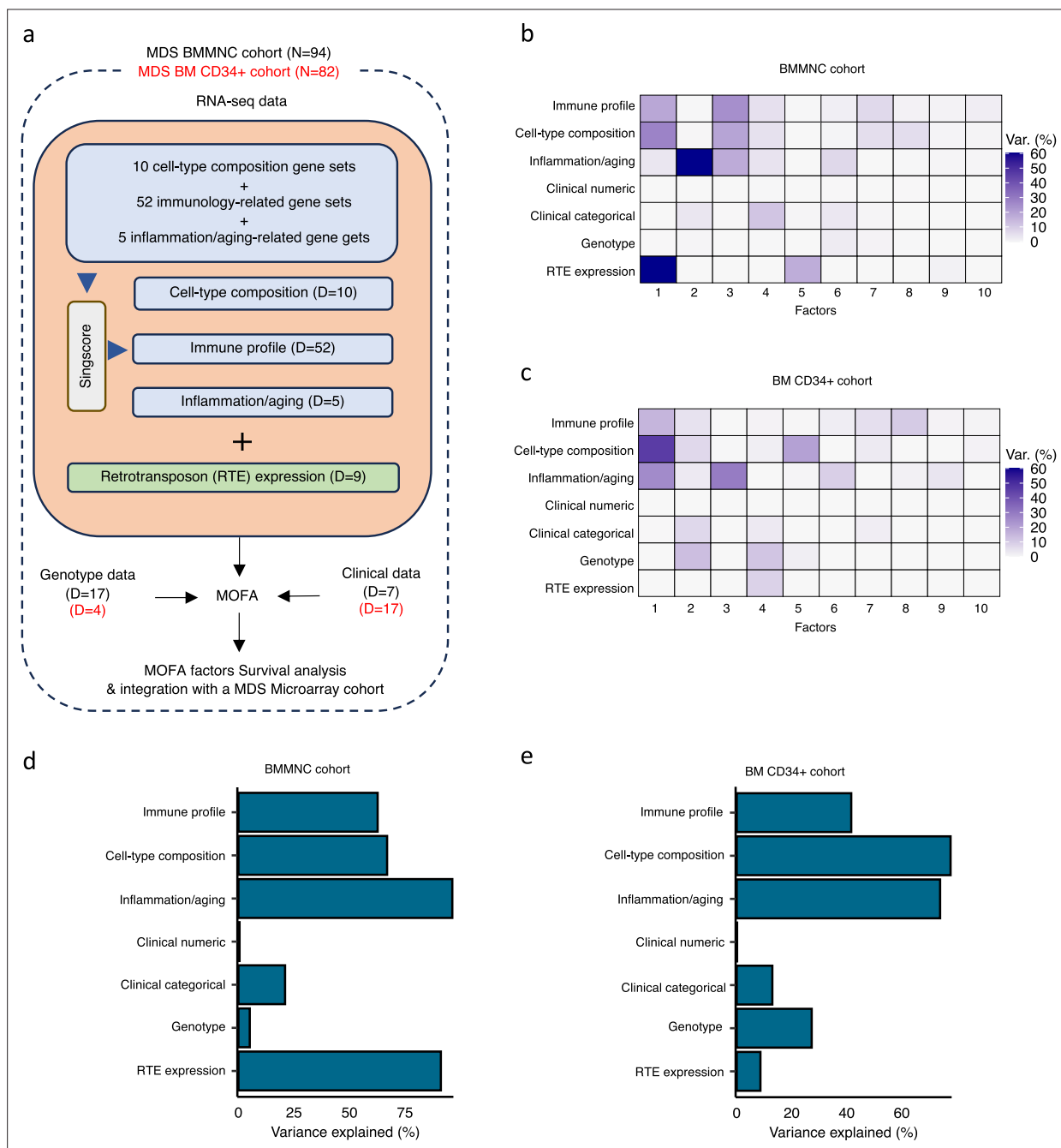


Figure 1. Schematic of multi-omics factor analysis (MOFA) workflow overview, downstream analyses, and factor determination in the bone marrow mononuclear cell (BMMNC) and bone marrow (BM) CD34 + cohorts. **(a)** RNA-seq, genotype, and clinical data were obtained from BMMNC samples of 94 myelodysplastic syndromes (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, and inflammation/aging. The other four views were clinical numeric and categorical (**Supplementary file 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.

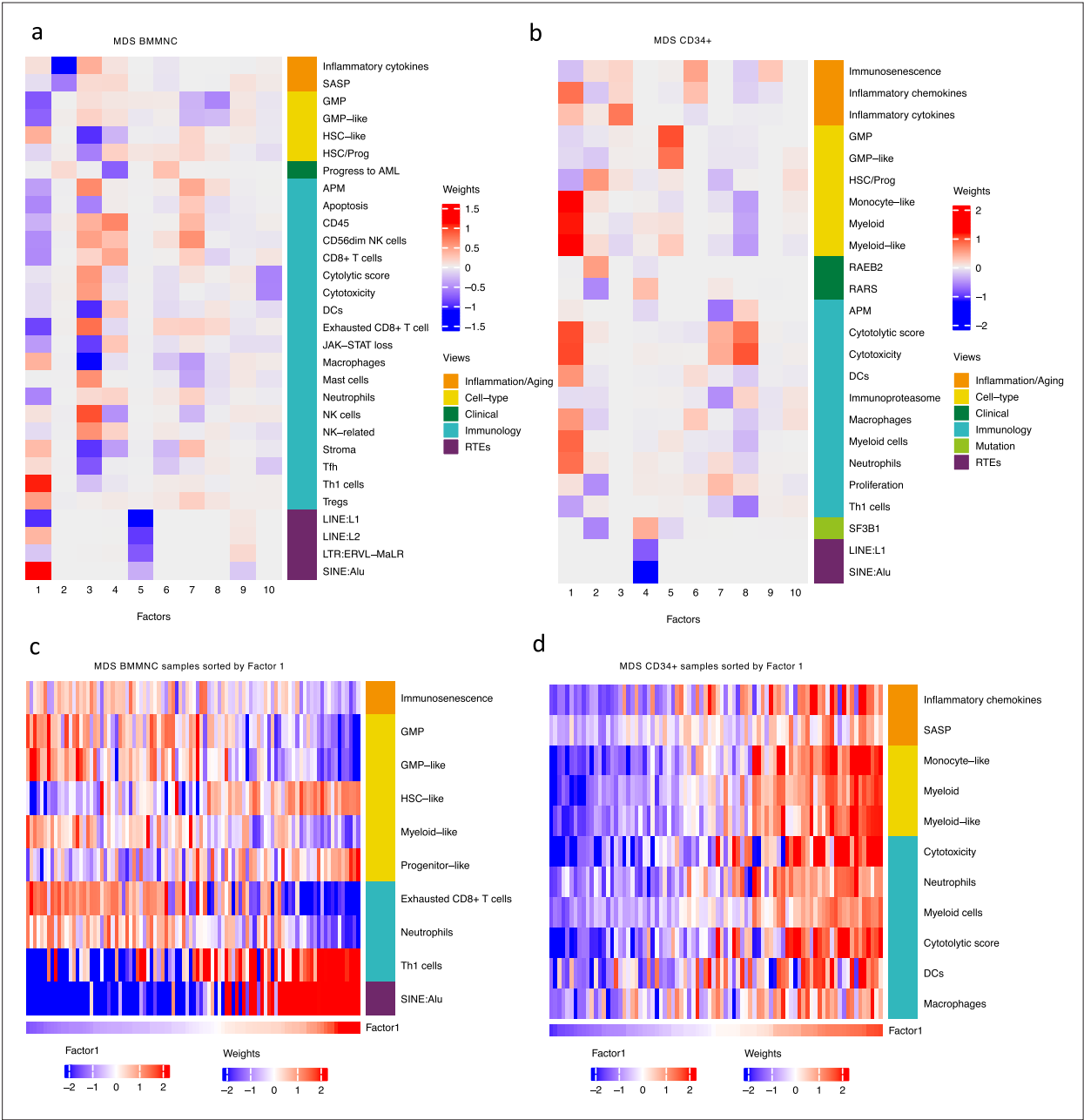


Figure 2. Breakdown of important features for each factor generated by multi-omics factor analysis (MOFA). **(a, b)** The important features with high weights for each biological view per factor were shown for the BMMNC and bone marrow (BM) CD34+. Blue represents features with the inverse correlation 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.

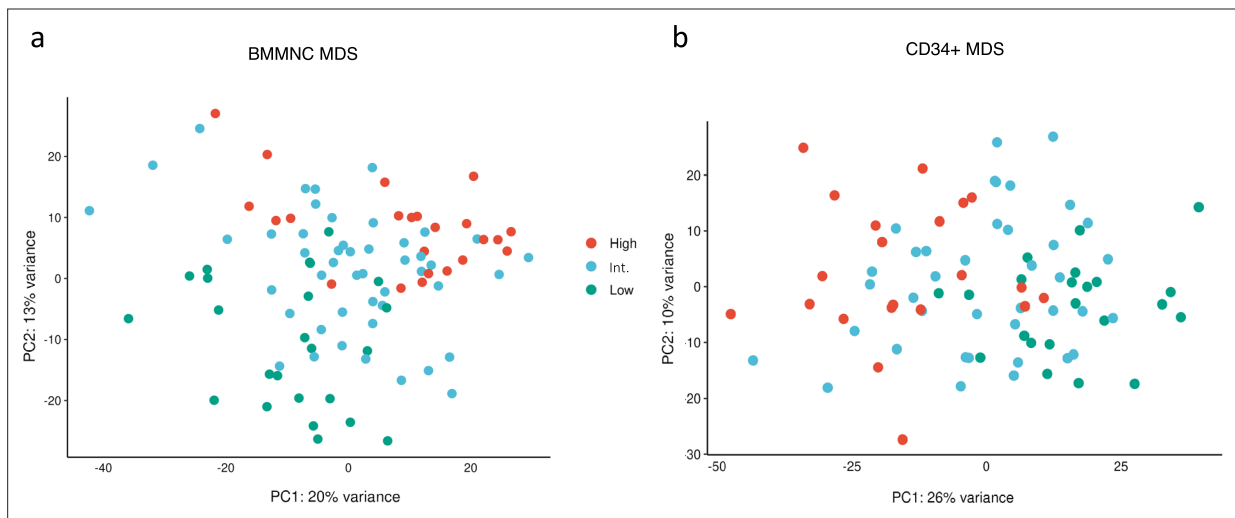


Figure 2—figure supplement 1. Factor 1 from the bone marrow mononuclear cell (BMMNC) and CD34 + myelodysplastic syndromes (MDS) cohorts stratified the patients in the gene expression principal component analysis (PCA) plots. **(a, b)** Principal component analysis on gene expression values of the BMMNC and CD34 + cohorts revealed separate clusters for patients split based on low (first quartile), intermediate (second and third quartile), and high (fourth 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.

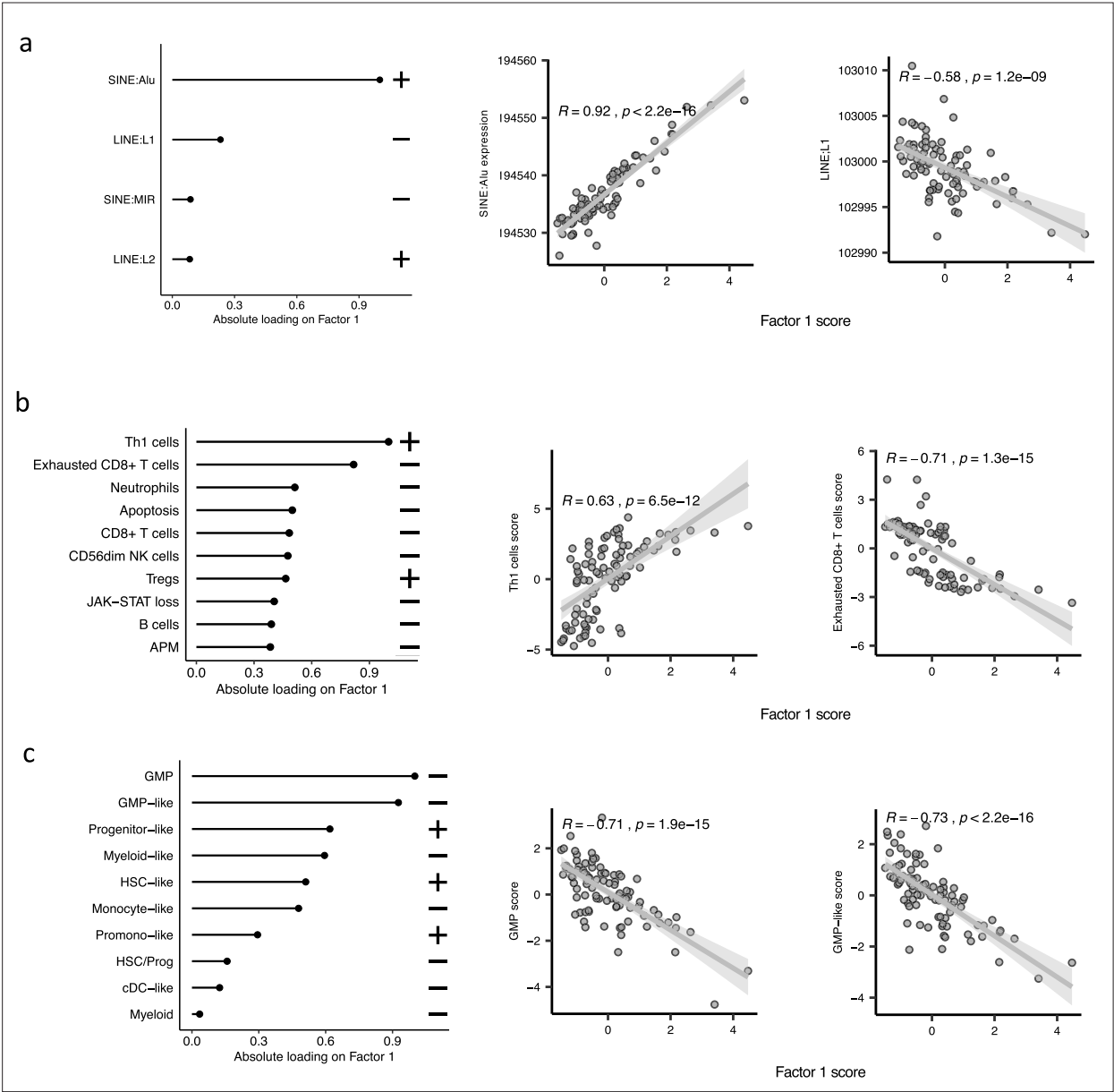


Figure 2—figure supplement 2. The top features influencing Factor 1 form various biological views in the bone marrow mononuclear cell (BMMNC) cohort. (a–c) Absolute loadings of the top features and their positive or inverse correlations with Factor 1 from retrotransposable element (RTE) expression, immune profile, and cell-type biological views, respectively. Pearson correlation coefficients (R) and p-values were displayed on top of the scatter plots.

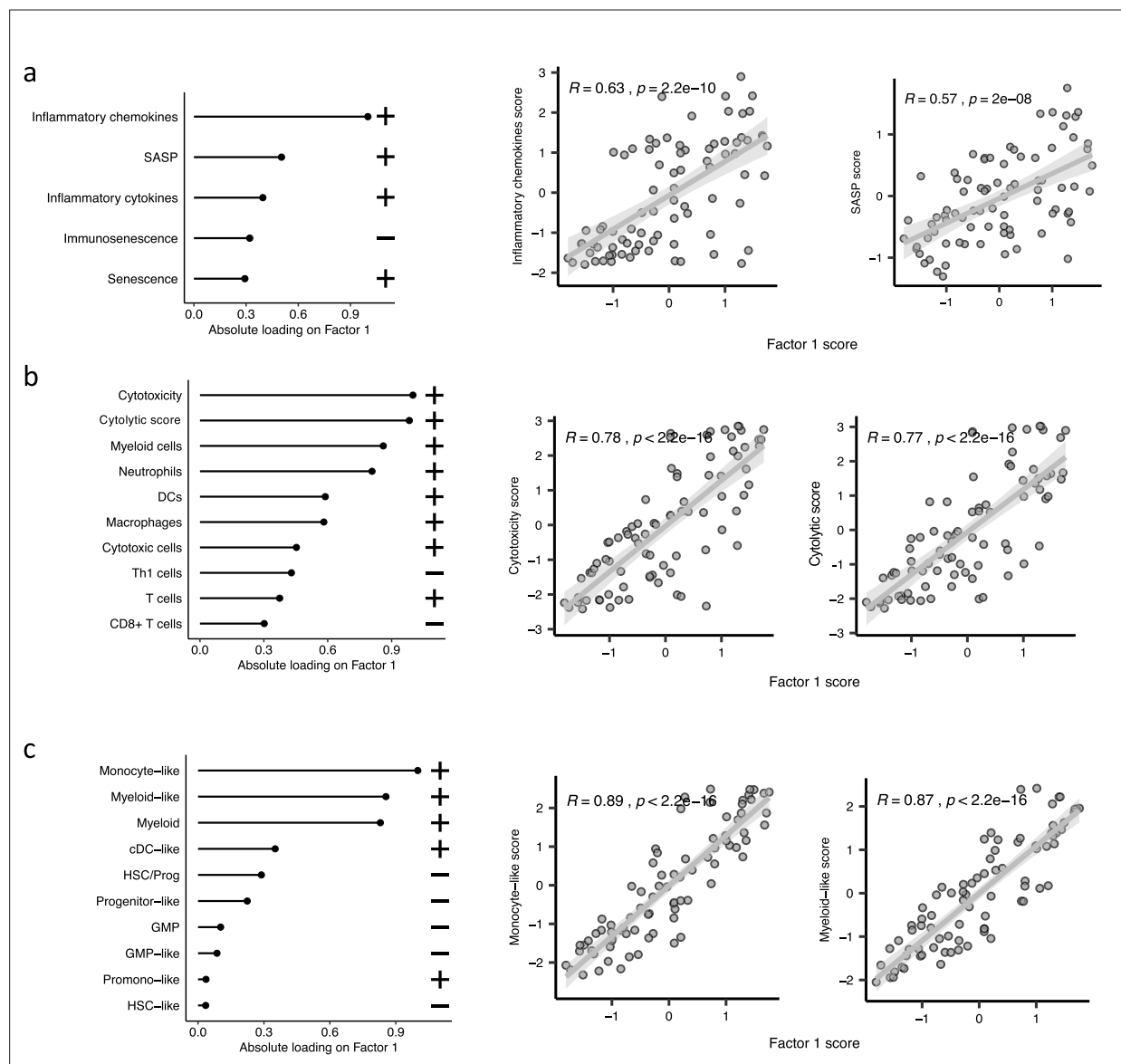


Figure 2—figure supplement 3. The top features influencing Factor 1 from various biological views in the CD34 + cohort. (a–c), Absolute loadings of the top features and their positive or inverse correlations with Factor 1 from inflammation/aging, immune profile, and cell-type biological views, respectively. Pearson correlation coefficients (R) and p-values were displayed on top of the scatter plots.

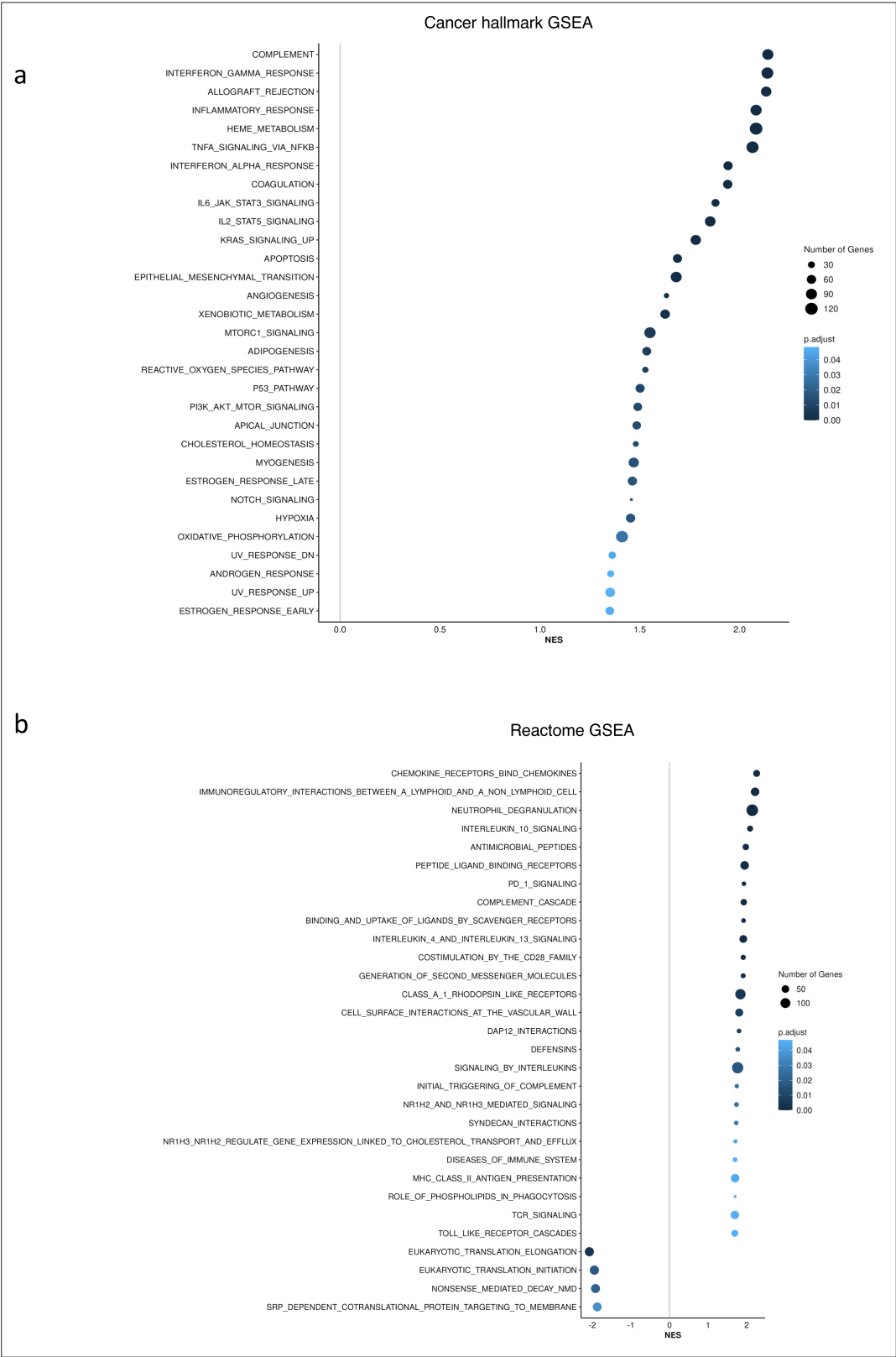


Figure 2—figure supplement 4. Gene set enrichment analysis (GSEA) identified upregulation of inflammation related pathways in patients having high (fourth quartile) versus low (first quartile) levels of Factor 1 in the CD34 + cohort. **(a)** Upregulation of inflammatory, Interferons, TNFA, and JAK-STAT 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.

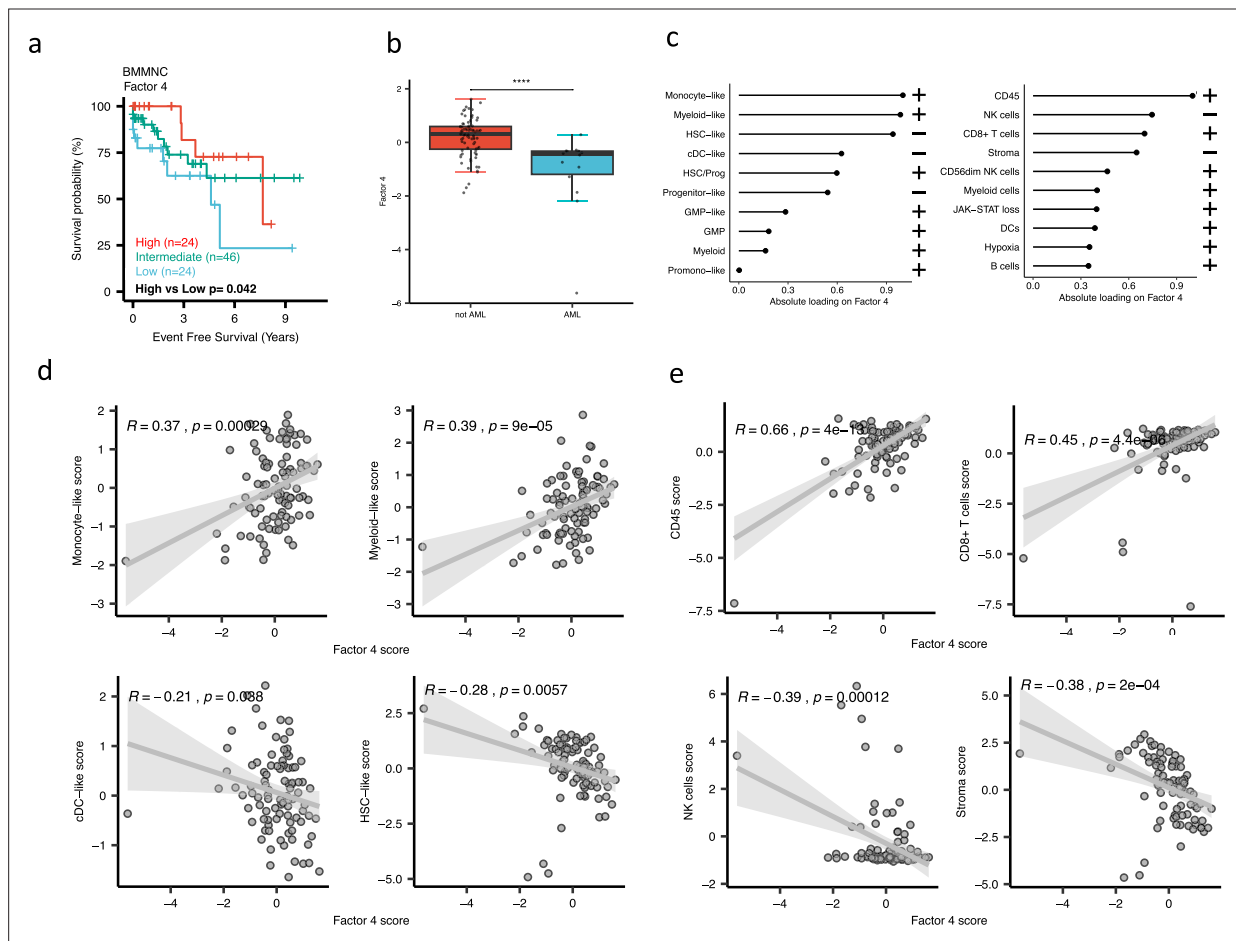


Figure 2—figure supplement 5. Characterisation of Factor 4 in the bone marrow mononuclear cell (BMMNC) cohort using gene signatures from multiple biological views. **(a)** Kaplan-Meier plots for the BMMNC cohort where patients were split based on low (first quartile), intermediate (second and third quartile), and high (fourth quartile) levels of Factor 4. The p-value was calculated using a log-rank test on event-free survival values of high versus low groups. **(b)** Boxplot depicting a 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.

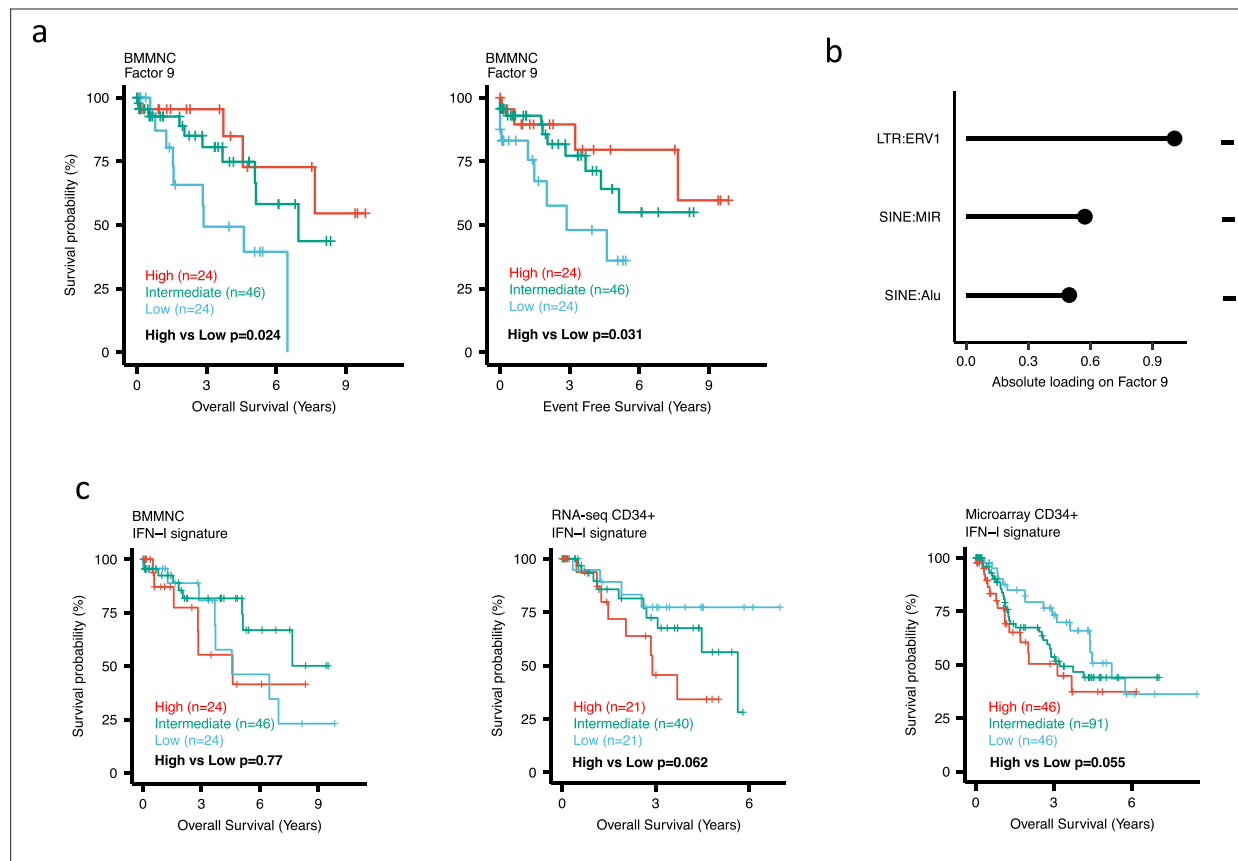


Figure 3. Impact of Factors 9 and IFN-I levels on myelodysplastic syndromes (MDS) prognosis. **(a)** Kaplan-Meier plots for the bone marrow mononuclear cell (BMMNC) cohort where patients were split based on low (first quartile), intermediate (second and third quartile), and high (fourth quartile) levels of Factor 9. **(b)** The absolute loading of the top three features affecting Factor 9 in retrotransposable element (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 a log-rank test on overall and event-free survival values of high versus low groups.

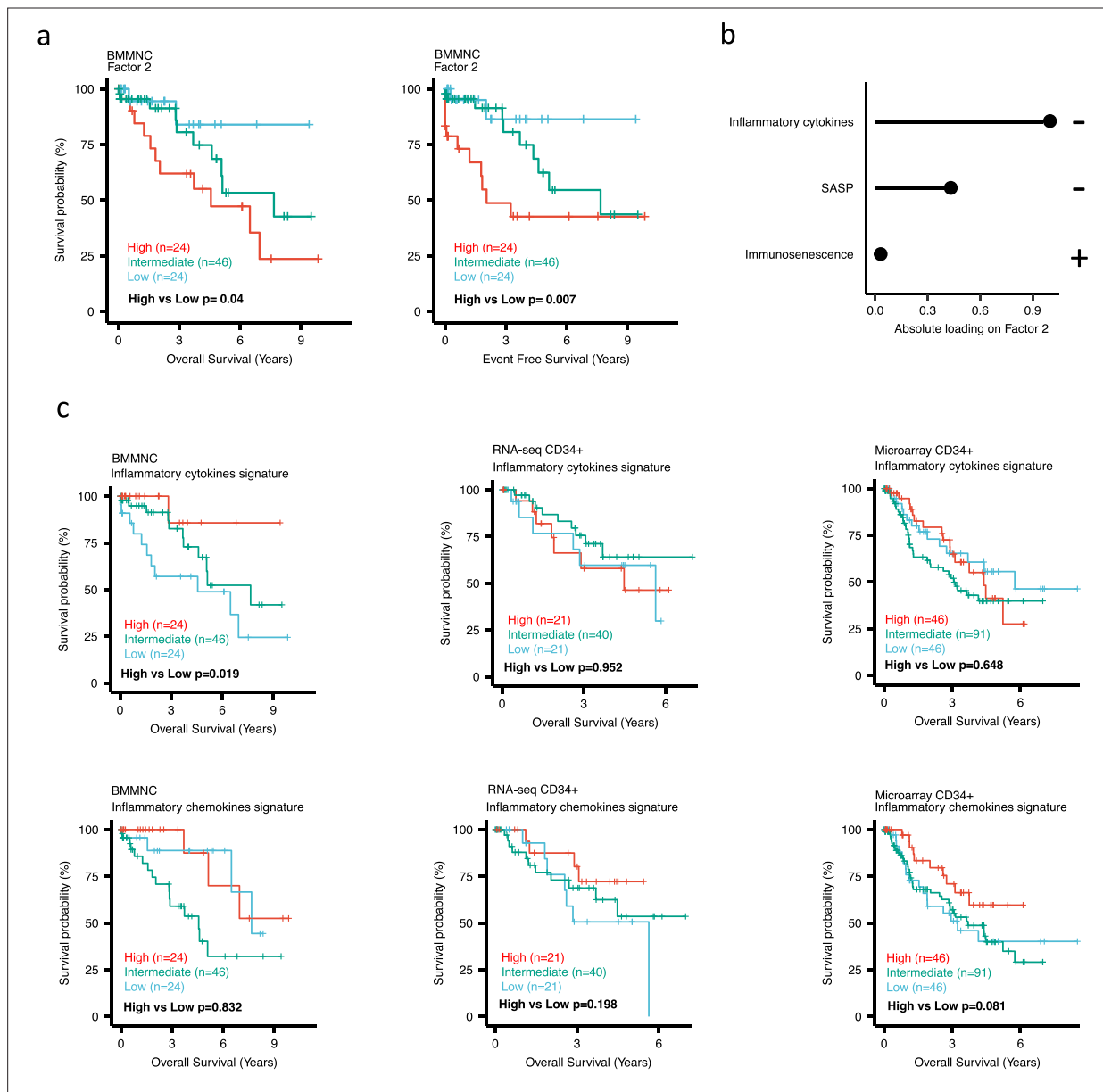


Figure 4. Impact of Factors 2 and inflammation levels on myelodysplastic syndromes (MDS) prognosis. **(a)** Kaplan-Meier plots for the bone marrow mononuclear cell (BMMNC) cohort where patients were split based on low (first quartile), intermediate (second and third quartile), and high (fourth quartile) 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 a log-rank test on overall and event-free survival values of high versus low groups.

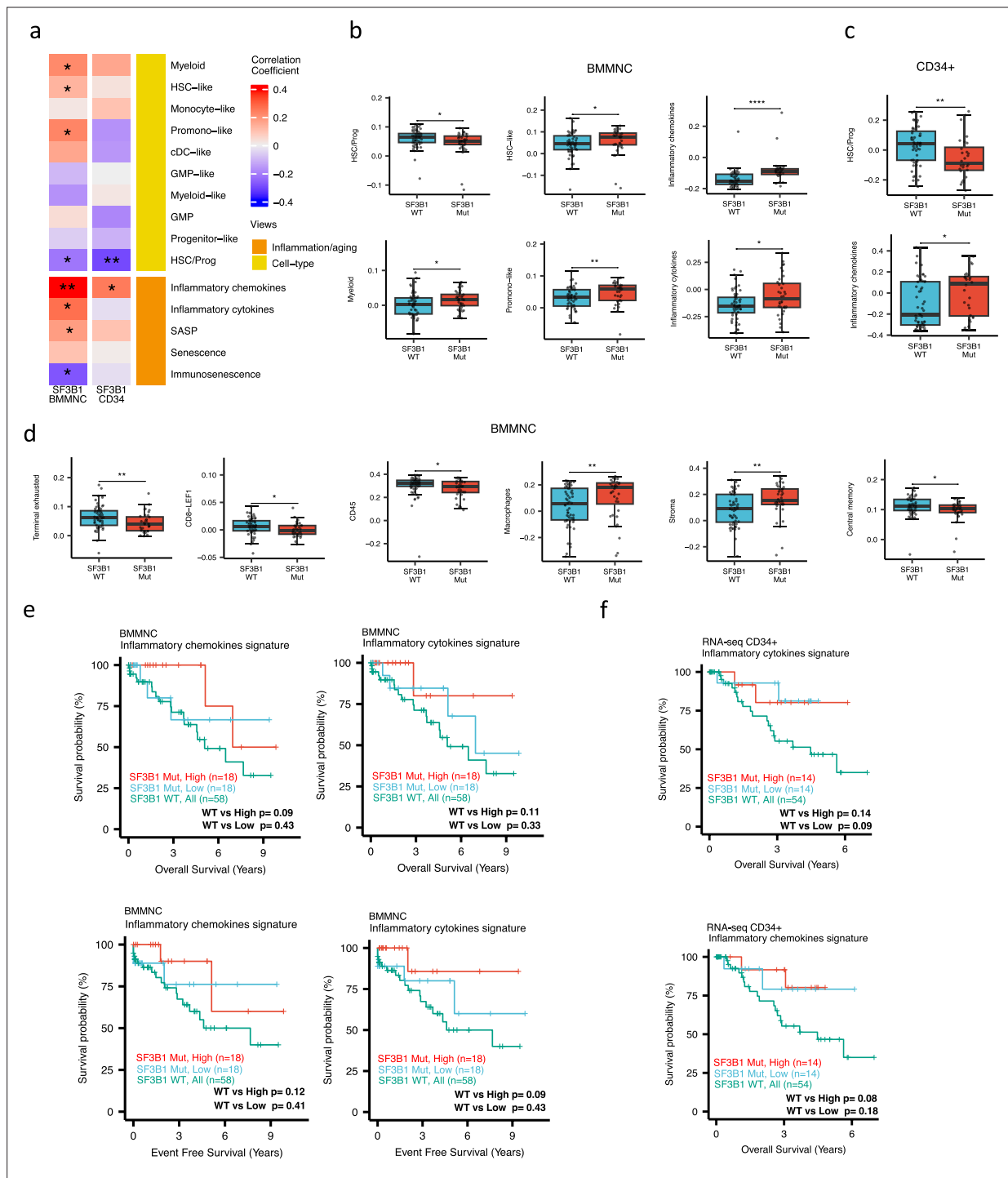


Figure 5. Characterisation of SF3B1 mutant myelodysplastic syndrome (MDS) using gene signatures from multiple biological views. **(a)** Association of a subset of inflammation/aging and cell-type features with SF3B1 mutation in the BMMNC and bone marrow (BM) CD34+ cohorts, with red depicting a positive 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 wild-type (WT) cases in the BMMNC ($n_{\text{Mut}} = 36$ vs. $n_{\text{WT}} = 58$) and CD34+ ($n_{\text{Mut}} = 28$ vs. $n_{\text{WT}} = 54$) 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 a log-rank test on overall and event-free survival values of SF3B1 mutant high and low versus wild-type (WT) groups.

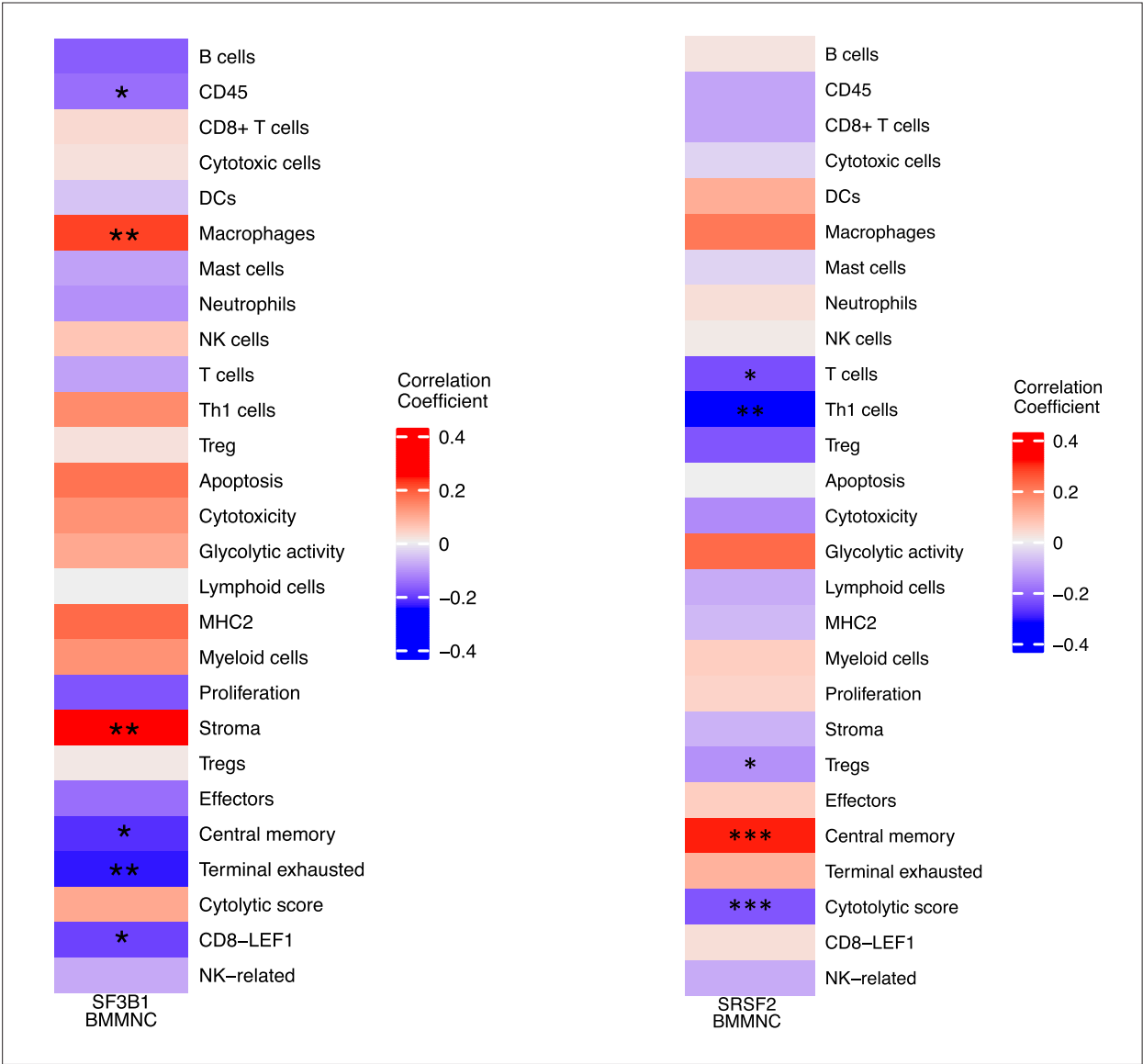


Figure 5—figure supplement 1. Correlation of selected immune profile gene sets with SF3B1 and SRSF2 mutations in the myelodysplastic syndromes (MDS) bone marrow mononuclear cell (BMMNC) cohort. **(a–b)** Correlation of a subset of immune profile features with SF3B1 and SRSF2 mutations in the BMMNC cohort, with red depicting a positive correlation and blue an inverse correlation with these mutations. The significances were calculated with the Wilcox rank-sum test, and the significant associations were shown by *(p<0.05), **(p<0.01), or ***(p<0.001).

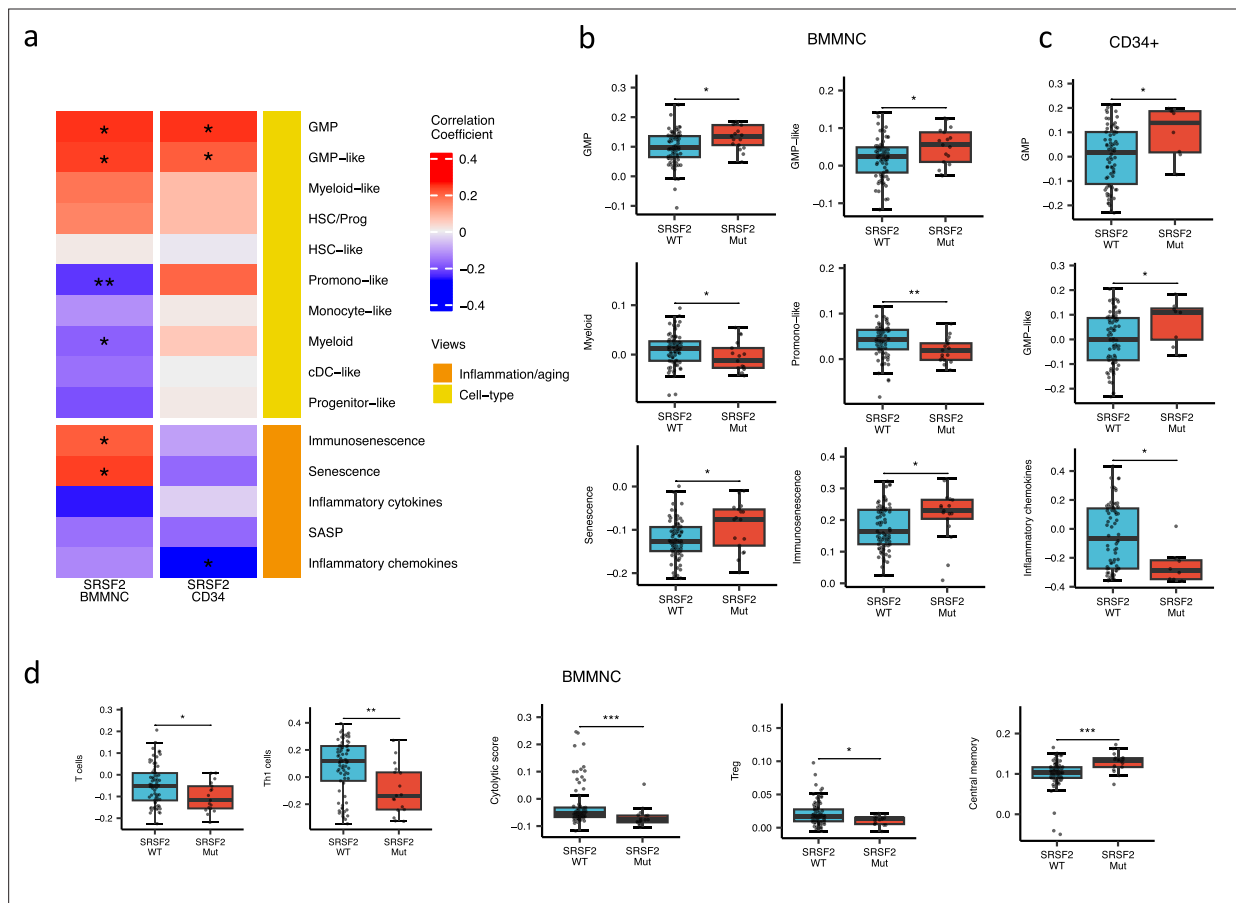


Figure 6. SRSF2 mutant myelodysplastic syndrome (MDS) is characterized by high GMP content and high levels of senescence and immunosenescence. **(a)** Association of a subset of inflammation/aging and cell-type features with SRSF2 mutation in the bone marrow mononuclear cell (BMMNC) and bone marrow (BM) 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 wild-type (WT) cases in the BMMNC (nMut = 17 vs. nWT = 77) and CD34+ (nMut = 8 vs. nWT = 74) 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.

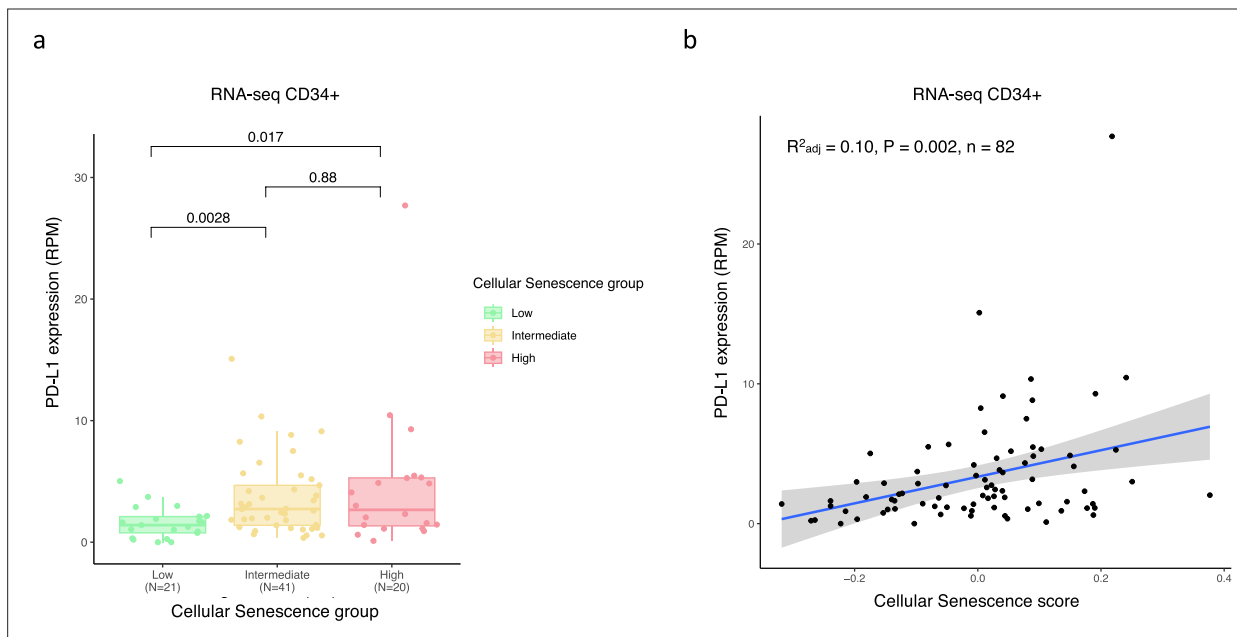


Figure 6—figure supplement 1. Positive correlation between programmed death-ligand 1 (PD-L1) expression and senescence score in bone marrow (BM) CD34 + RNA seq cohort. **(a)** Boxplot showing higher expression of PD-L1 gene in patient groups with high (fourth quartile) and intermediate (second and third quartile) vs low (first 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.