

Different inflammatory, fibrotic, and immunological signatures between pre-fibrotic and overt primary myelofibrosis

Seung-Hyun Jung,^{1-3*} Sung-Eun Lee,^{4*} Sujin Yun,³ Da-Eun Min,¹ Youngjin Shin,⁵ Yeun-Jun Chung^{2,3,5,6} and Sug Hyung Lee^{3,7,8}

¹Department of Biochemistry, ²Precision Medicine Research Center/Integrated Research Center for Genome Polymorphism, ³Department of Medical Sciences, ⁴Department of Internal Medicine, ⁵Basic Medical Science Facilitation Program, ⁶Department of Microbiology, ⁷Cancer Evolution Research Center, and ⁸Department of Pathology, College of Medicine, The Catholic University of Korea, Seoul, South Korea

**S-HJ and S-EL contributed equally as first authors.*


Correspondence: S.H. Lee
suhulee@catholic.ac.kr

Y-J. Chung
yejun@catholic.ac.kr

S-H. Jung
hyun@catholic.ac.kr

Received: April 12, 2024.
Accepted: September 27, 2024.
Early view: October 10, 2024.

<https://doi.org/10.3324/haematol.2024.285598>

©2024 Ferrata Storti Foundation
Published under a CC BY-NC license 

Supplementary Materials

Table of Contents

Supplementary Methods.....	page 2
Supplementary Figure 1.....	page 5
Supplementary Figure 2.....	page 7
Supplementary Figure 3.....	page 8
Supplementary Figure 4.....	page 9
Supplementary Figure 5.....	page 10
Supplementary Figure 6.....	page 11
Supplementary Figure 7.....	page 13
Supplementary Table 1.....	page 14
Supplementary Table 2.....	page 15
Supplementary References.....	page 16

Supplemental Methods

Bone marrow samples

The PMF diagnosis was made with BM biopsies based on the 2016 WHO criteria and involved a composite assessment of clinical and laboratory features.¹ The diagnosis of post-ET-MF and post-PV-MF adheres to the criteria published by the International Working Group for MPN Research and Treatment.² For scRNA-seq, we used BM aspirates from 33 patients with MPN. The 33 patients consisted of five ETs, one PV, 17 PMF (five pre-PMFs and 12 overt PMFs), and ten post-MF (six post-ET-MFs and four post-PV-MFs). Among the 12 patients with overt PMF, six were treated with the JAK inhibitor ruxolitinib, while the remaining six were not treated with any JAK inhibitor. To minimize the risk of PB dilution in BM aspirate samples, we repeated the aspiration at a different site if a dry tap was encountered during the procedure. We also performed microscopic examination of the BM aspirates to assess their quality and composition. The patients' BM fibrosis was confirmed by the BM biopsy, indicating that our aspirates represented remaining fluidic areas that the fibrotic PMF-BM surrounded. We obtained 16 mL (8 mL x 2 bottles) of BM aspirates from each patient. The mean of mononuclear cell yield for the MF patient aspirates was 1.1×10^8 cells (range, 1.9×10^7 – 2.5×10^8). This study was approved by the institutional review board of the Catholic University of Korea, College of Medicine (KC20TISI0206). All specimens from the patients in this study were obtained with appropriate consent in accordance with the declaration of Helsinki.

BM aspirates were filtered through a 100 μ m Cell Strainer (SPL Life Sciences), then layered onto a Ficoll-Paque PLUS (GE Healthcare) gradient. The sample was centrifuged at 2,500 rpm for 30 min. The isolated mononuclear cells were subsequently washed with Phosphate Buffered Saline (HyClone). After removing red blood cells using RBC lysis buffer (Miltenyi Biotec), the cells were counted and cryopreserved in freezing medium (10% DMSO + 90% FBS) at -80 °C until use.

scRNA-seq library construction

Chromium Single Cell 3' v3 Reagent kit (10x Genomics, Pleasanton, CA, USA) was used for the library construction of scRNA-seq according to the manufacturer's protocol. In brief, single-cell suspension was counted by hemocytometer (Thermo Fisher Scientific, Waltham, MA) and loaded into a Chromium instrument system targeting 10,000 cells. The cells were then partitioned into gel beads in emulsion in the Chromium instrument (10x Genomics), where cell lysis and barcoded reverse transcription of RNA occurred. Complementary DNA (cDNA) was synthesized and amplified for 14 cycles. cDNA clean-up was performed using a SPRIselect Reagent Kit (Beckman Coulter, Brea, CA). 50 ng of the amplified cDNA was used for each sample to construct indexed sequencing libraries. Sequencing libraries were sequenced on the Illumina NovaSeq platform (Illumina, San Diego, CA). Raw sequencing data generated for scRNA-sequencing have been deposited in Sequence Read Archive under accession number PRJNA1070224.

scRNA-seq data analysis

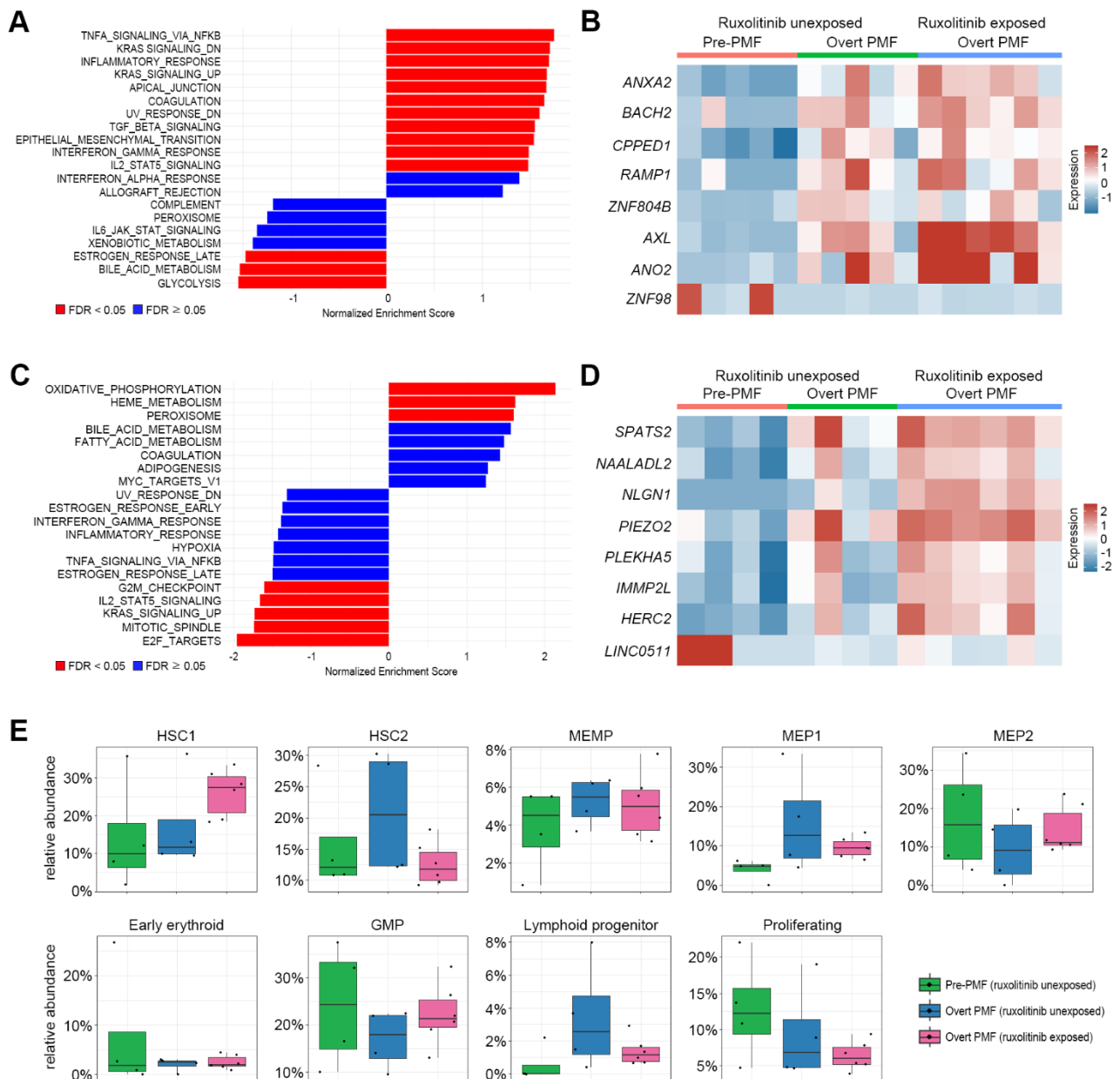
The sequenced data were processed into the expression matrices with the CellRanger 3.1.0 (10x Genomics). Sequencing reads were mapped to the GRCh38 reference genome, followed by unique molecular identifier (UMI) and barcode counting, and the UMI count matrices were constructed. Bioinformatics processing of the scRNA-sequencing data was performed with the R package Seurat.³ To exclude low-quality cells in scRNA-sequencing, we filtered cells with an expressed gene count of fewer than 200 or cells with more than 20% of reads corresponding to mitochondrial genes. Additionally, doublets were excluded using Scrublet.⁴ Data was log-normalized, and highly variable features were identified based on a variance stabilizing transformation method. All individual datasets were then integrated using the Harmony

algorithm.⁵ Principal components analysis and graph-based clustering were performed on the integrated datasets, and the clustering data were then applied to the uniform manifold approximation and projection. Each cell cluster was annotated for its cell type using the SingleR⁶ and the well-known cell-type-specific markers.

Cell cycle analysis was performed by using the ‘CellCycleScoring’ function in Seurat. Differentially expressed gene (DEG) analysis was used to identify significant DEGs within each cluster using the ‘FindMarkers’ function in Seurat. We kept only genes with an average \log_2 fold-change value greater than 1 and an adjusted *P* value (false discovery rate) lower than 0.2 in the analysis. Gene signature scores were calculated based on the Mann-Whitney U statistic using UCell.⁷ Gene set enrichment analysis was performed using fgsea with the ‘HALLMARK’ gene set downloaded from MSigDB.⁸ To compare MK subset abundance between PMF and other MPN diseases, MK cells of 16 MPN patients (five ET, one PV, six post-ET-MF, and four post-PV-MF) were projected onto the MK subsets of PMF patients using the “TransferData” function in Seurat.

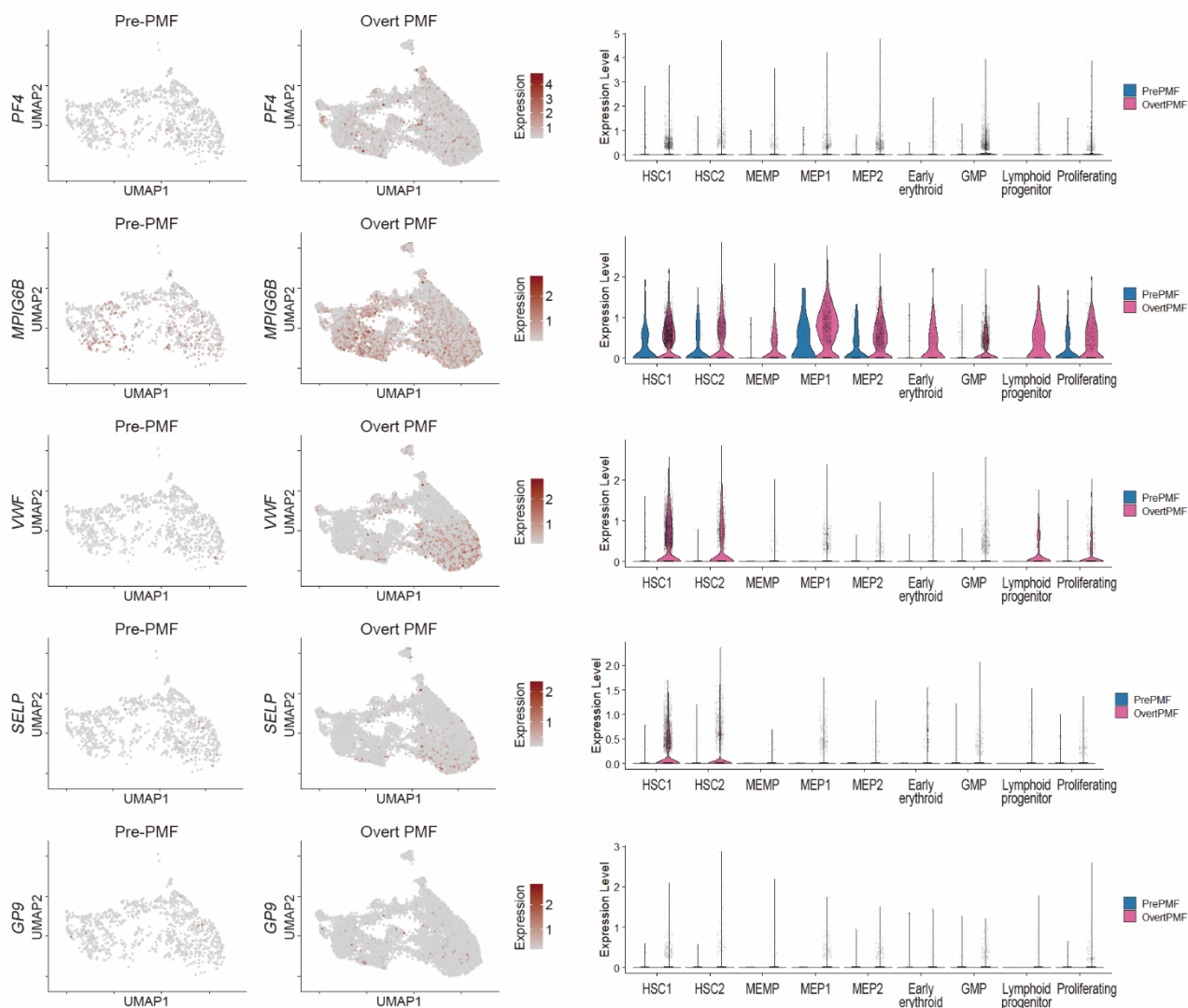
To examine the cell-to-cell communication between different cell subsets in the PMF-BM, receptor-ligand (LR) interactions were analyzed using CellChat.⁹ The LR pairs in CellChat retrieved from the previous studies were divided into four groups: cytokine/chemokine, immune checkpoint, growth factor, and others.⁹ Four rare cell populations (pDC, pre-B, and two HSPC subsets [early erythroid and lymphoid progenitor]) were excluded from the analysis. To simplify the data, some cell subsets were merged as follows; naïve T (CD8⁺Naïve T, CD4⁺Naïve T, and CD3⁺CD4⁺Naïve T), NK (NK1, NK2, NK3, and NK4), and monocyte (mono1, mono2, mono4, mono5, mono6, and mono7). To compare the CellChat results between normal BM and MM BM, the ‘liftCellChat’ function was used. Cell-cell communications, if there are only a few cells (< 10) in certain cell groups, were filtered out.

Supplementary Figures

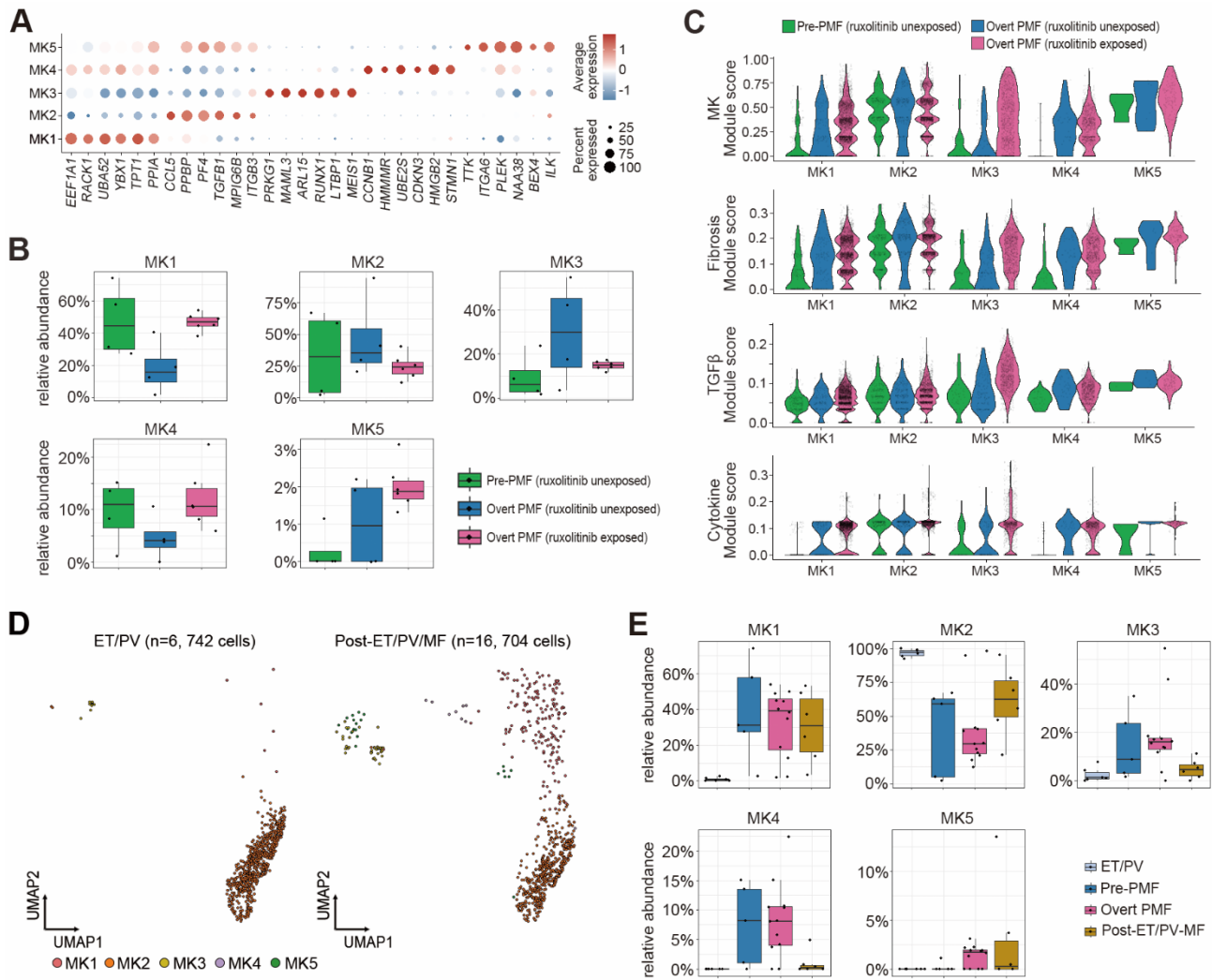


Supplementary Fig. 1. HSPC subsets and gene signatures. (A) The top 20 statistically-enriched ‘HALLMARK’ gene sets for the HSC1 subset. The x-axis represents the normalized enrichment score. (B) The heatmap shows the eight DEGs in the HSC1 subset between pre-PMF and overt PMF (seven up-regulated and one down-regulated in overt PMF). Red and blue colors indicate up-regulated and down-regulated genes, respectively. (C) The top 20 statistically enriched ‘HALLMARK’ gene sets for MEP1 subset. GSEA analysis revealed enrichment in the MK-lineage differentiation,¹⁰⁻¹² including ‘oxidative phosphorylation’, ‘peroxisome’, and ‘coagulation’. The x-axis represents the normalized enrichment score. (D) The heatmap shows the eight DEGs in the MEP1 subset between pre-PMF and

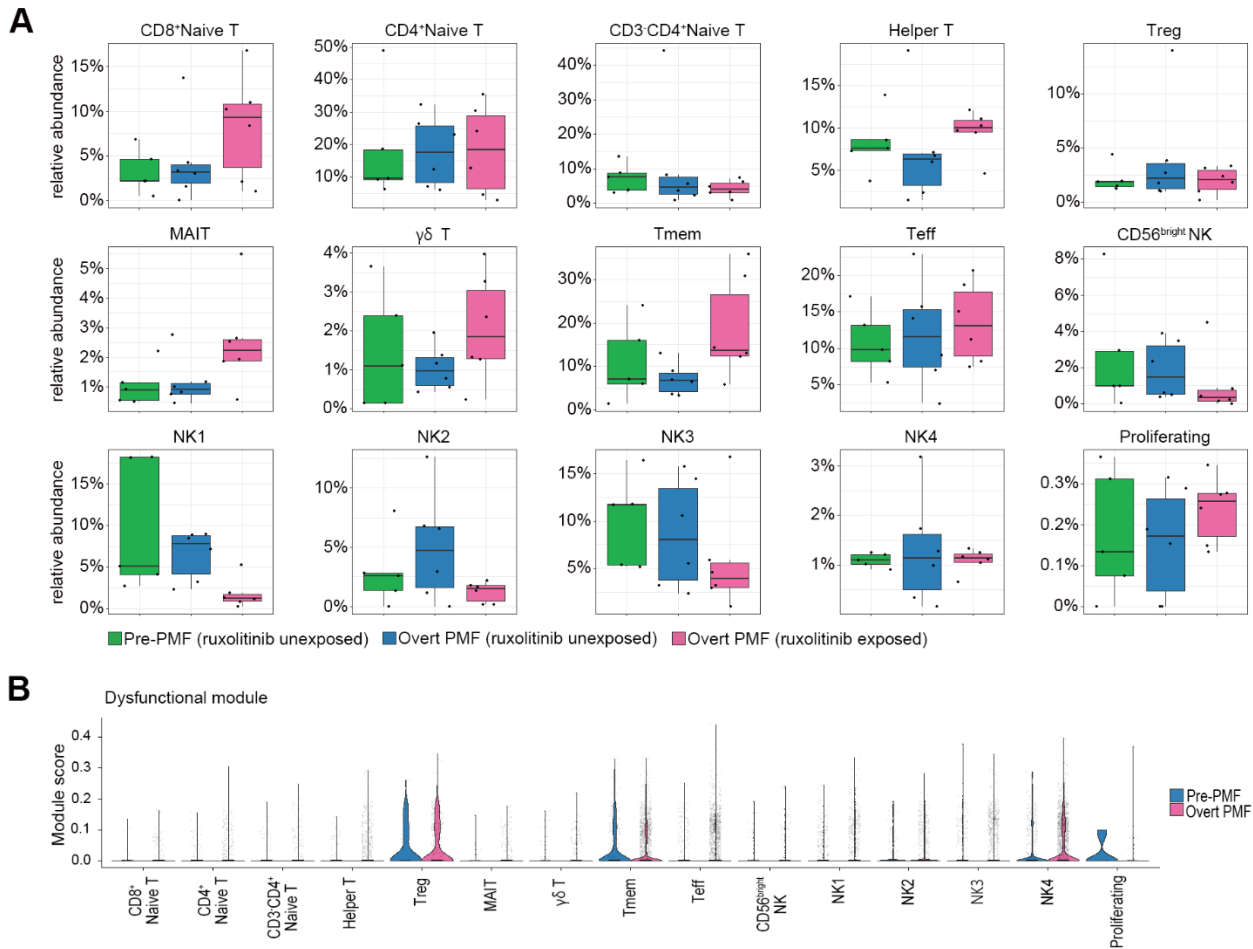
overt-PMF (seven up-regulated and one down-regulated in overt PMFs). Red and blue colors indicate up-regulated and down-regulated genes, respectively. (E) Box plots representing the proportion of each cell type between pre-PMF (n = 5), ruxolitinib-unexposed overt PMF (n=6), and ruxolitinib-exposed overt PMF (n = 6). The mean and 95% confidence interval are represented with black lines.



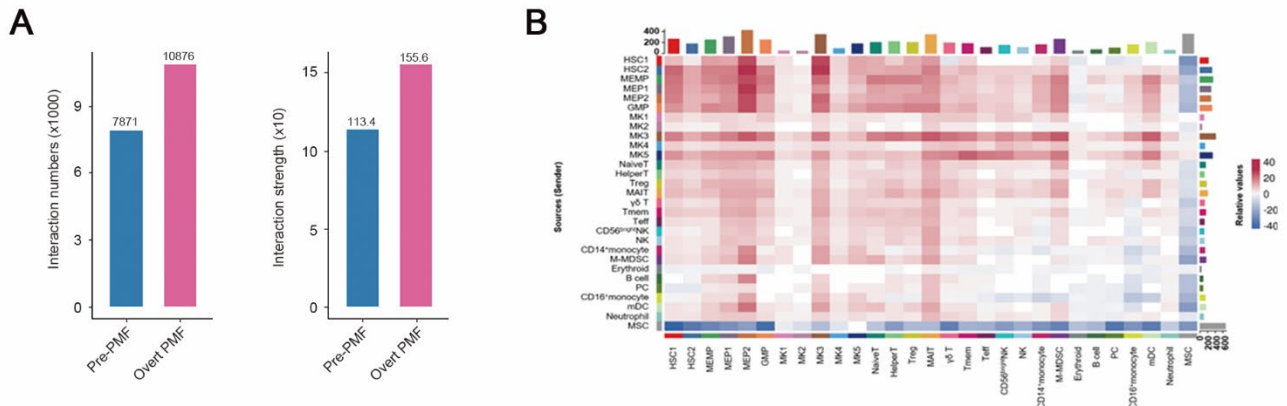
Supplementary Fig. 2. HSPC subsets and their MK signature genes. The expression level of MK signature genes is shown in the UMAP plot by the clinical groups (left panels) and the violin plot by HSPC subsets (right panels).



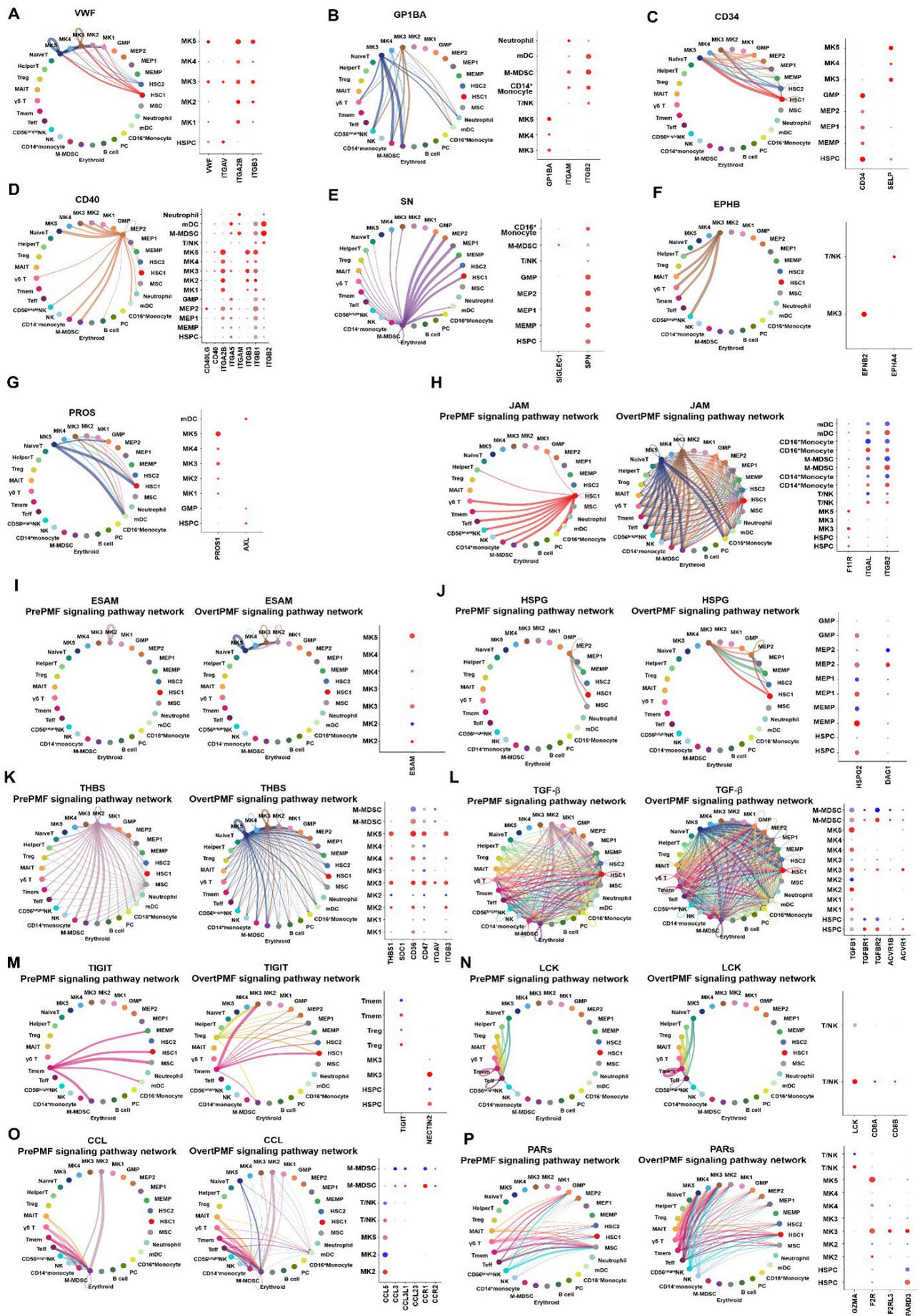
Supplementary Fig. 3. MK subsets and their gene signatures. (A) Dot plot of subset-specific marker genes per MK subset. Dot intensity (from blue to red) represents the average expression value of all cells per MK subset, whereas dot size represents the proportion of cells expressing the genes. (B) Box plots represent the proportion of each MK subset between pre-PMF (n = 5), ruxolitinib unexposed overt PMF (n = 6), and ruxolitinib exposed overt PMF (n = 6). The mean and 95% confidence interval are represented with black lines. (C) Signature scores are shown in the violin plot by MK subsets. (D) MK subsets in non-PMF MPNs. UMAP plot colored by MK subsets. (E) Box plots represent the proportion of each MK subset between ET/PV (n = 6), pre-PMF (n = 5), overt PMF (n = 12), and post-ET/PV-MF (n = 10). The mean and 95% confidence interval are represented with black lines. The overt-PMF-specific MK5 subset was observed in post-ET/PV-MF patients (three out of six patients with more than 10 MK cells), but not in ET/PV.



Supplementary Fig. 4. The subset abundance and dysfunctional module score of T and NK cell populations. (A) Box plots represent the proportion of each T and NK subset between pre-PMF (n = 5), ruxolitinib unexposed overt PMF (n=6), and ruxolitinib exposed overt PMF (n = 6). The mean and 95% confidence interval are represented with black lines. (B) The expression level of dysfunctional signature score is shown in the violin plot by T and NK subsets.

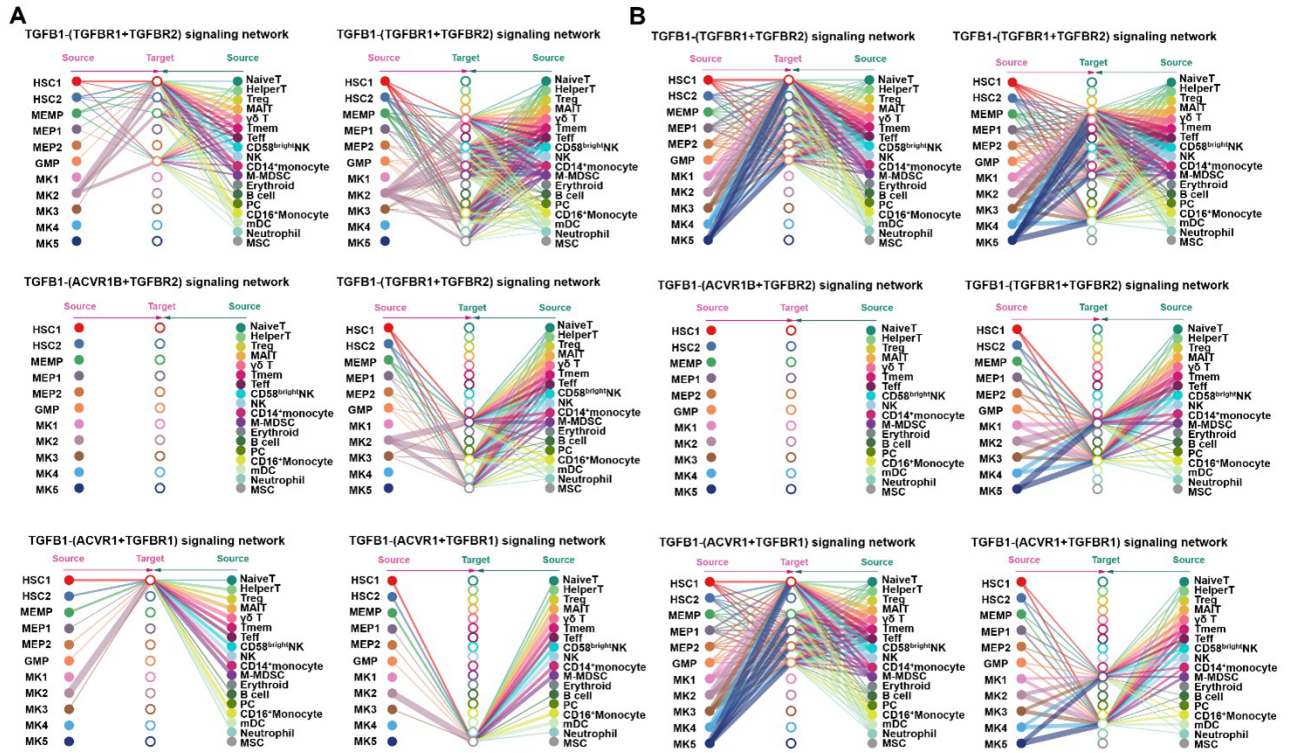


Supplementary Fig. 5. Visualization of cell-cell communication. (A) Comparison of the total number of interactions and strength of these interactions between pre-PMF and overt PMF. (B) Differential number of ligand-receptor (LR) interactions between pre-PMF and overt PMF. The total number of enriched LR interactions is shown as a bar on the x-axis and y-axis, and the relative strength of the interactions (pre-PMF vs overt PMF) is shown in the heatmap. Red and blue indicate enrichment of overt PMF and pre-PMF, respectively.



Supplementary Fig. 6. Cell-cell communications enriched in overt PMF compared with pre-PMF.

Circle plots (left panel) represent the inferred interaction pathways of (A) VWF, (B) GP1BA, (C) CD34, (D) CD40, (E) SN, (F), EPHB, and (G) PROS, which were exclusively identified in overt PMF, as well as those of (H) JAM, (I) ESAM, (J) HSPG, (K) THBS, (L) TGF- β , (M) TIGIT, (N) LCK, (O) CCL, and (P) PARs, identified in both pre-PMF and overt PMF. Edge width represents the communication probability (strength of the interactions) between cell populations. Edge colors are consistent with the signaling source. The expression of ligands and receptors for each signaling pathway in each cell subset from pre-PMF (blue) and overt PMF (red) is shown. Genes corresponding to ligands in each signaling pathway are indicated in bold. MK5 from pre-PMF was excluded due to a low number of cells.



Supplementary Fig. 7. The inferred TGF- β signaling network. The inferred TGF- β signaling network from (A) pre-PMF and (B) overt PMF. Edge width represents the communication probability.

Supplementary Table 1. Significantly enriched genes in each BM cell subset.
The contents of Supplementary Table 1 are provided in a separate Excel file.

Supplementary Table 2. Pseudobulk DEGs of HSPC and MK between pre- and overt-PMFs

Subset	Gene	Avg_log2FC*	p_val**	p_val_adj**	pct.1 [§]	pct.2 [§]
HSC1	<i>ANXA2</i>	6.84	9.1x10 ⁻⁰⁹	2.3x10 ⁻⁰⁴	1.00	0.40
HSC1	<i>BACH2</i>	5.54	1.3x10 ⁻⁰⁸	3.2x10 ⁻⁰⁴	1.00	1.00
HSC1	<i>CPPED1</i>	5.46	3.1x10 ⁻⁰⁷	0.008	1.00	0.80
HSC1	<i>RAMP1</i>	6.62	6.0x10 ⁻⁰⁷	0.015	1.00	0.40
HSC1	<i>ZNF804B</i>	8.11	8.0x10 ⁻⁰⁷	0.021	1.00	0.60
HSC1	<i>ZNF98</i>	-1.16	1.1x10 ⁻⁰⁶	0.028	0.55	0.60
HSC1	<i>AXL</i>	7.47	3.8x10 ⁻⁰⁶	0.098	0.91	0.40
HSC1	<i>ANO2</i>	9.47	6.7x10 ⁻⁰⁶	0.173	0.91	0.40
MEP1	<i>SPATS2</i>	6.14	2.2x10 ⁻⁰⁹	5.6x10 ⁻⁰⁵	1.00	1.00
MEP1	<i>NAALADL2</i>	6.11	3.2x10 ⁻⁰⁸	8.2x10 ⁻⁰⁴	1.00	1.00
MEP1	<i>NLGN1</i>	12.90	3.3x10 ⁻⁰⁸	8.6x10 ⁻⁰⁴	1.00	0.00
MEP1	<i>PIEZO2</i>	6.71	3.5x10 ⁻⁰⁸	9.1x10 ⁻⁰⁴	1.00	0.75
MEP1	<i>LINC00511</i>	-1.41	4.8x10 ⁻⁰⁸	0.001	0.80	0.50
MEP1	<i>PLEKHA5</i>	5.25	2.1x10 ⁻⁰⁶	0.054	1.00	1.00
MEP1	<i>IMMP2L</i>	5.38	2.9x10 ⁻⁰⁶	0.073	1.00	1.00
MEP1	<i>HERC2</i>	5.87	4.9x10 ⁻⁰⁶	0.126	1.00	1.00
MK	<i>PRTN3</i>	-1.14	1.1x10 ⁻¹¹	2.7x10 ⁻⁰⁷	0.58	0.60
MK	<i>COL24A1</i>	10.46	3.9x10 ⁻⁰⁹	1.0x10 ⁻⁰⁴	0.92	0.20
MK	<i>FBXL7</i>	-1.61	3.2x10 ⁻⁰⁷	0.008	0.50	0.80
MK	<i>MAML3</i>	6.87	3.4x10 ⁻⁰⁷	0.009	1.00	1.00
MK	<i>KIAA1217</i>	-1.05	3.7x10 ⁻⁰⁷	0.010	0.58	1.00
MK	<i>SCPEP1</i>	1.78	4.2x10 ⁻⁰⁷	0.011	0.92	0.80
MK	<i>STXBP5</i>	6.75	9.2x10 ⁻⁰⁷	0.024	0.92	1.00
MK	<i>YES1</i>	7.07	1.4x10 ⁻⁰⁶	0.036	0.92	0.80
MK	<i>ARHGEF3</i>	6.69	1.6x10 ⁻⁰⁶	0.042	0.92	0.80
MK	<i>LTBP1</i>	8.17	1.6x10 ⁻⁰⁶	0.042	0.92	1.00
MK	<i>FAM30A</i>	6.73	1.9x10 ⁻⁰⁶	0.050	0.92	0.60
MK	<i>INPP4B</i>	7.73	2.0x10 ⁻⁰⁶	0.052	0.92	1.00
MK	<i>CXCL2</i>	8.67	2.210x ⁻⁰⁶	0.057	0.83	0.40
MK	<i>GRB10</i>	7.46	2.6x10 ⁻⁰⁶	0.066	0.92	0.40
MK	<i>EFCAB13</i>	9.17	3.5x10 ⁻⁰⁶	0.089	0.83	0.60
MK	<i>UBASH3B</i>	7.34	3.8x10 ⁻⁰⁶	0.097	0.92	0.80
MK	<i>STEAP1B</i>	1.44	4.9x10 ⁻⁰⁶	0.126	0.83	1.00
MK	<i>GFPT1</i>	7.20	6.5x10 ⁻⁰⁶	0.166	0.92	0.60
MK	<i>SCFD2</i>	6.75	6.7x10 ⁻⁰⁶	0.173	0.92	1.00

*Log fold-change of the average expression between the selected cell populations and all other cell groups.

**p_val and p_val_adj represent the original *p*-value (unadjusted) and adjusted *p*-value based on Bonferroni correction using all features in the dataset, respectively.

§pct.1 and pct.2 represent the percentage of cells, where the feature is detected in the selected cell population and all other cell groups, respectively.

Supplemental References

1. Barbui T, Thiele J, Gisslinger H, et al. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. *Blood Cancer J.* 2018;8(2):15.
2. Tefferi A. Primary myelofibrosis: 2021 update on diagnosis, risk-stratification and management. *Am J Hematol.* 2021;96(1):145-162.
3. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell.* 2019;177(7):1888-1902 e1821.
4. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* 2019;8(4):281-291 e289.
5. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* 2019;16(12):1289-1296.
6. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* 2019;20(2):163-172.
7. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring. *Comput Struct Biotechnol J.* 2021;19(3796-3798).
8. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011;27(12):1739-1740.
9. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* 2021;12(1):1088.
10. Lannan KL, Sahler J, Kim N, et al. Breaking the mold: transcription factors in the anucleate platelet and platelet-derived microparticles. *Front Immunol.* 2015;6(48).
11. Nakamura-Ishizu A, Matsumura T, Stumpf PS, et al. Thrombopoietin Metabolically Primes Hematopoietic Stem Cells to Megakaryocyte-Lineage Differentiation. *Cell Rep.* 2018;25(7):1772-1785 e1776.
12. Poirault-Chassac S, Nivet-Antoine V, Houvert A, et al. Mitochondrial dynamics and reactive oxygen species initiate thrombopoiesis from mature megakaryocytes. *Blood Adv.* 2021;5(6):1706-1718.