Expression profiles analysis identifies specific interferon-stimulated signatures as potential diagnostic and predictive indicators of *JAK2V617F*+ myelofibrosis

Yanhong Zhao, Di Wang, Yipeng Liang, Changlu Xu, Lihong Shi and Jingyuan Tong*

State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin, China

Objective: This study aimed to identify specific dysregulated genes with potential diagnostic and predictive values for *JAK2V617F*+ myelofibrosis.

Methods: Two gene expression datasets of CD34+ hematopoietic stem and progenitor cells (HSPCs) from patients with *JAK2V617F*+ myeloproliferative neoplasm (MPN) [n = 66, including polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF)] and healthy controls (HC) (n = 30) were acquired from the GEO (Gene Expression Omnibus) database. The differentially expressed genes (DEGs) were screened between each *JAK2V617F*+ MPN entity and HC. Subsequently, functional enrichment analyses, including Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, and Gene Set Enrichment Analysis (GSEA), were conducted to decipher the important biological effects of DEGs. Protein−protein interaction (PPI) networks of the DEGs were constructed to identify hub genes and significant modules. Another two gene expression profiles of patients with *JAK2V617F*+ MPN [n = 23, including PV, ET, secondary myelofibrosis (SMF), and PMF] and HC (n = 6) from GEO were used as external validation datasets to prove the reliability of the identified signatures.

Results: KEGG analysis revealed the upregulated genes in three *JAK2V617F*+ MPN entities compared with HC were essentially enriched in inflammatory pathways and immune response signaling pathways, and the number of these pathways enriched in PMF was obviously more than that in PV and ET. Following the PPI analysis, 10 genes primarily related to inflammation and immune response were found upregulated in different *JAK2V617F*+ MPN entities. In addition, Reactome enrichment analysis indicated that interferon signaling pathways were enriched specifically in PMF but not in PV or ET. Furthermore, several interferon (IFN)-stimulated genes were identified to be uniquely upregulated in *JAK2V617F*+ PMF. The external datasets validated the

upregulation of four interferon-related genes (*OAS1*, *IFITM3*, *GBP1*, and *GBP2*) in *JAK2V617F*[+] myelofibrosis. The receiver operating characteristic (ROC) curves indicate that the four genes have high area under the ROC curve (AUC) values when distinguishing *JAK2V617F*[+] myelofibrosis from PV or ET.

Conclusion: Four interferon-stimulated genes (*OAS1*, *IFITM3*, *GBP1*, and *GBP2*) exclusively upregulated in *JAK2V617F*[+] myelofibrosis might have the potential to be the auxiliary molecular diagnostic and predictive indicators of myelofibrosis.

Introduction

Polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF) are a range of heterogeneous hematopoietic diseases classified as Philadelphia chromosome-negative myeloproliferative neoplasms (Ph- MPNs), which also include chronic neutrophilic leukemia (CNL), chronic eosinophilic leukemia not otherwise specified (CEL-NOS), and unclassifiable MPN according to 2016 WHO classification (Arber et al., 2016). PV, ET, and PMF are more frequent than other subtypes of Ph-MPNs (Anderson and McMullin, 2014). Among PV, ET, and PMF, the most frequent driver mutation is *JAK2V617F*, which could be observed in the great majority of patients with PV and most patients with ET or PMF (O'Sullivan and Mead, 2019). Although sharing the same *JAK* mutation, the three *JAK2V617F*[+] MPN entities have distinct clinical manifestations: PV with an elevated hemoglobin/hematocrit potentially accompanied by leukocytosis or thrombocytosis, ET with thrombocytosis, and PMF with bone marrow fibrosis, extramedullary hematopoiesis, and even progressive cytopenia (Arber et al., 2016). Among these three subtypes, PMF is the more malignant and more aggressive myeloproliferative neoplasm, while PV and ET are relatively benign and indolent diseases (Grinfeld et al., 2018). However, with the progression, both PV and ET have a chance to develop into secondary myelofibrosis (SMF), sharing similar clinical features to PMF; therefore, these patients could be named post-PV-MF or post-ET-MF (Mesa et al., 2007).

The pathogenesis and progression mechanism of MPN is multifactorial. The constitutive activation of the JAK/STAT pathway due to the *JAK2V617F* acquired by the hematopoietic stem cell (HSC) plays an essential role in MPN pathogenesis (Baxter et al., 2005; Kralovics et al., 2005; Levine et al., 2005). Somatic mutations of non-driver genes related to epigenetic regulation, splicing factors, and other signaling pathways have been discovered in a large number of MPN patients, particularly in individuals with PMF (Ayalew et al., 2016; Li et al., 2016). In addition, an altered bone marrow environment and chronic inflammation could also participate in the pathogenesis and progression mechanism of MPN (Hasselbalch, 2013; Schepers et al., 2013; Bjorn and Hasselbalch, 2015; Di Buduo et al., 2015;

Schmitt-Graeff et al., 2015; Zhan et al., 2018). Despite many possible causes being discovered, the underlying mechanisms of MPN onset and progression are still not entirely understood, especially in myelofibrosis.

Patients with myelofibrosis generally have an overall poor prognosis, and the median survival period is about 6 years (Rumi and Cazzola, 2017). A series of thrombohemorrhagic complications and severe constitutional symptoms are presented during disease progression. Additionally, approximately 10–20% of patients with myelofibrosis may progress to leukemic transformation (Hong et al., 2019; Vallapureddy et al., 2019). Most patients with myelofibrosis died from thrombohemorrhagic complications, cardiovascular events, and leukemic transformation (Campanelli et al., 2021).

Current treatment for myelofibrosis is still limited. Hydroxyurea (HU) has been used as first-line treatment of PMF; however, the disease course was not improved (Reilly et al., 2012). *JAK2* inhibitors such as ruxolitinib could improve the clinical symptoms of myelofibrosis (Harrison et al., 2012; Harrison et al., 2016). Unfortunately, limited molecular responses were achieved (Wang et al., 2014). Also, there are problems of drug resistance or intolerance due to long-term treatment and side effects. Allogeneic HSC transplantation is the solely possibly curative remedy for patients with myelofibrosis; however, the potential complications are more severe and complex. Therefore, it is urgently necessary to explore novel targets of myelofibrosis with the potential value of early diagnosis, prediction of progression, and accurate treatment.

Nowadays, gene expression analysis is increasingly becoming a valuable method to identify the potential targets for early diagnosis, prediction of disease progression, and treatment. Here, in this study, two gene expression datasets [GSE103237 (Zini et al., 2017) and GSE53482 (Norfo et al., 2014)] of CD34[+] HSPCs from patients with *JAK2V617F*[+] MPN (PV, ET, PMF) and healthy controls (HC) were analyzed to identify important genes specifically dysregulated in *JAK2V617F*[+] PMF. Another two external datasets [GSE174060 (Baumeister et al., 2021) and GSE120362 (Schubert et al., 2019)] were applied to validate the results. The important genes specifically dysregulated in *JAK2V617F*[+] myelofibrosis might contribute to differential diagnosis and prediction of disease progression in myelofibrosis.

Materials and methods

Data collection

Two microarray datasets [GSE103237 (Zini et al., 2017) and GSE53482 (Norfo et al., 2014)] of patients with MPN and HC from the GEO database were retrieved. $JAK2V617F^+$ PV ($n = 26$), $JAK2V617F^+$ ET ($n = 17$), $JAK2V617F^+$ PMF ($n = 23$), and HC ($n = 30$) were selected for analysis. The gene expression profile from the peripheral blood/bone marrow CD34$^+$ cells was annotated by the data of the GPL13667 platform.

$JAK2V617F^+$ PV ($n = 8$), $JAK2V617F^+$ ET ($n = 2$), $JAK2V617F^+$ SMF ($n = 3$, including 2 post-PV-MF, 1 post-ET-MF), $JAK2V617F^+$ PMF ($n = 10$), and HC ($n = 6$), which came from GSE174060 (Baumeister et al., 2021) and GSE120362 (Schubert et al., 2019), were used for external validation. The gene expression profile from the peripheral blood/bone marrow CD34$^+$ cells was annotated by the data of the GPL17586 platform.

Data processing and differentially expressed gene screening

The gene expression data from two datasets [GSE103237 (Zini et al., 2017) and GSE53482 (Norfo et al., 2014)] were normalized by the limma package (version 3.50.0) (Ritchie et al., 2015) in R software. Then, "ComBat" from the sva package (version 3.42.0) was used to remove batch effects after these two datasets were merged. The merged expression matrix included samples from four groups: HC, $JAK2V617F^+$ PV, $JAK2V617F^+$ ET, and $JAK2V617F^+$ PMF. PCA plots were performed with the Factoextra package and FactoMineR package (Lê et al., 2008; Kassambara and Mund, 2020). Then, the limma package was used to process the expression matrix and identify the DEGs between different $JAK2V617F^+$ MPN subtypes (PV, ET, and PMF) and HC. The adjusted p-value less than 0.05 and |log2 fold change (logFC)| more than 1 were regarded as significant DEGs for downstream analysis. Heatmaps were performed with the pheatmap package (Kolde, 2019). Volcano plots were generated with the ggplot2 package (Wickham, 2016). Venn plots were made with the VennDiagram package (Chen, 2021).

Kyoto encyclopedia of genes and genomes and Reactome enrichment analysis

The clusterProfiler package (version 4.2.0) (Yu et al., 2012) was utilized to perform the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis to decipher the detailed functions of DEGs in biology and disease. The clusterProfiler package (Yu et al., 2012) and ReactomePA package (version 1.38.0) (Yu and He, 2016) were used to perform Reactome pathway analysis. The significantly enriched pathways were selected with the cut-off threshold of BH adjusted p-value less than 0.05. The bubble diagrams of KEGG results were drawn with the enrichplot package (version 1.14.1) (Yu, 2022).

Gene set enrichment analysis

GSEA software was utilized to run GSEA, using reference gene sets from the Molecular Signatures Database (Subramanian et al., 2005). The expression matrix of all genes from patients with MPN and HC was uploaded for analysis. The parameters of permutation number and permutation type were set as "1000" and "phenotype," respectively. NOM p-value less than 0.05 was regarded as statistically significant.

Analysis of protein–protein interaction networks and identification of hub genes
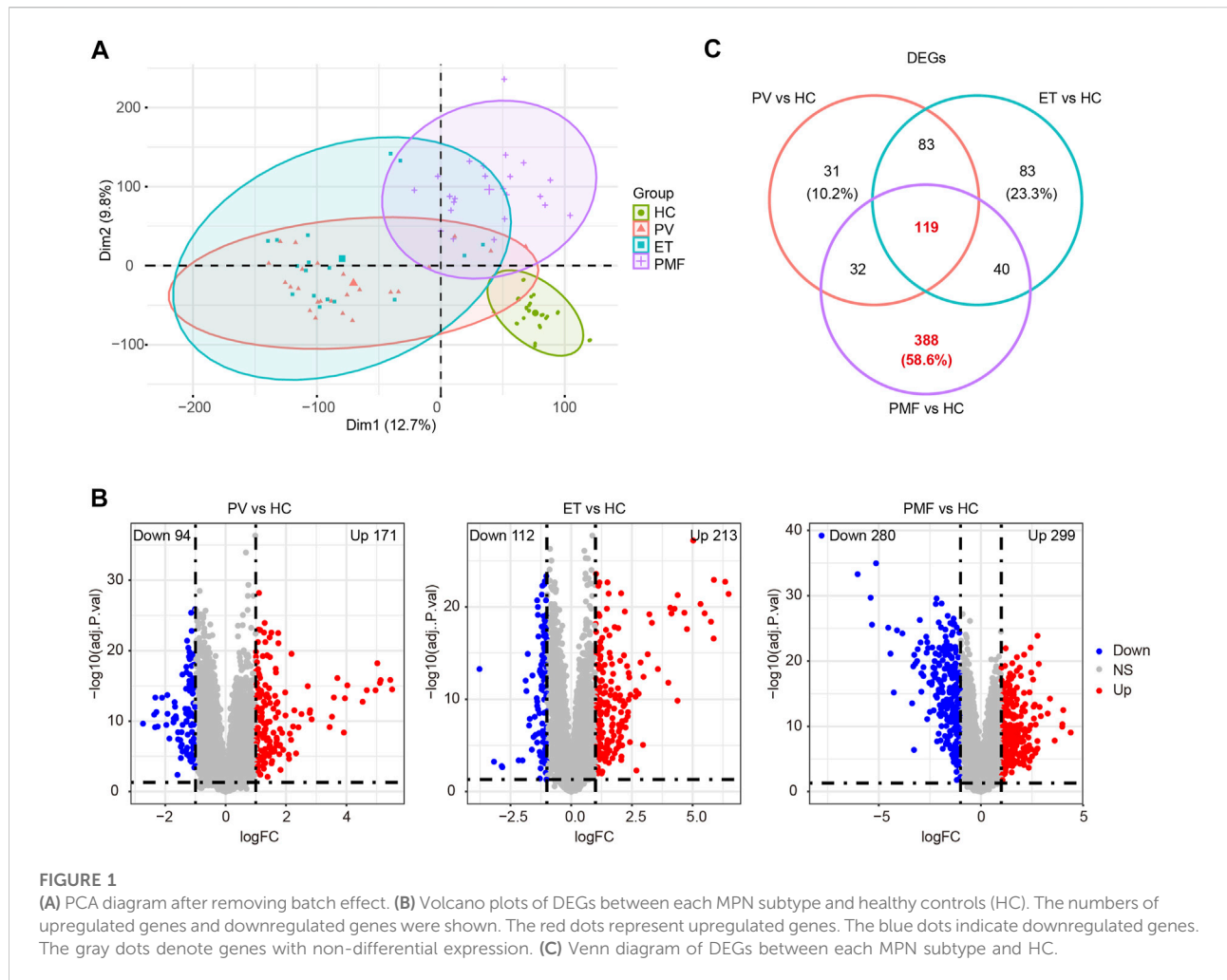
Protein–protein interaction (PPI) analysis of DEGs was constructed based on the STRING database (Szklarczyk et al., 2015). Cytoscape (Shannon et al., 2003) was utilized to further process the results from PPI networks. The cytoHubba app (Chin et al., 2014) was utilized for obtaining the hub DEGs (top ten DEGs) by the MCC algorithm, and the MCODE app (Bader and Hogue, 2003) was applied to identify the sub-networks (top two modules).

External validation of differentially expressed genes

Another two datasets [GSE174060 (Baumeister et al., 2021) and GSE120362 (Schubert et al., 2019)] from the GEO database were retrieved. Samples, including HC ($n = 6$), $JAK2V617F^+$ PV ($n = 8$), $JAK2V617F^+$ ET ($n = 2$), $JAK2V617F^+$ SMF ($n = 3$), and $JAK2V617F^+$ PMF ($n = 10$), were selected for external validation. The gene expression matrix of CD34$^+$ HSPCs from bone marrow/peripheral blood was annotated by the GPL17586 platform data. The pipeline of data processing is similar to the methods described above. The expression levels of DEGs between patients with MPN and HC were compared with Wilcoxon rank-sum test from the ggpubr package (version 0.4.0) (Kassambara, 2020). The p value less than 0.05 was regarded as having statistical significance. The GraphPad Prism (version 9.3.1) was used to calculate the value of AUC and draw the ROC curve.

Statistical analysis

R software and GraphPad Prism were utilized for data processing and statistical analysis. The ggpubr package in R software was applied to draw boxplots and perform statistical

FIGURE 1
(A) PCA diagram after removing batch effect. **(B)** Volcano plots of DEGs between each MPN subtype and healthy controls (HC). The numbers of upregulated genes and downregulated genes were shown. The red dots represent upregulated genes. The blue dots indicate downregulated genes. The gray dots denote genes with non-differential expression. **(C)** Venn diagram of DEGs between each MPN subtype and HC.

analysis. The Wilcoxon rank-sum test was applied to compare the difference in expression levels between distinct groups. The p value less than 0.05 at two-sided was regarded as statistically significant. "*" denotes "$p < 0.05$", "**" denotes "$p < 0.01$", "***" denotes "$p < 0.001$", "****" denotes "$p < 0.0001$", and "ns" denotes "not significant".

Results

CD34+ hematopoietic stem and progenitor cells from *JAK2V617F+* primary myelofibrosis exhibited a greater number of dysregulated genes than those from *JAK2V617F+* polycythemia vera and *JAK2V617F+* essential thrombocythemia

A total of 96 samples, including HC ($n = 30$), *JAK2V617F+* PV ($n = 26$), *JAK2V617F+* ET ($n = 17$), and *JAK2V617F+* PMF

($n = 23$), were collected from two datasets [GSE103237 (Zini et al., 2017) and GSE53482 (Norfo et al., 2014)] based on the same GPL platform. After removing the batch effect, the PCA result showed that the features of three *JAK2V617F+* MPN subtypes could be distinguishable from those of HC (Figure 1A), suggesting the signatures of HSPCs under disease state changed significantly compared to those in normal state. In addition, the features of *JAK2V617F+* PV and ET were similar to each other but distinct from *JAK2V617+* PMF (Figure 1A), which might be explained by that they were indeed in the different stages ("indolent" and "aggressive" stages, respectively) of MPN. Furthermore, we also noticed that the features from patients with three *JAK2V617F+* MPN entities were less uniform than those of HC (Figure 1A), indicating the heterogeneity of CD34+ HSPCs existed not only between different MPN subtypes but also in the same MPN subtype.

Next, the gene expression matrix of *JAK2V617F+* PV ($n = 26$) and HC ($n = 30$) was processed to discover the differentially expressed genes (DEGs). A total of 265 DEGs were identified

from the matrix, including 171 upregulated genes and 94 downregulated genes in patients with *JAK2V617F*[+] PV compared to HC (Figure 1B). A similar DEG screening procedure was also performed in *JAK2V617F*[+] ET (*n* = 17) and *JAK2V617F*[+] PMF (*n* = 23). A total of 213 upregulated DEGs and 112 downregulated DEGs were found in patients with *JAK2V617F*[+] ET relative to HC (Figure 1B), and 299 upregulated DEGs and 280 downregulated DEGs were obtained in patients with *JAK2V617F*[+] PMF relative to HC (Figure 1B). The number of dysregulated genes identified from PMF was much more abundant than that from PV and ET, indicating the more complex and severe disease course of PMF.

Integration of all DEGs between each MPN subtype and HC showed that among the different MPN subtypes, the DEG expression patterns from *JAK2V617F*[+] PV and ET patients were distinct from those from PMF patients (Supplementary Figures S1A–D). However, Partial patterns shared by the three entities were also observed (Supplementary Figures S1A–D). These shared and different signatures might provide important clues for the pathogenesis and heterogeneity of MPN. Therefore, the overlapping or unique DEGs derived from the integrated matrix were identified (Figure 1C and Supplementary Table S1). A total of 119 overlapping DEGs in patients with three different MPN subtypes were obtained, and 31 DEGs were merely acquired in *JAK2V617F*[+] PV, while 83 DEGs were uniquely obtained in *JAK2V617F*[+] ET (Figure 1C). Among three MPN entities, *JAK2V617F*[+] PMF had the greatest number of specific DEGs, with 388 DEGs (58.6%) exclusively dysregulated in PMF (Figure 1C). These specific dysregulated genes might play an essential part in the onset and progression of myelofibrosis.

In summary, the results corroborated different patterns of "mild and indolent subtype" and "aggressive and fibrotic subtype" in *JAK2V617F*[+] MPN, suggesting that the heterogeneity of HSPCs is obvious between different MPN subtypes and *JAK2V617F*[+] PMF might have more complicated pathogenesis than PV or ET.

CD34[+] hematopoietic stem and progenitor cells from *JAK2V617F*[+] primary myelofibrosis enriched a larger number of inflammatory pathways and immune response signaling pathways than those from *JAK2V617F*[+] polycythemia vera and *JAK2V617F*[+] essential thrombocythemia

The KEGG database was utilized to investigate the detailed function of DEGs and identify the shared and distinct enriched pathways between three MPN subtypes. Considering the upregulated genes were more potentially prominent in biological function than the downregulated genes, we focused on the upregulated genes in each MPN subtype compared to HC to perform the downstream pathway enrichment analysis. The results suggested that most upregulated genes were enriched

mainly in pathways correlated with immune response and inflammation. Some of the inflammatory pathways and immune response signaling pathways were enriched in three MPN subtypes, such as chemokine signaling pathway, NF-kappa B signaling pathway, NOD-like receptor signaling pathway, neutrophil extracellular trap formation, phagosome, as well as IL-17 signaling pathway (Figure 2), demonstrating that the inflammation and related immune response may participate in the pathogenesis of three MPN entities. However, the number of these enriched pathways in PMF was obviously more than that in PV and ET. For example, the TNF signaling pathway, toll-like receptor signaling pathway, and cytokine–cytokine receptor interaction were also enriched in PMF, but not in PV or ET (Figure 2). The results suggested that the more broad and severe state of inflammation and immune response might contribute to the aggressive progression of myelofibrosis.

To further discover the potential biological function of upregulated genes, we performed the venn diagram and identified the overlapping and unique upregulated genes among three MPN subtypes (Figure 3A and Supplementary Table S1). The results showed the number of upregulated genes uniquely in PMF was much more than that exclusively in PV or ET and even more than the number of overlapping genes (Figure 3A), suggesting that beyond some common factors underlying the pathogenesis, such as inflammation previously mentioned, there were additional elements leading to myelofibrosis. Considering the fundamental and essential roles of the common factors involved in the pathogenesis of three MPN subtypes, we studied these common characteristics before identifying the heterogeneous features. The overlapping upregulated genes (*n* = 78) were shown (Figure 3B), and KEGG analysis indicated that most of them were enriched in inflammatory pathways and immune response pathways, such as the NF-kappa B signaling pathway (Figure 3C). The expression levels of the relative gene set and the GSEA results between each MPN subtype and HC also corroborated the upregulation of the NF-kappa B signaling pathway in distinct MPN subtypes (Figure 3D, Supplementary Figure S2).

In summary, these results indicated that inflammation and immune response constitute the fundamental and essential parts of the pathogenesis in MPN, and the difference in degree and scope of them might be one of the explanations for more severe manifestation in myelofibrosis.

The overlapping upregulated genes were identified in distinct *JAK2V617F*[+] myeloproliferative neoplasm subtypes and primarily related to the inflammatory pathways and immune response signaling pathways

To resolve the interaction between proteins encoded by overlapping upregulated genes (*n* = 78) in three *JAK2V617F*[+]
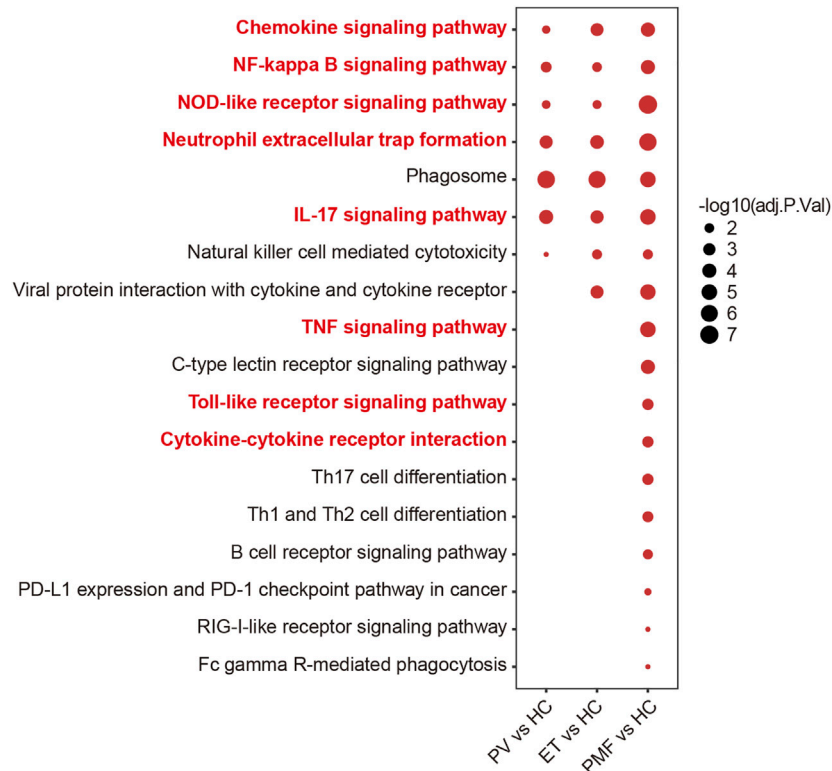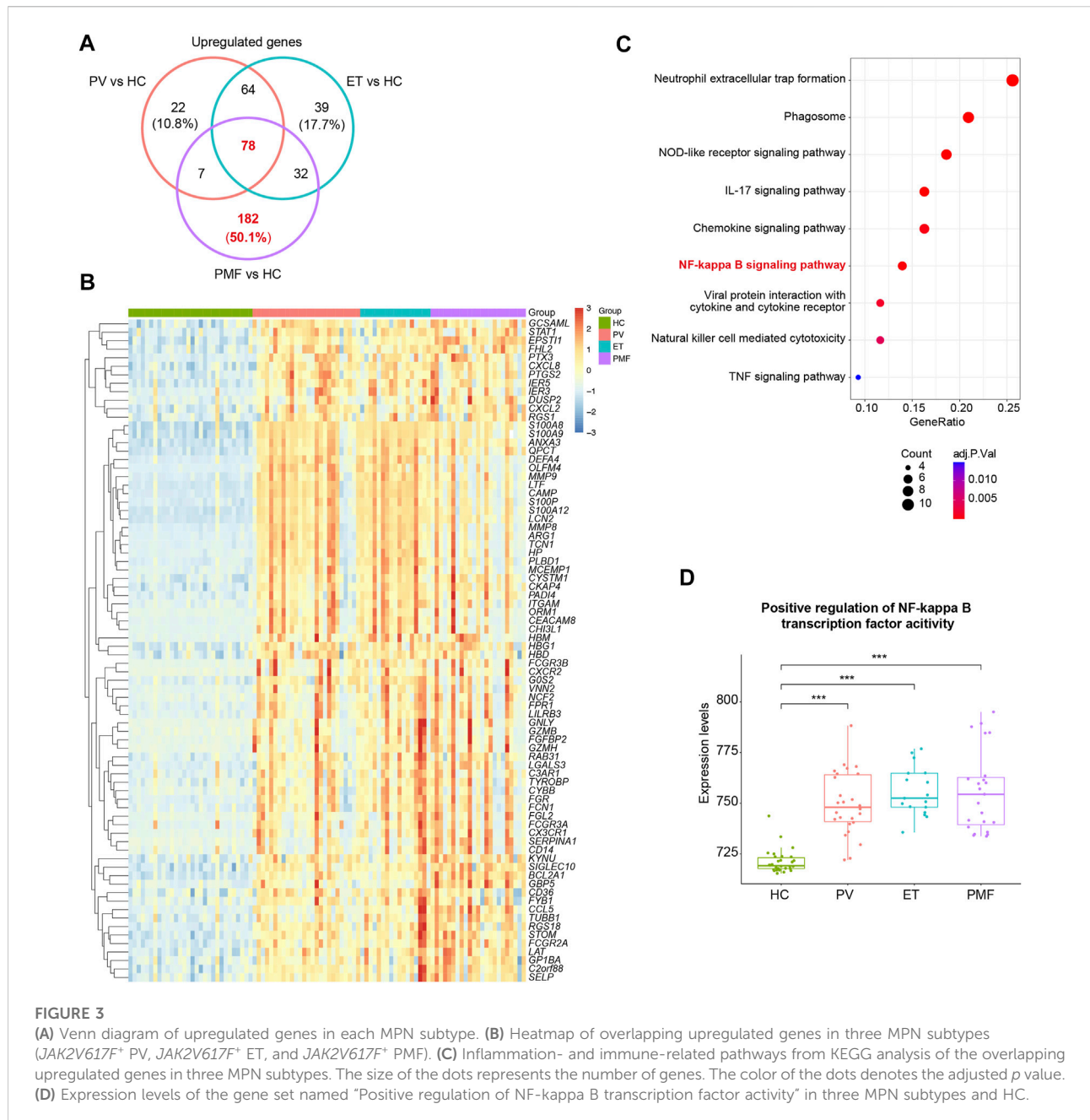
FIGURE 2
Inflammation- and immune-related pathways from KEGG analysis of upregulated genes in each MPN subtype compared to HC. The size of the dots represents −log10 (adjusted *p* value).

MPN subtypes, the STRING database was utilized to conduct the analysis. The results indicated that, in total, 405 edges (protein interaction pairs) and 78 nodes (encoded proteins) were included. Ten hub genes of overlapping upregulated genes in three *JAK2V617F*[+] MPN subtypes were obtained by the cytoHubba plugin from the PPI network (Figure 4A). Also, the hub genes, such as *ITGAM*, *FPR1*, Fc Gamma Receptor genes (*FCGR3A* and *FCGR3B*), and S100 family genes (*S100A8/9/12*), were related to inflammatory pathways and immune response signaling pathways (Figure 4A). In addition, two significant modules were acquired by the MCODE plugin, module 1 containing 12 nodes and 64 edges and module 2 containing 11 nodes and 30 edges, including genes related to immune response signaling pathways, such as *FPR1*, *NCF2*, *ARG1*, *CCL5*, and *HP* (Figure 4B). Furthermore, we found that the expression levels of some identified genes (*ITGAM*, *FPR1*, Fc Gamma Receptor genes, *TYROBP*, *NCF2*, *ARG1*, *CCL5*, and *HP*) in three *JAK2V617F*[+] MPN subtypes were significantly higher than those in HC (Figure 4C). And, most of their expression levels were also positively correlated with each other (Figure 4D).

Two external datasets [GSE174060 (Baumeister et al., 2021) and GSE120362 (Schubert et al., 2019)] derived from the same GPL platform were utilized and analyzed to verify the inflammation- and immune-related genes overlapping upregulated in different *JAK2V617F*[+] MPN subtypes, with CD34[+] HSPC samples of *JAK2V617F*[+] PV (*n* = 8), *JAK2V617F*[+] ET (*n* = 2), *JAK2V617F*[+] SMF (*n* = 3), *JAK2V617F*[+] PMF (*n* = 10), and HC (*n* = 6) included. Due to the paucity of CD34[+] samples in patients with *JAK2V617F*[+] ET, we integrated the samples of *JAK2V617F*[+] PV and *JAK2V617F*[+] ET into one group named *JAK2V617F*[+] PV/ET, reflecting the chronic and indolent stage of MPN. The expression levels of these inflammation- and immune-related genes (*ITGAM*, *FPR1*, Fc Gamma Receptor genes, *TYROBP*, *NCF2*, *ARG1*, *CCL5*, and *HP*) were significantly higher in both chronic and aggressive fibrotic phases of *JAK2V617F*[+] MPN than those in HC (Figure 5), which confirmed the results described above. Taken together, these results suggested that the upregulation of these genes associated with inflammation and immune response might play fundamental roles in the pathogenesis of MPN.

FIGURE 3
(A) Venn diagram of upregulated genes in each MPN subtype. **(B)** Heatmap of overlapping upregulated genes in three MPN subtypes (*JAK2V617F⁺* PV, *JAK2V617F⁺* ET, and *JAK2V617F⁺* PMF). **(C)** Inflammation- and immune-related pathways from KEGG analysis of the overlapping upregulated genes in three MPN subtypes. The size of the dots represents the number of genes. The color of the dots denotes the adjusted *p* value. **(D)** Expression levels of the gene set named "Positive regulation of NF-kappa B transcription factor activity" in three MPN subtypes and HC.

Interferon signaling pathways were upregulated significantly and specifically in CD34⁺ hematopoietic stem and progenitor cells of *JAK2V617F⁺* primary myelofibrosis

To further explore the potential detailed function of upregulated genes uniquely in PMF and identify the signaling pathways solely enriched in myelofibrosis, the Reactome database including more various signaling pathway

information was used to resolve the roles of upregulated genes in each MPN subtype compared to HC. The Reactome enrichment analysis identified the overlapping enriched pathways (*n* = 19) (Figure 6A, Supplementary Table S2), most of which were related to inflammation and immune response, indicating the identical conclusion to the analysis described above. More importantly, we identified the uniquely enriched signaling pathways (*n* = 21) in PMF (Figure 6A, Supplementary Table S2). Interestingly, we surprisingly found that interferon signaling pathways were enriched exclusively in CD34⁺ HSPCs of
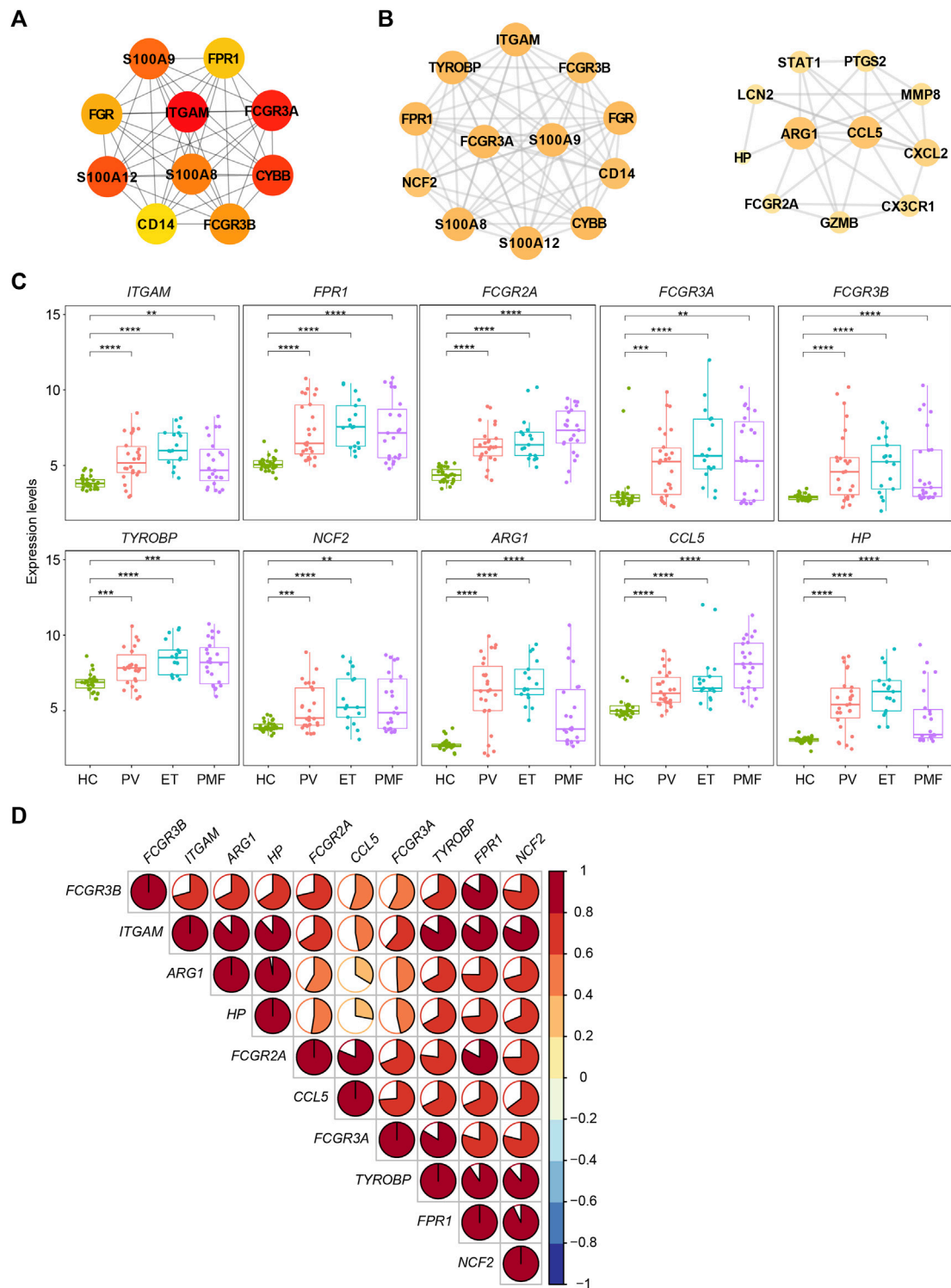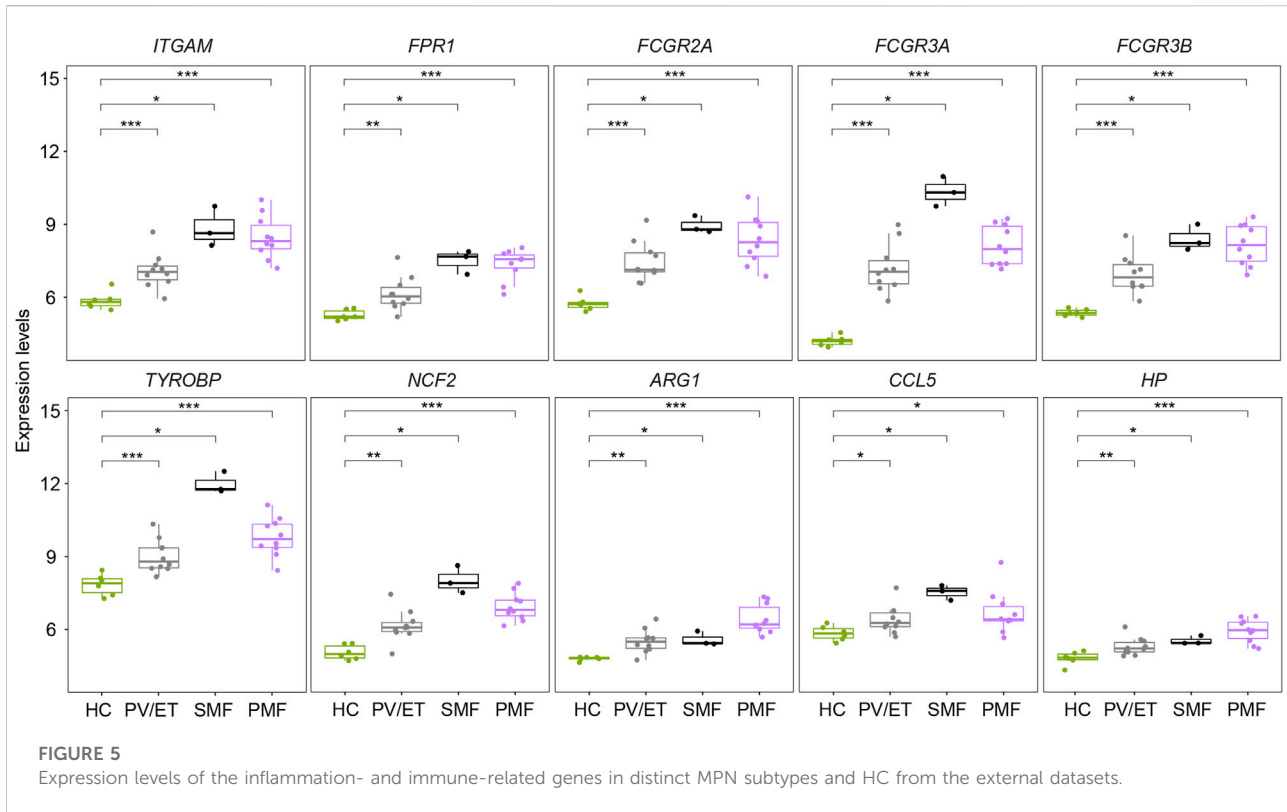
FIGURE 4
(A) Top 10 hub genes from the PPI network analysis for the overlapping upregulated genes in three MPN subtypes. Nodes denote encoded proteins, and edges denote the interaction between encoded proteins. The color represents the scores ranked by the MCC method. The deeper color denotes the more important genes having higher scores. **(B)** Top two significant modules from the PPI network analysis for the overlapping upregulated genes in three MPN subtypes. The color and size represent the degrees of the nodes. The bigger size and deeper color denote the more important genes having higher degrees. **(C)** Expression levels of the inflammation- and immune-related genes upregulated in three MPN subtypes. **(D)** Correlation of the expression levels of the upregulated genes in **(C)**.

FIGURE 5
Expression levels of the inflammation- and immune-related genes in distinct MPN subtypes and HC from the external datasets.

PMF, but not in those of PV or ET (Figure 6A). In addition, when we selected the upregulated genes (*n* = 182) uniquely in PMF to perform the Reactome enrichment analysis, interferon signaling pathways were also enriched significantly, such as interferon signaling, interferon-alpha/beta signaling, and interferon-gamma signaling (Figure 6B). Furthermore, the GSEA analysis was conducted, and the result also showed that multiple gene sets of interferon signaling pathways were enriched significantly and exclusively in CD34+ HSPCs of *JAK2V617F*+ PMF, but not in those of PV or ET (Figures 6C,D), corroborating the previous results. The expression levels of gene set "interferon alpha response" also verified that the upregulation of the interferon signaling pathway was significantly and specifically in *JAK2V617F*+ PMF (Figure 6E).

Therefore, the upregulation of interferon signaling pathways might participate in the pathogenesis of myelofibrosis.

Four interferon-stimulated genes (*OAS1, IFITM3, GBP1,* and *GBP2*) were upregulated specifically in CD34+ hematopoietic stem and progenitor cells of myelofibrosis

The PPI network of these upregulated DEGs (*n* = 182) exclusively in *JAK2V617F*+ PMF was constructed with

179 nodes (encoded proteins) and 322 edges (protein interaction pairs). Through the PPI network, the top 10 hub genes were discovered, namely, *MX1, IFIT1, IFIT3, OAS1, RSAD2, XAF1, MX2, ISG15, IFI6,* and *IRF9*, most of which were strongly associated with interferon-gamma signaling (Figure 7A). Two significant modules were also discovered (Figure 7B). Module 1 had 18 nodes and 153 edges, and module 2 contained 6 nodes and 15 edges, and the involved genes were almost related to the interferon signaling pathway (Figure 7B). Additionally, we found that the interferon-related genes (*OAS1, MX1, MX2, RSAD2, IFI6, IRF9, IFITM2, IFITM3, GBP1, GBP2,* and *IFI44L*) were upregulated in *JAK2V617F*+ PMF compared to not only HC but also *JAK2V617F*+ PV and ET (Figure 7C). Furthermore, there was a positive correlation in the expression levels between most genes (Figure 7D).

Two external datasets [GSE174060 (Baumeister et al., 2021) and GSE120362 (Schubert et al., 2019)] described above were used as validation datasets. Considering the additional three samples from patients with *JAK2V617F*+ SMF, we finally identified four genes (*OAS1, IFITM3, GBP1,* and *GBP2*) with increased expression levels not only in *JAK2V617F*+ PMF but also in *JAK2V617F*+ SMF in comparison with HC and PV/ET (Figure 8A). In addition, their expression levels are also positively correlated well with each other (Figure 8B).
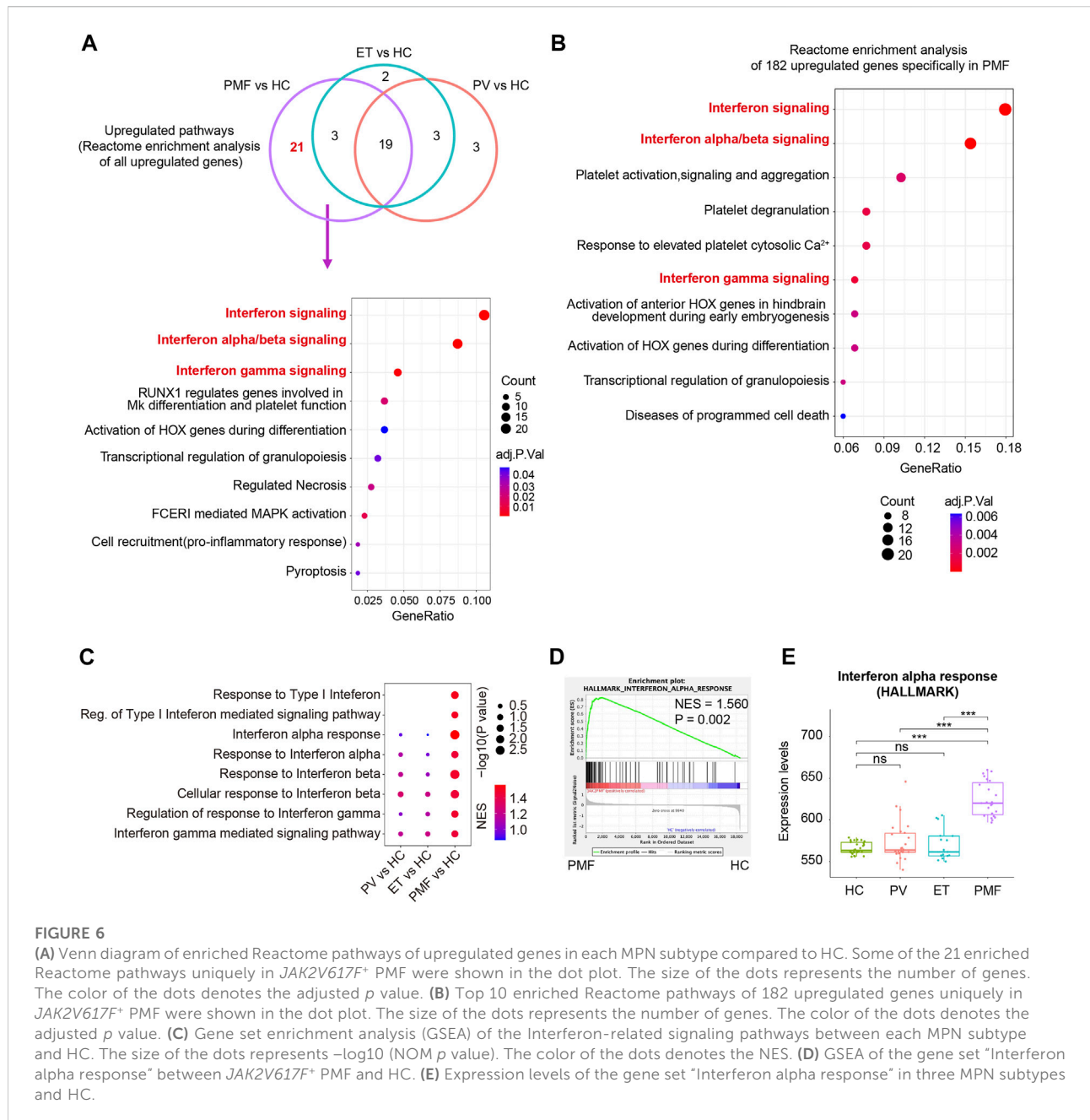
FIGURE 6
(A) Venn diagram of enriched Reactome pathways of upregulated genes in each MPN subtype compared to HC. Some of the 21 enriched Reactome pathways uniquely in *JAK2V617F*[+] PMF were shown in the dot plot. The size of the dots represents the number of genes. The color of the dots denotes the adjusted *p* value. **(B)** Top 10 enriched Reactome pathways of 182 upregulated genes uniquely in *JAK2V617F*[+] PMF were shown in the dot plot. The size of the dots represents the number of genes. The color of the dots denotes the adjusted *p* value. **(C)** Gene set enrichment analysis (GSEA) of the Interferon-related signaling pathways between each MPN subtype and HC. The size of the dots represents −log10 (NOM *p* value). The color of the dots denotes the NES. **(D)** GSEA of the gene set "Interferon alpha response" between *JAK2V617F*[+] PMF and HC. **(E)** Expression levels of the gene set "Interferon alpha response" in three MPN subtypes and HC.

To investigate the differential diagnostic ability of the four genes (*OAS1*, *IFITM3*, *GBP1*, and *GBP2*) for patients with myelofibrosis, the ROC curves and AUC were performed. In the ROC curves here, the true positive rate means the percentage of that a *JAK2V617F*[+] myelofibrosis patient is correctly predicted as *JAK2V617F*[+] myelofibrosis, while the false positive rate is the proportion of that a *JAK2V617F*[+] PV or ET individual is classified falsely as *JAK2V617F*[+] myelofibrosis. The AUC represents the area under the ROC curve, indicating the

accuracy and reliability of diagnostic prediction. The results showed high AUC values of the four genes (*OAS1*, *IFITM3*, *GBP1*, and *GBP2*) when distinguishing *JAK2V617F*[+] myelofibrosis from *JAK2V617F*[+] PV or ET (Figure 8C). Among the four genes, *OAS1* had the highest AUC value. Therefore, these four genes (*OAS1*, *IFITM3*, *GBP1*, and *GBP2*), especially *OAS1*, might possess the potential to be novel auxiliary diagnostic and predictive indicators of myelofibrosis, but further research is still necessary in the future.
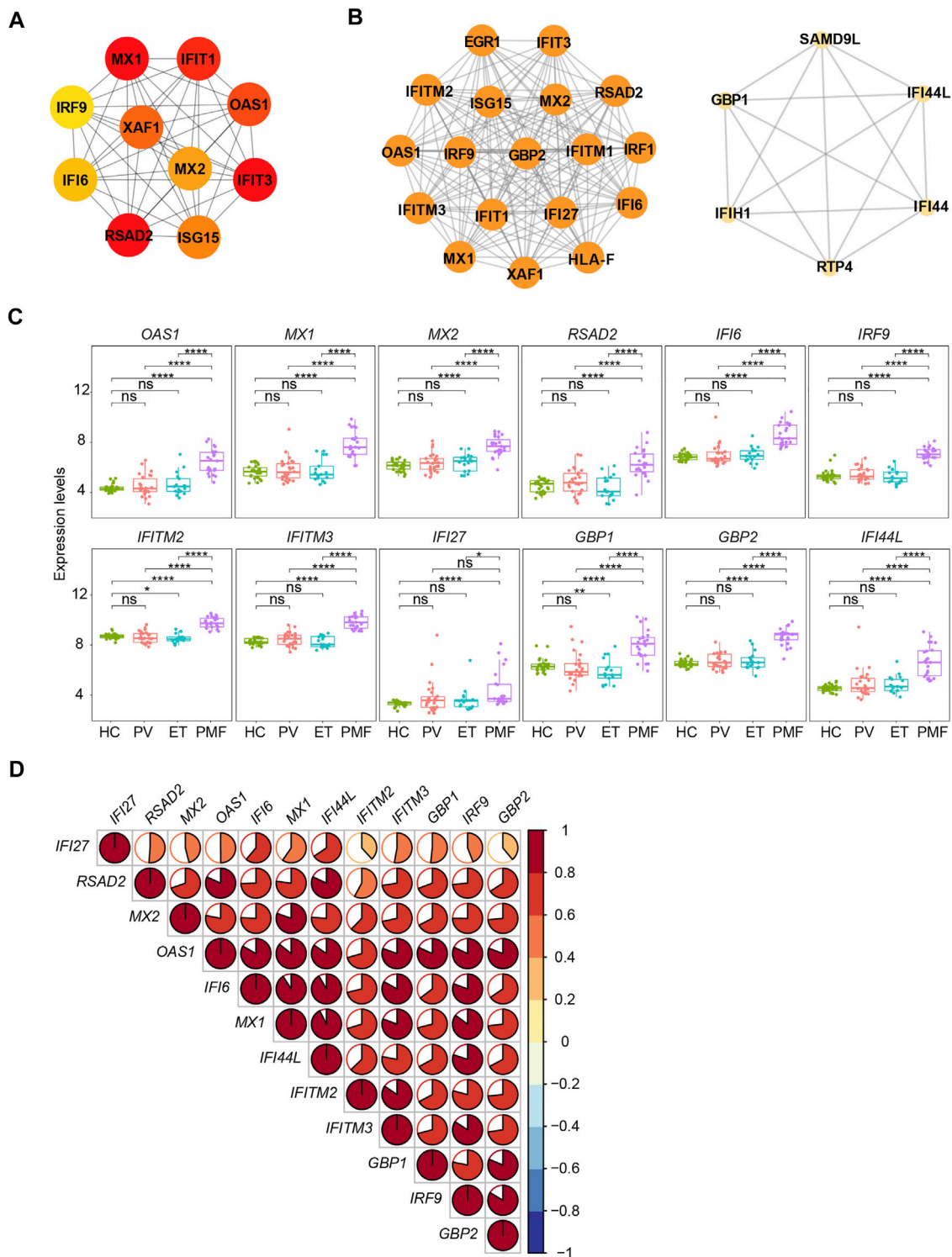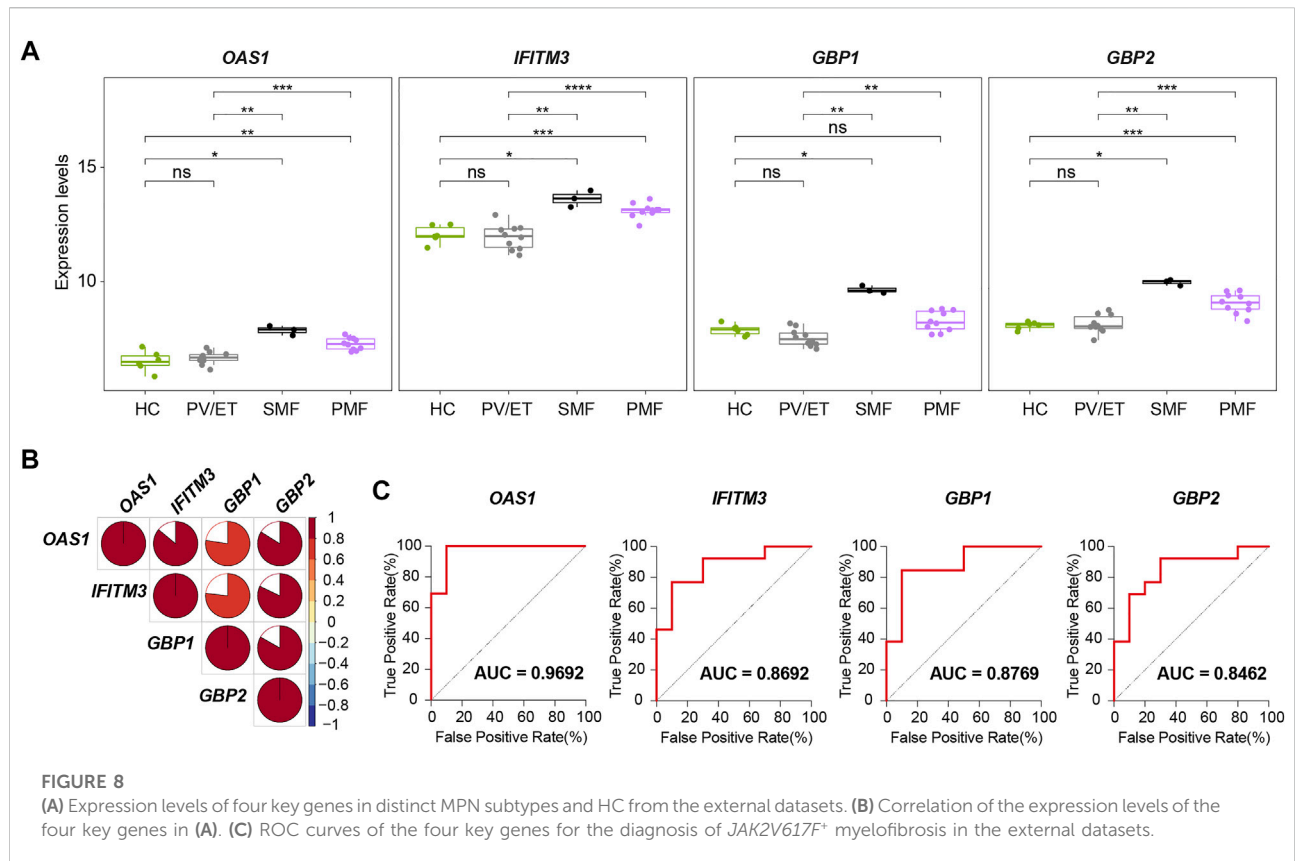
FIGURE 7
(A) Top 10 hub genes from the PPI network analysis for the upregulated genes uniquely in *JAK2V617F*[+] PMF. Nodes denote encoded proteins, and edges denote the interaction between two encoded proteins. The color represents the scores ranked by the MCC method. The deeper color denotes the more important genes having higher scores. **(B)** Top two significant modules from the PPI network analysis for the upregulated genes uniquely in *JAK2V617F*[+] PMF. The color and size represent the degrees of the nodes. The bigger size and deeper color denote the more important genes having higher degrees. **(C)** Expression levels of the key genes uniquely upregulated in *JAK2V617F*[+] PMF. **(D)** Correlation of the expression levels of genes in **(C)**.

FIGURE 8
(A) Expression levels of four key genes in distinct MPN subtypes and HC from the external datasets. **(B)** Correlation of the expression levels of the four key genes in **(A)**. **(C)** ROC curves of the four key genes for the diagnosis of *JAK2V617F*⁺ myelofibrosis in the external datasets.

Discussion

Although abundant research about Ph- MPNs has been performed, the mechanisms of disease onset and progression in myelofibrosis are still not fully understood. Considering there are significantly diverse manifestations and prognoses between the chronic stage and aggressive stage of MPN, it is urgently needed to identify some reliable biomarkers to predict myelofibrosis as early as possible, which could also contribute to early intervention and treatment. Here, we precisely analyzed the gene expression profiles in different subtypes of *JAK2V617F*⁺ MPN (including PV, ET, SMF, and PMF) to provide new information on pathogenesis and identify novel indicators for diagnosis and prediction of myelofibrosis.

Inflammation could be an important determinant that promotes the development and progression of MPN. Recent studies found that the acquisition of *JAK2V617F* in HSPCs, leading to the cytokine-independent activation of JAK-STAT signaling, may occur even during early childhood and *in utero*, and there is a very long asymptomatic period (Van Egeren et al., 2021; Williams et al., 2022). Chronic inflammation might be one required factor for transformation from asymptomatic clonal hematopoiesis to overt MPN or even for the progression from PV/ET to myelofibrosis. Studies about inflammatory cytokines indicated that the levels of IL6, IL2, and sIL2a were elevated during the progression from PV/ET to myelofibrosis (Panteli et al.,

2005). Here, we also observed the upregulated pathways and genes strongly associated with inflammation and immune response. For instance, the chemokine signaling pathway, NF-kappa B signaling pathway, as well as pro-inflammatory genes (such as S100 family genes) were upregulated in three *JAK2V617F*⁺ MPN subtypes, consistent with some published findings (Fisher et al., 2017; Fisher et al., 2019; Baumeister et al., 2021). Some inflammation- and immune-related upregulated genes (*ITGAM*, *FPR1*, Fc Gamma Receptor genes, *TYROBP*, *NCF2*, *ARG1*, *CCL5*, and *HP*), which were undetected in previously published studies, were also identified in distinct *JAK2V617F*⁺ MPN subtypes in comparison with HC. Additionally, among three subtypes, *JAK2V617F*⁺ PMF has more numbers of inflammation-related upregulated pathways and genes, suggesting that the more prominent inflammatory condition is a pronounced feature of myelofibrosis.

Furthermore, we identified four potentially important genes (*OAS1*, *IFITM3*, *GBP1*, and *GBP2*) that were upregulated uniquely in myelofibrosis. These genes belong to interferon-stimulated genes, related closely to interferon signaling (Schneider et al., 2014). Previous studies suggested that interferon (IFN) and IFN-stimulated genes might be involved in pulmonary fibrosis (Neville et al., 1997; Berkmana et al., 2001; Christmann et al., 2014). The low dose of IFN-α could promote bleomycin-induced lung fibrosis in mice and hamsters, though this effect was complicated and might also be influenced by IFN

preparations (Neville et al., 1997; Berkmana et al., 2001). The expression levels of interferon-regulated genes (OAS1 and IFI44) were found to be positively correlated with progressive lung fibrosis in patients with systemic sclerosis (SSc)–related interstitial lung disease (ILD) (Christmann et al., 2014). Additionally, a recent study demonstrated that a high expression level of IFITM3 could reflect the adverse prognosis in AML (Liu et al., 2020). However, the roles of IFN and interferon-stimulated genes in the pathogenesis of myelofibrosis are still largely unclear. Our previously published research showed that JAK2 mutant HSCs displayed increased IFN signaling, which might promote Mk (megakaryocyte)-biased hematopoiesis (Tong et al., 2021). There was another study showing that GBP2 overexpression suppressed the erythroid differentiation and increased the level of matrix metalloproteinase-9 in TF-1 cells (Lin et al., 2013), which was also found to be increased in patients with idiopathic myelofibrosis (Xu et al., 2005). These results suggested the characteristics including thrombocythemia, anemia, and even bone marrow failure in myelofibrosis might be related to the increased IFN signaling, but more evidence should be required in the future. Overall, the four interferon-stimulated genes (OAS1, IFITM3, GBP1, and GBP2) exclusively upregulated in myelofibrosis might not only provide important novel clues to the MPN field but also offer special insights into the effects of IFN signaling on the pathogenesis of myelofibrosis.

Inevitably, this study has limitations. Though quite large numbers of samples and repeated validations from multiple datasets were applied to minimalize the individual variation, this study was limited by the lack of detailed clinical information in these GEO datasets such as variant allele frequency (VAF) of JAK2V617F, other somatic mutations, disease duration, and treatment, which remains to be further investigated.

In summary, our study precisely analyzed the gene expression profiles of CD34[+] HSPCs across the different JAK2V617F[+] MPN subtypes. The gene expression signatures in JAK2V617F[+] PMF were to a large extent distinct from those in JAK2V617F[+] PV or ET, although there were also similar features among them. The upregulation of inflammation- and immune-related signature genes and especially IFN signaling might be involved in the pathogenesis and progression of myelofibrosis. Additionally, the four genes (OAS1, IFITM3, GBP1, and GBP2) we identified in this study might have the potential to be auxiliary diagnostic and predictive indicators of myelofibrosis, but further investigations are still necessary in the future.

Data availability statement

Publicly available datasets were analyzed in this study. The datasets presented in this study can be found in GEO database (https://www.ncbi.nlm.nih.gov/geo/). The accession numbers of the datasets can be found in the article/Supplementary Material.

Author contributions

YZ analyzed the data and wrote the manuscript. DW and YL reviewed and revised the manuscript. JT, LS, and CX guided the process of analysis and writing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.927018/full#supplementary-material

SUPPLEMENTARY FIGURE S1
(A) Heatmap of all DEGs between each MPN subtype and HC. The blue color indicates the downregulated genes, and the red color denotes the upregulated genes. **(B)** Heatmap of top 20 upregulated genes between JAK2V617F[+] PV and HC. **(C)** Heatmap of top 20 upregulated genes between JAK2V617F[+] ET and HC. **(D)** Heatmap of top 20 upregulated genes between JAK2V617F[+] PMF and HC.

SUPPLEMENTARY FIGURE S2
GSEA of the gene set "Positive regulation of NF-kappa B transcription factor activity" between each MPN subtype and HC.

SUPPLEMENTARY TABLE S1
(A) DEGs commonly or uniquely dysregulated in JAK2V617F[+] PV, JAK2V617F[+] ET or JAK2V617F[+] PMF. **(B)** DEGs commonly or uniquely upregulated in JAK2V617F[+] PV, JAK2V617F[+] ET or JAK2V617F[+] PMF.

SUPPLEMENTARY TABLE S2
Reactome pathways commonly or uniquely enriched in JAK2V617F[+] PV, JAK2V617F[+] ET, or JAK2V617F[+] PMF.

References

Anderson, L. A., and McMullin, M. F. (2014). Epidemiology of MPN: what do we know? *Curr. Hematol. Malig. Rep.* 9 (4), 340–349. doi:10.1007/s11899-014-0228-z

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., et al. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127 (20), 2391–2405. doi:10.1182/blood-2016-03-643544

Ayalew, Tefferi, Lasho, Terra L., Finke, Christy M., Elala, Y., Hanson, C. A., Ketterling, R. P., et al. (2016). Targeted deep sequencing in primary myelofibrosis. *Blood Adv.* 1 (2), 105–111. doi:10.1182/bloodadvances.2016000208

Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinforma.* 4, 2. doi:10.1186/1471-2105-4-2

Baumeister, J., Maié, T., Chatain, N., Gan, L., Weinbergerova, B., de Toledo, M. A. S., et al. (2021). Early and late stage MPN patients show distinct gene expression profiles in CD34(+) cells. *Ann. Hematol.* 100 (12), 2943–2956. doi:10.1007/s00277-021-04615-8

Baxter, E. J., Scott, L. M., Campbell, P. J., East, C., Fourouclas, N., Swanton, S., et al. (2005). Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 365 (9464), 1054–1061. doi:10.1016/S0140-6736(05)71142-9

Berkmana, N., Kremera, S., Orb, R., Lossos, I., Christensen, T., Goldstein, R., et al. (2001). Human recombinant Interferon-·2a and interferon-·A/D have different effects on bleomycin-induced lung injury. *Respiration* 68 (2), 169–177. doi:10.1159/000050488

Bjorn, M. E., and Hasselbalch, H. C. (2015). The role of reactive Oxygen species in myelofibrosis and related neoplasms. *Mediat. Inflamm.* 2015, 648090. doi:10.1155/2015/648090

Campanelli, R., Massa, M., Rosti, V., and Barosi, G. (2021). New markers of disease progression in myelofibrosis. *Cancers (Basel)* 13 (21), 5324. doi:10.3390/cancers13215324

Chen, H. (2022). *VennDiagram: Generate high-resolution Venn and euler plots*. Available at: https://CRAN.R-project.org/package=VennDiagram.

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., Lin, C. Y., et al. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8 (Suppl. 4), S11. doi:10.1186/1752-0509-8-S4-S11

Christmann, R. B., Sampaio-Barros, P., Stifano, G., Borges, C. L., de Carvalho, C. R., Kairalla, R., et al. (2014). Association of Interferon- and transforming growth factor beta-regulated genes and macrophage activation with systemic sclerosis-related progressive lung fibrosis. *Arthritis Rheumatol.* 66 (3), 714–725. doi:10.1002/art.38288

Di Buduo, C. A., Wray, L. S., Tozzi, L., Malara, A., Chen, Y., Ghezzi, C. E., et al. (2015). Programmable 3D silk bone marrow niche for platelet generation *ex vivo* and modeling of megakaryopoiesis pathologies. *Blood* 125 (14), 2254–2264. doi:10.1182/blood-2014-08-595561

Fisher, D. A. C., Malkova, O., Engle, E. K., Miner, C. A., Fulbright, M. C., Behbehani, G. K., et al. (2017). Mass cytometry analysis reveals hyperactive NF Kappa B signaling in myelofibrosis and secondary acute myeloid leukemia. *Leukemia* 31 (9), 1962–1974. doi:10.1038/leu.2016.377

Fisher, D. A. C., Miner, C. A., Engle, E. K., Hu, H., Collins, T. B., Zhou, A., et al. (2019). Cytokine production in myelofibrosis exhibits differential responsiveness to JAK-STAT, MAP kinase, and NFκB signaling. *Leukemia* 33 (8), 1978–1995. doi:10.1038/s41375-019-0379-y

Grinfeld, J., Nangalia, J., Baxter, E. J., Wedge, D. C., Angelopoulos, N., Cantrill, R., et al. (2018). Classification and personalized prognosis in myeloproliferative neoplasms. *N. Engl. J. Med.* 379 (15), 1416–1430. doi:10.1056/NEJMoa1716614

Harrison, Claire, Kiladjian, Jean-Jacques, HaifaAl-Ali, Kathrin, Gisslinger, H., Waltzman, R., Stalbovskaya, V., et al. (2012). JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N. Engl. J. Med.* 366 (9), 787–798. doi:10.1056/NEJMoa1110556

Harrison, C. N., Vannucchi, A. M., Kiladjian, J. J., Al-Ali, H. K., Gisslinger, H., Knoops, L., et al. (2016). Long-term findings from COMFORT-II, a phase 3 study of ruxolitinib vs best available therapy for myelofibrosis. *Leukemia* 30 (8), 1701–1707. doi:10.1038/leu.2016.148

Hasselbalch, H. C. (2013). Chronic inflammation as a promotor of mutagenesis in essential thrombocythemia, polycythemia vera and myelofibrosis. A human inflammation model for cancer development? *Leuk. Res.* 37 (2), 214–220. doi:10.1016/j.leukres.2012.10.020

Hong, J., Lee, J. H., Byun, J. M., Lee, J. Y., Koh, Y., Shin, D. Y., et al. (2019). Risk of disease transformation and second primary solid tumors in patients with myeloproliferative neoplasms. *Blood Adv.* 3 (22), 3700–3708. doi:10.1182/bloodadvances.2019000655

Kassambara, A. (2020). *ggpubr: 'ggplot2' based publication ready plots*. Available at: https://CRAN.R-project.org/package=ggpubr.

Kassambara, A., and Mundt, F. (2020). Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Available at: https://CRAN.R-project.org/package=factoextra.

Kolde, R. (2019). *pheatmap: Pretty heatmaps*. Available at: https://CRAN.R-project.org/package=pheatmap.

Kralovics, R., Passamonti, F., Buser, A. S., Teo, S. S., Tiedt, R., Passweg, J. R., et al. (2005). A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N. Engl. J. Med.* 352 (17), 1779–1790. doi:10.1056/NEJMoa051113

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for multivariate analysis. *J. Stat. Softw.* 25 (1), 1–18. doi:10.18637/jss.v025.i01

Levine, R. L., Wadleigh, M., Cools, J., Ebert, B. L., Wernig, G., Huntly, B. J. P., et al. (2005). Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 7 (4), 387–397. doi:10.1016/j.ccr.2005.03.023

Li, B., Zhang, L., Bai, J., Xu, Z., Qin, T., Zhang, Y., et al. (2016). Non-driver mutations profile identified by a 206-gene NGS panel in patients with primary myelofibrosis and post-polycythaemic/essential Thrombocythaemia myelofibrosis in a single center from China. *Blood* 128 (22), 1942. doi:10.1182/blood.v128.22.1942.1942

Lin, X., Rice, K. L., Buzzai, M., HExnEr, E., Costa, F. F., Kilpivaara, O., et al. (2013). miR-433 is aberrantly expressed in myeloproliferative neoplasms and suppresses hematopoietic cell growth and differentiation. *Leukemia* 27 (2), 344–352. doi:10.1038/leu.2012.224

Liu, Y., Lu, R., Cui, W., Pang, Y., Liu, C., Cui, L., et al. (2020). High IFITM3 expression predicts adverse prognosis in acute myeloid leukemia. *Cancer Gene Ther.* 27 (1-2), 38–44. doi:10.1038/s41417-019-0093-y

Mesa, R. A., Verstovsek, S., Cervantes, F., Barosi, G., Reilly, J. T., Dupriez, B., et al. (2007). Primary myelofibrosis (PMF), post polycythemia vera myelofibrosis (post-PV MF), post essential thrombocythemia myelofibrosis (post-ET MF), blast phase PMF (PMF-BP): Consensus on terminology by the international working group for myelofibrosis research and treatment (IWG-MRT). *Leuk. Res.* 31 (6), 737–740. doi:10.1016/j.leukres.2006.12.002

Neville, Berkman., Goldstein., Ronald H., and Breuer, R. (1997). Bleomycin-induced lung injury is enhanced by interferon-alpha. *Life Sci.* 60 (26), PL415–421. doi:10.1016/s0024-3205(97)00332-9

Norfo, R., Zini, R., Pennucci, V., Bianchi, E., Salati, S., Guglielmelli, P., et al. (2014). miRNA-mRNA integrative analysis in primary myelofibrosis CD34+ cells: role of miR-155/JARID2 axis in abnormal megakaryopoiesis. *Blood* 124 (13), e21–32. doi:10.1182/blood-2013-12-544197

O'Sullivan, J., and Mead, A. J. (2019). Heterogeneity in myeloproliferative neoplasms: Causes and consequences. *Adv. Biol. Regul.* 71, 55–68. doi:10.1016/j.jbior.2018.11.007

Panteli, K. E., Hatzimichael, E. C., Bouranta, P. K., Katsaraki, A., Seferiadis, K., Stebbing, J., et al. (2005). Serum interleukin (IL)-1, IL-2, sIL-2Ra, IL-6 and thrombopoietin levels in patients with chronic myeloproliferative diseases. *Br. J. Haematol.* 130 (5), 709–715. doi:10.1111/j.1365-2141.2005.05674.x

Reilly, J. T., McMullin, M. F., Beer, P. A., Butt, N., Conneally, E., Duncombe, A., et al. (2012). Guideline for the diagnosis and management of myelofibrosis. *Br. J. Haematol.* 158 (4), 453–471. doi:10.1111/j.1365-2141.2012.09179.x

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Rumi, E., and Cazzola, M. (2017). Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms. *Blood* 129 (6), 680–692. doi:10.1182/blood-2016-10-695957

Schepers, K., Pietras, E. M., Reynaud, D., Flach, J., Binnewies, M., Garg, T., et al. (2013). Myeloproliferative neoplasia remodels the endosteal bone marrow niche into a self-reinforcing leukemic niche. *Cell Stem Cell* 13 (3), 285–299. doi:10.1016/j.stem.2013.06.009

Schmitt-Graeff, A. H., Nitschke, R., and Zeiser, R. (2015). The hematopoietic niche in myeloproliferative neoplasms. *Mediat. Inflamm.* 2015, 347270. doi:10.1155/2015/347270

Schneider, W. M., Chevillotte, M. D., and Rice, C. M. (2014). Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* 32, 513–545. doi:10.1146/annurev-immunol-032713-120231

Schubert, C., Allhoff, M., Tillmann, S., Maie, T., Costa, I. G., Lipka, D. B., et al. (2019). Differential roles of STAT1 and STAT2 in the sensitivity of JAK2V617F- vs.

BCR-ABL-positive cells to interferon alpha. *J. Hematol. Oncol.* 12 (1), 36. doi:10. 1186/s13045-019-0722-9

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., 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 (43), 15545–15550. doi:10.1073/pnas.0506580102

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43 (Database issue), D447–D452. doi:10.1093/nar/gku1003

Tong, J., Sun, T., Ma, S., Zhao, Y., Ju, M., Gao, Y., et al. (2021). Hematopoietic stem cell heterogeneity is linked to the initiation and Therapeutic response of myeloproliferative neoplasms. *Cell Stem Cell* 28 (3), 502–513.e6. e6. doi:10.1016/j.stem.2021.01.018

Vallapureddy, R. R., Mudireddy, M., Penna, D., Lasho, T. L., Finke, C. M., Hanson, C. A., et al. (2019). Leukemic transformation among 1306 patients with primary myelofibrosis: risk factors and development of a predictive model. *Blood Cancer J.* 9 (2), 12. doi:10.1038/s41408-019-0175-y

Van Egeren, D., Escabi, J., Nguyen, M., Liu, S., Reilly, C. R., Patel, S., et al. (2021). Reconstructing the lineage histories and differentiation Trajectories of individual cancer cells in myeloproliferative neoplasms. *Cell Stem Cell* 28 (3), 514–523.e9. e519. doi:10.1016/j.stem.2021.02.001

Wang, X., Ye, F., Tripodi, J., Hu, C. S., Qiu, J., Najfeld, V., et al. (2014). JAK2 inhibitors do not affect stem cells present in the spleens of patients with myelofibrosis. *Blood* 124 (19), 2987–2995. doi:10.1182/blood-2014-02-558015

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer.

Williams, N., Lee, J., Mitchell, E., Moore, L., Baxter, E. J., Hewinson, J., et al. (2022). Life histories of myeloproliferative neoplasms inferred from phylogenies. *Nature* 602 (7895), 162–168. doi:10.1038/s41586-021-04312-6

Xu, M., Bruno, E., Chao, J., Huang, S., Finazzi, G., Fruchtman, S. M., et al. (2005). Constitutive mobilization of CD34+ cells into the peripheral blood in idiopathic myelofibrosis may be due to the action of a number of proteases. *Blood* 105 (11), 4508–4515. doi:10.1182/blood-2004-08-3238

Yu, G. enrichplot: Visualization of functional enrichment result. Available at: https://yulab-smu.top/biomedical-knowledge-mining-book/. 2022.

Yu, G., and He, Q. Y. (2016). ReactomePA: an R/bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* 12 (2), 477–479. doi:10. 1039/c5mb00663e

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhan, H., Lin, C. H. S., Segal, Y., and Kaushansky, K. (2018). The JAK2V617F-bearing vascular niche promotes clonal expansion in myeloproliferative neoplasms. *Leukemia* 32 (2), 462–469. doi:10.1038/leu. 2017.233

Zini, R., Guglielmelli, P., Pietra, D., Rumi, E., Rossi, C., Rontauroli, S., et al. (2017). CALR mutational status identifies different disease subtypes of essential thrombocythemia showing distinct expression profiles. *Blood Cancer J.* 7 (12), 638. doi:10.1038/s41408-017-0010-2