



## OPEN ACCESS

## EDITED BY

Abhijeet R. Sonawane,  
Brigham and Women's Hospital and  
Harvard Medical School, United States

## REVIEWED BY

Cadiele Oliana Reichert,  
Federal University of Santa Catarina, Brazil  
Mehdi Pirooznia,  
Johnson & Johnson, United States  
Zhijun Meng,  
Shanxi Provincial People's Hospital, China

## \*CORRESPONDENCE

Jixuan Zou  
✉ 18811532579@163.com

<sup>†</sup>These authors have contributed equally to  
this work and share first authorship

RECEIVED 20 May 2024

ACCEPTED 25 July 2024

PUBLISHED 12 August 2024

## CITATION

Wang Z and Zou J (2024) Potential biomarkers  
and immune characteristics for polycythemia  
vera-related atherosclerosis using bulk RNA  
and single-cell RNA datasets: a combined  
comprehensive bioinformatics and machine  
learning analysis.

Front. Cardiovasc. Med. 11:1426278.  
doi: 10.3389/fcvm.2024.1426278

## COPYRIGHT

© 2024 Wang and Zou. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC  
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Potential biomarkers and immune characteristics for polycythemia vera-related atherosclerosis using bulk RNA and single-cell RNA datasets: a combined comprehensive bioinformatics and machine learning analysis

Ziqing Wang<sup>1†</sup> and Jixuan Zou<sup>2\*†</sup>

<sup>1</sup>Beijing Friendship Hospital, Capital Medical University, Beijing, China, <sup>2</sup>Graduate School, Beijing University of Chinese Medicine, Beijing, China

**Background:** Polycythemia vera (PV) is a myeloproliferative disease characterized by significantly higher hemoglobin levels and positivity for JAK2 mutation. Thrombosis is the main risk event of this disease. Atherosclerosis (AS) can markedly increase the risk of arterial thrombosis in patients with PV. The objectives of our study were to identify potential biomarkers for PV-related AS and to explore the molecular biological association between PV and AS.

**Methods:** We extracted microarray datasets from the Gene Expression Omnibus (GEO) dataset for PV and AS. Common differentially expressed genes (CGs) were identified by differential expression analysis. Functional enrichment and protein-protein interaction (PPI) networks were constructed from the CG by random forest models using LASSO regression to identify pathogenic genes and their underlying processes in PV-related AS. The expression of potential biomarkers was validated using an external dataset. A diagnostic nomogram was constructed based on potential biomarkers to predict PV-related AS, and its diagnostic performance was assessed using ROC, calibration, and decision curve analyses. Subsequently, we used single-cell gene set enrichment analysis (GSEA) to analyze the immune signaling pathways associated with potential biomarkers. We also performed immune infiltration analysis of AS with "CIBERSORT" and calculated Pearson's correlation coefficients for potential biomarkers and infiltrating immune cells. Finally, we observed the expression of potential biomarkers in immune cells based on the single-cell RNA dataset.

**Results:** Fifty-two CGs were identified based on the intersection between up-regulated and down-regulated genes in PV and AS. Most biological processes associated with CGs were cytokines and factors associated with chemotaxis of immune cells. The PPI analysis identified ten hub genes, and of these, CCR1 and MMP9 were selected as potential biomarkers with which to construct a diagnostic model using machine learning methods and external dataset validation. These biomarkers could regulate Toll-like signaling, NOD-like signaling, and chemokine signaling pathways associated with AS. Finally, we determined that these potential biomarkers had a strong correlation with macrophage M0 infiltration. Further, the potential biomarkers were highly expressed in macrophages from patients with AS.

**Conclusion:** We identified two CGs (CCR1 and MMP9) as potential biomarkers for PV-related AS and established a diagnostic model based on them. These results may provide insight for future experimental studies for the diagnosis and treatment of PV-related AS.

#### KEYWORDS

atherosclerosis, biomarkers, immune, polycythemia vera, machine learning analysis

## 1 Introduction

Polycythemia vera (PV) is a myeloproliferative disorder characterized by significantly increased hemoglobin levels and JAK2 gene mutation. PV is the most common myeloproliferative neoplasm, with an incidence of 10.9 per million. It is often accompanied by hematological features such as increased white blood cell (WBC) and platelet (PLT) counts, and clinical manifestations such as thrombosis, headache, gastrointestinal ulcer, pulmonary hypertension, and splenomegaly (1–3). Thrombosis is the main risk event for this disease, with an incidence of approximately 46%, and arterial thrombosis occurs 2–3 times more frequently than venous thrombosis (4, 5). Arterial thrombosis-related cardiovascular and cerebrovascular disorders pose a major risk to patient safety and quality of life, as well as an economic and public health burden to society. Atherosclerosis (AS) can significantly increase the risk of arterial thromboembolism in patients with PV, according to recent research (6). To prevent and reduce the incidence of thrombosis in patients with PV, it is crucial to evaluate the impact of AS in these individuals.

Patients with PV have a higher risk of AS and plays a significant role in arterial thrombosis in patients with PV. The growth and rupture of an atherosclerotic plaque is frequently a prerequisite for arterial thrombosis, acting as a prepathological stage of the disease. Arterial thrombosis may result from JAK2 mutation-induced leukocytosis and immunological dysfunction (7–9). Furthermore, the expression of cytokines by bone marrow cells in patients with PV can interfere with endothelial system homeostasis and further supports the potential that endothelial damage in patients with PV leads to AS (10–12). Although numerous studies have shown an intricate connection between PV and AS, additional studies are required to fully understand the implications and underlying biological processes between the two diseases (13–15).

Our hypothesis was that the pathogenic mechanisms underlying PV-related AS could be unraveled at the molecular biological level by employing bulk-RNA and single-cell RNA datasets and bioinformatics approach to thoroughly investigate abnormally expressed genes of the two diseases. In this study, we first explored the potential cellular and molecular pathways involved and thoroughly analyzed the association between PV and AS based on bulk-RNA and single-cell RNA datasets using various bioinformatics tools. We used a variety of advanced statistical algorithms to identify potential biomarkers of PV and AS and assessed their interaction and infiltrating immune cells. Additionally, the latent value of potential biomarkers in the

diagnosis of disease was evaluated and validated in different cohorts (Figure 1).

## 2 Materials and methods

### 2.1 Data collection

Six human gene expression profile datasets were obtained from the GEO databases (<https://www.ncbi.nlm.nih.gov/geo/>) (16). Bulk transcriptome data of controls and patients with PV were contributed by GSE26049 (whole blood from 41 PV individuals and 21 control individuals), GSE61629 (whole blood from 21 PV individuals and 21 control individuals), and GSE103237 (bone marrow CD34+ cells from 21 PV individuals and 21 control individuals). Bulk transcriptome data of the control and AS patients was contributed by GSE100927 (35 control arteries and 69 atherosclerotic lesions) and GSE43292 (32 control arteries and 32 atherosclerotic lesions). Single-cell transcriptome data of AS patients were extracted from GSE159677 (3 AS patients).

### 2.2 Analysis of differentially expressed genes

Integrated PV expression data was obtained by the batch correction of GSE26049 and GSE61629 based on the “SVA” package (17) in R software (version 4.2.1), which contained 62 PV samples and 42 control samples. Gene symbol conversion, background correction, and normalization were performed on the integrated PV and the AS data sets (GSE100927). Subsequently, the differentially expressed genes (DEGs) in the PV and AS datasets were identified with the “Limma” package (18) in R software. The DEGs were filtered with a fold change  $\geq 1.5$  and  $p.adjust < 0.05$ . Next, DEGs were summarized as volcano plots and heatmaps with the “ggplot2” and “pheatmap” packages in R software. As a result, the up- and down-regulated genes of PV and AS were intersected with the “dplyr” package in R software to identify CG.

### 2.3 Enrichment analysis of CGs from patients with AS and PV

Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes, Genomes (KEGG) analyses were performed to identify possible functions for the CGs from patients with AS and PV by

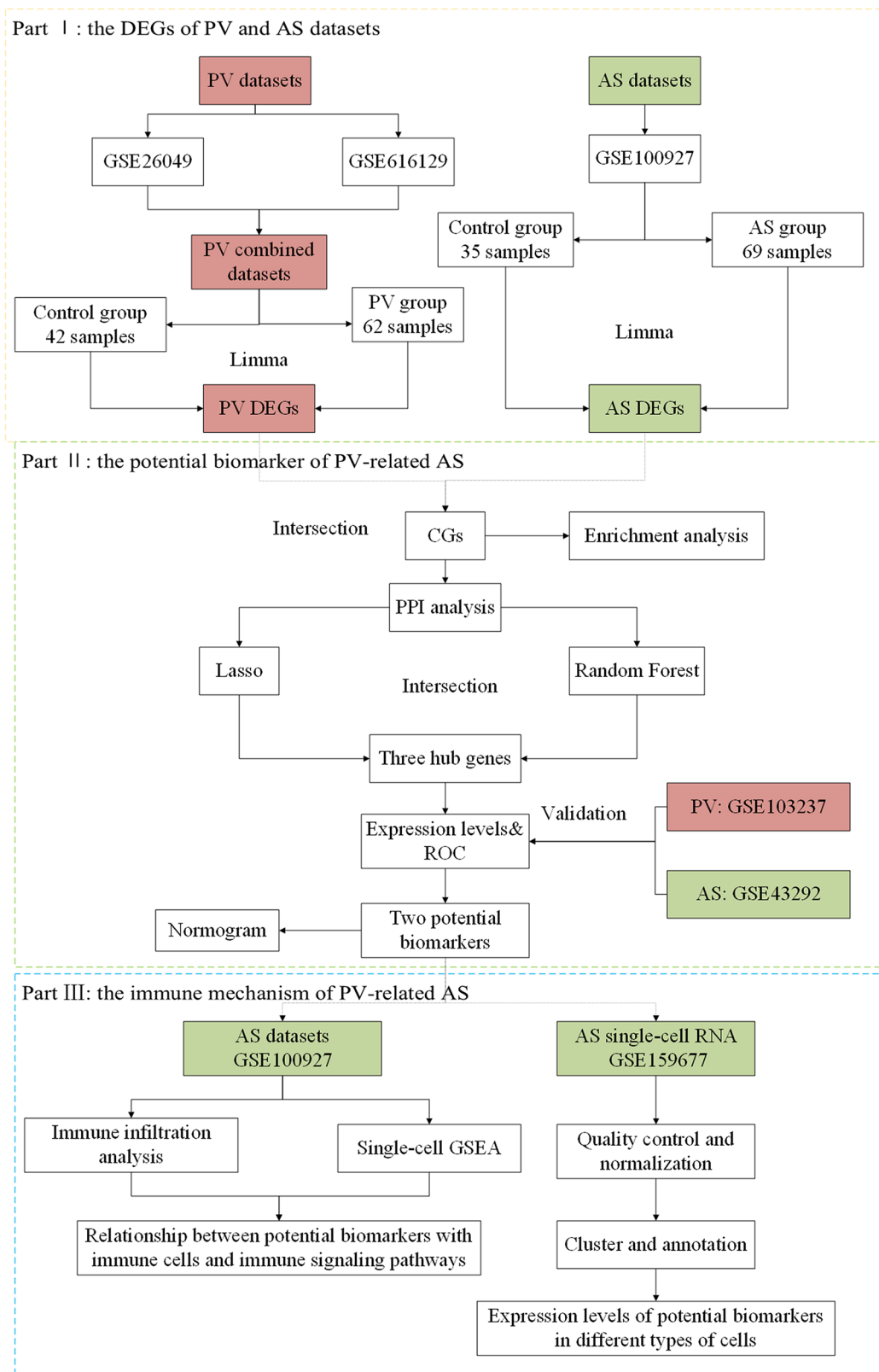


FIGURE 1  
Flow chart of this study design.

“clusterProfiler” packages (19) in the R software. A  $P$ -value  $<0.05$  was considered a significant statistical difference in the GO and KEGG analysis. Visualizations of GO and KEGG analysis were depicted with bar plot and chord diagrams by “ggplot2” package in the R software.

## 2.4 Construction of a protein-protein interaction network

To identify interactions among the CGs from patients with AS and PV, a protein-protein interaction (PPI) network was established on the basis of STRING data (<https://www.string-db.org>) (20), with a medium confidence score of  $>0.4$ . Later, the PPI network was visualized using Cytoscape software (version 3.8.0). We also used the Cytoscape plug-in molecular complex detection (MCODE) to detect the significant modules. The CGs in the modules with the highest scores were chosen for further analysis.

## 2.5 Machine learning

The least absolute shrinkage and selection operator (LASSO) method was used to screen for possible biomarkers using the “glmnet” package (21) in R software. Subsequently, the random forest (RF) modelling was used to further refine potential biomarkers using the “randomForest” package (22) in R software. The overlapping genes from the LASSO model and the RF model (MeanDecreaseGini  $>2$ ) were defined as hub genes for the establishment of a diagnostic model of PV-related AS.

## 2.6 Expression of hub genes and evaluation of receiver operating characteristic

The expression patterns of the hub gene were first evaluated in the GSE100927 and integrated PV datasets, and then confirmed in the GSE103237 and GSE43292 datasets. The Wilcoxon test was used for comparison with a significance level of  $P < 0.05$ . ROC curves were produced to evaluate the diagnostic value of the hub gene for the diagnosis of PV and AS, respectively. AUC values  $>0.7$  suggested a substantial difference. The AUC values and their accompanying 95% confidence intervals were calculated to separate the illness group from the control group.

## 2.7 Nomogram creation and the potential marker prediction model evaluation

The nomogram was created based on the hub genes using the “rms” package in R software. The performance of each hub gene and the nomogram in the diagnosis of AS was assessed by the area under the ROC curve. Furthermore, the ROC curve analysis was performed to assess the suitability of the nomogram-based diagnosis of AS. Lastly, the prediction efficiency of nomograms

in PV-related AS was evaluated with calibration curves and decision curve analysis (DCA).

## 2.8 Implementation of gene set enrichment analysis (GSEA) for potential biomarkers

After acquiring potential biomarkers, we used the “clusterProfiler” (23) program to perform single-gene gene set enrichment analysis (GSEA) for every potential biomarker in AS datasets. Our objective was to analyze the signaling pathways regulated by the potential biomarkers in AS with GSEA. The MSigDB (c5.go.bp.v7.5.1.entrez.gmt) was used to obtain gene sets. A  $P$ -value  $<0.05$  was considered statistically significant. The immunological pathways for every gene in the AS datasets were represented using the “enrichplot” package in R software.

## 2.9 Immune infiltration analysis

The AS gene expression profile was used to determine the amount of immune cell infiltration based on “CIBERSORT” program (24). A bar plot representing the amount and percentage of immune infiltration was displayed for every sample with the “ggplot2” package. The proportions of 22 different immune cell types in the calcified and control aortic valve samples were compared using “kruskal-test”. A  $P$ -value  $<0.05$  was considered statistically significant, and the results were displayed in a stacked histogram created with the “ggplot2” software. Finally, the association between the expression of potential biomarkers and the amount of immune cells infiltration was then analyzed using the Pearson’s rank correlation coefficient, with  $P < 0.05$  regarded as statistically significant. A coefficient of correlation  $>0.7$  was considered a strong correlative.

## 2.10 Single-cell RNA analysis

We used the “Seurat” and “SingleR” packages (25) to analyze AS single-cell RNA datasets. To preserve high-quality data, we excluded cells with a mitochondrial gene percentage  $>5\%$ , cell counts  $<3$ , and cells expressing  $<300$  and  $>5,000$  genes. We used the “NormalizeData” function to normalize the gene expression of the included cells. Principal component analysis (PCA) was used to extract the top 20 principal components (PC) based on the top 2,000 highly variable genes. Finally, the “FindVariableFeatures” function was used to save these PCs for additional analysis. The “FindNeighbors”, “FindClusters” (resolution = 0.6), and “RunUMAP” programs were used to cluster cell subpopulations in an unsupervised and unbiased manner. The marker genes for each cluster were screened using the “FindAllMarkers” function with fold change  $\geq 1.2$  and  $p.adjust < 0.05$ . Lastly, the “SingleR” package annotated cell types. Finally, “Featureplot” function was used to show the expression of potential biomarkers in the type of cells in AS.

### 3 Results

#### 3.1 Identification of CGs in PV and AS datasets

GSE26049 and GSE61629 were combated after batch correction (Figures 2A,B), the differences between two data sets were significantly decreased after batch effect removal. In the PV datasets (GSE26049 and GSE61629), a total of 535 DEGs, consisting of 319 upregulated DEGs and 216 downregulated DEGs, were identified (Figure 2C). In the AS data set GSE100927, a total of 1,687 DEGs, consisting of 1,067 upregulated DEGs and 620 downregulated DEGs, were identified (Figure 2D). The heatmaps show the top 50 expression pattern of CGs in PV and AS cohorts (Figures 2E,F). As shown in Figures 2G,H, there were 48 overlapping upregulated CGs and 4 overlapping downregulated CGs between PV and AS cohorts.

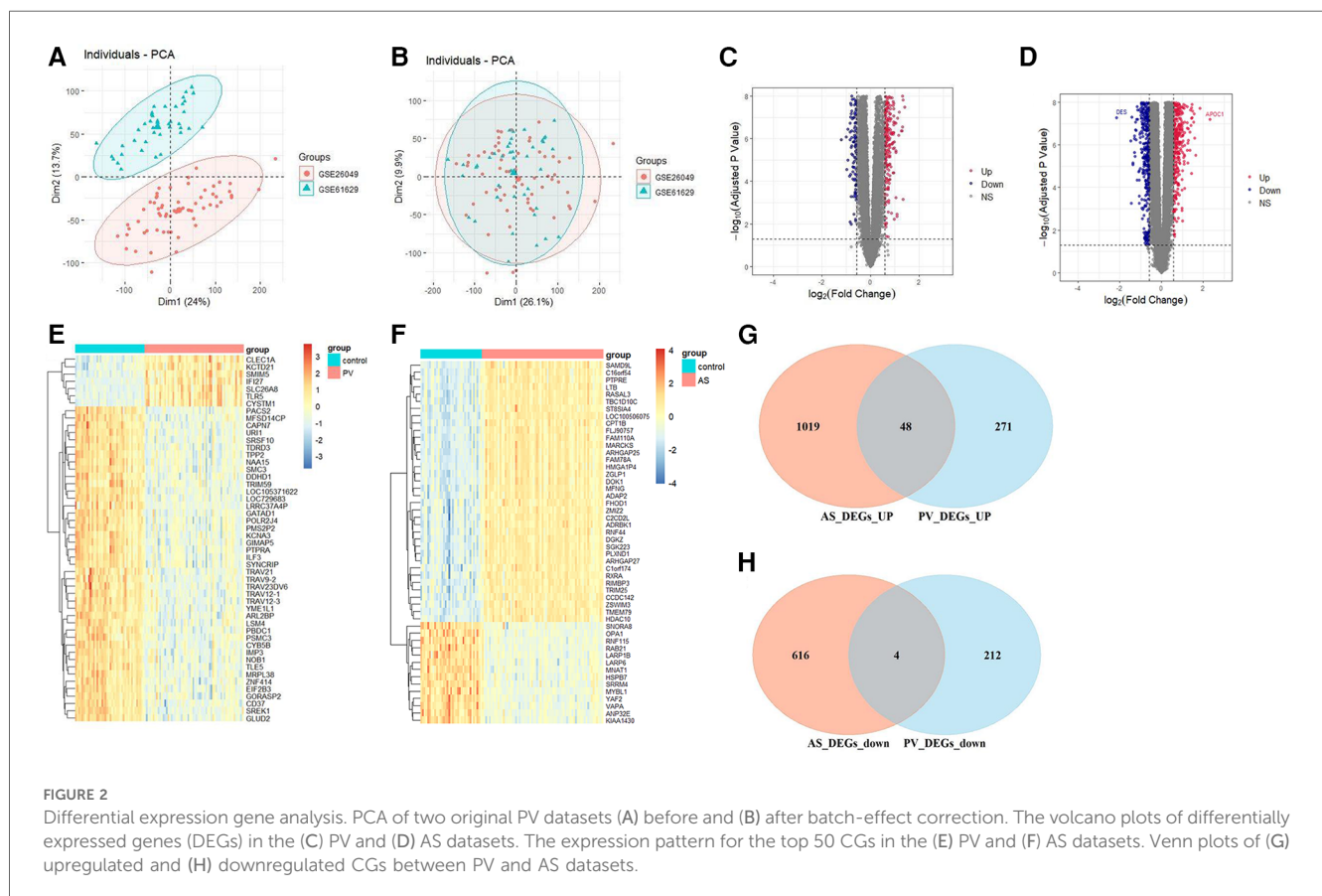
#### 3.2 Biological function and pathways of identified CGs

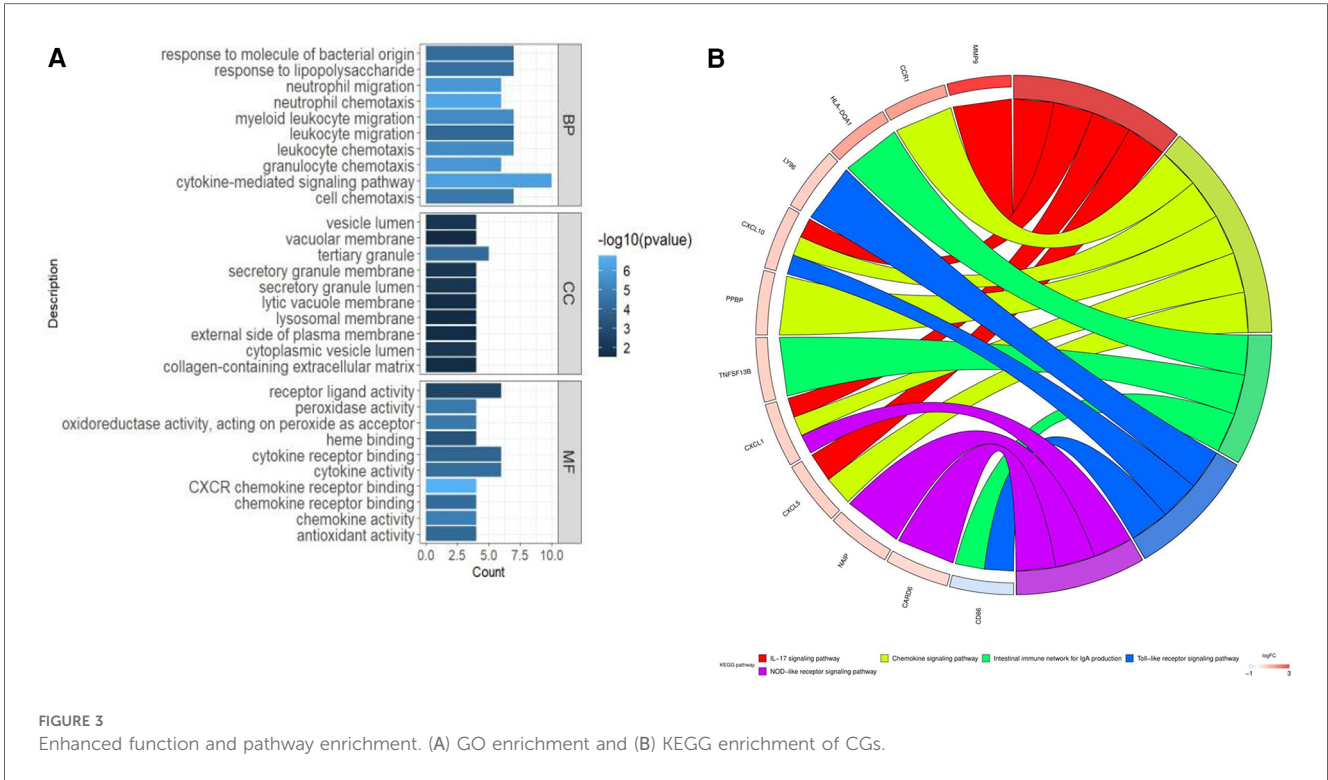
The biological role of CGs was examined using GO and KEGG enrichment analysis. The enriched biological processes (BPs) of CGs included neutrophil migration, neutrophil chemotaxis, myeloid leukocyte migration, leukocyte migration, leukocyte chemotaxis, granulocyte chemotaxis, and cytokine-mediated signaling pathways,

and cell chemotaxis (Figure 3A). The enriched molecular function (MF) of CGs included receptor ligand activity, peroxidase activity, oxidoreductase activity, acting on peroxide as acceptor, heme binding, cytokine receptor binding, cytokine activity, CXCR chemokine receptor binding, chemokine receptor binding, chemokine activity, and antioxidant activity (Figure 3A). The enriched cell component (CC) of CGs included the vesicle lumen, vacuolar membrane, tertiary granule, secretory granule membrane, secretory granule lumen, lytic vacuole membrane, lysosomal membrane, and external side of plasma membrane (Figure 3A). The immunological and inflammatory pathways in which CGs were the most enriched were chemokine signaling, IL-17, Toll-like receptor, and NOD-like receptor signaling pathways, and the intestinal immune network for IgA production (Figure 3B).

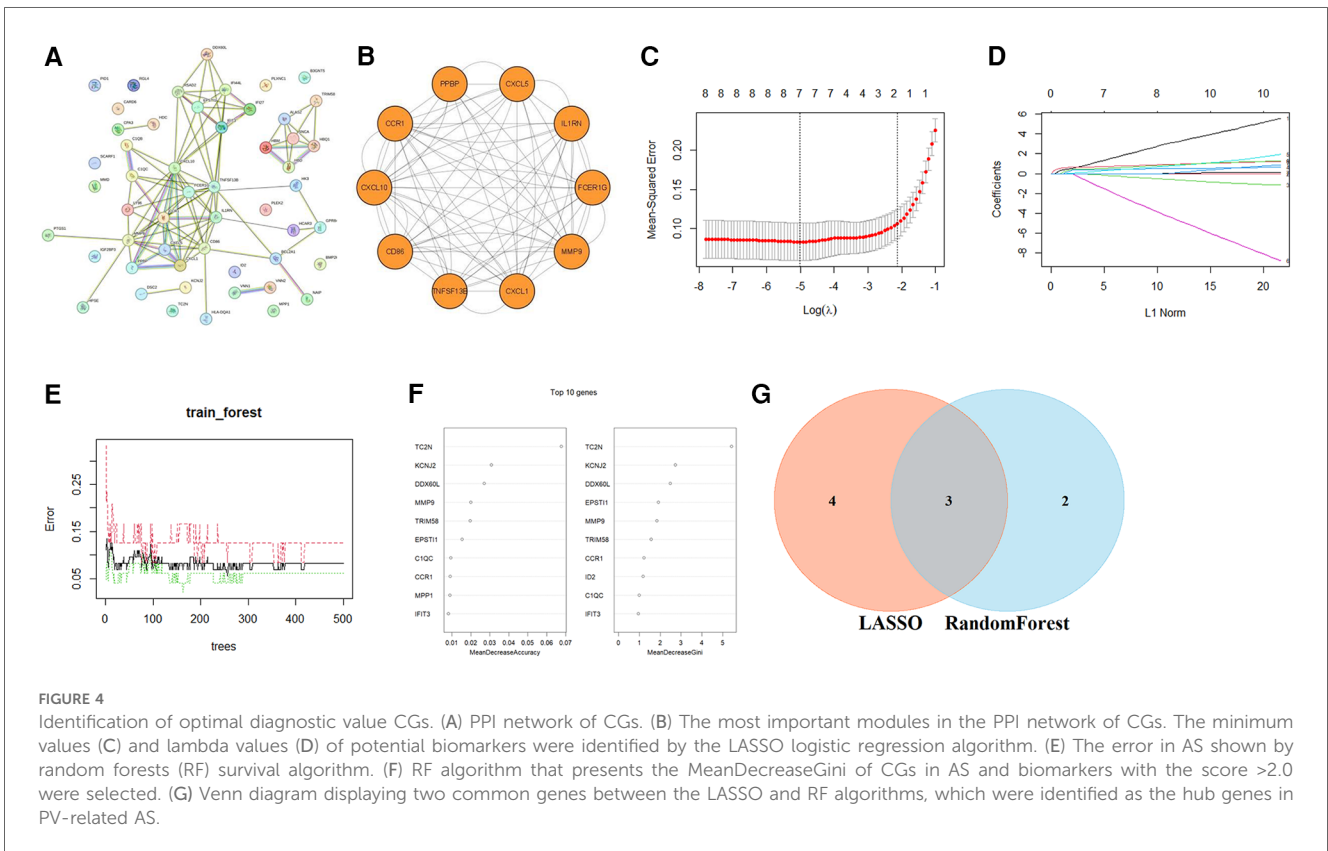
#### 3.3 Identification of optimal diagnostic value CGs

To reveal the potential pathogenic genes and underlying mechanisms in PV-related AS, the interaction of the key genes in PV and AS was collected by the STRING database to identify probable pathogenic genes and the underlying mechanism in PV-related AS. The most important modules were identified through MCODE, which included the 10 genes that have been recognized as pathogenic genes in PV-related AS. (Figures 4A,B). Subsequently, seven possible candidate genes among ten CG were identified





**FIGURE 3** Enhanced function and pathway enrichment. (A) GO enrichment and (B) KEGG enrichment of CGs.



**FIGURE 4** Identification of optimal diagnostic value CGs. (A) PPI network of CGs. (B) The most important modules in the PPI network of CGs. The minimum values (C) and lambda values (D) of potential biomarkers were identified by the LASSO logistic regression algorithm. (E) The error in AS shown by random forests (RF) survival algorithm. (F) RF algorithm that presents the MeanDecreaseGini of CGs in AS and biomarkers with the score >2.0 were selected. (G) Venn diagram displaying two common genes between the LASSO and RF algorithms, which were identified as the hub genes in PV-related AS.

with the LASSO regression algorithm, which had a significant impact on the diagnosis of patients with PV who presented AS (Figures 4C,D). Five CGs were retrieved using the RF machine

learning technique, which was also used to rank the ten CGs according to the varying relevance of each gene to further narrow the potential biomarkers (Figures 4E,F). Three potential biomarkers

overlapped in both groups after superposing the seven candidate genes from LASSO and the five probable genes from RF modelling, including motif chemokine receptor 1 (CCR1), C-X-C motif chemokine ligand 5 (CXCL5), and matrix metalloproteinase 9 (MMP9); all potential biomarkers were all up-regulated (Figure 4G).

Subsequently, we examined the expression of potential biomarkers in patient groups from the AS dataset (GSE100927) and PV datasets (combined of GSE26049 and GSE61629) and compared them with control groups. The expression of potential biomarkers was significantly different between the patient and control groups ( $P < 0.05$ ), and the ROC analysis supported the potential biomarkers as the most promising diagnostic marker for this condition (Figures 5A–D). Because the AUC of CXCL5 in GSE43292 was  $< 0.7$ , CXCL5 was not suitable as a biomarker. Furthermore, the results of the external validation analysis demonstrated a significant difference in the expression of CCR1 and MMP9 ( $P < 0.05$ ) between the patient and control groups (Figures 5E–H).

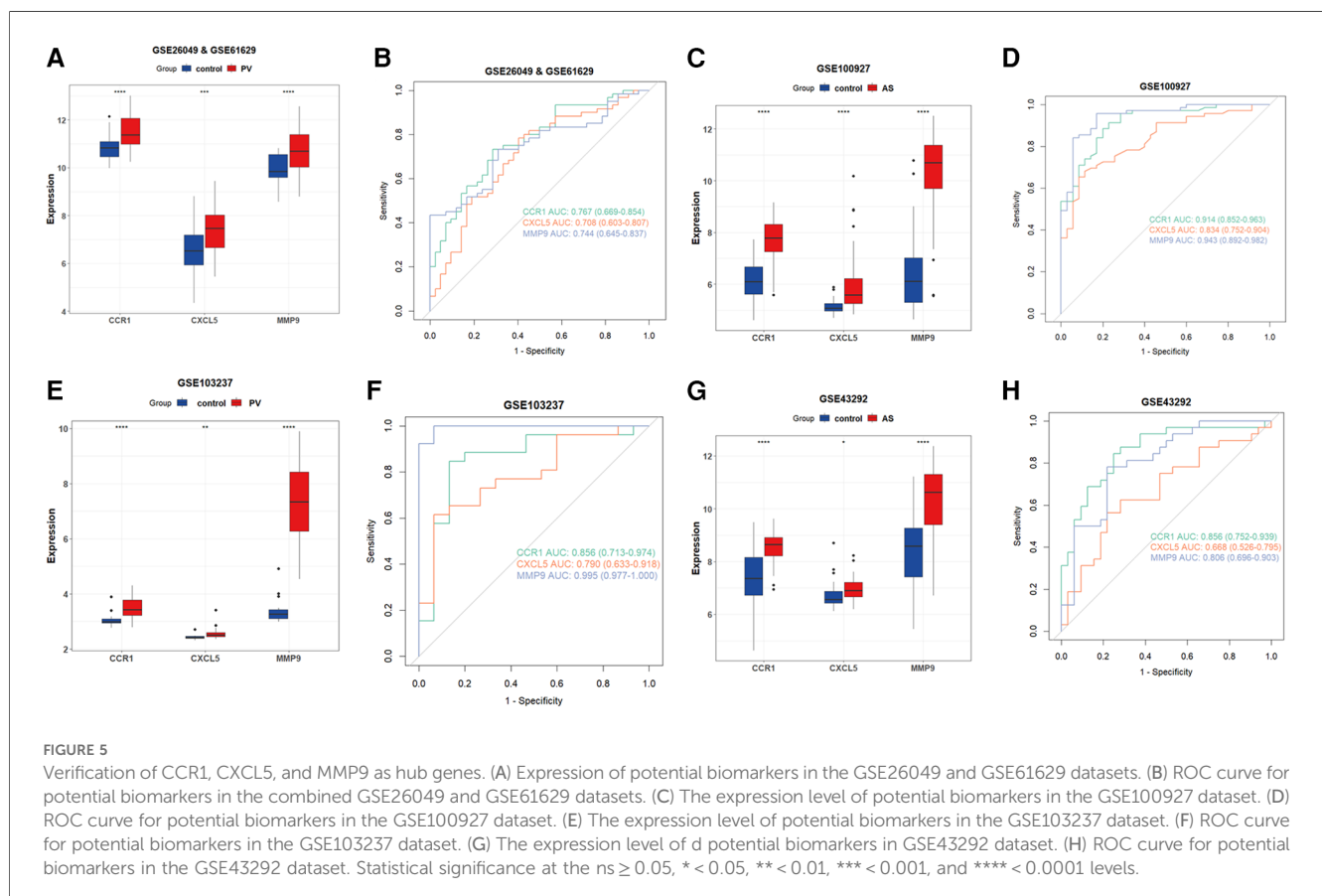
### 3.4 Construction and evaluation of a diagnostic model in PV-related AS

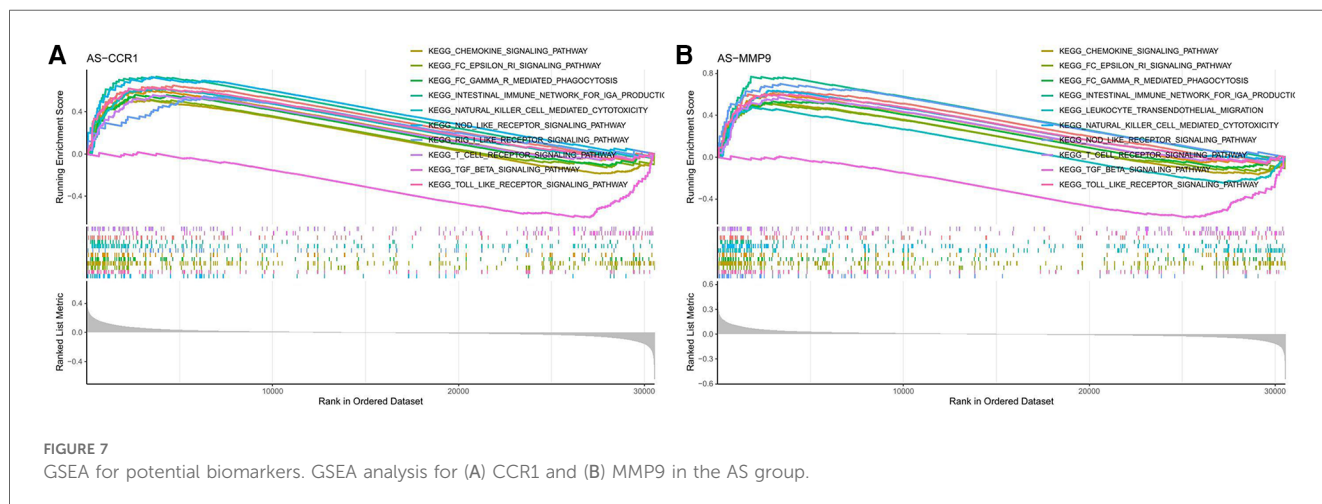
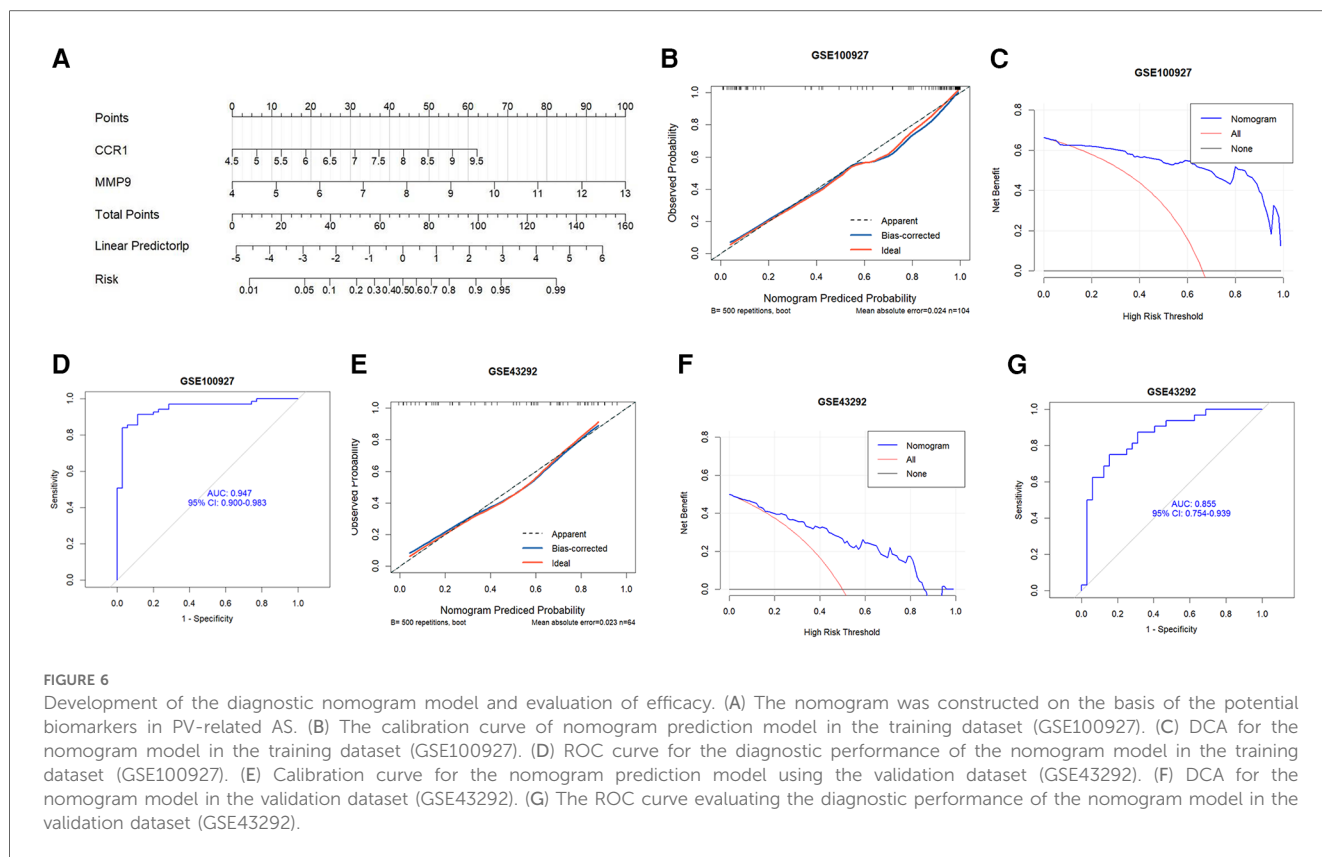
A nomogram was created based on the two hub genes using logistics regression analysis to predict the possibility of developing AS in patients with PV (Figure 6A). The calibration curves revealed that the probability of the diagnostic nomogram model was

approximately the same as that of the ideal model (Figure 6B). Furthermore, the DCA of the nomogram demonstrated that the nomogram model could be advantageous for the diagnosis of PV-related AS (Figure 6C). The nomogram had a high AUC value, indicating that it could be a useful diagnostic tool for AS-related PV (Figure 6D). Furthermore, the nomogram exhibited a great predictive value in the GEO GSE43292 dataset (Figures 6E–G), which means that the nomogram also had good predictive performance in the external cohorts.

### 3.5 Connection between potential biomarkers and inflammatory and immune processes in AS

Based on the GO and KEGG analysis of CGs in the PV and AS datasets, we found that CGs were closely associated with inflammatory and immune processes. Therefore, we used single-gene GSEA analysis of the two potential biomarkers in the AS group to determine the regulatory status of potential biomarkers for AS through immune pathways. CCR1 and MMP9 were both involved in immune pathways such as chemokine signaling, NOD-like receptor signaling, TOLL-like receptor signaling, FcεRI signaling, and in FCγR mediated phagocytosis (Figures 7A,B).



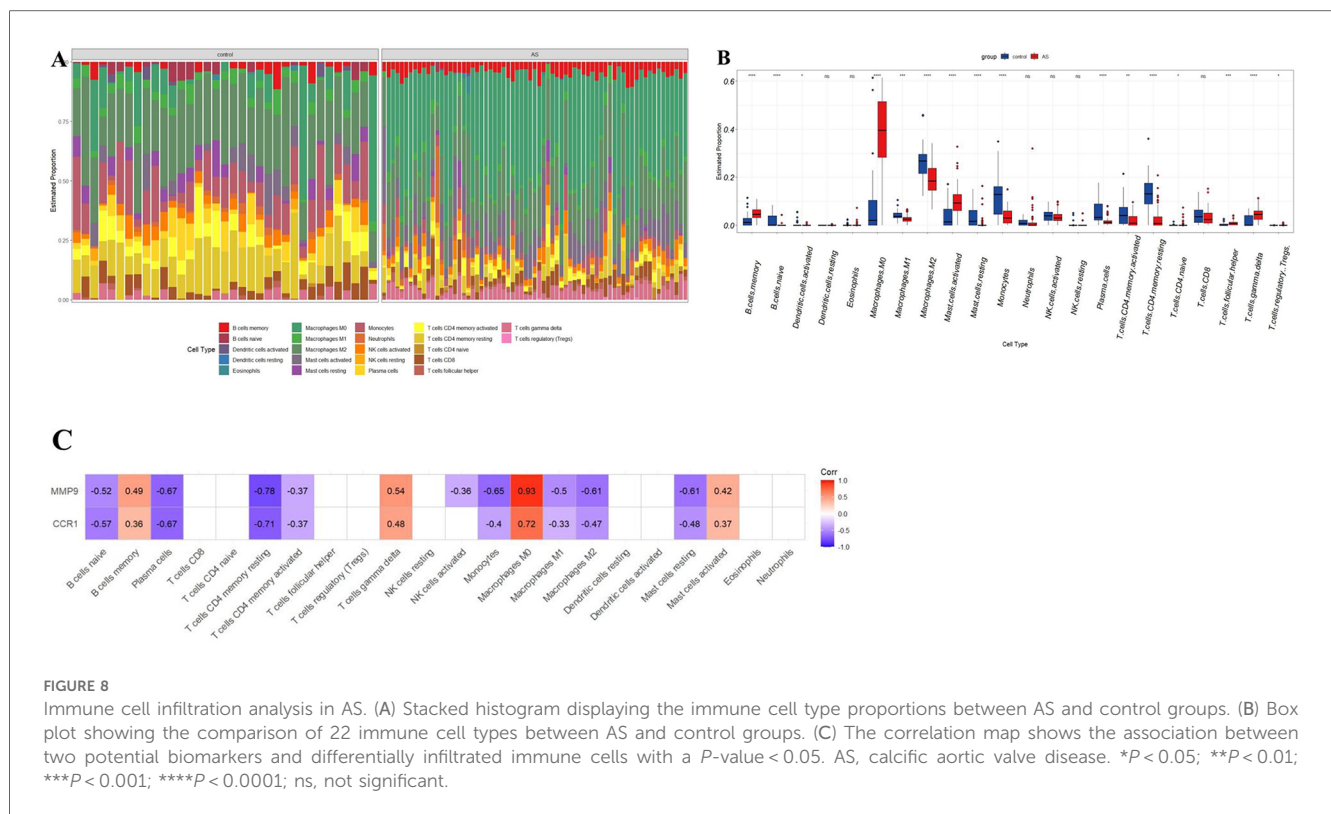


### 3.6 Immune cell infiltration and correlation analysis of potential biomarkers with invading immune cells in AS

The potential biomarkers CCR1 and MMP9 showed a close association with inflammatory and immune processes in AS. Sequentially, the CIBERSORT algorithm was used to determine the properties of immune cells and to investigate the association between potential biomarkers and immune cell type in AS. Unlike the control group, AS exhibited

higher proportions of macrophages M0, memory B-cells, activated mast cells, naive CD4T cells, follicular helper T cells,  $\gamma\delta$ T cells, and regulatory T cells (Tregs), whereas lower proportions of naive B cells, activated dendritic cells, M1 macrophages, M2 macrophages, resting mast cells, monocytes, plasma cells, and active memory CD4T cells, and resting memory CD4T cells (Figures 8A,B). Furthermore, the expression of two potential biomarkers demonstrated a strong correlation with accumulation of M0 macrophages in AS (Figure 8C).





### 3.7 Expression of hub genes in single-cell RNA-seq datasets

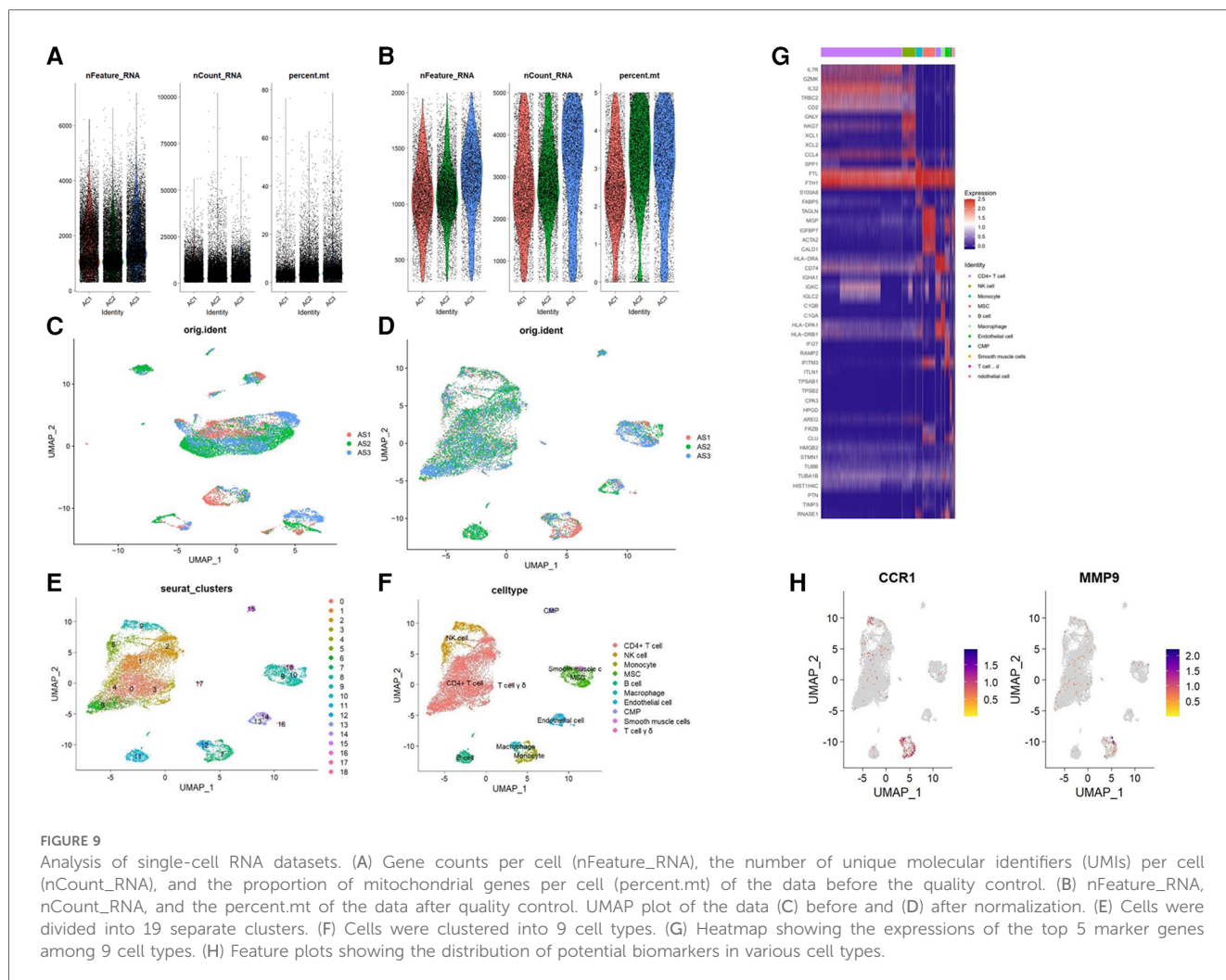
The gene expression profiles of 38,611 cells from three AS samples were obtained from the GSE159677 dataset. We filtered the data according to the depth of sequencing and the number of genes found (Figures 9A,B), and normalized the data (Figures 9C,D). For further examination, the top 2,000 highly variable genes were chosen. The “RunPCA” function was used to reduce dimensionality and a total of 19 clusters were identified (Figure 9E). Subsequently, we annotated and visualized nine different cell types (CD4+ T cell, NK cell, monocyte, MSC, B cell, macrophage, endothelial cell, CMP, smooth muscle cells, and T cell  $\gamma\delta$ ) with the “SingleR” function (Figure 9F). The top five genes of every cell type validated the annotation using the “SingleR” function (Figure 9G). Unlike other immune cell types, we observed that hub genes were substantially expressed in macrophages when we examined the location and expression of CCR1 and MMP9 (Figure 9H).

## 4 Discussion

PV related to arterial thrombosis is a significant public health concern and economic burden and is a major risk factor and event for patients with PV. AS is a significant risk factor for arterial thrombosis in patients with PV (6). Furthermore, acquired alterations in immune cells are directly responsible for the worsening of atherosclerotic plaque inflammation, which in turn causes arterial thrombosis in patients with PV (7–9). Genes

associated with atherosclerosis have also been confirmed to be highly dysregulated in MPNs (26). This study is the first to integrate bulk RNA and single-cell RNA data and to apply bioinformatics and machine learning methods to discover potential biomarkers for PV-associated AS and to explore the association between inflammation and immunity and PV-associated AS.

Our study identified 52 CGs from PV and AS datasets. Subsequently, we constructed a PPI network of CGs using the String database. Next, ten CGs were chosen using MCOD. Finally, three core CGs (CCR1, CXCL5, and MMP9) were obtained by combining the results of the LASSO regression and RF algorithms. CCR1 and MMP9 were finally identified via external data sets as core potential biomarkers for PV-associated AS. We constructed the diagnosis model based on these two potential biomarkers. CCR1 and MMP9 are both important genes involved in inflammation and immunoreaction and have a close connection with PV and AS. CCR1 is an important leukocyte chemokine receptor for several ligands, including CCL3 and MIP-1 $\alpha$ . A large increase in serum CCL3 expression has been reported in patients with PV, and a significant increase in CCR1 expression has been described in a study of the neutrophil transcriptome of these patients (27). CCR1 is expressed in AS-related cell types, such as T lymphocytes and monocytes/macrophages, and mediates CCL5-dependent inhibition and transepithelial infiltration (28, 29). MMP9 is a zinc-dependent endopeptidase that is crucial for leukocyte motility and local proteolysis of the extracellular matrix. Using an animal model, bone marrow-derived JAK2-V617F mutant mice macrophages had a noticeably higher level of MMP9 mRNA expression (30). According to a different clinical study, the polymorphism of the



MMP9 gene was associated with PV and may have a significant role in the increase in the risk of thrombosis (31). MMP-9 controls cholesterol metabolism via the MMP-9-plasma secreted phospholipase A2 axis, and MMP expression is markedly up-regulated in regions where macrophages are concentrated (around the lipid-rich core) (32, 33). The MMP9 of macrophages is crucial for the formation of susceptible regions of the AS plaque.

According to our GO enrichment analysis of 52 CGs, immune cell migration and chemotaxis were the primary biological processes of PV and AS. According to the KEGG enrichment analysis, CGs were enriched primarily in Toll-like receptors, NOD-like receptors, and chemokine signaling pathways. With single-gene GSEA, potential biomarkers included Toll-like receptor signaling pathways, NOD-like receptor signaling pathways, and chemokine signaling pathways in AS.

Toll-like receptors (TLRS) are membrane-bound receptors that generate innate immune responses, and aberrant activation of TLR3 is common in inflammatory or autoimmune diseases. Studies have shown that TLR polymorphisms may be protective against PV disease. Furthermore, the JAK2-Akt signaling pathway contributes to the pathophysiology of AS and is involved in the activation of monocyte chemoattractant protein-1 (MCP-1), which is necessary for monocyte migration into blood arteries and is

induced by TLR (34, 35). TLR3 up-regulates *P*-selectin and vascular cell adhesion molecule-1 (VCAM-1) in injured endothelial cells, facilitating leukocyte adherence (mostly lymphocytes and monocytes) in blood and endothelial cell infiltration (36). TLR may also exacerbate the inflammatory state of thrombosis by stimulating the PLT and monocyte-macrophage system (37–39). NOD-like receptors (NLRs) are a specific family of pattern recognition receptors that are responsible for the generation of innate immune responses and play a key role in the recognition of intracellular ligands. NLRP3 is highly expressed in bone marrow cells from MPN patients, and its increased expression is associated with the JAK2V617F mutation, WBC counts and splenomegaly (40). In advanced AS, the NLRP3 inflammasome can induce premature macrophage death and massive lipid release, thus increasing plaque susceptibility (41). Various chemokines can activate the JAK/STAT, Ras, ERK, and Akt pathways through their receptors to induce directional chemotaxis of immune cells. Chemokine levels (CCL2, CCL5, CXCL8, CXCL12, CXCL10) were elevated in the bone marrow of patients with PV, indicating a highly inflammatory environment (42). Chemokines and their receptors may play a key role in the pathogenesis of AS. Chemokine receptors such as CCR1, CCR5, and CXCR2 can promote plaque development by recruiting monocytes into the blood. Elevated CXCL10

concentration is associated with an increased risk of vulnerable plaque and atherothrombosis (43).

Additionally, our study observed a strong connection between potential biomarkers and the infiltration of M0 macrophages in AS. Potential biomarkers were predominantly expressed in macrophages in AS plaque tissue, according to single-cell annotation and tissue marker gene analysis. This suggests that one of the primary mechanisms of PV-associated AS may be macrophage infiltration. Peripheral blood from PV patients contains considerably higher levels of inflammatory CD14 and CD16 monocytes/macrophages in various studies (44, 45). Macrophages are crucial at every stage of AS. During the inflammatory phase of AS, injured endothelium cells emit chemokines that attract circulating monocytes, which then undergo macrophage differentiation in response to growth factors and pro-inflammatory cytokines. These macrophages can quickly identify oxidized LDL, phagocytose it, and transform into foam cells, which form the first atherosclerotic lesions (46). The late stage of AS is characterized by a progressive increase in M1 macrophage counts and an increase in the release of pro-inflammatory factors, which increases the risk of plaque rupture (47). Furthermore, M1 macrophages have the potential to release MMPs, including MMP2 and MMP9, which cause the extracellular matrix of plaques to degrade and eventually rupture, increasing the risk of acute cardiovascular events (48).

However, our study has several limitations that should be addressed. Firstly, our study was constructed using retrospective data from public databases. Some bias may occur due to the limited sample size and multiple data analyses. Therefore, data from real-world studies are needed to verify the model's clinical applicability. In the future, we plan to collect blood samples from patients with initial onset of disease to confirm the expression and probable function of CGs.

## 5 Conclusions

This study identified two CGs (CCR1 and MMP9) as potential biomarkers for PV-related AS and established a diagnostic model based on these genes. Our results showed that Toll-like receptor signaling, NOD-like receptor signaling, and chemokine signaling may be closely associated connection with PV-related AS. Macrophage infiltration may be one of the primary mechanisms underlying PV-related AS. This study describes two potential markers for assessing the risk of developing AS in patients with

PV and provides a new perspective on the common molecular mechanisms underlying PV and AS, which may provide clues for designing experimental studies, determining diagnosis, and widening treatment options for PV-related AS.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: (<https://www.ncbi.nlm.nih.gov/geo/>) GSE26049 GSE61629 GSE100927 GSE43292 GSE103237 GSE159677.

## Author contributions

ZW: Writing – review & editing. JZ: Writing – original draft.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

The authors are grateful for Gene Expression Omnibus.

## 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.

## References

- Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the world health organization classification of myeloid neoplasms and acute leukemia. *Blood*. (2016) 127(20):2391–405. doi: 10.1182/blood-2016-03-643544
- Srou SA, Devesa SS, Morton LM, Check DP, Curtis RE, Linet MS, et al. Incidence and patient survival of myeloproliferative neoplasms and myelodysplastic/myeloproliferative neoplasms in the United States, 2001–12. *Br J Haematol*. (2016) 174(3):382–96. doi: 10.1111/bjh.14061
- Barbui T, Thiele J, Gisslinger H, Kvasnicka HM, Vannucchi AM, Guglielmelli P, et al. The 2016 WHO classification and diagnostic criteria for myeloproliferative

neoplasms: document summary and indepth discussion. *Blood Cancer J*. (2018) 8(2):15. doi: 10.1038/s41408-018-0054-y

- Szuber N, Mudireddy M, Nicolosi M, Penna D, Vallapureddy RR, Lasho TL, et al. 2023 Mayo clinic patients with myeloproliferative neoplasms: risk-stratified comparison of survival and outcomes data among disease subgroups. *Mayo Clin Proc*. (2019) 94(4):599–610. doi: 10.1016/j.mayocp.2018.08.022

- Cerquozzi S, Barraco D, Lasho T, Finke C, Hanson CA, Ketterling RP, et al. Risk factors for arterial versus venous thrombosis in polycythemia vera: a single center experience in 587 patients. *Blood Cancer J*. (2017) 7(12):662. doi: 10.1038/s41408-017-0035-6

6. Herbreteau L, Couturaud F, Hoffmann C, Bressollette L, Pan-Petes B, Rio L, et al. Atrial fibrillation and peripheral arterial disease define MPN patients with very high risk of thrombosis. *Thromb Res.* (2023) 226:93–9. doi: 10.1016/j.thromres.2023.04.021
7. Carobbio A, Finazzi G, Antonioli E, Guglielmelli P, Vannucchi AM, Delaini F, et al. Thrombocytosis and leukocytosis interaction in vascular complications of essential thrombocythemia. *Blood.* (2008) 112(8):3135–7. doi: 10.1182/blood-2008-04-153783
8. Carobbio A, Finazzi G, Guerini V, Spinelli O, Delaini F, Marchioli R, et al. Thrombocytosis and leukocytosis interaction in essential thrombocythemia: interaction with treatment, standard risk factors, and Jak2 mutation status. *Blood.* (2007) 109(6):2310–3. doi: 10.1182/blood-2006-09-046342
9. Falanga A, Marchetti M, Vignoli A, Balducci D, Russo L, Guerini V, et al. V617F JAK-2 mutation in patients with essential thrombocythemia: relation to platelet, granulocyte, and plasma hemostatic and inflammatory molecules. *Exp Hematol.* (2007) 35(5):702–11. doi: 10.1016/j.exphem.2007.01.053
10. Medinger M, Skoda R, Gratwohl A, Theocharides A, Buser A, Heim D, et al. Angiogenesis and vascular endothelial growth factor-receptor expression in myeloproliferative neoplasms: correlation with clinical parameters and JAK2-V617F mutational status. *Br J Haematol.* (2009) 146(2):150–7. doi: 10.1111/j.1365-2141.2009.07726.x
11. Karakantza M, Giannakoulas NC, Zikos P, Sakellaropoulos G, Kouraklis A, Aktipi A, et al. Markers of endothelial and *in vivo* platelet activation in patients with essential thrombocythemia and polycythemia vera. *Int J Hematol.* (2004) 79(3):253–9. doi: 10.1532/IJH97.E0316
12. Musolino C, Calabro' L, Bellomo G, Martello F, Loteta B, Pezzano C, et al. Soluble angiogenic factors: implications for chronic myeloproliferative disorders. *Am J Hematol.* (2002) 69(3):159–63. doi: 10.1002/ajh.10020
13. Kwon SS, Yoon SY, Jeong SY, Lee MY, Kim KH, Lee N, et al. Neutrophil/lymphocyte ratio and carotid plaque burden in patients with essential thrombocythemia and polycythemia vera. *Nutr Metab Cardiovasc Dis.* (2022) 32(8):1913–6. doi: 10.1016/j.numecd.2022.04.013
14. Hasselbalch HC. Perspectives on chronic inflammation in essential thrombocythemia, polycythemia vera, and myelofibrosis: is chronic inflammation a trigger and driver of clonal evolution and development of accelerated atherosclerosis and second cancer? *Blood.* (2012) 119(14):3219–25. doi: 10.1182/blood-2011-11-394775
15. Haybar H, Khodadi E, Shahjehani M, Saki N. Cardiovascular events: a challenge in JAK2-positive myeloproliferative neoplasms. *Cardiovasc Hematol Disord Drug Targets.* (2017) 17(3):161–6. doi: 10.2174/1871529X17666171030122345
16. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* (2013) 41(Database issue):D991–5. doi: 10.1093/nar/gks1193
17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
19. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
20. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING V11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* (2019) 47(D1):D607–13. doi: 10.1093/nar/gky1131
21. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01
22. Liaw A, Wiener M. Classification and regression by randomForest. *R News.* (2002) 2:18–22. Available online at: <https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf>
23. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb).* (2021) 2(3):100141. doi: 10.1016/j.xinn.2021.100141
24. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. *Methods Mol Biol.* (2020) 2117:135–57. doi: 10.1007/978-1-0716-0301-7\_7
25. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* (2019) 20(2):163–72. doi: 10.1038/s41590-018-0276-y
26. Skov V, Thomassen M, Kjaer L, Larsen MK, Knudsen TA, Ellervik C, et al. Whole blood transcriptional profiling reveals highly deregulated atherosclerosis genes in Philadelphia-chromosome negative myeloproliferative neoplasms. *Eur J Haematol.* (2023) 111(5):805–14. doi: 10.1111/ejh.14081
27. Slezak S, Jin P, Caruccio L, Ren J, Bennett M, Zia N, et al. Gene and microRNA analysis of neutrophils from patients with polycythemia vera and essential thrombocytosis: down-regulation of micro-RNA-1 and -133a. *J Transl Med.* (2009) 7:39. doi: 10.1186/1479-5876-7-39
28. Krohn R, Raffetseder U, Bot I, Ren J, Bennett M, Zia N, et al. Y-box binding protein-1 controls CC chemokine ligand-5 (CCL5) expression in smooth muscle cells and contributes to neointima formation in atherosclerosis-prone mice. *Circulation.* (2007) 116(16):1812–20. doi: 10.1161/CIRCULATIONAHA.107.780816
29. Brauersreuther V, Zernecke A, Arnaud C, Liehn EA, Steffens S, Shagdarsuren E, et al. Ccr5 but not Ccr1 deficiency reduces development of diet-induced atherosclerosis in mice. *Arterioscler Thromb Vasc Biol.* (2007) 27(2):373–9. doi: 10.1161/01.ATV.0000253886.44609.ae
30. Hernandez-Anzaldo S, Brglez V, Hemmerlyckx B, Leung D, Filep JG, Vance JE, et al. Novel role for matrix metalloproteinase 9 in modulation of cholesterol metabolism. *J Am Heart Assoc.* (2016) 5(10):e004228. doi: 10.1161/JAHA.116.004228
31. Sag SO, Gorukmez O, Ture M, Gorukmez O, Topak A, Sahinturk S, et al. MMP2 gene-735 C/T and MMP9 gene -1562 C/T polymorphisms in JAK2V617F positive myeloproliferative disorders. *Asian Pac J Cancer Prev.* (2015) 16(2):443–9. doi: 10.7314/APJCP.2015.16.2.443
32. Shah PK, Falk E, Badimon JJ, Fernandez-Ortiz A, Mailhac A, Villareal-Levy G, et al. Human monocyte-derived macrophages induce collagen breakdown in fibrous caps of atherosclerotic plaques. Potential role of matrix-degrading metalloproteinases and implications for plaque rupture. *Circulation.* (1995) 92(6):1565–9. Available online at: <http://intl-circ.ahajournals.org/cgi/content/abstract/92/6/1565>
33. Yokokawa T, Misaka T, Kimishima Y, Wada K, Minakawa K, Sugimoto K, et al. Crucial role of hematopoietic JAK2 V617F in the development of aortic aneurysms. *Haematologica.* (2021) 106(7):1910–22. doi: 10.3324/haematol.2020.264085
34. Quirino MG, Macedo LC, Pagnano KBB, Pagliarini-E-Silva S, Sell AM, Visentainer JEL. Toll-like receptor gene polymorphisms in patients with myeloproliferative neoplasms. *Mol Biol Rep.* (2021) 48(6):4995–5001. doi: 10.1007/s11033-021-06238-8
35. Park DW, Lee HK, Jeong TW, Kim JS, Bae YS, Chin BR, et al. The JAK2-akt-glycogen synthase kinase-3 $\beta$  signaling pathway is involved in toll-like receptor 2-induced monocyte chemoattractant protein-1 regulation. *Mol Med Rep.* (2012) 5(4):1063–7. doi: 10.3892/mmr.2012.741
36. Gimbrone MA Jr, Garcia-Cardena G. Endothelial cell dysfunction and the pathobiology of atherosclerosis. *Circ Res.* (2016) 118(4):620–36. doi: 10.1161/CIRCRESAHA.115.306301
37. Rivadeneyra L, Carestia A, Etulain J, Pozner RG, Fondevila C, Negrotto S, et al. Regulation of platelet responses triggered by toll-like receptor 2 and 4 ligands is another non-genomic role of nuclear factor-kappaB. *Thromb Res.* (2014) 133(2):235–43. doi: 10.1016/j.thromres.2013.11.028
38. Damien P, Cognasse F, Payrastra B, Spinelli SL, Blumberg N, Arthaud CA, et al. NF- $\kappa$ B links TLR2 and PAR1 to soluble immunomodulator factor secretion in human platelets. *Front Immunol.* (2017) 8:85. doi: 10.3389/fimmu.2017.00085
39. Watanabe T, Kitani A, Murray PJ, Strober W. NOD2 is a negative regulator of toll-like receptor 2-mediated T helper type 1 responses. *Nat Immunol.* (2004) 5(8):800–8. doi: 10.1038/ni1092
40. Zhou Y, Yan S, Liu N, He N, Zhang A, Meng S, et al. Genetic polymorphisms and expression of NLRP3 inflammasome-related genes are associated with Philadelphia chromosome-negative myeloproliferative neoplasms. *Hum Immunol.* (2020) 81(10–11):606–13. doi: 10.1016/j.humimm.2020.09.001
41. Zeng W, Wu D, Sun Y, Suo Y, Yu Q, Zeng M, et al. The selective NLRP3 inhibitor MCC950 hinders atherosclerosis development by attenuating inflammation and pyroptosis in macrophages. *Sci Rep.* (2021) 11(1):19305. doi: 10.1038/s41598-021-98437-3
42. Ho CL, Lasho TL, Butterfield JH, Tefferi A. Global cytokine analysis in myeloproliferative disorders. *Leuk Res.* (2007) 31(10):1389–92. doi: 10.1016/j.leukres.2006.12.024
43. Rolin J, Maghazachi AA. Implications of chemokines, chemokine receptors, and inflammatory lipids in atherosclerosis. *J Leukoc Biol.* (2014) 95(4):575–85. doi: 10.1189/jlb.1113571
44. Fan W, Cao W, Shi J, Gao F, Wang M, Xu L, et al. Contributions of bone marrow monocytes/macrophages in myeloproliferative neoplasms with JAK2V617F mutation. *Ann Hematol.* (2023) 102(7):1745–59. doi: 10.1007/s00277-023-05284-5
45. Bassan VL, Barretto GD, de Almeida FC, Palma PVB, Binelli LS, da Silva JPL, et al. Philadelphia-negative myeloproliferative neoplasms display alterations in monocyte subpopulations frequency and immunophenotype. *Med Oncol.* (2022) 39(12):223. doi: 10.1007/s12032-022-01825-6
46. Khatana C, Saini NK, Chakrabarti S, Saini V, Sharma A, Saini RV, et al. Mechanistic insights into the oxidized low-density lipoprotein-induced atherosclerosis. *Oxid Med Cell Longev.* (2020) 2020:5245308. doi: 10.1155/2020/5245308
47. Kallou-Laschet J, Varthaman A, Fornasa G, Compain C, Gaston AT, Clement M, et al. Macrophage plasticity in experimental atherosclerosis. *PLoS One.* (2010) 5(1):e8852. doi: 10.1371/journal.pone.008852
48. Seifert R, Kuhlmann MT, Eligehausen S, Kiefer F, Hermann S, Schäfers M, et al. Molecular imaging of MMP activity discriminates unstable from stable plaque phenotypes in shear-stress induced murine atherosclerosis. *PLoS One.* (2018) 13(10):e0204305. doi: 10.1371/journal.pone.0204305