Journal Pre-proof

Plasma proteomic signature for preoperative prediction of microvascular invasion in HCC

Xinrui Shi, Yunzheng Zhao, Ke Li, Qingyu Li, Yifeng Cui, Yuhang Sui, Liang Zhao, Haonan Zhou, Yongsheng Yang, Jiajun Li, Meng Zhou, Zhaoyang Lu

PII: S2589-5559(25)00159-4

DOI: https://doi.org/10.1016/j.jhepr.2025.101481

Reference: JHEPR 101481

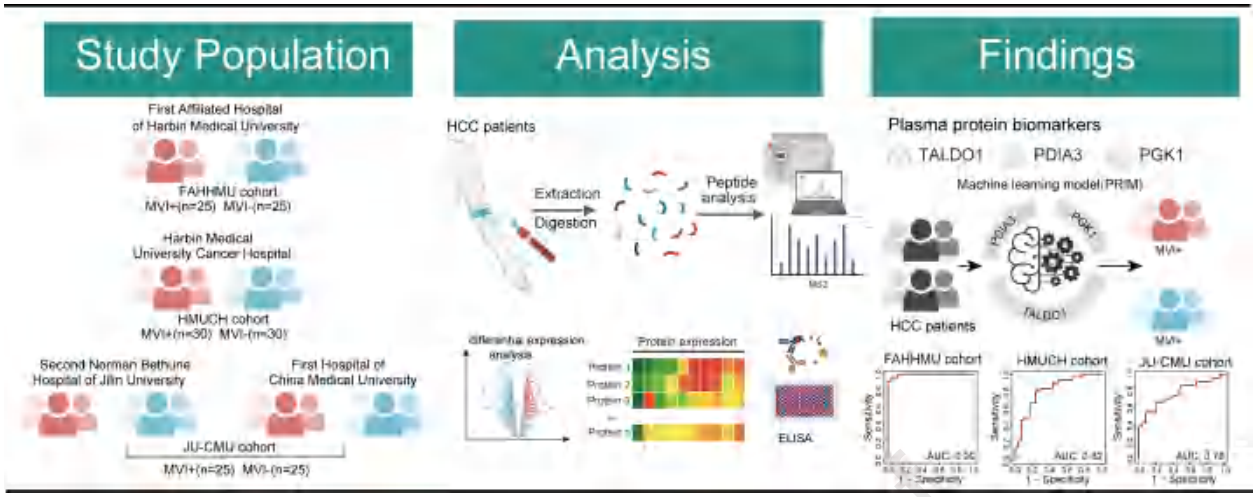
To appear in: *JHEP Reports*

Received Date: 12 March 2025

Revised Date: 2 June 2025

Accepted Date: 4 June 2025

Please cite this article as: Shi X, Zhao Y, Li K, Li Q, Cui Y, Sui Y, Zhao L, Zhou H, Yang Y, Li J, Zhou M, Lu Z, Plasma proteomic signature for preoperative prediction of microvascular invasion in HCC, *JHEP Reports*, https://doi.org/10.1016/j.jhepr.2025.101481.



Plasma proteomic signature for preoperative prediction of microvascular invasion in HCC

Short Title: Plasma Proteomics for MVI

Xinrui Shi[2#], Yunzheng Zhao[1,2#], Ke Li[2#], Qingyu Li[1], Yifeng Cui[1], Yuhang Sui[1], Liang Zhao[3], Haonan Zhou[4], Yongsheng Yang[5], Jiajun Li[1], Meng Zhou[2*], Zhaoyang Lu[1*]

[1]Department of Hepatic Surgery, Key Laboratory of Hepatosplenic Surgery, Ministry of Education, The First Affiliated Hospital of Harbin Medical University, Harbin, China

[2.] Institute of Genomic Medicine, School of Biomedical Engineering, Wenzhou Medical University, Wenzhou 325027, P. R. China.

3Department of Hepatopancreatobiliary Surgery, Harbin Medical University Cancer Hospital, Harbin, China

[4] Department of Hepatobiliary Surgery, The First Hospital of China Medical University, Shenyang, China

[5] Department of Hepatobiliary and Pancreatic Surgery, The Second Hospital of Jilin University, Changchun, China

[#]These authors contributed equally.

[*]Corresponding authors: zhoumeng@wmu.edu.cn (M.Z); lzy76772005@hrbmu.edu.cn (Z.Y.L.)

Keywords: Plasma proteomics, microvascular invasion, risk stratification, HCC management, tumor microenvironment

Electronic word count: 4258

Number of figures and tables: 6

Conflict of interest statement: The authors have declared no conflicts of interest.

Financial support statement: This work was supported by the National Natural Scientific Foundation of China (Grant No. 81972230), Heilongjiang Province key research and development plan project (Grant No. 2022ZX06C17), The Heilongjiang Postdoctoral Science Foundation (Grant No. LBH-Z20178), The Scientific Foundation of the First Affiliated Hospital of Harbin Medical University (Grant No. 2021B03) and The Excellent Youth Science Fund of the First Affiliated Hospital of Harbin Medical University (Grant No. 2021Y01).

Author contributions: M.Z. and Z.Y.L. conceived, designed and organized the study. Y.Z.Z., Q.Y.L., Y.F.C., Y.H.S., H.N.Z. and J.J.L. contributed to sample and clinical information collection. X.R.S. and K.L. conducted data analysis and visualization. Y.Z.Z. conducted ELISA verification. X.R.S., K.L., Y.Z.Z. and M.Z. interpreted the results and drafted the manuscript. All authors reviewed the manuscript and approved the submitted version.

Abstract

Background & Aims: Microvascular invasion (MVI) is a major determinant of poor prognosis in hepatocellular carcinoma (HCC). However, reliable noninvasive biomarkers for the preoperative evaluation and diagnosis of MVI are urgently needed in clinical practice.

Methods: Plasma samples were collected from 160 HCC patients (80 MVI-positive and 80 MVI-negative patients) from four medical centers. Plasma proteomic profiling was obtained using data-independent acquisition mass spectrometry (DIA-MS). Principal component analysis and differential protein abundance analysis were used to assess the proteomic changes between the two groups of patients. Protein biomarker candidates were further quantitatively validated by enzyme-linked immunosorbent assay (ELISA).

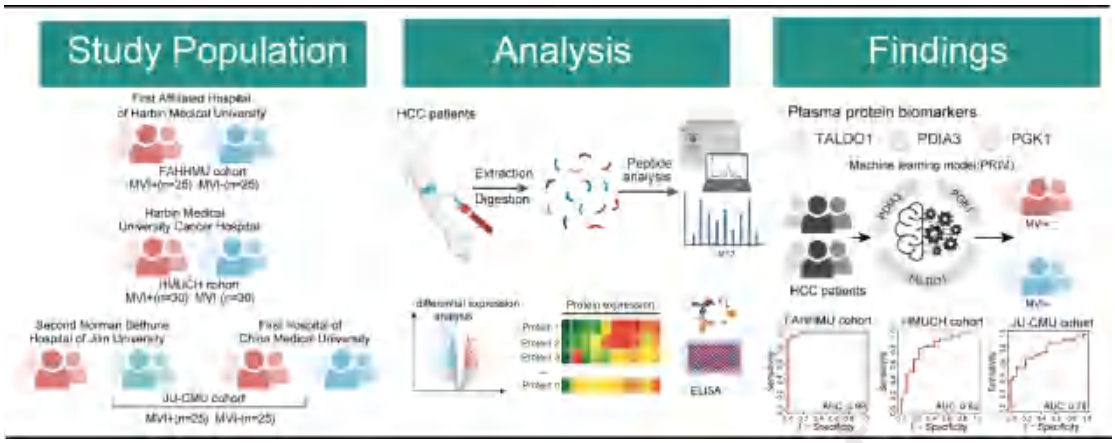
Results: Proteomic analysis of 50 HCC patients (25 MVI-positive and 25 MVI-negative) identified three plasma protein biomarkers (TALDO1, PDIA3, and PGK1) which are significantly upregulated in MVI-positive patients (FDR-adjusted $p < 0.05$) and subsequently were cross-validated by ELISA. A machine learning-based **P**lasma p**R**otein MV**I** risk **M**odel (PRIM) was developed for the preoperative prediction of MVI. The PRIM model demonstrated excellent discriminatory ability, with areas under the receiver operating characteristic curve (AUROC) values ranging from 0.78 to 0.99 across three independent cohorts. Single-cell RNA sequencing of five HCC tumors provided a cell type-resolved atlas of biomarker expression, showing their predominant presence in malignant cells and macrophages within the MVI+ tumor microenvironment compared to MVI- tumors.

Conclusions: This study provides a comprehensive analysis of the plasma proteomic landscape in HCC and presents a promising blood-based tool for preoperative MVI risk stratification.

Impact and implications

This study highlights the transformative potential of plasma proteomic profiling in improving the preoperative prediction of microvascular invasion in hepatocellular carcinoma. By integrating data-independent acquisition mass spectrometry and with machine learning, we identified three plasma protein biomarkers (TALDO1, PDIA3, and PGK1) and developed the Plasma pRotein MVI risk Model (PRIM), which demonstrated robust diagnostic accuracy across multicenter validation cohorts. These findings pave the way for preoperative risk stratification and personalized therapeutic strategies in HCC management.

Graphical abstract



Introduction

Hepatocellular carcinoma (HCC) accounts for approximately 90% of primary liver cancers and remains one of the leading causes of cancer-related mortality worldwide [1]. Microvascular invasion (MVI) is a crucial pathological feature characterized by the infiltration of tumor cells into adjacent microvessels [2, 3]. MVI is known to be strongly associated with aggressive disease progression, increased recurrence and metastasis, and poor survival rates following surgery or liver transplantation [3]. Early preoperative identification of MVI enables more accurate risk stratification and more personalized treatment strategies, which could significantly improve patient survival [4-6]. However, the current diagnostic methods for MVI rely predominantly on postoperative histopathological examination, highlighting the urgent need for reliable and non-invasive biomarkers that can accurately predict the presence of MVI preoperatively.

Plasma proteomics has emerged as a powerful strategy for biomarker discovery, offering a minimally invasive means into systemic pathophysiological changes. The proteins circulating in the plasma reflect the underlying biological processes, and their levels can be modulated by tumor presence, stage, and even the dynamics of the tumor microenvironment (TME). Recent advances in mass spectrometry (MS) and data-independent acquisition (DIA) workflows have enabled deep proteome profiling, as demonstrated in studies identifying panels of proteins associated with various cancer types [7-15]. In HCC, previous plasma proteomic efforts have focused primarily on HCC diagnosis, but have not specifically addressed biomarkers for MVI [16-18].

In this study, we conducted a multi-stage, multi-center investigation of plasma proteome dynamics in HCC patients with MVI. By integrating DIA-MS proteomics, enzyme-linked immunosorbent assay (ELISA), and machine learning, we identified

102 and cross-validated a three-protein panel that could serve as a preoperative biomarker
103 signature for MVI risk stratification.

104 **Materials and Methods**

105 **Study population and study design**

106 This multi-center case-control study was conducted between January 2023 and
107 December 2024 across four hospitals in China: the First Affiliated Hospital of Harbin
108 Medical University (FAHHMU), Harbin Medical University Cancer Hospital
109 (HMUCH), the Second Norman Bethune Hospital of Jilin University (JU), and the First
110 Hospital of China Medical University (CMU). The study was approved by the Ethics
111 Committees of each participating institution, and written informed consent was
112 obtained from all participants prior to any study-related procedures. All procedures
113 adhered to the principles of the Declaration of Helsinki.

114 The inclusion criteria were as follows: (1) Imaging examinations demonstrating a
115 solitary lesion; (2) Normal or well-compensated liver function, corresponding to Child-
116 Pugh grades A-B; (3) Adequate tolerance for curative hepatic resection; (4)
117 Postoperative pathological diagnosis confirming HCC, with the grading of MVI
118 confirmed; (5) Availability of comprehensive preoperative imaging data and related
119 laboratory test results; (6) Voluntary participation in the study, with signed informed
120 consent and agreement to follow-up and data collection. The exclusion criteria
121 included: (1) Prior to surgery, any systemic anti-HCC treatments (e.g., liver
122 transplantation, transarterial chemoembolization, radiotherapy, chemotherapy,
123 molecular targeted therapy, or immunotherapy); (2) Current treatment with medications
124 known to potentially cause liver injury; (3) Severe liver dysfunction, decompensated
125 cirrhosis, or active hepatitis; (4) A history or concurrent occurrence of ruptured and
126 bleeding HCC; (5) A history or coexistence of other malignancies; (6) Severe diseases

127 affecting critical organs such as the heart, lungs, kidneys, brain, or blood system; (7)
128 Comorbid autoimmune diseases, metabolic disorders, or severe neurological or
129 psychiatric conditions; (8) Substance abuse.

130 Patients were consecutively recruited at each participating center during the study
131 period to ensure representative sampling and minimize selection bias. All HCC patients
132 who met the inclusion criteria and had no exclusion criteria were evaluated. After
133 pathological confirmation of HCC and MVI status assessment, patients were classified
134 as cases (MVI-positive) or controls (MVI-negative). The consecutive enrollment
135 continued at each center until approximately 25-30 cases and 25-30 controls were
136 obtained from each hospital, resulting in a total of 160 patients across all four centers.
137 This recruitment strategy ensured that the sample was representative of the general
138 HCC patient population undergoing surgical resection at each institution, avoiding
139 selective sampling that might introduce bias. All eligible patients were approached for
140 study participation.

141 Based on this recruitment process, the study cohorts were organized as follows:
142 (1) the discovery cohort included 50 plasma samples (25 MVI+ and 25 MVI-) from the
143 FAHHMU cohort. (2) Validation cohort 1 consisting of 60 HCC patients (30 MVI+ and
144 30 MVI-) from the HMUCH cohort; (3) Validation cohort 2 composing of 50 patients
145 (25 MVI+ and 25 MVI-) from the JU-CMU cohort.

146 **Plasma sample preparation**

147 All participants underwent radical surgical resection for HCC, and fasting blood
148 samples were collected in the early morning prior to surgery into EDTA-containing
149 tubes. The samples were then centrifuged at 1000×g for 15 minutes at 4°C and the
150 resulting plasma supernatant was stored at -80°C for further analysis. Postoperative
151 MVI status was independently assessed and pathologically confirmed by three

experienced pathologists.

Plasma samples were transferred to fresh centrifuge tubes, and magnetic nanomaterials (PTM-00F13303, PTM Bio) were added. The samples were incubated at 1200 rpm and 37°C for one hour. After incubation, the magnetic beads were washed three times with washing buffer. Next, 70 μL of enzyme digestion buffer was added to the beads, and the mixture was heated at 95°C for 10 minutes, allowing it to cool to room temperature. Trypsin was added to a final concentration of 20 ng/μL for overnight digestion. The digestion solution was reduced with 5 mM dithiothreitol (DTT) at 56°C for 30 minutes, followed by alkylation with 11 mM iodoacetamide (IAM) for 15 minutes at room temperature in the dark. The resulting peptides were desalted using C18 ZipTips (Millipore) according to the manufacturer's instructions and then lyophilized for subsequent MS analysis.

LC-MS/MS analysis

For MS analysis, tryptic peptides were dissolved in solvent A and loaded onto a homemade reversed-phase analytical column (15 cm length, 100 μm i.d.). The mobile phase consisted of solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid, 80% acetonitrile in water). Peptides were separated using the following gradient: 0-1.6 min, 4%-22.5% B; 1.6-2.0 min, 22.5%-35% B; 2.0-2.6 min, 35%-55% B; 2.6-2.7 min, 55%-99% B; 2.7-6.8 min, 99% B; 6.8-7.6 min, 99% B, at a constant flow rate of 300 nL/min using a Vanquish Neo UPLC system (ThermoFisher Scientific). The separated peptides were analyzed using an Orbitrap Astral mass spectrometer with a nano-electrospray ion source and applying an electrospray voltage of 1900 V. Full MS scans were acquired in the Orbitrap detector with a resolution of 240,000, scanning the mass-to-charge ratio (m/z) range of 380–980. MS/MS scans were conducted in the Astral detector with a resolution of 80,000, using a fixed first mass of 150.0 m/z and

high-energy collision-induced dissociation (HCD) fragmentation at a normalized collision energy (NCE) of 25%. The automatic gain control (AGC) target was set at 500%, with a maximum injection time of 3ms.

DIA data were processed using the DIA-NN (v1.8), with tandem mass spectra searched against the Homo sapiens reference database (Homo_sapiens_9606_SP_20231220.fasta) concatenated with a reverse decoy database. Data analysis was performed with Trypsin/P as the cleavage enzyme, allowing one missed cleavage, and fixed carbamidomethylation of cysteine and N-terminal methionine excision. A false discovery rate (FDR) of <1% was used.

Missing value imputation

Proteins detected in fewer than 25% of the samples were excluded from further analysis to ensure that only reliable, widely detectable proteins were included. For the remaining proteins, missing values were imputed using the k-nearest neighbors (KNN) algorithm to generate a more robust and accurate dataset for subsequent analyses [19]. Missing data imputation was performed using the impute R package (version 1.62.0), with default parameters as recommended by the package documentation. This KNN algorithm identifies the k most similar samples (neighbors) to each incomplete sample based on non-missing features, with similarity measured by Euclidean distance. The missing values are then imputed as a weighted average of the corresponding values from these neighbors, with the weights inversely proportional to the distance between the target sample and its neighbors. By default, the number of neighbors (k) was set to 10. Before calculating the distances, each feature (row) is standardized to zero mean and unit variance to ensure that features with larger dynamic ranges do not disproportionately influence neighbor selection. The imputation process is performed iteratively, with the algorithm continuing until convergence or up to a maximum of

1000 iterations.

Differential abundance analysis

Differential protein abundance between MVI+ and MVI- patients was assessed using the Mann-Whitney U test. Proteins that were significantly upregulated or downregulated in the MVI+ group compared to the MVI- group were defined as differentially abundant with a p-value < 0.05 and a log2 fold change (FC) greater than 0.58 (upregulated) or less than -0.58 (downregulated).

Weighted gene co-expression network analysis (WGCNA)

WGCNA analysis was used to identify co-regulated protein modules in an unsupervised manner [20]. A signed gene co-expression network was constructed with a soft-thresholding power of 3 to achieve a scale-free topology model fit ($R^2 = 0.8$). Groups of co-regulated genes (modules) were detected using the blockwiseModules function, with a minimum module size set to 20. The robustness of the identified protein modules was evaluated by applying t-distributed stochastic neighbor embedding (t-SNE) to the proteins in the top 30% based on module membership (kME) values within each module using the Rtsne package.

Functional enrichment analysis

Functional enrichment analysis was performed to identify enriched biological processes for the group of proteins, using the clusterProfiler R package (v3.16.1) [21] and ClueGO [22], with an adjusted p-value <0.05.

ELISA detection

The plasma concentrations of TALDO1, HSPA8, HSPA1A, PDIA3, PGK1, IDH1, PGM3 and HSPA6 were quantified using commercially available ELISA kits (TALDO1, HSPA8, HSPA1A, PDIA3, PGK1 and IDH1: JONLNBIO; PGM3: ABEBIO; HSPA6: ABCLONAL.) The working standards, biotin-conjugated antibody,

227 streptavidin-HRP, and wash buffer were prepared in accordance with the manufacturer's
228 instructions, and the test samples were appropriately diluted. A 100 μL aliquot of either
229 the standards or the test samples were added to each well and incubated at 37°C for 2
230 hours. After incubation, the wells were washed three times. Next, 100 μL of the working
231 biotin-conjugated antibody was added, and the mixture was incubated for 1 hour at
232 37°C, followed by three additional wash cycles. Then, 100 μL of working streptavidin-
233 HRP was added, and the mixture was incubated for 30 minutes at 37°C, followed by
234 five wash cycles. Subsequently, 90 μL of substrate mixture was added, and the reaction
235 was incubated for 20 minutes at 37°C in the dark. Finally, 50 μL of Stop Solution was
236 added to terminate the reaction, and absorbance was measured at 450 nm within 5
237 minutes.

238 **Machine learning model development**

239 To construct a predictive model for MVI status (MVI+ or MVI-), we developed
240 an ensemble KNN model to estimate the probability of MVI risk using ELISA data.
241 The ensemble model integrates predictions from six base KNN models with K-values
242 of 1, 3, 5, 7, 9, and 11. The final prediction score is obtained by averaging the outputs
243 from all six models, thereby balancing local sensitivity and global stability. This
244 design addresses the limitations of single K-value models, where small K values (e.g.,
245 K=1) may overfit noise, and larger K values (e.g., K=11) may oversimplify the decision
246 boundaries.

247 **Preprocessing and analysis of scRNA-seq data**

248 The scRNA-seq data from 5 HCC samples, including 3 MVI+ tumor samples, and
249 2 MVI- tumor samples, were obtained from our previous study [23]. The Seurat
250 workflow (v4.0) with default parameters was used for downstream analyses[24].
251 Quality control was performed to filter out cells with fewer than 500 detected genes or

those with more than 30% mitochondrial gene expression. Data normalization was carried out using scTransform, with regression for mitochondrial gene expression, UMIs and detected genes. Dimensionality reduction was performed using principal component analysis (PCA) in the seurat, and the top 3000 highly variable genes were selected for downstream analysis. UMAP was used for visualization using the top 30 principal components. Cell clustering was performed using the Louvain algorithm with a resolution parameter of 0.3, and cell identities were assigned based on the clustering categories from our previous study [23].

Statistical analysis

All statistical analyses were performed using R (v4.4.1). Principal component analysis (PCA) was performed using the FactoMineR package [25]. Complete clustering was performed using the Euclidean distance on the group average protein quantitation data. The discriminatory ability of the model was assessed using receiver operating characteristic (ROC) curves, and area under the curve (AUC) values. Additionally, precision-recall (PR) curves were plotted to evaluate model performance in detecting positive cases using the PRROC package [26]. The diagnostic performance of the model was evaluated by calculating key performance metrics using the cvms package (https://github.com/LudvigOlsen/cvms), including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score. Decision curve analysis (DCA) was conducted to assess the net clinical benefit of the model at varying threshold probabilities using the dcurves package [27]. All statistical tests were two-sided, and statistical significance was considered when p-value < 0.05.

Results

Plasma proteomic landscape of MVI+ and MVI- HCC

A detailed description of the study population and design is provided in Figure 1.

Plasma samples from 50 HCC patients in the discovery cohort were subjected to quantitative proteomic analysis using the DIA strategy. DIA-MS quantified 24,994 peptides and 3,107 proteins across all plasma samples (Figure 2A), with an average of 2,453 and 2,508 proteins detected in MVI+ and MVI- plasma samples, respectively (Figure 2B). No outliers were observed, ensuring the suitability of all samples for further analysis. Among the identified proteins, 1,216 were detected with 100% completeness, 2,162 with 75% completeness, and 2,603 with 50% completeness (Figure 2C). As the sample size increased, the number of identified proteins plateaued, indicating deep proteomic coverage and stable protein detection (Figure 2D). Additionally, the number of identified protein was not influenced by age or gender (Supplementary Figure 1A).

Protein intensities spanned eight orders of magnitude, with the top 10 most abundant proteins accounting for 43.84% and 42.66% of total plasma protein abundance in the MVI+ and MVI- groups, respectively (Figure 2E). No significant differences in protein abundance were observed between the MVI+ and MVI- groups, and this consistency was maintained across different age and gender groups (Figure 2F, Supplementary Figure 1B and C). According to the Human Protein Atlas, the majority of proteins were localized to the cytoplasm (28%) and the extracellular space (27.45%) (Figure 2G). Focusing on secreted proteins, 899 of the identified proteins were classified as secreted, with 42.38% secreted into the blood, 19.13% into intracellular and membrane compartments, and 12.1% into the extracellular matrix (Figure 2H). The correlation among plasma samples was consistently above 0.90, indicating high repeatability across the samples and stability of the MS platform (Supplementary Figure 1D).

MVI-relevant functional protein module

To identify clinically relevant functional protein modules associated with MVI, we performed WGCNA on the plasma proteomic data. We began by conducting a sample clustering analysis to assess variations among the 50 samples, confirming the absence of outliers and enabling the inclusion of all samples in subsequent analyses (Supplementary Figure 2A). The network was constructed with a power of 3, achieving a scale-free topology (Figure 3A). A total of 14 protein functional modules were identified, with sizes ranging from 28 to 530 proteins (Figure 3B). These modules could also be identified independently of the WGCNA algorithm using t-SNE analysis, which demonstrated the robustness of the protein communities identified by the WGCNA algorithm (Supplementary Figure 2B). Biological functions of 14 protein modules were annotated using the GO functional enrichment analysis (Figure 3C). To explore whether any of the co-expression modules was specifically related to MVI, we correlated the module eigenproteins (MEs)—the first principal component of each module's protein expression—to the MVI phenotype across the samples. All p-values were adjusted for multiple testing using the Benjamini-Hochberg procedure to control the false discovery rate (FDR). After FDR correction ($q < 0.05$), we observed that only one module, ME09, was significantly correlated with MVI. The biological functions associated with ME09 were primarily characterized by lipid metabolic processes (Figure 3D). To further validate the consistency and robustness of this correlation, we examined the differential expression patterns of module MEs across MVI+ and MVI- subgroups. As expected, ME09, which was significantly correlated with MVI, showed downregulation in MVI+ patients when mapped onto the WGCNA network (Figure 3E).

Identification of plasma proteomic biomarkers for MVI

PCA analysis of the plasma proteomics data revealed distinct clustering between MVI+ and MVI- patients, suggesting underlying differences in protein expression

327 patterns and biological processes (Figure 4A). To further explore the molecular features
328 associated with MVI, we conducted differential protein expression analysis, and
329 identified 83 differentially expressed proteins (DEPs) between MVI+ and MVI-
330 patients. Among these, 46 proteins were significantly upregulated, and 37 were
331 downregulated in the MVI+ patients (Figure 4B). The differential abundance of these
332 83 DEPs effectively distinguished between the MVI+ and MVI- patients (Figure 4C).
333 GO enrichment analysis revealed that upregulated proteins were significantly involved
334 in ATP metabolism, immunoglobulin mediated immune response, and protein refolding
335 (Figure 4D), while downregulated proteins were associated with regulation of lipase
336 activity, leukocyte chemotaxis, and ribonucleoprotein complex biogenesis (Figure 4E).
337 To identify potential protein biomarkers for MVI, we selected the top eight
338 upregulated proteins based on stringent criteria (FDR < 0.05, log2FC > 0.58). To
339 validate the stability and reproducibility of these candidates, we performed ELISA on
340 plasma samples from the same cohort (Figure 4G). Among the eight candidates, the
341 levels of three proteins (TALDO1, PDIA3, and PGK1) measured by ELISA showed
342 consistent elevation in MVI+ patients, aligning with the trends observed in the DIA-
343 MS results (Figure 4F and G, and Supplementary Table 1). The elevated plasma levels
344 of TALDO1, PDIA3, and PGK1 in MVI+ patients compared to MVI- patients suggest
345 their relevance with MVI and highlights its potential as robust biomarkers for assessing
346 MVI status in HCC.

347 **Plasma protein-based machine learning model for preoperative MVI prediction**

348 To evaluate the clinical utility of the identified protein biomarkers in predicting
349 MVI, we developed a machine learning-based **P**lasma p**R**otein MV**I** risk **M**odel
350 (PRIM), selected based on systematic algorithm comparison (Supplementary Figure
351 S3), which integrates the expression levels of three protein biomarkers measured by

352 ELISA in the discovery cohort. PRIM was constructed using an ensemble model that
353 combines predictions from six base KNN models with values of 1, 3, 5, 7, 9, and 11
354 (Supplementary Figure 4A). PRIM achieved an AUC of 0.99 (95% CI:0.98-1.00) in
355 distinguishing between MVI+ and MVI- patients, with both specificity and sensitivity
356 at 96%, outperforming the individual biomarkers (AUCs <0.70) (Figure 5A,
357 Supplementary Figure 4B and E). MVI+ patients revealed significantly higher PRIM
358 risk scores compared to MVI- patients (Figure 5B).

359 Next, we assessed the effectiveness and robustness of PRIM by evaluating its
360 performance on the two independent validation cohorts using the ELISA-measured
361 protein levels (Supplementary Table 2 and Supplementary Table 3). PRIM achieved
362 consistent high performance in discriminating MVI+ patients from MVI- patients, with
363 AUCs of 0.82 (95% CI:0.71-0.93) and 0.78 (95% CI:0.65-0.91) in the HMUCH and
364 JU-CMU validation cohorts, respectively (Figure 5C and E, Supplementary Figure 4F
365 and G). Consistently, the performance of PRIM was superior to that of the individual
366 protein biomarkers. The PRIM risk scores were significantly higher in the MVI+
367 patients than in the MVI-patients (Figure 5D and F, Supplementary Figure 4C and D).
368 The expression levels of the three proteins (TALDO1, PDIA3 and PGK1) in MVI+
369 patients were significantly or marginally significantly higher than in MVI-patients
370 (Figure 5D and F). Additionally, PRIM maintained robust performance even in patients
371 with small tumors (<3 cm) (Figure S5A). In terms of etiology-specific evaluation,
372 PRIM performed well in both HBV-related and non-HBV/non-HCV-related patients, as
373 well as in HCV-related cases, demonstrating its broad applicability across different
374 disease backgrounds (Figure S5B-D). Notably, PRIM consistently outperformed
375 conventional clinical predictors, including tumor size, tumor grade, and serum AFP
376 levels, highlighting its superior diagnostic value (Supplementary Figure S6).

377 Furthermore, the DCA curve demonstrated that PRIM achieved higher net benefits for
378 distinguishing MVI status in HCC across a range of threshold probabilities in all three
379 cohorts (Figure 5G). These results confirm the robust and high diagnostic performance
380 of the PRIM for predicting MVI status.

381 **ScRNA-seq analysis reveals cell type- specific expression of MVI biomarkers**

382 To elucidate the cellular origins and microenvironmental dynamics of the identified
383 MVI biomarkers (TALDO1, PDIA3, PGK1), we performed scRNA-seq analysis on
384 tumor tissues surgically from five HCC patients (three MVI+, two MVI-). After quality
385 control, a total of 22,679 cells were analyzed, including 9,975 cells from MVI+ patients
386 and 12,704 from MVI- patients. Dimensionality reduction and clustering identified 10
387 major cell types according to canonical marker genes (Figure 6A and Supplementary
388 Figure 7). Consistent with our plasma proteomic findings, PDIA3, TALDO1, and
389 PGK1 showed significantly elevated expression in MVI+ tumors compared to MVI-
390 tumors (Figure 6B). Strikingly, cell type-specific analysis revealed that these
391 biomarkers were predominantly enriched in malignant cells and macrophages within
392 the MVI+ tumor microenvironment (Figure 6C). The concordance between plasma
393 proteomic elevations of TALDO1/PDIA3/PGK1 and their transcriptional upregulation
394 in MVI+ TME cell populations suggests a bidirectional crosstalk between systemic
395 circulation and local tumor biology. These findings align with emerging paradigms in
396 HCC biology, where metabolic symbiosis between tumor cells and TAMs fosters
397 invasive phenotypes [28].

398 **Discussion**

399 MVI is a critical determinant of HCC prognosis, significantly influencing post-
400 treatment recurrence rates and overall survival following curative interventions such as
401 surgical resection or radiofrequency ablation. Patients with MVI are at a higher risk of

402 the cancer recurrence after treatment, making MVI a key factor that oncologists
403 consider when planning post-treatment surveillance and management strategies.
404 Despite its clinical significance, current methods for assessing MVI remain limited to
405 postoperative histopathological evaluation, which restricts the ability to perform
406 preoperative risk stratification. In this study, we comprehensively investigated the
407 plasma proteomic landscape of HCC patients with and without MVI and successfully
408 identified distinct plasma proteomic alterations associated with MVI, which offer
409 promising biomarkers for the pretreatment prediction of MVI. To date, this is the largest
410 and most comprehensive plasma proteomic study to determine noninvasive biomarkers
411 for MVI.

412 Unlike tissue biopsy, which is limited by sampling bias and invasiveness, our DIA-
413 MS-based quantitative proteomic profiling achieved comprehensive proteome
414 coverage, identifying over 2,000 proteins with high reproducibility and stability across
415 plasma samples. The unbiased nature of this approach enabled the discovery of novel
416 biomarkers, including TALDO1, PDIA3, and PGK1. Importantly, we cross-validated
417 these biomarkers by measuring their levels in plasma by ELISA, confirming their
418 detectability in raw biological fluids, ensuring their potential translation into clinics, as
419 ELISA is a cost-effective and widely available platform in routine diagnostics. To
420 translate these findings into clinical practice, we integrated TALDO1, PDIA3, and
421 PGK1 into PRIM, a plasma protein-based diagnostic tool that demonstrated high
422 diagnostic performance across multiple cohorts. By enabling accurate preoperative
423 prediction of MVI, PRIM offers a non-invasive alternative to histopathological
424 diagnosis, facilitating personalized treatment decisions and potentially improving
425 outcomes for HCC patients.

426 Furthermore, the functional roles of these biomarkers may provide mechanistic

427 insights into MVI pathogenesis. PDIA3, a thiol-oxidoreductase chaperone, plays a
428 multifaceted role in cancer biology, including protein folding and immune regulation
429 [29, 30]. Its elevated expression in HCC, as observed in our study and supported by
430 previous reports [31], may reflect enhanced endoplasmic reticulum stress responses and
431 immune evasion mechanisms in MVI+ tumors. TALDO1, a central enzyme in the
432 pentose phosphate pathway, is integral to metabolic reprogramming in cancer [32]. Its
433 deficiency has been linked to spontaneous liver tumorigenesis, highlighting its critical
434 role in HCC progression [33]. PGK1, a key player in glycolysis, supports the rapid
435 energy demands of metastatic cells by enhancing glycolytic flux, a hallmark of cancer
436 metabolism [34]. Its upregulation in MVI+ HCC likely supports the increased energy
437 demands of invasive tumor cells, consistent with the Warburg effect [35]. Beyond their
438 mechanistic insights, these biomarkers also present promising therapeutic targets for
439 personalized treatment strategies in MVI+ patients. Recent studies have identified
440 specific inhibitors targeting two of these markers: Ilicicolin H, a non-ATP competitive
441 inhibitor of PGK1, has demonstrated dose-dependent inhibition of HCC cell
442 proliferation and apoptosis induction [36], while AO-022, an allosteric inhibitor of
443 TALDO1, has shown capacity to block tumor invasion and metastasis in cancer models
444 [37]. Although specific PDIA3 inhibitors remain underdeveloped, its role in immune
445 evasion makes it an attractive target for future drug development. These findings not
446 only elucidate the biological mechanisms underlying MVI but also provide a rational
447 foundation for potential neoadjuvant targeted therapies in patients identified with high
448 MVI risk through our proteomic approach. ScRNA-seq analysis further elucidates the
449 cellular origins of these biomarkers and revealed that these biomarkers were
450 significantly upregulated in both macrophages and malignant cells, highlighting a
451 potential synergistic relationship between tumor cells and immune cells. This

interaction may drive vascular invasion through coordinated metabolic and immune regulatory mechanisms. Specifically, the upregulation of PDIA3 in malignant cells may enhance immune evasion, while TALDO1 and PGK1 likely support the metabolic demands of both cell types, fostering a pro-invasive microenvironment. These findings align with previous reports suggesting that macrophage-tumor cell interactions promote inflammation, tumor growth, and extracellular matrix (ECM) remodeling [38].

Despite these promising findings, several limitations must be acknowledged. First, while our multicenter design enhances generalizability in Chinese patient population, further validation in more geographically and ethnically diverse populations is needed. Second, the retrospective nature of biomarker selection may have overlooked additional proteins potentially involved in MVI, suggesting the value of integrating multi-omics data in future studies to identify novel biomarkers and enhance the predictive capacity of the model.

In conclusion, our study provides a comprehensive analysis of the plasma proteomic landscape and identifies three plasma protein biomarkers specific to MVI-positive patients. By developing and validating a robust, non-invasive plasma protein diagnostic panel, we present a promising tool for the early identification of MVI in HCC, with potential clinical implications for patient management and therapeutic decision-making.

Abbreviations

AUROC, areas under the receiver operating characteristic curve; DCA; decision curve analysis; DEP, differentially expressed protein; DIA-MS, data-independent acquisition mass spectrometry; ELISA, enzyme-linked immunosorbent assay; HCC, hepatocellular carcinoma; MVI, microvascular invasion; TME, tumor microenvironment; WGCNA, weighted gene co-expression network analysis

Statement of Ethics

Written informed consent was obtained from all participants, and the study was

479 approved by the Ethics Committees of each participating institution, including the First
480 Affiliated Hospital of Harbin Medical University (Approval No. 202476), Harbin
481 Medical University Cancer Hospital (Approval No. KY-2024-65), the Second Norman
482 Bethune Hospital of Jilin University (Approval No.2024-97-3), and the First Hospital
483 of China Medical University (Approval No. 2024-083). All procedures adhered to the
484 principles of the Declaration of Helsinki. Written informed consent was obtained from
485 all participants prior to any study-related procedures.

486 **Data availability Statement**

487 The datasets used and/or analysed during the current study are available from the
488 corresponding author on reasonable request. All datasets and software used for analysis
489 are listed in the CTAT table. Further inquiries can be directed to the corresponding
490 authors.

491 **References**

492 1. Zheng, R.S., et al., *[Cancer incidence and mortality in China, 2022].* Zhonghua Zhong Liu
493 Za Zhi, 2024. **46**(3): p. 221-231.
494 2. Huo, T.I., S.Y. Ho, and J.I. Liao, *Predicting post-resection early recurrence of hepatocellular*
495 *carcinoma: Defining the role of microvascular invasion.* Liver Int, 2023. **43**(12): p. 2826-
496 2827.
497 3. Beaufrere, A., et al., *Gene expression signature as a surrogate marker of microvascular*
498 *invasion on routine hepatocellular carcinoma biopsies.* J Hepatol, 2022. **76**(2): p. 343-352.
499 4. Han, J., et al., *The impact of resection margin and microvascular invasion on long-term*
500 *prognosis after curative resection of hepatocellular carcinoma: a multi-institutional study.*
501 HPB (Oxford), 2019. **21**(8): p. 962-971.
502 5. Ma, Y.N., et al., *Neoadjuvant therapies in resectable hepatocellular carcinoma: Exploring*
503 *strategies to improve prognosis.* Biosci Trends, 2024. **18**(1): p. 21-41.
504 6. Yang, P., et al., *Liver resection versus liver transplantation for hepatocellular carcinoma*
505 *within the Milan criteria based on estimated microvascular invasion risks.* Gastroenterol
506 Rep (Oxf), 2023. **11**: p. goad035.
507 7. Alvez, M.B., et al., *Next generation pan-cancer blood proteome profiling using proximity*
508 *extension assay.* Nat Commun, 2023. **14**(1): p. 4308.
509 8. Bech, J.M., et al., *Proteomic Profiling of Colorectal Adenomas Identifies a Predictive Risk*
510 *Signature for Development of Metachronous Advanced Colorectal Neoplasia.*
511 Gastroenterology, 2023. **165**(1): p. 121-132 e5.
512 9. Davies, M.P.A., et al., *Plasma protein biomarkers for early prediction of lung cancer.*

EBioMedicine, 2023. **93**: p. 104686.

10. Harel, M., et al., *Longitudinal plasma proteomic profiling of patients with non-small cell lung cancer undergoing immune checkpoint blockade.* J Immunother Cancer, 2022. **10**(6).
11. Lapitz, A., et al., *Liquid biopsy-based protein biomarkers for risk prediction, early diagnosis, and prognostication of cholangiocarcinoma.* J Hepatol, 2023. **79**(1): p. 93-108.
12. Li, Y., et al., *Longitudinal plasma proteome profiling reveals the diversity of biomarkers for diagnosis and cetuximab therapy response of colorectal cancer.* Nat Commun, 2024. **15**(1): p. 980.
13. Liang, H., et al., *LcProt: Proteomics-based identification of plasma biomarkers for lung cancer multievent, a multicentre study.* Clin Transl Med, 2025. **15**(1): p. e70160.
14. Qu, Y., et al., *Plasma proteomic profiling discovers molecular features associated with upper tract urothelial carcinoma.* Cell Rep Med, 2023. **4**(9): p. 101166.
15. Zhang, Z., et al., *Proteogenomic Characterization of High-Grade Lung Neuroendocrine Carcinoma Deciphers Molecular Diversity and Potential Biomarkers of Different Histological Subtypes in Chinese Population.* Research (Wash D C), 2025. **8**: p. 0671.
16. Xing, X., et al., *Proteomics-driven noninvasive screening of circulating serum protein panels for the early diagnosis of hepatocellular carcinoma.* Nat Commun, 2023. **14**(1): p. 8392.
17. Yi, X., et al., *Proteome Landscapes of Human Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma.* Mol Cell Proteomics, 2023. **22**(8): p. 100604.
18. Zhang, X., et al., *Prediagnostic plasma proteomics profile for hepatocellular carcinoma.* J Natl Cancer Inst, 2024. **116**(8): p. 1343-1355.
19. Zhang, Z., *Introduction to machine learning: k-nearest neighbors.* Ann Transl Med, 2016. **4**(11): p. 218.
20. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.
21. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* OMICS, 2012. **16**(5): p. 284-7.
22. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.* Bioinformatics, 2009. **25**(8): p. 1091-3.
23. Li, K., et al., *Single-cell dissection of the multicellular ecosystem and molecular features underlying microvascular invasion in HCC.* Hepatology, 2024. **79**(6): p. 1293-1309.
24. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587 e29.
25. Lê, S., J. Josse, and F. Husson, *FactoMineR: An R Package for Multivariate Analysis.* Journal of Statistical Software, 2008. **25**(1): p. 1 - 18.
26. Grau, J., I. Grosse, and J. Keilwagen, *PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R.* Bioinformatics, 2015. **31**(15): p. 2595-7.
27. Vickers, A.J. and E.B. Elkin, *Decision curve analysis: a novel method for evaluating prediction models.* Med Decis Making, 2006. **26**(6): p. 565-74.
28. Chen, D., et al., *Metabolic regulatory crosstalk between tumor microenvironment and tumor-associated macrophages.* Theranostics, 2021. **11**(3): p. 1016-1030.
29. Clement, C.C., et al., *PDIA3 epitope-driven immune autoreactivity contributes to hepatic damage in type 2 diabetes.* Sci Immunol, 2022. **7**(74): p. eabl3795.

30. Mahmood, F., et al., *PDIA3: Structure, functions and its potential role in viral infections.* Biomed Pharmacother, 2021. **143**: p. 112110.
31. Takata, H., et al., *Increased expression of PDIA3 and its association with cancer cell proliferation and poor prognosis in hepatocellular carcinoma.* Oncol Lett, 2016. **12**(6): p. 4896-4904.
32. Jin, L. and Y. Zhou, *Crucial role of the pentose phosphate pathway in malignant tumors.* Oncol Lett, 2019. **17**(5): p. 4213-4221.
33. Hanczko, R., et al., *Prevention of hepatocarcinogenesis and increased susceptibility to acetaminophen-induced liver failure in transaldolase-deficient mice by N-acetylcysteine.* J Clin Invest, 2009. **119**(6): p. 1546-57.
34. Xie, H., et al., *PGK1 Drives Hepatocellular Carcinoma Metastasis by Enhancing Metabolic Process.* Int J Mol Sci, 2017. **18**(8).
35. DeBerardinis, R.J. and N.S. Chandel, *Fundamentals of cancer metabolism.* Sci Adv, 2016. **2**(5): p. e1600200.
36. Li, M., et al., *A novel inhibitor of PGK1 suppresses the aerobic glycolysis and proliferation of hepatocellular carcinoma.* Biomed Pharmacother, 2023. **158**: p. 114115.
37. Xu, G., et al., *Proteomic Profiling of Serum Extracellular Vesicles Identifies Diagnostic Signatures and Therapeutic Targets in Breast Cancer.* Cancer Res, 2024. **84**(19): p. 3267-3285.
38. Kloosterman, D.J. and L. Akkari, *Macrophages at the interface of the co-evolving cancer ecosystem.* Cell, 2023. **186**(8): p. 1627-1651.

580 **Figure Legends**

581 **Figure 1. Study design of the analyzed cohort and experiment workflow.** Created
582 by Biorender.com.

583 **Figure 2. Plasma proteomic landscape of MVI+ and MVI- HCC.** A. Number of
584 proteins identified by quality control. B. Number of proteins identified in the two
585 groups. C. Data completeness curve. The curve highlights the data completeness at
586 thresholds of 50%, 75% and 100%, with arrows marking these key values. D.
587 Cumulative number of identified proteins. E. The protein abundance distributions in
588 MVI+ and MVI- samples. F. Density plot of protein abundance in MVI+ and MVI-
589 samples. G. Radar plot of protein subcellular localization. H. Annotation of secreted
590 proteins.

591 **Figure 3. Functional protein module associated with MVI**. A. Soft-threshold plot for
592 WGCNA. B. Gene dendrogram with different colors showing the modules identified
593 by WGCNA. C. WGCNA identified 14 functional protein modules (ME01–14)
594 enriched in proteomic data. Each network node represents a protein, color-coded
595 according to the different functional modules. D. The relationship between gene
596 modules and the MVI phenotype. The strengths of the positive (red) and negative (blue)
597 correlations are shown in the two-color heatmap. Pearson correlation coefficients and
598 FDR were calculated using the WGCNA package. E. Bar plot showing the module score
599 of the 14 protein modules in the MVI+ and MVI- groups. P-values were calculated
600 using the Mann–Whitney U test and adjusted for multiple comparisons using the FDR
601 correction.

602 **Figure 4. Plasma protein biomarkers of MVI.** A. Principal component analysis (PCA)
603 of proteins in plasma samples from MVI+ and MVI- groups. B. Volcano plot showing
604 differentially abundant proteins between MVI+ and MVI- samples. C. Heat map of

differentially expressed proteins downregulated (blue) and upregulated (red) in MVI+ samples. D. Enrichment network for upregulated proteins in MVI+ samples. E. Enrichment network for downregulated proteins in MVI+ samples. F. Box plot illustrating the abundance of potential biomarker proteins in MVI+ and MVI- samples based on proteomics data. G. Box plot illustrating the abundance of potential biomarker proteins in MVI+ and MVI- samples based on ELISA data.

Figure 5. Plasma protein-based machine learning model for preoperative MVI prediction. A. ROC analysis for three protein biomarkers and PRIM to predict MVI in the FAHHMU cohort (up left). Confusion matrix showing the classification results of the model in the FAHHMU cohort (up right), performance metrics including AUC, specificity, sensitivity for the model in the FAHHMU cohort (bottom). B. PRIM score and expression of three protein biomarkers in MVI+ and MVI- samples in the FAHHMU cohort. C. ROC analysis for three protein biomarkers and PRIM in identifying MVI+ patients in HMUCH cohort (up left). Confusion matrix showing the classification results of the model in the HMUCH cohort (up right), performance metrics including AUC, specificity, sensitivity for the model in the HMUCH cohort (bottom). D. PRIM score and expression of three protein biomarkers in MVI+ and MVI- samples in the HMUCH cohort. E. ROC analysis for three protein biomarkers and PRIM in identifying MVI+ patients in the JU-CMU cohort (up left). Confusion matrix showing the classification results of the model in the JU-CMU cohort (up right), performance metrics including AUC, specificity, sensitivity for the model in JU-CMU cohort (bottom). F. PRIM score and expression of three protein biomarkers in MVI+ and MVI- samples in the JU-CMU cohort. G. Decision curve analysis to assess clinical benefit (left: FAHHMU cohort, middle: HMUCH cohort, right: JU-CMU cohort).

Figure 6. A cell type-resolved atlas of biomarker expression by ScRNA-seq. A. UMAP visualization showing the major cell types of HCC tumors (left: total samples, middle: MVI+ samples, right: MVI- samples). B. UMAP of single-cell transform–normalized PDIA3, TALDO1, and PGK1 expression, and Bar plot showing gene expression of three marker proteins in all cells in MVI+ and MVI- groups. ***P < 0.001; **P < 0.01; *P < 0.05. C. Bar plot showing cell type gene expression of three marker proteins in MVI+ and MVI- groups. ***P < 0.001; **P < 0.01; *P < 0.05.

Patients cohorts

First Affiliated Hospital of Harbin Medical University

FAHHMU cohort
MVI+(n=25) MVI-(n=25)

Harbin Medical University Cancer Hospital

HMUCH cohort
MVI+(n=30) MVI-(n=30)

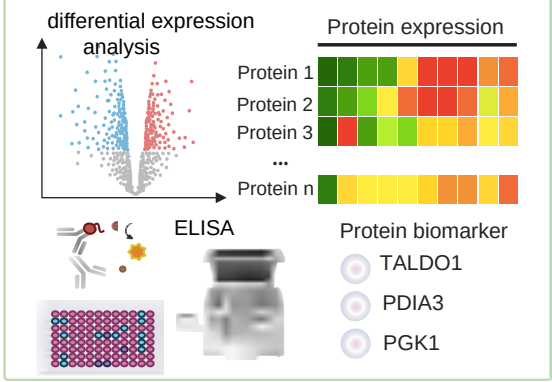
Second Norman Bethune Hospital of Jilin University

First Hospital of China Medical University

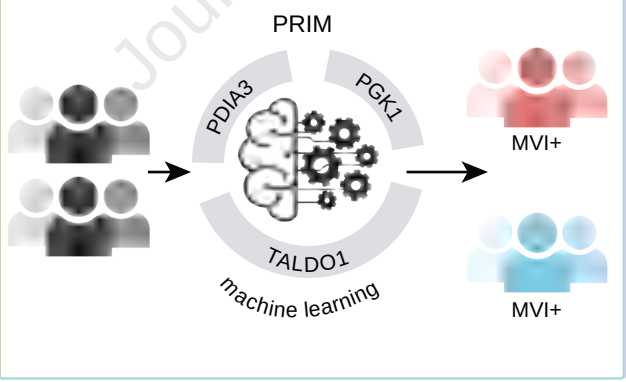
JU-CMU cohort
MVI+(n=25) MVI-(n=25)

Plasma proteomics

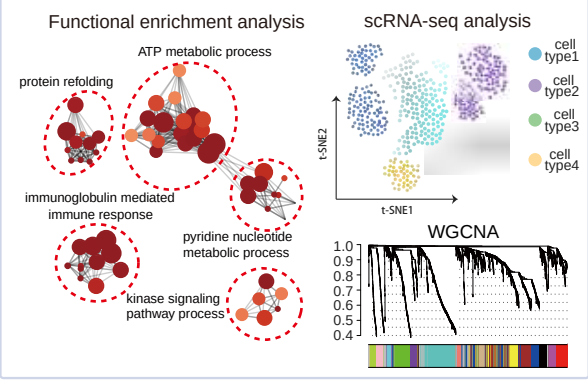
Biomarker identification

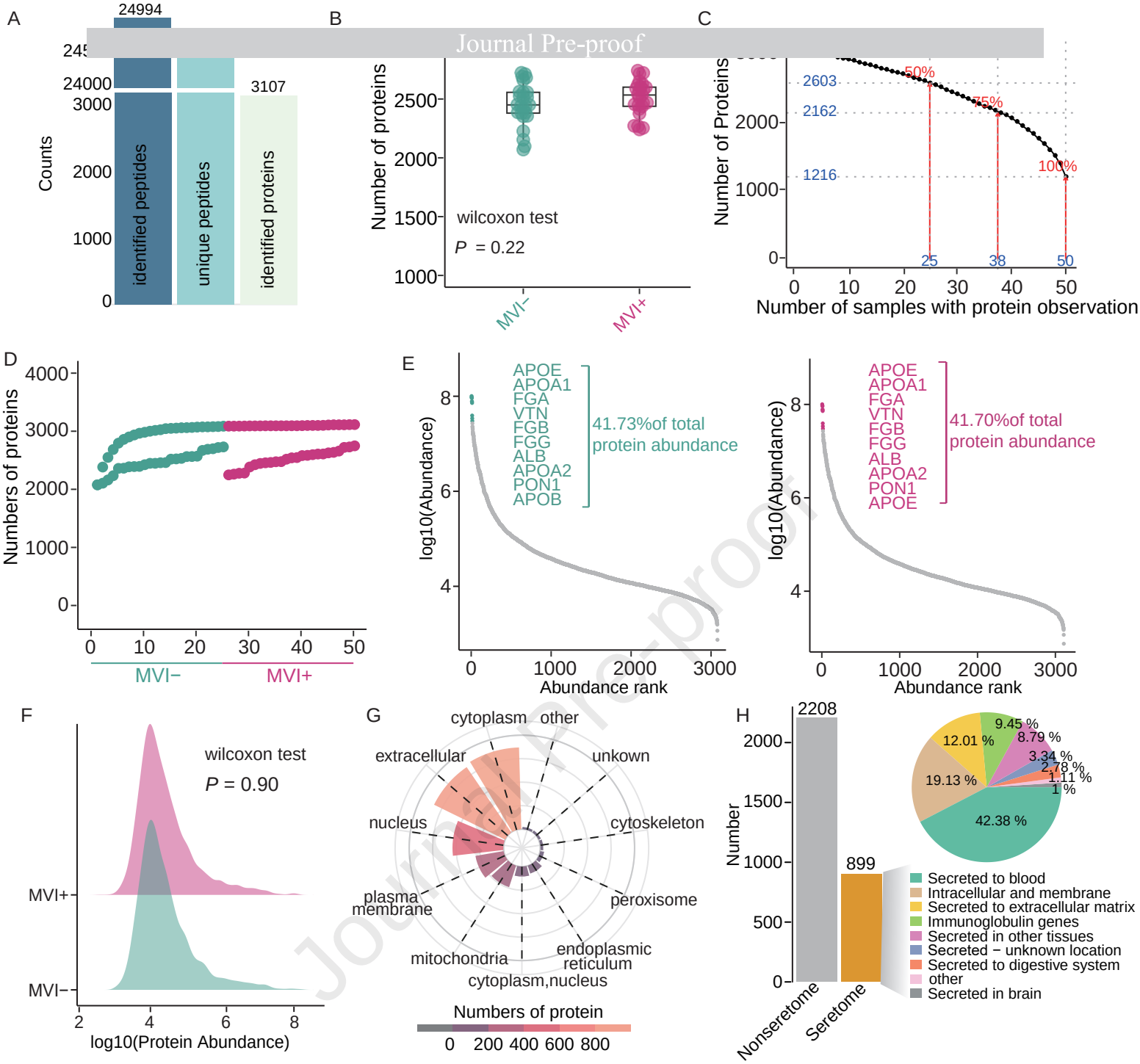


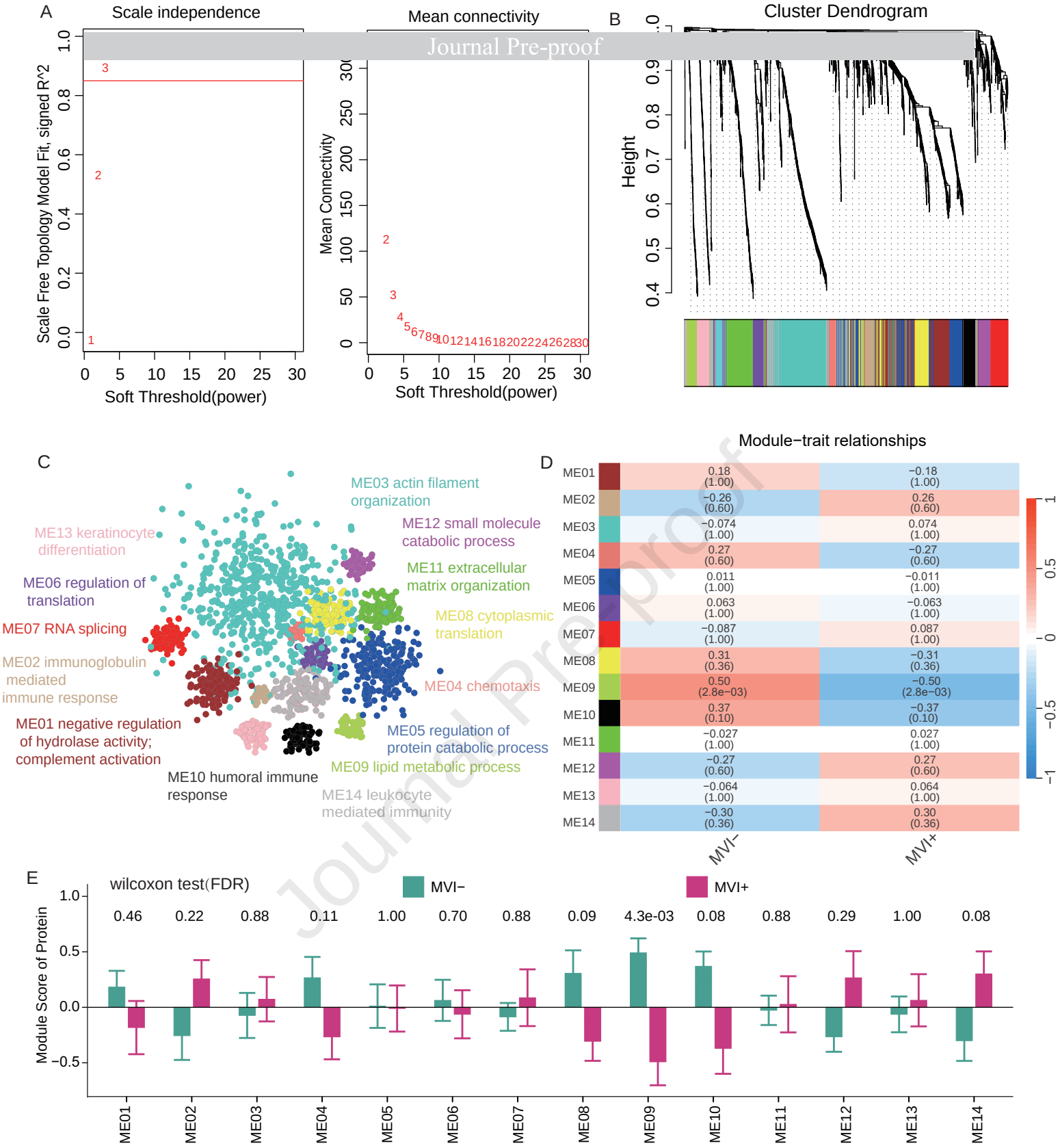
Model establishment & validation

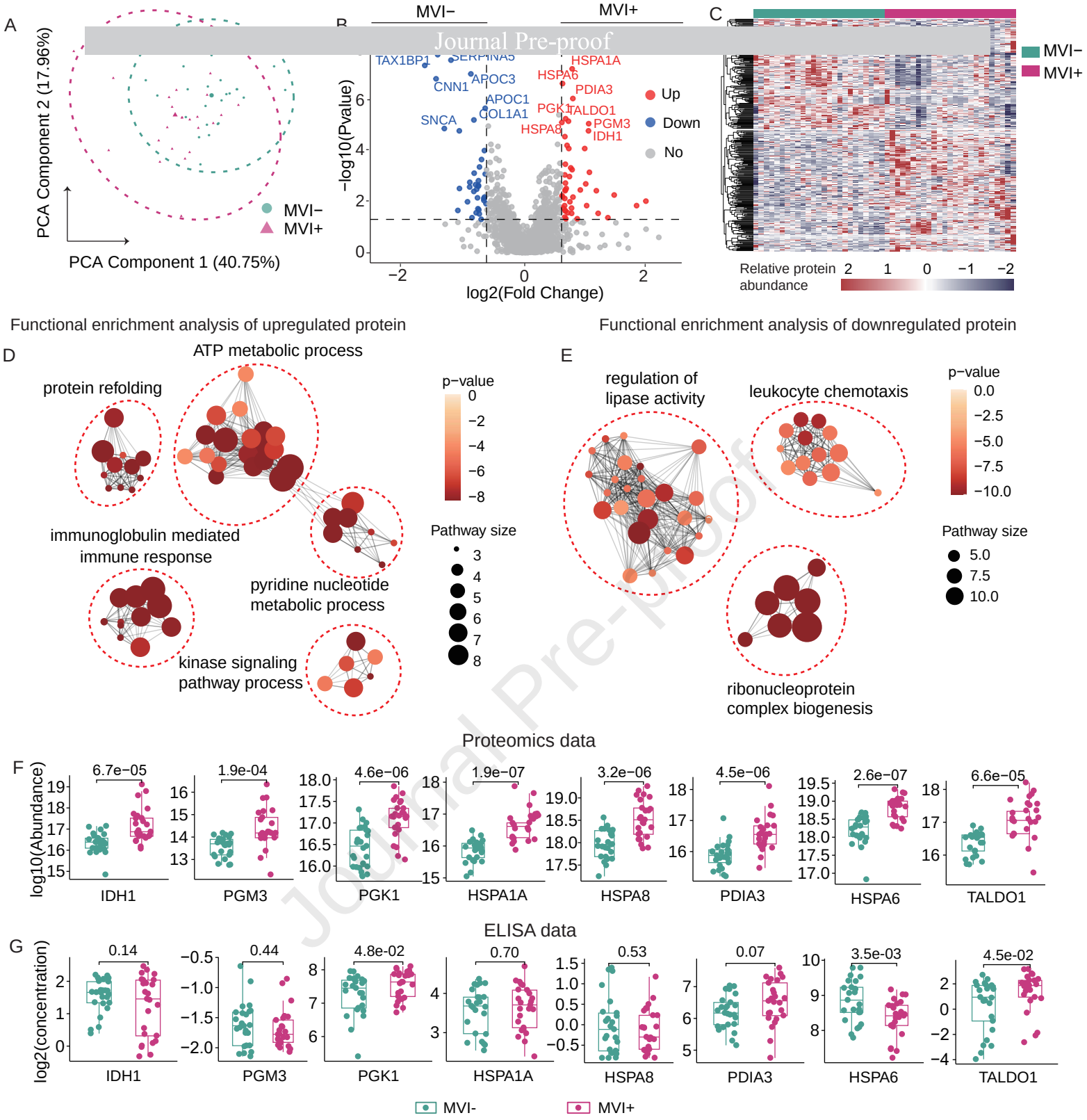


Biological interpretation

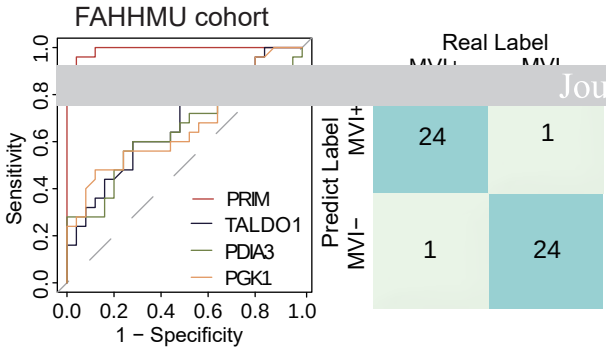






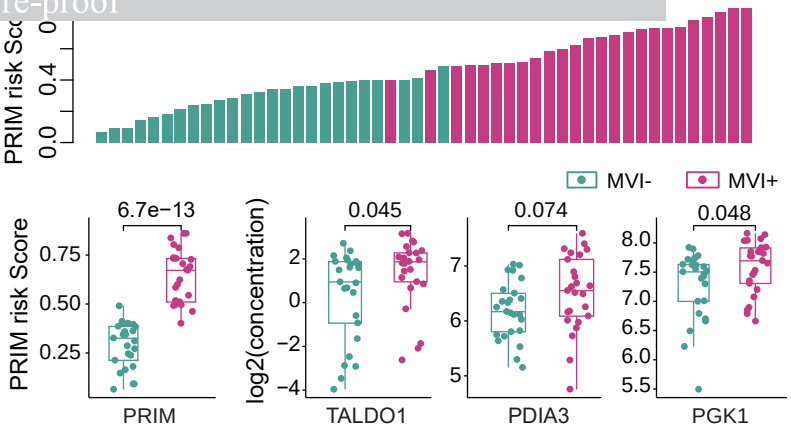


A

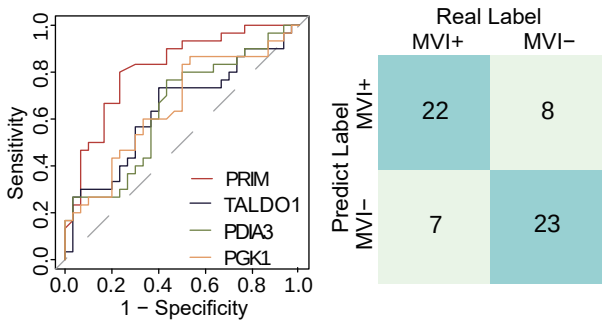


Protein & Model	MVI+ (n = 25) *vs.* MVI- (n = 25)		
	AUC	SEN	SPE
PRIM	0.99	0.96	0.96
TALDO1	0.69	0.60	0.60
PDIA3	0.66	0.60	0.60
PGK1	0.67	0.56	0.56

B

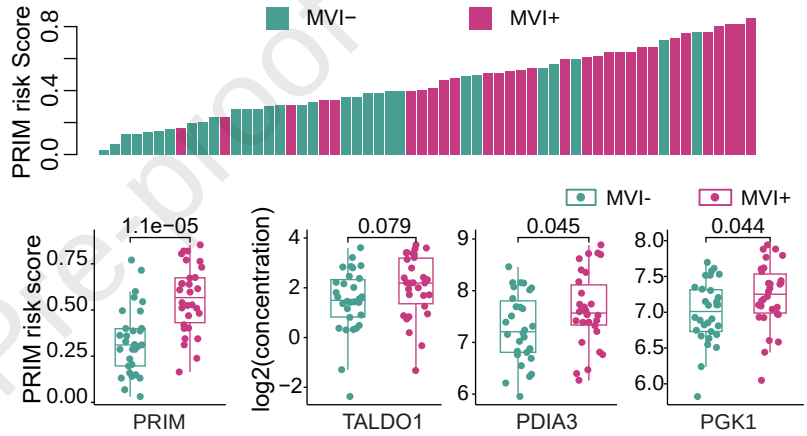


C

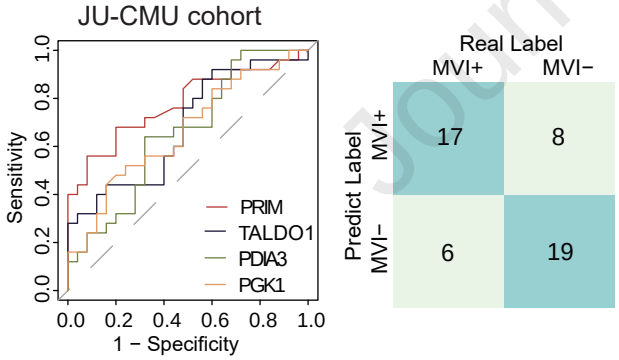


Protein & Model	MVI+ (n = 30) *vs.* MVI- (n = 30)		
	AUC	SEN	SPE
PRIM	0.82	0.73	0.77
TALDO1	0.64	0.73	0.43
PDIA3	0.64	0.60	0.60
PGK1	0.66	0.70	0.53

D

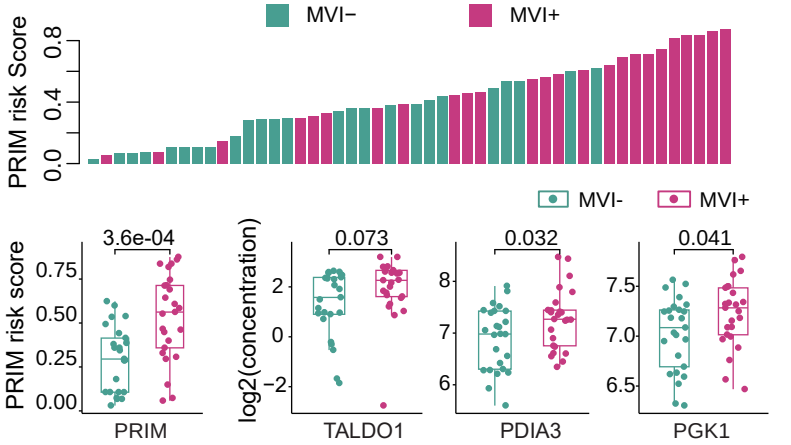


E

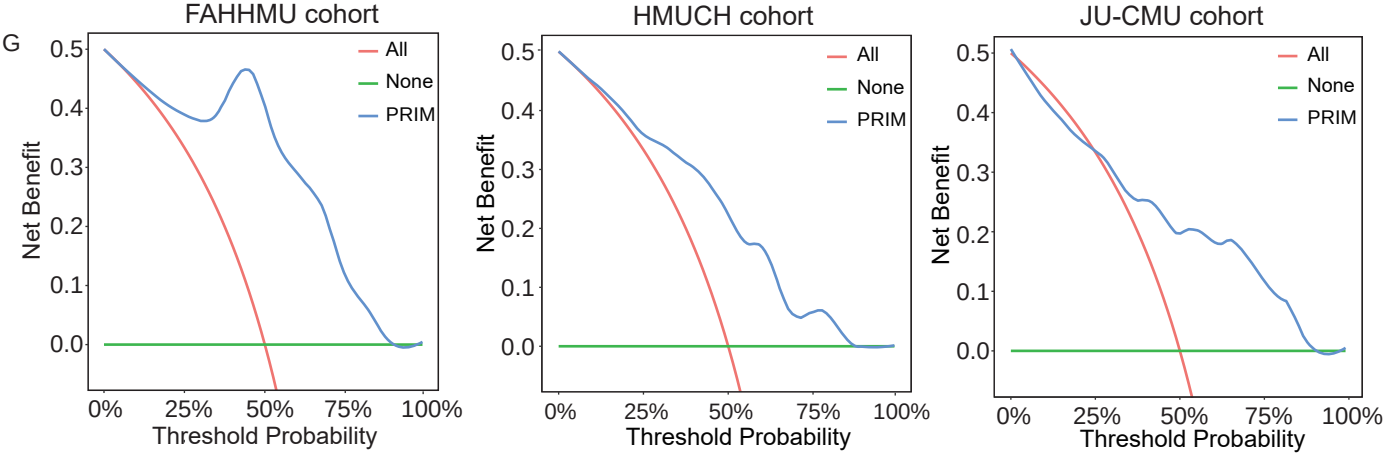


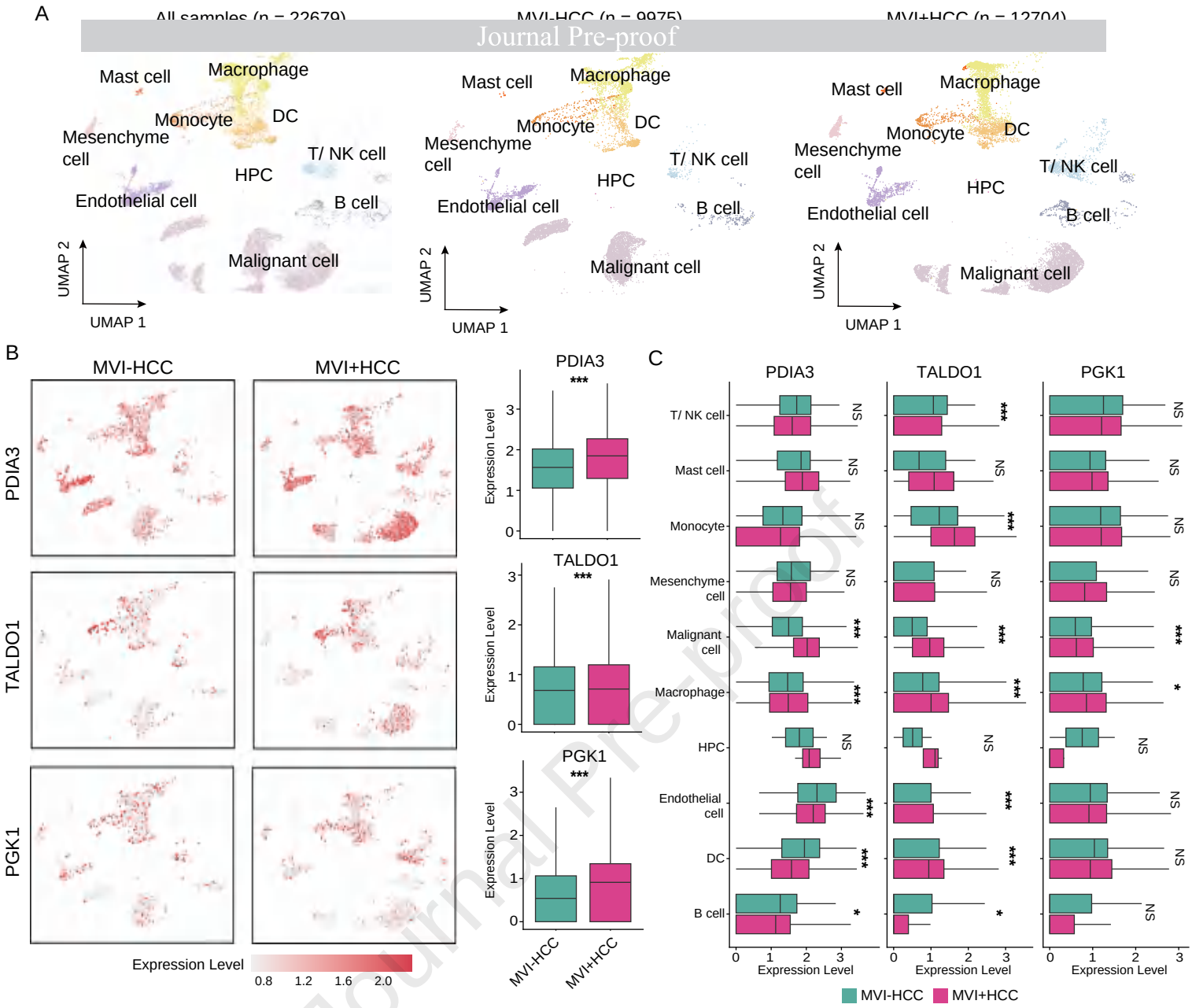
Protein & Model	MVI+ (n = 25) *vs.* MVI- (n = 25)		
	AUC	SEN	SPE
PRIM	0.78	0.68	0.76
TALDO1	0.68	0.56	0.56
PDIA3	0.66	0.56	0.64
PGK1	0.66	0.60	0.56

F



G





MVI-HCC MVI+HCC

Highlights

• Plasma proteome profiling can differentiate between MVI+ and MVI- patients.

• A plasma protein-based model was developed for preoperative MVI prediction.

• ScRNA-seq analysis reveals cell-type-specific expression of MVI biomarkers.