BioData Mining

RESEARCH

Open Access

Integrating transcriptomics and proteomics to analyze the immune microenvironment of cytomegalovirus associated ulcerative colitis and identify relevant biomarkers

Yang Chen[1,2,3†], Qingqing Zheng[2,3†], Hui Wang[2,3†], Peiren Tang[2,3], Li Deng[2,3], Pu Li[4], Huan Li[2,3], Jianhong Hou[6*], Jie Li[5*], Li Wang[2,3*] and Jun Peng[6*]

†Yang Chen, Qingqing Zheng and Hui Wang contributed equally to this work.

*Correspondence:
Jianhong Hou
hjh@kust.edu.cn
Jie Li
lijie@kmmu.edu.cn
Li Wang
2001WL@163.com
Jun Peng
9y140211@kust.edu.cn

Full list of author information is available at the end of the article

Abstract

Background In recent years, significant morbidity and mortality in patients with severe inflammatory bowel disease (IBD) and cytomegalovirus (CMV) have drawn considerable attention to the status of CMV infection in the intestinal mucosa of IBD patients and its role in disease progression. However, there is currently no high-throughput sequencing data for ulcerative colitis patients with CMV infection (CMV + UC), and the immune microenvironment in CMV + UC patients have yet to be explored.

Method The xCell algorithm was used for evaluate the immune microenvironment of CMV + UC patients. Then, WGCNA analysis was explored to obtain the co-expression modules between abnormal immune cells and gene level or protein level. Next, three machine learning approach include Random Forest, SVM-rfe, and Lasso were used to filter candidate biomarkers. Finally, Best Subset Selection algorithms was performed to construct the diagnostic model.

Results In this study, we performed transcriptomic and proteomic sequencing on CMV + UC patients to establish a comprehensive immune microenvironment profile and found 11 specific abnormal immune cells in CMV + UC group. After using multi-omics integration algorithms, we identified seven co-expression gene modules and five co-expression protein modules. Subsequently, we utilized various machine learning algorithms to identify key biomarkers with diagnostic efficacy and constructed an early diagnostic model. We identified a total of eight biomarkers (PPP1R12B, CIRBP, CSNK2A2, DNAJB11, PIK3R4, RRBP1, STX5, TMEM214) that play crucial roles in the immune microenvironment of CMV + UC and exhibit superior diagnostic performance for CMV + UC.

Conclusion This 8 biomarkers model offers a new paradigm for the diagnosis and treatment of IBD patients post-CMV infection. Further research into this model will be significant for understanding the changes in the host immune microenvironment following CMV infection.

Keywords Inflammatory bowel disease, CMV+UC, Multi-omics, Immune microenvironment, Machine learning, Diagnostic biomarkers

Background

In North America, Oceania, and numerous European countries, inflammatory bowel disease (IBD), comprising ulcerative colitis (UC) and Crohn's disease (CD), is a prevalent condition.IBD rates have been increasing in Asia, South America, and Africa in recent years [1–3]. IBD patients may experience a decrease in their quality of life due to various issues such as frequent hospital stays, side effects from medications, surgeries, and stoma formation [4]. Therefore, IBD patients require long-term and systematic treatment. The exact cause of IBD is still not completely understood, but it is commonly thought to be a developing long-lasting intestinal condition influenced by ongoing environmental, genetic, infectious, and immunological factors [5]. Compared to the research on the role of gut microbiota in IBD pathogenesis, the involvement of gut viruses in IBD has recently garnered attention [6, 7]. Epstein-Barr virus (EBV) and cytomegalovirus (CMV) infections are prevalent and usually contracted during childhood [8]. Researchers have shown significant interest in the relationship between EBV and CMV infections in the intestinal mucosa of IBD patients and their impact on disease advancement. Reported frequencies of CMV infection in severe acute UC range from 21 to 34%, with refractory cases showing frequencies of 33–36% [9, 10]. EBV infection rates are even higher, ranging from 33 to 81% [11, 12]. While there is a connection between CMV or EBV infection and severe colitis, the link between these viruses and IBD is not as straightforward as with other factors. The involvement of these viruses in IBD, whether as participants or passive observers, continues to be a subject of ongoing debate.

Human CMV, a type of β-herpesvirus and also referred to as human herpesvirus 5 (HHV-5), is the largest virus in the human herpesvirus group, with widespread prevalence in the population. The infection rate varies by country and region, reaching 70–100%, and 40–100% of adults are infected before the age of 40 [13, 14]. CMV has a double-stranded linear DNA genome approximately 235 kb in length, encoding about 165 proteins. CMV gene products are present in various human malignancies, typically associated with tumor cells and the tumor vasculature [15]. CMV can infect multiple organs and persist in various adult stem cells, inducing the expression of latency-associated CMV proteins, promoting stem cell proliferation, causing genomic instability, and leading to immune evasion [15]. Healthy hosts are usually asymptomatic during primary CMV infection, although mild symptoms such as fever and lymphadenopathy may occur [16]. However, like many other herpesviruses, CMV remains latent in the host and can reactivate later in life. In immunocompromised patients, it may cause severe systemic diseases such as pneumonia, hepatitis, and colitis [17]. Patients with IBD may experience reactivation even while undergoing immunosuppressive treatment [18]. The gastrointestinal tract, especially the colon, is commonly affected during CMV reactivation, leading to acute colitis. The link between CMV and IBD has been recorded in publications dating back to 1961 [19]. However, the nature of this relationship remains hotly debated. Early studies suggested that CMV infection might perpetuate IBD, but it is now generally accepted that CMV colitis predominantly occurs in patients with preexisting IBD [9]. Patients with severe comorbid IBD and CMV infection exhibit significantly higher

morbidity and mortality rates [18, 20]. For instance, compared to colitis patients without CMV, those with concurrent IBD and CMV infection require more surgical interventions and have higher in-hospital mortality rates [18]. Additionally, studies in UC suggest that CMV could cause acute colitis that does not respond to steroids and worsen the prognosis of the disease [21–23]. The prevalence of CMV colitis in IBD patients is estimated to be 0.53–4% [18]. Yet, in individuals with severe steroid-resistan, the estimated frequency is believed to be significantly greater, approximately 36% [18]. Indeed, routine histology and immunohistochemistry (IHC) reveal that approximately 11.7% of colectomy specimens from adult UC patients are CMV-positive [24].

Immunosuppression plays a crucial role in increasing the risk of CMV colitis in IBD patients. In UC patients, the use of high-dose systemic corticosteroids for more than one month is an independent risk factor for CMV-associated colitis [25]. Other studies have also confirmed the association of corticosteroids and other immunomodulators (thiopurines and methotrexate) with CMV-positive IBD [26, 27]. According to the 2019 guidelines from the American Gastroenterological Association (AGA), it is recommended that adults with acute severe UC undergo sigmoidoscopy to assess for the presence of CMV colitis [28]. Likewise, the European Society for Paediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) consensus statement advises that children with steroid-refractory IBD should undergo sigmoidoscopy and pathological assessment for CMV infection [29]. The seropositivity rate of CMV in IBD patients varies with age: 38% in children under 10 years, 22-25% in children aged 11–19 years, and nearly 90% in adult IBD patients [30, 31]. Currently, the infection rate of CMV in individuals with IBD shows substantial variation because of discrepancies in methodologies or diagnostic techniques. Particularly in China, research on viral infections in the intestinal mucosa of patients with IBD is limited. The specific mechanisms by which CMV infection exacerbates IBD progression remain underexplored, and currently, there are no sequencing data targeting IBD patients following CMV infection. Thus, this study aims to characterize the immune microenvironment of IBD patients with CMV infection, further explore immune-related molecular markers in these patients using high-throughput sequencing data, identify new biomarkers, and provide novel insights and strategies for the diagnosis and treatment of CMV-positive IBD.

Methods

Data sources

Ulcerative colitis patients with CMV infection (CMV+UC) and without CMV infection (CMV- UC), and normal control with CMV infection (CMV+N) and without CMV infection (CMV- N), were pathologically confirmed after colonoscopy from January 2021 to April 2023 in the Department of Pathology of the First People's Hospital of Yunnan Province. Intestinal tissue from the above 4 groups (3 patients per group) were collected and then performed transcriptome and proteome sequencing. Ethical approval for this study was obtained from the Ethics Committee of the First People's Hospital of Yunnan Province (KHLL2024-KY199). In addition, all patients or their legal guardians who participated in the study provided written informed consent. The research was carried out following the guidelines of the Helsinki Declaration (2013 Update).

Transcriptome

Sample extraction and pre-processing were performed as previously described [32]. Afterward, TrimGalore and FastQC software were used to perform quality control and preprocessing on raw data of ulcerative colitis patients. After removing the adapter and ploy-N reads from raw data, clean data (clean reads) were obtained. Using STAR software, paired-end clean reads were aligned to the reference genome (hg38), and then the expression counts were calculated. The RPKM expression value of each gene was calculated based on the count number. Here, we obtained the expression levels of 58,375 transcripts for further analysis.

Proteome

FFPE samples of 12 patients used for proteome sequencing were obtained from the sample bank of the Department of Pathology, First People's Hospital of Yunnan Province. Samples were then underwent total protein extraction by the following procedure: Initially, the samples were dewaxed with octane and hydrated using graded ethanol. Following hydration, they were washed twice with phosphate-buffered saline (PBS). After removing the PBS, an appropriate amount of protein lysis buffer (4% SDS, 100 mM Tris) was added. The samples were then incubated at 95 °C for 10 min, shaken, and mixed thoroughly, followed by sonication for 5 min in an ice water bath. The samples were decrosslinked at 95 °C for 60 min, then reduced by adding an appropriate amount of TCEP and alkylated with CAA at 95 °C for 5 min. The samples were centrifuged at 12,000 g for 15 min at 4 °C, after which the supernatant was collected. To the supernatant, four times its volume of -20 °C pre-cooled acetone was added, and the sample was precipitated at -20 °C for at least 4 h. This was followed by centrifugation at 12,000 g for 15 min at 4 °C. The resulting precipitate was collected air-dried and then dissolved in a protein solution containing 8 M urea and 100 mM triethylammonium bicarbonate (TEAB) at pH 8.5. The extracted total protein was quantified using the Bradford assay. Samples that passed the protein quality check were used to generate raw data for mass spectrometry using the Q Exactive™ HF-X mass spectrometer in data-independent acquisition (DIA) mode.

The mass spectrometry downlink data were used for protein sequence identification based on the Uniport database using Spectronaut-Pulsar (Biognosys) software. The search parameters were set as follows: 10 ppm mass tolerance for the precursor ion and 0.02 Da mass tolerance for the production. A maximum of two missed sites were allowed. Only peptide spectral matches (PSMs) with more than 99% confidence were identified as PSMs. Identified proteins contained at least one unique peptide. Then, the DIA data were imported into Spectronaut (Biognosys) software to generate a DDA library and extract ion-pair chromatographic peaks. Ions were matched and peak areas were calculated for peptide characterization and quantification. The iRT was added to the samples to correct the retention time, and the precursor ion Q cut-off value was set at 0.01. A total of 4564 proteins were obtained for further analysis.

Immune microenvironment analysis

The xCell algorithm was used for evaluate the immune microenvironment of these patients. The method can assess the level of infiltration of up to 64 cell types, including multiple adaptive and innate immune cells, haematopoietic progenitor cells, epithelial cells and extracellular matrix cells, based on gene expression data. The expression

profiles of immune cells and stromal cells were first extracted from the bulk gene expression data as cell Signatures. Enrichment scores for the samples on each cell type Signature were calculated using ssGSEA. The enrichment scores for each cell type were converted to the corresponding cell type scores using a fitting formula. Finally, xCell performed a compensating correction for scores of closely related cell types, reducing the effect of possible covariance/correlation between different cell types.

Differential expression analysis

The comparison of gene level between the CMV+IBD and control groups (CMV+IBD vs CMV- IBD, CMV+IBD vs CMV+N, CMV- IBD vs CMV- N, CMV+N vs CMV- N) were conducted with the calculation of the 'limma' R package. The differentially expressed genes (DEGs) were detected based on an absolute log2 fold change greater than or equal to 0.585 and an adjusted P-value less than 0.05.

We also screened for immune cells with abnormal abundance in these groups using an adjusted P-value threshold of under 0.05.

On the other hand, the differentially expressed proteins (DEPs) in these groups were determined by an adjusted P-value below 0.05 and an absolute log2 fold change of at least 0.263.

Weighted correlation network analysis (WGCNA)

The co-expression networks of mRNA expression level and protein expression level in CMV+IBD was built using the 'WGCNA' R package and the automatic network construction function. Next, hierarchical clustering was used to identify gene/protein modules with similar expression patterns. Then, characteristics of abnormal immune cells in CMV+IBD patients were ultimately linked to these modules, and the key genes/proteins from the selected module were investigated for further analysis.

Correlation network analysis

Using Pearson analysis to construct the regulatory networks between gene modules and protein modules associated with the same abnormal immune cells. The threshold of selected correlationship between genes and proteins was set as the absolute correlation coefficient ≥ 0.5 and P-value < 0.05.

Machine learning analysis

Analysis of different co-expression networks using three machine learning algorithms (Random Forest, SVM-rfe, Lasso) to screen for crucial genes/proteins that regulating different aberrant immune cells in CMV+IBD patients. Then, Best Subset Selection regression model was used for final filter. Notably, oversampling is done to resolve the sample imbalance before the machine learning. Here, oversampling is performed by the ovun.sample function of the ROSE R package.

Random forest

Feature selection was developed using random forest (RF) approach. Random Forest is a collection of decision tree models created by randomly selecting features from a subset of the training data. The model was built using the randomForest R package with 1000

trees in the oversampled dataset, and internal validation was performed with 10-fold cross-validation.

SVM-rfe

Support Vector Machine Recursive Feature Extraction (SVM-rfe) is an iterative algorithm that starts with a set of features and gradually eliminates them. During each iteration, a basic linear SVM is initially applied, and the features are sorted according to their weights in the SVM solution. Subsequently, the feature with the smallest weight is removed. Here, we used the e1071 R package to perform the SVM-rfe andalysis and to filter the CMV+IBD prediction features.

Lasso

Lasso is a regularised form of linear regression that achieves model sparsity by introducing an L1-paradigm penalty term, which helps in feature selection. Lasso regression adds an L1-paradigm penalty term to the vector of coefficients, which is equal to λ times the sum of the absolute values of all the regression coefficients (λ is the penalty coefficient), on top of the least squares method. The purpose of this is to shrink some unimportant regression coefficients to zero, thus enabling feature selection. Here, the LASSO regression analyses were performed by the glmnet R package.

Best subset selection

Best Subset Selection can fit models for all possible combinations of predictor variables and then filter the best model conditional on the existing variables according to some criterion (e.g. R2, corrected R2, BIC, etc.). In Best Subset Selection, a larger value of adjusted R^2 and smaller value of BIC suggests a better model. Here, the best diagnosis model of CMV+IBD was constructed by Best Subset Selection using Leaps R package.

Function enrichment analysis

The study extracted all differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) for additional functional enrichment using the metascape webserver. Enrichment analysis was performed on the KEGG pathways and Hallmarks. Functions with a false discovery rate < 0.05 was selected.

Statistical analyses

Statistical analyses were performed using R software (version 4.2.2). A T-test was utilized to assess the variances between the two chosen groups. A p-value less than 0.05 was considered to be statistically significant.

Results

The immune microenvironment in CMV+IBD patients is different from that in other IBD patients

We used the xCell algorithm to assess the abundance of four levels of immune cells, specifically in the following comparisons: CMV+UC vs. CMV- UC, CMV+UC vs. CMV+N, CMV- UC vs. CMV- N, and CMV+N vs. CMV- N. In the CMV+UC vs. CMV- UC group, we found abnormal abundances of CD8+T cells, CD8+Tem, chondrocytes, HSC, keratinocytes, myocytes, neurons, neutrophils, platelets, and sebocytes.

Specifically, CD8+Tem, HSC, keratinocytes, myocytes, neurons, neutrophils, platelets, and sebocytes were significantly increased in CMV+UC patients, while CD8+T cells and chondrocytes were significantly decreased (Fig. 1A). In the CMV+UC vs. CMV+N group, we observed abnormal levels of adipocytes, CD4+Tem, GMP, neurons, neutrophils, platelets, and Tgd cells. Apart from CD4+Tem, which were significantly decreased in CMV+UC patients, all other abnormal cells showed increased abundance (Fig. 1B). We also evaluated the immune microenvironment in the IBD group without CMV infection (CMV- UC vs. CMV- N). The results indicated abnormal levels of astrocytes, CD8+T cells, ly endothelial cells, mesangial cells, mv endothelial cells, NKT, pDC, and Tregs. Notably, astrocytes and CD8+T cells were more abundant in UC patients, while the other immune cells showed decreased abundance (Fig. 1C). Additionally, in the non-UC group, the abundance of ly endothelial cells and Tregs decreased significantly post-CMV infection (Fig. 1D). These findings suggest that CMV infection significantly alters the immune microenvironment in UC patients. Therefore, we further investigated the impact of CMV infection on UC patients.

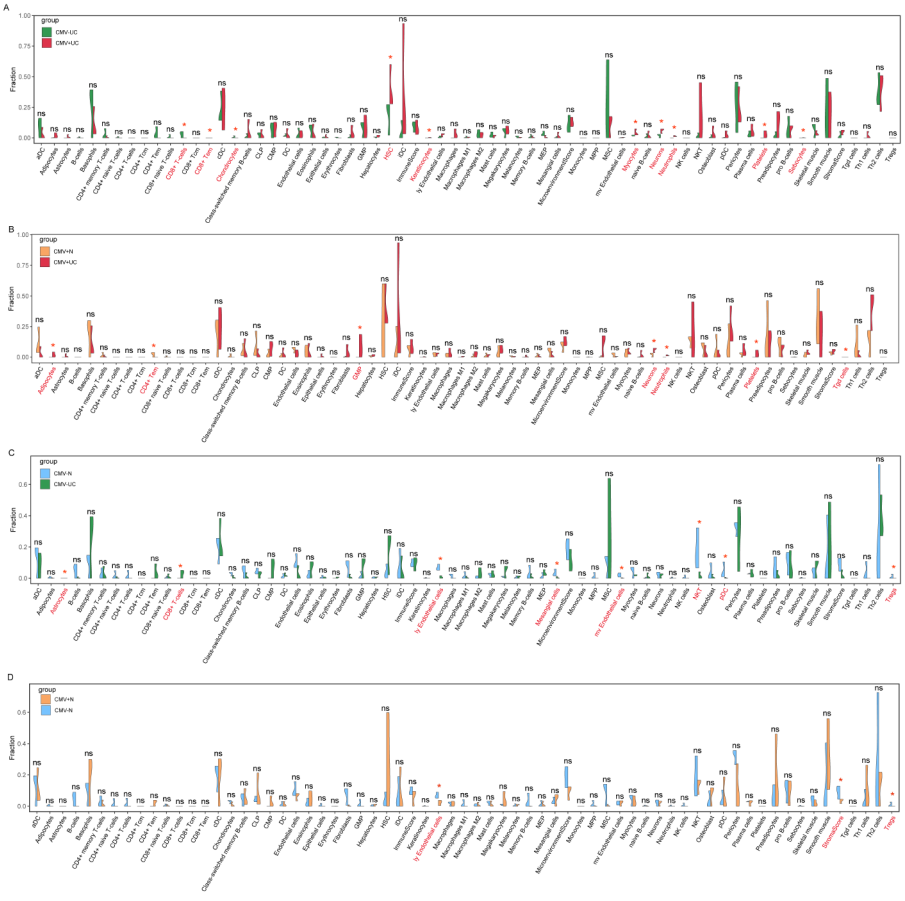


Fig. 1 Comprehensive assessment of the immune microenvironment in IBD patients with or without CMV virus infection. (**A**) We used the xCell algorithm to assess the abundance of CMV+UC vs. CMV- UC groups. (**B**) The immune cell abundance of CMV+UC vs. CMV+N group. (**C**) The immune cell abundance of CMV- UC vs. CMV- N. (**D**) The immune cell abundance of CMV+N vs. CMV- N

Identify CMV + UC-specific immune microenvironment-related molecular modules

We subsequently conducted differential analysis of transcription and protein levels across four groups: CMV + UC vs. CMV- UC, CMV + UC vs. CMV + N, CMV- UC vs. CMV- N, and CMV + N vs. CMV- N. In transcriptome (Fig. 2A), the comparison between CMV-infected and uninfected UC patients (CMV + UC vs. CMV- UC) showed 300 genes exhibited abnormal expression (|log2FC|>0.585, adjusted $P<0.05$), with 116 genes downregulated and 184 genes upregulated. Enrichment analysis of these differentially expressed genes revealed their involvement in KRAS signaling, the intestinal immune network for IgA production, inflammatory bowel disease, and the TGF-beta signaling pathway (Supplementary Fig. 1A). In the comparison between CMV-infected UC patients and CMV-infected non-UC individuals (CMV + UC vs. CMV + N), we identified 786 differentially expressed genes, with 492 upregulated and 294 downregulated. These genes were primarily enriched in the IL2/STAT5 signaling, p53 signaling pathway, NF-kappa B signaling pathway, B cell receptor signaling pathway, and ECM-receptor interaction pathways (Supplementary Fig. 1B). In the group of UC patients without

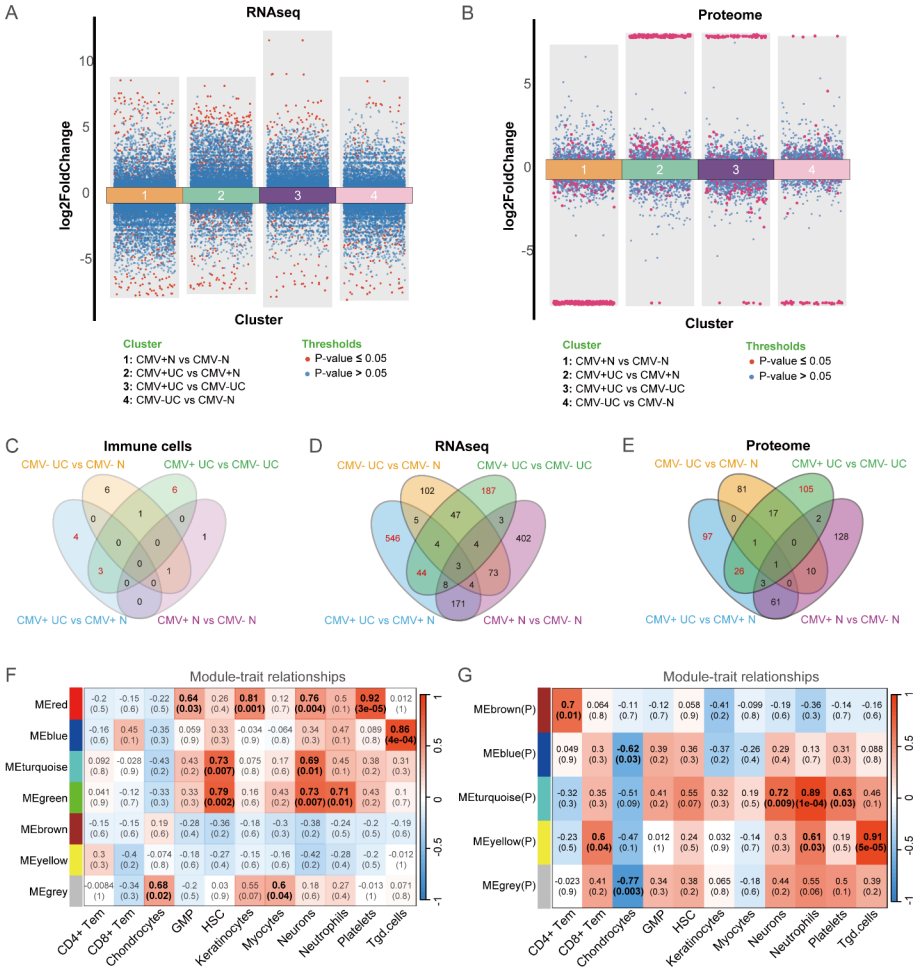


Fig. 2 Transcriptome- and proteome-based screening of CMV + UC associated co-expression modules. (**A**) Volcano plot of four groups (CMV + UC vs. CMV- UC, CMV + UC vs. CMV + N, CMV- UC vs. CMV- N, and CMV + N vs. CMV- N) comparison in transcriptome. (**B**) Volcano plot of these four groups comparison in proteome. (**C**) Venn plot showed the abnormal immune cells distribution in the four groups. (**D**) Venn plot showed the DEGs in the four groups. (**E**) Venn plot showed the DEPs in the four groups. (**F**) CMV + UC associated co-expression modules in transcriptome. (**G**) CMV + UC associated co-expression modules in proteome

CMV infection (CMV- UC vs. CMV- N), we identified 242 differentially expressed molecules, with 99 upregulated and 143 downregulated. These genes were mainly enriched in the IL2/STAT5 signaling, inflammatory response, KRAS signaling, and MTORC1 signaling pathways (Supplementary Fig. 1C). In the comparison of CMV-infected versus uninfected non-UC individuals (CMV+N vs. CMV- N), we identified 668 differentially expressed genes, with 367 upregulated and 301 downregulated. These genes were predominantly enriched in the IL-17 signaling pathway, p53 signaling pathway, IL6/JAK/STAT3 signaling pathways and so on (Supplementary Fig. 1D).

At the protein level, we conducted differential analysis across the same four groups (Fig. 2B). The results showed that in the CMV+UC vs. CMV- UC group, there were 155 differentially expressed proteins ($|\log2FC|>0.263$, adjusted $P<0.05$), which were mainly enriched in Protein processing in endoplasmic reticulum, MTORC1 Signaling and so on (Supplementary Fig. 2A). In the CMV+UC vs. CMV+N group, we identified 189 differentially expressed proteins, which were involved in various signaling pathways, including IL2 STAT5 Signaling, ECM−receptor interaction, and Endocytosis pathways (Supplementary Fig. 2B). In the CMV- UC vs. CMV- N group, 116 proteins were differentially expressed and enriched in pathways associated with Inflammatory bowel disease and Antigen processing and presentation (Supplementary Fig. 2C). Lastly, in the CMV+N vs. CMV- N group, 205 differentially expressed proteins were identified, enriched in pathways related to TCA cycle, Endocytosis and so on (Supplementary Fig. 2D).

Subsequently, we conducted an integrated analysis of abnormal immune cells, DEGs, and DEPs across the four groups to identify immune cells and biomarkers specific to CMV+UC patients. Our criteria focused on those abnormalities present exclusively in CMV+UC patients. Then, we identified 13 CMV+UC-specific cells, 11 of which are immune-related: CD4+Tem, CD8+Tem, chondrocytes, GMP, HSC, keratinocytes, myocytes, neurons, neutrophils, platelets, and Tgd cells (Fig. 2C). Additionally, we identified 777 DEGs (Fig. 2D) and 228 DEPs (Fig. 2E).

Next, we used WGCNA to identify co-expression modules of DEGs and DEPs specific to CMV+UC patients and their association with the immune cells. This allowed us to pinpoint biomarkers related to the immune microenvironment. The analysis revealed seven co-expression gene modules (MEred, MEblue, MEturquoise, MEgreen, MEbrown, MEyellow, and MEgrey) in CMV+UC patients. Specifically, MEred was associated with GMP, keratinocytes, neurons, and platelets; MEblue with Tgd cells; MEturquoise with HSC and neurons; MEgreen with HSC, neurons, and neutrophils; and MEgrey with chondrocytes and myocytes (Fig. 2F).

Similarly, five co-expression protein modules were identified (MEbrown(P), MEblue(P), MEturquoise(P), MEyellow(P), and MEgrey(P)). In which MEbrown(P) was associated with CD4+Tem; MEblue(P) with chondrocytes; MEturquoise(P) with neurons, neutrophils, and platelets; MEyellow(P) with CD8+Tem, neutrophils, and Tgd cells; and MEgrey(P) with chondrocytes (Fig. 2G).

Constructing the immune cells related molecular regulatory network

We further integrated gene co-expression modules and protein co-expression modules within the same abnormal immune cells, identifying potential regulatory relationships. Our analysis revealed that CD4+Tem cells are associated only with the MEbrown(P) module; CD8+Tem cells are associated only with the MEyellow(P) module;

chondrocytes are associated with the MEgrey, MEgrey(P), and MEblue(P) modules; GMP cells are associated only with the MEred module; HSC cells are associated with the MEgreen and MEturquoise modules; keratinocytes are associated only with the MEred module; myocytes are associated only with the MEgrey module; neurons are associated with the MEgreen, MEred, MEturquoise, and MEturquoise(P) modules; neutrophils are associated with the MEgreen, MEturquoise(P), and MEyellow(P) modules; platelets are associated with the MEred and MEturquoise(P) modules; and Tgd cells are associated with the MEblue and MEyellow(P) modules (Fig. 3A).

Upon further refinement, we identified four potential regulatory relationships: (I) MEgrey and MEgrey(P): related to chondrocytes (Fig. 3B); (II) MEgreen, MEblue, and MEyellow(P): related to neutrophils and Tgd cells (Fig. 3C); (III) MEgreen, MEred, MEturquoise, and MEturquoise(P): related to neurons, neutrophils, and platelets (Fig. 3D); (IV) MEgrey and MEblue(P): related to chondrocytes (Fig. 3E). Subsequently, we constructed co-expression regulatory networks between the gene co-expression modules and protein co-expression modules for different regulatory relationships. Only regulatory relationships with an absolute correlation coefficient greater than 0.5 and a p-value less than 0.05 were considered significant. These regulatory networks potentially reflect that the abnormal immune microenvironment is influenced by multiple factors (Fig. 3B-E).

Identifying potential diagnostic biomarkers for CMV + UC using machine learning algorithms

To identify early diagnostic biomarkers for CMV+UC, we employed three different machine learning algorithms across five levels: the four regulatory networks mentioned above and all modules associated with abnormal immune cells.

In the MEgrey and MEgrey(P) regulatory network, the Random Forest algorithm identified a set of 54 genes/proteins with the lowest classification error rate (Fig. 4A). The SVM-RFE algorithm selected 79 genes/proteins with the best classification accuracy (Fig. 4B), and the LASSO algorithm identified 9 highly efficient genes/proteins (Fig. 4C).

In the MEgreen, MEblue, and MEyellow(P) regulatory network, the Random Forest algorithm selected a set of 14 genes/proteins with optimal classification performance (Fig. 4D). The SVM-RFE algorithm identified 16 genes/proteins (Fig. 4E), and the LASSO algorithm found 6 highly efficient genes/proteins (Fig. 4F).

In the MEgreen, MEred, MEturquoise, and MEturquoise(P) regulatory network, the Random Forest algorithm identified 24 genes/proteins with classification efficacy (Fig. 4G). The SVM-RFE algorithm selected 30 genes/proteins (Fig. 4H), while the LASSO algorithm identified 7 genes/proteins (Fig. 4I).

In the MEgrey and MEblue(P) regulatory network, the Random Forest algorithm identified 18 genes/proteins (Fig. 4J), the SVM-RFE algorithm selected 22 genes/proteins (Fig. 4K), and the LASSO algorithm identified 10 genes/proteins (Fig. 4L).

Finally, considering all modules, the Random Forest algorithm identified 25 genes/proteins with classification efficacy (Fig. 4M), the SVM-RFE algorithm selected 132 relevant genes/proteins (Fig. 4N), and the LASSO algorithm identified 7 genes/proteins (Fig. 4O).

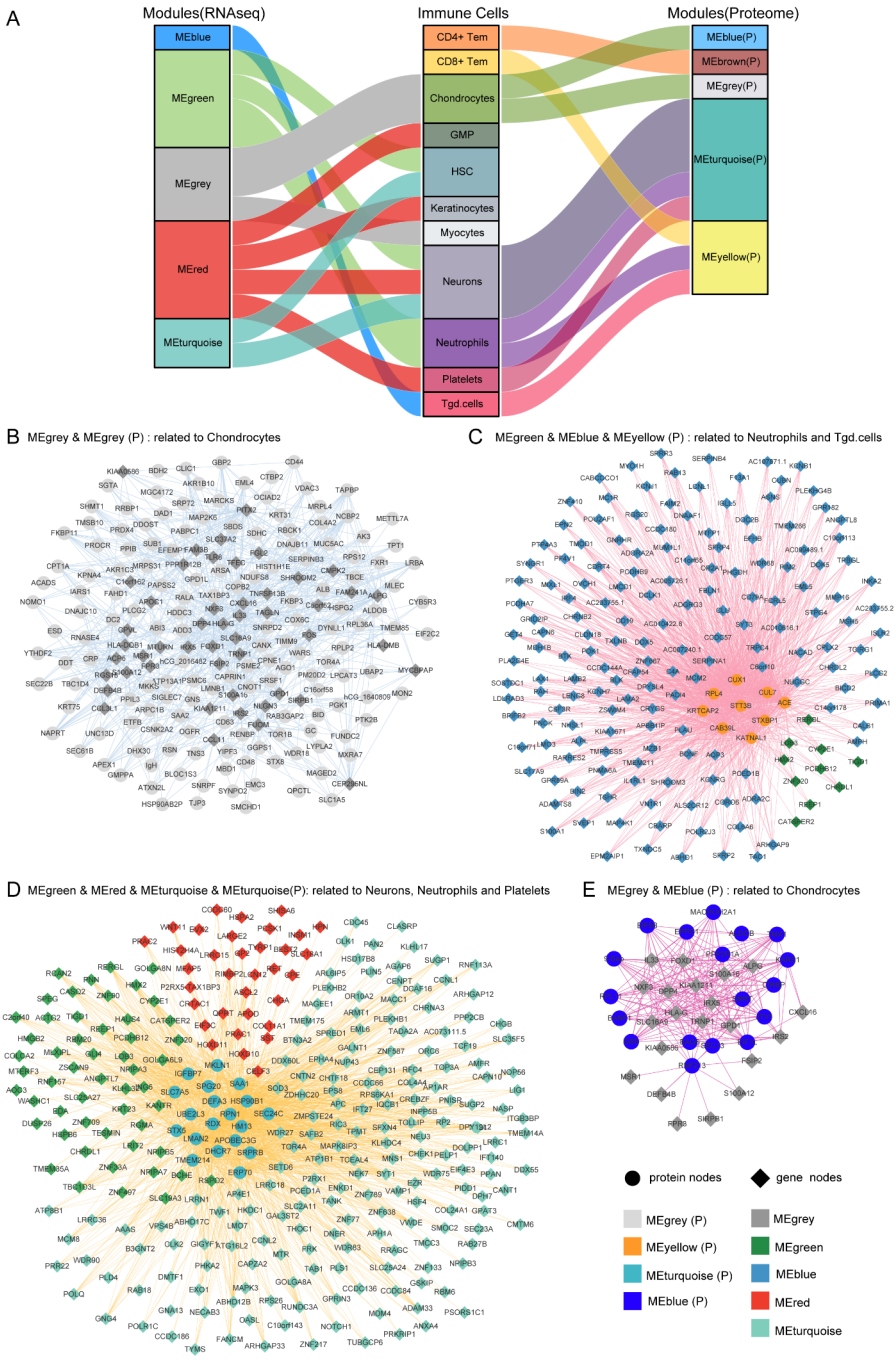


Fig. 3 Integrating the transcriptome and proteome to construct relevant regulatory networks. (**A**) Sankey plot showing association between co-expressed modules and aberrant immune cells. (**B**) Co-expression network of MEgrey and MEgrey (P) module. (**C**) Co-expression network of MEgreen, MEblue, and MEyellow (P) module. (**D**) Co-expression network of MEgreen, MEred, MEturquoise, and MEturquoise(P) module. (**E**) Co-expression network of MEgrey and MEblue (P) module

Constructing a diagnostic model for CMV + UC

We further integrated the results from three machine learning algorithms. At the MEgrey and MEgrey(P) levels, we identified 8 molecules that consistently exhibited effective classification across all three algorithms. In the MEgreen, MEblue, and MEyellow(P) group, we identified 3 molecules with stable classification performance.

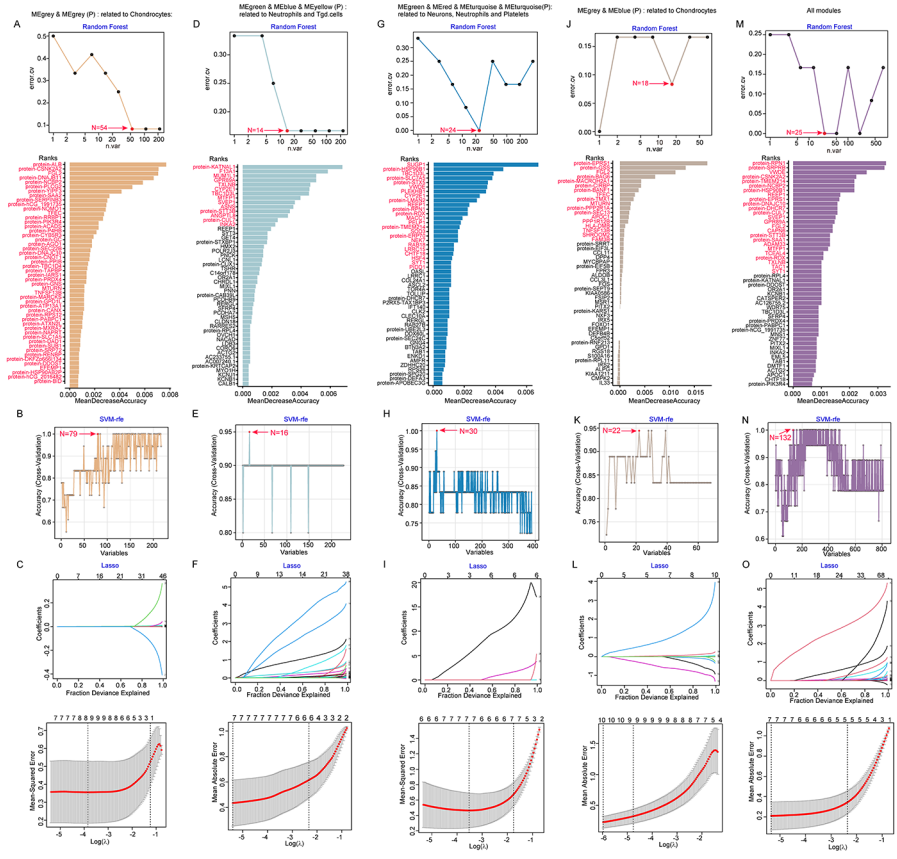


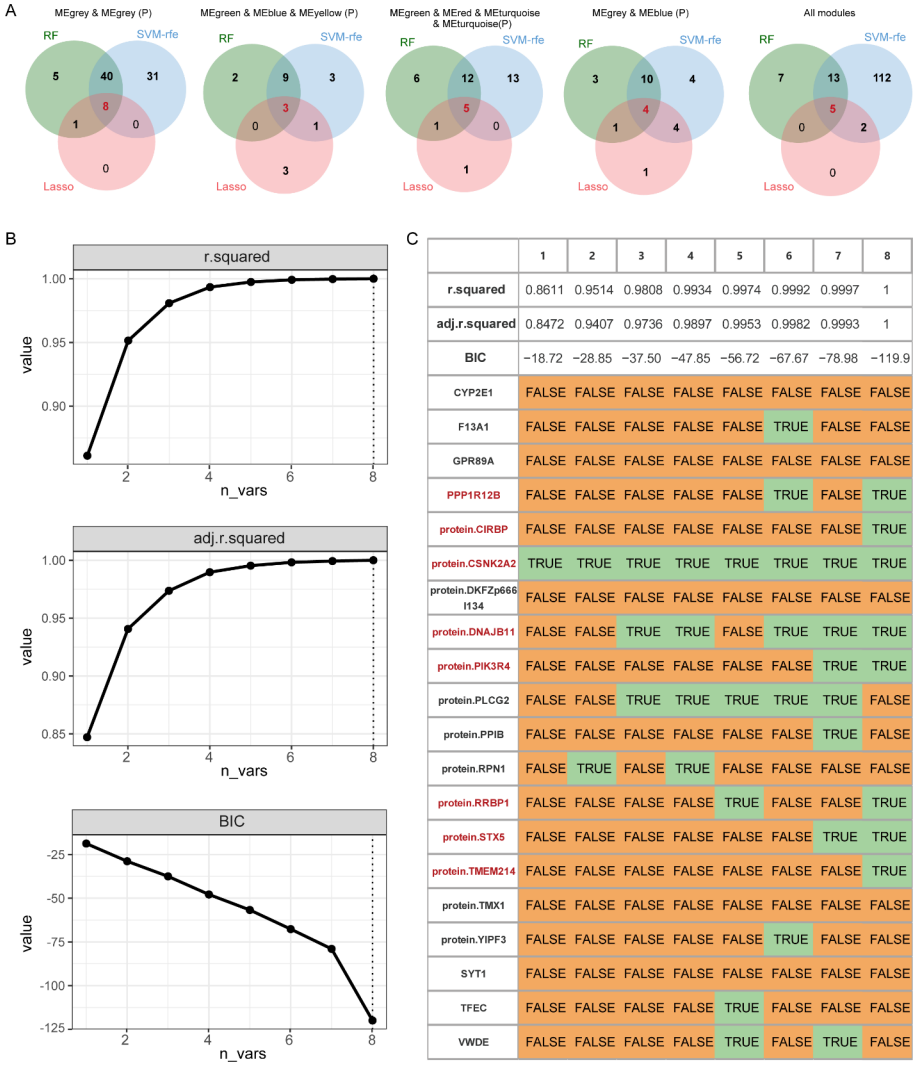
Fig. 4 Screen candidate diagnostic biomarkers by machine learning algorithms. (**A**) Random Forest analysis in MEgrey and MEblue (P) groups. (**B**) SVM-rfe analysis in MEgrey and MEblue (P) groups. (**C**) Lasso analysis in MEgrey and MEblue (P) groups. (**D**) Random Forest analysis in MEgreen, MEblue, and MEyellow (P) groups. (**E**) SVM-rfe analysis in MEgreen, MEblue, and MEyellow (P) groups. (**F**) Lasso analysis in MEgreen, MEblue, and MEyellow (P) groups. (**G**) Random Forest analysis in MEgreen, MEred, MEturquoise, and MEturquoise(P) groups. (**H**) SVM-rfe analysis in MEgreen, MEred, MEturquoise, and MEturquoise(P) groups. (**I**) Lasso analysis in MEgreen, MEred, MEturquoise, and MEturquoise(P) groups. (**J**) Random Forest analysis in MEgrey and MEblue (P) groups. (**K**) SVM-rfe analysis in MEgrey and MEblue (P) groups. (**L**) Lasso analysis in MEgrey and MEblue (P) groups. (**M**) Random Forest analysis in all modules. (**N**) SVM-rfe analysis in all modules. (**O**) Lasso analysis in all modules

Similarly, at the MEgreen, MEred, MEturquoise, and MEturquoise(P) levels, we found 5 such molecules. Within the MEgrey and MEblue(P) group, we selected 4 molecules showing stable classification across the algorithms. Finally, across all modules, we identified 55 molecules that consistently demonstrated effective classification across the three algorithms (Fig. 5A).

Subsequently, we integrated these molecules, resulting in 20 candidate biomarkers that displayed stable classification performance across different levels. Using optimal subset regression, we constructed a diagnostic model for CMV+UC. The analysis revealed that an 8 biomarkers model achieved the maximum R^2 and adjusted R^2 values, with the lowest Bayesian Information Criterion (BIC) (Fig. 5B). These findings indicate that the 8 biomarkers model (Table 1) performs optimally in diagnosing CMV+UC (Fig. 5C).

Discussion

So far, diagnosing and treating CMV+UC in IBD patients remains a challenge. The diagnosis and treatment of IBD are often either underutilized or overused, potentially leading to adverse disease progression and increased cost [33]. Studies in IBD patients



	1	2	3	4	5	6	7	8
r.squared	0.8611	0.9514	0.9808	0.9934	0.9974	0.9992	0.9997	1
adj.r.squared	0.8472	0.9407	0.9736	0.9897	0.9953	0.9982	0.9993	1
BIC	−18.72	−28.85	−37.50	−47.85	−56.72	−67.67	−78.98	−119.9
CYP2E1	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
F13A1	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
GPR89A	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
PPP1R12B	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE
protein.CIRBP	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
protein.CSNK2A2	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
protein.DKFZp666I134	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
protein.DNAJB11	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
protein.PIK3R4	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
protein.PLCG2	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE
protein.PPIB	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
protein.RPN1	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
protein.RRBP1	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE
protein.STX5	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
protein.TMEM214	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
protein.TMX1	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
protein.YIPF3	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
SYT1	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
TFEC	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
VWDE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE

Fig. 5 Constructing the CMV + UC Diagnostic Model. (**A**) Venn plot showed the stable diagnostic biomarkers in both three machine learning algorithms. (**B**) Parameters of Best Subset Selection regression model, including R^2, adjusted R^2 and BIC. (**C**) Final diagnostic biomarkers selected by Best Subset Selection

Table 1 Information of 8 crucial diagnostic biomarkers

ID	Full name	Type	Modules	Associated Immune cells
PPP1R12B	Protein phosphatase 1 regulatory subunit 12B	Gene	MEgrey	Chondrocytes
CIRBP	Cold inducible RNA binding protein	Protein	MEblue (P)	Chondrocytes
CSNK2A2	Casein kinase 2 alpha 2	Protein	MEgrey (P)	Chondrocytes
DNAJB11	DnaJ heat shock protein family member B11	Protein	MEgrey (P)	Chondrocytes
PIK3R4	Phosphoinositide-3-kinase regulatory subunit 4	Protein	MEgrey (P)	Chondrocytes
RRBP1	Ribosome binding protein 1	Protein	MEgrey (P)	Chondrocytes
STX5	Syntaxin 5	Protein	MEturquoise(P)	Neurons, Neutrophils and Platelets
TMEM214	Transmembrane protein 214	Protein	MEturquoise(P)	Neurons, Neutrophils and Platelets

have shown a correlation between CMV infection and refractory disease [34]. Research indicates a connection between the presence of CMV and refractory Crohn's colitis and ulcerative colitis [35]. The recently updated ECCO guidelines (2021) for opportunistic infections recommend baseline CMV screening for all IBD patients, especially before initiating immunosuppressive therapy [36]. Early detection of CMV infection is crucial in IBD patients. It is recommended that symptomatic IBD patients with more than four positive cells per biopsy or plasma CMV DNA levels ≥ 1000 IU/mL, and who have any of the following conditions—steroid-refractory disease, splenomegaly, or absence of leukocytosis—should receive antiviral treatment [35].

Given that colonic mucosal viral detection is invasive and serum CMV PCR is not a reliable indicator of end-organ disease, with lower sensitivity in peripheral blood testing for clinical viral monitoring [37], there is a need for high-sensitivity, non-invasive diagnostic methods. Therefore, in this research, our goal was to develop a new diagnostic approach for CMV + UC using multi-omics integration analysis of transcriptomics and proteomics, combined with machine learning algorithms. We identified an 8 biomarkers diagnostic model (PPP1R12B, CIRBP, CSNK2A2, DNAJB11, PIK3R4, RRBP1, STX5, TMEM214) to provide useful insights into CMV + UC diagnosis. Here, PPP1R12B has been found to play a significant role in intestinal diseases, particularly colorectal cancer [38]. Upregulation of CIRBP has been shown to promote Th1 cell-mediated mucosal inflammation in IBD [39]. CSNK2A2 is associated with oxaliplatin resistance in colorectal cancer cells [40]. DNAJB11 encodes a soluble glycoprotein in the ER lumen that regulates immunoglobulin binding activity by stimulating ATPase activity [41], and our study found it to be crucial in IBD. PIK3R4 is involved in intestinal autophagy and is related to aggravated intestinal inflammation [42]. RRBP1 plays a role in ER proliferation, the secretory pathway, and ER-microtubule interactions, with relevance to IBD found in this study. Some viral transductions depend on STX5 for retrograde transport to the trans-Golgi network [43]. TMEM214 mediates ER stress-induced Caspase 4 activation and apoptosis [44], also playing a significant role in IBD as identified in our study.

Notably, the small sample size is a limitation of this study. Here, to avoid the influence of this, we have used highly sensitive and specific methods for detailed and accurate observation and all machine learning analyses use oversampling methods to balance the data distribution and improve the predictive performance of the model. Certainly, further studies with larger patient groups are necessary to confirm the promising results.

Conclusion

In summary, we constructed a high-throughput sequencing profile of CMV + IBD and our data show that CMV + UC patients have a distinct immune microenvironment compared to other patients, with these abnormal microenvironments being associated with multiple molecular dimensions. These molecules have the potential to become diagnostic and therapeutic markers for CMV + UC patients. While real-time quantitative PCR to detect viral load in fresh mucosal specimens is a feasible method for monitoring CMV infection in IBD patients, our 8-molecule model offers new avenues for exploring other non-invasive diagnostic methods.

Abbreviations

IBD	Inflammatory bowel disease
UC	Ulcerative colitis
CD	Crohn's disease

EBV	Epstein-Barr virus
CMV	Cytomegalovirus
IHC	Immunohistochemistry
AGA	American Gastroenterological Association
ESPGHAN	European Society for Paediatric Gastroenterology Hepatology and Nutrition
CMV + UC	Ulcerative colitis patients with CMV infection
CMV- UC	Ulcerative colitis patients without CMV infection
CMV + N	Normal control with CMV infection
CMV-N	Normal control without CMV infection
DEGs	Differentially expressed genes
DEPs	Differentially expressed proteins
WGCNA	Weighted correlation network analysis
RF	Random forest
SVM-rfe	Support Vector Machine Recursive Feature Extraction

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13040-024-00382-0.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
Y.C., J.H.H., J.L., L.W. and J.P. conceived the study and designed the experiments. Y.C. conducted the majority of the work and wrote the paper. Q.Q.Z and H.W. pre-processing of data. P.R.T, L.D, and P.L. collected samples. J.L., H.L. and Y.C. performed bioinformatic analysis. J.H.H., L.W. and J.P. supervised and coordinated the study and reviewed the final version of the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by Open Foundation of Yunnan Provincial Laboratory of Clinical Virology (2023A4010403-02), and supported in part of the Health Commission Foundation of Yunnan Province (2023-KHRCBZ-B15), Major Science and Technology Projects of Yunnan Province (202402AA310016), Basic Research Science and Technology Foundation of Yunnan Province (202201AS070009) and Xing Dian Foundation of Yunnan Province (XDYC-MY-2022-0029).

Data availability
The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Ethical approval for this study was obtained from the Ethics Committee of the First People's Hospital of Yunnan Province (KHLL2024-KY199).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Yunnan Provincial Laboratory of Clinical Virology, The First People's Hospital of Yunnan Province, Kunming, Yunnan 650032, China
[2]Department of Pathology, The First People's Hospital of Yunnan Province, Kunming, Yunnan 650032, China
[3]Department of Pathology, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan 650032, China
[4]Department of General Practice, The First People's Hospital of Yunnan Province, Kunming, Yunnan 650032, China
[5]Academy of Biomedical Engineering, Kunming Medical University, Kunming, Yunnan 650500, China
[6]Department of Surgery, The First People's Hospital of Yunnan Province, Kunming, Yunnan 650032, China

Received: 15 July 2024 / Accepted: 22 August 2024
Published online: 27 August 2024

References

1. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nat Rev Gastroenterol Hepatol. 2015;12(12):720–7.
2. Hodges P, Kelly P. Inflammatory bowel disease in Africa: what is the current state of knowledge? Int Health. 2020;12(3):222–30.
3. Windsor JW, Kaplan GG. Evolving epidemiology of IBD. Curr Gastroenterol Rep. 2019;21(8):40.

4. Mohsenizadeh SM, Manzari ZS, Vosoghinia H, Ebrahimipour H. Family caregivers' burden in inflammatory bowel diseases: an integrative review. J Educ Health Promot. 2020;9:289.
5. Chang JT. Pathophysiology of Inflammatory Bowel diseases. N Engl J Med. 2020;383(27):2652–64.
6. Lopetuso LR, Ianiro G, Scaldaferri F, Cammarota G, Gasbarrini A. Gut virome and inflammatory bowel disease. Inflamm Bowel Dis. 2016;22(7):1708–12.
7. Ungaro F, Massimino L, D'Alessio S, Danese S. The gut virome in inflammatory bowel disease pathogenesis: from metagenomics to novel therapeutic approaches. United Eur Gastroenterol J. 2019;7(8):999–1007.
8. Wang W, Chen X, Pan J, Zhang X, Zhang L. Epstein-Barr Virus and human cytomegalovirus infection in intestinal mucosa of Chinese patients with inflammatory bowel disease. Front Microbiol. 2022;13:915453.
9. Kandiel A, Lashner B. Cytomegalovirus colitis complicating inflammatory bowel disease. Am J Gastroenterol. 2006;101(12):2857–65.
10. Kim JJ, Simpson N, Klipfel N, Debose R, Barr N, Laine L. Cytomegalovirus infection in patients with active inflammatory bowel disease. Dig Dis Sci. 2010;55(4):1059–65.
11. Li X, Chen N, You P, Peng T, Chen G, Wang J, Li J, Liu Y. The Status of Epstein-Barr Virus infection in intestinal mucosa of Chinese patients with inflammatory bowel disease. Digestion. 2019;99(2):126–32.
12. Ryan JL, Shen YJ, Morgan DR, Thorne LB, Kenney SC, Dominguez RL, Gulley ML. Epstein-Barr virus infection is common in inflamed gastrointestinal mucosa. Dig Dis Sci. 2012;57(7):1887–98.
13. Krech U. Complement-fixing antibodies against cytomegalovirus in different parts of the world. Bull World Health Organ. 1973;49(1):103–6.
14. Teo WH, Chen HP, Huang JC, Chan YJ. Human cytomegalovirus infection enhances cell proliferation, migration and upregulation of EMT markers in colorectal cancer-derived stem cell-like cells. Int J Oncol. 2017;51(5):1415–26.
15. Fornara O, Bartek J Jr., Rahbar A, Odeberg J, Khan Z, Peredo I, Hamerlik P, Bartek J, Stragliotto G, Landazuri N, et al. Cytomegalovirus infection induces a stem cell phenotype in human primary glioblastoma cells: prognostic significance and biological impact. Cell Death Differ. 2016;23(2):261–9.
16. Sissons JG, Carmichael AJ. Clinical aspects and management of cytomegalovirus infection. J Infect. 2002;44(2):78–83.
17. Shieh AC, Guler E, Tirumani SH, Dumot J, Ramaiya NH. Clinical, imaging, endoscopic findings, and management of patients with CMV colitis: a single-institute experience. Emerg Radiol. 2020;27(3):277–84.
18. Kwon J, Fluxa D, Farraye FA, Kroner PT. Cytomegalovirus-related colitis in patients with inflammatory bowel disease. Int J Colorectal Dis. 2022;37(3):685–91.
19. Garrido E, Carrera E, Manzano R, Lopez-Sanroman A. Clinical significance of cytomegalovirus infection in patients with inflammatory bowel disease. World J Gastroenterol. 2013;19(1):17–25.
20. Streetz MD, Buhr T, Wedemeyer H, Bleck J, Schedel I, Manns MP, Goke MN. Acute CMV-Colitis in a patient with a history of Ulcerative Colitis. Scand J Gastroenterol. 2003;38(1):119–22.
21. de Saussure P, Lavergne-Slove A, Mazeron MC, Alain S, Matuchansky C, Bouhnik Y. A prospective assessment of cytomegalovirus infection in active inflammatory bowel disease. Aliment Pharmacol Ther. 2004;20(11–12):1323–7.
22. Kambham N, Vij R, Cartwright CA, Longacre T. Cytomegalovirus infection in steroid-refractory ulcerative colitis: a case-control study. Am J Surg Pathol. 2004;28(3):365–73.
23. Roblin X, Pillet S, Oussalah A, Berthelot P, Del Tedesco E, Phelip JM, Chambonniere ML, Garraud O, Peyrin-Biroulet L, Pozzetto B. Cytomegalovirus load in inflamed intestinal tissue is predictive of resistance to immunosuppressive therapy in ulcerative colitis. Am J Gastroenterol. 2011;106(11):2001–8.
24. Buck Q, Cho S, Mehta Walsh S, Schady D, Kellermayer R. Routine histology-based diagnosis of CMV colitis was Rare in Pediatric patients. J Pediatr Gastroenterol Nutr. 2022;75(4):462–5.
25. Hirayama Y, Ando T, Hirooka Y, Watanabe O, Miyahara R, Nakamura M, Yamamura T, Goto H. Characteristic endoscopic findings and risk factors for cytomegalovirus-associated colitis in patients with active ulcerative colitis. World J Gastrointest Endosc. 2016;8(6):301–9.
26. McCurdy JD, Jones A, Enders FT, Killian JM, Loftus EV Jr., Smyrk TC, Bruining DH. A model for identifying cytomegalovirus in patients with inflammatory bowel disease. Clin Gastroenterol Hepatol. 2015;13(1):131–7. quiz e137.
27. Shukla T, Singh S, Tandon P, McCurdy JD. Corticosteroids and thiopurines, but not tumor necrosis factor antagonists, are Associated with Cytomegalovirus Reactivation in Inflammatory Bowel Disease: a systematic review and Meta-analysis. J Clin Gastroenterol. 2017;51(5):394–401.
28. Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG Clinical Guideline: Ulcerative Colitis in adults. Am J Gastroenterol. 2019;114(3):384–413.
29. Turner D, Travis SP, Griffiths AM, Ruemmele FM, Levine A, Benchimol EI, Dubinsky M, Alex G, Baldassano RN, Langer JC, et al. Consensus for managing acute severe ulcerative colitis in children: a systematic review and joint statement from ECCO, ESPGHAN, and the Porto IBD Working Group of ESPGHAN. Am J Gastroenterol. 2011;106(4):574–88.
30. Hradsky O, Copova I, Zarubova K, Durilova M, Nevoral J, Maminak M, Hubacek P, Bronsky J. Seroprevalence of Epstein-Barr Virus, Cytomegalovirus, and Polyomaviruses in Children with Inflammatory Bowel Disease. Dig Dis Sci. 2015;60(11):3399–407.
31. Thavamani A, Umapathi KK, Sferra TJ, Sankararaman S. Cytomegalovirus infection is Associated with adverse outcomes among hospitalized Pediatric patients with inflammatory bowel disease. Gastroenterol Res. 2023;16(1):1–8.
32. Chen Y, He LX, Chen JL, Xu X, Wang JJ, Zhan XH, Jiao JW, Dong G, Li EM, Xu LY. L2Delta13, a splicing isoform of lysyl oxidase-like 2, causes adipose tissue loss via the gut microbiota and lipid metabolism. iScience. 2022;25(9):104894.
33. Hendler SA, Barber GE, Okafor PN, Chang MS, Limsui D, Limketkai BN. Cytomegalovirus infection is associated with worse outcomes in inflammatory bowel disease hospitalizations nationwide. Int J Colorectal Dis. 2020;35(5):897–903.
34. Khan TV, Toms C. Cytomegalovirus Colitis and subsequent new diagnosis of inflammatory bowel disease in an Immunocompetent host: a Case Study and Literature Review. Am J Case Rep. 2016;17:538–43.
35. Temtem T, Whitworth J, Zhang J, Bagga B. Cytomegalovirus in pediatric inflammatory bowel disease patients with acute severe colitis. Clin Res Hepatol Gastroenterol. 2021;45(6):101625.
36. Kucharzik T, Ellul P, Greuter T, Rahier JF, Verstockt B, Abreu C, Albuquerque A, Allocca M, Esteve M, Farraye FA, et al. ECCO Guidelines on the Prevention, diagnosis, and management of infections in inflammatory bowel disease. J Crohns Colitis. 2021;15(6):879–913.

37. Kim JW, Boo SJ, Ye BD, Kim CL, Yang SK, Kim J, Kim SA, Park SH, Park SK, Yang DH, et al. Clinical utility of cytomegalovirus antigenemia assay and blood cytomegalovirus DNA PCR for cytomegaloviral colitis patients with moderate to severe ulcerative colitis. J Crohns Colitis. 2014;8(7):693–701.
38. Tan J, Lu T, Xu J, Hou Y, Chen Z, Zhou K, Ding Y, Jiang B, Zhu Y. MicroRNA-4463 facilitates the development of colon cancer by suppression of the expression of PPP1R12B. Clin Transl Oncol. 2022;24(6):1115–23.
39. He Q, Gao H, Chang YL, Wu X, Lin R, Li G, Lin J, Lu H, Chen H, Li Z, et al. ETS-1 facilitates Th1 cell-mediated mucosal inflammation in inflammatory bowel diseases through upregulating CIRBP. J Autoimmun. 2022;132:102872.
40. Yun CW, Lee JH, Lee SH. Casein kinase 2alpha augments Oxaliplatin Resistance in Colorectal Cancer cells by increasing ABCE1 expression. Anticancer Res. 2023;43(6):2519–25.
41. Otero JH, Lizak B, Feige MJ, Hendershot LM. Dissection of structural and functional requirements that underlie the interaction of ERdj3 protein with substrates in the endoplasmic reticulum. J Biol Chem. 2014;289(40):27504–12.
42. Wang SL, Shao BZ, Zhao SB, Chang X, Wang P, Miao CY, Li ZS, Bai Y. Intestinal autophagy links psychosocial stress with gut microbiota to promote inflammatory bowel disease. Cell Death Dis. 2019;10(6):391.
43. Nonnenmacher ME, Cintrat JC, Gillet D, Weber T. Syntaxin 5-dependent retrograde transport to the trans-golgi network is required for adeno-associated virus transduction. J Virol. 2015;89(3):1673–87.
44. Li C, Wei J, Li Y, He X, Zhou Q, Yan J, Zhang J, Liu Y, Liu Y, Shu HB. Transmembrane protein 214 (TMEM214) mediates endoplasmic reticulum stress-induced caspase 4 enzyme activation and apoptosis. J Biol Chem. 2013;288(24):17908–17.

Publisher's note