

SOFTWARE

Open Access



Immune status assessment based on plasma proteomics with meta graph convolutional networks

Min Zhang¹, Nan Xu², Qi Cheng¹, Jing Ye², Shiwei Wu¹, Haoliang Liu¹, Chengkui Zhao^{1,3*}, Lei Yu^{2,3*} and Weixing Feng^{1*}

Abstract

Plasma proteins, especially immune-related proteins, are vital for assessing immune health and predicting disease risks. Despite their significance, the link between these proteins and systemic immune function remains unclear. To bridge this gap, researchers developed ProMetaGCN, a model integrating meta-learning, graph convolutional networks, and protein-protein interaction (PPI) data to evaluate immune status via plasma proteomics. This framework identified 309 immune-related factors with associated biological functions and pathways. Using six machine learning methods, four algorithms (Random Forest, LightGBM, XGBoost, Lasso) were selected for immune profiling and aging analysis, revealing ADAMTS13, GDF15, and SERPINF2 as key biomarkers. Validation across two COVID-19 cohorts confirmed the model's robustness, showing immune status correlates with infection progression and recovery. Furthermore, the study proposed ImmuneAgeGap, a novel metric linking immune profiles to survival rates in non-small-cell lung cancer (NSCLC) patients. These insights advance personalized immune health strategies and disease prevention.

Keywords Meta-learning graph convolutional network, Plasma proteomics, Immune-related proteins, Machine learning, Immune status score

Introduction

The increasing emphasis on health has highlighted the importance of assessing immune status. A key aspect of this is immunosenescence, the gradual decline in immune function with aging, which is closely linked to age-related diseases such as cancer, cardiovascular diseases, and neurodegenerative disorders [1, 2]. Immunosenescence is characterized by weakened immune responses to infections and vaccines, as well as increased chronic inflammation [3–6]. Changes in plasma proteins, particularly immunoglobulins, cytokines, and complement proteins, reflect this decline and are integral to immune and inflammatory responses. As these proteins undergo significant alterations with aging, they can serve as indicators of immune system health [7, 8]. Therefore,

*Correspondence:

Chengkui Zhao
zhaochengkui@hrbeu.edu.cn

Lei Yu
yulei@nbic.ecnu.edu.cn

Weixing Feng
fengweixing@hrbeu.edu.cn

¹College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, China

²Institute of Biomedical Engineering and Technology, Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, China

³Shanghai Unicar-Therapy Bio-medicine Technology Co., Ltd., Shanghai, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

identifying reliable immune-related plasma protein biomarkers and developing effective immune assessment models based on these biomarkers are crucial for managing immune health and preventing diseases [9].

Currently, many studies focus on using plasma proteins to explore the relationship between chronological age and biological age, which reflects individual health and physiological function [10–13]. However, existing research methods have limitations. Typically, correlation tests identify age-related proteins as predictive model features, using chronological age as a label to predict age. However, this approach cannot comprehensively reflect the immune status. As people age, immune function declines nonlinearly, complicating age prediction models and making it harder to interpret individual immune status differences. Meanwhile, some studies use other data types to detect immune status. For example, complete blood counts (CBC) measure routine immune cell counts and proportions but can't classify immune cell subtypes, failing to capture the broader molecular interactions or the complex network of immune responses [14]. Similarly, another study's flow cytometry can analyze T-cell and NK-cell subtypes in more detail but focuses only on these two cell types, ignoring other immune cells' roles in immunosenescence and failing to show the whole immune system changes. Also, the random forest binary classification model simplifies immunosenescence into a binary classification, making it hard to accurately reflect the continuous immunosenescence process [15]. Meanwhile, using serum protein data to assess immune health has limitations. Studies rely on random forest models to distinguish healthy participants from patients but cover limited diseases, failing to include all diseases and find as many immune-related proteins as possible [7]. Although progress has been made in identifying the roles of certain plasma proteins (like immunoglobulins, cytokines, and complement proteins), many other potential immune-related protein functions remain unexplored [16–20]. Therefore, there's an urgent need for a research method that can comprehensively capture the complexity of the immune system, so as to achieve a comprehensive and reliable assessment of the immune status.

To address this challenge, we introduce the ProMetaGCN model, aiming to accurately identify immune-related proteins and comprehensively assess immune health. We selected GCN to study potential immune-related proteins because it is designed for graph topology data, where proteins are nodes and interactions are edges, efficiently capturing complex relationships and topological information in protein-protein interaction (PPI) networks. Unlike traditional machine learning, GCN more accurately models interactions between biomolecules, crucial for understanding immune-related protein functions and synergies. These advantages enable

GCN to identify potential immune-related proteins with higher precision and efficiency, thus helping us discover more reliable immune-related proteins. To enhance the model, we combined GCN with meta-learning, significantly improving learning and generalization with limited labeled data, and systematically identifying key plasma proteins linked to immune status. Additionally, ProMetaGCN integrates advanced algorithms like Random Forest, LightGBM, XGBoost, and Lasso Regression, boosting prediction and robustness for precise reflection of immune aging's continuous changes, avoiding information loss and inaccuracy from simplistic binary classification. In practice, ProMetaGCN performed well. Validated externally on samples from healthy donors and COVID-19 patients, it can monitor immune responses in systemic infectious diseases. We also introduced a new metric, ImmuneAgeGap, quantifying the difference between actual and immune ages, valuable for immune aging and health intervention research. Notably, ProMetaGCN's application in NSCLC patients revealed a significant correlation between immune aging and survival rates (and death risks), offering new insights and biomarkers for cancer prognosis and personalized immunotherapy, potentially advancing precision medicine in oncology. In conclusion, combining GCN, meta-learning, and diverse machine learning algorithms, we aim to provide an efficient, accurate, and innovative tool for immune-related protein research and immune state assessment, offering new ideas for disease diagnosis, treatment, and prevention.

Methods

The architecture of ProMetaGCN

ProMetaGCN is an interpretable and robust tool for assessing human immune status by utilizing plasma protein data collected from healthy individuals. The workflow comprises two principal steps: the prediction of immune-related proteins and the assessment of immune status, as illustrated in Fig. 1. In Step 1, ProMetaGCN employs a semi-supervised meta-learning graph convolutional network (Meta-GCN) model, enhanced by literature mining, to predict potential immune-related proteins. These predicted proteins serve as features and are subsequently input into various machine learning models. In Step 2, the proteins are utilized to evaluate immune status. Additionally, the system conducts a series of downstream analyses and is validated using external datasets to ensure the accuracy and reliability of the results.

Datasets

Details about the datasets employed in this study are detailed as follows.

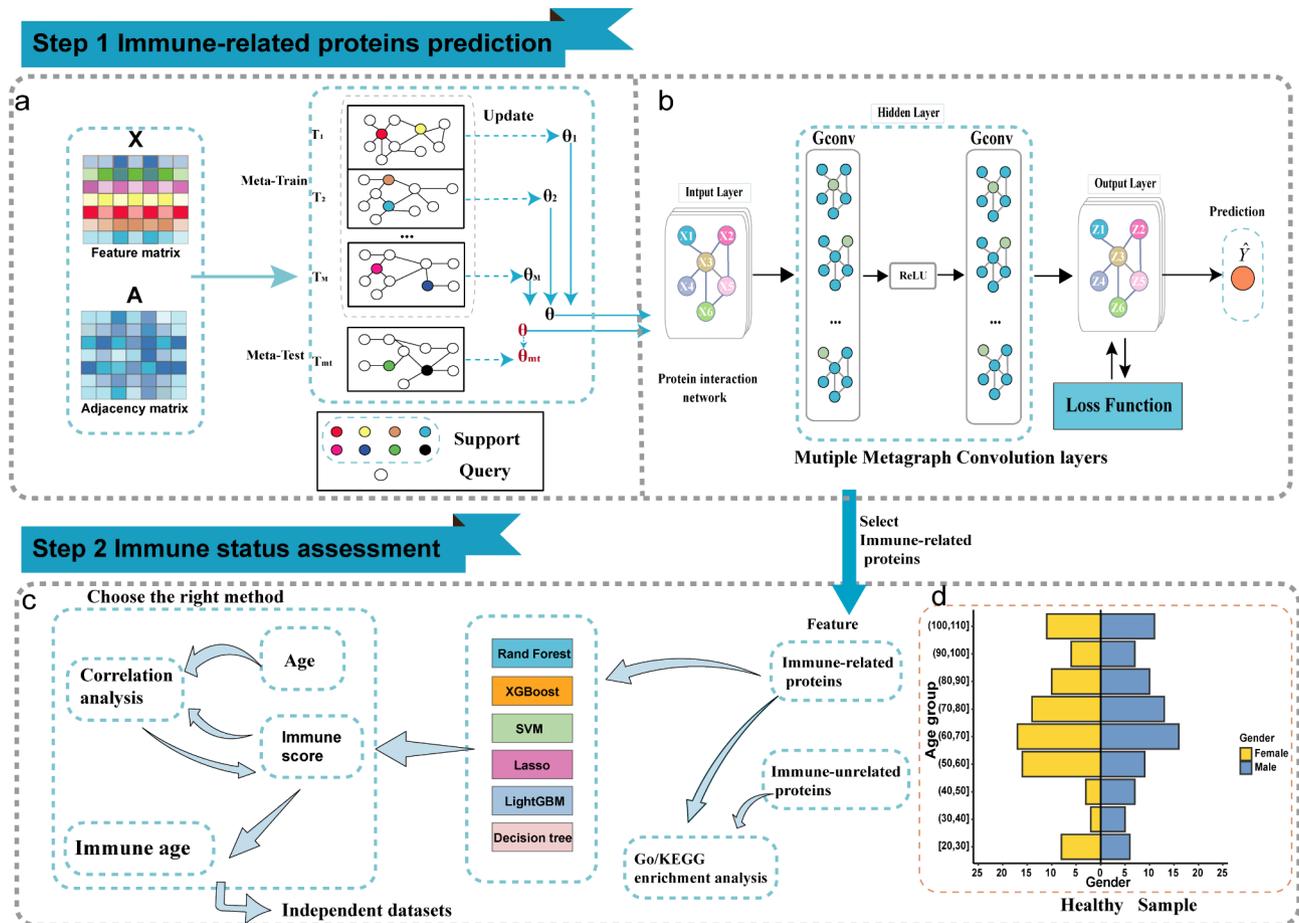


Fig. 1 The Overall workflow of the ProMetaGCN model, which consists of two steps: the prediction of immune-related proteins and the assessment of immune status. In Step 1, **(a)** The meta-learner module optimizes the model parameters, with training divided into m metagraphs; **(b)** The working mechanism of the GCN in immune-related protein prediction. The input to the Meta-GCN model includes feature matrices and adjacency matrices, which facilitate the classification of nodes based on the graph's topological structure. In Step 2, **(c)** The selection of multiple machine learning methods and downstream analyses (including GO/KEGG, correlation analysis, and validation with independent test datasets). **(d)** the age and gender distribution pyramid chart for the healthy sample dataset utilized in this study

Healthy dataset

The dataset is derived from the research conducted by Lehallier et al. [21], based on plasma proteomic data generated using the SOMAscan platform from four independent cohorts in the United States and Europe (VASEattle, PRIN06, PRIN09, and GEHA). It includes plasma samples from 171 participants, aged between 21 and 107 years, comprising 84 males and 87 females. The relative fluorescence units (RFU) of 1,305 plasma proteins were measured and subsequently log₁₀-transformed to form the dataset.

COVID-19 dataset1

This dataset comes from the study conducted by Wang et al. [22], in which levels of 803 plasma proteins were measured using the TMTpro 16-plex platform. To ensure normalization, the data underwent log₂ transformation. This dataset includes individual samples from three stages: the healthy group ($n = 35$, age range 25–64 years),

the acute infection group ($n = 26$, age range 25–67 years), and the post-acute group ($n = 32$, age range 19–69 years).

COVID-19 dataset2

This dataset originates from the study by Zhong et al. [23], which recruited 50 patients aged 19 to 66 years who tested positive for SARS-CoV-2 via PCR. Blood samples were collected for analysis within 24 h of confirming COVID-19 infection (Day 0) and on Day 14. The analysis included NPX values for 1,459 quality-controlled proteins, where NPX represents a relative protein quantification unit on a log₂ scale. The study tracked a cohort of individuals with COVID-19 and compared their plasma protein profiles with those of a healthy control group. All treatments commenced on the day of diagnosis, and by Day 14, all patients tested negative on PCR. The COVID-19 cohort comprised individuals with mild to moderate symptoms who did not require hospitalization.

NSCLC dataset

This dataset is derived from the study by Harel et al. [24], which employed Quantibody multiplex ELISA antibody array technology to measure plasma protein levels. Data were gathered from 58 patients with non-small cell lung cancer (NSCLC) aged between 49 and 85 years, with plasma samples obtained from the Sheba Medical Center in Israel. All data utilized in this research were derived from pre-treatment plasma samples to exclude potential effects of treatment on immune status. The study measured levels of 810 proteins, and the data underwent log2 transformation for normalization.

Semi-supervised meta-learning graph convolutional network

Semi-supervised learning effectively integrates the advantages of labeled and unlabeled data, minimizing annotation costs while enhancing model performance, stability, and adaptability in real-world scenarios. The Meta-GCN combines graph convolutional networks (GCNs) [25, 26] with meta-learning techniques [27] to improve the prediction of immune-related plasma proteins in protein-protein interaction networks. This approach encompasses three key components: adjacency matrix construction, GCN training, and meta-learning optimization.

Initially, we utilized the names of 1,305 plasma proteins from the Healthy dataset as input and analyzed interaction data among 1,207 plasma proteins, which included 82,828 interactions obtained from the STRING database (refer to Table S1 for details). These interactions were selected based on a medium confidence threshold of 0.4. We constructed the protein interaction adjacency matrix using a “combined_score” that integrates various factors, including chromosomal proximity, gene fusion, phylogenetic co-occurrence, homology, co-expression, experimental validation, database annotations, and text mining. This combined_score approach enhances the accuracy of the adjacency matrix, providing a robust foundation for subsequent bioinformatics research. In the protein interaction network, the adjacency matrix A describes the interaction relationships between nodes (proteins). For each pair of proteins (i, j) , we set A_{ij} in the adjacency matrix A to the value of the combined_score between them, ranging from 0 to 1, where a higher value indicates a greater likelihood of interaction. The normalized adjacency matrix \hat{A} is computed as follows:

$$\hat{A} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} \quad (1)$$

Let $\tilde{A} = A + I$, where I is the identity matrix. \tilde{D} represents the degree matrix of \tilde{A} , with each element being the sum of the corresponding row of \tilde{A} . The normalized adjacency matrix \hat{A} facilitates the capture of relationships

between nodes and enhances the representational power of Graph Convolutional Networks (GCNs). Subsequently, GCNs are employed for feature learning on graph-structured data. The graph convolution operation in GCNs is defined as follows:

$$H^{(l+1)} = \sigma(\hat{A}H^{(l)}W^{(l)}) \quad (2)$$

Where $H^{(l)}$ represents the node feature matrix at the l -th layer, W^l denotes the weight matrix at the l -th layer, and σ refers to the activation function *ReLU*.

In this study, the GCN aggregates neighborhood information through two convolutional layers to progressively extract features from the graph. The adjacency matrix is normalized using FirstOrderGCN. Since each protein is represented solely by its name and lacks actual features, we utilize an identity matrix as the feature matrix. The entire forward propagation process from input features X to output predictions Z can be described as follows:

$$Z = f(X, \hat{A}) = \text{Softmax}(A \cdot \text{ReLU}(\hat{A}XW^{(0)})W^{(1)}) \quad (3)$$

where $W^{(0)}$ and $W^{(1)}$ represent the weight matrices for the first and second layers, respectively. *ReLU* serves as the activation function, while *Softmax* is employed as the activation function for the output layer to produce the classification probability distribution.

In semi-supervised learning, we leverage a limited number of labeled nodes alongside a substantial quantity of unlabeled nodes for training. Through literature mining, 35 immune-related proteins (including IFNB1, IL1A, IL23R, TGFB1, C2, C3, C5, C6, C7, C9, CCL13, CCL2, CCL20, CCL28, CXCL1, CXCL5, CXCL6, CXCL8, IFNA2, IFNG, IL10, IL13, IL17A, IL1B, IL2, IL22, IL34, IL4, IL5, IL6, IL9, TNF, CXCL9, GDF15, and CSF1) were designated as positive labels. Subsequently, 100 proteins were randomly selected from the remaining proteins to serve as negative labels. The interaction data of these proteins were utilized for model training and prediction. After training, we obtained the predicted probabilities for the immune-related nodes. While maintaining the positive labels unchanged, we relabeled the 100 proteins with the lowest predicted probabilities as negative and retrained the model until convergence of the prediction results. The training set D_{train} comprises the feature matrix X and the labels Y of the labeled nodes. The model is trained by optimizing the cross-entropy loss function, which is defined as:

$$L_{ce} = - \sum_{i \in L} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (4)$$

where L represents the set of labeled nodes, y_i denotes the true label (0 or 1), and \hat{y}_i is the probability predicted to be 1.

Finally, meta-learning, also known as “learning to learn”, enhances the model’s adaptability across different tasks or environments, particularly in few-shot learning scenarios. It improves model performance when few labeled samples are available by acquiring learning experiences from a small set of samples [27, 28]. The meta-learning framework optimizes the initialization of model parameters to enable rapid adaptation to new tasks. The meta-learning process consists of two stages: model training and model testing. During the model training phase, we randomly select several nodes from each class in the training set D_{train} to form a support set S_i . The remaining nodes constitute the query set Q_i . Each meta-learning task $T_i = S_i + Q_i$ is thus constructed, and by repeating this process M times, we generate M meta-learning tasks. In each meta-learning task, the objective function is:

$$L_{meta} = \sum_{T_i \in \mathcal{T}} L_{task}(f_{\theta^*}(T_i)) \quad (5)$$

where f_{θ^*} is the model optimized through meta-learning, and L_{task} is the loss function for task T_i . The model is updated using the Adam optimizer with the following update rule:

$$\theta_{t+1} = \theta_t - \alpha \cdot \nabla_{\theta} L(\theta_t) \quad (6)$$

where θ_t represents the model parameters at the t -th iteration, α is the learning rate, and $\nabla_{\theta} L(\theta_t)$ is the gradient of the loss function with respect to the parameters.

Using this method, we trained an efficient model based on a small number of known immune-related proteins, capable of comprehensively predicting additional immune-related proteins for further assessments of immune status.

Prediction of immune status

To predict immune status, we evaluated six commonly utilized machine learning methods: Lasso regression, Support Vector Machine (SVM), LightGBM, Random Forest, XGBoost, and Decision Tree. Each method was systematically applied to forecast immune status, with the best-performing technique identified based on a range of performance metrics and the Pearson correlation between predicted and actual values.

We initiated this process by training and testing each method, randomly partitioning the dataset in an 8:2 ratio. To facilitate comparability of prediction results across different methods and ensure experimental reproducibility, the “random_state” parameter was set to 42. For

feature selection, we incorporated 309 immune-related proteins as input features. The output was determined by immune status scores derived from our previous study of 16,705 healthy individuals, which were mapped onto a scale from 60 to 100, taking into account age-related factors; further details can be found in our prior research [14]. During model training, we executed multiple training iterations for each method while optimizing relevant hyperparameters to ensure an optimal fit on the training set and robust generalization capabilities. To evaluate the performance of each method, we employed several common error metrics, including mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R^2). Additionally, we calculated the Pearson correlation coefficient between the predicted and actual values to assess the predictive accuracy of the models further. By comparing the correlation coefficients (r values) of each method, we identified those with a high correlation (≥ 0.9) as the final predictive model.

Ultimately, we amalgamated the predictions from various methods by calculating the mean of the predicted values to derive the final immune status score, aiming to enhance the accuracy and stability of the predictions through the consolidation of multiple machine learning models.

Immune age gap

Immune age more accurately reflects an individual’s aging and immune status. Previous research has established the relationship between age and immune status score [14], deriving a formula shown in the left part of Figure S1. In this study, focusing on healthy individuals, we set the immune status score to be no less than 60 to make it more intuitive. The original score range of 0–1 was remapped to 60–95 based on interval proportions, as shown in the right part of Figure S1. Based on current data, we’ve set the reference range from 19 to 110 years old and assigned an immune status score for each age in this span in Table S2. Compare the individual’s calculated immune status score with those in Table S2. The age with the smallest difference in scores is the person’s immune age. Subsequently, this immune age is compared with the chronological age to assess whether the individual’s immune status is better or worse relative to their chronological age. The difference between the immune age and chronological age is defined as the “ImmuneAgeGap”, as shown in Eq. (7):

$$ImmuneAgeGap = Immune\ age - Chronological\ age \quad (7)$$

If $ImmuneAgeGap < 0$, it indicates that the individual’s immune system is performing more youthfully than their chronological age, reflecting immune rejuvenation (immu

nerjuvenation). Conversely, if *ImmuneAgeGap* > 0, it suggests that the individual's immune system is older in comparison to their chronological age, indicating immunosenescence. Immunosenescence refers to the gradual decline and functional deterioration of the immune system with age, leading to weakened immune capacity. This deterioration increases susceptibility to infections, diseases, tumors, and other immune-related issues. The degradation of the immune system is reflected in both cellular and humoral immunity, impacting antibody production and the functionality and responsiveness of immune cells. This evaluation method facilitates the quantification and in-depth analysis of an individual's immune system health status.

Statistical analysis

In this study, we applied z-score normalization to the protein data from various datasets during the construction of the predictive model. This method addresses disparities in feature scales across datasets, thereby ensuring the model's effective applicability to data from different studies.

In this study, we applied z-score normalization to the protein data from various datasets during the construction of the predictive model. This method addresses disparities in feature scales across datasets, thereby ensuring the model's effective applicability to data from different studies.

The protein-protein interaction (PPI) network of immune-related proteins was constructed using STRING (<https://cn.string-db.org/>) and Cytoscape version 3.7.0. In our study, we employed the Mann–Whitney U test and Pearson correlation analysis. The Mann–Whitney U test was applied to compare the expression differences of immune-related proteins between sample groups, with statistical significance defined as $p < 0.05$, to uncover the potential functions of these differences in immune processes. Pearson correlation analysis was utilized to assess linear correlations between variables and to explore interactions among immune proteins. Statistical significance thresholds were established as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, while $p \geq 0.05$ was designated as “ns” (not significant).

Results

Identifying potential immune-related proteins in plasma using Meta-GCN

In predicting immune-related proteins in plasma using the Meta-GCN model, the model parameters were configured for 20 iterations and 50 training epochs, with a learning rate of 0.0001. During training, 35 immune-related proteins were designated as positive, while 100 randomly selected proteins served as negative samples; the remaining nodes were classified into other categories.

In each iteration, 20 random selections of nodes were performed, each selection comprising 10 negative and 5 positive protein nodes to ensure comprehensive coverage of all known positive samples. This approach facilitated adequate representation of both positive and negative samples during training. In the testing phase, a pre-trained model was employed to predict outcomes for all nodes, with the 100 proteins exhibiting the lowest predicted probabilities being relabeled as negative and subsequently used for retraining. This process was iterated until the predictions converged. To enhance the reliability of the results, the entire procedure was repeated 100 times, and the median predicted probability for each protein was recorded as its immune-related probability.

Utilizing this methodology, we identified 307 proteins with prediction probabilities exceeding 0.95, as presented in Table S3, with the probability range illustrated in Fig. 2. Furthermore, through a literature mining, we augmented the dataset with two proteins whose prediction probabilities were below 0.95—one with a probability greater than 0.9 and the other exceeding 0.75—bringing the total count of candidate immune-related proteins to 309. In subsequent validations using external datasets, these candidate proteins, along with the intersection of all protein types, were utilized as input features for the immune status evaluation model.

Biological significance and potential functions of the predicted immune-related plasma proteins

To investigate the biological significance of the predicted immune-related proteins, we conducted GO/KEGG enrichment analysis on plasma proteins with prediction scores exceeding 0.95, as presented in Fig. 3. The results indicated that the ten most significantly enriched biological processes and pathways were primarily associated with immune responses and inflammation. These included cytokine-mediated signaling pathways, leukocyte migration, cytokine-cytokine receptor interactions, and the JAK-STAT signaling pathway (see Fig. 3a and c). These findings suggest that proteins with prediction scores above 0.95 are indeed linked to immune functions. Additional details regarding the GO/KEGG biological processes and pathways can be found in Table S4.

Concurrently, we performed GO/KEGG enrichment analysis on proteins that were predicted to be unrelated to immune function, with scores below 0.05. The results showed that the ten most significantly enriched biological processes and pathways were primarily connected to growth and development, with axon development and axon guidance pathways being particularly prominent (refer to Fig. 3b and d). This further substantiates that proteins with predicted values below 0.05 are indeed unrelated to immune functions. Detailed descriptions of

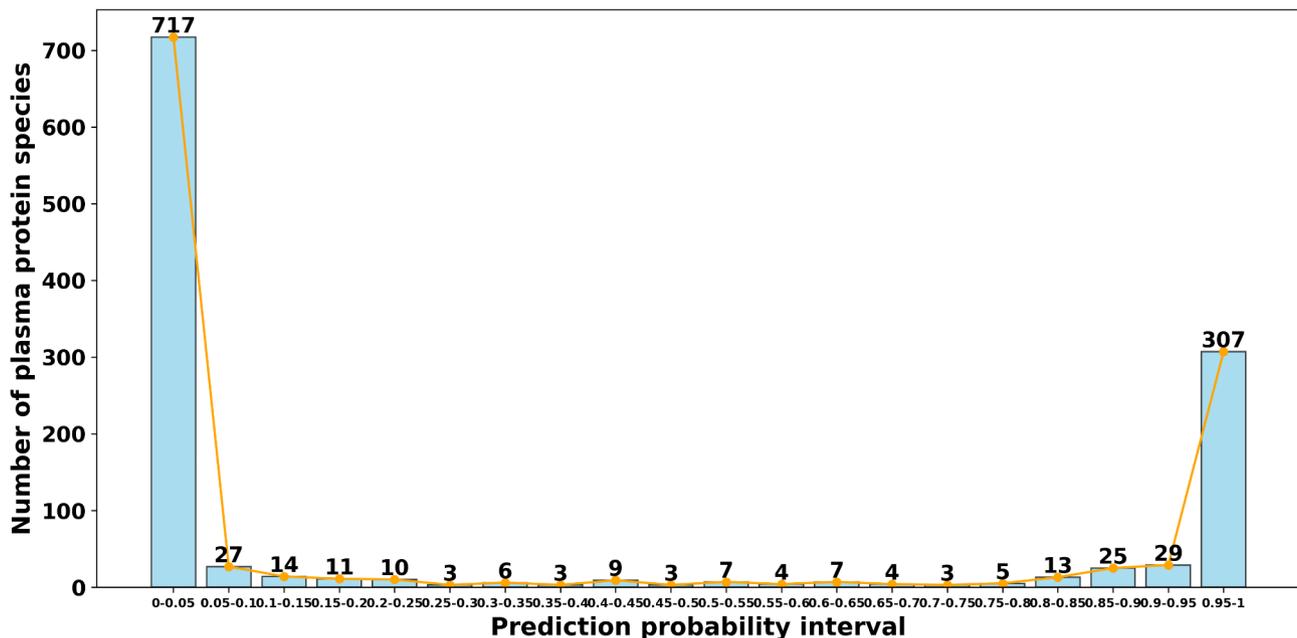


Fig. 2 The probability distribution of immune-related protein predictions by Meta-GCN

the enriched GO/KEGG biological processes and pathways are provided in Table S5.

In our analysis of the top ten GO biological processes, we identified four proteins (CCL19, IL1B, IL6, and XCL1) that collectively contribute to these processes, as illustrated in Fig. 3e. These four proteins play critical roles in the immune system. Specifically, CCL19 is primarily responsible for the migration of T cells and dendritic cells, thus facilitating immune responses. It also participates in thymocyte development and the functions of regulatory and memory T cells [29]. IL1 β , predominantly produced by monocytes and macrophages, is a crucial factor in innate immunity, regulating inflammatory responses and promoting the activation of inflammatory and immune cells [30]. IL6 (interleukin-6) is mainly synthesized by macrophages, T cells, and B cells; it is involved in the differentiation of Th17 cells and enhances the proliferation and differentiation of B cells, thereby boosting humoral immune responses [17]. XCL1, an important C-type chemokine, is primarily secreted by activated T cells and natural killer (NK) cells. It facilitates the migration of immune cells and the establishment of self-tolerance, thus strengthening cytotoxic immune responses [16].

These cytokines are crucial for assessing immune status and studying immune-related diseases. Variations in their expression levels can yield valuable insights into immune status and disease progression.

Interaction analysis of immune-related proteins

We conducted an analysis of interactions among immune-related proteins (Fig. 4), which highlighted

the close connections between various protein families. These six modules are not isolated; rather, they are interconnected by specific proteins that form a larger network, as illustrated in Fig. 4a. Within the chemokine network (Fig. 4b), cytokines such as CCL11 and CCL2 regulate the migration and chemotaxis of immune cells, directing leukocytes to sites of inflammation and tumors to enhance immune responses. The interleukin network (Fig. 4c) encompasses essential functions related to the proliferation, differentiation, and activation of immune cells; IL-6 and IL-10 play vital roles in modulating immunity and inflammation. The tumor necrosis factor network (Fig. 4d) influences cell death and immune activation, with TNF- α being crucial for apoptosis and immune responses through its receptors (TNFRSF1A and TNFRSF1B). The complement system (Fig. 4e) bolsters immune responses by identifying and eliminating pathogens, featuring critical components such as C3 and C5 that augment antibody and leukocyte function. The colony-stimulating factor network (Fig. 4f) supports the proliferation and differentiation of hematopoietic stem cells, which are essential for the generation of immune cells; CSF1 and CSF3 facilitate the differentiation of bone marrow cells. Finally, the interferon network (Fig. 4g) is pivotal for antiviral defense, with interferons (e.g., IFNG and IFNA2) enhancing immune responses against viral infections. Collectively, these cytokines and their receptors coordinate the body's adaptive strategies to confront external threats and manage disease.

In the protein interaction network, “degree” denotes the number of interactions a specific protein has with other proteins. This metric reflects the protein’s centrality

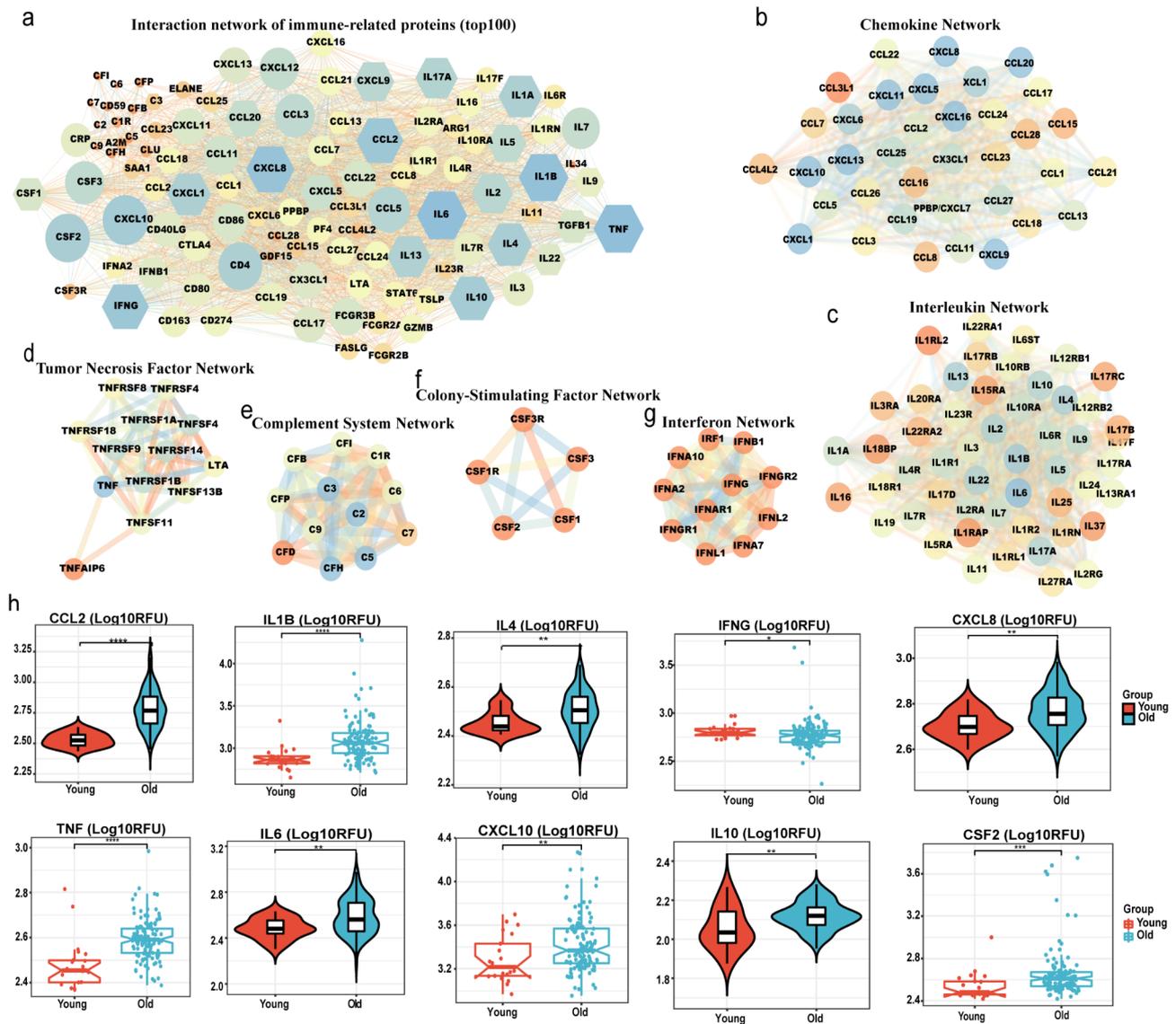


Fig. 4 Protein interaction network and its analysis. (a) Interaction network of immune-related proteins (top100); node size is proportional to the degree, with cooler colors indicating higher degree values; (b) Chemokine Network; (c) Interleukin Network; (d) Tumor Necrosis Factor Network; (e) Complement System Network; (f) Colony-Stimulating Factor Network; (g) Interferon Network; (h) Differential expression analysis of the top 10 proteins by interaction degree in different age groups (Mann-Whitney U test)

analysis of the top ten proteins ranked by degree between the young and old groups. The young group consisted of individuals aged 20–40, while the old group included individuals aged 60 and above. The results, presented in Fig. 4h, reveal significant differences in CSF2, IL1B, CCL2, CXCL8, CXCL10, IL6, TNF, IL4, IL10, and IFNG between the two age groups.

This shows that immune system regulatory mechanisms change significantly with age, especially in the elderly, where alterations in protein expression are closely related to immune dysfunction and chronic inflammation. For instance, increased CSF2 and IL1B expression may indicate chronic inflammation and immune cell dysfunction. Higher levels of pro-inflammatory cytokines

like IL1B, IL6, and TNF suggest exacerbated chronic inflammation in the elderly. Changes in proteins such as CCL2, CXCL8, and CXCL10 may be linked to increased immune cell recruitment at inflammation and injury sites. Also, altered expression of proteins like IL4, IL10, and IFNG, which are crucial for immune cell proliferation, differentiation, and regulation, may be associated with immune system decline. These key regulatory protein expression changes reveal the immune system’s complex regulatory network during aging. The chronic inflammation common in the elderly may stem from reduced immune system function and regulatory capacity, and changes in these key regulatory proteins may worsen this condition. Moreover, the decline in immune

cell function is closely related to these protein expression changes. Overall, analyzing age-related differences in immune-related protein expression not only reveals significant immune system changes with age but also provides insights into the mechanisms of immune decline and chronic inflammation during aging, offering potential targets for further research on age-related immune regulation.

Prediction of immune status score

This study aims to evaluate the effectiveness of various machine learning algorithms in predicting immune status scores, thereby providing an efficient approach for assessing individual immune status. We selected six algorithms—Lasso, Decision Tree, LightGBM, Support Vector Machine (SVM), XGBoost, and Random Forest—primarily due to their exceptional performance in managing high-dimensional features and complex non-linear relationships. The extensive application of these algorithms in the fields of medicine and biology establishes a robust theoretical foundation for our research.

In this experiment, we utilized immune-related proteins as input features and immune status scores as labels for model training, with the results and analyses presented in Fig. 5. To evaluate the predictive performance of the six algorithms, we primarily employed the Pearson correlation coefficient to measure the linear relationship between predicted and true values. Additionally, we computed several error metrics, including the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results indicate (Fig. 5a) that, considering the five evaluation metrics, the Pearson correlation coefficients for the predictions made by Lasso, LightGBM, XGBoost, and Random Forest all exceeded 0.9, demonstrating that these algorithms significantly outperformed Decision Tree and SVM in this study. This finding suggests that Lasso, LightGBM, XGBoost, and Random Forest are more accurate in capturing variations in immune status scores. To enhance the stability of the experimental results, we used the average of the predicted values from these four algorithms as the final immune status score for each individual.

All four machine learning methods exhibit interpretability, allowing for the assessment of each feature's importance in relation to the predictive outcomes. Figure 5b illustrates that among the top ten important features identified by these four methods, three proteins—ADAMTS13, GDF15, and SERPINF2—appeared repeatedly, suggesting these proteins may play a significant role in predicting immune status scores. To further validate this conclusion, we conducted a differential analysis based on age and gender (Fig. 5c and d). The results indicate no significant differences between genders;

however, a notable difference was observed between the younger group (ages 20–40) and the older group (60 years and older). Specifically, the expression level of GDF15 was higher in the older group, whereas ADAMTS13 and SERPINF2 exhibited higher expression levels in the younger group.

It is worth noting that there is a certain functional association between these three proteins and the potential key regulatory proteins identified by the interaction network analysis. In the interaction network analysis, we focused on the interactions between proteins and identified proteins with high “degree” values in the immune-related protein network, such as CSF2, IL1B, CCL2, CXCL8, CXCL10, IL6, TNF, IL4, IL10, and IFNG. These proteins exhibited significant expression differences between the younger and older groups, reflecting changes in immune regulatory mechanisms with age, associated with immune function decline and chronic inflammation.

ADAMTS13, identified as an important feature in the machine learning model and highly expressed in the younger group, participates in cellular responses to bacterial-derived molecules (GO:0071219, GO:0002237), which is functionally related to multiple proteins in the interaction network (e.g., TNF, CCL2, CXCL10, CXCL8, IL6, IL1B, CSF2), potentially playing a role in regulating inflammatory responses and immune cell activation. Additionally, its involvement in cellular responses to interleukin-4 (GO:0071353) is related to the interleukin network in the interaction network (e.g., IL-4), which may be important for immune cell proliferation, differentiation, and regulation.

GDF15, also an important feature in the machine learning model and highly expressed in the older group, participates in cytokine-cytokine receptor interaction (hsa04060), which is functionally related to multiple cytokines in the interaction network (e.g., TNF, CCL2, CXCL10, CXCL8, IL6, IL1B, CSF2), potentially playing a role in immune regulation and inflammatory responses. Moreover, its role in the acute-phase response (GO:0006953) may involve regulating inflammatory responses and immune cell activation in response to tissue damage or infection.

SERPINF2, identified as an important feature in the machine learning model and highly expressed in the younger group, participates in the complement and coagulation cascades (hsa04610), which is functionally related to the complement system network in the interaction network (e.g., C3, C5), crucial for the normal functioning of the immune system. Additionally, its role in the acute-phase response (GO:0006953) may involve regulating inflammatory responses and immune cell activation in response to tissue damage or infection, with related proteins including TNF, SERPINF2, IL6, and IL1B.

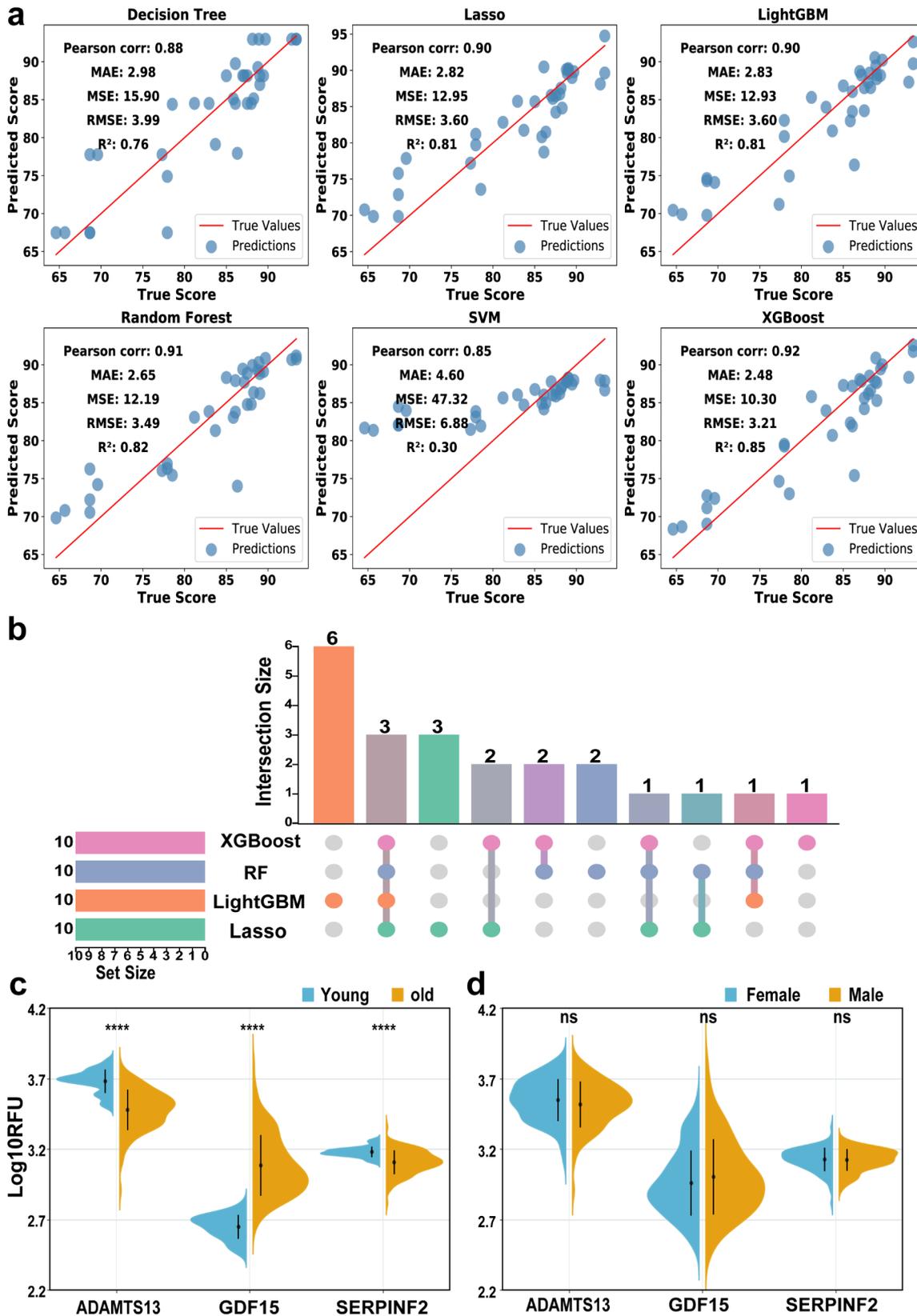


Fig. 5 Predictions and protein expression differences across various machine learning models. **(a)** Correlation analysis between the predicted results of Lasso, Decision Tree, LightGBM, SVM, XGBoost, and Random Forest and the true age. **(b)** Intersection of the top ten important proteins identified by the four machine learning methods (upset plot). **(c)** Analysis of protein expression differences across different age groups (Mann–Whitney U test). **(d)** Analysis of protein expression differences between different genders (Mann–Whitney U test)

In summary, although ADAMTS13, GDF15, and SERPINF2 were not directly identified in the interaction network analysis, their immune system functions are associated with the mechanisms of key proteins in the interaction network, suggesting they may play important roles in immune regulation and aging.

External validation of model performance

In this study, two external datasets were employed to validate the model and evaluate its effectiveness and reliability. Firstly, the immune status assessment model was validated using samples from healthy individuals. Secondly, the model was assessed by comparing immune status scores across groups with varying immune statuses, thereby examining differences in immune status scores between different stages of infection and healthy individuals. The analysis of immune status scores from these

two datasets is detailed below, with results illustrated in Fig. 6.

For the COVID-19 dataset1, due to its external origin, not all known immune-related proteins were included. Therefore, we identified the intersection of immune-related proteins with all measured proteins in the dataset, resulting in 81 common features for model input. These features were utilized in the immune status score prediction model to calculate scores for each sample. To confirm the model's effectiveness in assessing the immune status of healthy individuals, we conducted a correlation analysis between immune status scores and age within the healthy individuals of the COVID-19 dataset1. The results, presented in Fig. 6a, showed a correlation coefficient of $r = -0.444$ with a p-value less than 0.01, indicating a significant negative correlation between immune status scores and age. This finding demonstrates the validity

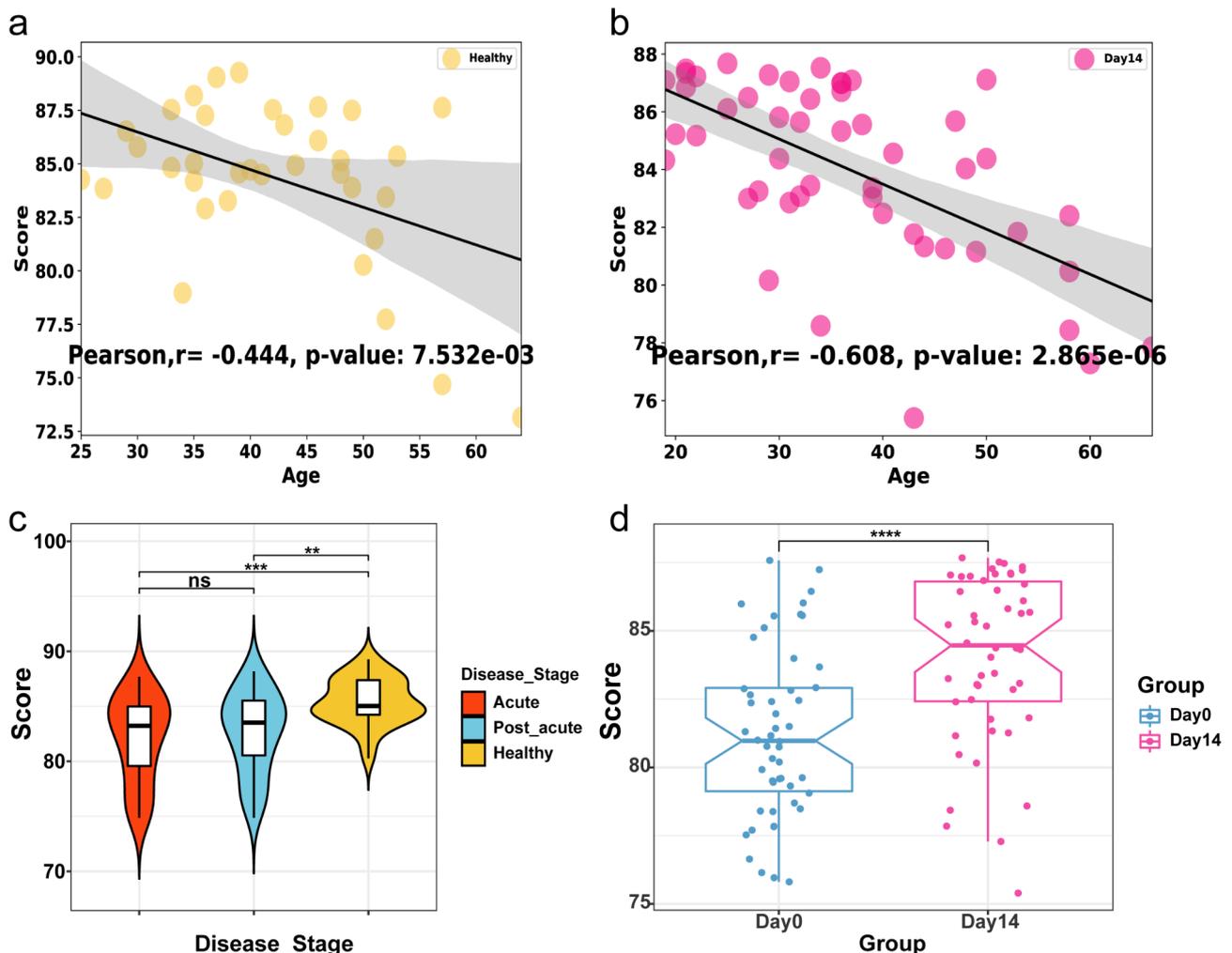


Fig. 6 Analysis of immune status in two external datasets. (a) Correlation analysis between immune status scores and age in the healthy control group of the COVID-19 dataset1. (b) Correlation analysis between immune status scores and age on Day 14 (negative nucleic acid test) in the COVID-19 dataset2. (c) Comparison of immune status scores across different immune status groups in the COVID-19 dataset1. (d) Comparison of immune status scores at different time points (Day 0 vs. Day 14) post-infection in the COVID-19 dataset2

of the model in evaluating the immune status of healthy individuals.

In the COVID-19 dataset2, we identified the intersection of immune-related proteins with all measured proteins in the dataset, resulting in 185 common features for model input. These features were used to compute immune status scores for individuals at early infection (Day 0) and recovery (Day 14) stages. Day 0 and Day 14 correspond to infection day 0 and day 14 of the same cohort of 50 samples, respectively, maintaining a one-to-one relationship. Further analysis revealed a significant negative correlation between immune status scores at Day 14 and age, with a correlation coefficient of $r = -0.608$ and a p -value less than 0.0001, indicating statistical significance, as shown in Fig. 6b. This result suggests that age significantly impacts immune status, indicating the necessity of considering age when evaluating immune recovery. This finding is consistent with the negative correlation observed in the healthy individuals of the COVID-19 dataset1, further validating the stability of the immune status assessment model.

To assess differences in immune status scores across various immune states, we validated the model's application in infectious diseases. In the COVID-19 dataset1, we compared immune status scores among three groups: healthy individuals, patients in the acute infection phase, and those in the post-acute infection phase. Due to the lack of a one-to-one correspondence between samples in different groups (only samples XGO_81, XGO_91, and XGO_92 have data available for both acute and post-acute infection phases, while other samples appear in only one group), individual differences may influence the results. In the Mann-Whitney U test, shown in Fig. 6c, although there was no statistically significant difference between acute and post-acute infection phases, the median immune status score of the post-acute group was slightly higher than that of the acute group. Compared to the healthy control group, the immune status score was significantly higher in the healthy group than in the post-acute group ($p < 0.01$), while the difference between the acute infection group and the healthy control group was even more significant ($p < 0.001$).

Table 1 Immune status scores of the same sample at different stages and days of infection

Sample	Age	Disease Stage	score	Time (days)
XGO_81	36	Acute	84.379	9
XGO_81	36	Acute	84.784	12
XGO_81	36	Post acute	85.673	18
XGO_91	48	Acute	74.868	12
XGO_91	48	Post acute	76.372	22
XGO_92	60	Acute	75.611	14
XGO_92	60	Post acute	77.858	25

In the COVID-19 dataset2, we calculated immune status scores for the early infection phase (Day 0) and the recovery phase (Day 14). The Mann-Whitney U test results, shown in Fig. 6d, indicated that the immune status score on Day 14 was significantly higher than on Day 0 ($p < 0.0001$), suggesting a significant improvement after 14 days of recovery.

To further investigate the effect of infection stages on immune status scores, we examined scores at different infection days in the COVID-19 dataset1 (see Table 1). For example, in sample XGO_81, infection days during the acute phase were 9 and 12 days, with a noticeable score increase at Day 12, suggesting gradual improvement as infection progresses. For samples XGO_91 and XGO_92, post-acute phase scores were higher than those in the acute phase. These results indicate that, following acute infection, the patient's immune status significantly improves as recovery progresses, underscoring the importance of timely monitoring and assessment of immune status.

Impact of the immune age gap

To further evaluate the immune status of individuals, we predicted the immune score to determine each individual's immune age and utilized the Immune Age Gap (ImmuneAgeGap) to assess the quality of their immune status, as illustrated in Fig. 7. The test set comprised 34 healthy individuals, 29 of whom were over the age of 50, including 12 individuals aged over 80. This distribution reflects the characteristics of our dataset, which predominantly consists of elderly individuals with longevity. In Fig. 7a, segments with ImmuneAgeGap < 0 indicate immune rejuvenation, while ImmuneAgeGap > 0 signifies immunosenescence. The ImmuneAgeGap values are classified into four categories: Low, Moderate, Severe, and Extreme.

Our analysis revealed that the ImmuneAgeGap values for healthy individuals in the test set were primarily concentrated within the range of -15 to 15, suggesting a degree of discrepancy between an individual's immune age and chronological age. This discrepancy may be associated with the adaptive capacity of the immune system in individuals of advanced age (> 90 years). Research indicates that the immune systems of longevity individuals often display enhanced resistance and recovery capabilities, implying that their immune status may be relatively younger than their chronological age, as demonstrated in Fig. 7b. Furthermore, longevity individuals typically benefit from factors such as lifestyle, genetic predispositions, and nutritional intake, all of which may contribute to improved immune function. Therefore, the phenomenon observed in our study, where many longevity individuals have lower immune ages than their chronological ages, is reasonable. This finding underscores the necessity

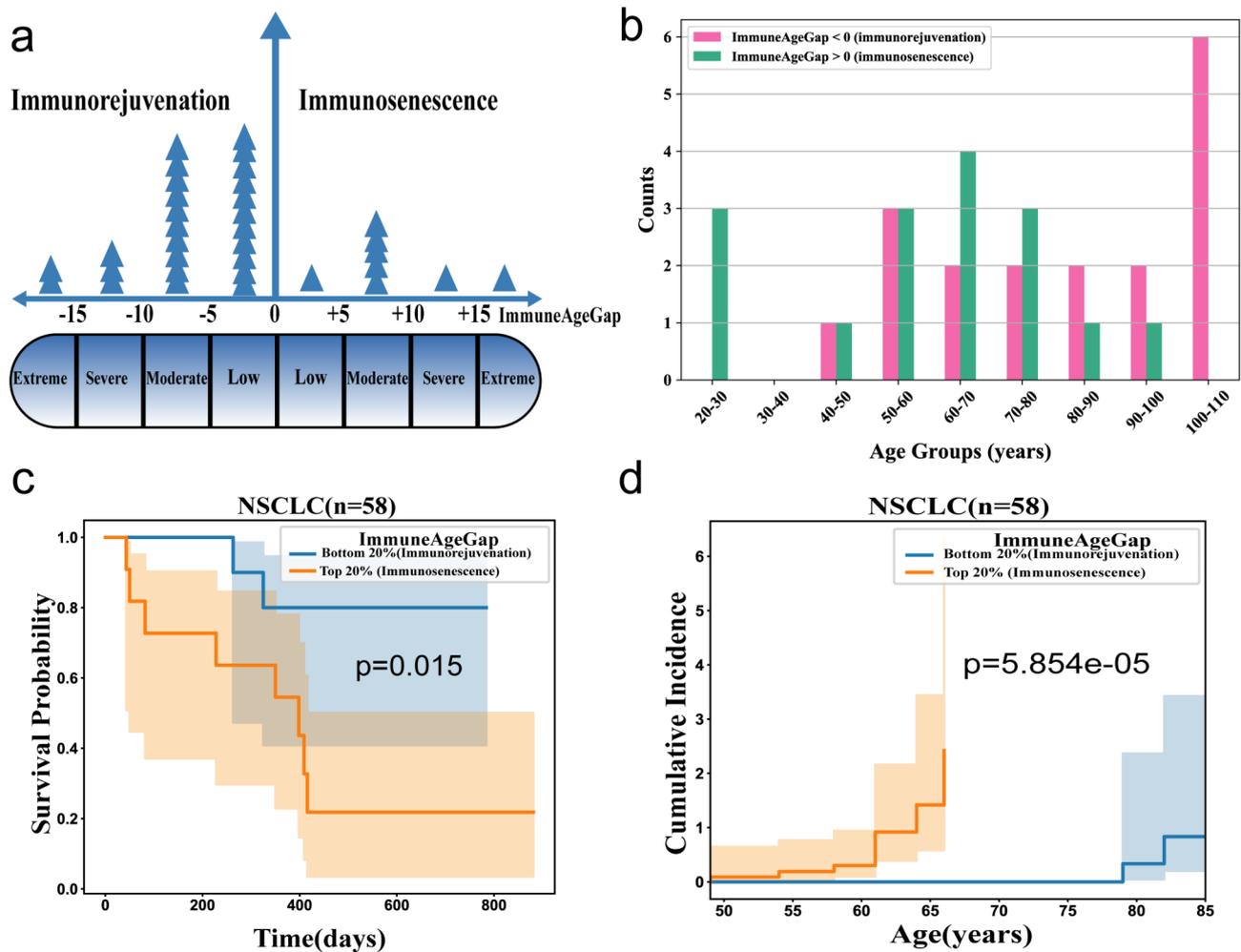


Fig. 7 Association analysis of ImmuneAgeGap with survival prognosis and cancer risk. **(a)** Distribution of the age difference (ImmuneAgeGap) between immune age and chronological age in the healthy individuals test set; **(b)** The count of ImmuneAgeGap > 0 and ImmuneAgeGap < 0 in each age group within the healthy individual test set; **(c)** Survival analysis of NSCLC (pre-treatment) patients grouped by ImmuneAgeGap. The survival curves for the immune rejuvenation group (Bottom 20%) and the immunosenescence group (Top 20%) are displayed, accompanied by statistical testing. The lighter shaded area in the figure represents the 95% confidence interval. **(d)** Trajectory analysis of ImmuneAgeGap in NSCLC patients concerning age-specific cancer risk, where the x-axis indicates age and the y-axis represents the cumulative incidence of risk

for further investigation into the immune health status of elderly individuals.

In a study on non-small cell lung cancer (NSCLC) employing an external dataset [24], we conducted survival analyses on the top and bottom 20% of the Immune Age Gap (ImmuneAgeGap) groups. The analysis revealed a significant difference in survival rates between these two groups. The survival curve for the immune rejuvenation group (bottom 20%) was markedly higher than that of the immunosenescence group (top 20%), indicating that individuals with immune rejuvenation had a higher likelihood of survival (see Fig. 7c). Statistical analyses further confirmed that the difference in survival rates between these groups was statistically significant ($p < 0.05$), highlighting the critical role of ImmuneAgeGap in survival prognosis. This finding underscores the

potential value of immune status in predicting survival rates and lays a theoretical foundation for immuno-status-based intervention strategies. Additionally, we examined the age-specific cumulative incidence of NSCLC diagnosis within both the immunosenescence group (top 20%) and the immune rejuvenation group (bottom 20%). The results revealed a significant difference between the two groups (see Fig. 7d). The immunosenescence group displayed an earlier age of onset for NSCLC (death) and a higher risk of mortality, while the immune rejuvenation group exhibited a later age of onset and a lower risk of incidence. This outcome further supports the relationship between the ImmuneAgeGap and NSCLC risk. The biological significance of these findings is that immune system function is crucial in the aging process. Immune rejuvenation may indicate a stronger immune

system with better adaptability and resistance, which could be more effective in combating diseases and aging. This could be linked to factors such as lifestyle, genetics, and nutrition that influence immune function. Thus, a smaller ImmuneAgeGap might be associated with better immune function, enabling individuals to better fight diseases, such as through more efficient pathogen recognition and clearance or better-regulated immune responses to prevent excessive inflammation.

In this study, we also analyzed the relationship between immune scores and prognosis in NSCLC patients. Higher immune scores imply better immune function. We divided patients into Max 20% and Min 20% groups based on immune scores. Results in Figure S2 show that the high immune score group had a significantly higher survival rate ($p=0.036$) and a lower cumulative risk of death ($p=0.019$) compared to the low immune score group. This indicates that patients with higher immune scores tend to have a better prognosis. This finding is in line with the conclusions drawn from the ImmuneAgeGap analysis, further highlighting the importance of immune status in predicting patient survival rates and disease risk. The biological significance lies in the immune system's crucial role in disease resistance and overall health. Patients with higher immune scores may have immune systems that function more effectively, enabling better pathogen recognition, clearance, and regulation of immune responses to prevent excessive inflammation, thus contributing to a better prognosis.

Discussion

In this study, we developed the ProMetaGCN framework to explore immune-related plasma proteins and predict immune status scores, thereby facilitating effective assessment of individual immune health. Integrating external validation datasets, we thoroughly investigated these proteins and evaluated the robustness of the immune status scores, yielding several significant findings.

ProMetaGCN demonstrated outstanding performance in predicting immune-related proteins, successfully identifying 309 proteins linked to immune responses, including key regulatory proteins such as CCL19, IL-6, IL-1 β , and XCL1. Protein interaction analyses further corroborated these findings and provided new insights into immune mechanisms. The framework's integration of diverse machine learning techniques enhanced the accuracy of immune status score predictions. Validation results revealed significant variations in immune status across different stages of infectious diseases, including COVID-19. These results underscored the model's efficacy in evaluating immune recovery, as scores exhibited a negative correlation with age.

Feature importance analysis revealed significant roles for ADAMTS13, GDF15, and SERPINF2 among the immune-related proteins. Literature indicates ADAMTS13 is involved in thrombosis regulation and inflammatory response suppression, alleviating excessive inflammation through interactions with endothelial cells [31, 32]. GDF15, an important regulatory factor, promotes inducible T regulatory cell generation by binding to the CD48 receptor on T cells while enhancing immunosuppressive functions of natural Treg cells. Additionally, GDF15 inhibits dendritic cell maturation and function, contributing to tumor immune evasion by promoting TGF β 1 secretion and suppressing T cell activation. Recent studies show GDF15 inhibits dendritic cell function via interaction with CD44, facilitating immune evasion in ovarian cancer [33–35]. Therefore, GDF15 is crucial in regulating immune responses, suppressing inflammation, and maintaining immune tolerance. Literature also indicates SERPINF2 protects by regulating inflammatory factors and modulating complement and coagulation cascades [36, 37].

In exploring immune status scores to assess immune age, we introduce the concept of ImmuneAgeGap to evaluate individual immune quality. Results show that ImmuneAgeGap effectively distinguishes immune rejuvenation from aging, demonstrating relevance in both healthy individuals and cancer patients. Longevity cohorts often display signs of immune rejuvenation, potentially linked to adaptability, lifestyle, and genetics. In NSCLC patients, lower ImmuneAgeGap values correlate with higher survival rates and lower mortality, while immune aging corresponds to poorer prognoses. These findings highlight immune age disparities' potential in assessing immune status, providing theoretical foundations for personalized interventions in immune-related diseases, including cancer. Future integration of immune age with biomarkers could lead to precise diagnostics and treatment strategies for immunological diseases.

This study presents a novel model for immune protein screening and immune status assessment by combining machine learning with biological data. ProMetaGCN not only enhances the predictive accuracy of immune-related proteins but effectively evaluates immune health through immune status scoring. The model holds promise for early diagnosis of immune diseases, particularly in cancer immunotherapy and personalized treatment strategies.

Despite these achievements, several limitations still exist. Although multiple external datasets were used for validation, the limited sample size and data types may affect the model's generalizability, particularly when verifying outcomes across different diseases. Additionally, the relationship between immune status scores and factors such as genetics and lifestyle requires further investigation. Future research could focus on increasing

sample size, extending follow-up time and dataset diversity (including the number and proportion of immune cells and subtypes, and the diversity of TCR and BCR) to validate the ProMetaGCN model's universality and stability. Furthermore, incorporating a wider range of immunological markers could improve predictive accuracy, and integrating concepts like immune age with other biomarkers could facilitate a comprehensive assessment. As clinical data accumulate, validating the link between immune status scores, disease progression, treatment efficacy, and prognosis will be crucial for advancing personalized immunotherapy.

In summary, the ProMetaGCN model provides an innovative tool for screening immune-related proteins and assessing immune status from plasma protein data, thus advancing immunological research. While the model has limitations, future studies will aim to optimize its performance, offering more precise guidance for the prevention and treatment of immune-related diseases.

Conclusion

The ProMetaGCN model introduced in this study demonstrates substantial advantages in predicting immune-related proteins and assessing immune status utilizing plasma protein data from healthy individuals. This model not only effectively identifies critical immune-related proteins but also clarifies their essential roles in immune responses and inflammatory processes. By integrating literature mining, meta-learning graph convolutional networks, and diverse machine learning techniques, ProMetaGCN establishes new theoretical foundations and practical tools for personalized immunotherapy and the assessment of immune health.

Abbreviations

PPI	Protein-protein interaction
LASSO	Least Absolute Shrinkage and Selector Operator
LightGBM	Light Gradient Boosting Machine
XGBoost	Extreme Gradient Boosting
NSCLC	Non-small-cell lung cancer
GCN	Graph Convolutional Networks
Meta-GCN	Meta-learning graph convolutional network
SVM	Support Vector Machine
MAE	Mean absolute error
MSE	Mean squared error
RMSE	Root mean squared error
R ²	The coefficient of determination
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
CCL19	C-C motif chemokine 19
IL1B	Interleukin 1 beta
IL6	Interleukin-6
XCL1	Lymphotactin
CSF2	Granulocyte-macrophage colony-stimulating factor
CCL2	C-C motif chemokine 2
CXCL8	Interleukin-8
CXCL10	C-X-C motif chemokine 10
TNF	Tumor necrosis factor
IL4	Interleukin-4
IL10	Interleukin-10
IFNG	Interferon gamma

ADAMTS13	A disintegrin and metalloproteinase with thrombospondin motifs 13
GDF15	Growth/differentiation factor 15
SERPINF2	Alpha-2-antiplasmin

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11537-6>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6

Acknowledgements

We would like to sincerely thank Shanghai Unicar-Therapy Bio-medicine Technology Company for their guidance in AI and immunology.

Author contributions

M.Z wrote the main manuscript text and prepared figures and tables. S.W and Q.C conducted code debugging. J.Y and N.X provided biological significance guidance. H.L performed partial data processing. C.Z, L.Y, and W.F reviewed and revised the manuscript. All authors reviewed the manuscript.

Funding

This work was supported by the China National Natural Science Foundation (62172121, 82073800) and Natural Science Foundation of Heilongjiang Province of China (LH2022F012).

Data availability

This study leverages publicly accessible data, with access identifiers detailed within the manuscript. Specifically, the interaction data pertaining to plasma proteins were sourced from the STRING database, accessible at <https://string-db.org/>. Furthermore, we have referenced the original study (PMID: 31806903, 37257450, 34844191, 35718373), which encompasses comprehensive information about this dataset. For reproducibility and transparency, the sample data and code are publicly available online at <https://github.com/zhangbeibei-min/ProMetaGCN>. Additionally, our project utilizes Python as the programming language, with specific version requirements including Python 3.7.6, sklearn 0.22.1, numpy 1.21.6, scipy 1.5.2, pandas 1.0.1, lightgbm 3.2.0, and xgboost 1.5.2. The operating system used is Windows, and the project is licensed under the MIT License. There are no specified restrictions for non-academic use.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 15 February 2025 / Accepted: 27 March 2025

Published online: 10 April 2025

References

1. Dammer EB, Ping L, Duong DM, Modeste ES, Seyfried NT, Lah JJ, Levey AI, Johnson ECB. Multi-platform proteomic analysis of Alzheimer's disease

- cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. *Alzheimers Res Ther* 2022;14(1):174.
2. Feng E, Balint E, Poznanski SM, Ashkar AA, Loeb M. Aging and interferons: impacts on inflammation and viral disease outcomes. *Cells* 2021;10(3):708.
 3. Alpert A, Pickman Y, Leipold M, Rosenberg-Hasson Y, Ji X, Gaujoux R, Rabani H, Starosvetsky E, Kveler K, Schaffert S, et al. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat Med*. 2019;25:487–95.
 4. Goudsmit J, van den Biggelaar AHJ, Koudstaal W, Hofman A, Koff WC, Schenkberg T, Alter G, Mina MJ, Wu JW. Immune age and biological age as determinants of vaccine responsiveness among elderly populations: the human immunomics initiative research program. *Eur J Epidemiol*. 2021;36:753–62.
 5. Shive C, Pandiyan P. Inflammation, immune senescence, and dysregulated immune regulation in the elderly. *Front Aging* 2022, 3.
 6. Trowbridge JJ, Starczynowski DT. Innate immune pathways and inflammation in hematopoietic aging, clonal hematopoiesis, and MDS. *J Exp Med* 2021;218(7):e20201544.
 7. Sparks R, Rachmaninoff N, Lau WW, Hirsch DC, Bansal N, Martins AJ, Chen J, Liu CC, Cheung F, Failla LE, et al. A unified metric of human immune health. *Nat Med*. 2024;30:2461–72.
 8. Williams SA, Kivimaki M, Langenberg C, Hingorani AD, Casas JP, Bouchard C, Jonasson C, Sarzynski MA, Shipley MJ, Alexander L, et al. Plasma protein patterns as comprehensive indicators of health. *Nat Med*. 2019;25:1851–7.
 9. Altay O, Arif M, Li X, Yang H, Aydin M, Alkurt G, Kim W, Akyol D, Zhang C, Dinler-Doganay G et al. Combined metabolic activators accelerates recovery in Mild-to-Moderate COVID-19. *Adv Sci* 2021;8(17):e2101222.
 10. Argentieri MA, Xiao S, Bennett D, Winchester L, Nevado-Holgado AJ, Ghose U, Albukhari A, Yao P, Mazidi M, Lv J, et al. Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nat Med*. 2024;30:2450–60.
 11. Oh HS-H, Rutledge J, Nachun D, Pálóvics R, Abiose O, Moran-Losada P, Channappa D, Urey DY, Kim K, Sung YJ, et al. Organ aging signatures in the plasma proteome track health and disease. *Nature*. 2023;624:164–72.
 12. Sathyan S, Ayers E, Gao T, Weiss EF, Milman S, Verghese J, Barzilai N. Plasma proteomic profile of age, health span, and all-cause mortality in older adults. *Aging Cell* 2020;19(11):e13250.
 13. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, Candia J, Zhang P, Cheung F, Fantoni G et al. Plasma proteomic signature of age in healthy humans. *Aging Cell* 2018;17(5):e12799.
 14. Zhang M, Zhao C, Cheng Q, Xu J, Xu N, Yu L, Feng W. A score-based method of immune status evaluation for healthy individuals with complete blood cell counts. *BMC Bioinformatics*. 2023;24:467.
 15. Jia Z, Ren Z, Ye D, Li J, Xu Y, Liu H, Meng Z, Yang C, Chen X, Mao X, et al. Immune-Ageing evaluation of peripheral T and NK lymphocyte subsets in Chinese healthy adults. *Phenomics*. 2023;3:360–74.
 16. Lei Y, Takahama Y. XCL1 and XCR1 in the immune system. *Microbes Infect*. 2012;14:262–7.
 17. Tete S, Saggini A, Maccauro G, Rosati M, Conti F, Cianchetti E, Tripodi D, Toniato E, Fulcheri M, Salini V, et al. Interleukin-9 and mast cells. *J Biol Regul Homeost Agents*. 2012;26:319–26.
 18. Delanghe JR, Speeckaert R, Speeckaert MM. Complement C3 and its polymorphism: biological and clinical consequences. *Pathology*. 2014;46:1–10.
 19. Justiz Vaillant AA, Jamal Z, Patel P, Ramphul K. Immunoglobulin. In *StatPearls*. 2024.
 20. Pasricha C, Bansal N, Kaur R, Kumari P, Jangra S, Singh R. Immunoglobulins: mechanistic approaches in moderation of various inflammatory and Anti-Inflammatory pathways. *Curr Pharm Biotechnol* 2024;25:1–20.
 21. Lehallier B, Gate D, Schaum N, Nanasi T, Lee SE, Yousef H, Moran Losada P, Berdnik D, Keller A, Verghese J, et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med*. 2019;25:1843–50.
 22. Wang H, Liu C, Xie X, Niu M, Wang Y, Cheng X, Zhang B, Zhang D, Liu M, Sun R, et al. Multi-omics blood atlas reveals unique features of immune and platelet responses to SARS-CoV-2 Omicron breakthrough infection. *Immunity*. 2023;56:1410–e14281418.
 23. Zhong W, Altay O, Arif M, Edfors F, Doganay L, Mardinoglu A, Uhlen M, Fagerberg L. Next generation plasma proteomic profiling of COVID-19 patients with mild to moderate symptoms. *eBioMedicine* 2021;74:103723.
 24. Harel M, Lahav C, Jacob E, Dahan N, Sela I, Elon Y, Raveh Shoval S, Yahalom G, Kamer I, Zer A et al. Longitudinal plasma proteomic profiling of patients with non-small cell lung cancer undergoing immune checkpoint blockade. *J Immunother Cancer* 2022;10(6):e004582.
 25. Yang R, Dai W, Li C, Zou J, Xiong H. NCGNN: Node-Level capsule graph neural network for semisupervised classification. *IEEE Trans Neural Networks Learn Syst*. 2024;35:1025–39.
 26. Zhang Y, Chu Y, Lin S, Xiong Y, Wei D-Q. ReHoGCNES-MDA: prediction of miRNA-disease associations using homogenous graph convolutional networks based on regular graph with random edge sampler. *Brief Bioinform* 2024;25(2):bbae103.
 27. Finn C, Abbeel P, Levine S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. 2017.
 28. Zhou F, Cao C, Zhang K, Trajcević G, Zhong T, Geng J. Meta-GNN: on Few-shot node classification in graph Meta-learning. in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pages.2357–2360.
 29. Comerford I, Harata-Lee Y, Bunting MD, Gregor C, Kara EE, McColl SR. A myriad of functions and complex regulation of the CCR7/CCL19/CCL21 chemokine axis in the adaptive immune system. *Cytokine Growth Factor Rev*. 2013;24:269–83.
 30. Kang S, Narazaki M, Metwally H, Kishimoto T. Historical overview of the interleukin-6 family cytokine. *J Exp Med* 2020;217(5):jem.2019034704212020c.
 31. Chen X, Cheng X, Zhang S, Wu D. ADAMTS13: an emerging target in stroke therapy. *Front Neurol*. 2019;10:772.
 32. Feng Y, Li X, Xiao J, Li W, Liu J, Zeng X, Chen X, Chen S. ADAMTS13: more than a regulator of thrombosis. *Int J Hematol*. 2016;104:534–9.
 33. Gao Y, Xu Y, Zhao S, Qian L, Song T, Zheng J, Zhang J, Chen B. Growth differentiation factor-15 promotes immune escape of ovarian cancer via targeting CD44 in dendritic cells. *Exp Cell Res* 2021;402(1):112522.
 34. Zhou Z, Li W, Song Y, Wang L, Zhang K, Yang J, Zhang W, Su H, Zhang Y. Growth differentiation Factor-15 suppresses maturation and function of dendritic cells and inhibits Tumor-Specific immune response. *PLoS ONE* 2013;8(11):e78618.
 35. Wang Z, He L, Li W, Xu C, Zhang J, Wang D, Dou K, Zhuang R, Jin B, Zhang W et al. GDF15 induces immunosuppression via CD48 on regulatory T cells in hepatocellular carcinoma. *J Immunother Cancer* 2021;9(9):e002787.
 36. Hou M. Exploring novel independent prognostic biomarkers for hepatocellular carcinoma based on TCGA and GEO databases. *Med (Baltim)*. 2022;101:e31376.
 37. Xu Y, Chen Q, Jiang Y, Liang X, Wang T, Xu Y. UMI-77 modulates the complement cascade pathway and inhibits inflammatory factor storm in sepsis based on TMT proteomics and inflammation array glass chip. *J Proteome Res*. 2023;22:3464–74.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.