



OPEN ACCESS

EDITED BY

Julia Kzhyshkowska,
Heidelberg University, Germany

REVIEWED BY

Cemil Çolak,
İnönü University, Türkiye
Mi He,
Army Medical University, China

*CORRESPONDENCE

Dong-Qing Wei
✉ dqwei@sjtu.edu.cn
Guangfu Ji
✉ cyfjfk@caep.cn

SPECIALTY SECTION

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

RECEIVED 21 June 2022

ACCEPTED 04 January 2023

PUBLISHED 26 January 2023

CITATION

Zhu J, Chen T, Mao X, Fang Y, Sun H,
Wei D-Q and Ji G (2023) Machine learning
of flow cytometry data reveals the delayed
innate immune responses correlate with
the severity of COVID-19.
Front. Immunol. 14:974343.
doi: 10.3389/fimmu.2023.974343

COPYRIGHT

© 2023 Zhu, Chen, Mao, Fang, Sun, Wei and
Ji. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Machine learning of flow cytometry data reveals the delayed innate immune responses correlate with the severity of COVID-19

Jing Zhu¹, Tunan Chen², Xueying Mao³, Yitian Fang³, Heqi Sun³,
Dong-Qing Wei^{3,4*} and Guangfu Ji^{1*}

¹National Key Laboratory for Shock Wave and Detonation Physics Research, Institute of Fluid Physics, Chinese Academy of Engineering Physics, Mianyang, China, ²Department of Neurosurgery and Key Laboratory of Neurotrauma, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing, China, ³State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Cooperation in Metabolic and Developmental, Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, ⁴Peng Cheng Laboratory, Shenzhen, China

Introduction: The COVID-19 pandemic has posed a major burden on healthcare and economic systems across the globe for over 3 years. Even though vaccines are available, the pathogenesis is still unclear. Multiple studies have indicated heterogeneity of immune responses to SARS-CoV-2, and potentially distinct patient immune types that might be related to disease features. However, those conclusions are mainly inferred by comparing the differences of pathological features between moderate and severe patients, some immunological features may be subjectively overlooked.

Methods: In this study, the relevance scores (RS), reflecting which features play a more critical role in the decision-making process, between immunological features and the COVID-19 severity are objectively calculated through neural network, where the input features include the immune cell counts and the activation marker concentrations of particular cell, and these quantified characteristic data are robustly generated by processing flow cytometry data sets containing the peripheral blood information of COVID-19 patients through PhenoGraph algorithm.

Results: Specifically, the RS between immune cell counts and COVID-19 severity with time indicated that the innate immune responses in severe patients are delayed at the early stage, and the continuous decrease of classical monocytes in peripheral blood is significantly associated with the severity of disease. The RS between activation marker concentrations and COVID-19 severity suggested that the down-regulation of IFN- γ in classical monocytes, Treg, CD8 T cells, and the not down-regulation of IL-17a in classical monocytes, Tregs are highly correlated with the occurrence of severe

disease. Finally, a concise dynamic model of immune responses in COVID-19 patients was generalized.

Discussion: These results suggest that the delayed innate immune responses in the early stage, and the abnormal expression of IL-17a and IFN- γ in classical monocytes, Tregs, and CD8 T cells are primarily responsible for the severity of COVID-19.

KEYWORDS

delayed innate immune responses, relevance scores, COVID-19, neural network, dynamics model

1 Introduction

Coronavirus disease 2019 (COVID-19), caused by the novel human pathogen severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a serious disease that has resulted in widespread global morbidity and mortality. Clinical manifestations of COVID-19 are heterogeneous, with about 80% people experiencing asymptomatic or moderate symptom, the other patients develop severe symptom which may progress to acute respiratory distress syndrome (ARDS) (1, 2). Even though vaccines are available, the pathogenesis of COVID-19 is still unclear. To optimally manage the pandemic, there is an urgent need to understand the host immune responses in COVID-19 patients.

High-throughput single-cell technologies such as flow cytometry and mass cytometry, which can measure features on millions of individual cells, are well suited to support studies of the heterogeneity of immune responses and of how immune cells interact with other host cells and pathogens. Identification of host immunological correlated factors for disease severity is one of the most common application of single-cell technologies (3). Innate immune cells like basophils (4), monocytes (5), plasma DCs (4, 6), and NK cells (6) were reported with reduced abundances in peripheral blood of COVID-19 patients, and with greater reductions in individuals with severe disease than those with moderate disease. While other innate immune cells like neutrophils (7, 8), eosinophils (9) have been shown increased abundances in COVID-19 patients, especially severe patients. And the neutrophil-to-lymphocyte ratio (10, 11) was also reported to be associated with severity of illness. What's more, the numbers of SARS-CoV-2 specific B cells were also found increased from 1-3 months (12) after symptom onset, but with abnormal expansion of antibody-secreting cells in severe patients rather than moderate patients (13), which raised the question about the role of B cell responses in COVID-19 patients. Nevertheless, T cell responses in COVID-19 patients are more controversially, there were evidences of terminally differentiated T cells in severe disease (10, 14), other study (15) suggested that CD8 T cells in severe patients might in a hyperactive state by expressing high level of nature killer cell related markers and increased cytotoxicity. It is not clear whether the T cells in severe patients are exhausted or just highly activated. Multiple studies (14, 16–18) have indicated heterogeneity of immune responses

to SARS-CoV-2, and potentially distinct patient immune types that might be related to disease features.

However, those immunological correlated factors for disease severity in previous studies were mainly inferred by comparing the differences of cell counts or bio-marker expression levels between moderate and severe patients, some immunological features may be subjectively overlooked. In this study, the relevance scores (RS) (19–26) between immunological features and the severity of COVID-19 are objectively calculated through neural network, this calculation method belongs to the feature importance explainability approaches (explanation for AI system's decisions) (25), these values reflect which features played a more critical role in the decision-making process. To the best of our knowledge, this is the first time that COVID-19 patients' cytometry data are analyzed by the explainability approach of AI system. Firstly, we collected two publicly available flow cytometry data sets containing peripheral blood information of COVID-19 patients from the Flow Repository website (27). Secondly, we used the PhenoGraph algorithm (28) to robustly cluster these patients' cells into phenotypically distinct subpopulations. Thirdly, we constructed a neural network with these immune cell counts or activation marker concentrations of particular immune cells as the input neurons, the disease severity as the output neuron. Fourthly, we calculated the RS value between input neurons and output neuron through the Layer-wise Relevance Propagation (LRP) algorithm (20), then we compared these RS between immune cell counts and disease severity at different stages, and analyzed the RS between activation marker concentrations and disease severity on particular immune cells. Finally, we generalized a concise dynamic model of immune responses in COVID-19 patients. These results suggested that the delayed innate immune responses in the early stage is primarily responsible for the severity of COVID-19.

2 Methods

2.1 Acquisition of data sets

To understand the host immune responses to SARS-CoV-2 infection, the publicly available individual flow cytometry data sets were selected from the Flow Repository (<http://flowrepository.org/>)

(27) under accession number FR-FCM-Z36F (29) and FR-FCM-Z2KP (30). Detailed information of samples in these data sets can be found in the original research papers and on the Flow Repository website. A total of 145 samples were obtained from these data sets and a summary of these data sets can be found in Table 1. Relevance data sets were identified from the query “COVID-19”. Selection was primarily focus on the integrity of severity categories: health control, mild/moderate and severe, the specific of patient’s illness time, the uniformity of the patient’s condition distribution, and the staining strategy(which could identify the lymphocyte subsets). Due to the different staining strategies of the data sets in Flow Repository, it is infeasible to merge them into a large and consistent data set. Finally, one mass cytometry and one flow cytometry data set were selected from 22 COVID-19 related data sets. The data set FR-FCM-Z36F was collected from a cohort of hospitalized COVID-19 patients and healthy controls to identify dynamic disease-associated changes in circulating immune cell frequency and phenotype, it will be used for calculating the RS between immune cell counts and the severity of COVID-19 in time. The data set FR-FCM-Z2KP was collected to analysis the activation markers produced by PBMC from COVID-19 patients, it will be used for calculating the RS between activation marker concentrations produced by PBMC and the severity of COVID-19.

2.2 Data pre-processing (Clustering by PhenoGraph algorithm)

The PhenoGraph algorithm (28) was used for robustly clustering the peripheral blood cells of COVID-19 patients into phenotypically distinct subpopulations. The algorithm was run on the R-based (31) application. Samples(in.fcs files) were first pre-processed: margin events were filtered out, live single cells were gated (11). Then these cleaned data were used for PhenoGraph training. To address patient specific variability and to understand immune cells dynamics shared between samples, PhenoGraph clusters were merged, and data were transformed with an arcsinh transformation with cofactor 5. The k-nearest neighbor was set to be 30 for data set Z36F, and 100 for data set Z2KP.

After the PhenoGraph clustering, the percentage of peripheral blood cells(cell counts) for each cluster (phenotypically distinct

subpopulations) of patients are recorded(see in S1 File), these data are tensors of floating point numbers distributed between 0 and 1, which can be directly used as the input of neural network. And the expression matrix of each cluster(activation marker concentration of cells) are recorded also(see in S2 File), all data were compressed with an arcsinh transformation with cofactor 5, the missing value of patients is set to 0, these data will be used as the input of neural network as well.

2.3 Relevance scores calculated by neural network

Given a trained neural network that models a scalar-valued prediction score for each target output, and given an input vector, we are interested in computing for a RS quantifying the relevance of input vector to a considered target of interest (25). In other words, we want to analyze which features of input vector are important for the neural network’s decision toward the target.

The RS can be computed by the Layer-wise Relevance Propagation(LRP) algorithm proposed by Bach et al (19), these derivations go from upper-layer neurons to lower-layer neurons. Let z_j be an upper-layer neuron, whose value in the forward pass is computed as $z_j = \sum_i z_i w_{ij} + b_j$ where z_i is one neuron of the lower-layer, and w_{ij} , b_j are the connection weight and biases. The relevance redistribution onto lower-layer neurons z_i is performed in two steps:

Step one, computing relevance messages $R_{i \leftarrow j}$ going from upper-layer neuron z_j to lower-layer neuron z_i .

$$R_{i \leftarrow j} = \frac{z_i \cdot w_{ij} + \frac{\epsilon \cdot \text{sign}(z_j) + \delta \cdot b_j}{N}}{z_j + \epsilon \cdot \text{sign}(z_j)} \cdot R_j \quad (1)$$

(20, 21)

where N is the total number of lower-layer neurons to which z_j is connected, ϵ is a small positive number which serves as a stabilizer, and $\text{sign}(z_j) = (|z_j|_{\geq 0} - |z_j|_{< 0})$ is defined as the sign of z_j . Moreover, δ is a multiplicative factor that is either set to 1.0, in which case the total relevance of all neurons in the same layer is conserved, or else it is set to 0.0, which implies that a part of total relevance is “absorbed” by the biases and that the relevance propagation rule is approximately conservative.

TABLE 1 Publicly available data sets from the Flow repository database included in analysis.

Repository ID	FCS File Numbers	Study Design	Sample Source	Sample Description	Detecting Instrument	The RS calculated by neural network
FR-FCM-Z36F (29)	96	Apply a streamlined CYTOF workflow to characterize whole blood samples to identify dynamic disease-associated changes in circulating immune cell frequency and phenotype.	Human whole blood	Severe: 36 (Early stage: 11 Middle stage: 10 Late stage: 15) Moderate: 48 (Early stage: 28 Middle stage: 13 Late stage: 7) Healthy: 12	Mass cytometry	The RS between immune cell counts and COVID-19 severity.
FR-FCM-Z2KP (30)	49	Analysis of cytokine secretion in PBMC of patients with COVID-19.	PBMC	Severe: 17 Severe_to_Moderate: 6 Moderate: 20 Healthy: 6	BD Symphony	The RS between activation marker concentrations of immune cells and COVID-19 severity.

Since ϵ is a small stabilizer, formula (1) actually equals to

$$R_{i \leftarrow j} = \left(\frac{z_i \cdot w_{ij}}{z_j} + \frac{\delta \cdot b_j}{N \cdot z_j} \right) \cdot R_j \tag{2}$$

In our experiment, we set $\delta=0$ to ignore the effect of b_j to RS. Formula (2) finally reduced to

$$R_{i \leftarrow j} \approx \frac{z_i \cdot w_{ij}}{z_j} \cdot R_j \tag{3}$$

(19, 22)

Step two, computing relevance R_i going from all the neurons in upper-layer to lower-layer neuron z_i .

$$R_i = \sum_j R_{i \leftarrow j} \tag{4}$$

(20, 21)

Taking a regression neural network with two hidden layers as an example, the structure of this neural network is [In, H1, H2, Out] (Figure 1). To calculate $R_{k \leftarrow out}$ going from neuron 'out' in output layer to neuron 'k' in H2 layer. Since there is only one single output neuron z_{out} , its relevance R_{out} is set to its value z_{out} (20). Plug them into formula (3):

$$R_{k \leftarrow out} = z_k \cdot w_{k out} \tag{5}$$

where z_k means neuron 'k' in H2 layer, $w_{k out}$ means the connection weight between neuron 'k' and neuron 'out'.

To calculate the vector \tilde{R}_k going from all the neurons in output layer(the actual number of neurons in output layer is ignored here to unify the representation) to neuron 'k' in H2 layer:

$$\tilde{R}_k = \sum_{out} R_{k \leftarrow out} = z_k \cdot \widetilde{W_{H2 Out}}[k, :] \tag{6}$$

where $\widetilde{W_{H2 Out}}[k, :]$ means the connection vector between neuron k and all the neurons in output layer, its value is obtained by taking the kth row of vector $\widetilde{W_{H2 Out}}$.

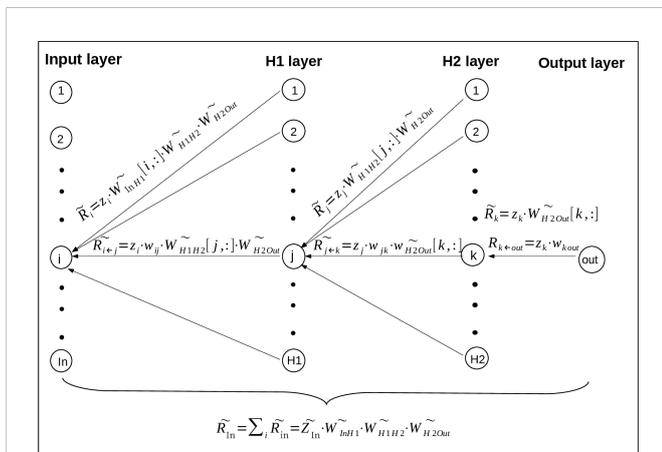


FIGURE 1
The derivation of RS value goes from upper-layer neurons to lower-layer neurons (23) based on LRP algorithm. Where the neural network has two hidden layers H1 and H2, and the output layer has only one neuron. Specifically, 'i', 'j', 'k' and 'out' represent one neuron of the input, H1, H2 and output layer, and the neuron number of these layers are 'In', 'H1', 'H2' and 'Out' respectively. $R_{k \leftarrow out}$ means the relevance going from the neuron out in output layer to neuron k in H2 layer, \tilde{R}_k means the relevance vector going from all the neurons in output layer to neuron k in H2 layer, and so on for the other vectors.

\tilde{R}_j and \tilde{R}_i are obtained by repeating formula (5) and (6) as shown in Figure 1. $\tilde{R}_{In} = \sum_i \tilde{R}_i$, which represents the relevance score going from all the neurons in H1 layer to all the neurons in input layer, is finally expressed as

$$\tilde{R}_{In} = \sum_i \tilde{R}_i = \tilde{Z}_{In} \cdot \widetilde{W_{In H1}} \cdot \widetilde{W_{H1 H2}} \cdot \widetilde{W_{H2 Out}} \tag{7}$$

where \tilde{Z}_{In} means the vector of all the neurons in input layer, $\widetilde{W_{In H1}}$ means the connection vector between Input layer and H1 layer, $\widetilde{W_{H1 H2}}$ means the connection vector between H1 layer and H2 layer, $\widetilde{W_{H2 Out}}$ means the connection vector between H2 layer and Output layer. \tilde{R}_{In} actually contains all the relevance messages from output neurons to input neuron because of the transmission between neurons, it will be used as the RS value between all the input neurons and output neurons.

2.4 Optimization of neural network

To easily calculate the RS corresponding to disease severity, the publicly available free toolbox Pyrenn (32) from Technische Universität München has been used to implement the neural network learning, one neuron was assigned in the output layer to construct a regression model (Figure 1). The targets is from 1 to 3, where 1 is for healthy person, 2 is for moderate patients, and 3 is for severe patients. When calculating the RS between immune cell counts and disease severity, the input vector is the cluster percentage of whole blood cells(red blood cells have been lysed) of patients (S1 File). When calculating the RS between the concentration of activation markers and disease severity, the input vector is the activation marker concentrations(all data were compressed with an arcsinh transformation with cofactor 5) of one cell type of patients (S2 File).

Because the disease severity is not an exact number, but a range of values, the accuracy was set as in Figure S1 (in S3 File) to meet the actual situation of this regressive neural network. For instance, if Yt (the target)=1, when yt (the predictive value)- $Yt < 0.5$, the forecast is deemed accurate, otherwise, it is considered inaccurate, the whole algorithm is shown in Figure S1. The K-fold cross validation was used in the optimization process of neural network, in order to make the validation set contain about 20% of the sample data, the value of k is set to 5. The average accuracy of 20 epoches(20~40 epoches) was used to evaluate the performance of the neural network. The dropout regularization has been used to prevent the neural network from overfitting. The detail of the optimization processes of the neural network are recorded in S3 File.

After the neural network learning, the value of RS will be achieved. For neural network with two hidden layers created by Pyrenn (30), the connection matrix $\widetilde{IW}^{1,1}$, $\widetilde{LW}^{2,1}$ and $\widetilde{LW}^{3,2}$ in Pyrenn are actually the connection weight of input layer to hidden layer 1, hidden layer 1 to hidden layer 2, and hidden layer 2 to output layer respectively. According to formula (7) described in methods, the RS is calculated by the following formula:

$$RS = \widetilde{IW}^{1,1} \cdot \widetilde{LW}^{2,1} \cdot \widetilde{LW}^{3,2} \tag{8}$$

In order to measure the contribution of each input neuron per se to the result, \tilde{Z}_{In} is not included in this formula.

The source code for optimization of neural network and the RS calculation have been uploaded on the website: https://github.com/Zhu-0010/hello_world/branches.

2.5 Welch's t-test

The Welch's t-test was used to compare whether the difference between the two averages of RSs (between active marker expression of cells and severity of COVID-19) between group HC_W and group HC_ICU is significant. This test assumes that both groups of data are sampled from populations that follow a normal distribution, but it does not assume that those two populations have the same variance.

2.6 The pipeline for RS calculation

The pipeline for RS calculation is carried out in the following order (Figure 2): firstly, the flow cytometry data of COVID-19 patients are prepared for PhenoGraph clustering, these preparations include filtering the margin events and gating the live single cells. Secondly, the PhenoGraph algorithm is used to robustly cluster these prepared data into phenotypically distinct subpopulations for each patient, and

generates two files which will be used as the input data for neural network learning: the 'Cluster_Percentage with group' file and the 'PhenoGraphX_Acsinh_Expr' file, the former mainly contains the information of the immune cell (subpopulations) counts of each patient, the later mainly contains the activation marker concentrations on particular cells of each patient. Thirdly, these two files are transformed into the data format that is suitable for neural network learning. The main task is to determine the input and output neurons, where the data of cell counts or the activation marker concentrations are used as the input neurons, the disease severity are used as the output neurons. Fourthly, depending on the different kinds of input neurons, two neural networks are constructed. The RS between the input neurons and output neurons will be calculated according to the method described in subsection 4.2.

3 Results

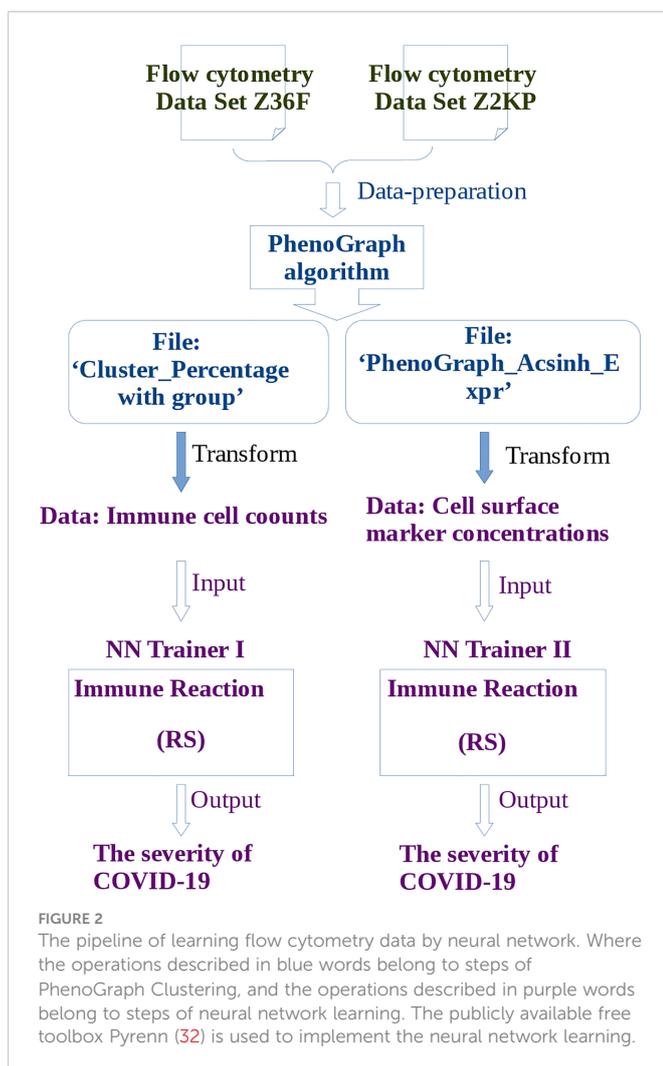
3.1 Clustering by PhenoGraph algorithm

To investigate the phenotype of immune cells in COVID-19 patients, the PhenoGraph algorithm (28) was used for robustly partitioning the two flow cytometry data set (Table 1) of COVID-19 patients into phenotypically distinct clusters. The algorithm identified 38 clusters for each patient in data set Z36F (Figure 3A), these clusters mainly included I B cells, plasmacytoid dendritic cELLS(pDC), basophils, plasma B cells, CD16 low NK cells, CD57 high memory CD4 T cells, CD57 high CD8 TEMInaive CD8 TIs, naive CD4 T cells, CXCR3+ CCR6- memory CD4 T cells, $\gamma\delta$ T cells, CD161+ effector memory CD8 T cells, effector memory CD8 T cells, neutrophils, inducible eosinophils(iEos) (33), resident eosinophils(rEos), classical monocytes and non-classical monocytes. And the cell counts of these clusters for each patient have been gotten as well (S1 File), which will be used as the input data for deep learning of RS between immune cell counts and the severity of COVID-19, where the cell counts of immune cells will be used as the input neurons, and the severity of the patients will be used as the output neurons.

The PhenoGraph algorithm also identified 52 clusters for each patient in data set Z2KP (Figure 3B), the subtypes of these clusters included B cells, classical monocytes, CD8 T cells, conventional CD4 T cells (Tconv), regulatory CD4 T cells (Treg), double-negative T cells (DNT). And the activation marker concentrations of these clusters for each patient have been gotten as well (the activation markers' information of cluster 1 are shown in S2 File), specifically, these activation markers are IL-17a, IL-2, GATA3, IFN- γ , IL-4, IL-10, PD_1, CD40L, ROR γ t, CTLA_4, CCR7, TNF- α and IL-6. These files will be used as the input data for deep learning of RS between activation marker concentrations and the severity of COVID-19, where the activation marker concentrations will be used as the input neurons, and the severity of patients will be used as the output neurons. These clusters' cell types were defined by manual gating as described in Daniel et al. (34).

3.2 The RS between immune cell counts and the severity of COVID-19 at different stages

Patients at different stages (in data set Z36F) and different severities are divided into six groups: health control and moderate patients (HC_W)/health control and severe patients(HC_ICU) at the early



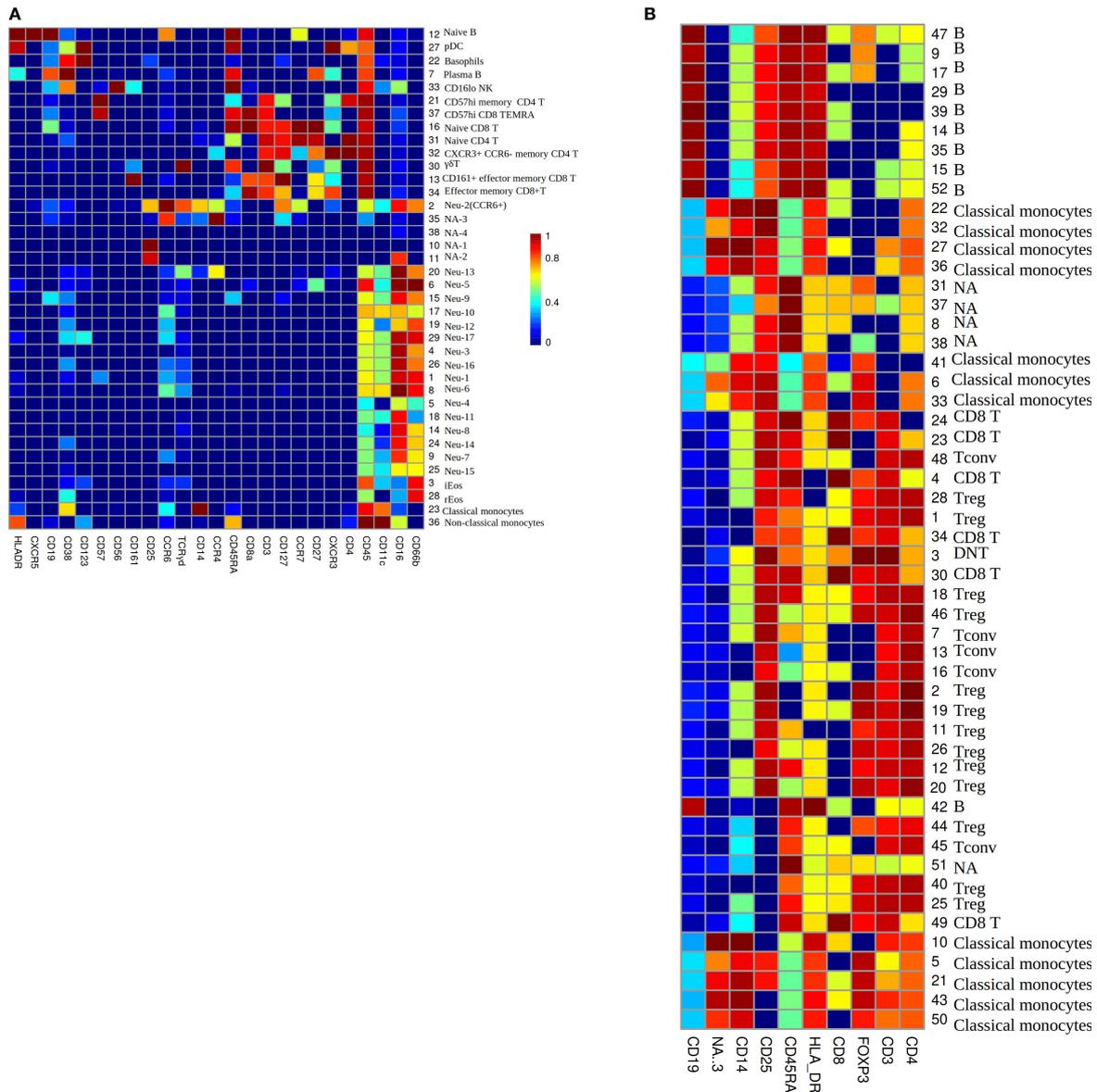


FIGURE 3 (A) The heatmap of median lineage marker expressed on automatically gated immune clusters of data set Z36F. (B) The heatmap of median lineage markers expressed on automatically gated immune clusters of data set Z2KP. The horizontal axis represent the types of markers. The vertical axis show the ID and manually gated cell type for each cluster, NA means there is no cell type matches this cluster by comparing the expression heatmap. The colors on the Mosaics from red to blue indicate the strongest to weakest marker intensity, the intensity of each kind of marker is normalized to 0-1.

stage (day 1 since symptom onset), HC_W/HC_ICU at the middle stage (day 4), HC_W/HC_ICU at the late stage (day 7-12). The RS between immune cell counts of peripheral blood and severity of COVID-19 are calculated separately at these six stages. The hyperparameters of neural network are optimized the way in subsection 2.3, the network structure $[n_{in}, 5, 4, 1]$ with dropout rate = 0.1 was used for HC_W groups, and $[n_{in}, 2, 2, 1]$ with dropout rate = 0.2 was used for HC_ICU groups. The RS value is calculated through formula (8). The results are shown in Figure 4, a positive value of RS means that the cell counts have a positive correlation to the severity of COVID-19, the higher the RS value, the stronger the correlation. Relatively, a negative value of RS indicates that the cell counts have a negative correlation to the disease severity, the higher the absolute value, the stronger the negative correlation. What's

more, a RS value close to 0 is considered neutral, which means the cell counts has little influence on the severity of COVID-19 (20).

16 most correlated cell types (including 8 positive and 8 negative cell types) for each stage are recorded in descending order in Figure 4. We can get four aspects of information from this picture. Firstly, for moderate patients, the cell counts of immune cells in peripheral blood may increase or decrease during disease, it indicates that the decrease of immune cells in peripheral blood does not prevent the recovery of illness, it is more like a regular pathological process of COVID-19. Secondly, with the development of disease, the cell types work with positive/negative RS show the tendency of inheritance and development. For instance, in the early stage, CCR6+ neutrophils, classical monocytes, plasma B, CD16 lowIcells, naive CD8 T cells,

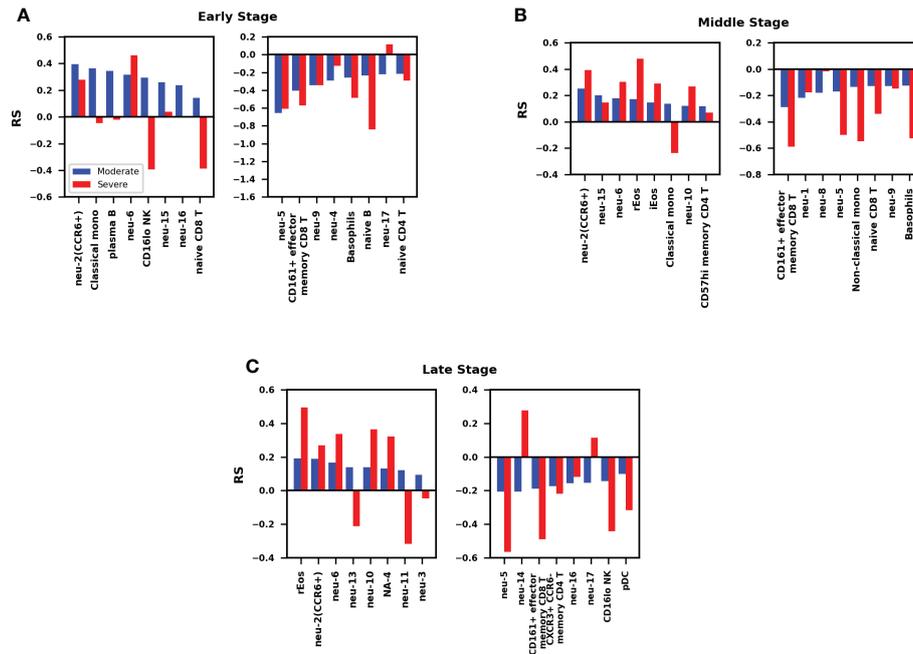


FIGURE 4

(A) Comparison of RS between moderate and severe patients at early stage. The RS represent the relevance score between cell counts and the severity of COVID-19. The mean accuracy in moderate patients is 69.5%, in severe patients is 61.2%. (B) Comparison of RS between moderate and severe patients at middle stage. The mean accuracy in moderate patients is 72.3%, in severe patients is 67.0%. (C) Comparison of RS between moderate and severe patients at late stage. The mean accuracy in moderate patients is 70.1%, in severe patients is 66.6%.

neu-6, neu-15, and neu-16 show positive RS, in the middle stage, CCR6+ neutrophils, classical monocytes, neu-6, neu-15 still have positive RS with the severity of illness, while rEos, iEos, neu-10, and CD57 high memory CD4 T cells become outstanding, in the late stage, rEos, CCR6+ neutrophils, neu-6, neu-10 continue the positive relationship, the other neutrophils begin to work. Thirdly, as the disease progresses, immune cells with a positive RS value and those with a negative value rarely appear in the opposite camp, that suggests the orderliness of the immune response in moderate patients.

Fourthly and importantly, the most significant difference of RS between moderate and severe patients occurred at the early stage (Figure 4A). Cells that are positively correlated with the severity of COVID-19 in moderate patients but have no remarkable correlation in severe patients, they are classical monocytes, plasma B, and some subtypes of neutrophils, some even have negative correlation in severe patients, they are CI NK and naive CD8 T cells. The minor differences happened at the middle and late stage (Figures 4B, C), when the function of immune cells in moderate patients tended to be gentle, most of these cells in severe patients remained at a high level. In addition to the overall differences by stages, the classical monocyte, which increase significantly in moderate patients in the early and middle stage, but remained at a low level or even decreases in the severe patients.

3.3 The RS between activation marker concentrations and the severity of COVID-19

The RS between activation marker concentrations and the severity of COVID-19 are calculated by training data set Z2KP,

where the activation markers are IL-17a, IL-2, GATA3, IFN- γ , IL-4, IL-10, PD_1, CD40L, ROR γ t, CTLA_4, CCR7, TNF- α and IL-6, respectively. The hyperparameters of neural network are optimized the way in subsection 2.3, the network structure [n_{in}, 11, 1] was used for both HC_W and HC_ICU groups. The RS value is calculated through formula (8) as well. In Figure 5, each data point represent the RS of a subtype of one kind of immune cells, these immune cells are classical monocyte, Treg, Tconv, CD8 T and B cells. And the Welch's t-test was used to determine whether the difference between the two averages of HC_W group and HC_ICU group is significant.

The information reflected in Figure 5 can be interpreted in three ways. We hypothesize that the immune response in moderate patients is normal pathological and that in severe patients is abnormal pathological. Therefore, firstly, we focused on the expression of activation marker on immune cells in moderate patients. For classical monocytes, IL_4 and CTLA_4 are markedly positively correlated with the development of disease, while IL_17a shows a negative correlation. For Treg, IL_4 is positively correlated with this disease, IL_17a is negatively correlated. For both Tconv and CD8 T cells, IL_4 still shows a remarkable positive correlation. For B cells, IL_4 continues its significant positive correlation with disease, while IF_2 and TNF α show negative correlation with this disease. It is obvious that IL_4 is up-regulated in all cells, IL_17a, IF_2 and TNF α are down-regulated depending on cell types.

Secondly, we focus on the expression of activation markers that are significantly associated with moderate patients but not severe patients. In severe patients, IL_17a is not down-regulated on classical monocytes; none of the IL_17a and IL_4 is down or up regulated on Treg; and there are no obvious expression difference on Tconv, CD8 T, and B cells. These phenomenon suggest that the not down-

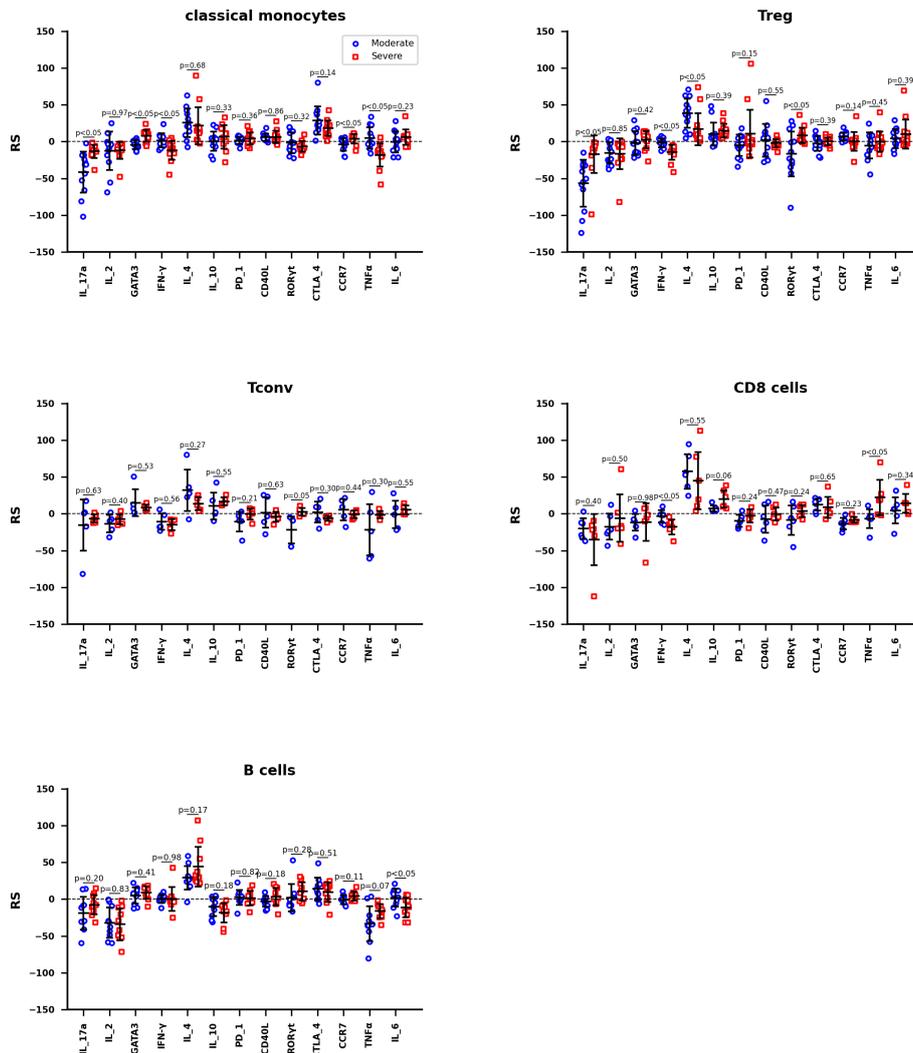


FIGURE 5
The RS between activation marker concentrations and the severity of COVID-19 of Classical monocytes, Treg, Tconv, CD8 T cells, and B cells, respectively. The mean accuracy of these neutral networks is 73.4% for moderate patients and 81.5% for severe patients. Differences are tested using Welch's t-test.

regulation of IL_17a on classical monocytes and Treg are more associated with the occurrence of severe disease.

Thirdly, we focus on the expression of activation markers that are significantly associated with severe patients but not moderate patients. In severe patients, IFN- γ and TNF α are significantly down-regulated on classical monocytes; IFN- γ is down-regulated on Treg; IFN- γ is down-regulated but TNF α is up-regulated on CD8 T cells. These phenomenon imply that the down-regulation of IFN- γ on classical monocytes, Treg, and CD8 T cells are highly correlated with the occurrence of severe disease.

3.4 Dynamics of immune response in COVID-19 patients

The dynamics of immune response in COVID-19 patients are summarized as a concise mode in Figure 6. This model suggests that in moderate patients (Figure 6A) the innate immune responses are rapidly activated on a large scale at the early stage, they occurred

within one day since symptom onset. Then the adaptive immune responses are primed by the innate immune responses (35), it takes about several days (36) to generate enough virus-specific immune cells. Subsequently, the innate immune responses are down regulated after an early peak, and then these responses slowly decline and continue into late stage of disease. While in severe patients (Figure 6B), the innate immune responses are delayed till the middle stage of disease, and this leads to the delayed priming of adaptive immune responses as well. Once activated, the innate immunity remains highly activity(compared to moderate patients at the same time) till the late stage of disease. Then the adaptive responses are activated as well.

4 Discussion

This study calculated the RS between immune cell counts and the severity of COVID-19, and the RS between activation marker (transcription factors and cytokines) concentrations of immune

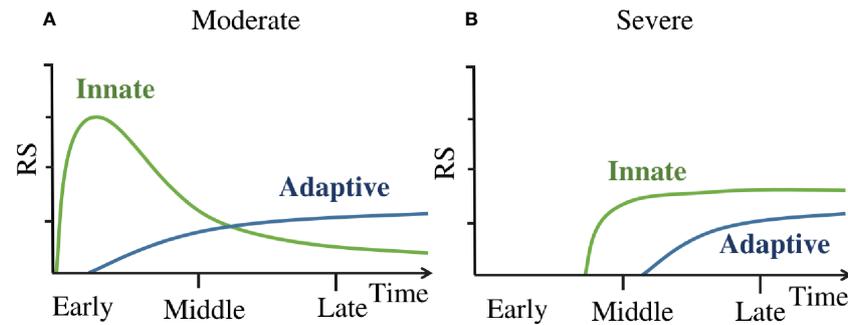


FIGURE 6

(A) The dynamics of immune response in moderate patients. (B) The dynamics of immune response in severe patients. 'Innate' line refers to the kinetics of innate immune cell counts detectable in peripheral blood and the activation marker concentrations of these cells. 'Adaptive' line refer to the kinetics of adaptive immune cell counts detectable in peripheral blood and the activation marker concentrations of these cells.

cells and the severity of COVID-19. By comparing the RS of immune cells counts to the severity of COVID-19 between moderate and severe patients at different stages, we found that the innate immune responses in severe patients are delayed till the middle stage of disease, and this leads to the delayed priming of adaptive immune responses as well. The dynamics of immune response of COVID-19 found in our work are in consistent with the model reviewed by Alessandro Sette et al. (35) This model suggested that the immune evasion of SARS-CoV-2 makes it evade the triggering of early innate immune responses in severe patients, and this delay in innate immune responses is correlate with the severity of illness by failing to prime an adaptive immune response, what's worse, the innate immune system (especially the neutrophils) tries to fill the vacuum left by absence of adaptive immune responses in the late stage, this finally result in excessive lung immunopathology. While our result show that not only the proliferation of a large number of neutrophils in the peripheral blood, but also the reduction of classical monocytes are significantly correlated with severe illness of COVID-19, this is consistent with other works (37, 38), both the apoptosis of classical monocytes in the circulatory system and migration to tissues are thought to influence the reduction of cell counts in the peripheral blood.

In addition, Daniel K. Beyer et al. (39) have reviewed that the delayed type I and type III IFN responses are associated with risk of severe COVID-19, and SARS-CoV-2 is thought to be effective at evading the triggering of early innate immune responses. And this delayed innate immune response subsequently failed to prime an adaptive immune response, the study of Carolina Lucas et al. (40) indicated that COVID-19 mortality did not correlated with the cross-sectional antiviral antibody levels per se but, rather, with the delayed kinetics of neutralizing antibody (NAb) production. What's more, advanced age has been widely recognized as a significant factor associated with severe disease, multi-omics profiling (41) suggests that age may delay or impair antiviral cellular immune responses and delay efficient return to immune homeostasis. The above data linked the general evidence for the pathogenesis of SARS-CoV-2, for instance, the target: type I and type III IFN; the mechanism: delaying the innate immune responses; the immunological characteristics: the delayed innate immune cell counts and the delayed kinetics of NAb production; and the clinical characteristics: advanced age.

Apart from the dynamics of immune response clarified by our study, there are also some details worth noting. Firstly, the negative value of RS in Figure 4 means that the corresponding cell counts have a negative correlation with the severity of COVID-19, these correlations corresponding to lymphopenia. An apoptosis and migration scoring system studied by Ji-Yuan Zhang et al. (42) suggested that cell death and lymphocyte migration (into infected site or adhesion to inflamed vascular endothelium) may be both associate with lymphopenia. So in our work, this negative value of RS mainly represent the decrease of the corresponding cells in the peripheral blood, the specific direction (apoptosis or migration) of these cells is till unknown.

Secondly, the lymphopenia is not only occurred in severe patients, but also occurred in moderate patients (Figure 4). In addition, lymphopenia is also occurred in patients with respiratory viral infections, such as the A/H3N2 virus, the human rhinovirus (HRV) and respiratory syncytial virus (RSV) (43). The above phenomenon imply that the lymphopenia is not the cause of severe illness.

Thirdly, it can be seen from Figure 4 that the neutrophil subsets show strong correlation at different disease stages, but have different modes of action, including positive or negative correlations. This is not surprising because different neutrophil subsets act in heterogeneous manners have already been reviewed (44), the reactive oxygen species (ROS) and neutrophil extracellular traps (NETs) produced by neutrophils are thought to contribute to cell death (45), and neutrophils have been characterized in the lungs and tracheal aspirates of COVID-19 patients (8).

Deep learning of disease characteristics are often limited by the sample size and the complexity of patient's own physical condition, its accuracy is usually not high (46–48). Our research is affected by the same factors, generally speaking, there are two main reasons that limited the accuracy of this study, one is the regression neural network model, the other is the sample size. Generally, the accuracy of classifier is higher than regression neural network, but one neuron was set in the output layer for facilitating the RS calculation, this made regressive neural network an appropriate choice for our work, and led to a partial sacrifice of accuracy. This work totally analyzed 145 cases, although it is sufficient for clinical analysis of diseases, it is indeed a small sample for neural network. Flow cytometry data set from the Flow Repository website often contains only dozens to hundreds of cases. Meanwhile, different cell staining strategies limit the possibility of merging these data sets to study homogeneity, they greatly limit the

size of sample collection. Therefore, further validation of our results is warranted when additional data are released, as well as immunological data on infected sites, to help accurately interpreting the biological significance of negative RS.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

TC assist in the analysis of immunological parameters. XM, YF and HS help validate machine learning results. D-QW assist in reviewing this article. GJ provide experimental guidance. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by grants from the National Nature Science Foundation of China (Grant No. 82173388, 32070662, 61832019, 32030063), the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the Science and Technology Commission of Shanghai Municipality (Grant No.: 19430750600), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2021ZD02).

References

- Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in wuhan, China. *JAMA* (2020) 323:1061. doi: 10.1001/jama.2020.1585
- Torres Acosta MA, Singer BD. Pathogenesis of COVID-19-induced ARDS: Implications for an ageing population. *Eur Respir J* (2020) 56:2002049. doi: 10.1183/13993003.02049-2020
- Tian Y, Carpp LN, Miller HER, Zager M, Newell EW, Gottardo R. Single-cell immunology of SARS-CoV-2 infection. *Nat Biotechnol* (2022) 40:30–41. doi: 10.1038/s41587-021-01131-y
- Laing AG. A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat Med* (2020) 26:1623–35. doi: 10.1038/s41591-020-1038-6
- Knoll R, Schultze JL, Schulte-Schrepping J. Monocytes and macrophages in COVID-19. *Front Immunol* (2021) 12:720109. doi: 10.3389/fimmu.2021.720109
- Zhou R, To KK-W, Wong Y-C, Liu L, Zhou B, Li X, et al. Acute SARS-CoV-2 infection impairs dendritic cell and T cell responses. *Immunity* (2020) 53:864. doi: 10.1016/j.immuni.2020.07.026
- Meizlish ML, Pine AB, Bishai JD, Goshua G, Nadelmann ER, Simonov M, et al. A neutrophil activation signature predicts critical illness and mortality in COVID-19. *Blood Adv* (2021) 5:1164–77. doi: 10.1182/bloodadvances.2020003568
- Veras FP, Pontelli MC, Silva CM, Toller-Kawahisa JE, de Lima M, Nascimento DC, et al. SARS-CoV-2-triggered neutrophil extracellular traps mediate COVID-19 pathology. *J Exp Med* (2020) 217:e20201129. doi: 10.1084/jem.20201129
- Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol* (2020) 38:970–9. doi: 10.1038/s41587-020-0602-4
- Kuri-Cervantes L, Pampena MB, Meng W, Rosenfeld AM, Ittner CAG, Weisman AR, et al. Comprehensive mapping of immune perturbations associated with severe COVID-19. *Sci Immunol* (2020) 5:eabd7114. doi: 10.1126/sciimmunol.abd7114
- The CONTAGIOUS consortium, Penttilä PA, Van Gassen S, Panovska D, Vanderbeke L, Van Herck Y, et al. High dimensional profiling identifies specific immune types along the recovery trajectories of critically ill COVID19 patients. *Cell Mol Life Sci* (2021) 78:3987–4002. doi: 10.1007/s00018-021-03808-8

Acknowledgments

The computations were partially performed at the Pengcheng Lab. and the Center for High-Performance Computing, Shanghai Jiao Tong University.

Conflict of interest

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

The reviewer MH declared a shared parent affiliation with the author TC to the handling editor at the time of review.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.974343/full#supplementary-material>

S1 FILE

Cluster Percentage. The cluster counts(%) for each patient robustly clustered by PhenoGraph algorithm.

S2 FILE

PhenoGraph1_Acsinh_Expr. The activation marker concentrations(simpleAsinh: asinh(x/5)) on PhenoGraph1(Cluster1) of each patient robustly clustered by PhenoGraph algorithm.

S3 FILE

Optimization of the neutral network.

12. Rodda LB, Netland J, Shehata L, Pruner KB, Morawski PA, Thouvenel CD, et al. Functional SARS-CoV-2-Specific immune memory persists after mild COVID-19. *Cell* (2021) 184:169–183.e17. doi: 10.1016/j.cell.2020.11.029
13. Woodruff MC. Extrafollicular b cell responses correlate with neutralizing antibodies and morbidity in COVID-19. *Nat Immunol* (2020) 21:1506–16. doi: 10.1038/s41590-020-00814-z
14. Mathew D, Giles JR, Baxter AE, Oldridge DA, Greenplate AR, Wu JE, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science* (2020) 369:eabc8511. doi: 10.1126/science.abc8511
15. Yu K, He J, Wu Y, Xie B, Liu X, Wei B, et al. Dysregulated adaptive immune response contributes to severe COVID-19. *Cell Res* (2020) 30:814–6. doi: 10.1038/s41422-020-0391-9
16. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, Akinosoglou K, Antoniadou A, Antonakos N, et al. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe* (2020) 27:992–1000.e3. doi: 10.1016/j.chom.2020.04.009
17. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Strålin K, Gorin J-B, Olsson A, et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* (2020) 183:158–168.e14. doi: 10.1016/j.cell.2020.08.017
18. Coleman M, Zimmerly K, Yang X. Accumulation of CD28null senescent T-cells is associated with poorer outcomes in COVID19 patients. *Biomolecules* (2021) 11:1425. doi: 10.3390/biom11101425
19. Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* (2015) 10:e0130140. doi: 10.1371/journal.pone.0130140
20. Arras L, Montavon G, Müller K-R, Samek W. Explaining recurrent neural network predictions in sentiment analysis. *ArXiv* (2017) 159–68. doi: 10.18653/v1/W17-5221
21. Arras L, Horn F, Montavon G, Müller K-R, Samek W. Explaining predictions of non-linear classifiers in NLP. *ArXiv Preprint* (2016) 1–7. doi: 10.18653/v1/W16-1601
22. Arras L, Horn F, Montavon G, Müller K-R, Samek W. “What is relevant in a text document?”: An interpretable machine learning approach. *PLoS One* (2017) 12:e0181142. doi: 10.1371/journal.pone.0181142
23. Lapuschkin S, Binder A, Montavon G, Müller K-R, Samek W. The LRP toolbox for artificial neural networks. *J Mach Learn Res* (2016) 17:1–5. Available at: <https://jmlr.org/papers/v17/15-618.html>.
24. Arras L, Osman A, Mueller K-R, Samek W. Evaluating recurrent neural network explanations. In: *Blackboxnlp workshop on analyzing and interpreting neural networks for nlp at acl 2019*. Stroudsburg: Assoc Computational Linguistics-Acl (2019). p. 113–26.
25. Kaur D, Uslu S, Rittichier KJ, Durresi A. Trustworthy artificial intelligence: A review. *ACM Comput Surv* (2023) 55:1–38. doi: 10.1145/3491209
26. Aketi SA, Roy S, Raghunathan A, Roy K. Gradual channel pruning while training using feature relevance scores for convolutional neural networks. *IEEE Access* (2020) 8:171924–32. doi: 10.1109/ACCESS.2020.3024992
27. Spidlen J, Breuer K, Rosenberg C, Kotecha N, Brinkman RR. FlowRepository: A resource of annotated flow cytometry datasets associated with peer-reviewed publications. *Cytometry* (2012) 81A:727–31. doi: 10.1002/cyto.a.22106
28. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir ED, Tadmor MD, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* (2015) 162:184–97. doi: 10.1016/j.cell.2015.05.047
29. Geanon D, Lee B, Gonzalez-Kozlova E, Kelly G, Handler D, Upadhyaya B, et al. A streamlined whole blood CyTOF workflow defines a circulating immune cell signature of COVID-19. *Cytometry* (2021) 99:446–61. doi: 10.1002/cyto.a.24317
30. Neumann J, Prezzemolo T, Vanderbeke L, Roca CP, Gerbaux M, Janssens S, et al. Increased IL-10-producing regulatory T cells are characteristic of severe cases of COVID-19. *Clin Transl Immunol* (2020) 9:e1204. doi: 10.1002/cti2.1204
31. Team RC. R: A language and environment for statistical computing. *MSOR Connections* (2014) 1.
32. Atabay D. *Pyrenn: First release* (2015). doi: 10.5281/ZENODO.45022.
33. Kanda A, Yun Y, Bui DV, Nguyen LM, Kobayashi Y, Suzuki K, et al. The multiple functions and subpopulations of eosinophils in tissues under steady-state and pathological conditions. *Allergy Int* (2021) 70:9–18. doi: 10.1016/j.alit.2020.11.001
34. Geanon D, Lee B, Kelly G, Handler D, Upadhyaya B, Leech J, et al. A streamlined CyTOF workflow to facilitate standardized multi-site immune profiling of COVID-19 patients. *Allergy Immunol* (2020). doi: 10.1101/2020.06.26.20141341
35. Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* (2021) 184:861–80. doi: 10.1016/j.cell.2021.01.007
36. Palmer EM, Holbrook BC, Arimilli S, Parks GD, Alexander-Miller MA. IFN γ -producing, virus-specific CD8+ effector cells acquire the ability to produce IL-10 as a result of entry into the infected lung environment. *Virology* (2010) 404:225–30. doi: 10.1016/j.virol.2010.05.004
37. Romão PR, Teixeira PC, Schipper L, da Silva I, Santana Filho P, Júnior LCR, et al. Viral load is associated with mitochondrial dysfunction and altered monocyte phenotype in acute severe SARS-CoV-2 infection. *Int Immunopharmacol* (2022) 108:108697. doi: 10.1016/j.intimp.2022.108697
38. Wang C, Yu R, Zhang S, Zhao Y, Qi C, Zhu Z, et al. Genome-wide mendelian randomization and single-cell RNA sequencing analyses identify the causal effects of COVID-19 on 41 cytokines. *Briefings Funct Genomics* (2022) 21:423–32. doi: 10.1093/bfgp/elac033
39. Beyer DK, Forero A. Mechanisms of antiviral immune evasion of SARS-CoV-2. *J Mol Biol* (2022) 434:167265. doi: 10.1016/j.jmb.2021.167265
40. Lucas C, Klein J, Sundaram ME, Liu F, Iwasaki A. Delayed production of neutralizing antibodies correlates with fatal COVID-19. *Nat Med* (2021) 27:1178–86. doi: 10.1038/s41591-021-01355-0
41. Speranza E, Purushotham JN, Port JR, Schwarz B, Flagg M, Williamson BN, et al. Age-related differences in immune dynamics during SARS-CoV-2 infection in rhesus macaques. *Life Sci Alliance* (2022) 5:e202101314. doi: 10.26508/lsa.202101314
42. Zhang J-Y, Wang X-M, Xing X, Xu Z, Zhang C, Song J-W, et al. Single-cell landscape of immunological responses in patients with COVID-19. *Nat Immunol* (2020) 21:1107–18. doi: 10.1038/s41590-020-0762-x
43. McClain MT, Park LP, Nicholson B, Veldman T, Zaas AK, Turner R, et al. Longitudinal analysis of leukocyte differentials in peripheral blood of patients with acute respiratory viral infections. *J Clin Virol* (2013) 58:689–95. doi: 10.1016/j.jcv.2013.09.015
44. Hirschfeld J. Neutrophil subsets in periodontal health and disease: A mini review. *Front Immunol* (2020) 10:3001. doi: 10.3389/fimmu.2019.03001
45. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. *Science* (2004) 303:1532–5. doi: 10.1126/science.1092385
46. Zhang S, Poon SK, Vuong K, Sneddon A, Loy CT. A deep learning-based approach for gait analysis in huntington disease. In: *MEDINFO 2019: Health and wellbeing e-networks for all*. IOS Press (2019). 264:477–81. doi: 10.3233/SHTI190267
47. Kumamaru KK, Fujimoto S, Otsuka Y, Kawasaki T, Kawaguchi Y, Kato E, et al. Diagnostic accuracy of 3D deep-learning-based fully automated estimation of patient-level minimum fractional flow reserve from coronary computed tomography angiography. *Eur Heart J - Cardiovasc Imaging* (2019), 21:437–45. doi: 10.1093/ehjci/jez160
48. Lee SJ, Rho M. Multimodal deep learning applied to classify healthy and disease states of human microbiome. *Sci Rep* (2022) 12:824. doi: 10.1038/s41598-022-04773-3