Cytokine production profiles can predict COVID-19 severity

Hiroshi Morimoto

Abstract—**Mortality in COVID-19 patients is related to the presence of a "cytokine storm" induced by the virus. Most patients developed mild symptoms, whereas some patients develop severe disease. Predicting the course of disease is necessary to mitigate or prevent COVID-19 disease severity. Carefully monitoring specific cytokines during the management of COVID-19 patients might improve patients' survival rates and reduce mortality from COVID-19. For example, IL-6 levels in patients with COVID-19 had been considered a relevant parameter in predicting the most severe course of the disease.**

The purpose of this study is to investigate whether a patient's cytokine levels would predict the course of disease, and to describe the characteristic differences in cytokine levels between patients with no symptoms and those with severe disease.

We applied a probabilistic method, naive Bayes classifier, to RNA-sequencing data extracted from GEO with the accession number GSE178967. We predicted a patient's disease course, i.e. either deterioration or improvement, and calculated the comprehensive accuracy of our prediction.

There were characteristic cytokine level patterns preceding a severe state of disease. Some important cytokines were identified other than IL-6 and IL-17, which are already known as key cytokines associated with a cytokine storm.

Our methodology shows that the systematic observation of cytokine levels in patients with COVID-19 can yield important information in predicting the most severe course of disease and thus the need for appropriate and intensive care.

***Index Terms*—COVID-19, SARS-CoV-2, cytokine storms, IL-6, IL-17, IL-16.**

I. Introduction

COVID-19 continuously causes a serious threat to public health, thus there is an urgent need to identify biomarkers for disease severity. Excessive production of inflammatory cytokines might aggravate ARDS (acute respiratory distress syndrome), resulting in widespread tissue damage and multi-organ failure. If we can predict the course of disease among patients with COVID-19, appropriate intensive care may improve survival rates.

Network of cytokines of COVID-19 patients is complex [1], but could be a relevant parameter in predicting most severe course of disease. For example, the increasing IL-6 levels in patients with COVID-19 was considered as a parameter in predicting most severe course of disease [2]. More systematic observation of cytokine levels during the management of COVID-19 patients may contribute to improve survival rates and reduce mortality.

We executed a Naive Bayes classification method to the RNA-sequencing data extracted from GEO with the accession number GSE178967. In the data of GSE178967, most patients with SARS-CoV-2 infection developed mild symptoms, whereas some patients later developed severe disease symptoms. We performed a comprehensive use of Naive Bayes classification method to these data. Since the number of samples from patients with severe disease was limited, our method was executed through the random and comprehensive use of samples. The method selected enabled us to predict whether a patient would deteriorate or improve. We calculated the accuracy of prediction of the severe course of disease.

We then described characteristic features of cytokine production levels for each severity level, and found some patterns of cytokine levels led to a severe state of disease. Although IL-6 and IL-17 had been linked to cytokine storm, we identified other cytokines as key factors that could distinguish the difference of disease severity states.

Therefore, our method could predict whether a patient would get severe disease or not. The systematic observation of cytokine levels in patients with COVID-19 may yield important information in predicting the most severe course of disease, and thus facilitate availing the most appropriate care.

II. Methods

The RNA-sequencing data was extracted from GEO under the accession number GSE178967. The data was supplied by [3]. They performed a randomized clinical trial of Pegylated PegInterferon Lambda for the treatment of SARS-CoV-2 infected patients conducted in the Stanford COVID-19 CTRU. They recruited a total of 108 patients between the ages of 18 to 75 years who tested PCR positive for SARS-CoV-2. The patients were treated and nursed at their first visit (called "day 0"). The study team conducted follow-up visits on Day 1, 3, 5, 7, 10, 14, 21, and 28, where they also collected oropharyngeal swabs for SARS-CoV-2 testing. RNA-sequencing assays were done using blood samples collected on day 0 and day 5 following enrollment. The day 0 and day 5 RNA expression data was listed in CEO.

We applied a naive Bayes classifier to predict whether a patient infected with COVID-19 will take a course to a severe state or not. A naive Bayes classifier is originally a method for classifying or distinguishing several categories. It is a simple probabilistic method based on the application of Bayes' theorem with an independence assumption of variables [4, 5, 1,6].

Let $X=(X_1, \ldots, X_n)$ be an array of random variables denoting observed attribute values (e.g. levels of expression of cytokines) and C be a random variable denoting the class of an instance (e.g. severe, moderate, asymptomatic). Thus C is a particular class label and $X = (X_1,\ldots,X_n)$ is a particular

Hiroshi Morimoto, Professor emeritus of Nagoya University, Japan (e:mail: h.morimoto@nagoya-u.jp)

observed attribute value. To simplify the calculation problem, the following independence condition is normally assumed: the attributes $X_1, \ldots, X_n$ are all conditionally independent of each other given C. This assumption dramatically simplifies the representation of the conditional probability P(X | C) and the problem of estimating it from the training data.

Given a vector of observed values for the predictive attributes $X=(X_1, \ldots, X_n)$, the probability P(C | X) of a class C can be calculated using the Bayes' theorem:

$$P(C|X) = \frac{P(C)P(X|C)}{P(X)}$$

As the attributes are assumed conditionally independent, the following equation holds:

$$P(X|C) = \prod_i P(X_i|C)$$

Therefore, we obtain the following:

$$P(C|X_1, \cdots, X_n) = \frac{P(C)P(X_1|C) \cdots P(X_n|C)}{P(X_1, \cdots, X_n)}$$

In this work, we set X as an array of the expression levels of cytokines. The variable C is defined as one of three classes of "severe", "moderate", and "asymptomatic". The above equation means that if we are given the array of observed values of expressions of cytokines, then the probability of having severe (or moderate, asymptomatic) disease can be predicted by the above equation.

III. Results

We attempted to predict the course of SARS-CoV-2 infection, using a stochastic method, i.e., Bayes Classification Method. We first calculated the probability that a patients might later develop symptoms among one of three states of severity, i.e., "Severe", "Moderate" and "Asymptomatic". Secondly, we described the profiles of levels of cytokines, that could predict the course of disease.

We extracted data regarding patients' symptoms from the file "GSE178967_series_matrix.txt" with GEO accession number GSE178967 described using the three states, "Asymtomatic", "Moderate" and "Severe". The RNA expression data was extracted from the file "GSE178967_RNAseq_counts.csv" with the same GEO accession number.

The expression data consisted of RNA-sequencing assays using blood samples collected on day 0 and day 5 following enrollment. We used data named by L_***_00 or L_***_05 , which were deduced from blood samples collected on day 0 or day 5 respectively.

Data at day 0 was a set of 105 samples of 35587 expressions of RNA, forming a matrix of 105 rows and 35587 columns. The severities were grouped into 7 asymptomatic, 90 moderate, 8 severe. Each of the samples were named "L_***_00. We restricted the columns (i.e., expression data) to interleukins and chemokines, excluding those with very low expression.

For the Bayes classification method, it is necessary to select two groups from samples, the train data and the test data. The train data is used to make a prediction model, and the test data is to calculate the precision of prediction. To proceed with this method, we attempted three scenarios namely (a), (b) and (c):

(a) We took all the data as train data and we considered all the states of severity (asymptomatic, moderate and severe). This means that we used all samples and all classes of symptoms.
(b) We took all the data as train data, and restricted the severity to two classes, asymptomatic and severe.
(c) We randomly and comprehensively identified train data and test data, and calculated the mean of all the precision numbers.

The scenario in (c) was the most relevant because the majority of samples belonged to patients with moderate disease with a limited number of patients with either severe disease states. Therefore, we used a comprehensive method that randomly selected patients with both severe and asymptomatic states. We randomly selected 2 samples from 7 patients with severe disease and 2 samples from 8 patients with asymptomatic states for the test data. All the remaining samples were used for the train data.

The predictive probability was calculated for all these scenarios, and the following results was obtained:

(a)0.62, (b)0.87, (c)0.71.

The result from the scenario (c) represented that our preferred prediction method would be useful in predicting the course of disease.

We then extracted the cytokine profiles to evaluate the course to disease. For this purpose, we compared expression levels on day 0 (files L_***_00,) to that on day 5 (files L_***_05) for different interleukins and chemokines, using the method of Log2FC. From this method, we observed an increase and a decrease in expression with time as shown in Fig.1.

Figure 1. Changes of logFC of cytokines between 00 (the first day) and 05 (after five days) calculated for each group of severe and moderate severity.

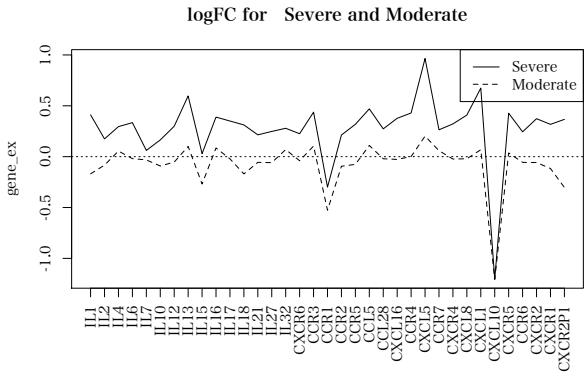
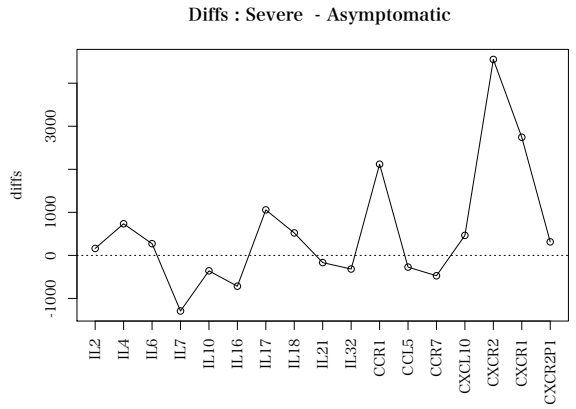


Fig.1 shows that the cytokines, CXCL5, CXCL1, CXCR2, IL13, IL6, IL16, IL17 were highly expressed in patients with severe disease than in patients with moderate disease. Here, IL6 and IL17 are well-known cytokines that play a key role in cytokine storms [7,8]. CXCL1, CXCL5, IL17, CXCR2 are related to each other. CXCL1, CXCL5 are ligands that combine with receptors CXCR2 in neutrophils and macrophage. CXCL5 is related to IL17 in a therapy to decrease joint vascularization [9].

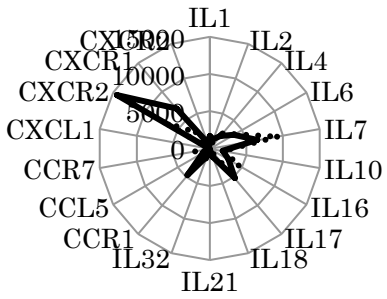
We then investigated the difference in RNA expression between patients with severe disease and asymptomatic patients on day 0 to describe the features of severe disease state. Fig.2 shows the result of our investigation

Figure 2. Comparison of cytokines between severe and asymptomatic at day 0 (the first day).



To visualize the profiles of cytokines more clearly we examined the differences in basic cytokines on day 0 between patients with severe disease and asymptomatic patients. We then illustrated our results using a "radar graph" as shown in Fig.3.

Figure 3. Profiles of cytokines on day 0 (the first day) for each group of severe and asymptomatic severity.



In Fig.3, we observed that the prominent cytokines, CXCR2, CXCR1, CCR1 were higher among patients with severe disease when compared to asymptomatic patients. Remarkably, IL16 is decreasing among patients with severe disease, as subsequently presented in the discussion section. Since IL-7 expression changes intermittently [10], IL-7 may not be a useful parameter to monitor disease progression.

IV. Discussion

If we are given a list of a patient's DNA (or RNA) expression levels of cytokines of a patient, can we tell whether the disease will proceed to a severe state or remain in an asymptomatic state? In other words, is it possible to predict a patient's severity? We answered this question by applying Bayes classification method to patients' cytokine expressions in the results' section.

We then described the characteristic features of cytokine levels associated with varying levels of severity, and identified some cytokines that played an important role in disease severity. Patients with severe disease had higher IL-6 expression than those with moderate disease (Fig.1). IL-17 was the most prominently expressed cytokine in Fig.1 and Fig.2. Both IL-6 and IL-17 were key cytokines in cytokine storms [7,8]. IL-6 is mainly involved in acute inflammation due to its role in regulating the acute phase response [11]. IL-6 could possibly be involved in the disease exacerbation and potential therapeutic approaches could be based on anti-IL-6 biologics [12]. IL-17 inhibitor has been used as a specific treatment for patients with severe COVID-19 pneumonia to control Th17 cell activation [13]. High Levels of IL-17 have also been observed in patients with Rheumatoid Arthritis [8]. IL17 generally enhances the expression of IL-1、IL-6、IL-8、IL-21、TNF-β、and MCP-1, and plays a key role in inducing cytokine storms.

CXCL1, CXCL5, CXCR2, were also highly expressed in patients with severe disease as shown in Fig.1 and Fig.2. These cytokines also played an important role in recognizing the deterioration of Covid-19 in Fig.3. CXCL1, CXCL5 are ligands that combine with the receptors CXCR2 in neutrophils and macrophages. The release of proinflammatory chemokines CXCL5 was elicited by the attenuation of ACE2, and lung inflammation-induced injury was exaggerated in the mouse lung [14]. Anti-CXCL5 therapy improved IL-17-induced arthritis by decreasing joint vascularization [9].

CCR1 and CXCR1 are highly expressed in Fig1 and Fig.2, and were also prominent in Fig.3. The ligands of theses receptors include CCL3, CCL5, CCL7, and CCL23. They are critical in the recruitment of effector immune cells (monocytes or neutrophils) to the exact sites of inflammation. It is also known that the chemokine receptor CCR1 is a potential target for the treatment of rheumatoid arthritis [15]. CCR1 plays a key role in changing the structure of the cytokine network [15]. CXCL1, CXCL5 are ligands that are combined with receptors CXCR2 in neutrophils and macrophage.

A distinct difference in cytokine profiles was noted between patients with severe disease and asymptomatic patients at day 0 in Fig.3. CXCR2, CXCR1, CCR1 were highly expressed in patients with severe disease, whereas IL-16 was less expressed in patients with severe disease.

IL-16 is an eminent cytokine in Fig.3 and has a very interesting role in HIV disease. Fig.3 shows that the production of IL-16 drops in the group of patients with severe disease. Since IL-16 enhances the IL-17 cascade, this observation appears very contradictory. On the other hand, IL-16 inhibits HIV replication [16,17]. They reported that IL-16 inhibited human immunodeficiency virus replication in cells from infected subjects and serum IL-16 levels decreased with disease progression [16]. Our observation of decreased IL-16 levels is like that seen during the progression of HIV disease [16,17].

V. Conclusion

Features of cytokine production could be a relevant parameters in predicting most severe course of disease and the need for intensive care. Predicting the course of disease is stochastically possible from the information regarding a profile of DNA(or RNA) expression levels of cytokines. The described pattern could serve as an effective profile in predicting a severe course of disease. The present study suggests that CXCR2, CXCR1, CCR1, IL-16 may also play a role in changing the cytokine network, and further

intensive investigation of cytokine levels is necessary for the management of patients with COVID-19.

REFERENCES

[1] Morimoto H (2021) Complexity in correlated cytokine networks associated with COVID-19. International Journal of Engineering and Applied Sciences (IJEAS) ISSN: 2394-3661, Volume-8, Issue-9, September. https://dx.doi.org/10.31873/IJEAS.8.9.06.

[2] E.O. Gubernatorovaa, E.A. Gorshkovaa, A.I. Polinovaa, M.S. Drutskayaa. IL-6: Relevance for immunopathology of SARS-CoV-2 T E.O. Cytokine and Growth Factor Reviews 53 (2020) 13–24 https://doi.org/10.1016/j.cytogfr.2020.05.009

[3] Z. Hu, K. Ploeg, S. Chakraborty et al. Early immune responses have long-term associations with clinical, virologic, and immunologic outcomes in patients with COVID-19. National Institutes of Health, (preprint) Version 1. Res Sq. Preprint. (2022) Feb 2. http://dx.doi.org/10.21203/rs.3.rs-847082/v1

[4] J.H. Albert, and S. Chi Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association. Vol. 88, No. 422 (1993) 669-679. http://dx.doi.org/10.2307/2290350

[5] D. Soria, J.M. Garibaldi, F. Ambrogi, E. M.Biganzoli , O. Ellis. A 'non-parametric' version of the naive Bayes classifier. Knowledge-Based Systems 24, pp-775–784. https://doi.org/10.1016/j.knosys (2011) 02.014

[6] Morimoto.H. (2017) Trend analysis of cardiovascular events associated with meteorological factors by Bayes analysis. (2017) General Medicine Open, vol.1, 1-4, doi: http://dx.doi.org/10.15761/GMO.1000106

[7] Dina Ragab, Haitham Salah Eldin, Mohamed Taeimah, Rasha Khattab and Ramy Salem, The COVID-19 Cytokine Storm; What We Know So Far Frontiers in Immunology. Vol 11, Article 1446 2020 http://dx.doi.org/10.3389/fimmu.2020.01446

[8] M. Ziolkowska, A. Koc, G. Luszczykiewicz, K. Ksiezopolska-Pietrzak, E. Klimczak, H. Chwalinska-Sadowska and W. Maslinski, High Levels of IL-17 in Rheumatoid Arthritis Patients: IL-15 Triggers In Vitro IL-17 Production Via Cyclosporin A-Sensitive Mechanism. J. Immunol (2000) 164:2832-2838 http://dx.doi.org/10.4049/jimmunol.164.5.2832

[9] Sarah R. Pickens, Nathan D. Chamberlain, Michael V. Volin, Mark Gonzalez, Richard M. Pope, Arthur M. Mandelin II, Jay K. Kolls & Shiva Shahrara, Anti-CXCL5 therapy ameliorates IL-17-induced arthritis by decreasing joint vascularization. Angiogenesis volume 14, (2011) 443–455

[10] Motoko Y. Kimura, Leonid A. Pobezinsky, Terry I. Guinter, Julien Thomas, Anthony Adams, Jung-Hyun Park, Xuguang Tai, Alfred Singer, IL-7 signaling must be intermittent, not continuous, during CD8+ T cell homeostasis to promote cell survival instead of cell death. Nature Immunology, 14, (2013) 143-151 http://www.ncbi.nlm.nih.gov/pubmed/23242416

[11] F. Coperchinia, L. Chiovato, L.Croce, F. Magri, M. Rotondi, The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system. Cytokine and Growth Factor Reviews, 53, (2020) 25-32 https://doi.org/10.1016/j.cytogfr.2020.05.003

[12] Gubernatorova, E.A. Gorshkova, A.I. Polinova, M.S. Drutskaya, IL-6: Relevance for immunopathology of SARS-CoV-2 T E.O. Cytokine and Growth Factor Reviews 53 (2020) 13–24 https://doi.org/10.1016/j.cytogfr.2020.05.009

[13] X. Sun, T. Wanga,1, D. Caib, Z. et al. Cytokine storm intervention in the early stages of COVID-19 pneumonia Cytokine and Growth Factor Reviews 53 (2020) 38–42 https://doi.org/10.1016/j.cytogfr.2020.04.002

[14] M. Bonafè, F. Prattichizzob, A. Giulianic, G. Storcia, J. Sabbatinellic, F. Olivieric, Inflamm-aging: Why older men are the most susceptible to SARS-CoV-2 T complicated outcomes. Cytokine and Growth Factor Reviews 53 (2020) 33–37 https://doi.org/10.1016/j.cytogfr.2020.04.005

[15] M. Amat, C. F. Benjamim, L. M. Williams, N. Prats, E. Terricabras, J. Beleta, S. L. Kunkel, N. Godessart, Pharmacological blockade of CCR1 ameliorates murine arthritis and alters cytokine networks in vivo. British Journal of Pharmacology, Volume 149, Issue 6, (2009) 666-675 https://doi.org/10.1038/sj.bjp.0706912

[16] C. Amiel, E. Darcissac, M.J. Truong, J. Dewulf, M. Loyens, Y. Mouton, A.Capron, and G. M. Bahr, Interleukin-16 (IL-16) Inhibits Human Immunodeficiency Virus Replication in Cells from Infected Subjects, and Serum IL-16 Levels Drop with Disease Progression. JID(The journal of Infectious disease), Volume 179 Issue 1, January (1999) https://doi.org/10.1086/314550

[17] T. Idziorek, J. Khalife, O. Billaut-Mulot, E. Hermann, M. Aumercier, Y. Mouton, A. Capron, and G.M. Bahr, Recombinant human IL-16 inhibits HIV-1 replication and protects against activation-induced cell death (AICD). Clinnical Exp. Immunol. (1998) 112(1): 84–91 http://dx.doi.org/10.1046/j.1365-2249.1998.00550.x

Hiroshi Morimoto Professor emeritus of Nagoya University. He graduated from Mathematical Institute of Nagoya University. He was first employed by Nagoya University as a pure Mathematician. Then he explored many fields including biology, global climate change, human health and DNA, using methods of data mining.