medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A machine learning explanation of the pathogen-immune relationship of SARS-CoV-2 and machine learning models of prognostic biomarkers to predict asymptomatic or symptomatic infections

Eric Luellen, MSc, MPH

Abstract

Asymptomatic people infected during the SARS-CoV-2 pandemic have outnumbered symptomatic people by an approximate ratio of 4:1 with little understanding to date as to why; therefore, they have been difficult to identify. Moreover, studies indicate that most asymptomatic virus-positive patients are infectious, thereby creating a new public health danger via a plethora of "silent spreaders." This data science study identified four novel discoveries that may significantly impact our understanding of the pathogen-immune relationship: (1) Spearman rho correlation coefficients and associated P-values identified 33 of 55 common immune factors have statistically significant associations with SARS-CoV-2 morbidity, their direction (+/-) and strength to inform research and therapies; (2) five machine learning algorithms were applied to 74 observations of these 33 immunological variables and identified three models of prognostic biomarkers that can classify and predict who will be asymptomatic or symptomatic if infected with 94.8% to 100% accuracy; (3) a random forest of 200 decision trees ordinally ranked the 33 statistically significant independent predictor variables by their relative importance in predicting SARS-CoV-2 symptoms; and, (4) three different decision-tree algorithms separately identified and validated three immunological biomarkers and levels that nearly always differentiate asymptomatic patients – SCGF-β (> 127637), IL-16 (> 45), and M-CSF (> 57). The implications of these findings are they indicate a tool that can identify in advance the 20% of people who are at higher risk of morbidity from infection and suggests a specific stem-cell factor for therapeutics.

Keywords: SARS-CoV-2 (COVID-19), asymptomatic, immunological factors, stem cells growth factor-beta

Introduction

This section discusses what was known and unknown on this topic, and the resulting hypothesis. This introduction also puts the importance of these findings and the use of machine learning modeling into context.

While it has been speculated that stem cells may play a role in SARS-CoV-2 and other zoonoses' resistance, prior research has focused on different stem cell involvement than stem-cell growth factor-beta (Yu, 2020) (Golchin, 2020) (Chrzanowski, 2020). Previous research has also established that stem cells can inhibit viral growth by expressing interferon-gamma stimulated genes (ISGs) and have been particularly effective against influenza A H5N1 virus and resulting lung injuries (Wu, 2018) (Chan, 2016). Stem cell therapy (SCT) has been hypothesized as a treatment for SARS-CoV-2; however, there is no record in the literature specific as to which factors may influence SARS-CoV-2 infections, favorably or unfavorably, or to what degree until now (Florindo, 2020).

Elevated SCGF-β has also been associated with the specific disease states of hepatocellular cancer, Chagas' disease, cardiomyopathy, inflammation and insulin resistance, and unstable carotid plaques (Sukowati, 2018) (Wang, 2013) (Tarantino, 2020). Interleukin 16, the second most important variable in predicting SARS-CoV-2 immunity or resistance here, has been strongly associated with asthma (Mathy, 2000).

Prior studies on the biomarkers associated with SARS-CoV-2 immune response and morbidity include interferon-gamma (IFN-γ), interferon-beta (IFN-β), and interleukin-8 (IL-8) (Geifman, 2020). Other previous research on immune parameters associated with SARS-CoV-2 severity and prognosis have involved interleukin one beta (IL-1β) and interleukin six (IL-6). However, others found reduced immunoglobin G levels in asymptomatic patients (Jesenak, 2020) (Long, 2020). The general finding in prior research regarding the pathogen-immune relationship with SARS-CoV-2 is that symptomatic patients have considerably more inflammation and cytokine storm activity than asymptomatic patients (Long, 2020).

What has been unknown for SARS-CoV-2 are three questions to which the answers are suggested in this study. One, which immunological variables are statistically significant, and how important is each in predicting asymptomatic status? Two, which of those variables, if any, have a strong negative correlation, or relationship, with disease severity (i.e., asymptomatic patients' levels are significantly higher than symptomatic patients)? Three, is there an algorithmic or formulaic model of prognostic

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

biomarkers that can accurately predict morbidity – who will be asymptomatic if infected, and who is at risk of more severe symptoms and disease progression – and why?

Methods

This study was based on secondary data published as a supplement in Nature Medicine in June 2020. Therein immunological factors were measured in 74 patients in the Wanzhou District of China. They were diagnosed with SARS-CoV-2 positive by reverse transcriptase-polymerase chain reaction (RT-PCR) in the 14 days before the recordation of the observations. The median age of the 37 asymptomatic patients was 41 years (range 8-75 years), 22 were female, 15 were male. For comparison, 37 RT-PCR test-positive patients were selected and matched to the asymptomatic group by age, comorbidities, and sex (Long, 2020).

In this study, five algorithms, or types, of machine learning – a kind of artificial intelligence employing robust brute-force statistical calculations – were applied to a data set of 74 observations of 33 immunological factors to attempt to do three things: (1) develop a model to accurately predict the classification of which patients will be asymptomatic or symptomatic to SARS-CoV-2; (2) determine the relative importance of each immunological factor; and, (3) determine if there is any level of a subset of immunological factors that can accurately predict which patients are likely to be immune or resistant to SARS-CoV-2.

Minitab 19 (version 19.2020.1, Minitab LLC) was used to calculate means, 95% confidence intervals, P-values, and two-sample T-tests of statistical significance. Correlation coefficients were also computed using Minitab via Spearman rho because the data was distributed nonparametrically. A second classification and regression tree (CART) algorithm were also applied in Minitab to cross-validate decision tree results from R in Rattle. Minitab's CART methodology was initially described by Stanford University and the University of California Berkeley researchers in 1984 (Breiman, 1984).

The Rattle library (version 5.3.0, Togaware) in the statistical programming language R (version 3.6.3, CRAN) was used to apply five machine learning algorithms – a decision tree, extreme gradient boosting (XGBoost), linear logistic model (LLM), random forest, and support vector machine (SVM) – to learn which model, if any, could predict asymptomatic status, how accurately, and how. Rattle randomly partitioned the data to select and train on 80% (n=59), validate on 10% (7), and test on 10% (7) of observations. Two evaluation methods were used: (1) plots of linear fits of the predicted versus observed categorization; and, (2) a pseudo-R^2 measure calculated as the square root of the correlation between the predicted and observed values. Pseudo-R^2 measure results were evaluated twice, each using for evaluation data that were held back by being randomly selected during partitioning and averaging the two accuracy findings for the final results.

Rattle's rpart decision tree was also used to identify if any levels of one or more immunological factors that could accurately diagnose someone was asymptomatic (i.e., via rules). The decision tree results reported here used 20 and 12 as the minimum number of observations necessary in nodes before the split (i.e., minimum split). The trees used 7 and 4 as the minimum number of observations in a leaf node (i.e., minimum bucket).

The random forest analysis in Rattle began by running a series of differently sized random forest algorithms, ranging from 50 to 500 decision trees, to learn the optimum number of trees to minimize error. Each random forest consisted of a minimum of six variables, which was closest to the square root of the number of statistically significant variables, 33. The lowest error rate was approximately 200 decision trees, which was applied, using four variables at a time, which was the closest whole number to the square root of the number of predictors.

The five machine learning models and CART classification trees were run, including and excluding SCGF-β to identify if there were alternative prognostic biomarkers and levels in the immune profile that could accurately classify and predict SARS-CoV-2 immunity.

Results

Thirty-three (33) of the 55 immunological factors (60%) were indicated as statistically significant by P-values less than .05 from a Spearman correlation. Conversely, 40% of the 55 immune factors had no statistically significant role in mediating whether a patient was asymptomatic or symptomatic to SARS-CoV-2. The 22 factors positively correlated with being symptomatic ranged from a minimum of .205

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(MCP-3) to a maximum of .781 (TRAIL). The 11 factors negatively associated with being symptomatic ranged from a minimum of -.866 (SCGF-β) to a maximum of -.276 (IFNα2) (see Table 1). Researchers have recently found that symptomatic patients generally have a more robust immune response to SARS-CoV-2 infection, culminating in cytokine storms in the worst cases. Conversely, asymptomatic patients have been found to have a weaker immune response (Long, 2020). Because infections are causal to immune response, of particular interest in this study, were the most impactful immune-related variables that negatively correlated with asymptomatic status (i.e., variables that were greater for asymptomatic patients than symptomatic patients), which are highlighted in gray in Table 1.

When SCGF-β was included in the machine-learning analysis, two algorithms predicted and classified SARS-CoV-2 immunity or resistance by being asymptomatic with 100% accuracy: a decision tree, and XGBoost. When SCGF-β was excluded, a random-forest algorithm predicted and classified SARS-CoV-2 asymptomatic and symptomatic cases with 94.8% area under the ROC curve accuracy (95% CI 90.17% to 100%) (see Table 2).

Notably, both the rpart decision trees and CART classification trees independently identified three prognostic biomarkers at specific levels that could classify asymptomatic and symptomatic cases with 95-100% accuracy. When SCGF- β was included, all asymptomatic cases had levels > 127656.8, while all symptomatic cases had levels < 127656.8 (see Figure 1). When SCGF-β was excluded, as a type of contingency analysis to understand prognostic biomarker levels in other factors better, IL-16 accurately classified asymptomatic cases > 44.59 and symptomatic cases < 44.59 in 90.4% of the cases. In the remaining 9.6% of cases where IL-16 > 44.59, all of them had M-CSF > 57.13 (see Figure 2).

Two-sample T-tests for the four factors with the highest positive and negative correlation coefficients, interquartile ranges, outliers, and levels between asymptomatic and symptomatic patients that were statistically significant were computed to ordinally rank factors by their correlation coefficients (see Figure 3).

A random forest analysis of the most important variables to accurately classify and predict SARS-CoV-2 patients by binary morbidity ordinally ranked the 33 statistically significant factors. Unsurprisingly, SCGF-β, and IL-16, followed by GRO-α and TRAIL, respectively, were the most critical factors in predicting morbidity (see Figure 4).

Finally, the results suggest that IL-1β, 3, 4, 9, 12, 13, 17, and RANTES are of low importance, or comparative irrelevance, in the pathogen-immune relationship and, that SCGF-β, IL-16, HGF, INFNα2, LIF, CTACK, IL-1α, Eotaxin, GM-CSF, IL-1Rα, and IL-5 are critical in models to predict and classify asymptomatic or symptomatic SARS-CoV-2 cases accurately.

Discussion

This work's overarching importance is the identification of immunological factors for diagnoses, treatments, and pre-clinical prophylactic immune-based approaches to SARS-CoV-2 in the first seven months of a pandemic that experts now opine will last decades (Farrar, 2020). Immunostimulant approaches are especially valuable because, unlike antivirals and vaccines, they may be given later in the course of the disease to optimize outcomes (Florindo, 2020).

The primary importance of this work is machine learning algorithmic models that can predict with high accuracy, whether someone, once infected, will be asymptomatic or symptomatic from SARS-CoV-2. This knowledge gives clinicians new tools to identify populations in advance who appear to be at higher risk of danger from the virus. Such devices, especially once reproduced in a more extensive study, may also inform policy decisions as to who needs to shelter-in-place. Finally, because of the scale of this pandemic and practical constraints as to how many vaccination doses can be manufactured how quickly, such tools may become valuable in prioritizing vaccine administration to those in greatest need because they are at the higher biological and immunological risk.

This work's secondary importance is a description of the cytokine and chemokine profile that is associated with asymptomatic or symptomatic SARS-CoV-2 infections. It enables a better understanding of the pathogen-immune relationship. These profiles provide insights into the biological pathways critical for SARS-CoV-2 progression.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

As one example, stem cell factors secrete multiple factors that regulate immune cells and modulate them to restore tissue homeostasis. These results suggest that higher levels of SCF-β may better control immune responses to prevent the more robust reactions universally associated so far with highly symptomatic patients and, further, prevent high morbidity and mortality cytokine storms. A better understanding of the pathogen-immune relationship may enable researchers to prevent and treat SARS-CoV-2 patients more effectively with therapeutics currently untested and unused. This knowledge may also extend to similar zoonotic coronaviruses in the future.

The tertiary importance of this work is identifying three immune factors and precise levels that appear to be prognostic biomarkers as to whether someone, once infected with the SARS-CoV-2 virus, will be immune or resistant, as demonstrated by being asymptomatic, or not. These insights also suggest new candidates for therapeutic research focused on the relatively newly identified and ill-understood SCGF-β and its role in the immunological process.

The quaternary importance of this work is further proof that machine-learning methods can accurately and quickly identify critical elements of disease dynamics that accelerate understanding and improve outcomes during pandemics.

This study has several limitations. One, it is unknown from the dataset how many days passed between exposure to the virus and immunological testing, or whether it was universally the same number of days. Two, because immune profiles are temporally sensitive, ideally, several tests would have been taken over several days, which did not occur (Janford, 2020). Three, immunological signaling and processing are multifactorial and complex. Therefore, it is unclear why SCGF-Beta levels are categorically high in asymptomatic patients and low in symptomatic patients, or whether they are causal to SARS-CoV-2 response. Four, combinatorial and sequential analysis of these immunological elements may be an important future research area to optimize therapeutic research outcomes. Five, at least one study in a leading journal, *Lancet*, found that Chinese SARS-CoV-2 case data may have been misreported by as much as 400% (Tsang, 2020). That study, and much higher case and fatalities numbers in over 200 countries, have created distrust and skepticism of SARS-CoV-2-related data originating in China.

Future research could ameliorate these limitations and focus on a more extensive study group to attempt to reproduce the results. Moreover, a prospective case-control study of patients with decreased SCFG-β levels and supplementation was protective against SARS-CoV-2 severity and symptoms.

Conclusion

One implication of these findings is that if we can predict the 80% of society who may be immune or resistant to SARS-CoV-2, or asymptomatic, it may profoundly impact public health intervention decisions as to who needs to be protected and how much. If, for example, 80% of the shelter-in-place orders and the resultant dramatic reduction in economic and social activity could have been prevented by accurately predicting who is at low risk of infection, the economic benefits alone may have been valued in US$ trillions. The second implication of these findings is evidence that elevated levels of SCGF-β, IL-16, and M-CSF may have a causal relationship with SARS-CoV-2 immunity or resistance may have utility as diagnostic determinants to (a) inform public health policy decisions to prioritize and reduce shelter-in-place orders to minimize economic and social impacts; (b) advance therapeutic research; and, (c) prioritize vaccine distribution to benefit those with the greatest need and risks first.

Acknowledgments

The author wishes to thank Dr. Luka Fajs for his helpful comments on drafts of this article.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figures

Immunological Factor	Abbreviation		Pairwise Spearman Correlation to Asymptomatic (0) or Symptomatic (1) Status	95% CI	P-Value (< .05 target)
TNF-related apoptosis-inducing ligand	TRAIL		0.781	(0.654, 0.865)	< .000
Growth-regulated oncogene alpha	GRO-α		0.75	(0.611, 0.845)	< .000
Macrophage colony stimulating factor	M-CSF		0.748	(0.608, 0.843)	< .000
Interleukin-6	IL-6		0.705	(0.549, 0.813)	< .000
Granulocyte colony-stimulating factor	G-CSF		0.697	(0.539, 0.808)	< .000
Interleukin-2	IL-2		0.667	(0.499, 0.787)	< .000
Nerve growth factor beta	NGF-β		0.651	(0.479, 0.775)	< .000
Interleukin-10	IL-10		0.614	(0.431, 0.748)	< .000
Monocyte chemoattractant protein-1	MCP-1		0.594	(0.407, 0.733)	< .000
Stem cell factor	SCF		0.586	(0.397, 0.728)	< .000
Interleukin-15	IL-15		0.527	(0.325, 0.683)	< .000
Interleukin-8	IL-8		0.514	(0.311, 0.673)	< .000
Interferon gamma	IFN-γ		0.464	(0.252, .633)	< .000
Interleukin-7	IL-7		0.454	(0.240, 0.625)	< .000
Interferon gamma inducible protein-10	INF-γ-IP-10		0.451	(0.237, 0.623)	< .000
Interleukin-18	IL-18		0.438	(0.223, 0.613)	< .000
Platelet derived growth gactor-BB	PDGF-BB		0.436	(0.220, 0.611)	< .000
Interleukin-2 receptor alpha	IL-2Rα		0.388	(0.166, 0.572)	0.001
Immunoglobulin G (convalescing)	IgG Convalesce		0.366	(0.143, 0.544)	0.001
Monokine induced by gamma	MIG		0.364	(0.140,0.552)	0.001
Immunoglobulin G (acute)	IgG-Acute		0.33	(0.103, 0.524)	0.004
Macrophage migration inhibitory factor	MIF		0.237	(0.006, 0.444)	0.042
Monocyte chemotactic protein-3	MCP-3		0.205	(-.027, 0.416)	0.079
Vascular endothelial growth factor	VEGF		0.184	(-0.048, .397)	0.117
N gene	N		0.18	(-0.053, 0.394)	0.126
Interleukin-3	IL-3		0.163	(-0.070, 0.379)	0.166
Interleukin-12-p40	IL-12(p40)		0.151	(-0.082, 0.368)	0.199
Interleukin-9	IL-9		0.149	(-0.084, 0.366)	0.206
Interleukin-1 beta	IL-1β		0.125	(-0.107, 0.345)	0.287
Days shed virions	Days shed		0.122	(-0.110, 0.342)	0.3
Stromal cell-derived factor-1 alpha	SDF-1α		0.098	(-0.134, 0.320)	0.406
Interleukin-12-p70	IL-12(p70)		0.083	(-0.149, 0.306)	0.484
Interleukin-17	IL-17		0.067	(-0.164, 0.291)	0.57
Interleukin-4	IL-4		0.02	(-0.210, 0.247)	0.868
Interleukin-13	IL-13		-0.022	(-0.249, 0.208)	0.856
Fibroblast growth factor	FGF		-0.078	(-0.302, 0.153)	0.506
Regulated upon activation, normal T cell expressed and secreted	RANTES		-0.085	(-0.308, 0.146)	0.469
Macrophage inflammatory protein-1 beta	MIP-1β		-0.109	(-0.330, 0.123)	0.353
ORF1ab gene	ORF1ab		-0.113	(-0.334, 0.119)	0.337
Macrophage inflammatory protein-1 alpha	MIP-1α		-0.138	(-0.356, 0.095)	0.241
Tumor necrosis factor-alpha	TNF-α		-0.168	(-0.383, 0.065)	0.153
Tumor necrosis factor-beta	TNF-β		-0.197	(-0.409, 0.035)	0.093
Interferon alpha-2	IFNα2		-0.276	(-0.478, -0.046)	0.017
Leukemia inhibitory factor	LIF		-0.312	(-0.509, -0.84)	0.007
Interleukin-5	IL-5		-0.316	(-0.512, -0.089)	0.006
Interleukin-1 alpha	IL-1α		-0.332	(-0.526, -0.106)	0.004
Granulocyte-macrophage colony-stimulating factor	GM-CSF		-0.359	(-0.548, -0.134)	0.002
Interleukin-1 receptor alpha	IL-1Rα		-0.359	(-0.548, -0.135)	0.002
Eotaxin	Eotaxin		-0.39	(-0.576, -0.169)	0.001
Cutaneous T cell-attracting chemokine	CTACK		-0.456	(-0.627, -0.243)	< .000
Hepatocyte growth factor	HGF		-0.594	(-0.733, -0.407)	< .000
Interleukin-16	IL-16		-0.827	(-0.895, -0.721)	< .000
Stem cell growth factor beta	SCGF-β		-0.866	(-0.92,-0.78)	< .000

Table 1: Immunological factors associated with SARS-CoV-2 morbidity ranked by Spearman correlation coefficients with 95% confidence intervals and P-values (statistically insignificant and corresponding P-values in gray text; negative correlations highlighted in gray at bottom of the table)

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Machine Learning Model	Pseudo R2 (10% evaluation holdback sample 1)	Pseudo R2 (10% evaluation holdback sample 2)	Average Pseudo R2
With SCGF-β			
Decision tree	100.00%	100.00%	100.00%
XGBoost	100.00%	100.00%	100.00%
GLM (logistic)	100.00%	98.89%	99.45%
Random forest	99.46%	94.83%	97.15%
SVM	78.81%	96.99%	87.90%
Without SCGF-β			
Random forest	97.68%	91.91%	94.80%
GLM (logistic)	100.00%	85.96%	92.98%
SVM	77.76%	89.69%	83.73%
XGBoost	99.42%	54.27%	76.85%
Decision tree	100.00%	2.22%	51.11%

Table 2: Comparative accuracy of six machine learning algorithms in predicting SARS-CoV-2 asymptomatic status from immunological factors

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

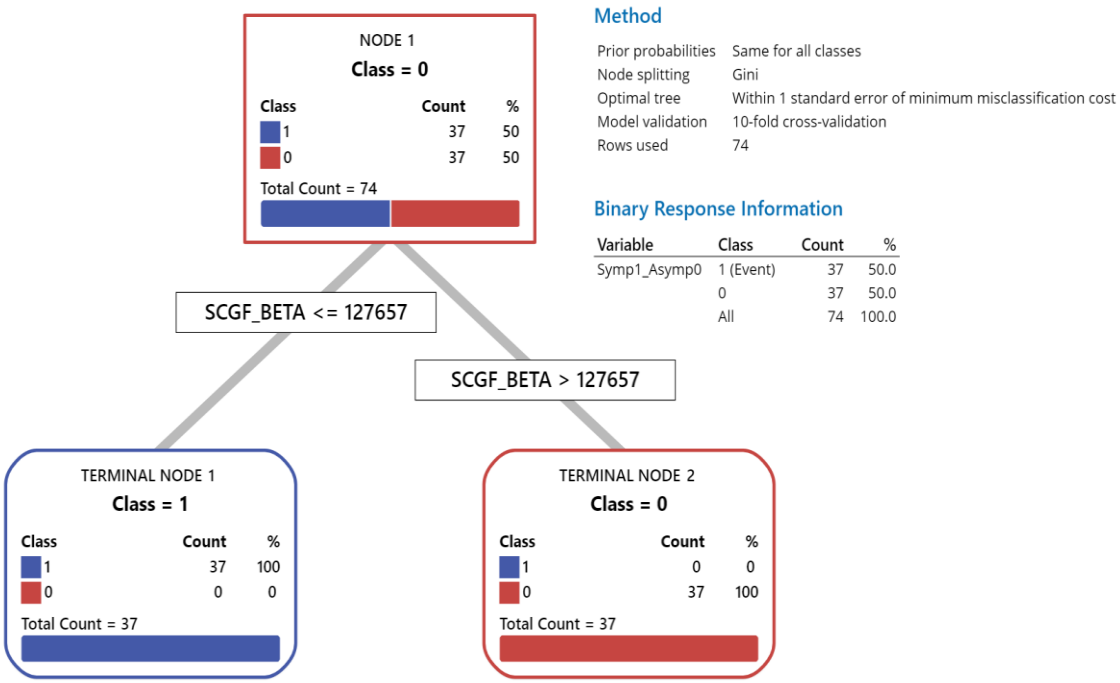


Figure 1: CART classification tree of role of SCGF-β in predicting SARS-CoV-2 morbidity

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

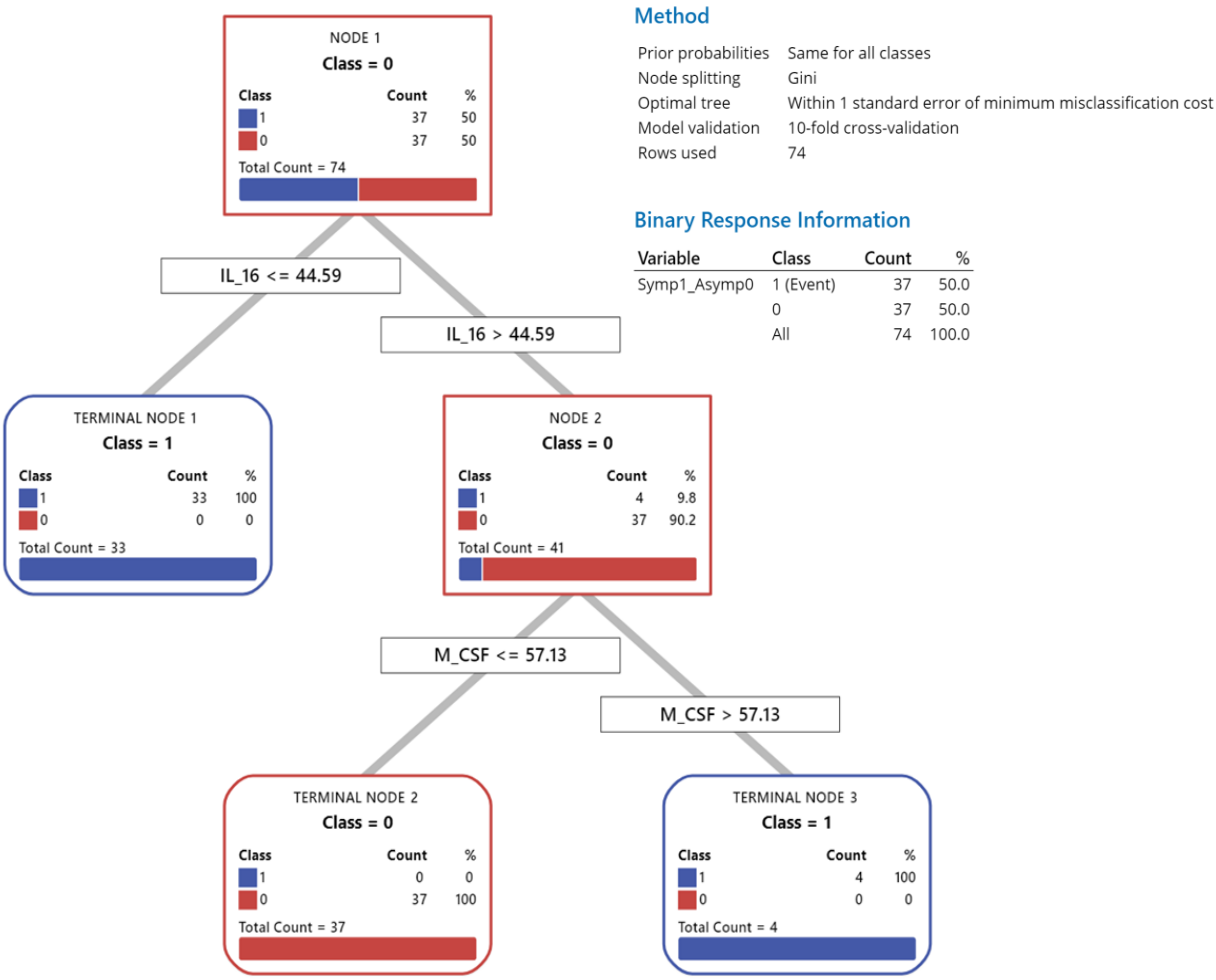
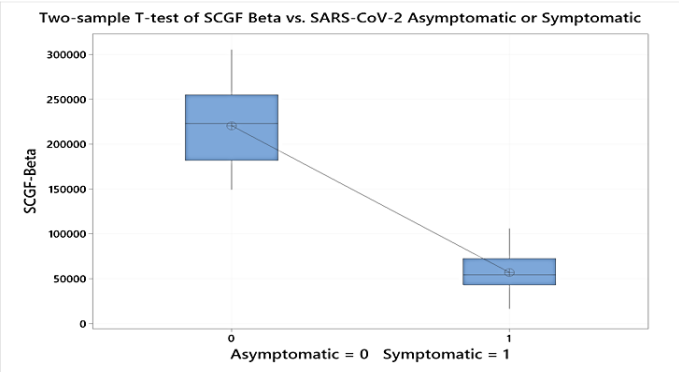


Figure 2: CART classification tree of rule of IL-16 and M-CSF in predicting SARS-CoV-2 morbidity in the absence of SCGF-β

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Descriptive Statistics: SCGF_BETA

Symp1_Asymp0	N	Mean	StDev	SE Mean
0	37	220743	41324	6794
1	37	57153	19189	3155

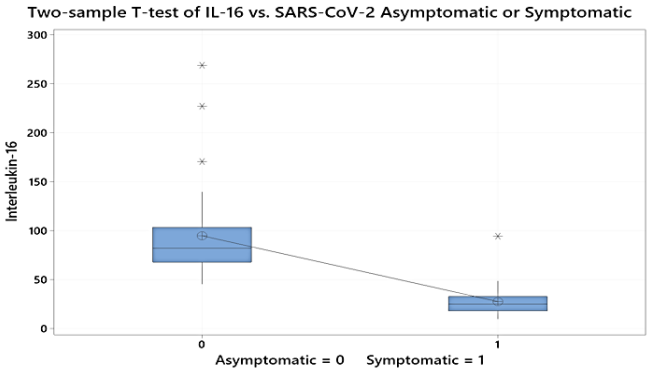
Estimation for Difference

Difference	95% CI for Difference
163590	(148545, 178634)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
21.84	50	0.000



Descriptive Statistics: IL_16

Symp1_Asymp0	N	Mean	StDev	SE Mean
0	37	94.8	46.2	7.6
1	37	27.5	15.3	2.5

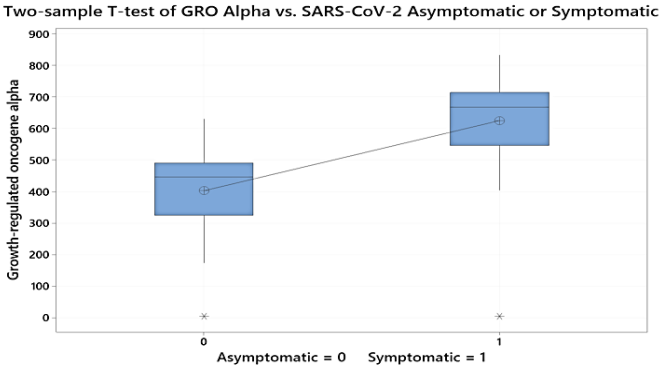
Estimation for Difference

Difference	95% CI for Difference
67.22	(51.07, 83.37)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
8.39	43	0.000



Descriptive Statistics: GRO_ALPHA

Symp1_Asymp0	N	Mean	StDev	SE Mean
0	37	403	121	20
1	37	625	144	24

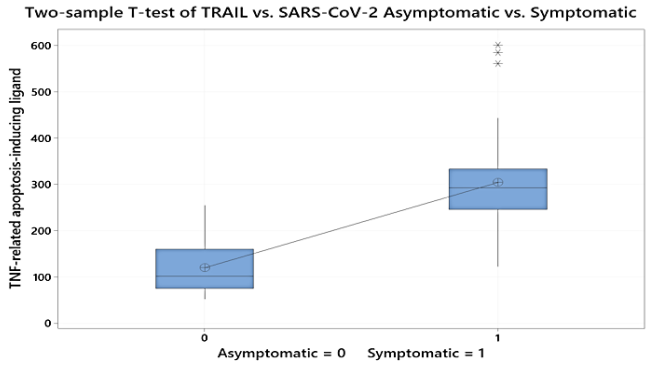
Estimation for Difference

Difference	95% CI for Difference
-222.4	(-284.1, -160.7)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-7.19	69	0.000



Descriptive Statistics: TRAIL

Symp1_Asymp0	N	Mean	StDev	SE Mean
0	37	119.9	57.6	9.5
1	37	304	114	19

Estimation for Difference

Difference	95% CI for Difference
-184.2	(-226.2, -142.3)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-8.80	53	0.000

Figure 3: Two-sample T-tests of statistical significance of difference in means of four leading prognostic biomarkers for asymptomatic or symptomatic SARS-CoV-2

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

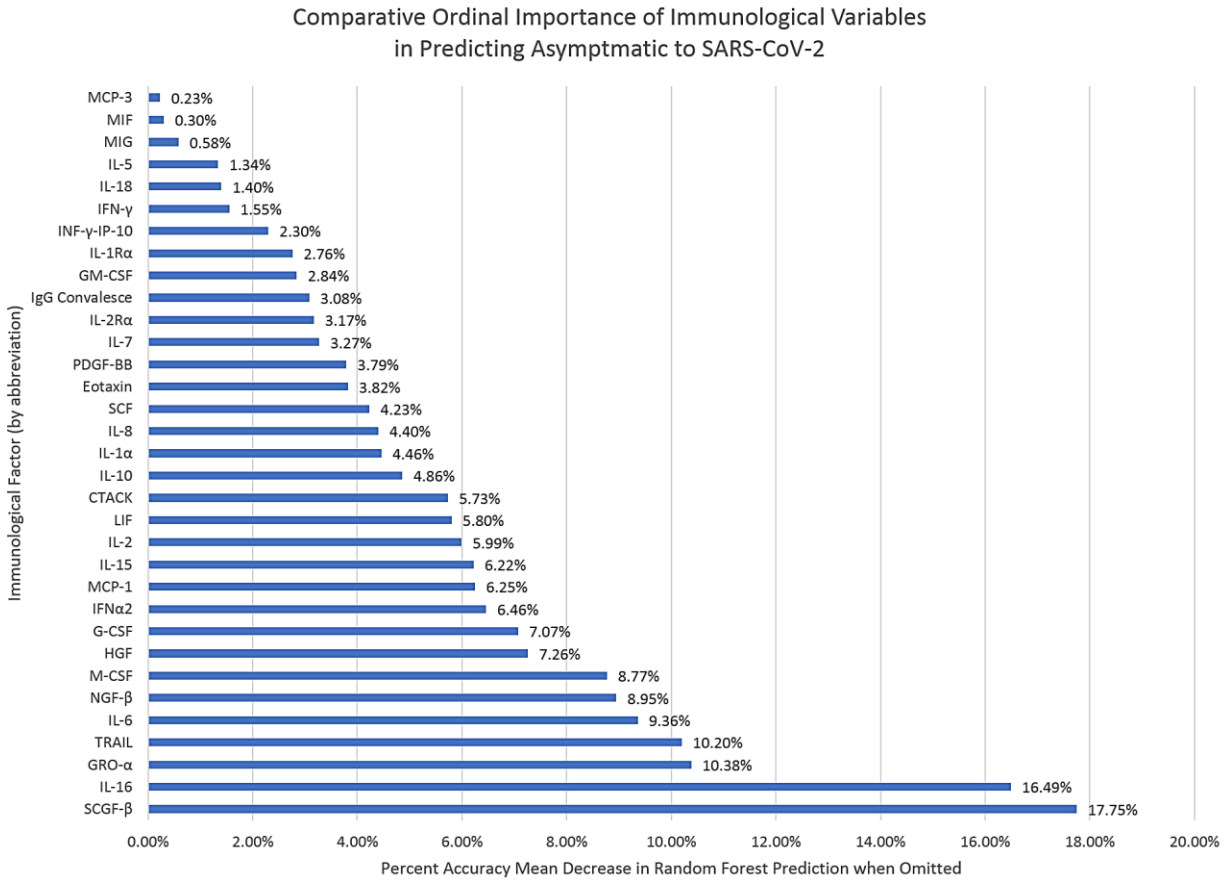


Figure 4: Relative importance of immunological variables from random forest analysis in predicting SARS-CoV-2 morbidity

medRxiv preprint doi: https://doi.org/10.1101/2020.07.27.20162867; this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

Breiman, L., Friedman, J., Olshen, R., Stone, C. (1984). *Classification and Regression Trees.* Boca Raton, Florida: Chapman & Hall/CRC.

Chan, M., Kuok, D., Leung, C., Hui, K., Valkenburg, S., Lau, E., & Nicholls, J., et al. (2016). Human mesenchymal stromal cells reduce influenza A H5N1-associated acute lung injury in vitro and in vivo. *Proceedings of the National Academy of Sciences*, 113(13): 3621-3626. DOI:https://doi.org/10.1073/pnas.1601911113

Chrzanowski, W., Kim, S., McClements, L. (2020). Can stem cells beat COVID-19: Advancing stem cells and extracellular vesicles toward mainstream medicine for lung injuries associated with SARS-CoV-2 infections. *Frontiers in Bioengineering and Biotechnology*. DOI:doi.org/10.3389/fbioe.2020.00554

Farrar, J. (2020, July 21). *Coronavirus: 'Infection here for many years to come.'* Retrieved from BBC: Health: https://www.bbc.com/news/uk-53488142

Geifman, N., Whetton, A. (2020). A consideration of publication-derived immune-related associations in Coronavirus and related lung-damaging diseases. *MedRxiv*. DOI:https://doi.org/10.1101/2020.04.15.20065425

Golchin, A., Seyedjafari, E., Ardeshirylajimi, A. (2020). Mesenchymal stem cell therapy for COVID-19: Present or future. *Stem Cell Reviews and Reports*, 13: 1–7. DOI:10.1007/s12015-020-09973-w

Janford, R. (2020, July 22). (E. Luellen, Interviewer)

Jesenak, M., Brndiarova, M., Urbancikova, I., Rennerova, Z., Vajtkova, J., Bobcakova, A., & Ostro, R., et al. (2020). Immune parameters and COVID-19 infection – associations with clinical severity and disease prognosis. *Frontiers in Cellular & Infection Microbiology*. DOI:https://doi.org/10.3389/fcimb.2020.00364

Liaw, A., Wie, M. (2002). Classification and regression by random forest. *R News*, 2-3: 18-22.

Long, Q., Tang, X., Shi, Q., Li, Q., Deng, H., Yuan, J. & Hu, J., et al. (2020). Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Medicine*. DOI:https://doi.org/10.1038/s41591-020-0965-6

Mathy, N., Scheuer, W., Lanzendörfer, M., Honold, K., Ambrosius, D., Norley, S. & Kurth, R. (2000). Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes. *Immunology*, 100(1): 63–69. DOI:10.1046/j.1365-2567.2000.00997.x

Schiro, A., Wilkinson, F., Weston, R., Smyth, J., Serracino-Inglott, F., Alexander, M. (2015). Elevated levels of endothelial-derived microparticles and serum CXCL9 and SCGF-β are associated with unstable asymptomatic carotid plaques. *Scientific Reports*, 5: 16658. DOI:10.1038/srep1665

Sukowati, C., Patti, R., Pascut, D., Ladju, R., Tarchi, P., Zanotta, N., & Comar, M., et al. (2018). Serum stem cell growth factor-beta for the prediction of therapy response in hepatocellular carcinoma. *BioMed Research International*, 6435482. DOI:10.1155/2018/6435482

Tarantino, G., Citro, V., Balsano, C., Capone, D. (2020). Could SCGF-beta levels be associated with inflammation markers and insulin resistance in male patients suffering from obesity-related NAFLD? *Diagnostics (Basel)*, 10(6): 395. DOI:10.3390/diagnostics10060395

Tsang, T., Wu, P., Lin, Y., Lau, E., Leung, G., Cowling B. (2020). Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modeling study. *The Lancet*, 5(5): e289-e296. DOI:https://doi.org/10.1016/S2468-2667(20)30089-X

Wang, Y. Khan, A., Heringer-Walther, S., Schultheiss, H., Consolação V Moreira, M., & Walther, T. (2013). Prognostic value of circulating levels of stem cell growth factor-beta in patients with Chagas' disease and idiopathic dilated cardiomyopathy. *Cytokine*, 61(3):728-31. DOI: 10.1016/j.cyto.2012.12.018

Wu, X., Thi, V., Huang, Y., Billerbeck, E., Saha, D., Hoffman, H., & Wang, Y., et al. (2018). Intrinsic immunity shapes the viral resistance of stem cells. *Cell*, 172(3):423-438. DOI:10.1016/j.cell.2017.11.018

Yu, F., Jia, R., Tang, Y., Liu, J., Wei, B. (2020). SARS-CoV-2 infection and stem cells: Interaction and intervention. *Stem Cell Research*, 46: 101859. DOI:10.1016/j.scr.2020.101859