

1 Early immune responses have long-term associations with clinical, virologic, and immunologic
2 outcomes in patients with COVID-19

3
4 Zicheng Hu^{1,2*}, Kattria van der Ploeg³, Saborni Chakraborty³, Prabhu Arunachalam⁴, Diego
5 Martinez Mori³, Karen B. Jacobson³, Hector Bonilla³, Julie Parsonnet^{3,5}, Jason Andrews³, Haley
6 Hedlin⁶, Lauren de la Parte³, Kathleen Dantzler³, Maureen Ty³, Gene S Tan⁷, Catherine A.
7 Blish^{3,8}, Saki Takahashi⁹, Isabel Rodriguez-Barraquer⁹, Bryan Greenhouse^{8,9}, Atul J. Butte¹,
8 Upinder Singh^{3,10}, Bali Pulendran^{4,10,11}, Taia T. Wang^{3,10}, Prasanna Jagannathan^{3,10*}

9
10 ¹ Bakar Computational Health Sciences Institute, University of California, San Francisco

11 ² Department of Microbiology and Immunology, University of California, San Francisco

12 ³ Department of Medicine, Stanford University

13 ⁴ Institute for Immunity, Transplantation, and Infection, Stanford University, Stanford, CA, USA

14 ⁵ Department of Epidemiology and Population Health, Stanford University, Stanford CA, USA

15 ⁶ Quantitative Sciences Unit, Stanford University, Stanford, CA, USA

16 ⁷ J. Craig Venter Institute, San Diego, USA

17 ⁸ Chan Zuckerberg Biohub, San Francisco, CA

18 ⁹ Department of Medicine, University of California, San Francisco

19 ¹⁰ Department of Microbiology and Immunology, Stanford University

20 ¹¹ Department of Pathology, Stanford University, Stanford CA, USA

21

22 *Corresponding authors

23 Zicheng Hu, 490 Illinois Street, San Francisco CA 94158

24 Prasanna Jagannathan, 240 Pasteur Drive Room 3456, Stanford CA 94305

25

26

1 Abstract:

2 The great majority of SARS-CoV-2 infections are mild and uncomplicated, but some
3 individuals with initially mild COVID-19 progressively develop more severe symptoms.
4 Furthermore, mild to moderate infections are an important contributor to ongoing
5 transmission. There remains a critical need to identify host immune biomarkers predictive of
6 clinical and virologic outcomes in SARS-CoV-2-infected patients. Leveraging longitudinal
7 samples and data from a clinical trial of Peginterferon Lambda for treatment of SARS-CoV-2
8 infected outpatients, we used host proteomics and transcriptomics to characterize the
9 trajectory of the immune response in COVID-19 patients within the first 2 weeks of symptom
10 onset. We define early immune signatures, including plasma levels of RIG-I and the CCR2
11 ligands (MCP1, MCP2 and MCP3), associated with control of oropharyngeal viral load, the
12 degree of symptom severity, and immune memory (including SARS-CoV-2-specific T cell
13 responses and spike (S) protein-binding IgG levels). We found that individuals receiving
14 BNT162b2 (Pfizer–BioNTech) vaccine had similar early immune trajectories to those
15 observed in this natural infection cohort, including the induction of both inflammatory
16 cytokines (e.g. MCP1) and negative immune regulators (e.g. TWEAK). Finally, we
17 demonstrate that machine learning models using 8-10 plasma protein markers measured
18 early within the course of infection are able to accurately predict symptom severity, T cell
19 memory, and the antibody response post-infection.

20

21

22

23

24

25

26

1 Although the clinical course of COVID-19 exhibits considerable heterogeneity, the great
2 majority of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) infections
3 are uncomplicated with mild symptoms throughout the course of the illness. However, some
4 initially mild infections progress to more severe and/or prolonged symptoms as well as
5 sustained disability¹. Moreover, mild infections are an important contributor to ongoing
6 transmission. An improved understanding of host determinants of clinical, virologic, and
7 immunologic outcomes of SARS-CoV-2 infection is essential to addressing the unmet clinical
8 need for novel therapies for COVID-19.

9 The early host response to acute SARS-CoV-2 infection likely plays a critical role in
10 determining disease outcome and generation of virus-specific memory immune responses.
11 Nucleic acid pattern recognition receptors (PRRs) mediate the early detection and host
12 response to viral infections, with RNA virus recognition thought to occur mainly in the
13 endosomal and/or cytosolic compartment by two different PRRs: Toll-like receptors (TLRs) and
14 RIG-I-Like Receptors (RLRs). Viral recognition by TLRs and RLRs typically triggers a signaling
15 cascade leading to induction of pro-inflammatory cytokines and type I and type III interferons
16 (IFN). Type I and Type III IFNs lead to induction of antiviral effectors (interferon-stimulated
17 genes, ISGs) which provide both a cell-intrinsic state of viral resistance and help coordinate the
18 generation of adaptive immune responses². Importantly, SARS-CoV-2 proteins interfere with
19 induction of these IFNs²⁻⁴, and deficient IFN production has been associated with increased
20 disease severity and poor outcomes^{5,6}. Most studies evaluating the early host immune
21 response have been performed in patients with severe disease^{5,7-13}; signatures associated with
22 patient outcomes among those with more mild to moderate disease remain less well defined¹⁴.

23 Naturally acquired immunity to SARS-CoV-2 results in protection from reinfection,
24 mediated in part by SARS-CoV-2-specific memory T cell and antibody responses. However,
25 there is considerable heterogeneity in the T cell and antibody response following natural SARS-
26 CoV-2 infection, providing valuable opportunities to identify key immune components that are

1 associated with protective immunity. The BNT162b2 (Pfizer–BioNTech) vaccine has been
2 widely used throughout the world and is highly effective in preventing SARS-CoV-2 infection, as
3 well as protecting patients from severe symptoms after infections¹⁵. Comparing responses
4 induced by vaccination with those induced by natural SARS-CoV-2 infection could potentially
5 guide researchers to better understand determinants of protective immunity and improve
6 vaccine design.

7 Recent advances have highlighted the power of the tools of systems biology to
8 comprehensively define the innate and adaptive immune response to infections, including
9 SARS-CoV-2^{5,5,7,9–13}. Such analyses have used host transcriptomics and proteomics, and
10 assessments of the adaptive immune response, coupled with computational approaches, to
11 define molecular signatures that predict clinical and immunologic outcomes of infection.

12 In this paper, we utilized a multi-omics approach to define early infection signatures
13 following SARS-CoV-2 infection that predict subsequent symptom severity, oropharyngeal viral
14 load, and memory immune responses. We leveraged longitudinal samples collected from
15 outpatients enrolled in a randomized controlled trial of a type III interferon, Peginterferon
16 Lambda-1a (Lambda, NCT04331899)¹⁶. In this trial, outpatients with initially mild to moderate
17 COVID-19 were recruited within 72 hours of diagnosis, and followed through 7 months post-
18 infection. We observed sequential activation of immune pathways in initially mild to moderate
19 COVID-19 patients within the first 2 weeks of symptom onset, including interferon responses,
20 NK cell activation, T cell activation and B cell responses. We identified variations in plasma
21 proteins, early interferon signaling, and downstream cytokines (MCP1, MCP-2 and MCP-3) that
22 were associated with multiple patient outcomes, including symptom severity, viral load, memory
23 T cell activity and S protein-binding IgG levels measured up to 7 months after enrollment. By
24 comparing the immune response in COVID-19 patients to the response to COVID-19 mRNA
25 vaccine, we show that the immune response after the first dose of vaccination largely mirrors
26 the trajectory of immune response after SARS-CoV-2 infection, including induction of the

1 negative immune regulatory TWEAK, while the response to the second dose of vaccine is
2 characterized by rapid activation of adaptive immunity and an absence of neutrophil response.
3 Finally, we demonstrate that a machine learning model is able to predict symptom severity, T
4 cell memory response and antibody response accurately using 8-10 plasma protein markers
5 measured early during infection.

6

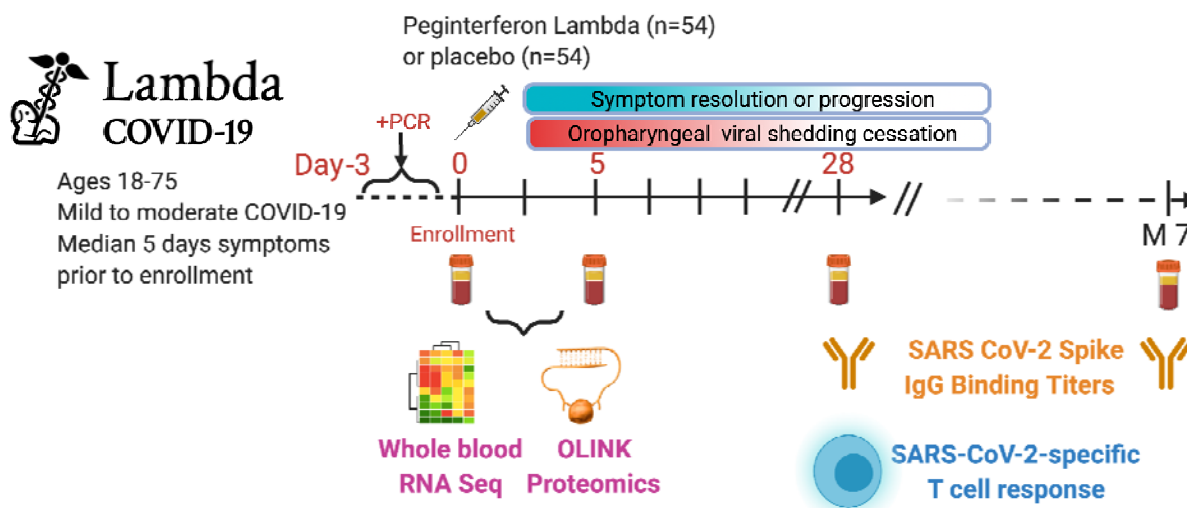
7 Results

8 **Transcriptomic and proteomic profiles correlate with the time to symptom onset in**
9 **COVID-19 patients.**

10 We recruited 108 participants with initially mild to moderate COVID-19 at diagnosis into this
11 study. The median age of participants was 37 years (range 18-71) with 57% male, and 62% of
12 Latinx ethnicity (Supplementary Table 1). Eight (6.7%) participants were asymptomatic at
13 baseline. Of those with symptoms, the median duration of symptoms prior to randomization was
14 5 days.

15

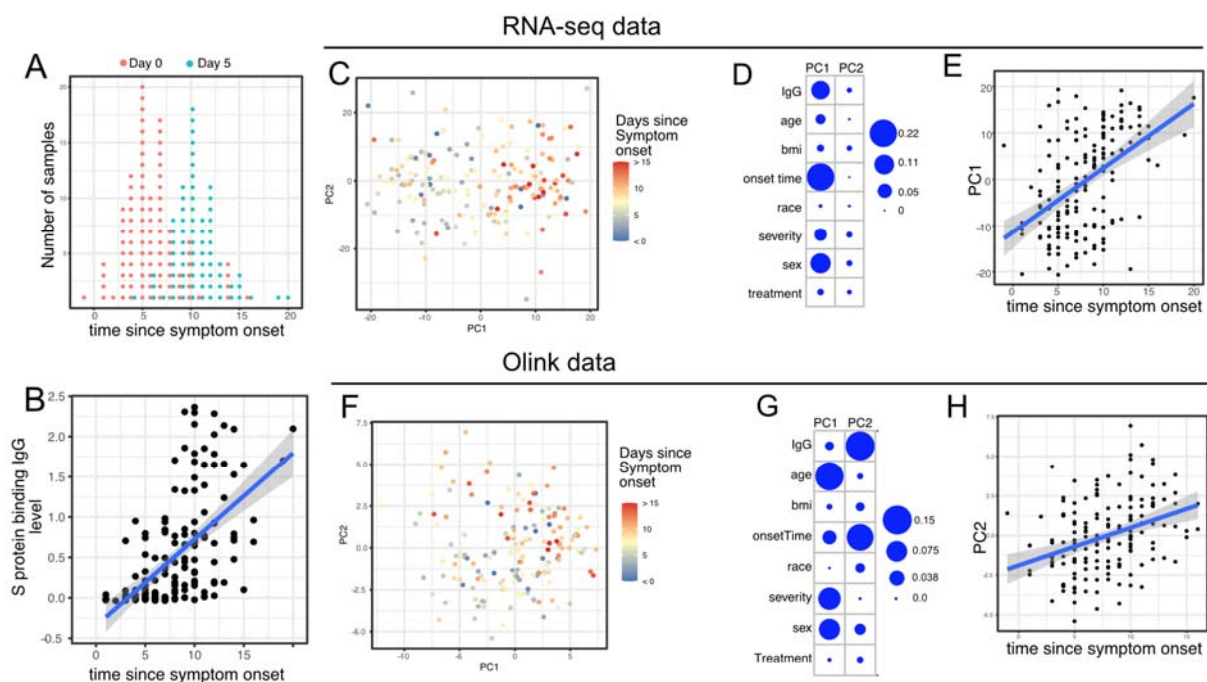
16 Subjects were randomized to receive a single dose of Peginterferon Lambda or placebo at their
17 first visit and followed up to 7 months post-enrollment (Figure 1). The median duration of viral
18 shedding post-enrollment was 7 days, and symptoms was 8 days, and this did not differ
19 between participants randomized to Lambda compared with placebo¹⁶. To profile the immune
20 response in the COVID-19 patients, we conducted whole blood RNA-sequencing and plasma
21 protein profiling with multiplex Olink panels (inflammation and immune response panels, n=184
22 proteins) using blood samples collected at day 0 and day 5 after enrollment. We assessed
23 SARS-CoV-2-specific CD4+ T cell responses by intracellular cytokine staining using PBMC
24 collected at day 28 after enrollment. We also measured IgG binding titers against the SARS-
25 CoV-2 full length spike protein (S) using plasma collected at day 0, day 5, day 28, and month 7
26 (Figure 1).



1
2 Figure 1: Study schema. Outpatients (n=108) with PCR-confirmed SARS-CoV-2 infection and swab obtained within 72 hours of
3 randomization were enrolled in a Phase 2 clinical trial of subcutaneous Peginterferon lambda vs. placebo. In-person follow-up visits
4 were conducted at day 1, 3, 5, 7, 10, 14, 21, 28, and month 7 post-enrollment, with assessment of symptoms and vitals, and
5 collection of oropharyngeal swabs for SARS-CoV-2 testing. Blood obtained at Day 0 and 5 were evaluated by whole blood
6 transcriptomics (RNA Sequencing), plasma proteomics (Olink), and SARS-CoV-2 specific antibodies. Clinical outcomes assessed
7 included duration of symptoms and duration of virologic shedding. Immunologic outcomes assessed including SARS-CoV-2-specific
8 T cell responses at day 28, and antibody responses at day 28 and month 7. Created with biorender.com.

9
10 We first examined antibody levels and transcriptomic profiles at day 0 and day 5 after
11 enrollment in both patients randomized to Peginterferon Lambda and placebo. Based on the
12 subject-reported symptom starting date, these samples were collected -1 to 20 days after
13 symptom onset, with most of the samples collected within the first 2 weeks of the symptom
14 onset (Figure 2A). As expected, we observed a positive correlation between the S protein
15 binding IgG levels at enrollment and the time since symptom onset¹⁷ (Figure 2B). We
16 performed principal component analysis of transcriptomic data and calculated the correlation
17 between the first two principal components (PC) and other clinical variables. We found that PC1
18 had the strongest association with the time since symptom onset and the IgG titer, suggesting
19 that these transcriptomic profiles capture the progression of the immune response in COVID-19
20 patients (Figure 2C-E). We also performed PCA analysis on the Olink data. Similar to results
21 from the analysis of transcriptomics data, Olink data were associated with disease progression,
22 as indicated by the high correlation between PC2 and the time since symptom onset (Figure 2F-

1 H). We also observed an association between PC1 and age, which captures the impact of age
2 on the plasma protein landscape in COVID-19 patients.
3
4 We previously reported that Peginterferon Lambda treatment neither shortened the duration of
5 SARS-CoV-2 viral shedding nor improved symptoms in outpatients with COVID-19¹⁶. PCA
6 analysis revealed that transcriptional and proteomics profiles at day 5 post-treatment were not
7 affected by Peginterferon Lambda treatment (Figure 2D, 2G, and Supplemental Figure 1). In
8 addition, we found no significant differences in the T cell responses (at day 28 after enrollment)
9 and antibody responses (at day 28 and month 7 after enrollment) between the two treatment
10 arms (Supplemental Figure 1), as reported previously¹⁷. Taken together, Peginterferon Lambda
11 treatment did not show noticeable effects on the immune response in COVID-19 outpatients.
12 Therefore, we combined the data from the control and treatment arms together for all
13 downstream analysis.



14
15 Figure 2: Transcriptomics and proteomics profiles correlate with the time to symptom onset in COVID-19 patients. (A) The
16 distribution of RNA-seq sample collection time in respect to symptom onset. The colors of the dots represent the sample collection
17 time from the enrollment. Asymptomatic cases are not shown. (B) Scatter plot showing the positive correlation between SARS-CoV-
18 2 spike (S) protein binding IgG antibody level and the time since symptom onset. (C) PCA plot of the RNAseq samples. The colors

1 of the dots represent the sample collection time from the enrollment. (D) the percent of the variances of PC1 and PC2 explained by
2 different clinical variables. (E) scatter plot showing the positive correlation between PC1 of the RNA-seq data and the time since
3 symptom onset. (F) PCA plot of the Olink proteomics data. The colors of the dots represent the sample collection time from the
4 enrollment. (G) the percent of the variances of PC1 and PC2 explained by different clinical variables. (H) scatter plot showing the
5 positive correlation between PC2 of the Olink data and the time since symptom onset.
6
7

8 **Trajectory analysis reveal sequential activation of immune pathways in COVID-19** 9 **patients**

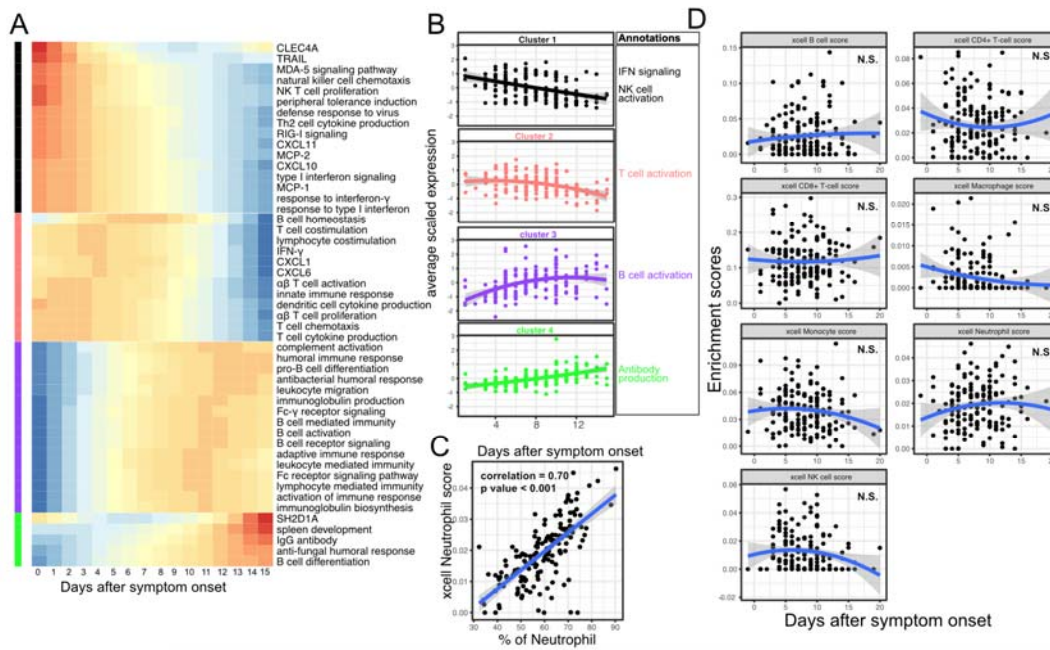
10 We next characterized the trajectory of early transcriptomic and proteomic responses
11 using the RNAseq and Olink data as a function of time since symptom onset. To reduce the
12 dimensionality and improve interpretability, we calculated the enrichment score of different
13 immune pathways (based on Gene Ontology¹⁸) from the RNAseq data. We then combined
14 pathway enrichment scores and Olink measurements into a single dataset for downstream
15 statistical analysis. We fitted the data with quadratic regression to capture the non-linear
16 dynamics of the pathways and proteins. We identified 38 immune pathways and 10 plasma
17 proteins that varied as a function of time since symptom onset (False Discovery Rate (FDR) <
18 0.05, Figure 3A and Supplemental Table 2). Among them, 16 immune pathways or proteins
19 showed nonlinear dynamics, as indicated by significant coefficients of the quadratic term
20 (Supplemental Table 2).

21
22 We performed clustering analysis and identified four clusters based on the trajectory of the
23 significant pathways and proteins (Figure 3A-B). Cluster 1 contains interferon-related pathways,
24 natural killer cell activation pathways and proteins known to be activated by interferon signaling,
25 including MCP-1, MCP-2, CXCL10 and CXCL11¹⁹⁻²². The trajectories in cluster 1 already
26 reached the peak at the time of symptom onset and monotonically decreased over time. The
27 trajectories in cluster 2 peaked at 1-5 days after symptom onset and contain Interferon- γ and
28 pathways related to T cell activation. Interestingly, it also contains several myeloid cell attracting
29 chemokines (CXCL1 and CXCL6) and the innate cell response pathway. Cluster 3 peaked

1 between 10 to 14 days after the symptom onset and is characterized by pathways related to B
 2 cell activation. Cluster 4 trajectories monotonically increase after symptom onset and are
 3 characterized by the increasing S protein binding IgG level and related B cell differentiation
 4 pathways. The trajectory analysis revealed the sequential activation of interferon signaling, NK
 5 cells, myeloid cells, Interferon- γ , T cells, B cell and antibody production within the first 15 days
 6 of symptom onset.

7

8 To characterize how the composition of blood immune cells change over time, we used a
 9 previously established tool named xCell to estimate the enrichment score of the major immune
 10 cells²³. As a positive control, we compared the neutrophil score with the neutrophil count data
 11 obtained from clinical lab tests and found high correlation between them (Figure 3C). Quadratic
 12 regression did not find significant associations between the major cell types and the time since
 13 symptom onset (Figure 3D). The results suggest that the trajectory of different immune
 14 pathways (Figure 3A) are mainly driven by the activation of corresponding immune cells rather
 15 than the composition change of major immune cell types.



16

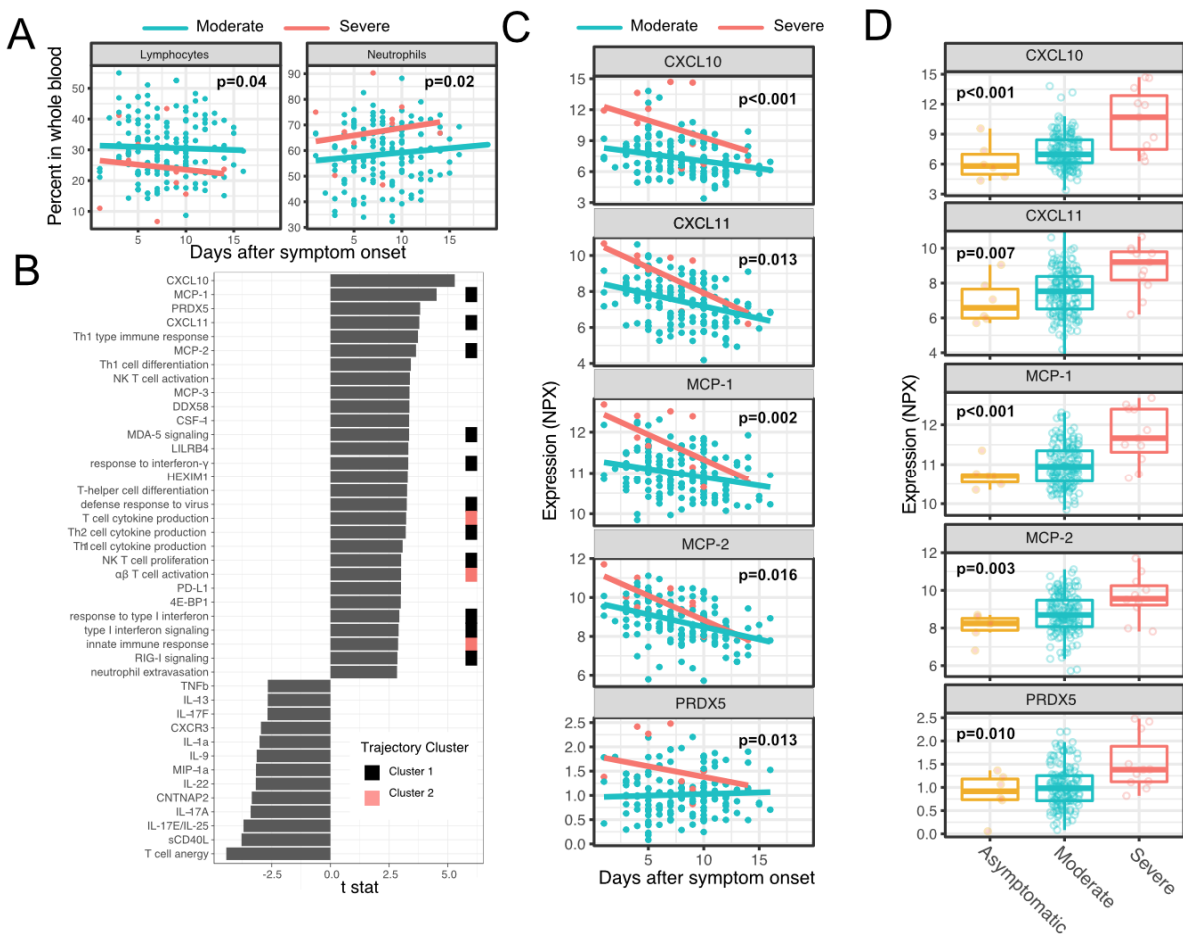
1 Figure 3: Trajectory analysis reveals sequential activation of immune pathways in COVID-19 patients. (A) The fitted expression level
2 of immune pathways and plasma proteins at 0-15 days after symptom onset. The values are calculated by fitting quadratic
3 regressions and are scaled to a mean of 0 and a standard deviation of 1. The color bar on the left side shows the clustering
4 membership of the pathways and plasma proteins. (B) the average trajectory of the clusters. We scaled the expression level of each
5 pathway and plasma proteins to a mean of 0 and a standard deviation of 1. We then calculated the average scaled expression of all
6 members in the clusters. Each dot represents the mean expression in each blood sample. The lines represent the fitted quadratic
7 regression. The grey areas represent the 95% confidence intervals. (C) We estimated the spearman correlation between the
8 neutrophil enrichment score using the xCell. The plot shows the correlation between the xCell score and the counted neutrophil
9 percentage in whole blood. (D) The relationship between xCell enrichment score and days after symptom onset.

10 11 **Variations in early immune responses are associated with disease severity in COVID-19** 12 **patients.**

13
14 We next sought to identify immune pathways and plasma proteins associated with symptom
15 severity in COVID-19 outpatients. At the time of sample collection (day 0 and day 5 after
16 enrollment), the majority of subjects showed either mild to moderate symptoms that
17 subsequently resolved (n=100) or were asymptomatic (n=8). However, 8 patients later
18 developed progressive and more severe symptoms and were hospitalized or presented to the
19 emergency department (median 2 days to progression, range 1-13 days). We defined these
20 individuals as severe COVID-19, and used regression models to identify immune pathways and
21 plasma proteins to compare these participants with those who didn't seek care at the hospital
22 (mild/moderate COVID19), while controlling for days after symptom onset.

23
24 As two positive controls, we confirmed well documented findings that lymphocyte percentages
25 were negatively correlated with symptom severity and neutrophil percentages were positively
26 correlated with symptom severity (Figure 4A)²⁴. In addition, our regression analysis identified 17
27 immune pathways and 24 plasma proteins that are significantly associated with symptom
28 severity (FDR<0.05, Figure 4B-C, and Supplemental table 3). The proteins and pathways from
29 cluster 1 (as identified above in Figure 3A) were significantly enriched (Fisher's exact test, $p <$
30 0.001), including pathways related to interferon response, Rig-I signaling, NK cell activation and
31 multiple protein markers known to be induced by interferon signaling (MCP-1, MCP-2 and
32 CXCL11). The result highlights the association between early immune responses and symptom

1 severity. Our regression analysis excluded the asymptomatic individuals, as their symptom
 2 onset date was unknown. To include the asymptomatic individuals, we performed one-way
 3 ANOVA analysis without adjusting for symptom onset time. The results from the ANOVA
 4 analysis were consistent with the regression analysis (Figure 4D).
 5



6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 Figure 4: Variations in early immune responses are associated with disease severity in COVID-19 patients. (A) Scatter plot comparing the percentage of lymphocytes and neutrophils in whole blood between moderate and severe cases. The lines represent the fitted linear relationship between the percentages and the time after symptom onset. We fitted regression models to test the relationship between the immune measurements and the disease severity while controlling for the time after symptom onset. The p values for the disease severity are reported. (B) We fitted regression models to test the relationship between the immune measurements and symptom severity while controlling for the time after symptom onset. The bar plot shows the t score of the regression coefficient for symptom severity. The colored squares represent the clusters each immune measurement belongs to. The clusters are defined in Fig. 3A. (C) Scatter plot comparing the plasma protein levels between moderate and severe cases. The lines represent the fitted linear relationship between the percentages and the time after symptom onset. The top 5 significant proteins are shown. Data from asymptomatic cases are omitted, as their symptom onset time was unknown. (D) Box plots comparing the plasma protein levels between asymptomatic moderate and severe cases.

18 **Early proteomic and transcriptomic signatures show long-term association with virologic**
 19 **and immunologic outcomes.**

1 We examined associations between plasma proteins measured early in the course of infection
2 and oropharyngeal viral load (measured by the area under the Ct curve from day 0 to 14 post
3 enrollment). We identified 36 plasma proteins significantly associated with oropharyngeal viral
4 load (top 10 significant proteins shown in Figure 5A, Supplemental table 3). Higher levels of
5 several of these proteins were inversely correlated with viral load, including the cytosolic RNA
6 sensor RIG-I (gene symbol DDX58), chemokines (CCL20 and CCL25), and other proteins
7 (Keratin 19 {KRT19}, amphiregulin{AREG}) previously shown to be upregulated in COVID-19
8 patients^{14,25}. Several immune transcriptomic pathways were also associated with viral control,
9 including complement and B cell activation and the humoral immune response pathway (top 10
10 significant pathways shown in Figure 5B, Supplemental Table 3).

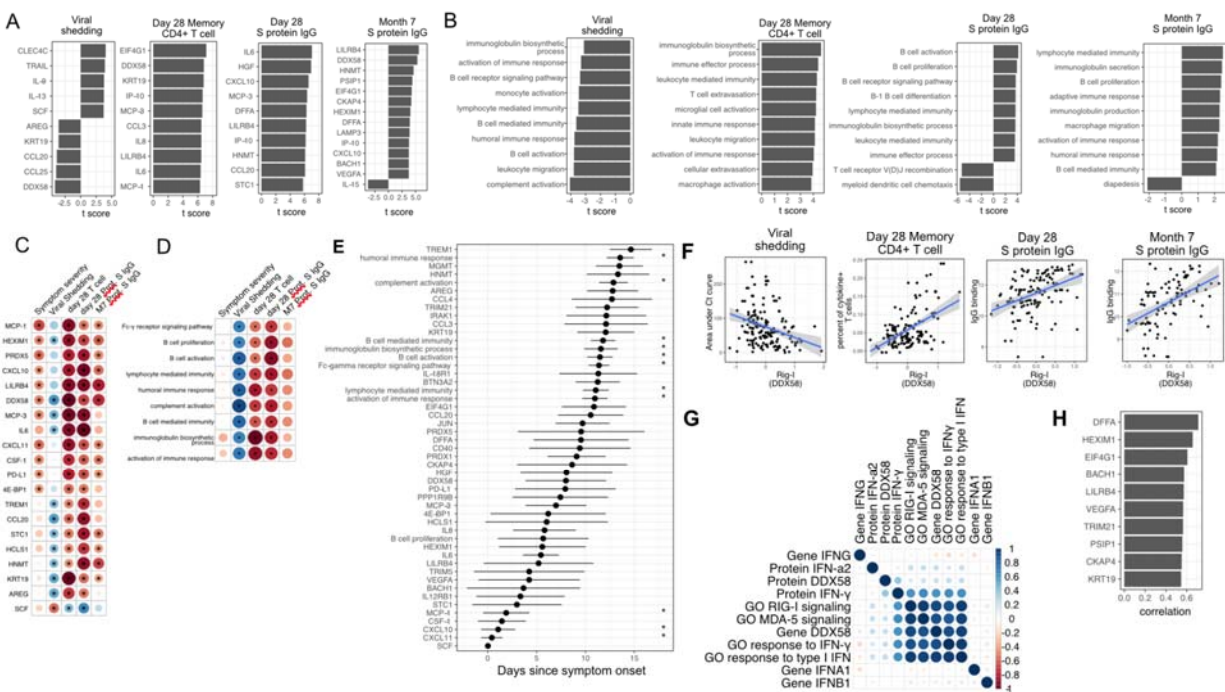
11
12 We next measured associations between early plasma proteins, gene expression and
13 subsequent immune memory, including cytokine-producing SARS-CoV-2-specific T cells
14 measured 28 days post-enrollment (Supplemental Figure 2; Supplemental Table 3), and SARS-
15 CoV-2-specific antibodies measured at 28 days and 7 months post-enrollment. We identified 87
16 plasma proteins that were significantly associated with SARS-CoV-2-specific T cell responses at
17 day 28, and 91 and 13 plasma proteins significantly associated with S protein-binding IgG at
18 day 28 days and month 7, respectively (top 10 significant proteins shown in Figure 5A,
19 Supplemental table 3). Several proteins were associated with higher levels of SARS-CoV-2-
20 specific T cells and the antibody response (Figure 5E), including RIG-I (gene symbol DDX58),
21 chemokines (CXCL11), and other proteins (KRT19, AREG) also associated with control of viral
22 load. Induction of early transcriptomic pathways were also associated with the development of
23 SARS-CoV-2 specific T cells and S protein-binding IgG at day 28 (top 10 significant pathways
24 shown in Figure 5B, Supplemental Table 3).

25

1 We further examined immune pathways and proteins associated with multiple clinical, virologic,
2 and immune outcomes in COVID-19 patients. We identified 21 plasma proteins and 15 immune
3 pathways that are correlated with three out of four aspects of the patient outcomes (Figure 5C
4 and D). These include expected direct links (e.g. the correlation between immunoglobulin
5 production pathway and S protein-binding IgG) and more indirect links (e.g. the correlation
6 between proinflammatory cytokines and S protein-binding IgG) between immune
7 measurements. To establish a sequential order of pathway activation and protein expression,
8 we fit a quadratic regression for each measurement, and then identified the time when the
9 measurement reached maximal expression (Supplemental Figure 2A). This revealed 4 plasma
10 proteins and 12 immune pathways whose expression were significantly associated with time.
11 Among them, the interferon-gamma response, T cell cytokine production and several interferon
12 induced cytokines (MCP1, MCP2, CXCL10 and CXCL11) reached maximal expression within
13 the first 5 day after symptom onset (Figure 5E). Pathways related to B cell activation and
14 antibody production reached maximum the latest, 10 to 15 days after symptom onset. A
15 previous GWAS study has shown that symptom severity is associated with genetic variations
16 near the interferon receptors (INFA2) and CCR2 (receptor for MCP1, MCP2 and MCP3) loci.
17 In addition, these variations were predicted to increase the expression of IFNAR2 and CCR2²⁶.
18 While the exact causal relationship cannot be established from our observational data, together
19 our results suggest that the early interferon-related response and downstream CCR2 signaling
20 shape later adaptive responses, and have long-term impact on the clinical, virological and
21 immunological outcomes in COVID-19 patients.

22
23 Interestingly, plasma levels of RIG-I (gene symbol DDX58) were significantly associated with all
24 examined virologic and immunologic outcomes (Figure 5F), as well as symptom severity (Figure
25 4B). Higher levels of plasma RIG-I were associated with less oropharyngeal viral load, more
26 severe symptoms, increased SARS-CoV-2 specific T cell responses, and increased levels of S

1 protein-binding IgG to SARS-CoV-2. Since RIG-I is a cytosolic PRR that, upon recognition of
 2 short viral double-stranded RNA during a viral infection, leads to upregulation of interferon
 3 signaling²⁷, we explored associations between plasma RIG-I levels and related immune
 4 measurements, including the mRNA-level and protein-level expression of RIG-I and interferons,
 5 as well as RIG-I and interferon-related pathways. We found that the plasma RIG-I levels were
 6 modestly correlated with mRNA-level expression of RIG-I (correlation = 0.23, p value = 0.004,
 7 Figure 5G), as well as RIG-I signaling and Interferon related pathways (Figure 5C). Interestingly,
 8 we found a strong correlation between plasma level of RIG-I and plasma level of DFFA, an
 9 intracellular protein known to be involved in apoptosis (Figure 5H)^{28,29}. In addition, the top 10
 10 plasma proteins correlated with the plasma level of RIG-I are all intracellular proteins (Figure
 11 5H). These data are consistent with the hypothesis that plasma RIG-I is associated with a cell
 12 death process that releases intracellular protein into the plasma. Furthermore, since plasma
 13 levels of RIG-I were not significantly associated with time (Fig 5E), these data suggest that
 14 plasma RIG-I levels might serve as a powerful and stable biomarker for predicting several
 15 clinical, viral and immunological outcomes in patients with COVID-19.



1 Figure 5: Plasma RIG-I is a biomarker for viral shedding, T cell activity, and spike (S)-binding IgG levels (A) The association
2 between plasma proteins and viral shedding, memory T cell activity, and anti-S binding IgG levels. Memory CD4+ T cell
3 activities are measured by the percent of cytokine positive T cells (TNF- α + or TNF γ + or IL21+) after spike protein
4 stimulation. T cells are collected from patients 28 days after enrollment. Spike protein-binding IgG levels are measured 28
5 days or 7 months after enrollment. We fitted regression models to test the relationship between the immune measurements
6 and viral shedding, memory T cell activity, and S protein binding IgG level while controlling for the time after symptom onset.
7 The bar plot shows the t score of the regression coefficient for viral shedding, memory T cell activity, and anti-S binding IgG
8 levels. (B) The association between immune pathways and viral shedding, memory T cell activity, and S protein binding IgG
9 levels. (C-D) Association between plasma proteins and immune pathways that are associated with multiple outcomes. The
10 heatmaps include immune measurements that are significantly associated (indicated by stars) with at least 3 outcomes. (E)
11 The estimated time during which an immune measure reaches the maximum level. The error bars represent standard
12 deviation derived from bootstrapping. The immune measures with stars are significantly associated with time. (F) Correlation
13 between plasma RIG-I (DDX58) and viral shedding, memory T cell activity, and S protein binding IgG levels. (G) Correlation
14 between plasma RIG-I protein and selected level of plasma proteins, genes, and pathways. (H) The top 10 plasma proteins
15 correlated with plasma RIG-I protein.

16

17 **Similar trajectories of immune responses induced by SARS-CoV-2 infection and COVID-** 18 **19 mRNA vaccine.**

19 The BNT162b2 (Pfizer–BioNTech) vaccine has been widely used throughout the world and is
20 highly effective in preventing SARS-CoV-2 infection, as well as protecting patients from severe
21 symptoms after infections¹⁵. We leveraged a recently published Olink proteomics dataset from a
22 BNT162b2 vaccine study³⁰ to compare the immune response induced by COVID-19 vaccine
23 and SARS-CoV-2 infections. Among the 66 protein markers shared between our SARS-CoV-2
24 infection dataset and the vaccination dataset, 8 proteins were significantly associated with time
25 in the SARS-CoV-2 infection dataset, and 22 proteins were significantly associated with time in
26 the vaccination dataset. It should be noted that the data collection in the vaccine study is highly
27 synchronized, with well-defined vaccination day (day 0). In contrast, the time of infections in the
28 SARS-CoV-2 infection dataset are only approximated by the self-reported time since symptom
29 onset.

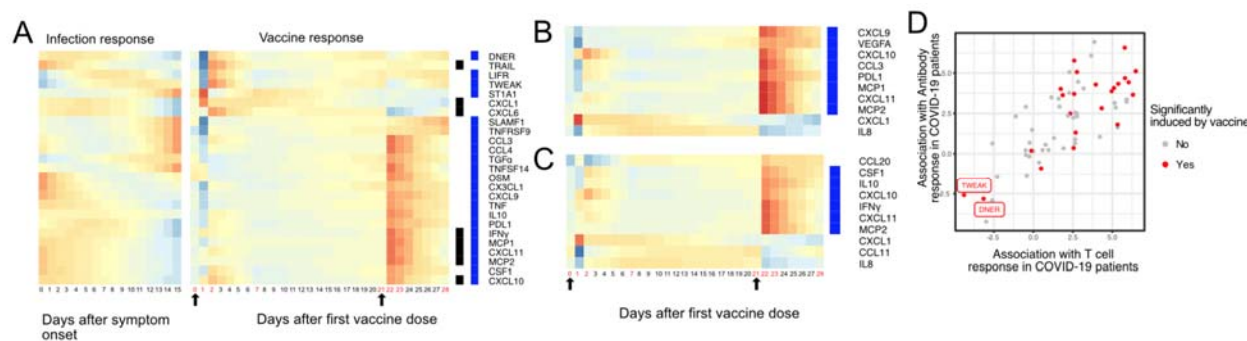
30

31 Comparison of the datasets reveals that the immune response after the first dose of vaccination
32 (day 0 to day 21) largely mirrors the trajectory of immune response after SARS-CoV-2 infection.
33 Early immune markers in the SARS-CoV-2 infection dataset, including IFN γ , MCP1, CXCL11,
34 MCP2 and CXCL10 are upregulated within the first 7 days of the vaccination. Late immune

1 markers in the SARS-CoV-2 infection dataset, including SLAMF1, TNFRSF9, CCL3, CCL4,
2 TGF α and TNFSF14 are upregulated much later and show highest levels 21 days after the
3 vaccination (Figure 6A). In contrast, the response after the second dose of vaccine (day 22 to
4 day 28) is characterized by fast upregulation of both early and late protein markers (Figure 6A).
5 Interestingly, three proteins that are significantly upregulated in SARS-CoV-2 patients were not
6 induced after the second dose of vaccine, including TRAIL, CXCL1 and CXCL6. All three
7 proteins are highly expressed in neutrophils³¹, and have been shown to regulate neutrophil
8 recruitment (CXCL1 and CXCL6) or apoptosis (TRAIL) during inflammation³²⁻³⁴. The lack of
9 TRAIL, CXCL1 and CXCL6 suggests an absence of neutrophil response to the second dose of
10 vaccine.

11
12 Our analysis of the SARS-CoV-2 infection dataset reveals multiple plasma markers that are
13 associated with T cell and antibody responses in COVID-19 patients (Figure 5). We next
14 examined the trajectory of these proteins in response to the BNT162b2 vaccine. Among the top
15 10 proteins positively associated with CD4+ T cell response, 8 proteins are significantly induced
16 by the vaccine (Figure 6B). Among the top 10 proteins positively associated with S protein-
17 binding antibody levels, 6 proteins are significantly induced by the vaccine (Figure 6C). Similar
18 to the result in Figure 6A, we found multiple neutrophil related proteins (CXCL1, IL8 and CCL11)
19 to be induced by the first vaccine dose, but are absent in response to the second vaccine dose.
20 Interestingly, we found that the vaccine also induced two proteins that are negatively associated
21 with T cell and antibody response following natural infection, including TWEAK and DNER
22 (Figure 6D). TWEAK has been known to attenuate the adaptive immunity by inhibiting STAT-1
23 and NF- κ B³⁵, suggesting that its induction could have a negative impact on the protective
24 immunity against COVID-19. Taken together, our comparative analysis shows that the
25 proteomic response of the BNT162b2 vaccine mirrors in many ways the proteomic response
26 after SARS-CoV-2 infection. At the same time, we found important distinctions, including fast

- 1 activation of adaptive immunity, an absence of neutrophil response in response to the second
- 2 dose of vaccine and the induction of TWEAK that may negatively affect the adaptive response.



3
4 Figure 6: Comparing the immune response induced by SARS-CoV-2 infection and COVID-19 vaccine (BNT162b2). (A) The heat
5 map shows the expression level of plasma proteins at 0-15 days after symptom onset in COVID-19 patients (left) and 0-28 days
6 after vaccination in healthy individuals (right). The values from the SARS-CoV-2 dataset are calculated by fitting quadratic
7 regressions and are scaled to a mean of 0 and a standard deviation of 1. The values from the vaccination dataset are computed by
8 fitting using linear interpolation between the measured time points (days in red color) and are scaled to a mean of 0 and a standard
9 deviation of 1. The black bar on the right side shows the protein markers that are significantly associated with time in COVID-19
10 patients (FDR<0.05). The blue bar on the right side shows the protein markers that are significantly associated with time after
11 vaccination (FDR<0.05). The black arrows indicate the time of the first and the second doses of vaccination. (B) The vaccine
12 response of the top 10 plasma proteins that are associated with CD4+ T cell response in SARS-CoV-2 patients. The blue bar on the
13 right side shows the protein markers that are significantly associated with time after vaccination (FDR<0.05, p values are adjusted
14 within the 10 protein set). (C) The vaccine response of the top 10 plasma proteins that are associated with spike (S)-binding IgG
15 levels in SARS-CoV-2 patients. The blue bar on the right side shows the protein markers that are significantly associated with time
16 after vaccination (FDR<0.05, p values are adjusted within the 10 protein set). (D) A scatter plot showing the plasma proteins profiled
17 in both natural infection and vaccination studies. The X axis shows the association (t score of the regression coefficient) between
18 plasma proteins and the SARS-CoV-2-specific memory CD4+ T cell response in natural infection. The Y axis shows the association
19 (t score of the regression coefficient) between plasma proteins and the Spike protein binding IgG in natural infection. Points in red
20 show if the protein is induced after vaccination.

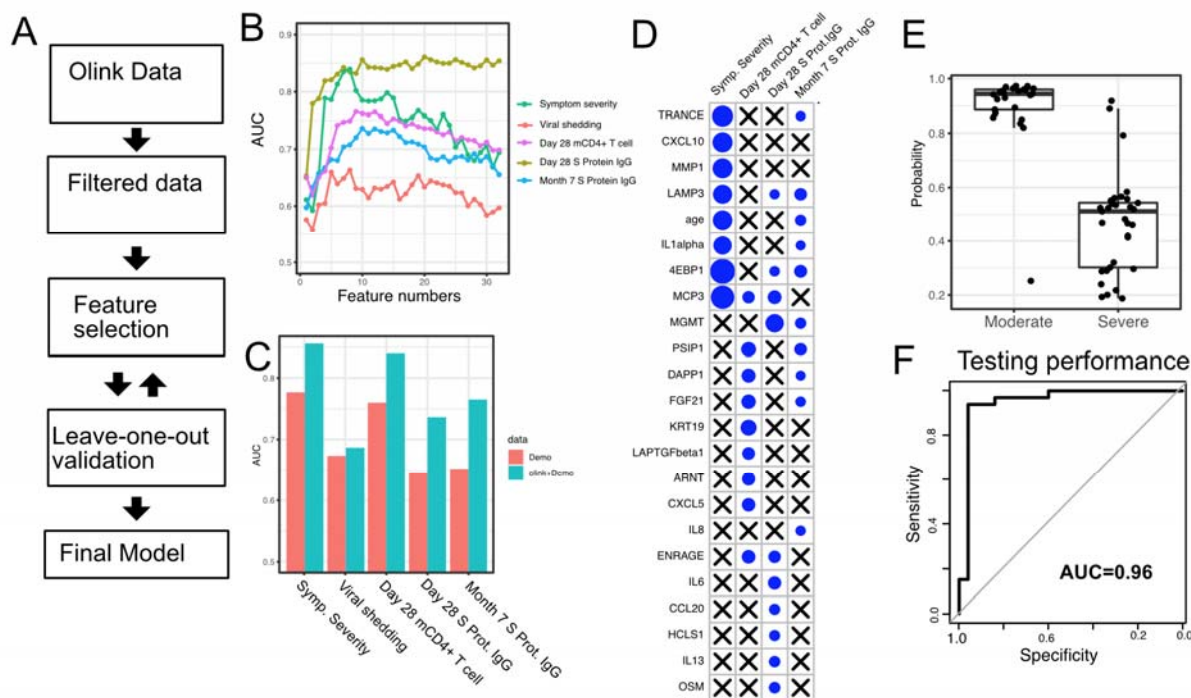
21
22 **Plasma proteins predict symptom severity, T cell response and Spike protein-binding IgG**
23 **levels in COVID-19 patients.**

24 We performed predictive modeling to test if plasma proteins measured early following infection
25 can accurately predict symptom severity, oropharyngeal viral load, and SARS-CoV-2 specific
26 memory T cell and antibody responses manifested later in the study. We adopted a computation
27 pipeline to select a small subset of predictive biomarkers from the 184 proteins measured by
28 Olink assays. We used a leave-one-out cross validation strategy to iteratively evaluate the
29 model performance. We used Random Forest for feature selection and for building the final
30 model (Figure 7A). Based on results from cross-validation, we selected between 8 to 10 protein
31 markers measured at early infection to predict each of the five outcomes. The final models
32 achieved cross-validation AUC of 0.84, 0.66, 0.77, 0.84 and 0.75 for predicting symptom

1 severity, oropharyngeal viral load, memory t cell activity, day 28 spike protein binding IgG levels
2 and month 7 spike protein binding IgG levels, respectively (Figure 7B).

3
4 We compared the final models to baseline models that use only demographic (age and gender)
5 data. The selected protein markers substantially improved the prediction of symptom severity,
6 spike protein binding IgG levels at day 28 and month 7, and memory T cell responses at day 28.
7 On the other hand, protein markers did not improve the prediction for oropharyngeal viral load.

8
9 We further tested if our model can accurately predict symptom severity in an independent
10 dataset. We identified a published dataset that characterized the plasma proteins from 58
11 COVID-19 patients (26 moderate cases and 34 severe cases)³⁶. Our model was able to
12 accurately identify severe cases in the independent dataset, achieving an AUC of 0.96. The
13 individuals in the test dataset already manifested severe symptoms while our training dataset
14 was collected before the severe symptom were shown, potentially explaining the higher model
15 performance in the test dataset than in the training dataset.



1
 2 Figure 7: Plasma protein markers predict symptom severity, T cell response, and S protein-binding IgG level in COVID-19
 3 patients. (A) machine-learning procedure for predicting COVID-19 patient outcomes using Olink proteomics data. (B)
 4 Random forest models were built to predict symptom severity, S protein-binding IgG level at 28 days and 7 months after
 5 enrollment, and cytokine+ memory CD4+ T cells 28 days after enrollment. The plot shows the leave-one-out cross-validation
 6 performance (measured by AUC of the receiver operator characteristics curve(ROC)) achieved by random forest models
 7 with different numbers of features. (C) The leave-one-out cross-validation performance of the best performing models and
 8 the models using demographical data (age and sex) only. (D) Feature importance of the final random forest models for
 9 predicting symptom severity, S protein-binding IgG level at 28 days and 7 months after enrollment, and cytokine+ memory
 10 CD4+ T cells at 28 days after enrollment. (E) We used the final model to predict severe cases in an independent dataset.
 11 The y axis of the plot shows the predicted probability. The x-axis shows the observed severity. (F) The performance of the
 12 final model measured by the ROC curve.

13
 14 **Discussion**

15 In this study, we longitudinally characterized the early immune response in patients who
 16 initially presented with mild to moderate COVID-19. With transcriptomic and proteomic profiling,
 17 we reveal a sequential activation of interferon signaling, NK cells, T cells and B cells within 2
 18 weeks of symptom onset. We also identified associations between early immune profiles and
 19 later clinical, virologic, and immunologic outcomes. These data suggest that variations in
 20 plasma RIG-I levels, early interferon signaling, and related cytokines (MCP1, MCP-2 and MCP-
 21 3) are associated with multiple aspects of patient outcomes, including symptom severity, viral
 22 shedding, and the SARS-CoV-2 specific T cell and antibody response measured up to 7 months

1 after enrollment. We also observed that the immune response after the first dose of SARS-
2 CoV-2 mRNA vaccination largely recapitulates the trajectory of immune response after SARS-
3 CoV-2 infection while the response to the second dose of vaccine is characterized by fast
4 upregulation of both early and late protein markers and absence of a neutrophil response.
5 Finally, we demonstrate that a machine learning model is able to predict symptom severity, T
6 cell memory response and antibody response accurately using 8-10 plasma protein markers.

7
8 We observed that high plasma RIG-I levels were associated with greater disease severity, T cell
9 activity, and the antibody response, suggesting that plasma RIG-I is a biomarker for increased
10 immune activity in COVID-19 patients. High plasma RIG-I levels were also associated with
11 lower SARS-CoV-2 viral loads, suggesting a potential role of this protein in restricting early virus
12 replication. RIG-I has been shown to be critically important in the response to several RNA
13 viruses, including influenza virus, typically via interactions with the adapter protein mitochondrial
14 antiviral-signaling protein (MAVS) and downstream Type I and Type III interferon upregulation.
15 RIG-I was recently shown to play an important role in both sensing SARS-CoV-2 RNA and
16 inhibiting SARS-CoV-2 replication in human lung cells, but not via downstream MAVS
17 induction³⁷. Rather, interactions between the RIG-I helicase domain and SARS-CoV-2 RNA
18 induced an inhibitory effect on viral replication, independent of downstream interferon
19 upregulation³⁷. This may explain the rather modest correlations observed between plasma RIG-I
20 and RIG-I signaling and interferon related pathways. In contrast, we observed significant
21 correlations between plasma RIG-I levels and plasma levels of DFFA, an intracellular protein
22 known to be involved in cell death²⁹, as well as other intracellular proteins, suggesting that
23 plasma RIG-I levels may reflect increased cellular apoptosis. This hypothesis is consistent with
24 a recent report which observed significant associations between gene expression signatures of
25 apoptosis in plasmacytoid dendritic cells with increased disease severity⁹.

26

1 Our analysis also suggests that the variations in the early immune response shape the long-
2 term outcome of COVID-19 patients. However, the cause of variations in the early responses
3 are not fully understood. We observed that higher expression of three CCR2 ligands (MCP1,
4 MCP2 and MCP3) were associated with multiple patient outcomes in COVID-19 patients,
5 including increased disease severity and higher T cell activity and S protein-binding IgG levels.
6 In addition MCP1 and MCP3 are negatively associated with oropharyngeal viral load. Our result
7 is consistent with previous studies that show CCR2 signaling is associated with symptom
8 severity in multiple viral infections, including SARS-Cov-2 and influenza^{38,39}. A previous GWAS
9 study has also identified an association between COVID-19 symptom severity and genetic
10 variations that leads to increased CCR2 expression (receptor to interferon induced MCP1,
11 MCP2 and MPC3)²⁶. On the other hand, mouse studies show that CCR2 is essential for the
12 survival of mice after pathogen challenge⁴⁰⁻⁴². Our study also shows that the CCR2 ligands are
13 associated with positive outcomes in patients with COVID-19, including reduced oropharyngeal
14 viral load, increased memory T cell and IgG antibody response. Taken together, the results
15 demonstrated the complex role of CCR2 signaling in regulating immune response. While it's
16 essential for an effective immune response, it also leads to severe symptoms and tissue
17 damage. Therapeutic strategies to balance the positive and negative effects of CCR2 may
18 benefit the management of COVID-19 patients.

19
20 Although naturally acquired SARS-CoV-2 infection results in protective antibody and T cell
21 immune responses, reinfections can occur, and the precise determinants driving susceptibility to
22 reinfection remain unclear⁴³. The BNT162b2 (Pfizer–BioNTech) vaccine has been shown to be
23 highly effective in preventing SARS-CoV-2 infection¹⁵, although breakthrough cases have been
24 increasingly reported since its approval⁴⁴. Comparing the vaccine response with the immune
25 response of natural infection may shed light on determinants of protective immunity to SARS-
26 CoV-2, and potential ways to improve COVID-19 vaccines. Our analysis reveals that the

1 proteomic response of the BNT162b2 vaccine mirrors in many ways the proteomic response
2 after SARS-CoV-2 infection, including the induction of protein markers that are positively
3 associated with T cell and antibody response in COVID-19 patients. At the same time, we found
4 important distinctions. In particular, proteomic signatures of a neutrophil response were absent
5 in response to the second dose of vaccine. Multiple studies have demonstrated the importance
6 of neutrophils in protection against both viral and bacterial reinfection^{45,46}, although it remains
7 unclear whether a neutrophil response may be beneficial, or detrimental, following vaccination.
8 In addition, we found that the vaccine also induced proteins that are negatively associated with
9 T cell and antibody response following natural infection, including TWEAK³⁵. Further studies will
10 be required to determine whether inhibiting TWEAK could potentially improve vaccine efficacy.

11
12 Our study has some limitations. First, while we identified multiple associations between early
13 immune measures and the outcome of COVID-19 patients, we did not establish causal
14 relationships between them. Future studies are needed to perturb key immune pathways in the
15 early immune response and test their effect on the patient outcomes. Second, our study
16 measured the immune response during the first 2 weeks of symptom onset in COVID-19
17 patients. Earlier immune responses between the initial infection and symptom onset have not
18 been characterized. This is due to the difficulty to detect pre-symptomatic COVID-19 infection.
19 Routine SARS-COV-2 monitoring in a select cohort will be required to acquire samples prior to
20 and immediately after the infection in order to assess whether pre-infection signatures predict
21 outcomes in COVID-19 patients. Third, our analysis focused on individual plasma proteins
22 (based on olink data) and immune related Gene Ontology pathways (based on RNA-seq data).
23 We used the Gene Ontology-based pathways to provide a high-level overview of the immune
24 response in COVID-19 patients. Caution should be taken when interpreting the Gene Ontology
25 pathways results, as the pathways are manually curated gene lists from literature and subject to
26 publication bias, curation errors and over-simplification of biological processes. We encourage

1 others to investigate the immune response of individual genes of interest using our shared RNA-
2 seq data. Finally, we have created machine learning models to predict multiple outcomes in
3 COVID-19 patients, including symptom severity, T cell response and antibody responses. While
4 we are able to validate the model for predicting symptom severity in an independent dataset,
5 additional datasets are needed to validate our model for predicting T cell and antibody
6 responses.

7
8 In this study, we identified multiple biomarkers for predicting clinical and immunological
9 outcomes in COVID-19 patients, including plasma level of RIG-I and the CCR2 ligands (MCP1,
10 MCP2 and MCP3). In addition, we demonstrate that machine learning models using 8-10
11 biomarkers are highly effective in predicting these outcomes. The models can potentially be
12 used to identify high-risk COVID-19 patients who will develop life-threatening symptoms, and to
13 predict the degree of immune memory development. In addition, these biomarkers and models
14 could also help explain variations in the response to COVID-19 vaccines, and to further identify
15 differences between natural infection and vaccine-induced immunity.

16

17 **Acknowledgements**

18 Support for the study was provided from NIH/NIAID (U01 AI150741-01S1 to ZH, ST, IRB, BG,
19 TW, and PJ), the Stanford's Innovative Medicines Accelerator, and NIH/NIDA DP1DA046089.
20 The Lambda clinical trial was funded by anonymous donors to Stanford University, and
21 Peginterferon Lambda provided by Eiger BioPharmaceuticals. The funders had no role in data
22 collection and analysis or the decision to publish.

23

24 We thank all study participants who participated in this study, the study team for their tireless
25 work, and Thanmayi Ranganath, Nancy Q. Zhao, Aaron J. Wilk, Rosemary Vergara, Julia L.
26 McKechnie, Giovanni J. Martínez-Colón, Arjun Rustagi, Geoff Ivison, Ruoxi Pi, Madeline J. Lee,

1 Taylor Hollis, Georgie Nahass, Kazim Haider, and Laura Simpson for assistance with
2 processing samples. We also thank our colleagues at Stanford University Occupational Health
3 and at San Mateo Medical Center for participant referrals. The Stanford REDCap platform
4 (<http://redcap.stanford.edu>) is developed and operated by Stanford Medicine Research IT team.
5 The REDCap platform services at Stanford are subsidized by a) Stanford School of Medicine
6 Research Office, and b) the National Center for Research Resources and the National Center
7 for Advancing Translational Sciences, National Institutes of Health, through grant UL1
8 TR001085.

9

10 **References**

- 11 1. Wu, Z. & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus
12 Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72,314 Cases From
13 the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239–1242 (2020).
- 14 2. Park, A. & Iwasaki, A. Type I and Type III Interferons - Induction, Signaling, Evasion, and
15 Application to Combat COVID-19. *Cell Host Microbe* **27**, 870–878 (2020).
- 16 3. Konno, Y. *et al.* SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is
17 Increased by a Naturally Occurring Elongation Variant. *Cell Rep* **32**, 108185 (2020).
- 18 4. Thorne, L. G. *et al.* Evolution of enhanced innate immune evasion by the SARS-CoV-2
19 B.1.1.7 UK variant. *bioRxiv* 2021.06.06.446826 (2021) doi:10.1101/2021.06.06.446826.
- 20 5. Arunachalam, P. S. *et al.* Systems biological assessment of immunity to mild versus severe
21 COVID-19 infection in humans. *Science* **369**, 1210–1220 (2020).
- 22 6. Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe
23 COVID-19 patients. *Science* **369**, 718–724 (2020).
- 24 7. Wilk, A. J. *et al.* A single-cell atlas of the peripheral immune response in patients with
25 severe COVID-19. *Nat Med* **26**, 1070–1076 (2020).
- 26 8. Abers, M. S. *et al.* An immune-based biomarker signature is associated with mortality in

- 1 COVID-19 patients. *JCI Insight* **6**, 144455 (2021).
- 2 9. Liu, C. *et al.* Time-resolved systems immunology reveals a late juncture linked to fatal
- 3 COVID-19. *Cell* **184**, 1836-1857.e22 (2021).
- 4 10. Lucas, C. *et al.* Longitudinal analyses reveal immunological misfiring in severe COVID-19.
- 5 *Nature* **584**, 463–469 (2020).
- 6 11. Schulte-Schrepping, J. *et al.* Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell
- 7 Compartment. *Cell* **182**, 1419-1440.e23 (2020).
- 8 12. Wilk, A. J. *et al.* Multi-omic profiling reveals widespread dysregulation of innate immunity
- 9 and hematopoiesis in COVID-19. *J Exp Med* **218**, e20210582 (2021).
- 10 13. Stephenson, E. *et al.* Single-cell multi-omics analysis of the immune response in COVID-19.
- 11 *Nat Med* **27**, 904–916 (2021).
- 12 14. Talla, A. *et al.* Longitudinal immune dynamics of mild COVID-19 define signatures of
- 13 recovery and persistence. *bioRxiv* 2021.05.26.442666 (2021)
- 14 doi:10.1101/2021.05.26.442666.
- 15 15. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New*
- 16 *England Journal of Medicine* **383**, 2603–2615 (2020).
- 17 16. Jagannathan, P. *et al.* Peginterferon Lambda-1a for treatment of outpatients with
- 18 uncomplicated COVID-19: a randomized placebo-controlled trial. *Nature Communications*
- 19 **12**, 1967 (2021).
- 20 17. Chakraborty, S. *et al.* Divergent early antibody responses define COVID-19 disease
- 21 trajectories. *bioRxiv* 2021.05.25.445649 (2021) doi:10.1101/2021.05.25.445649.
- 22 18. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29
- 23 (2000).
- 24 19. Lehmann, M. H. *et al.* CCL2 expression is mediated by type I IFN receptor and recruits NK
- 25 and T cells to the lung during MVA infection. *J Leukoc Biol* **99**, 1057–1064 (2016).
- 26 20. Lee, P. Y. *et al.* Type I interferon modulates monocyte recruitment and maturation in chronic

- 1 inflammation. *Am J Pathol* **175**, 2023–2033 (2009).
- 2 21. Dengel, L. *et al.* Differential effects of IFN- α and IFN- γ on upregulation of IFN-inducible
3 chemokines CXCL9, CXCL10 and CXCL11. *Cancer Res* **68**, 2149–2149 (2008).
- 4 22. Moll, H. P., Maier, T., Zommer, A., Lavoie, T. & Brostjan, C. The differential activity of
5 interferon- α subtypes is consistent among distinct target genes and cell types. *Cytokine* **53**,
6 52–59 (2011).
- 7 23. xCell: digitally portraying the tissue cellular heterogeneity landscape | Genome Biology | Full
8 Text. <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1349-1>.
- 9 24. Goyal, P. *et al.* Clinical Characteristics of Covid-19 in New York City. *New England Journal*
10 *of Medicine* **382**, 2372–2374 (2020).
- 11 25. Reyes-Gibby, C. C. *et al.* Oral microbiome and onset of oral mucositis in patients with
12 squamous cell carcinoma of the head and neck. *Cancer* **126**, 5124–5136 (2020).
- 13 26. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**,
14 92–98 (2021).
- 15 27. Matsumiya, T. & Stafforini, D. M. Function and Regulation of Retinoic Acid-Inducible Gene-I.
16 *Crit Rev Immunol* **30**, 489–513 (2010).
- 17 28. Thomas, D. A., Du, C., Xu, M., Wang, X. & Ley, T. J. DFF45/ICAD Can Be Directly
18 Processed by Granzyme B during the Induction of Apoptosis. *Immunity* **12**, 621–632 (2000).
- 19 29. Zhang, J. H. & Xu, M. DNA fragmentation in apoptosis. *Cell Research* **10**, 205–211 (2000).
- 20 30. Arunachalam, P. S. *et al.* Systems vaccinology of the BNT162b2 mRNA vaccine in humans.
21 *Nature* 1–10 (2021) doi:10.1038/s41586-021-03791-x.
- 22 31. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, (2015).
- 23 32. Jovic, S. *et al.* The neutrophil-recruiting chemokine GCP-2/CXCL6 is expressed in cystic
24 fibrosis airways and retains its functional properties after binding to extracellular DNA.
25 *Mucosal Immunol* **9**, 112–123 (2016).
- 26 33. Sawant, K. V. *et al.* Chemokine CXCL1 mediated neutrophil recruitment: Role of

- 1 glycosaminoglycan interactions. *Sci Rep* **6**, 33123 (2016).
- 2 34. Renshaw, S. A. *et al.* Acceleration of human neutrophil apoptosis by TRAIL. *J Immunol* **170**,
- 3 1027–1033 (2003).
- 4 35. Maecker, H. *et al.* TWEAK attenuates the transition from innate to adaptive immunity. *Cell*
- 5 **123**, 931–944 (2005).
- 6 36. Patel, H. *et al.* Proteomic blood profiling in mild, severe and critical COVID-19 patients.
- 7 *Scientific Reports* **11**, 6357 (2021).
- 8 37. Yamada, T. *et al.* RIG-I triggers a signaling-abortive anti-SARS-CoV-2 defense in human
- 9 lung cells. *Nat Immunol* **22**, 820–828 (2021).
- 10 38. Lin, K. L., Sweeney, S., Kang, B. D., Ramsburg, E. & Gunn, M. D. CCR2 Antagonist
- 11 Prophylaxis Reduces Pulmonary Immune Pathology and Markedly Improves Survival during
- 12 Influenza Infection. *J Immunol* **186**, 508–515 (2011).
- 13 39. Chen, Y. *et al.* IP-10 and MCP-1 as biomarkers associated with disease severity of COVID-
- 14 19. *Molecular Medicine* **26**, 97 (2020).
- 15 40. Pamer, E. G. Tipping the balance in favor of protective immunity during influenza virus
- 16 infection. *PNAS* **106**, 4961–4962 (2009).
- 17 41. Lim, J. K. *et al.* Chemokine Receptor CCR2 is Critical for Monocyte Accumulation and
- 18 Survival in West Nile Virus Encephalitis. *J Immunol* **186**, 471–478 (2011).
- 19 42. Defects in macrophage recruitment and host defense in mice lacking the CCR2 chemokine
- 20 receptor - PubMed. <https://pubmed.ncbi.nlm.nih.gov/9362535/>.
- 21 43. To, K. K.-W. *et al.* COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2
- 22 strain confirmed by whole genome sequencing. *Clin Infect Dis* ciaa1275 (2020)
- 23 doi:10.1093/cid/ciaa1275.
- 24 44. Birhane, M. *et al.* COVID-19 Vaccine Breakthrough Infections Reported to CDC — United
- 25 States, January 1–April 30, 2021. *MMWR Morb Mortal Wkly Rep* **70**, 792–793 (2021).
- 26 45. Fujisawa, H. Neutrophils Play an Essential Role in Cooperation with Antibody in both

- 1 Protection against and Recovery from Pulmonary Infection with Influenza Virus in Mice.
2 *Journal of Virology* **82**, 2772–2783 (2008).
- 3 46. Sjöstedt, A., Conlan, J. W. & North, R. J. Neutrophils are critical for host defense against
4 primary infection with the facultative intracellular bacterium *Francisella tularensis* in mice
5 and participate in defense against reinfection. *Infection and Immunity* **62**, 2779–2783
6 (1994).
- 7 47. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
8 quantification. *Nat Biotechnol* **34**, 525–527 (2016).
- 9 48. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv*
10 060012 (2019) doi:10.1101/060012.
- 11 49. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
12 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*
13 *(Methodological)* **57**, 289–300 (1995).
- 14
- 15
- 16 Methods
- 17 Lambda Study Design and Oversight
- 18 Data and samples were obtained from a Phase 2, single-blind, randomized placebo-controlled
19 trial to evaluate the efficacy of Lambda in reducing the duration of viral shedding in outpatients.
20 The trial was conducted within the Stanford Health Care System. Adults aged 18-65 years with
21 an FDA emergency use authorized reverse transcription-polymerase chain reaction (RT-PCR)
22 positive for SARS-CoV-2 within 72 hours from swab to the time of enrollment were eligible for
23 participation in this study. We included both symptomatic and asymptomatic patients based on
24 the previous finding that the detected infectious virus were similar in samples from
25 asymptomatic and symptomatic persons⁴⁴. Symptomatic individuals were eligible given the
26 presence of mild to moderate symptoms without signs of respiratory distress. Asymptomatic

1 individuals were eligible if infection was the initial diagnosis of SARS-CoV-2 infection. Exclusion
2 criteria included current or imminent hospitalization, respiratory rate >20 breaths per minute,
3 room air oxygen saturation <94%, pregnancy or breastfeeding, history of decompensated liver
4 disease, recent use of interferons, antibiotics, anticoagulants or other investigational and/or
5 immunomodulatory agents for treatment of COVID-19, and prespecified lab abnormalities. Full
6 eligibility criteria are provided in the study protocol. The protocol was amended on June 16th,
7 2020 after 54 participants were enrolled but before results were available to include adults up to
8 75 years of age and eliminate exclusion criteria for low white blood cell and lymphocyte count.
9 The trial was registered at ClinicalTrials.gov (NCT04331899). The study was performed as an
10 investigator-initiated clinical trial with the FDA (IND 419217), and approved by the Institutional
11 Review Board of Stanford University.

12

13 Participant Follow Up and Sample Collection

14 Participants completed a daily symptom questionnaire using REDCap Cloud version 1.5. In-
15 person follow-up visits were conducted at day 1, 3, 5, 7, 10, 14, 21, and 28, with assessment of
16 symptoms and vitals, and collection of oropharyngeal swabs (FLOQ Swabs; Copan
17 Diagnostics). Peripheral blood was collected at enrollment, day 5, day 28, and month 7 post
18 randomization. Whole blood was collected in Paxgene Tubes, and remaining blood was
19 processed for plasma and peripheral blood mononuclear cells.

20

21 Clinical Laboratory procedures

22 Laboratory measurements were performed by trained study personnel using point-of-care CLIA-
23 waived devices or in the Stanford Health Care Clinical Laboratory. Oropharyngeal swabs were
24 tested for SARS-CoV-2 in the Stanford Clinical Virology Laboratory using an emergency use
25 authorized, laboratory-developed, RT-PCR. Centers for Disease Control and Prevention
26 guidelines identify oropharyngeal swabs as acceptable upper respiratory specimens to test for

1 the presence of SARS-CoV-2 RNA, and detection of SARS-CoV-2 RNA swabs using
2 oropharyngeal swabs was analytically validated in the Stanford virology laboratory.

3

4 Whole blood transcriptomics

5 Whole blood transcriptomics were performed at Novogene Corporation, Inc. Briefly, whole blood
6 samples collected in Paxgene Tubes were first treated with Proteinase K, and then RNA
7 extraction performed using Quick-RNA MagBead Kit (R2132) on KingFisher followed by sample
8 quality control checks using a Qubit and Bioanalyzer 2100. Libraries were prepared using
9 ZymoSeq RiboFree Total RNA Library Kit (R3000). Sequencing took place on a Nova6000 on
10 an S4 lane, 30M paired reads, PE 150.

11

12 Whole blood transcriptomic data analysis

13 The transcript-level count data and transcript per million (TPM) data was calculated using
14 Kallist⁴⁷ (v0.46.2) and human cDNA index produced using kallisto on Ensembl v96
15 transcriptomes. We identified all gene ontology terms that are the child term of immune system
16 process (GO:0002376). We removed highly redundant gene ontology terms by grouping terms
17 with >80% overlap of genes and manually selected the representative terms within each group.
18 For each RNA-seq sample, we calculated the single-sample enrichment score of each gene
19 ontology term using the fgsea R package⁴⁸. The enrichment scores of the gene ontology terms
20 were normally distributed across samples and are treated as variables, similar to individual
21 protein markers, in the downstream analysis.

22

23 Plasma protein profiling using Olink panels

24 We measured proteins in plasma using Olink multiplex proximity extension assay (PEA)
25 inflammation panel and immune response panel (Olink proteomics, www.olink.com) according
26 to the manufacturer's instructions. The PEA is a dual-recognition immunoassay, where two

1 matched antibodies labeled with unique DNA oligonucleotides simultaneously bind to a target
2 protein in solution. This brings the two antibodies into proximity, allowing their DNA
3 oligonucleotides to hybridize, serving as a template for a DNA polymerase-dependent extension
4 step. This creates a double-stranded DNA “barcode” unique for the specific antigen and
5 quantitatively proportional to the initial concentration of target protein. The hybridization and
6 extension are immediately followed by PCR amplification and the amplicon is then finally
7 quantified by microfluidic qPCR.

8

9 T Cell Assays

10 SARS-CoV-2 specific T cell peptide pools were purchased from Miltenyi Biotec (PepTivator®
11 SARS-CoV-2 Prot_S, Prot_S1, Prot N, and Prot M) and resuspended in DMSO. These
12 PepTivator® reagents are pools of lyophilized peptides of 15 amino acid length with 11 amino
13 acid overlap, covering immunodominant sequence domains of the Spike (S and S1) (aa
14 sequence 1-1273), Nucleocapsid (N) or Membrane (M) proteins of SARS-CoV-2.

15

16 Antigen-specific T cell responses were measured using an intracellular cytokine staining assay.
17 Briefly, cryopreserved PBMCs were thawed, counted, and resuspended in complete RPMI
18 (RPMI (Corning) supplemented with 10% FBS (Gibco), 100 IU Penicillin (Corning), 100 ug/ml
19 Streptomycin (Corning), 1 mM HEPES (Corning) and 2 mM L-glutamine (Corning)). The cells
20 were plated in 96-well U bottom plates at 1×10^6 PBMCs per well and then rested overnight at
21 37°C in a CO₂ incubator. The following morning, cells were cultured in presence of either
22 SARS-CoV-2 peptides (1 µg/ml), PMA (300 ng/ml) and Ionomycin (1.5 µg/ml) as positive
23 control, or media as a negative control for 6 hours at 37°C. All conditions were in the presence
24 of brefeldin A (BD Pharmingen), monensin (BD Pharmingen), and CD107a. After a 6-hour
25 incubation, cells were washed and surface stained for CCR7 for 15 min at 37°C, followed by
26 the remaining surface stain for 30 min at room temperature (RT) in the dark. Thereafter, cells

1 were washed twice with PBS containing 0.5% BSA and 2 mM EDTA, then fixed/permeabilized
2 (FIX & PERM® Cell Permeabilization Kit, Invitrogen) and stained with intracellular antibodies for
3 20 min at RT in the dark. A complete list of antibodies are listed in Supplementary Methods. All
4 samples were analyzed on an Attune NXT flow cytometer and analyzed with FlowJo X (Tree
5 Star) software.

6

7 Antibody Assays

8 IgG antibody titers against the SARS-CoV-2 full length spike protein were assessed by enzyme-
9 linked immunosorbent assay (ELISA)²⁷. Briefly, 96 Well Half-Area microplates (Corning
10 (Millipore Sigma)) plates were coated with antigens at 2 µg/ml in PBS for 1h at RT. Next, the
11 plates were blocked for an hour with 3% non-fat milk in PBS with 0.1% Tween 20 (PBST).
12 Plasma was diluted fivefold starting at 1:50 in 1% non-fat milk in PBST. 25 µl of the diluted
13 plasma was added to each well and incubated for 2h at RT. Following primary incubation, 25 µl
14 of 1:5000 diluted horse radish peroxidase (HRP) conjugated anti-Human IgG secondary
15 antibodies (Southern Biotech) were added and incubated for 1h at RT. The plates were
16 developed by adding 25 µl/well of the chromogenic substrate 3,3',5,5'-tetramethylbenzidine
17 (TMB) solution (Millipore Sigma). The reaction was stopped with 0.2N sulphuric acid (Sigma)
18 and absorbance was measured at 450nm (iD5 SPECTRAmax, Molecular Devices). The plates
19 were washed 5 times with PBST between each step and an additional wash with PBS was done
20 before developing the plates. All data were normalized between the same positive and negative
21 controls and the binding AUC were calculated using GraphPad PRISM (Version 9).

22

23 Quantifying oropharyngeal viral load

24 We identified the cycle threshold (Ct) value using the fluorescence vs cycle data reported from
25 RT-PCR scanner. We subtract the Ct value from the detect limit (Ct=41) to quantify the viral

1 shedding in each OP swap. We plotted the viral shedding in each visit versus time, and
2 calculated the area under the curve using numerical integration based on the trapezoid rule.

3
4 Analysis of Olink data from the vaccine study
5 The olink data from the mRNA vaccine study was previously obtained and published ³⁰.We
6 tested if the level of the proteins are significantly altered after vaccination using ANOVA
7 (expression ~ time), where time is treated as a categorical variable to account for non-linear
8 behavior of the proteins. P values from the ANOVA models are adjusted using the False
9 Discovery Rate (FDR) method⁴⁹. To visualize the trajectory of the proteins, we imputed the
10 protein level in each day using linear interpolation with the 'approx' function in R.

11
12 Statistical analysis
13 Principal component analysis was conducted by applying the prcomp function in base R to the
14 whole Olink dataset or the top 500 genes with the highest variance. To access the association
15 between the principal components and clinical data, we fitted regression models (PC ~ clinical
16 variable). The percent of variances explained by the clinical variable is used to measure the
17 association.

18 We accessed the association between the expression of immune pathways or protein markers
19 with time using the regression model (expression ~ time + time²). It should be noted that our
20 study contains repeated measures of the same individuals in two time points (0 and 5 days after
21 enrollment). While including subjects as random effects in the regression model allows the
22 model to adjust for individual differences, it resulted in near-singular fits of the data for many of
23 the immune measurements. To avoid model over-fitting, we decided to only include the fixed
24 effects (time) in our model. To find significant associations, we compared the model with the
25 base model (expression ~ 1) and used the F test to calculate the p value. We adjusted the p
26 value using the False Discovery Rate (FDR) method. We performed a parallel analysis using

1 mixed-effect models [expression ~ time + time² + subject ID(random effect)] to fit the data and
2 found that all significant (FDR<0.05) variables identified using the fixed effect model were also
3 significant in the mixed-effect model (Supplemental Table 2).

4
5 We estimated the enrichment score of the major immune cell types using the xCell package²³.
6 The association between the xCell scores and time were tested using the same regression
7 method described above.

8
9 We tested the association between immune measurements and symptom severity using
10 regression models (measurements ~ symptom severity + time + time²) and the lm function in R.
11 The p value of the symptom severity term is adjusted using the FDR method. Similar regression
12 models were used to test the association between immune measurements and other outcomes,
13 including the oropharyngeal viral load, the memory CD4+ T cell activity and S protein binding
14 IgG levels. To test between immune measurements and symptom severity without adjusting the
15 time to symptom onset, we performed a one-way anova analysis using the lm function in R
16 (measurements ~ symptom severity) .

17
18 To estimate the time in which an immune measurement reaches the maximum level, we first fit
19 a quadratic regression model (measurement ~ time + time²). We then identified the day
20 (between 0 - 15 after symptom onset) in which the fitted regression model reached the maxim.
21 We repeated the process 100 times to estimate the variance of the time.

22
23 Predictive modeling

24 We used the protein measurements (measured by Olink assays) to predict the clinical,
25 virological and memory T cell activity and IgG antibodies. Since the outcomes are a mixture of
26 categorical (symptom severity) and continuous (viral load, T cell and antibody responses)

1 variables, we framed the all prediction tasks as classification problems by dichotomizing the
2 continuous variables using median as cutoffs. To prevent overfitting, we selected 30 proteins
3 with the highest variance as input data, as the highly variable proteins best capture the inter-
4 subject difference across the COVID-19 patients. We further select features using random forest
5 and leave-one-out validation. In step 1, we train a random forest model using data from all
6 samples but one left out sample. In step 2, we rank the feature importance of the 30 protein
7 markers based on the gini index reported by the random forest model. In step 3, We train
8 reduced random forest models with 1-29 most important proteins. In step 4, we predict the
9 outcomes using the data from the left out sample. We repeat steps 1-4 until we predict the
10 outcome of all samples. We calculate the model performance using the area under the receiver
11 operator characteristic curve (AUC). The variable combinations that give rise to the highest AUC
12 are selected as the optimal model. The optimal model for predicting symptom severity was
13 tested using Olink data from two independent studies.

14

15 Data availability

16 The RNA-sequencing data, Olink, clinical, virological, and immunological, as well as the
17 machine learning models, are available upon request.

18 All codes for data analysis are available upon request.

19

20