Host-Microbiome Associations in Saliva Predict COVID-19 Severity

Hend Alqedari*[1,2], Khaled Altabtbaei*[3], Josh L. Espinoza[6], Saadoun Bin-Hasan[4], Mohammad Alghounaim[5], Abdullah Alawady[4], Abdullah Altabtabae[4], Sarah AlJamaan[4], Sriraman Devarajan[2], Tahreer AlShammari[2], Mohammed Ben Eid[4], Michele Matsuoka[6], Hyesun Jang[6], Christopher L. Dupont[6], Marcelo Freire[6,7]

Affiliations

[1]Department of Oral Health Policy and Epidemiology, Harvard School of Dental Medicine, Boston, MA, 02115, USA; Dasman Diabetes Institute, Kuwait

[2]Dasman Diabetes Institute, 1180, Dasman, Kuwait

[3]School of Dentistry, Faculty of Medicine and Dentistry. University of Alberta. Edmonton AB, T6G 2L7, Canada

[4]Department of Pediatrics, Farwaniyah Hospital, Ministry of Health, Kuwait

[5]Department of Pediatrics, Amiri Hospital, Ministry of Health, Kuwait

[6]Department of Genomic Medicine and Infectious Diseases, J. Craig Venter Institute, La Jolla, CA 92037, USA

[7]Division of Infectious Diseases and Global Public Health Department of Medicine, University of California San Diego, La Jolla, CA, USA

Corresponding Author*

Marcelo Freire, D.D.S., Ph.D., D.Med.Sc.
Associate Professor
Genomic Medicine and Infectious Diseases
J. Craig Venter Institute
4120 Capricorn Lane, La Jolla, CA 92037, USA
Phone: 858-200-1846
Fax: 858-200-1880
Email: mfreire@jcvi.org

Author Contributions: Hend Alqaderi: conception and design, methodology, data acquisition, writing and data interpretation. Khaled Altabtbaei: software, formal analysis, data curation, revising and editing. Saadoun Bin-Hasan and Mohammad Alghounaim: data interpretation, revising and editing. Abdullah Alawady and Sriraman Devarajan: Data curation, data analysis, revising and editing. Abdullah Altabtabae, Sarah AlJamaan, and Tahreer AlShammari: Data collection. Mohammed Ben Eid: Data curation. Michele Matsuoka, Hyesun Jang, Josh Espinoza, Christopher Dupont: data analysis, software, data curation and interpretation, revising and editing. Marcelo Freire: conceptualization and design, methodology, data acquisition, writing original draft, visualization, data interpretation, and critically revised the manuscript. All authors gave final approval and agreed to be accountable for all aspects of the work.

Competing Interest Statement: All authors report there are no competing interests to disclose.

Classification: Biological Sciences/Immunology and Inflammation

Keywords: COVID-19, saliva microbiome, inflammatory cytokines, machine learning, network analysis

1

1

Abstract

Established evidence indicates that oral microbiota plays a crucial role in modulating host immune responses to viral infection. Following Severe Acute Respiratory Syndrome Coronavirus 2 – SARS-CoV-2 – there are coordinated microbiome and inflammatory responses within the mucosal and systemic compartments that are unknown. The specific roles the oral microbiota and inflammatory cytokines play in the pathogenesis of COVID-19 are yet to be explored. Here, we evaluated the relationships between the salivary microbiome and host parameters in different groups of COVID-19 severity based on their oxygen requirement. Saliva and blood samples (n = 80) were collected from COVID-19 and from non-infected individuals. We characterized the oral microbiomes using 16S ribosomal RNA gene sequencing and evaluated saliva and serum cytokines and chemokines using multiplex analysis. Alpha diversity of the salivary microbial community was negatively associated with COVID-19 severity, while diversity increased with health. Integrated cytokine evaluations of saliva and serum showed that the oral host response was distinct from the systemic response. The hierarchical classification of COVID-19 status and respiratory severity using multiple modalities separately (i.e., microbiome, salivary cytokines, and systemic cytokines) and simultaneously (i.e., multi-modal perturbation analyses) revealed that the microbiome perturbation analysis was the most informative for predicting COVID-19 status and severity, followed by the multi-modal. Our findings suggest that oral microbiome and salivary cytokines may be predictive of COVID-19 status and severity, whereas atypical local mucosal immune suppression and systemic hyperinflammation provide new cues to understand the pathogenesis in immunologically naïve populations.

Significance Statement

The oral mucosa is one of the first sites encountered by bacterial and viral infections, including SARS-CoV-2. It consists of a primary barrier occupied by a commensal oral microbiome. The primary function of this barrier is to modulate immunity and provide protection against invading infection. The occupying commensal microbiome is an essential component that influences the immune system's function and homeostasis. The present study showed that the host oral immune response performs unique functions in response to SARS-CoV-2 when compared to systemic responses during the acute phase. We also demonstrated that there is a link between oral microbiome diversity and COVID-19 severity. Additionally, the salivary microbiome was predictive of not only disease status but also severity.

Main Text

Introduction

The oral mucosa is one of the first sites where infections are encountered and is one of the primary barriers maintaining a low-level inflammatory response to the resident commensal microbiome[1]. This barrier effectively regulates immunity and protects the body against the breakdown of defense mechanisms leading to the development of oral or systemic diseases[1]. This defense mechanism involves complex interactions between the immune system, the microbiome, and the physical barriers in the oral cavity, such as saliva and mucosal tissue[2]. A balanced healthy microbiome is an essential component of the host and has often been considered an organ and protective component of the human body[3], which is involved in maintaining homeostasis, epithelial cell proliferation and differentiation, and in the regulation of immune response[4]. Well-established evidence showed that the immune defense of the mucosal barrier is built independently of commensal colonization and that disruption of homeostasis between the oral microorganisms leads to certain diseases[1]. The interplay between the microbiome and the host forms a complex ecosystem that serves as both a metabolic and physical barrier to outcompete external microorganisms[3], such as viral invaders.

The impact of the mucosal microbiome during coronavirus disease 2019 (COVID-19) is an emerging area with ongoing research. Recent evidence suggests that the severity of COVID-19 was related to microbiome changes, as SARS-CoV-2 replicates in cells expressing angiotensin-converting enzyme 2 (ACE2) and transmembrane protease serines 2 (TMPRSS2) receptors, both of which are found in the oral and nasal mucosa(4). These ongoing investigations are starting to reveal the relationship between the oral microbiome and disease pathogenesis. The decrease in oral microbial diversity was correlated to COVID-19 patients in ICU(5), while others finding that no changes in the nasal microbiome in mildly symptomatic COVID-19-positive individuals compared to controls(6). Although evidence is emerging, there are no molecular cues with predictive potential to understand symptoms from SARS-CoV-2 and risk to severity. While systemic immune responses can be monitored through blood sampling, it does not depict the mucosal responses from a host-microbial perspective(7). Emerging new immune response evidence (present in fluids such as saliva and plasma) has started to demonstrate changes in saliva from subjects that experienced and resolved from SARS-CoV-2 infection (8-12), yet cytokine responses is sparse(8, 9, 11, 12). It is consequently important to document molecular signatures in COVID-19 patients at the mucosal level to better define normal and pathogenic convalescence processes and to detect the possible initiation of aberrant innate immune activation, especially in fluids that are in direct contact with SARS-CoV-2. Although systemic immune responses can be monitored through blood sampling, understanding oral immune competence is complex(13).   Here, in a case-control study, we measured and compared SARS-CoV-2 oral mucosal responses in matched saliva and plasma samples and investigated how the host-microbiome in saliva can predict COVID-19 status and severity. We characterized host functional cytokines impaired by SARS-CoV-2 infection by comparing responses in acute COVID-19 subjects to healthy controls and by comparing responses in saliva to plasma. This case-control study aimed to evaluate whether host-microbiome in saliva can predict COVID-19 status in addition to severity. Overall, our findings highlight saliva as a stable and accessible biofluid to monitor SARS-CoV-2 host impact, including microbiome and cytokines associated with of disease severity. Saliva has the potential to reveal fundamental biological mechanisms of health and disease.

Results

Dysregulated systemic immune responses to SARS-CoV-2 infection impact COVID-19 disease severity, mortality, and the development of chronic diseases. Both hyperinflammatory systemic r and immune suppression responses have been found as viral evasion strategy. As body fluids are exposed to different antigens, we investigated how the body responds to SARS-CoV-2 in saliva compared to those in matched serum. Paired saliva and serum samples from COVID-19 donors in the acute phase were collected and subjected to comparative analyses among demographic factors (age, gender, and initial COVID-19 disease severity) (Table 1).

A total of 80 individuals were included in the study, consisting of 50 SARS-CoV-2 positive individuals confirmed by RT-PCR from nasopharyngeal swabs and 30 non-infected individuals. The severity of the disease was categorized as mild (hospitalized without oxygen therapy, n=11), moderate (hospitalized with low-flow oxygen <10 L/min, n=28), and severe (hospitalized with high-flow oxygen >10 L/min, n=11). The analysis combined the mild and moderate groups into a "non-severe" category. Of the participants, 55% were males and 45% were females. Among the SARS-CoV-2 positive group, 56% were older than 50, and 64% were older than 50, among the severe cases. Diabetes was present in 44% of the SARS-CoV-2 positive group, while only 9% of the subjects with severe symptoms had diabetes. Hypertension was present in 44% of SARS-CoV-2-positive individuals and 36% of those with severe symptoms. Obesity was observed in 37% of SARS-CoV-2-positive individuals and 40% of those with severe symptoms. All subjects with severe symptoms had a chronic cough.

Respiratory distress was reported by 82% of the SARS-CoV-2 positive group, while shortness of breath was reported by 82% of the severe cases at the time of data collection. Additionally, 45% of subjects with severe symptoms experienced persistent chest pain, 36% had sputum production, and 55% had a loss of taste and smell. Only 18% of subjects with severe symptoms had fever.

3

1   The demographic variables are presented in table 1.
2
3   *Microbial composition stratified samples by COVID-19 diagnosis better than cytokines*
4
5   We evaluated the microbiome composition detected in saliva to elucidate the effect of disease on
6   the microbiome. First, we conducted a principal coordinates analysis on Jaccard dissimilarity,
7   which showed distinct clusters between control, mild/moderate cases, and severe cases along
8   the first PC-axis, indicating that COVID-19 affects the microbiome composition during the acute
9   phase of infection (p<0.05, ADONIS on Jaccard distances with FDR correction, figure 1A). A
10   similar pattern was not seen in either salivary or blood cytokines (p>0.05, ADONIS,
11   Supplementary Figure S1). To understand what drives this microbial-enhanced separation
12   between the different conditions, we examined the alpha diversity. The control group had
13   statistically higher diversity compared to the mild/moderate group, which in turn had statistically
14   higher diversity than the severe group (p<0.05, observed features after FDR correction, figure 1B,
15   supplementary figure S2). The abundance-aware Shannon diversity demonstrated that the
16   relationship between the mild/moderate and control groups was not statistically significant
17   (P>0.05, figure 1C) therefore, the difference in observed ASVs was driven by rarer taxa. We next
18   conducted a principal coordinates analysis on Jaccard dissimilarity, which showed distinct
19   clusters between control, mild/moderate cases, and severe cases along the first PC-axis,
20   indicating that, indeed, COVID-19 affects the microbiome composition during the acute phase of
21   infection (p<0.05, ADONIS on Jaccard distances with FDR correction, figure 1A). To examine this
22   effect further, we broke Jaccard dissimilarity into its two components; replacement of taxa with
23   another (turnover) and whether one sample is a subset of another (nestedness)(14). Nestedness
24   was statistically significant between control and mild/moderate cases, but not turnover (p=0.012
25   and 0.299) respectively, ANOSIM after FDR correction). On the other hand, the difference
26   between mild/moderate and severe cases was due to turnover and not nestedness (p=0.003 and
27   0.365). Composition-aware quantitative analysis of the microbiome also showed a similar pattern
28   of clustering between control, mild/moderate and severe cases (p<0.05, ADONIS of PhILR
29   distances after FDR correction, figure 1D). In addition, alpha rarefaction curves demonstrated that
30   sufficient sample sequencing was achieved therefore the effect is not due to insufficient
31   sequencing depth (supplementary figure S3).
32
33   To gain insight into the capability of the oral variables in distinguishing COVID-19 status, we used
34   unsupervised clustering (Ward hierarchical clustering) on the samples using their salivary
35   microbiota and the salivary cytokines. The results revealed that the salivary microbiota clustered
36   most of the COVID-19-positive patients along the first tree bifurcation, while the salivary cytokines
37   were not able to show the same distinction (Supplementary figure S4). This finding was reflected
38   in the predictive accuracy of each variable, where the microbiota outperformed the cytokines
39   (Supplementary figure S4).
40
41   *Blood and saliva cytokines reflect different host responses to COVID-19*
42
43   We profiled saliva and serum samples to characterize oral mucosal and systemic responses
44   following SARS-CoV-2 infection more comprehensively. Out of a 65-cytokine assay, 15 salivary
45   and 45 blood cytokines were significantly different between SARS-CoV-2 positive and negative
46   patients (p<0.05, Kruskal-Wallis test, figure 2). Only 9 of these cytokines were shared between
47   the two profiles Furthermore, the salivary Hepatocyte Growth Factor (HGF) and Fibroblast
48   Growth Factor 2 (FGF2) were statistically significantly lower in COVID-19-positive patients
49   compared to the controls (<0.05), but not in blood. Additionally, salivary TNF-beta (Tumor
50   Necrosis Factor beta), IL10 (Interleukin 10), MCP2/CCL8 (Monocyte Chemotactic Protein-2/C-C
51   motif chemokine ligand 8), ENA78/CXCL5 (Epithelial cell-derived neutrophil-activating peptide
52   78/C-X-C motif chemokine ligand 5), CD40L (CD40 Ligand), IL2R (Interleukin 2 Receptor),
53   IL12p70 (Interleukin 12), and HGF (Hepatocyte Growth Factor) were found to be expressed less
54   in the mild/moderate condition compared to either severe or control (Supplementary figures S5-
55   6), possibly indicating an inverse relationship where the mucosal cytokine expression is
56   suppressed. We then determined the influence of the immune subclusters in relation to each fluid

4

by investigating correlations. To confirm this, we examined the relationship between blood and salivary cytokines in each condition, which confirmed that while the relationship in non-SARS-CoV-2 infected individuals was positive, the inverse relationship between salivary and blood cytokines is present in severe cases (Figure 2 C-F).

Our results indicate that measuring only serum levels during the COVID-19 convalescent phase does not provide a full picture of the host response mounted after SARS-CoV-2 infection. Indeed, while we confirmed that our acute COVID-19 subjects had mounted a systemic response in blood, our salivary sequencing analysis revealed novel immune signatures and altered microbiome in disease when compared to healthy controls. Overall, we demonstrated that population-based investigations of saliva can be used to map global host-microbial responses to local mucosa and begin to integrate to systemic functions.

Microbial perturbation and cytokines serve as a robust classifier for COVID-19

We next investigated the feature selection algorithm (*Clairvoyance*(15)) in a hierarchical manner (Figure 3A) to better understand the strength of the signal that the microbial perturbation and cytokine fluctuations are serving as a predictive/diagnostic tool for SARS-CoV-2. We examined the strength of the potential sources of salivary and blood biomarkers and their capability in four different manners: salivary microbiome only, blood cytokine only, salivary cytokine only, and multi-modal sample-specific perturbation networks (SSPN) including the three parameters stratified by COVID-19 status and severity. Feature selection analysis was performed on y1) uninfected vs. infected (N= 80 samples); and y2) mild/moderate vs. severe (N=50 samples). Our feature selection ended up with 42 features in y1 and 39 features in y2 which are fewer than the number of samples used for modeling and, thus, not subject to bias due arising from the curse of dimensionality(16). The feature selection results are shown in figure 3B and 3C. We demonstrated that the microbiome-only model had the highest classification capabilities, followed by the multi-modal SSPN (microbiome: 99.4%, 100%, multi-modal SSPN: 99.5%, 91.2% for submodal y1 and y2, respectively) compared to the salivary cytokine and blood cytokine alone models. The potential markers for the microbiome did not overlap between the two submodels, with y1 being represented by 10 ASVs, and y2 being represented by 6 ASVs (Supplementary table S2). Interestingly, most nodes in the SSPN were shared across the two submodels, inferring that the interactions (edges) between these nodes are drivers to the distinction between the two submodels. None of the microbial biomarkers were shared between sub-modals y1 (N = 10 ASVs) and y2 (N = 6 ASVs) while all the other modalities shared features between sub-modals y1 and y2 (Supplementary tables S2, S3). Central to these interactions is the systemic cytokine CXCL10/IP-10.

A major advantage of analyzing edge perturbations using feature selection algorithms is the ability to reconstruct edges into a network providing an opportunity to reap the benefits of graph theory. In this context, we reconstructed our aggregate networks (AN) using the multi-omic biomarkers and their fitted model weights (i.e., logistic regression coefficients) as edge weights with sub-models y1 and y2 representing AN_{y1} and AN_{y2}, respectively (Figure 3). As the edge weights represent model coefficients, they can be directly interpreted as their predictive capacity in classifying stepwise severity with positive coefficients translating to a higher likelihood of a sample to be classified as more severe and vice versa for negative coefficients.

Comparing overlap in network context can be performed at the node and edge level. For AN_{y1} and AN_{y2} there was a higher proportion of nodes shared (57.6%) compared to shared edges (20.9%) between both sub-models, indicating a common set of biomarkers useful for diagnosing COVID-19 phenotypes but the way those biomarkers interact change with severity.

 In both AN_{y1} and AN_{y2}, the most notable node is the systemic cytokine CXCL10/IP-10 which has disproportionately high connectivity relative to the other nodes and is highly centralized (Figure 4, S7). AN_{y1} and AN_{y2} have similar network structures but AN_{y1} node degree distribution is less homogenous than AN_{y2} with homogeneity indices of 80.7% and 88.3%, respectively. In other

1   words, the predictive capacity of systemic cytokine CXCL10/IP-10 is disproportionately higher
2   than the other nodes in classifying COVID-19 diagnosis (sub-model y1) relative to COVID-19
3   severity (sub-model y2). In sub-model y2, the classification of COVID-19 severity is more evenly
4   distributed across the other nodes (Figure S7).
5
6   For AN_{y1} the next highest degree nodes are ASV017 (Veillonella sp.), ASV019 (Pasteurellaceae
7   sp.), systemic HG, and ASV039 (Corynebacterium matruchotii) and, along with CXCL10/IP-10,
8   these 5 nodes represent 59.1% of the total predictive capacity in sub-model y1 in classifying
9   whether a patient has been diagnosed with COVID-19 at an accuracy of 99.5% (Figure S7, S8).
10 For AN_{y2} the next highest degree nodes are ASV015 (Gemella sp.), systemic CCL24|Exotaxin2,
11 ASV019 (Pasteurellaceae sp.), ASV013 (Veillonella sp.), and systemic HG and, along with
12 CXCL10/IP-10, these 6 nodes represent 58.4% of the total predictive capacity in sub-model y2 in
13 classifying COVID-19 severity at an accuracy of 91.2% (Figure S7B).
14
15
16 **Discussion**
17 Increasing evidence indicates that the immune system of COVID-19 subjects may be
18 compromised when compared to healthy controls(17). Most investigations have focused on
19 systemic body compartments such as blood and lung, and emerging investigations are dissecting
20 the nasal mucosal responses. We demonstrated that inflammatory response to viral infection is
21 regulated by microbiome changes locally and hypothesize that systemic responses are not
22 predictive of systemic changes. The present analysis demonstrated that the salivary microbial
23 community is highly susceptible to perturbation in the presence of SARS-CoV-2, leading to a
24 distinct separation of the salivary microbiome between healthy and COVID-19-infected
25 individuals. We also showed that the salivary microbiome was distinct between individuals with
26 severe and mild/moderate symptoms. Our findings also demonstrate that abnormal inflammatory
27 responses can be identified in both saliva of COVID-19 subjects, which was shown to be severity
28 specific. This indicates that even when SARS-CoV-2 systemic responses are mounted, the local
29 immune response does not necessarily define disease resolution. Here, we highlight saliva as an
30 important and accessible fluid that can be monitored to identify not just antibody responses, but
31 also diverse host-microbial pathways, including mucosal immunity, and innate immune
32 responses.
33
34 The variation of the microbial membership detected in individuals with severe and mild/moderate
35 symptoms can be attributed to the replacement of taxa with other taxa (turnover). However, the
36 separation between groups (health versus disease) was mainly attributed to different phenomena;
37 we found a decrease in the number of salivary bacterial members in the mild/moderate cases
38 compared to the control cases, with the taxa of mild/moderate being a subset of the control. This
39 decrease was not seen when adjusting for the overall abundance of the microbiome (Shannon
40 diversity). Therefore, our findings suggest that the transition from control to mild/moderate
41 disease status preserves most of the microbial membership and is mainly driven by members
42 with higher abundance; that is, the loss of microbial membership in the mild/moderate cases
43 occurred in less abundant taxa. This demonstrates that the salivary microbiome is generally well-
44 adapted to perturbations, where these bacteria can maintain the overall community structure
45 despite the change in the mucosal inflammatory environment during the acute phase of infection.
46 Conversely, the transition from mild/moderate to severe disease corresponds to a larger
47 environmental stress, causing a further collapse in microbial richness, thus allowing for the
48 replacement of some taxa with others (turnover). However, when accounting for the overall
49 collapse in membership and richness, this replacement of taxa was not enough to reverse the
50 trend of reduced alpha diversity in severe cases.
51
52 Carefully designed immune studies aimed at implementing cytokine testing by investigations of
53 blood-derived fluids, but not saliva. Although it is well-known that mucosal and systemic immunity
54 responds differently to pathogens(18), this study adds to the evidence that each immunity system
55 responds in a distinct manner to SARS-CoV-2 locally. Our findings showed that the oral response
56 was distinct from the systemic response. The present study demonstrated that systemic cytokines

showed higher levels in mild/moderate cases compared to control cases, which is expected. However, salivary cytokine levels were lower in mild/moderate cases compared to control cases. Even though the wealth of evidence agrees with the fact that cytokine production increases during the disease's status, it was suggested that reducing cytokine production can be a strategy to manage cytokine storms and other inflammatory reactions[18]. Thus, oral mucosal surfaces might lower cytokines in certain situations to balance between their two opposing roles: fighting pathogens and immune surveillance. It is worth noting that while saliva and serum samples were collected at the same time point from individuals within 48 hours of qualitative PCR-confirmed nasal pharyngeal COVID-19 diagnosis, delays in processing samples can lead to changes on the immune content. A study showed, that sample processing delay of plasma or serum caused perturbations of numerous cytokines (range -10.8 to 43.5%), frequencies of peripheral blood cytokines and gene expression [10]. In our study, while there is variability on the recruitment of subjects, we kept an uniform collection protocol of saliva and plasma with fast processing of samples prior to freezing and storage. When normalizing cytokine concentration, we did not detect major variance across the subjects within the groups.

Data collection was strategically conducted during standard working hours, from 8 am to 2 pm, to minimize variations in circadian rhythms. The timing of our data collection was chosen to reduce variability and to specifically exclude nighttime sampling, thereby enhancing the reliability of our findings. Both saliva and blood samples were collected at the same time, thereby capturing both local and systemic responses. Current evidence does not conclusively support the idea that circadian rhythms differentially affect cytokines in saliva compared to blood. Most research in this area has focused on blood samples and specific cytokines, such as tumor necrosis factor (TNF) and IL-6, with less attention paid to saliva [19].

We further demonstrated that salivary interleukin 8 responses to SARS-COV-2 could be involved in neutrophil interactions at the oral mucosal surface. Significant correlation between salivary IL8 with blood markers, such as IL23, IL2, IL15, GROalphaCXCL1, IL12p70, IL6, TNFalpha, IL18, SDF1alpha, IL13, and IFNalpha demonstrate the tight coupling required for immune surveillance where a variety of cytokines are expressed for immune-cellular recruitment and activation of innate immune cells. This relationship is not maintained during COVID-19 infections. Additionally, we showed that the oral mucosal response was not linear; for example, salivary Interferon gamma-induced protein 10 (IP-10) demonstrates lower expression in mild/moderate conditions compared to healthy and severe cases. The importance of IP10, a pro-inflammatory chemokine for T-lymphocytes, monocytes, and NK cells, has been shown in other viral infections such as HIV and HBV infections. For example, intra-organ (liver) IP10 mRNA expression and the IP10 blood levels are concurrent in patients with chronic hepatitis B infections][20]. While HIV disease progression is positively correlated with levels of IP10 in circulation[21]. the positive relationship between infection level and IP10 was not found in our study, thus indicating a different response path to infection. Understanding the effects of these cytokines helps us understand the host response to the infection. IL-10 and CD40L are responsible for IgA class switching of immunoglobulins[22, 23]. salivary IL12 was also reduced in mild/moderate cases compared to the severe group. IL12, a critical cytokine for the activation of dendritic cells, B-Cells and macrophages, stimulates the differentiation of T cells into Th1 cells. IL-12 also promotes the production of interferon-gamma (IFN-γ).

Taken together, the salivary cytokine profile of mild/moderate individuals supports the notion that the infection has not switched to a large-scale adaptive immune response to the virus. Further, it was suggested that the innate immune system in the oral cavity acts as the first line of defense against invading microorganisms[24], and several components of the innate immune system are involved in oral immunity, including saliva, oral epithelial cells, and other oral immune cells[24]. Saliva acts as a component of innate immunity with antimicrobial activity[1], and oral epithelial cells are actively implicated in immune regulation, promote defensive immune responses, and create a constitutive tolerogenic environment that maintains immune homeostasis[25].

1   The relationship between the oral cavity and systemic conditions is often neglected, despite
2   ample evidence of its connection to a multitude of diseases in distant organs, such as
3   pneumonia(26), low birth weight(27), obesity(28), diabetes(29), and coronary disease(30). The
4   pharynx structurally bridges the oral and nasal cavities, both of which express the
5   ACE2/TMPRRS2 receptors that play a fundamental role in the SARS-CoV-2's ability to infect
6   these niches. In addition, it also contains pharyngeal tonsil tissues, an instrumental induction site
7   for oral mucosal IgA response to respiratory viral infection. The salivary gland also plays a major
8   role as an effector site to produce T-cell-dependent antigen-specific oral IgA response. ACE2
9   receptors on highly differentiated epithelial cells are also strongly associated with cytokine levels
10 in these tissues(31, 32). Furthermore, mouse models have demonstrated that the resident
11 microbiota on mucosal surfaces can protect against rotavirus infections(33), and it has been
12 suggested to impact SARS-CoV-2 infections(34). The state of the mucosal tissues, such as the
13 oral cavity, by way of perturbation of the resident microbiome and local cytokine profile, reflects
14 the status of the host and can be used to predict and monitor the overall response. These
15 perturbations should be further understood to gain insight into the host response of the mucosal
16 surfaces of the upper respiratory tract.
17
18 While periodontal disease is known to influence levels of salivary biomarkers (35), conducting a
19 comprehensive periodontal examination during the COVID-19 pandemic posed significant
20 challenges. To circumvent this, our study employed a proxy questionnaire designed to assess
21 indicators of periodontal health, including bone loss, gum inflammation, and the presence of loose
22 teeth. This approach enabled us to establish a binary variable, "periodontal status," which
23 classified subjects based on their self-reported oral health conditions. However, upon conducting
24 a chi-square analysis, as detailed in Supplementary Tables 2 and 3, we found no significant
25 differences in periodontal status between subjects infected with COVID-19 and those who were
26 uninfected (p value > 0.05). Additionally, our analysis revealed no statistically significant variation
27 in periodontal status across different COVID-19 severity groups. This finding suggests that, within
28 the limitations of our study, periodontal disease as assessed by our questionnaire method may
29 not be a differentiating factor in COVID-19 infection or its severity (35).
30
31 Future studies would benefit from requiring convalescent COVID-19 subjects to report possible
32 chronic symptoms longitudinally. The use of saliva for health monitoring has received significant
33 attention due to its ease and simplicity of use. This study sought to explore this concept in relation
34 to the extreme case of an exogenous infection that has recently jumped into immunologically
35 naive humans, making it the ideal scenario for such examinations. This is particularly relevant
36 given the evidence that depending on CT values produced by RT-PCR reflect the host's
37 capability to clear the virus and not the infectiousness potential of the shedding particles
38 themselves(34, 36, 37). By closely studying this unique case, the potential of saliva monitoring as
39 a diagnostic tool can be further understood and improved upon.
40
41 In summary, our study has found the salivary microbiome to be a promising source to predict
42 COVID-19 status and severity and that the oral microbiome plays an important role in the immune
43 response by stimulating or suppressing the immune system. Our study supports the notion that
44 the salivary microbiome could be used as a potential diagnostic tool for COVID-19 and predict
45 severity(38). While these studies are compelling, more research is needed to determine the
46 clinical utility of using the salivary microbiome as a diagnostic tool for COVID-19. However, the
47 potential for a non-invasive, rapid, and convenient diagnostic test using salivary microbiome
48 analysis is exciting and warrants further investigation.
49
50 *Limitations*
51 This study had a small sample size and may limit our ability to detect the significant distinction
52 between different groups of symptom severity. Our COVID-19 cohort included only symptomatic
53 subjects with mild, moderate, and severe symptoms. Asymptomatic individuals with COVID-19
54 may exhibit different microbiome characteristics and might be included unintentionally in the
55 control uninfected group. In addition, the salivary microbiome can be influenced by periodontal
56 diseases that we did not control for in the present study. Our study had limitations and future

1   studies are needed to continue this work. While our samples were collected longitudinally, this
2   analysis was performed for the first time point to depict the acute phase. Microbiome and
3   inflammatory responses were not adjusted by the different baselines of each individual
4   intervisibility. While study subjects were able to report how many days, they manifested acute
5   COVID-19 illness (asymptomatic, mild, or moderate/severe), clinical recovery was complicated by
6   the use of national protocols and medications such as antibiotics and immunosuppressives.
7
8   It should be noted that the microbiome results may reflect the population's ethnicity and country
9   of residence(39), thereby are not generalizable to all populations. The feature selection was
10   performed on y1) Uninfected vs. Infected (80 samples); and y2) Mild/moderate vs. Severe (50
11   samples). Our feature selection ended up with 42 features in y1 and 39 features in y2 which are
12   fewer than the number of samples used for modeling. To further explore this, a multi-modal
13   analysis was conducted to identify biomarkers that could distinguish between health/disease and
14   disease severity, with the expectation that the host response to an external infection, as in
15   COVID-19, should be mostly conserved across different populations, with differences related to
16   quantity, rather than activation of completely different pathways. While this study offers
17   preliminary biomarker results, it is important to consider that the salivary microbiome is only one
18   of many factors that could influence the course of COVID-19 and further research, such as
19   longitudinal studies, is needed to fully understand the impact of other factors such as age,
20   underlying health conditions, and genetic susceptibility(40).
21
22   **Materials and Methods**
23
24   This study was a collaborative joint study between the Harvard School of Dental Medicine
25   (HSDM), J Craig Venter Institute (JCVI), the Ministry of Health in Kuwait, and the University of
26   Alberta. The study was approved by JCVI, Harvard, the Kuwait Ministry of Health, and the
27   University of Alberta. Informed consent was obtained from all enrolled participants (Kuwait
28   Ministry of health ethics approval: #2020/1462 Harvard: IRB21-1492, University of Alberta:
29   Pro00125245, JVCI: exempt due to secondary analysis of de-identified samples).
30
31   **Study Design**
32   A convenient sample strategy was used to recruit patients admitted at multiple Covid-19 centers
33   in Kuwait between July 24th and September 4th, 2020. The data collection took place in multiple
34   hospital sites at three hospitals in Kuwait: AlFarwaniyah Hospital, Jaber Al Ahmed Hospital, and
35   Kuwait Field Hospital. Data were collected from those who provided positive consent and who
36   tested positive for SARS-CoV-2 by RT-PCR from nasopharyngeal swabs (n=50). Non-infected
37   individuals were employees at these hospitals who were not in contact with any COVID-19 cases
38   (n=30). The hospital subjected them to a daily visual triage (temperature and symptoms check).
39   However, they did not have a negative PCR test. Saliva and serum samples were collected at the
40   same time point from individuals and within 48 hours of PCR-confirmed COVID-19 diagnosis. our
41   data collection team conducted the study during working hours (8 am – 2 pm), minimizing
42   variability, and excluding nighttime sample collection.
43
44   The basic demographic and clinical information (including medical history, medication,
45   periodontal health data, sleep data, weight, height, waist circumference, neck circumference,
46   blood group, respiratory rate, and oxygen supplementation in liters for those on oxygen) of the
47   study participants was obtained.
48
49   The severity of the disease was stratified into mild: hospitalized, no oxygen therapy (n=11);
50   moderate: hospitalized, low-flow oxygen (<10 L/min) (n=28); and severe: hospitalized, high-flow
51   oxygen (>10 L/min) (n=11). Mild and moderate groups were combined under "non-severe" in the
52   present analysis.
53
54   **Saliva collection**
55   15 mL plastic centrifuge tubes were prelabeled with the date and subject number. We then
56   marked the 4 mL line of the tube. A parafilm was used to stimulate saliva.

9

1 Prior to sample collection, the saliva collection tube was placed in a cup with ice. A nurse,
2 supervised by the research coordinator, would demonstrate how to provide saliva to the patient.
3 Each subject was instructed to take a sip of water and rinse their mouth, swallow the water, chew
4 the piece of parafilm, and then use their tongue to push saliva as it formed into the tube. They
5 were then instructed to place the tube back in the cup with the ice cube while they waited for
6 more saliva to form. The saliva was collected until it reached the line (4 mL) on the tube,
7 considering the liquid region of the saliva sample (not the foamy regions). Once the patient
8 finished providing the saliva sample, they notified the nurse. The nurse tightened the cap on the
9 tube, wiped it with alcohol, placed the tube in the collection rack in the cooler with ice and
10 discarded the other materials.
11
12 **Blood collection**
13 Serum and plasma samples were collected using standard venipuncture techniques in 7.5 mL BD
14 Vacutainer Serum marbled topped tubes with clot activator for serum, and plasma was collected
15 in a 5 mL lavender top tube. Samples were collected at the hospital for all samples.
16
17 **Sample processing**
18 The samples were transferred to the Jaber Alahmad Hospital laboratory in containers with dry ice.
19 The laboratory technician received the samples and processed them on the same day of sample
20 collection within no more than 3 hours. Saliva samples were centrifuged at 2000 x g for 5 min,
21 and the supernatants were separated from the pellets and transferred into different tubes. Plasma
22 and serum samples were allowed to sit upright in racks at room temperature for 30 min prior to
23 centrifuging at 2000xg for 10 min at room temperature. All the samples were stored at –80 °C and
24 were transferred from Kuwait to JCVI. The samples were placed on dry ice during shipment with
25 a monitoring device to ensure that the samples were frozen during the transfer.
26
27 **DNA extraction, library preparation, and sequencing**
28 DNA was extracted from samples using Qiagen's AllPrep Bacterial DNA/RNA/Protein Kit (Cat#
29 47054; QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Step 3 was
30 modified to use MP Biomedicals™ FastPrep-24™ Classic Bead Beating Grinder and Lysis
31 System (Cat# MP116004500; FisherScientific) for 1 minute instead of vortexing for 10 minutes.
32 All recommended products were used, and, at step 2, ß-ME was used instead of DTT.
33 Sequencing was done on the Illumina MiSeq platform. V4 hypervariable region was sequenced
34 using the following forward (MSV4Read1) and reverse primers (MSV34Index1), 5'-
35 TATGGTAATTGTGTGCCAGCMGCCGCGGTAA-3' and 5'-
36 ATTAGAWACCCBDGTAGTCCGGCTGACTGACT-'3.
37 Samples were deposited in NIH SRA under the accession number PRJNA948421.
38
39
40 **Cytokine abundance measurements**
41 Frozen (-80°C) serum and saliva samples were brought to room temperature over ice and
42 assayed using the Immune Monitoring 65-Plex Human ProcartaPlex Panel (Cat# EPX650-16500-
43 901; ThermoFisher Scientific, Vienna, Austria), according to manufacturer's instructions, and run
44 on the Luminex 200 system (Luminex Corporation, Austin, Texas, USA). This kit measured
45 immune function by analyzing 65 protein targets in a single well, including cytokine, chemokine,
46 growth factors/regulators, and soluble receptors. Assay targets included the following: APRIL,
47 BAFF, BLC, CD30, CD40L, ENA-78, Eotaxin, Eotaxin-2, Eotaxin-3, FGF-2, Fractalkine, G-CSF,
48 GM-CSFm Gro alpha, HGF, IFN-alpha, IFN-gamma, IL-1 alpha, IL-1 beta, IL-2, IL-2R, IL-3, IL-4,
49 IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12p70, IL-13, IL-15, IL-16, IL-17A, IL-18, IL-20, IL-21, IL-22,
50 IL-23, IL-27, IL-31, IP-10, I-TAC, LIF, MCP-1, MCP-2, MCP-3, M-CSF, MDC, MIF, MIG, MIP-1
51 alpha, MIP-1 beta, MIP-3, alpha, MMP-1, NGF beta, SCF, SDF-1 alpha, TNF alpha, TNF beta,
52 TNF-R2, TRAIL, TSLP, TWEAK, VEGF-A. The provided standard was diluted 4-fold to generate a
53 standard curve, and high and low controls were also included. Both saliva and serum samples
54 were diluted 1:1 with the included assay buffer. Results used in analysis from the Luminex assay
55 were displayed as Net MFI (Median Fluorescence Intensity).
56

1 Amplicon sequence variants (ASVs)

2 Analysis was performed at the amplicon sequence variant level: DNA sequences containing no
3 sequencing errors after algorithmic correction. The paired-end DADA2(41) workflow of QIIME2
4 v2022.2.0(42) was used for detecting ASVs and quantifying abundance implemented using the
5 amplicon.py module of VEBA(43). More specifically, this workflow uses the following protocol: 1)
6 qiime tools import of paired-end reads; 2) DADA2 denoising of paired reads and ASV detection
7 by qiime dada2 denoise-paired (forward_trim=251, reverse_trim=231, min-overlap=12); 3)
8 taxonomic classification of ASVs using the precompiled silva-138-99-nb-classifier.qza model
9 (Silva_v1383__SSURef_NR9); and 4) conversion into generalized machine-readable formats
10 (e.g., QIIME2 Artifact and BIOM formatted files → tab-separated value and Fasta-formatted files).
11

12 Cytokine and microbiome statistical analysis

13 Alpha diversity (i.e., richness) was calculated by summing each sample's detected ASVs. Beta
14 diversity (i.e., hierarchical clustering and PCoA) was calculated by computing the pairwise
15 Aitchison distance, center log-ratio followed by Euclidean distance, of samples and using this
16 distance matrix as input in the Agglomerative and PrincipleCoordinateAnalysis objects of
17 Soothsayer(15) which are wrappers around SciPy(44) and Scikit-Bio, respectively.
18

19 PCoA with Jaccard and PhILR(45) alpha diversity, probability density function for Jaccard
20 breakdown (betapart package in R) were conducted via the automated pipeline FALAPhyl:
21 https://github.com/khalidtab/falaphyl/. Cytokine Net MFI values were used for cytokine
22 measurements. Values were transformed by log10 transformation, then converted into z-scores,
23 and finally plotted with the R package ComplexHeatmap(46). Cytokines with negative MFI values
24 (i.e. lower signal than standard) had the lowest value plus one added to all the cytokine
25 measurements so that the lowest value equals one so that the values can undergo the same
26 log10 transformation and z-score standardization above. Correlations between blood and saliva
27 cytokines were done with Spearman correlation in R, and those found to be statistically significant
28 were then plotted with the R package corrplot (47).
29

30 Alpha diversity (i.e., richness) was calculated by summing each sample's detected ASVs. Beta
31 diversity (i.e., hierarchical clustering and PCoA) was calculated by computing the pairwise
32 Aitchison distance, center log-ratio followed by Euclidean distance, of samples and using this
33 distance matrix as input in the Agglomerative and PrincipleCoordinateAnalysis objects of
34 Soothsayer(15) which are wrappers around SciPy(44) and Scikit-Bio, respectively.
35

36 PCoA with Jaccard and PhILR(45) alpha diversity, probability density function for Jaccard
37 breakdown (betapart package in R) were conducted via the automated pipeline FALAPhyl:
38 https://github.com/khalidtab/falaphyl/. Cytokine Net MFI values were used for cytokine
39 measurements. Values were transformed by log10 transformation, then converted into z-scores,
40 and finally plotted with the R package ComplexHeatmap(46). Cytokines with negative MFI values
41 (i.e. lower signal than standard) had the lowest value plus one added to all the cytokine
42 measurements so that the lowest value equals one so that the values can undergo the same
43 log10 transformation and z-score standardization above. Correlations between blood and saliva
44 cytokines were done with Spearman correlation in R, and those found to be statistically significant
45 were then plotted with the R package corrplot (47).
46

47 For the abundance paradigm, we ran feature selection models using the following
48 transformations: 1) the quality assessed center log-ratio transformed ASV abundances
49 (multiplicative replacement of 1); and 2) the Z-score normalized cytokine levels. The predictive
50 features that were identified using Clairvoyance on the ASV and cytokine abundances were used
51 to build SSPNs (described below). For the inferred interaction paradigm, the perturbation matrix
52 that is made of all SSPNs was used as input into the final rounds of feature selection.
53

54 To avoid performing exhaustive feature selection using every combination of pairwise multi-modal
55 associations (~176k associations where the majority of associations would contain non-
56 informative features), we identified biomarkers for each modality individually and used the

11

1 resulting informative features (y1 + y2 feature sets for each modality) as input for the SSPN
2 analysis to produce a perturbation matrix as implemented in Nabwera & Espinoza et al. 2021 (15).
3
4
5 Briefly, a perturbation matrix with edge weights of the fully connected SSPNs was fed into a
6 feature selection algorithm to determine the minimal set of edges with the highest predictive
7 capacity. That is biomarkers for COVID-19 as multi-modal associations rather than the
8 abundances of specific features. Once the minimal edges predictive of COVID-19 are identified,
9 the Logistic Regression classification model is fit, and the coefficients representing predictive
10 capacity are used as edge weights to construct an aggregate network (Fig. 3). Positive
11 coefficients indicate that an increase in perturbation corresponds with an increased likelihood of a
12 COVID-19 diagnosis, while negative coefficients indicate that an increase in perturbation
13 corresponds with a decreased likelihood of a COVID-19 diagnosis. Ultimately, this procedure
14 resulted in 8 feature selection runs (y1 and y2 classifications for each scenario) that could be run
15 on a local machine: 1-2) ASV abundances; 3-4) Salivary cytokines abundances; 5-6) Systemic
16 cytokines abundances; and 7-8) multi-modal perturbations (i.e., SSPN).
17
18 For Clairvoyance, we used a logistic regression classifier with the following parameter space:
19 penalty (31) and C {1e-10,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0} with the liblinear solver and 1000
20 maximum iterations for convergence. The weights from the fitted classifiers are representative of
21 a feature's predictive capacity, that is, the coefficient of the fitted linear model. Linear models
22 have both positive and negative coefficients. In the case of our models, a positive coefficient for
23 a feature indicates that an increased value of the feature corresponds with an increased
24 likelihood of the query sample to be predicted as the following: y1) infected relative to uninfected;
25 and y2) severe relative to mild/moderate severity. Only feature sets with less than the number of
26 samples ($N_{Samples}$ = 80) were considered for downstream analysis to prevent overfitting.
27
28 Multi-omic sample-specific perturbation networks (SSPNs)
29 SSPNs were created to identify which inferred interactions were perturbed by adding a query
30 sample to a reference cohort. Cytokine Z-score normalized data was propagated from previous
31 analysis as they are not influenced by intra-sample features, while ASV abundance was
32 retransformed using the feature subset identified by Clairvoyance which is influenced by intra-
33 sample features. Pearson correlation was used for both continuous and multi-modal interactions
34 (i.e., cytokine-cytokine and cytokine-ASV), while rho proportionality was used for compositional
35 interactions (i.e., ASV – ASV). The reference category was the uninfected class, and in the y1
36 scenario, we used a dummy class to obtain perturbation measurements that reflect how adding
37 individual reference samples to the larger reference cohort perturbs the associations. SSPNs
38 were implemented using the Sample Specific Perturbation Network object in the Ensemble
39 NetworkX Python package (v2023.2.14) (15, 48) using 1000 iterations and a sample size of
40 0.618 with median and median-absolute deviation as the edge reduction statistic and variability
41 measurements, respectively.
42
43 Multi-omic aggregate networks
44 The logistic regression classifier used for constructing the aggregate networks was fit using the
45 perturbation of inferred interactions (i.e., the perturbation matrix from the SSPN analysis) with the
46 predictive feature subset identified by Clairvoyance (i.e., multi-modal edges in the perturbation
47 matrix) and the COVID-19 severity as classes. As each feature in our classification models
48 represented an edge, we reformatted our fitted models into undirected aggregate networks where
49 each node represents either an ASV or cytokine, each edge represents an inferred interaction,
50 and edge weight represents the predictive capacity of each edge when its associations are
51 perturbed. Salivary and systemic cytokines were kept separate during all analyses.
52
53 Node degree is calculated as the sum of absolute values of edge weights connected to a node.
54 Node degree homogeneity is a variant of Pielou's ecological index for evenness calculated via
55 H/llog2(Number of nodes) where H indicates Shannon entropy (49).

12

Supplementary Material

Supplementary materials available at PNAS Nexus online.

Data Availability

All analyzed data are included in the manuscript and Supplementary materials available at PNAS Nexus online. Raw microbial sequence reads are available at NCBI, PRJNA948421, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA948421/ Codes to compute the data, analyze the files, develop graphs are available at https://github.com/JCVenterInstitute and https://github.com/khalidtab/PNAS_nexus_paper/blob/main/pnas_nexus.Rmd

References

1.   Moutsopoulos NM, Konkel JE. Tissue-Specific Immunity at the Oral Mucosal Barrier. Trends Immunol. 2018;39(4):276-87.

2.   Şenel S. An Overview of Physical, Microbiological and Immune Barriers of Oral Mucosa. Int J Mol Sci. 2021;22(15).

3.   Salzano FA, Marino L, Salzano G, Botta RM, Cascone G, D'Agostino Fiorenza U, et al. Microbiota Composition and the Integration of Exogenous and Endogenous Signals in Reactive Nasal Inflammation. J Immunol Res. 2018;2018:2724951.

4.   Zhu F, Zhong Y, Ji H, Ge R, Guo L, Song H, et al. ACE2 and TMPRSS2 in human saliva can adsorb to the oral mucosal epithelium. J Anat. 2022;240(2):398-409.

5.   Haran JP, Bradley E, Zeamer AL, Cincotta L, Salive MC, Dutta P, et al. Inflammation-type dysbiosis of the oral microbiome associates with the duration of COVID-19 symptoms and long COVID. JCI Insight. 2021;6(20).

6.   Gupta A, Karyakarte R, Joshi S, Das R, Jani K, Shouche Y, et al. Nasopharyngeal microbiome reveals the prevalence of opportunistic pathogens in SARS-CoV-2 infected individuals and their association with host types. Microbes Infect. 2022;24(1):104880.

7.   Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet. 2020;395(10229):1033-4.

8.   Abdul-Kareem HH, Al-Maqtoofi MY, Burghal AA. Impact of COVID-19 vaccination on saliva immune barriers: IgA, lysozyme, and lactoferrin. Arch Virol. 2023;168(12):293.

9.   Bondareva M, Budzinski L, Durek P, Witkowski M, Angermair S, Ninnemann J, et al. Cross-regulation of antibody responses against the SARS-CoV-2 Spike protein and commensal microbiota via molecular mimicry. Cell Host Microbe. 2023;31(11):1866-81.e10.

13

10. Gottfried-Blackmore A, Rubin SJS, Bai L, Aluko S, Yang Y, Park W, et al. Effects of processing conditions on stability of immune analytes in human blood. Sci Rep. 2020;10(1):17328.

11. Jang H, Choudhury S, Yu Y, Sievers BL, Gelbart T, Singh H, et al. Persistent immune and clotting dysfunction detected in saliva and blood plasma after COVID-19. Heliyon. 2023;9(7).

12. Verdiguel-Fernández L, Arredondo-Hernández R, Mejía-Estrada JA, Ortiz A, Verdugo-Rodríguez A, Orduña P, et al. Differential expression of biomarkers in saliva related to SARS-CoV-2 infection in patients with mild, moderate and severe COVID-19. BMC Infect Dis. 2023;23(1):602.

13. Williams DW, Greenwell-Wild T, Brenchley L, Dutzan N, Overmiller A, Sawaya AP, et al. Human oral mucosa cell atlas reveals a stromal-neutrophil axis regulating tissue immunity. Cell. 2021;184(15):4090-104.e15.

14. Baselga A, Orme CDL. betapart : an R package for the study of beta diversity: <i>Betapart package</i>. Methods in Ecology and Evolution. 2012;3(5):808-12.

15. Espinoza JL, Dupont CL, O'Rourke A, Beyhan S, Morales P, Spoering A, et al. Predicting antimicrobial mechanism-of-action from transcriptomes: A generalizable explainable artificial intelligence approach. PLoS Comput Biol. 2021;17(3):e1008857.

16. Altman N, Krzywinski M. The curse(s) of dimensionality. Nat Methods. 2018;15(6):399-400.

17. Mathew D, Giles JR, Baxter AE, Oldridge DA, Greenplate AR, Wu JE, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science. 2020;369(6508).

18. Russell MW, Moldoveanu Z, Ogra PL, Mestecky J. Mucosal Immunity in COVID-19: A Neglected but Critical Aspect of SARS-CoV-2 Infection. Front Immunol. 2020;11:611337.

19. Reinhardt EL, Fernandes P, Markus RP, Fischer FM. Night work effects on salivary cytokines TNF, IL-1beta and IL-6. Chronobiol Int. 2019;36(1):11-26.

20. Willemse SB, Jansen L, de Niet A, Sinnige MJ, Takkenberg RB, Verheij J, et al. Intrahepatic IP-10 mRNA and plasma IP-10 levels as response marker for HBeAg-positive chronic hepatitis B patients treated with peginterferon and adefovir. Antiviral Res. 2016;131:148-55.

21. Lei J, Yin X, Shang H, Jiang Y. IP-10 is highly involved in HIV infection. Cytokine. 2019;115:97-103.

22. Cerutti A, Zan H, Schaffer A, Bergsagel L, Harindranath N, Max EE, et al. CD40 ligand and appropriate cytokines induce switching to IgG, IgA, and IgE and coordinated germinal center and plasmacytoid phenotypic differentiation in a human monoclonal IgM+IgD+ B cell line. J Immunol. 1998;160(5):2145-57.

23. Malisan F, Brière F, Bridon JM, Harindranath N, Mills FC, Max EE, et al. Interleukin-10 induces immunoglobulin G isotype switch recombination in human CD40-activated naive B lymphocytes. J Exp Med. 1996;183(3):937-47.

24. Kaczor-Urbanowicz KE, Martin Carreras-Presas C, Aro K, Tu M, Garcia-Godoy F, Wong DT. Saliva diagnostics - Current views and directions. Exp Biol Med (Maywood). 2017;242(5):459-72.

25. De Maio F, Posteraro B, Ponziani FR, Cattani P, Gasbarrini A, Sanguinetti M. Nasopharyngeal Microbiota Profiling of SARS-CoV-2 Infected Patients. Biol Proced Online. 2020;22:18.

26. Azarpazhooh A, Leake JL. Systematic review of the association between respiratory diseases and oral health. J Periodontol. 2006;77(9):1465-82.

27. Teshome A, Yitayeh A. Relationship between periodontal disease and preterm low birth weight: systematic review. Pan Afr Med J. 2016;24:215.

28. Godlewski AE, Veyrune JL, Nicolas E. [Obesity and oral health: risk factors of obese patients in dental practice]. Odontostomatol Trop. 2008;31(123):25-32.

29. Kudiyirickal MG, Pappachan JM. Diabetes mellitus and oral health. Endocrine. 2015;49(1):27-34.

30. Janket SJ, Baird AE, Chuang SK, Jones JA. Meta-analysis of periodontal disease and risk of coronary heart disease and stroke. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2003;95(5):559-69.

14

31. Zang R, Gomez Castro MF, McCune BT, Zeng Q, Rothlauf PW, Sonnek NM, et al. TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. Sci Immunol. 2020;5(47).

32. Stocker N, Radzikowska U, Wawrzyniak P, Tan G, Huang M, Ding M, et al. Regulation of angiotensin-converting enzyme 2 isoforms by type 2 inflammation and viral infection in human airway epithelium. Mucosal Immunol. 2023;16(1):5-16.

33. Dani N, Herbst RH, McCabe C, Green GS, Kaiser K, Head JP, et al. A cellular and spatial map of the choroid plexus across brain ventricles and ages. Cell. 2021;184(11):3056-74.e21.

34. Ngo VL, Gewirtz AT. Microbiota as a potentially-modifiable factor influencing COVID-19. Curr Opin Virol. 2021;49:21-6.

35. de Lima CL, Acevedo AC, Grisi DC, Taba M, Jr., Guerra E, De Luca Canto G. Host-derived salivary biomarkers in diagnosing periodontal disease: systematic review and meta-analysis. J Clin Periodontol. 2016;43(6):492-502.

36. Han MS, Byun JH, Cho Y, Rim JH. RT-PCR for SARS-CoV-2: quantitative versus qualitative. Lancet Infect Dis. 2021;21(2):165.

37. Shah S, Singhal T, Davar N, Thakkar P. No correlation between Ct values and severity of disease or mortality in patients with COVID 19 disease. Indian J Med Microbiol. 2021;39(1):116-7.

38. Sri Santosh T, Parmar R, Anand H, Srikanth K, Saritha M. A Review of Salivary Diagnostics and Its Potential Implication in Detection of Covid-19. Cureus. 2020;12(4):e7708.

39. Mason MR, Nagaraja HN, Camerlengo T, Joshi V, Kumar PS. Deep Sequencing Identifies Ethnicity-Specific Bacterial Signatures in the Oral Microbiome. PLOS ONE. 2013;8(10):e77287.

40. Hu J, Wang Y. The Clinical Characteristics and Risk Factors of Severe COVID-19. Gerontology. 2021;67(3):255-66.

41. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 2016;13(7):581-3.

42. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology. 2019;37(8):852-7.

43. Espinoza JL, Dupont CL. VEBA: a modular end-to-end suite for in silico recovery, clustering, and analysis of prokaryotic, microeukaryotic, and viral genomes from metagenomes. BMC Bioinformatics. 2022;23(1):419.

44. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-72.

45. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. Elife. 2017;6.

46. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.

47. SV. WT. R package 'corrplot': Visualization of a Correlation Matrix. 2021.

48. Hagberg A, Swart P, Chult D. Exploring Network Structure, Dynamics, and Function Using NetworkX2008.

49. Pielou EC. The measurement of diversity in different types of biological collections. Journal of Theoretical Biology. 1966;13:131-44.

50. KD. PS. Tachypnea StatPearls. StatPearls Publishing. 2023.

15

1 **Figure Legends**

2 **Figure 1. Analysis of salivary microbiome between the 3 groups, control (red),**
3 **mild/moderate (blue) and severe (yellow).** A: Principal coordinates analysis (PCoA) of Jaccard
4 dissimilarity demonstrating the three clusters formed by the microbial constituents. Ellipsoids
5 show the 1st standard deviation of spread of each group ($p<0.05$, ADONIS). B. Violin plots of
6 observed features diversity between the three groups ($p<0.05$, Wilcoxon rank-sum test). C.Violin
7 plots of Shannon diversity between the three groups ($p<0.05$, Wilcoxon rank-sum test). D.
8 Principal coordinates analysis (PCoA) of Euclidean distances of PhILR distances, demonstrating
9 the three clusters between the conditions.

10

11 **Figure 2. Heatmaps of salivary and blood cytokines showing statistical significance and**
12 **correlation.** A and B: Groups for control (black group), mild/moderate (pink group), and severe
13 (red group) show statistical significance by Kruskal-Wallis for blood (A) and saliva (B) cytokines.
14 C-F: Correlogram between blood (rows) and salivary (columns) cytokines for controls (C), all
15 COVID patients (D), mild/moderate category (E), and severe category (F). Areas were colored
16 based on the Spearman correlation Rho and any correlations not statistically significant ($P>0.05$)
17 were removed and replaced with a white square. See Supplementary Table 4 for the full list of
18 cytokines.

19

20 **Figure 3. Hierarchical feature selection for identifying biomarkers predictive of COVID-19**
21 **severity.** A) Hierarchical classification of COVID severity and the sub-models with relative
22 explained variance of each group derived from 1st principal component of PCA. Modality-specific
23 biomarker overlap between sub-models of B) abundance-based paradigms and C) inferred-
24 interaction paradigms. In the inferred interaction paradigm, features are edges in the SSPN
25 which can be decomposed into nodes which is not the case for the abundance-based paradigms
26 in B. *Clairvoyance* feature selection results for D) sub-model y1 and E) sub-model y2 color
27 coordinated by modality. Each marker on the scatter plot represents a unique
28 hyperparameter/feature set combination that yields a specific accuracy using 10-fold cross-
29 validation repeated with 10 different random states.

1   **Figure 4. Aggregate network representations of fitted sub-models showcasing predictive**

2   **capacity** – Aggregate network for sub-model A) y1 and B) y2. The edge weights can be

3   interpreted as predictive capacity for COVID severity. For y1, positive values indicate that an

4   increase in perturbation results in an increased likelihood that a sample is classified as y2 (i.e.,

5   infected) relative to uninfected. For y2, positive values indicate that an increase in perturbation

6   results in an increased likelihood that a sample is classified as severe relative to mild/moderate.

7   Node size is proportional to weighted degree as a measurement of network connectivity and, by

8   extension, biomarker importance. C) ASVs in the SSPN as proportions of presence in the various

9   conditions, demonstrating that the majority of the ASVs are due to differences in abundances of

10   common taxa across all conditions, and not rare species.

11

12   **Table 1. Clinical characteristics and demographics of the study patients according to**

13   **disease status and severity.** The chi-square test was conducted for categorical variables, and

14   the Mann-Whitney test was conducted for continuous variables. IQR: Interquartile Range. The

15   severity of the disease was stratified into mild: hospitalized, no oxygen therapy (n=11); moderate:

16   hospitalized, low-flow oxygen (<10 L/min); and severe: hospitalized, high-flow oxygen (>10 L/min)

17   (n=39). Mild and moderate groups were combined under "non-severe". Blood work**: ALP:

18   Alkaline phosphatase (liver function); ALT: alanine transaminase (liver function); AST: aspartate

19   transferees (liver function); GGT: gamma-glutamyl transferase (liver function); HB: hemoglobin;

20   PCT: procalcitonin; PLT: platelet count; eGFR: the measure of kidney function (the lower, the

21   worse). Reference for respiratory distress cut off is 20/minute(50).

22
23
24
25
26

1

Characteristics		SARS-CoV-2 Status					COVID-19 Severity				
		SARS-CoV-2(+)		SARS-COV-2 (-)			Non-severe		Severe		
		N=50 (%)		N=30 (%)		P-value	N=39 (%)		N=11 (%)		P-value
Variable 1	Variable 2	Variable 1	Variable 2	Variable 1	Variable 2		Variable 1	Variable 2	Variable 1	Variable 2	
Age <=50 yr	Age >50 yr	21/49 (43)	28/49 (57)	30/30 (100)	0/30 (0)	**P<0.001**	17/38 (45)	21/38 (55)	4/11 (36)	7/11 (64)	P=0.62
Female	Male	22/50 (44)	28/50 (56)	14/30 (47)	16/30 (53)	P=0.81	19/39 (49)	2039 (51)	3/11 (27)	8/11 (73)	P=0.20
Diabetic	Nondiabetic	22/50 (44)	28/50 (56)	1/30 (3)	29/30 (97)	**P<0.001**	21/39 (54)	18/39 (46)	1/11 (9)	10/11 (91)	**P=0.008**
Heart disease	No heart disease	8/50 (16)	42/50 (84)	0/30 (0)	30/30 (100)	**P=0.02**	6/39 (15)	33/39 (85)	2/11 (18)	9/11 (82)	P=0.82
Hypertension	No hypertension	22/50 (44)	28/50 (56)	0/30 (0)	30/30 (100)	**P<0.001**	18/39 (46)	21/39 (54)	4/11 (36)	7/11 (67)	P=0.56
BMI<=35	BMI>35	30/48 (62.5)	18/48 (37.5)	27/30 (90)	3/30 (10)	**P=0.008**	24/38 (63)	14/38 (37)	6/10 (60)	4/10 (40)	P=0.85
Smokers	Non-smokers	1/50 (2)	49/50 (98)	17/27 (63)	10/27 (37)	**P<0.001**	0/39 (0)	39/39 (100)	1/11 (9)	10/11 (91)	P=0.05
Chronic cough	No chronic cough	49/50 (98)	1/50 (2)	29/29 (100)	0/29 (0)	P=0.44	38/39 (97)	1/39 (3)	11/11 (100)	0/11 (0)	P=0.59
Chest pain	No chest pain	12/50 (24)	38/50 (76)	3/29 (10)	26/29 (90)	P=0.13	8/39 (21)	31/39 (79)	4/11 (36)	7/11 (64)	P=0.27
Arthritis	No arthritis	10/50 (20)	40/50 (80)	1/30 (3)	29/30 (97)	**P=0.03**	9/39 (23)	30/39 (77)	1/11 (9)	10/11 (91)	P=0.30
Cancer	No cancer	3/50 (6)	47/50 (94)	0/30 (0)	30/30 (100)	P=0.17	3/39 (8)	36/39 (92)	0/11 (0)	11/11 (100)	P=0.34
Asthma	No asthma	5/50 (10)	45/50 (90)	0/29 (0)	29/29 (100)	**P=0.07**	5/39 (13)	34/39 (87)	0/11 (0)	11/11 (100)	P=0.21
Medication in the past 2 years	No Medication in the past 2 years	20/50 (40)	6/50 (12)	3/29 (10)	26/29 (90)	**P<0.001**	18/39 (46)	4/39 (10)	2/11 (18)	2/11 (18)	P=0.24
Characteristics		**COVID-19 Symptoms**									
		SARS-CoV-2 Severity									

		Non-severe (mild/moderate)		Severe		P-value
		N=39 (%)		N=11 (%)		
Variable 1	Variable 2	Variable 1	Variable 2	Variable 1	Variable 2	
Fever	No fever	22/39 (56)	17/39 (44)	2/11 (18)	9/11 (82)	**P=0.02**
Shortness of breath	No Shortness of breath	23/39 (59)	16/39 (41)	9/11 (82)	2/11 (18)	P=0.16
Persistent pain	No Persistent pain	18/39 (46)	20/39 (51)	5/11 (45)	6/11 (55)	P=0.86
Respiratory distress <=20/ min	Respiratory distress >20/min	12/39 (31)	27/39 (69)	2/11 (18)	9/11 (82)	P=0.41
Sore throat	No Sore throat	12/39 (31)	27/39 (69)	2/11 (18)	9/11 (82)	P=0.41
Sputum production	No sputum production	19/39 (49)	20/39 (51)	4/11 (36)	7/11 (64)	P=0.46
Congestion	No Congestion	4/39 (10)	35/39 (90)	4/11 (36)	7/11 (64)	**P=0.03**
Loss of taste/smell	No Loss of taste/smell	12/39 (31)	27/39 (69)	6/11 (55)	5/11 (45)	P=0.14
Chills	No Chills	23/39 (59)	16/39 (41)	3/11 (27)	8/11 (73)	P=0.06
Fatigue	No Fatigue	25/39 (64)	14/39 (36)	8/11 (73)	3/11 (27)	P=0.59
Muscle ache	No Muscle ache	23/39 (59)	16/39 (41)	6/11 (55)	5/11 (45)	P=0.79
		COVID-19 Symptoms				
		SARS-CoV-2 Severity				
Characteristics		Non-severe (mild/moderate)		Severe		P-value
		N=39 (%)		N=11 (%)		
Bloodwork**		Median (IQR)		Median (IQR)		
ALP		75 (28)		63 (34)		P=0.29

19

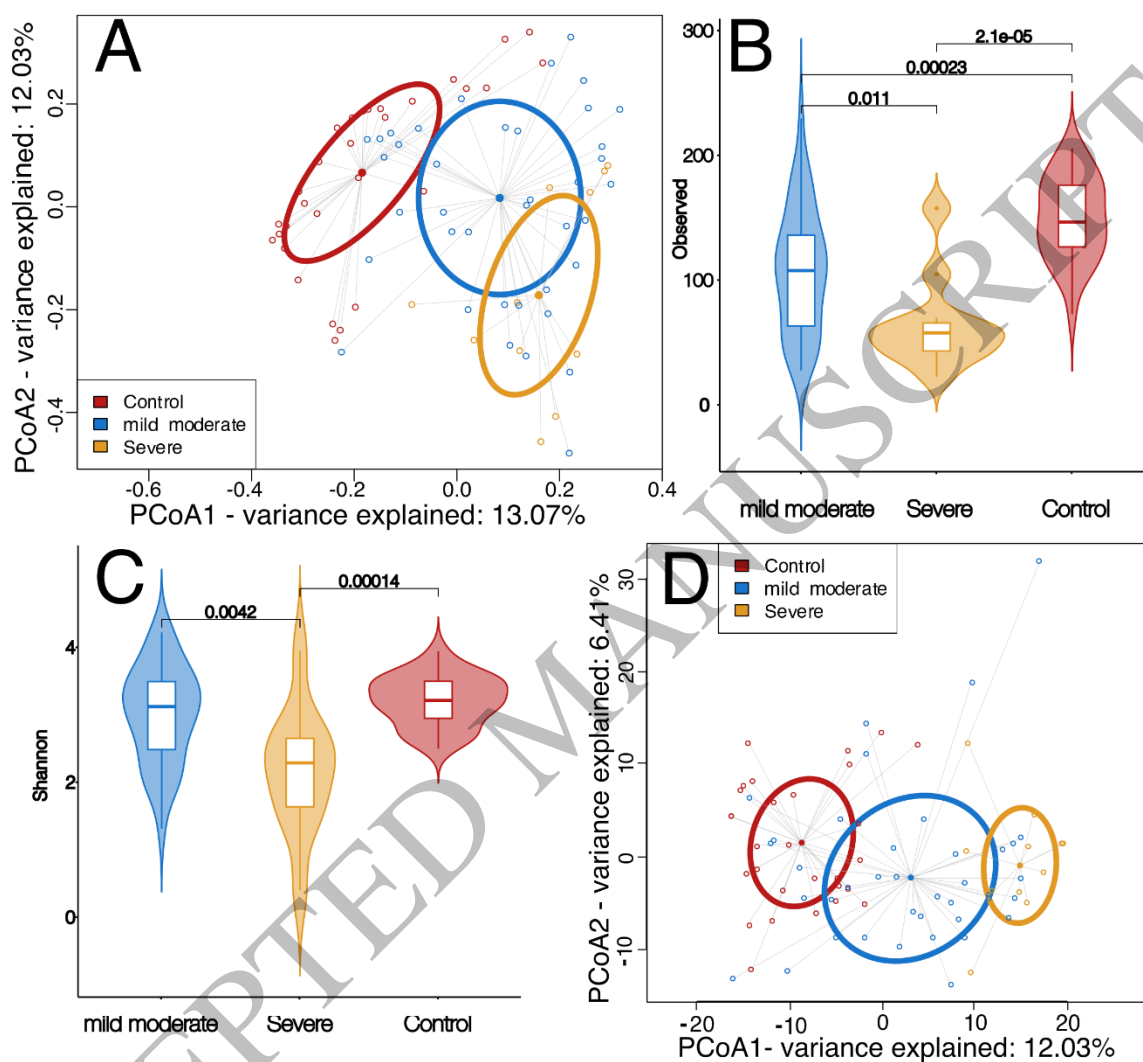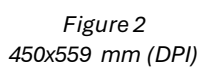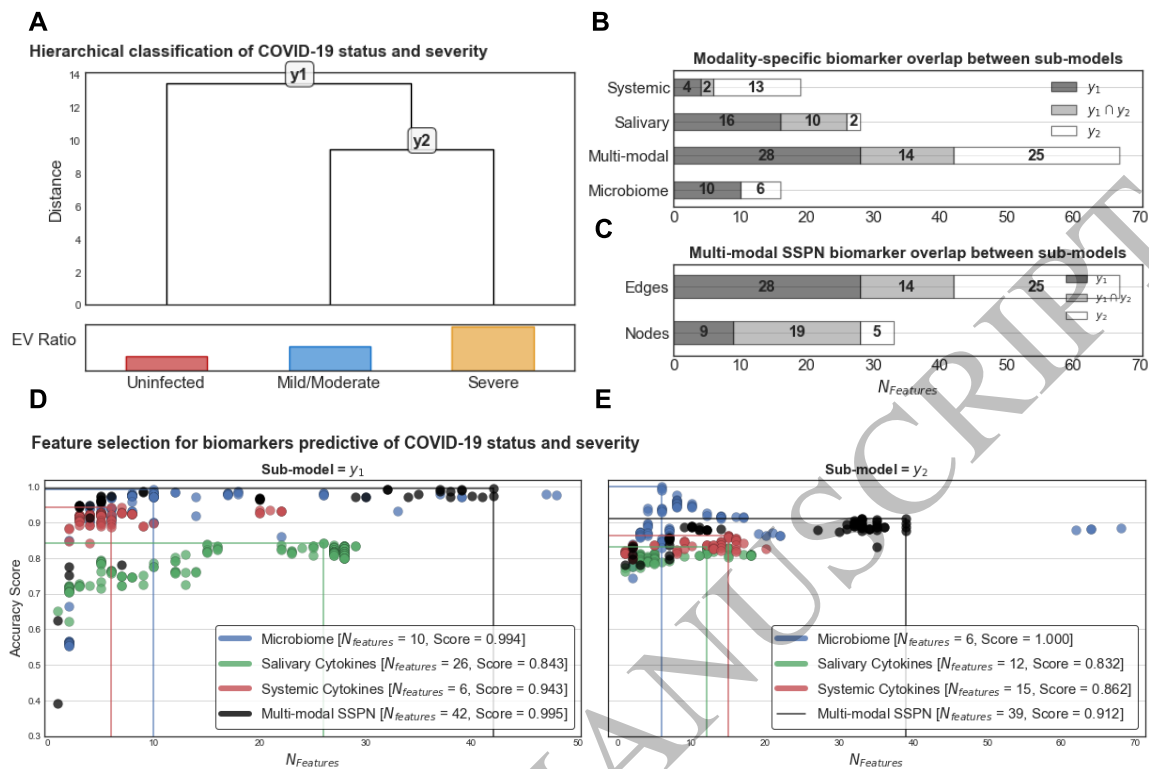|			
ALT	34 (28)	41 (17)	P=0.90
AST	41 (19)	41 (22)	P=0.77
GGT	67 (65)	62 (82)	P=0.80
HB	129 (26)	132 (18)	P=0.52
PCT	0.2 (0.1)	0.2 (0.17)	P=0.43
PLT	215 (90)	220 (131)	P=0.16
eGFR	87 (40)	64 (60)	P=0.17

1

1

2

3

4

5

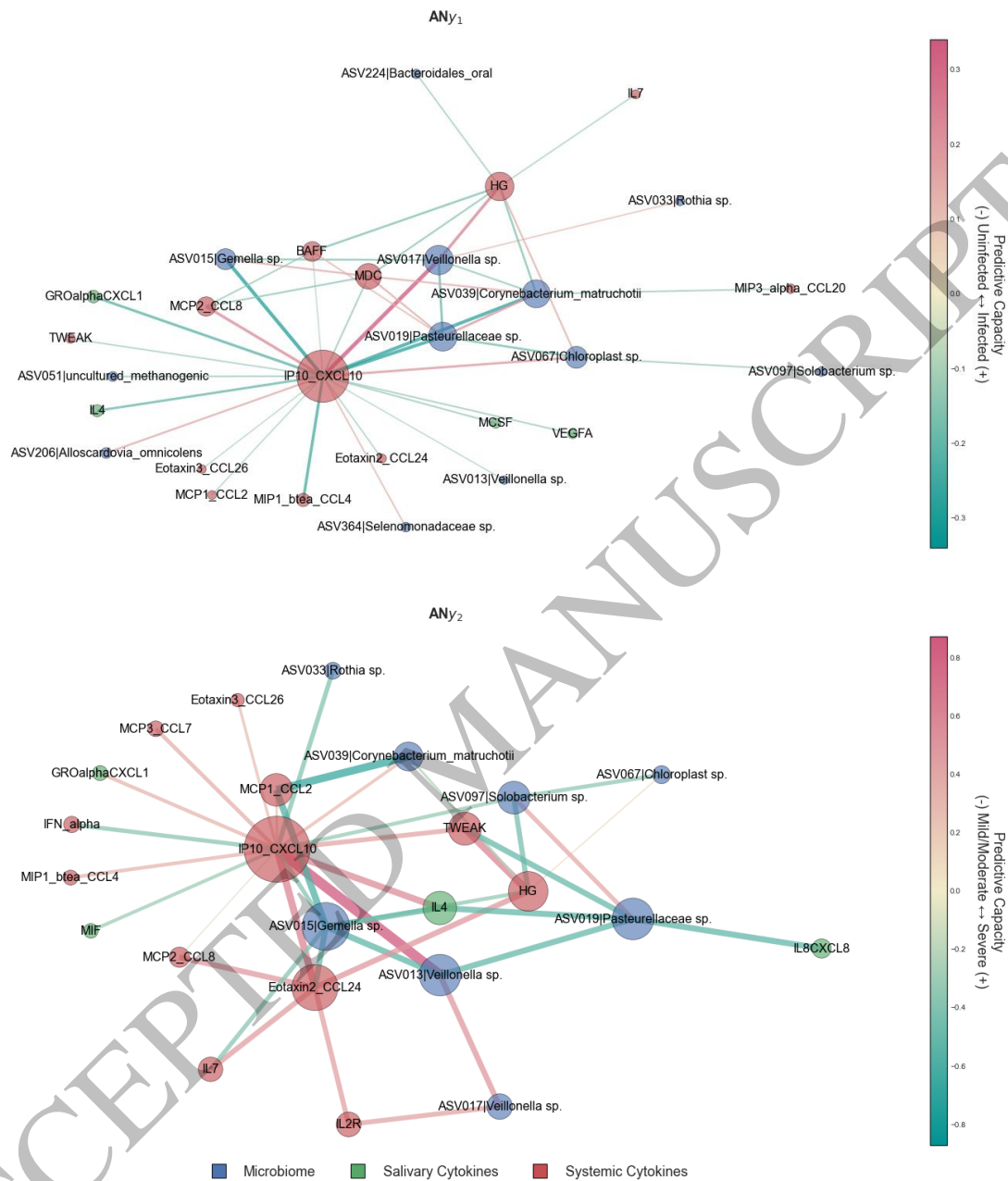*Figure 1*
559x522 mm (DPI)

6

Figure 2
450x559 mm (DPI)

1

2 *Figure 3*
3 *216x279 mm (DPI)*

4

Figure 4
216x279 mm (DPI)