

RESEARCH

Open Access



# Proteomic snapshot of saliva samples predicts new pathways implicated in SARS-CoV-2 pathogenesis

Elena Moreno<sup>1,2\*†</sup>, Sergio Ciordia<sup>3†</sup>, Santos Milhano Fátima<sup>3†</sup>, Daniel Jiménez<sup>1</sup>, Javier Martínez-Sanz<sup>1,2</sup>, Pilar Vizcarra<sup>1,2</sup>, Raquel Ron<sup>1,2</sup>, Matilde Sánchez-Conde<sup>1,2</sup>, Rafael Bargiela<sup>4</sup>, Sergio Sanchez-Carrillo<sup>5,7</sup>, Santiago Moreno<sup>1,2,6</sup>, Fernando Corrales<sup>3†</sup>, Manuel Ferrer<sup>5†</sup> and Sergio Serrano-Villar<sup>1,2†</sup>

## Abstract

**Background** Information on the microbiome's human pathways and active members that can affect SARS-CoV-2 susceptibility and pathogenesis in the salivary proteome is very scarce. Here, we studied a unique collection of samples harvested from April to June 2020 from unvaccinated patients.

**Methods** We compared 10 infected and hospitalized patients with severe ( $n=5$ ) and moderate ( $n=5$ ) coronavirus disease (COVID-19) with 10 uninfected individuals, including non-COVID-19 but susceptible individuals ( $n=5$ ) and non-COVID-19 and nonsusceptible healthcare workers with repeated high-risk exposures ( $n=5$ ).

**Results** By performing high-throughput proteomic profiling in saliva samples, we detected 226 unique differentially expressed (DE) human proteins between groups ( $q\text{-value} \leq 0.05$ ) out of 3376 unambiguously identified proteins (false discovery rate  $\leq 1\%$ ). Major differences were observed between the non-COVID-19 and nonsusceptible groups. Bioinformatics analysis of DE proteins revealed human proteomic signatures related to inflammatory responses, central cellular processes, and antiviral activity associated with the saliva of SARS-CoV-2-infected patients ( $p\text{-value} \leq 0.0004$ ). Discriminatory biomarker signatures from human saliva include cystatins, protective molecules present in the oral cavity, calprotectins, involved in cell cycle progression, and histones, related to nucleosome functions. The expression levels of two human proteins related to protein transport in the cytoplasm, DYNC1 ( $p\text{-value}$ , 0.0021) and MAPRE1 ( $p\text{-value}$ , 0.047), correlated with angiotensin-converting enzyme 2 (ACE2) plasma activity. Finally, the proteomes of microorganisms present in the saliva samples showed 4 main microbial functional features related to ribosome functioning that were overrepresented in the infected group.

**Conclusion** Our study explores potential candidates involved in pathways implicated in SARS-CoV-2 susceptibility, although further studies in larger cohorts will be necessary.

**Keywords** Proteomics, SARS-CoV-2, Saliva, Pathogenesis, Functional analysis

<sup>†</sup>Elena Moreno, Sergio Ciordia and Santos Milhano Fátima have contributed equally to this work and share the first authorship.

<sup>†</sup>Fernando Corrales, Manuel Ferrer and Sergio Serrano-Villar share senior authorship.

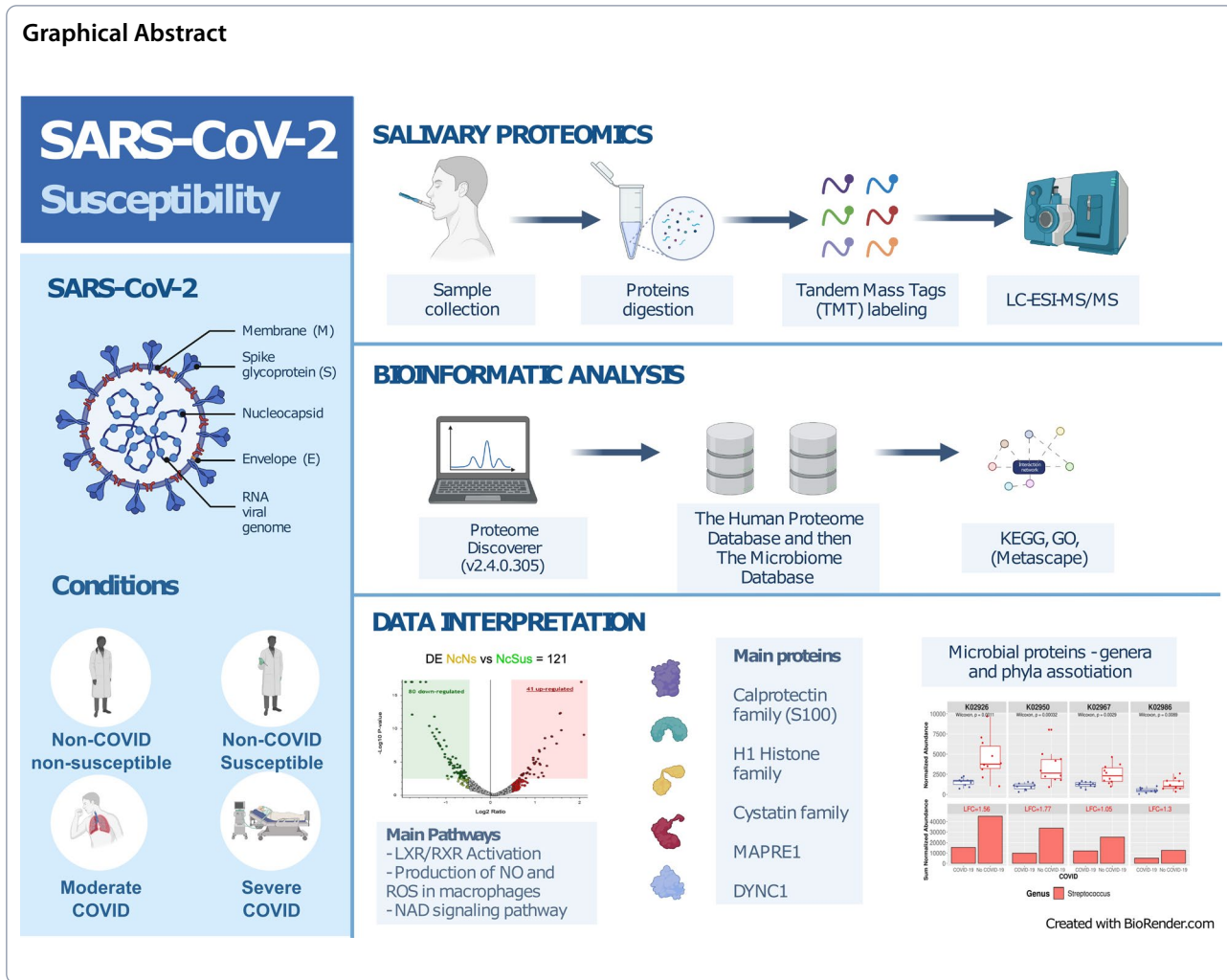
\*Correspondence:

Elena Moreno

emolmo@salud.madrid.org

Full list of author information is available at the end of the article





## Introduction

Although 3 years have passed since the beginning of the pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and a myriad of publications have tried to explain the mechanisms underlying its infectious process, only a few studies have focused on human and bacterial proteomic analysis in saliva to understand SARS-CoV-2 pathogenesis. As a respiratory virus, SARS-CoV-2 primary target cells reside in the nasopharyngeal and oropharyngeal mucosa. Saliva samples are noninvasive and inexpensive to obtain, providing an ideal site to discriminate between physiological and pathological conditions [1].

The characterization of the human salivary proteome led to the identification of more than 3000 proteins, mostly of microbial origin, supporting the notion that the oral microbiota is a major contributor to the whole salivary proteome [2]. Indeed, previous metagenomic studies have detected more than 10,000 taxa in saliva [3–6].

The functional relevance of the microbiota is supported by other groups that have identified microbial peptides from patients with a specific disease and described the functions associated with those proteins in that context [7, 8]. This approach seems especially relevant in the case of a respiratory virus since the probability of direct interactions with salivary proteins is high.

The pathophysiology of SARS-CoV-2 has been associated with an abnormal immune response, and a severe response to the infection can be facilitated by several comorbidities, such as diabetes, pulmonary disease, or cardiovascular disease [9, 10]. However, the role of saliva composition in disease pathogenesis remains unclear, and research has been carried out on other types of samples, such as plasma, nasopharyngeal swabs, feces, and semen, as will be detailed in the Discussion section. To address this question, we performed an exploratory study comparing unvaccinated patients hospitalized with different degrees of severe COVID-19 disease and healthcare

workers involved in COVID-19 care during the first wave in Madrid (Spain) without clinical or serologic evidence of developing infection despite repeated high-risk exposures to SARS-CoV-2. Saliva samples were collected and analyzed using a tandem mass tag (TMT)-based quantitative proteomics technique, and their proteomics profile was correlated with previously described SARS-CoV-2 susceptibility and COVID-19 clinical progression traits. Furthermore, we searched for alterations in saliva related to microbiota, which allowed us to find relevant functional bacteria associated with specific conditions. By using these complementary approaches, we show here some molecular principles describing the underlying mechanisms responsible for different susceptibilities to COVID-19, which can be relevant for the clinical management of COVID-19 patients.

## Results

### General characteristics of the study population

The study population included 20 individuals, 10 non-COVID individuals with negative polymerase chain reaction (PCR) test results and 10 hospitalized patients due to COVID-19 disease symptoms. Of the non-COVID individuals, 5 were healthcare workers considered non-susceptible to COVID-19 (NcNs), since they remained seronegative despite repeated high-risk exposures, and 5 were healthy controls but susceptible to COVID-19

(NcSus), i.e., they became infected during the follow-up. Of the infected patients, 5 were considered moderate cases (Mcov) (i.e., they did not require respiratory support by intubation), and 5 developed severe disease (Scov) (i.e., they needed intubation). The general characteristics of the study participants are shown in Table 1, and more details are displayed in the materials and methods.

### Human proteins in saliva related to different SARS-CoV-2 infection sensitivities

A TMT-based proteomics approach by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) was applied to analyze and compare the expression levels of proteins in saliva samples of the 20 individuals conforming to the four target COVID-19 susceptibility groups (Table 1; Additional file 1: Table S1A). Combining the information obtained from the four search engines, a total of 3378 human proteins were unambiguously identified with an FDR < 1% (Additional file 1: Table S1B). Similar to other studies focused on the analysis of biological fluids by mass spectrometry during COVID-19 infection [2], SARS-CoV-2 proteins were not detected in this study. Of the 3378 proteins, a group of 2721 proteins were quantified in two TMT experiments with two Internal Standards (IS) (Additional file 1: Table S1C), and different unique

**Table 1** General characteristics of the study population

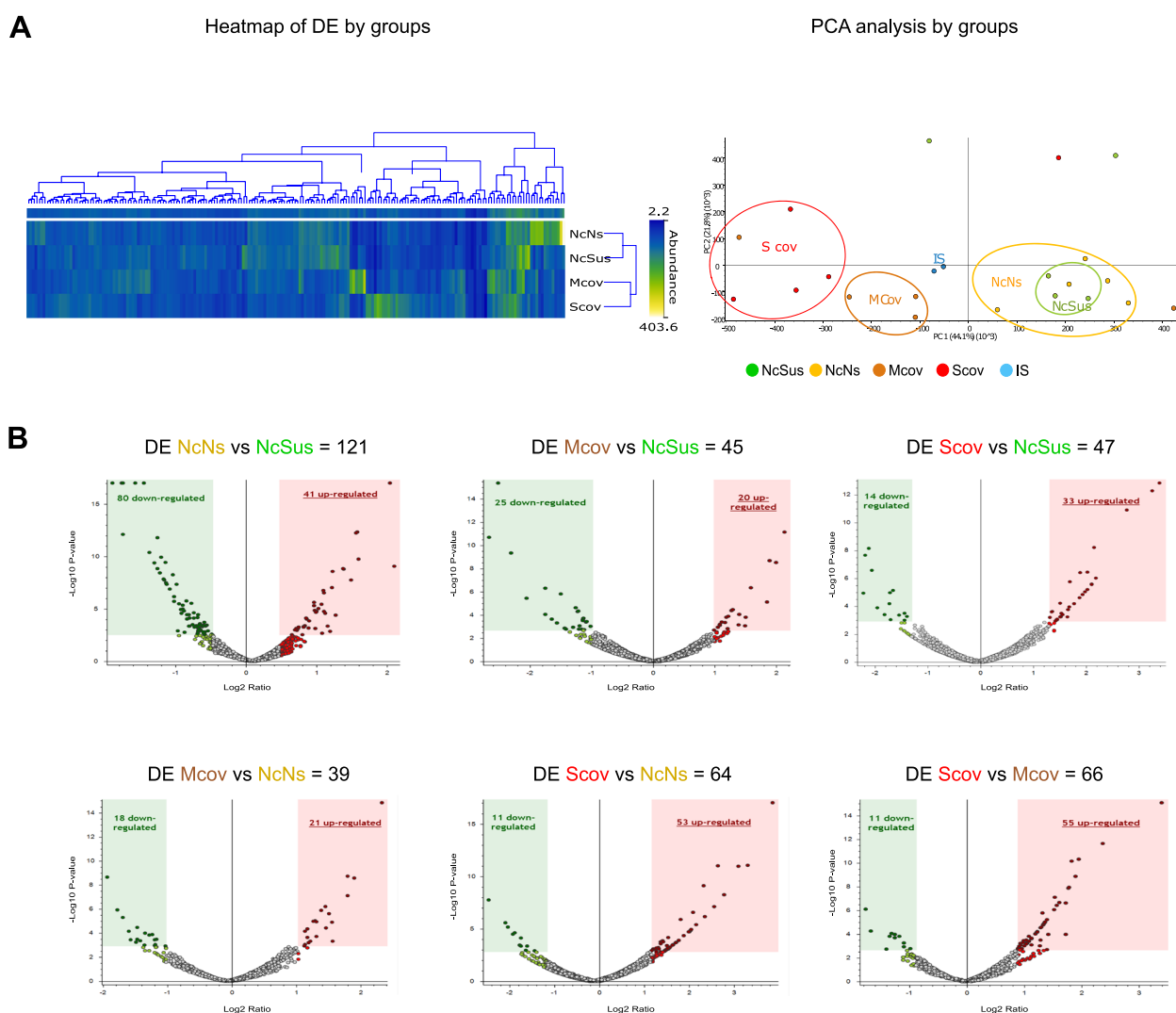
Study id	Group	Age (years)	Sex	Diabetes	ObeYest	Lung disease	Cardiac disease	Outcome
24	Non-COVID nonsusceptible (NcNs)	34	Female	No	No	No	No	Healthy
31	Non-COVID nonsusceptible (NcNs)	24	Male	No	No	No	No	Healthy
16	Non-COVID nonsusceptible (NcNs)	39	Male	No	No	No	No	Healthy
22	Non-COVID nonsusceptible (NcNs)	38	Female	No	No	No	No	Healthy
59	Non-COVID nonsusceptible (NcNs)	37	Female	No	No	No	No	Healthy
65	Non-COVID susceptible (NcSus)	44	Female	No	No	No	No	Healthy
71	Non-COVID susceptible (NcSus)	38	Female	No	No	No	No	Healthy
66	Non-COVID susceptible (NcSus)	57	Female	No	No	No	No	Healthy
43	Non-COVID susceptible (NcSus)	48	Female	No	No	No	No	Healthy
45	Non-COVID susceptible (NcSus)	51	Female	No	No	No	No	Healthy
26	Moderate COVID (Mcov)	91	Female	Yes	No	No	No	Survived/no intubation
3	Moderate COVID (Mcov)	69	Male	Yes	Yes	No	No	Survived/no intubation
5	Moderate COVID (Mcov)	89	Female	No	No	No	No	Survived/no intubation
74	Moderate COVID (Mcov)	78	Male	No	No	No	Yes	Survived/no intubation
68	Moderate COVID (Mcov)	75	Female	No	No	No	No	Survived/no intubation
157	Severe COVID (Scov)	58	Male	Yes	No	No	No	Survived/intubation
162	Severe COVID (Scov)	47	Male	No	No	No	No	Survived/intubation
35	Severe COVID (Scov)	60	Female	Yes	No	Yes	No	Survived/intubation
20	Severe COVID (Scov)	92	Male	Yes	No	No	No	Deceased
177	Severe COVID (Scov)	61	Male	No	No	No	No	Survived/intubation

proteins were found to be differentially expressed (DE) depending on the comparisons of the groups (Additional file 1: Table S1D–I). Both the representation of DE proteins on a heatmap and unsupervised principal component analysis (PCA) showed a clear separation between the non-COVID individuals (NcN and NcSus groups) and the infected patients (moderate COVID, Mcov and severe COVID, Scov groups) (Fig. 1A). Furthermore, Volcano plots show the best differences in the comparison between the NcNs and the NcSus groups (121 proteins), followed by Scov vs Mcov (66 proteins), Scov vs NcNs groups (64 proteins), Scov vs

NcSus (47 proteins), Mcov vs NcSus (45 proteins), and Mcov vs NcNs (39 proteins) (Fig. 1B; Additional file 1: Table S1D–I).

**Functional analysis of DE human proteins: biological meaning of different COVID-19 disease outcomes**

To understand the different roles of the DE human proteins found in saliva, a functional analysis was performed using Ingenuity Pathway Analysis (IPA) software (Qiagen). This analysis inferred the most significant pathways, upstream regulators, diseases, and biological functions associated with the DE proteins by comparison



**Fig. 1** Analysis of DE proteins in the saliva of patients with COVID-19. **A** Heatmap and PCA of DE proteins (see Additional file 1: Table S1) by samples divided into 4 groups [NcSus (n = 5; green), NcNs (n = 5; yellow); Mcov (n = 5; orange); Scov (n = 5; red)] plus the IS (internal standard). A heatmap was created using normalized abundance data, and branches represent grouping by similarities. PCA for overrepresented proteins was depicted using the log<sub>2</sub>-transformed abundances and explained ~44,1% of the variance in PC1 and 21,8% in PC2. **B** Volcano plots show the proteins that are DE between the different study groups (upregulated proteins are represented in dark red and downregulated proteins in dark green). The statistical thresholds used for these analyses are detailed in the Materials and Methods section

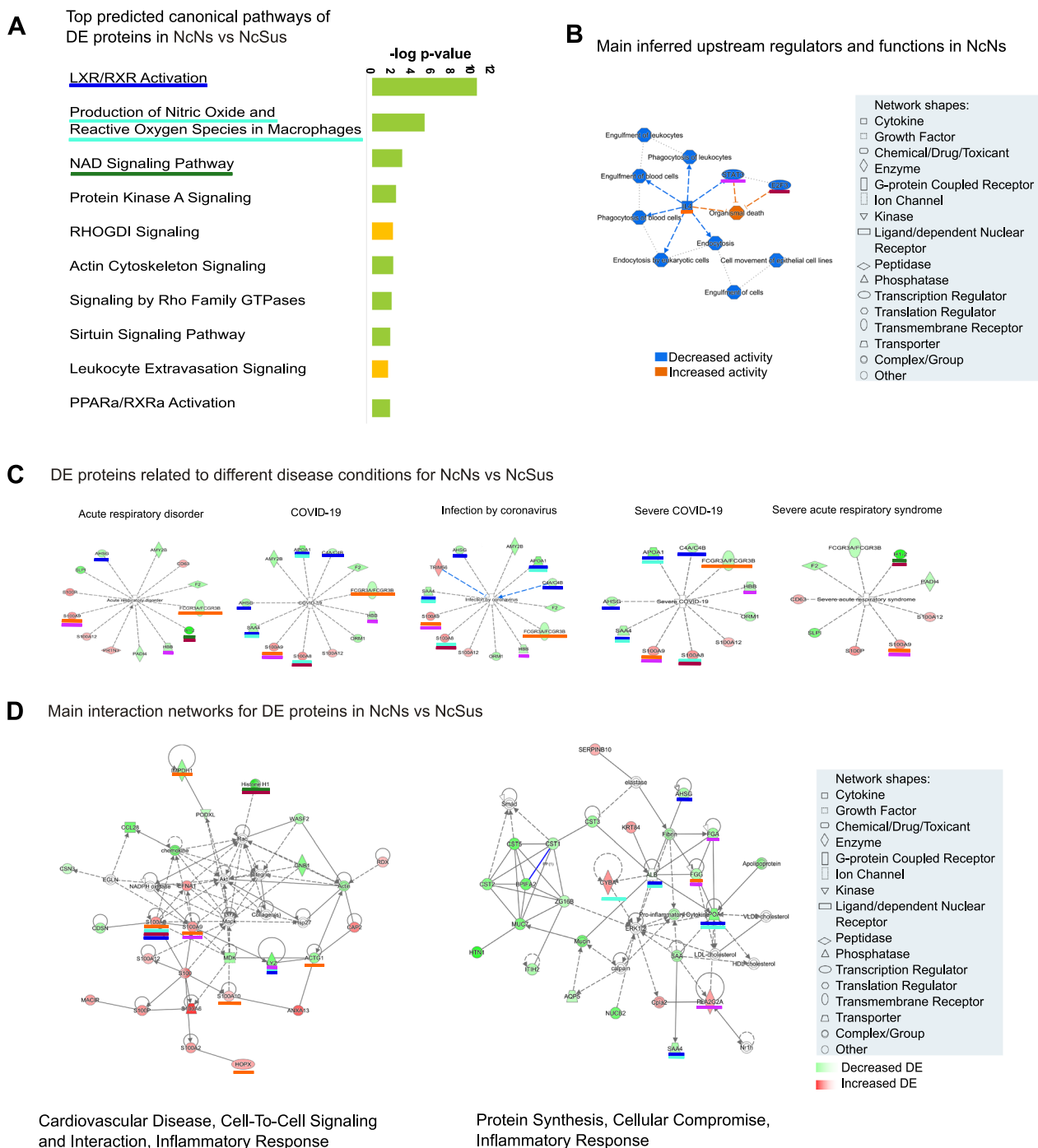
with the Ingenuity Knowledge database. This tool has been applied before to investigate canonical pathways of human proteomes in nasopharyngeal swabs from COVID-19 patients [11]. This analysis was performed for the DE proteins found in paired comparisons for all the groups (Additional file 1: Figure S1-5, Additional file 1: Table S2A–F). However, to understand the underlying mechanisms that could explain the different susceptibilities to COVID-19 disease, we focused on the proteins found to be DE in the non-COVID-19 nonsusceptible condition (NcNs) compared to the non-COVID-19 susceptible condition (NcSus) (Fig. 2 and Additional file 1: Table S2B). In this analysis, 10 canonical pathways showed relevant differences ( $-\log_{10}(\text{p-value}) > 1.8$  and z-score value  $> 0$ , highlighted in yellow in Additional file 1: Table S2). B), such as those related to LXR/RXR activation, production of nitric oxide and reactive oxygen species (ROS) in macrophages and NAD signaling (Fig. 2A). From this same analysis, we also inferred the main upstream regulators of those DE proteins with a predicted decreased activation state in the NcNs group. These include interleukin-4 (IL-4), signal transducer and activator of transcription 3 (STAT3), and transcription factor E2F3 (E2F3), with regulatory functions related to cellular mechanisms (endocytosis and cell movement), blood cell processing (e.g., phagocytosis of leukocytes), and regulation of cell apoptosis (Fig. 2B).

Furthermore, IPA analysis of these proteins showed a pattern associated with a low susceptibility to SARS-CoV-2. Some of the proteins specifically decreasingly or increasingly expressed, which are represented by different colors, are related to COVID-19 or acute respiratory disorder in general, as shown in Fig. 2C. Of these, the expression of apolipoprotein A-I (APOA1), alpha-2-HS-glycoprotein (AHSG), hemoglobin beta chain (HBB), serum amyloid A-4 (SAA4), alpha-1-acid glycoprotein 1 (ORM1), and complement C4A (C4A) was found to be significantly reduced in the NcNs group compared to the NcSus group. Furthermore, some of these proteins were also found to be upregulated in the saliva of patients with severe and moderate COVID-19 (Scov and Mcov) when compared with susceptible and/or nonsusceptible individuals (Additional file 1: Table S1D, E, H, I and Additional file 1: Figure S1-5). Interestingly, calprotectin protein family (S-100) expression, including S100A8, S100A12, S100A9, and S100P, was found to be increased in the NcNs group compared with the NcSus group (red in Fig. 2D), in contrast to their decreased expression during COVID-19 infection conditions (Additional file 1: Table S1C-Scov or Mcov conditions abundance ratio with adj p-values ranging from  $2.28 \times 10^{-2}$  to  $1.96 \times 10^{-5}$ ). Furthermore, related to the predicted canonical pathways by the IPA analysis, it is worth mentioning the LXR/RXR

activation pathway with APOA1, AHSG, SAA4, ORM1, S100A8, and C4A being involved and the STAT3 pathway with AHSG, HBB, and S100A9 and APOA1, SAA4, and C4A (Additional file 1: Table S2, sheet B, highlighted in yellow and Fig. 2C and D, blue underlined). These proteins are related to cholesterol transport to tissues and lipoprotein synthesis (HDL), which acts as a scavenger of viruses, immune modulator, and mediator of viral entry.

To further understand the mechanisms related to the DE proteins in the nonsusceptible group, we selected the two main interaction networks predicted in the analysis according to the IPA score and showed the main processes associated with them: they range from cell-to-cell signaling and interaction to protein synthesis, cellular compromise, and inflammatory response (Fig. 2D). Furthermore, we underlined in matching colors the main protein families in these interaction networks with the predicted canonical pathways (Fig. 2A) and/or upstream regulators (Fig. 2B). To summarize these results, the main DE proteins and predicted functions related to SARS-CoV-2 susceptibility obtained in the IPA analysis are described in Table 2.

Next, to further understand the potential role specifically in the COVID-19 context of some of the DE human proteins in nonsusceptible patients, we compared the 121 DE proteins obtained in the analysis of NcNs vs NcSus with all the -omic studies published that identified direct interactions and biological pathways related to SARS-CoV-2 by using Metascape [12]. The main discriminatory pathways were specific for microbial infections, immune response, and cellular growth and were even specific for salivary secretion (Fig. 3A). Clustering coefficients calculated using enrichment analysis and the Cytoscape and MCODE algorithms for subnet analysis identified three main clusters of proteins (Fig. 3B, Additional file 1: Table S3). The first cluster contains the above mentioned proteins related to LXR/RXR activation and cholesterol metabolism and other related proteins, such as cystatin C (CST3), interalpha-trypsin inhibitor heavy chain H2 (ITIH2), fibrinogen gamma chain (FGG), and pseudokinase FAM20A (FAM20A). Therefore, the levels of expression of these proteins follow the same tendency as the others. The second cluster is related to apoptosis-induced DNA fragmentation, nucleosome assembly and positioning and includes proteins found to be downregulated in NcNs, including several isoforms of histone H1, amine oxidase [flavin-containing] B (MAOB), and high mobility group protein B2 (HMGB2). The last cluster includes proteins involved in Rho GTPase-mediated regulation and organization of polymeric cytoskeleton rearrangement. Cytoplasmic dynein 1 (DYNC1), microtubule-associated protein RP/EB family member 1 (MAPRE1), and cytoplasmic actin, 2 (ACTG1) were downregulated



**Fig. 2** Results of the predicted functional analysis of the proteins found DE in saliva of NcNs compared NcSus patients by IPA software. **A** Top 10 canonical pathways common for the two groups by the top 10 were selected from Additional file 1: Table S2 (highlighted in yellow) by selecting the ones with the highest  $\log_{10}$  ratio (observed/expected) and adjusted p-value cutoff  $\leq 0.5$ . **B** Graphical summary with the main upstream regulators with predicted increased (orange) and decreased (blue) activity. **C** Diseases and biological functions related to DE proteins, including infectious and respiratory diseases associated with infection by SARS-CoV-2. **D** Main interaction networks and related upstream regulators/molecules predicted by IPA (white background). Upregulated molecules are depicted in red and downregulated molecules in green according to adjusted p-values of overlap calculated by IPA software and detailed in Additional file 1: Table S2

**Table 2** Summary of the main functions and proteins related with SARS-CoV-2 susceptibility

Proteins	Pathway or function related
S100A8, LYZ, AHSG, APOA1, and SAA4	LXR/RXR activation and production of nitric oxide and ROS in macrophages
CYBA	oxidative stress-related pathways
H1 family	NAD signaling pathway and E2F3 (upstream regulator)
S100A9, LYZ, fibrinogen chains (FGA and FGG) and PLA2G2A	STAT3 immune pathway
S100 protein family, IMPDH1, ACTG1	IL-4 (act as regulator)

in the NcNs compared to NcSus, in contrast with axonemal dynein 17 (DNAH17), whose levels were significantly higher in NcNs compared to NcSus and COVID<sup>+</sup> groups. All the proteins in these subnetworks were also highlighted by IPA analysis, although some of them were not in the main networks (FAM20A, MAOB, DYNC1, MAPRE1, and DNAH17) (Fig. 3).

Furthermore, by comparison with other databases (TRRUST, Transcription Factor Targets, PaGenBase), different specific functions were found to be altered. Three relevant transcription factors, namely, SP1 (TRRUST database; also identified in the IPA analysis), which has been described as involved in viral transcription [13–16], LMO2, a regulator of T-cell translocation [17, 18], and AP1, are related to viral infection and ROS [19, 20] (Transcription Factor Targets database). Finally, the proteins were associated to some specific tissues and cell types, such as salivary gland and lung or skeletal muscle and lung cells (Figure S6;  $-\log_{10}(p\text{-value}) > 4.0$ ).

#### Microbiome composition and role in SARS-CoV-2 infection susceptibility

Because different microbiota features have been described as relevant actors for disease outcomes, we decided to explore some of these traits to understand the link between the active salivary microbiome and COVID-19 disease. According to the strict confidence intervals detailed in the materials and methods section, we identified and quantified a total of 500 microbial proteins (Additional file 1: Table S4B). We compared, using the ANOSIM test, the abundance of individual microbial proteins, as well as their KO and taxonomy, in the saliva microbiome according to the presence of SARS-CoV-2 infection: COVID-19<sup>+</sup> (Mcov & Scov) and COVID-19<sup>-</sup> (NcNs & NcSus). While no statistically significant difference was found when examining protein expression profiles, we identified 4 translation-related Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) genes overexpressed in the COVID-19<sup>-</sup> group: K02926, K02950, K02967, and K02986 (Fig. 4,  $p\text{-value} \leq 0.001$ ). Their expression was decreased in the COVID-19<sup>+</sup> group, suggesting that the translational machinery, and

thus microbial protein biosynthesis, of the microbiota in COVID-19<sup>+</sup> patients might be impaired. This result agrees with the fact that protein synthesis has been found to be altered in nasopharyngeal swabs from COVID-19 patients [11]. Notably, proteins associated with KOs overexpressed in the COVID-19<sup>-</sup> group were assigned to bacteria from the *Streptococcus* genus. This suggests that although this taxonomic group has been found to be highly abundant in patients with COVID-19 [21], at the functional level, its involvement in pathogenesis should be limited.

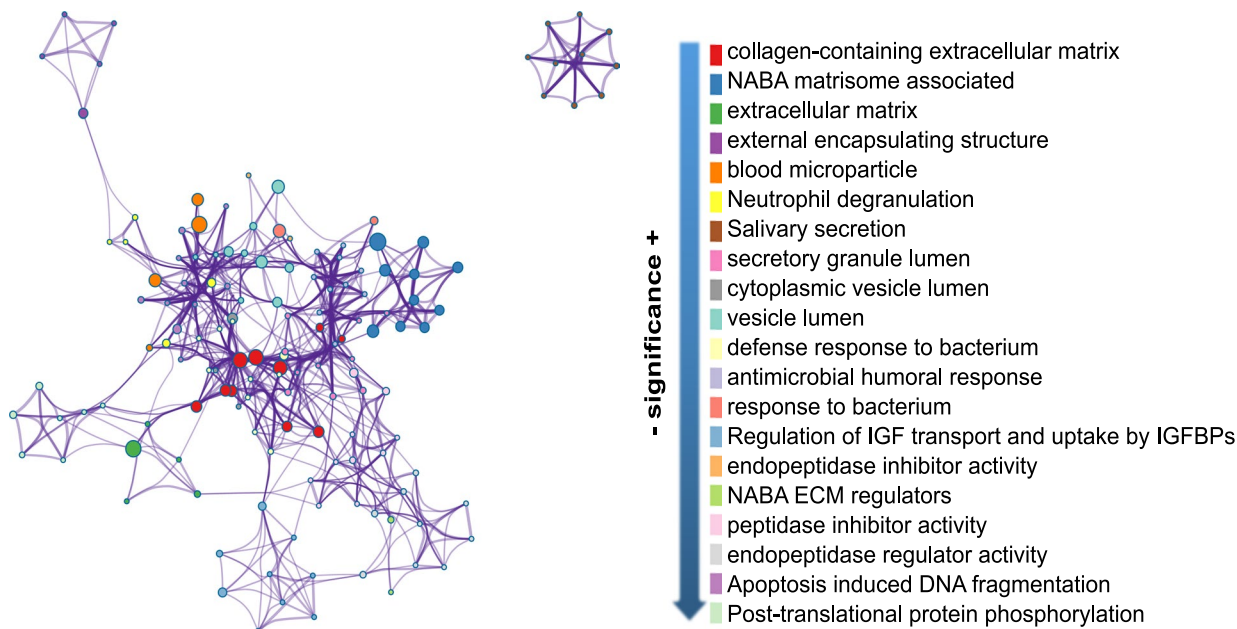
#### Targeted correlations of protein expression with ACE2 activity in plasma

Finally, because angiotensin-converting enzyme 2 (ACE2) is the molecular target for SARS-CoV-2 cell entry, we analyzed the correlations between the DE proteins consistently identified in the Metascape analysis and matching with the IPA analysis (FAM20A, MAOB, DYNC1, MAPRE1, and DNAH17) with ACE2 activity. In Fig. 5, we show the positive and significant correlations found between the DYNC1 and MAPRE1 proteins and ACE2 activity. These two proteins are related to protein transport through microtubules in the cytoplasm of cells, possibly affecting viral replication by allowing viral proteins to reach the Golgi apparatus or endoplasmic reticulum to replicate and translate their proteins. Furthermore, both proteins were decreasingly expressed in the Non-COVID and nonsusceptible groups (Additional file 1: Table S1), indicating a potential role of DYNC1 and MAPRE1 in SARS-CoV-2 susceptibility by affecting ACE2 levels in plasma.

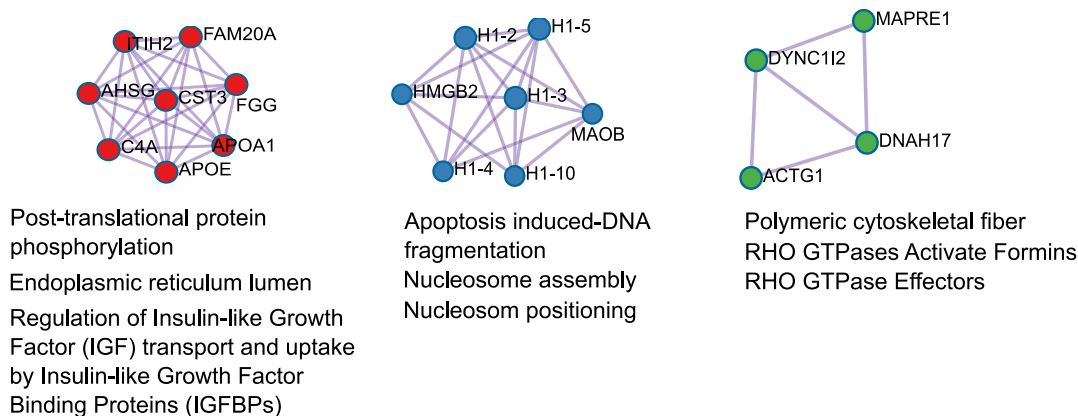
#### Discussion

Patients with COVID-19 have been reported to have significant alterations in oral (tongue-coating) and gut microbial diversity [21–23], as well as alterations in the expression level of proteins from fecal microbiomes [24, 25], plasma human proteomes [26], nasopharyngeal swabs [27–29] and semen samples [30]. Proteome profiling of nasopharyngeal swabs from COVID-19 patients further revealed alterations in proteins

**A** Main interaction networks for NcNs vs NcSus compared to all COVID-19 omic databases



**B**

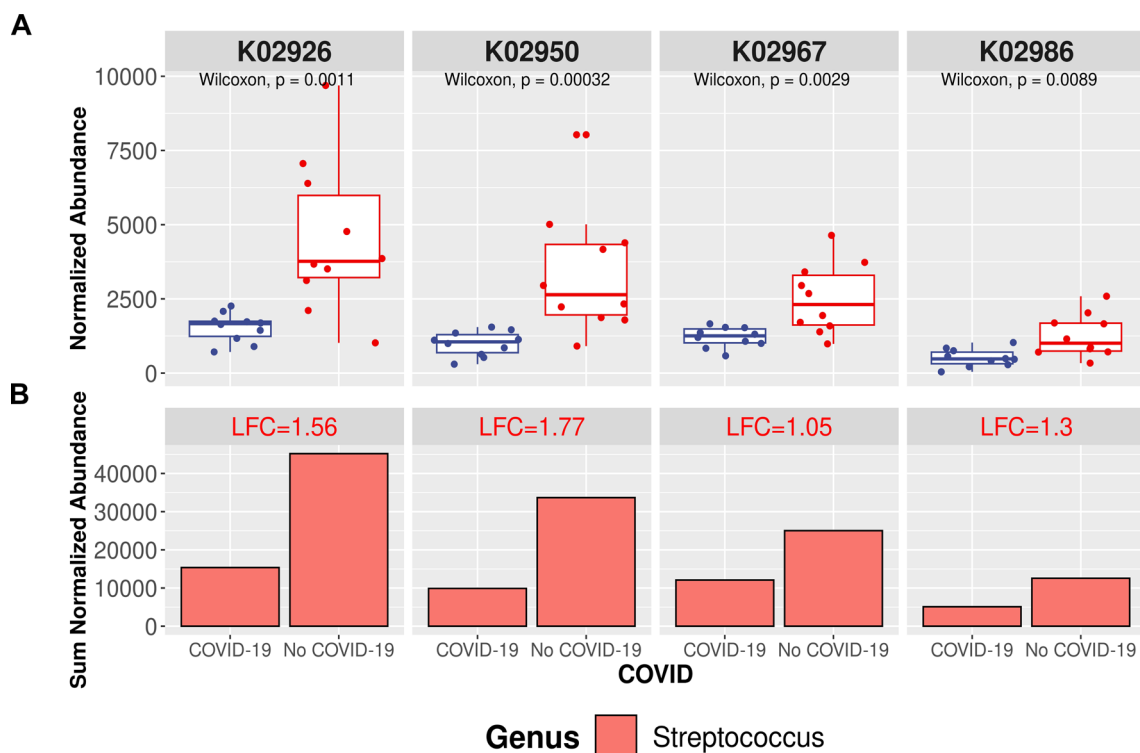


**Fig. 3** Comparison of DE proteins in the saliva of the NcNs vs NcSus groups with all omics COVID-19 databases by Metascape. **A** Main interaction network predicted by Cytoscape (left) and list ordered by more probable functions related to the DE proteins selected (right). The colors of the functions match the colors in the network. The size of the nodes represents the number of proteins associated with that function (the bigger the node, the higher the number of the selected proteins involved in that function). **B** Main subnets of the protein interactions inferred by the MCODE algorithm and their main predicted biological functions (below) according to a p-value < 0.01. Detailed information about the enrichment analysis, statistical values and specific proteins associated with the functions is further detailed in Additional file 1: Table S3 and its caption

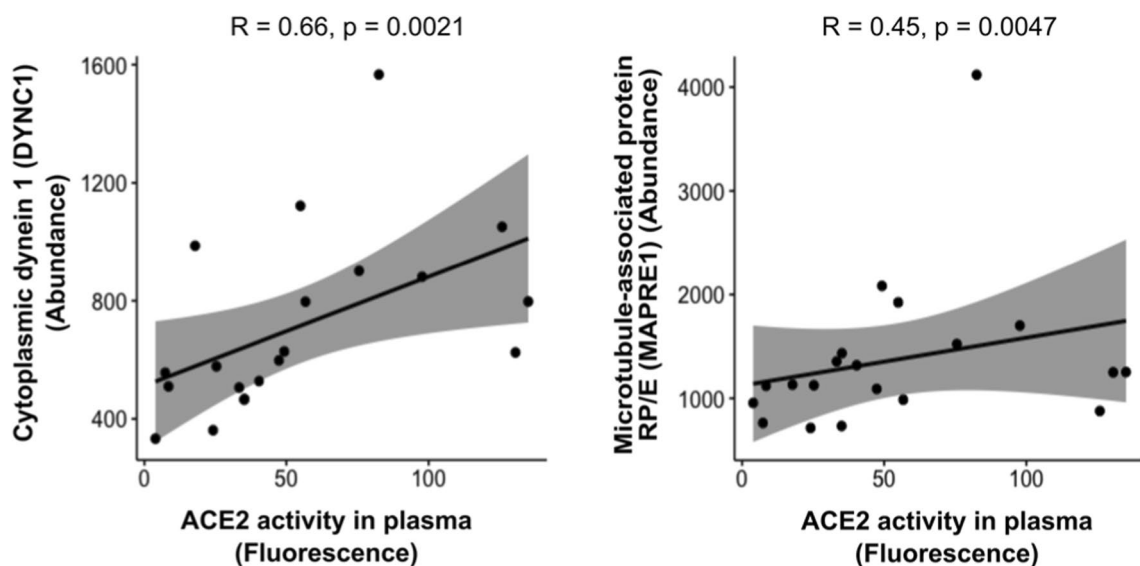
implicated in the innate immune response, viral assembly, and exocytosis [31]. To the best of our knowledge, only one study has focused on the application of untargeted mass spectrometry to compare the saliva

proteome of SARS-CoV-2-infected and non-COVID-19 individuals. Although differences in the expression of a human protein related to the defense/immunity protein, an immunoglobulin, a translational protein, and a protease inhibitor were observed, no clear link with the





**Fig. 4** Comparison of DE KOs in saliva of patients affected by COVID-19 and not affected in terms of Wilcoxon test (upper panel) and bacteria-contributing *genera* (lower panel). Box plots show the differentially presented KOs according to a p-value < 0.01. Bar plots represent the sum of the normalized abundance of the proteins associated with them (A) and the taxonomy at the *genus* level. B LFC (log<sub>2</sub>-fold change (FC)) means log<sub>2</sub>(A/B), where A is the sum of the normalized abundance of the proteins contributing to the KO for patients affected by COVID-19 and not affected in the case of B



**Fig. 5** Significant correlations of factors identified in Metascape and IPA analysis with previous ACE2 fluorescence levels quantified in plasma. Graphs of correlations for Dynein 1 (DYNC1) and microtubule-associated protein RP/E (MAPRE1) abundances (normalized) with ACE2 fluorescence are shown. As described in Material and Methods, abundances were calculated based on TMT label intensities normalized using an Internal Standard control (Additional file 1: Table S1) and fluorescence was measured at 420 nm (arbitrary units). Linear trends were analyzed, and their fit is depicted by the Pearson R coefficient and the p-values in each graph

mechanisms of pathogenesis and response to SARS-CoV-2 could be established [32].

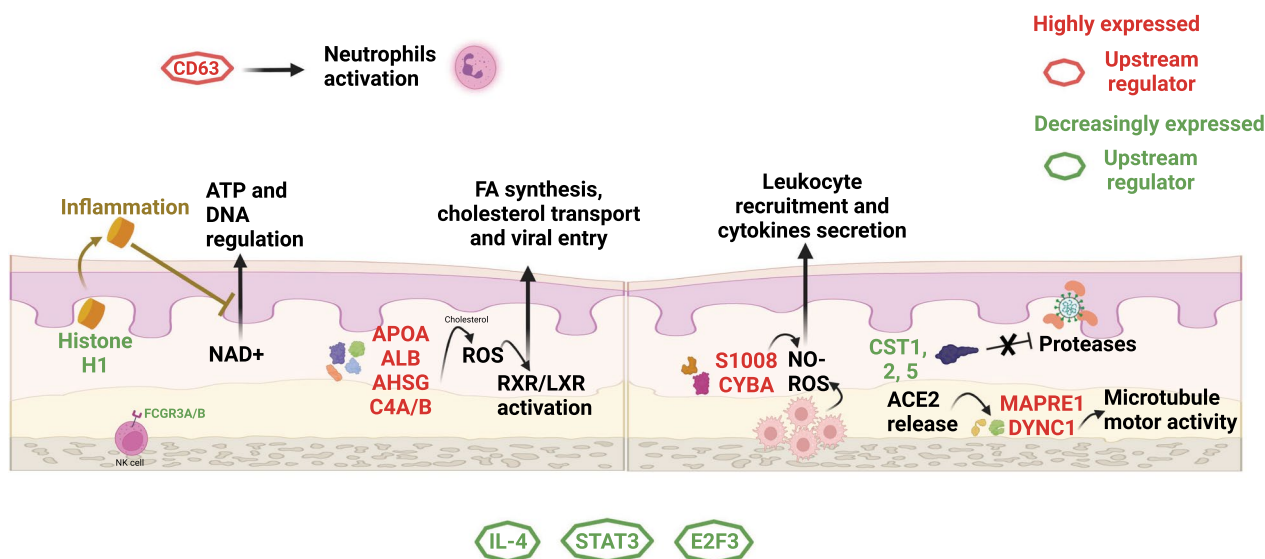
This exploratory study aimed to describe DE proteins that could be related to COVID-19 disease susceptibility. Several proteomic analyses have been published to date to find causal candidates or drug targets to defeat SARS-CoV-2 infection, but very few have been performed in saliva, the only sample from the main infection sites of the virus. Furthermore, our study includes an extraordinary group of study, the nonsusceptible group, formed by a small group of heavily exposed healthcare workers who were not infected by SARS-CoV-2. This is important to understand possible specific mechanisms that might be responsible for some protective status of certain people who do not get infected although they are heavily exposed to the virus.

We are aware of the main limitations of our study: the sample size was small due to the difficulty at that time in selecting patients for each specific condition; it is limited to the first viral variant and cannot be extrapolated to other variants; and there might be confounding factors such as diabetes, age and sex. Diabetes has been shown to affect the salivary proteome in previous studies [33–36] and 5 subjects in our study had this condition. Age is also an important factor, since the proteome of adults over 60 years is different from younger individuals [37–39] and we had age differences in the moderate COVID-19 group. However, we did not see these differences in the main comparison of groups: NcNs vs NcSus. Furthermore, as a quality control of our study, to determine the

identified human proteins with the highest confidence we filtered and selected only the proteins appearing in two experiments with two internal standards, and even doing that, we obtained similar or even higher numbers compared to previous studies in nasopharyngeal swabs [11] and in saliva samples [32]). It also constitutes one of the best studies reporting the combined human proteome and microbial meta-proteome in saliva samples, since no previous study reported alterations in canonical human and microbial pathway analysis in saliva samples. However, other technical validations, such as western blot, were not performed, since they are out of the scope of this particular study and samples were limited.

When we analyzed the DE proteins of the nonsusceptible group compared to the susceptible group, we found some interesting findings. The most important observations are summarized in Fig. 6 for clarification and will be described in the following paragraphs.

When analyzing the described groups, the NcNs and NcSus comparison yielded the most significant differences (Fig. 1). Focusing on this comparison, we found that the main canonical pathway was the LXR/RXR activation pathway. The liver X receptor forms a heterodimer with the retinoid X receptor, and they participate in the regulation of lipid metabolism, inflammation, and cholesterol transportation [40, 41]. Furthermore, its inhibition has been associated with prolonged viral RNA shedding [42]. Related to this pathway, we found molecules mainly synthesized in the liver and related to HDL synthesis and cholesterol transport, immune modulators, or regulators



**Fig. 6** Summary of the main pathways and DE proteins in nonsusceptible individuals identified in this study. Highly expressed proteins are depicted in red, and decreasingly expressed proteins are depicted in green. An octagon shape has been added to the proteins that have roles as upstream regulators of processes involved in SARS-CoV-2 pathogenesis and that came up in our analysis. (This figure can go to Additional file 1)

of viral entry, such as APOA1, AHSG, or C4A. Another important canonical pathway in our analysis is the NO production pathway. The production of nitric oxide and ROS by activated macrophages is essential to the control of infections and is regulated by IFN-gamma. Specifically, this pathway has been related to the pathophysiology of the cytokine storm and severe COVID-19 [43, 44]. Some DE proteins related to these functions were increasingly expressed in the nonsusceptible group, such as S100A8 and CYBA. S100 proteins are part of the calprotectin dimer, which has been related to COVID-19 severity [45, 46], and CYBA has been described as an interactor of the Nef protein of HIV-1 [47]. The dimer SA8/A9 is released by neutrophils and macrophages during inflammation, so its overexpression before infection could have a protective effect [48]. Another significant canonical pathway was the NAD signaling pathway. NAD is the active metabolite of vitamin B3 and is essential for cellular processes of energy metabolism, cell protection, and biosynthesis [49]. Moreover, its function was previously related to inflammation and coagulation present in COVID-19 [50, 51]. Decreased expression of histone H1 appears to be related to this pathway in our analysis. This could be due to a protective role, since histone H1 acts as a sensor of NAD<sup>+</sup>, which is related to ATP production and DNA regulation. It has also been shown that extracellular histone H1 is released in response to inflammatory challenges, reducing NAD<sup>+</sup> levels and thus maintaining low NAD<sup>+</sup> protection roles [52, 53]. Furthermore, in our previous study with these same samples, we found a link with energy metabolism and biosynthesis, where some compounds, such as citrulline, citric acid, and histidine decrease, and BAIBA, phenylalanine, and 2-AB increase were related to mitochondrial and liver dysfunction as a consequence of hypoxemia associated with SARS-CoV-2 severity [54], and some of the DE proteins found in this study are also involved in metabolic pathways, such as cholesterol-NO-ROS, cholesterol-LXR or tryptophan-NAD<sup>+</sup> pathways.

The main upstream regulators identified in our analysis and associated with organismal death were IL-4, STAT3, and E2F3, whose activity was predicted to be decreased by IPA analysis based on the expression of the proteins in our dataset. IL-4 is related to IFN-gamma, IL-17A, and STAT6, all factors involved in T-cell differentiation to helper T cell and vitamin D3 receptor [55, 56]. STAT3 is a factor related to IL-6 and NF-KB, is known as an antiviral and proinflammatory mediator [57], and has been largely associated with COVID-19 disease severity [57–63]. E2F3 is a transcription factor of genes involved in the cell cycle and is known to be regulated by some viruses, such as HIV, HCV, and EBV [64–66]. Moreover, E2F3 is usually downregulated by miRNAs to avoid cell proliferation,

which is usually very related to viral infections [64, 67, 68]. Previous analysis of nasopharyngeal swabs from COVID-19 patients found altered expression of proteins involved in the innate immune response, cell death and inflammation in COVID-19-infected patients. Among these proteins, only IFN and STAT proteins were also found to be associated with COVID-19 in saliva samples, which suggests different responses in oral (saliva) and nasopharyngeal environments during SARS-CoV-2 infection [11].

Regarding specific proteins, in nonsusceptible individuals, we observed higher expression in general of the S100 proteins (calprotectin) family, which are calcium- and zinc-binding proteins produced by immune cells and involved in several cellular processes, including cycle progression and differentiation, immune regulation, inflammation, etc. [48, 69, 70]. Increased expression of S100A6 is related to signal transduction, ion transmembrane transport, positive regulation of fibroblast proliferation, or protein binding [71]. Specifically, elevated levels of proteins of the calprotectin family have been related to COVID-19 severity [44, 45]. Decreased expression of the histone H1 family is related to NAD signaling, protein kinase A signaling, and sirtuin signaling pathways [72], which are all related to cellular metabolism and growth. The histone H1 family acts as a regulator of transcription through chromatin remodeling, nucleosome spacing, and DNA methylation, which are associated with the NAD signaling pathway. Decreased expression of FCGR3A/B, which are IgG receptors of NK cells and polymorphonuclear neutrophils, is implicated in antibody-dependent enhancement of viral infections. Highly expressed TRIM56, a previously described antiviral host that confers resistance to human coronavirus OC43, dengue virus, and yellow fever virus (15). Mucin and AQP5 are related to the calpain protease family, which has been used as a target against SARS-CoV-2 infection by M protease inhibitors (29).

Another relevant group of proteins found to be associated with NcNs is the cystatin family, including cystatin-SN (CST1), cystatin-SA (CST2), CST3, and cystatin-D (CST5). All these proteins were found to be significantly downregulated in the NcNs and COVID+ groups when compared with NcSus, except for CST3, which was only downregulated in the NcNs group. Cystatin 5 (or D) is an active cysteine protease inhibitor that plays a protective role against specific proteases present in the oral cavity [73]. It has also been previously described as overexpressed during other virus infections, such as hepatitis B infection [74]. High concentrations of some cystatins have been associated with higher COVID-19 severity and mortality [75]. Thus, the decrease in the levels of cystatins in nonsusceptible individuals may be a protective

mechanism that helps to maintain the peptidase activity level, which could help to fight potential viral infections.

Another interesting protein that appeared as highlighted in the networks related to acute respiratory disorder and syndrome in our analysis for the nonsusceptible group was the CD63 protein. CD63+ activated neutrophils have been related to the acute phase of SARS-CoV-2 infection [76] and to ZIKV infection protection as a marker in EVs [77].

Finally, a correlation analysis with ACE2 levels in plasma showed a significant correlation with DYNC1 and MAPRE1. DYNC1 acts as a motor for the intracellular retrograde motility of vesicles and organelles along microtubules through ATPase activity. This activity has been related to RNA binding and microtubule motor activity and is associated with neuropathies [78]. Regarding SARS-CoV-2, the use of motor proteins, kinesins, and dynein, prominent in neuronal pathways, has been shown to be essential for viral migration through olfactory nerves to the brain [79]. The olfactory lobe, a part of the brain that translates smell responses, is the starting point of dysfunction, which partly explains the early signs of inability to smell, i.e., anosmia [80, 81]. Dynein is part of the endosome unit that mediates viral entry and then localizes and releases it in the right place of the cell, facilitating viral replication [82]. MAPRE1 regulates microtubule polymerization and mediates cargo transport. This role has been shown during HIV infection [83]. The fact that the expression of these proteins is decreased in the nonsusceptible group in our analysis and that they correlate positively with plasma ACE2 activity, which has been described before as increased following SARS-CoV-2 infection [84], supports that they might play a specific role during SARS-CoV-2 infection.

In summary, this study is the first to explore, through the analysis of DE proteins, possible alterations in canonical human and microbial pathways in saliva samples and their association with SARS-CoV-2 infection [64–66]. Although several studies from plasma, semen or cell proteomes have been published, the biological processes found to be altered in saliva differ from those in other types of samples, and thus, this study provides a complementary view into COVID-19 pathogenesis. Furthermore, since microbiota proteins are being revealed as potentially relevant factors in the outcome of some diseases [85–87], our exploration of a potential role in COVID-19 pathogenesis might seed some interesting facts.

## Materials and methods

### Study design and setting

Participants confirmed SARS-CoV-2 (COVID-19<sup>+</sup> group) infection by polymerase chain reaction (PCR)

from nasopharyngeal swabs, sputum, or lower respiratory tract secretions within the first 7 days from the onset of symptoms. Then, they were classified according to clinical severity as follows: moderate disease, defined as the presence of bilateral radiologic infiltrates or opacities and clinical assessment requiring supplemental oxygen; and severe disease, defined as the development of acute respiratory distress syndrome [31]. Participants without SARS-CoV-2 (non-COVID group) were asymptomatic subjects with a negative PCR from nasopharyngeal swabs. These individuals were considered “susceptible” to COVID-19 since they had positive IgG for SARS-CoV-2 or previous COVID-19 confirmed by PCR from nasopharyngeal exudate (non-COVID susceptible, NcSus). In contrast, the group considered as nonsusceptible adults were healthy healthcare workers who were on duty for at least 3 months in COVID-19 wards or intensive care units and reported at least three high-risk exposures to SARS-CoV-2 [88] without having experienced symptoms suggestive of SARS-CoV-2 infection, were persistently negative for SARS-CoV-2 PCR testing and did not have SARS-CoV-2 IgM and IgG in plasma (non-COVID nonsusceptible, NcNs). Additionally, this nonsusceptible group differs from the Non-COVID group in that they did not get infected either at any time in the next 5 months, in contrast to the Non-COVID group, in which the 5 individuals eventually became infected at some point during the next 5 months. The most frequent exposure was largely unprotected exposure to aerosol-generating procedures or patient secretions and close contact without facemasks with other confirmed cases of COVID-19. We measured SARS-CoV-2 antibodies by indirect chemiluminescence immunoassay (Vircell, Granada, Spain). Samples were coded for further proteomic analysis as indicated in the “Coding” sheet in Additional file 1: Table S1.

### Sample collection and treatment

Unstimulated saliva samples (5 ml approx) were collected consistently across all participant, from fasting patients early in the morning, to ensure minimal variation in the saliva composition. Any additional support or device were used to avoid contaminations with other secretions, such as sputum.

Cryopreserved (− 80 °C) saliva samples were inactivated, reduced and alkylated with a 1:1 (v:v) volume of 2X denaturing buffer (10% sodium dodecyl sulfate (SDS)+200 mM triethylammonium bicarbonate (TEAB)+10 mM tris(2-carboxyethyl)phosphine (TCEP)+10 mM chloroacetamide (CAA)), followed by incubation at 60 °C for 30 min. The total amount of protein was quantified by PIERCE 660 nm reagent (compatible with ionic detergents). Protein digestion on S-Trap

columns (Protifi) was performed following the manufacturer's instructions with minor changes [89]: 50 µg of each protein sample was digested at 37 °C overnight using trypsin protease (ratio 1:15, enzyme-protein). A pool of all digested samples was included as an internal standard (IS) in each experiment. After overnight incubation, the peptides were eluted, dried in a speed-vac and quantified by fluorimetry (QuBit). For TMT-11plex labeling, the dried digests were distributed into two labeling experiments: 20 µg of each of the 11 samples per experiment were labeled with the corresponding TMT 11-plex tag. After two hours, the labeled digests from each batch were pooled and fractionated.

#### Basic-pH fractionation using SDB-RPS STAGE Tips

The fractionation of TMT-labeled peptides was performed using in-house-made STAGE tips prepared from SDB-RPS solid-phase extraction disks (Empore™) similar to previously reported protocols [90, 91]. The STAGE tip was prepared according to the protocol previously described by our research group [92]. Briefly, 12 pieces were cut off from SDB-RPS solid-phase extraction disks with the assistance of a 16-gauge blunt end needle and packed into a 200 µL tip. The STAGE tip was inserted onto the top of a 2-mL tube using an in-house-made adapter and activated with 100 µL MeOH and centrifuged at 900×g for 3 min. The tip was then conditioned with steps of 100 µL 50% acetonitrile (ACN) and 0.1% formic acid (FA) and centrifuged, followed by three equilibration steps with 0.1% FA. After that, TMT-labeled peptides (80 µg) were reconstituted in 100 µL 1% FA (pH < 3) and loaded onto the STAGE tip. The sample was centrifuged at 900×g for 5 min, and the collected flow-through was loaded again to improve the peptide recovery yield. The STAGE tip was washed with 100 µL 0.1% FA, followed by 100 µL H<sub>2</sub>O. The elution was carried out using a 10-stepwise elution with 100 µL of 5 mM ammonium formate buffer and increasing acetonitrile concentrations (0, 5.0%, 7.5%, 10.0%, 12.5%, 15.0%, 17.5%, 20%, 25%, and 60%). Fractions were dried in a speed vacuum and frozen until further processing.

#### Analysis by liquid chromatography coupled to mass spectrometry OE-240

Each fraction (10 fractions in total) of the TMT-labeled experiments was quantified by fluorimetry (QuBit), and a 1 µg aliquot of each fraction was subjected to 1D-nano LC-ESI-MS/MS analysis using an Ultimate 3000 nano HPLC system (Thermo Fisher Scientific) coupled online to an Orbitrap Exploris™ 240 mass spectrometer (Thermo Fisher Scientific). Tryptic peptides (200 ng/µL) were eluted onto a 50 cm×75 µm Easy-spray Pep-Map C18 analytical column at 45 °C and separated at a

flow rate of 300 nL/min using a 90 min gradient ranging from 2 to 95% mobile phase B (mobile phase A: 0.1% FA; mobile phase B: 80% ACN in 0.1% FA). The loading solvent was 2% ACN in 0.1% FA. In any case, we ruled out any carry-over effect since between the fractions of each experiment, two 40 min blanks were injected and systematically checked.

Data acquisition was performed using a data-dependent top-20 method in full scan positive mode, scanning 375 to 1200 m/z. Survey scans were acquired at a resolution of 60,000 at m/z 200, with a normalized automatic gain control (AGC) target (%) of 300 and a maximum injection time (IT) in AUTO. The top 20 most intense ions from each MS1 scan were selected and fragmented via higher-energy collisional dissociation (HCD). The resolution for HCD spectra was set to 45,000 at m/z 200, with an AGC target of 100 and a maximum ion injection time in AUTO. Isolation of precursors was performed with a window of 0.7 m/z, exclusion duration (s) of 45 and HCD collision energy of 30. Precursor ions with single, unassigned, or six and higher charge states from fragmentation selection were excluded.

Data obtained by mass spectrometry were analyzed with Proteome Discoverer (v2.4.0.305) using four search engines (Mascot (v2.7.0), MsAmanda (v2.4.0), MsFragger (v3.1.1) and Sequest HT) and a target/decoy database built on one side from sequences in the *Homo sapiens* and *SARS-CoV-2* proteomes at UniProt Knowledgebase (20210225, 20,464 sequences, respectively) and on the other side, the high-quality reference human gut microbiome database (87), formerly called Integrated Genome Catalog, which contains approximately 10 million human microbiome protein sequences with functional and taxonomic annotations. All searches were configured with dynamic modifications for TMT reagents (+ 229.163 Da) on lysine and N-termini of the peptide, pyrrolidone from Q (− 17.027 Da) and oxidation of methionine residues (+ 15.9949 Da), and static modification as carbamidomethyl (+ 57.021 Da) on cysteine, monoisotopic masses, and trypsin cleavage (max 2 missed cleavages). The peptide precursor mass tolerance was 10 ppm, and the MS/MS tolerance was 0.02 Da. The false discovery rate (FDR) cutoff for proteins, peptides, and peptide spectral match (PSM) peptides was kept at 1%.

Quantitation was also performed in Proteome Discoverer using the "Reporter Ions Quantifier" feature in the quantification workflow using the following parameters: unique+razor peptides were used for quantitation, the coisolation threshold was set at 50%, the signal-to-noise ratio of reporter ions was 10, and normalization and scaling were performed considering the total peptide amount and the control (IS) average, respectively. Protein groups (master proteins) with an FDR lower than 1% and with

abundance values in both IS were considered for quantitation. No imputation method was used because only proteins quantified in all channels and present in both internal standards (IS) were included. The protein ratio was calculated considering the protein abundance, and the hypothesis test was based on a t test (background based). Then, an adjusted p-value  $\leq 0.05$  of the abundance ratio using Benjamini–Hochberg was set to determine the DE proteins (up- and downregulated) in paired comparisons (Additional file 1: Table S1D–I), Volcano plots and PCA (performed in Proteome Discover software).

To further understand the specific roles related to SARS-CoV-2 infection of the DE human proteins identified in our analysis, we used the following tools: Qiagen's IPA software [93] and the Metascape website [12]. Ingenuity Pathway Analysis (IPA) software from Qiagen uses a proprietary algorithm to generate causal networks from input data. The algorithm takes into account various factors, including published literature, pathway databases, and experimental evidence, to identify causal relationships between molecules in the input dataset. It estimates whether proteins in a given network are increased or decreased based on the direction of change in the expression levels of their corresponding genes. Metascape compares user-introduced gene or protein lists with all analyses from -omic technologies related to SARS-CoV-2 samples collected from different databases. For this study, we uploaded our DE protein list from the comparison of non-COVID-susceptible vs non-COVID-susceptible groups to the Metascape website and selected the best enrichment from RNA-seq and proteomics analysis.

Preprocessing of the protein intensities for the selected microbial proteomes was performed as follows: (a) the protein abundance was calculated according to the Peptide Spectrum Matches (PSMs), which has been commonly used as a relative quantitation score of the proteins in a complex mixture based on protein coverage by the peptide matches in a database search result; (b) the identified proteins were assigned to KO pathways (87–89), and the relative abundance of each KO was obtained by the sum of the relative abundances of all proteins assigned to each of the identified KO [94]; (c) the differential abundance of proteins and KO among groups were obtained using Wilcoxon test p-values, using R programming language [95]. We considered DE KOs between groups those with a p-value  $\leq 0.01$ .

#### ACE-2 activity measurement

Plasma samples were collected in EDTA tubes and cryopreserved at  $-80^{\circ}\text{C}$ . After thawing, ACE2 activity was measured in batches through a fluorometric assay using a synthetic ACE2-specific substrate (Mca-APK(Dnp);

Reactomix S.L., Granada, Spain) that is metabolized to a fluorescent compound in the presence of a functional enzyme. Samples were incubated with the buffer solution (150 nM Tris–HCl pH 7.5, 200 nM NaCl, 10  $\mu\text{M}$  ZnCl<sub>2</sub>, protease inhibitor) and the substrate (portions 10-99-1) at room temperature for 16 h. We quantified each sample's fluorescence using Varioskan lux (Thermo), a fluorescence reader with an excitation of 320 nm and an emission of 420 nm. Measurements in duplicate showed a coefficient of variation of 5.1%. The protein concentration, measured colorimetrically by the Pierce™ 660 nm Protein Assay reagent (Thermo Scientific, USA) for all samples, was on average  $1.58 \pm 0.07$   $\mu\text{g}$  protein/ $\mu\text{L}$ , with no significant differences across samples. Therefore, the differences in ACE2 activity have biological significance and are not due to a bias in the amount of protein in each sample. According to this technique, increased fluorescence indicates increased ACE2 activity.

We used R software to calculate Spearman's coefficients for the targeted correlation analyses for ACE2 plasma activity and the six proteins consistently identified in the METASCAPE analysis (library corrplot).

#### Significance of the study

Although several proteomics analyses have been performed in plasma samples from COVID-19 patients to understand the role of host factors in disease progression, information about the role of the human salivary proteome is scarce. Furthermore, although several studies have addressed the differences between different degrees of the disease, little is known about those people who are less susceptible to SARS-CoV-2 infection even having been heavily exposed to the virus. In this study, we analyzed a unique cohort of unvaccinated people to address these two specific questions. The salivary proteome of nonsusceptible to COVID-19 patients compared to that of susceptible patients or infected patients showing different degrees of the disease (moderate and severe) was obtained. Then we compared our data with previously published datasets related to COVID-19 and analyzed bacterial-produced proteins, which have been shown to have regulatory roles in different diseases, regarding The Microbiome Database. Finally, we studied correlations with the plasma activity of the molecular target of SARS-CoV-2, ACE2, and found a link with two proteins related to protein transport through microtubules in the cytoplasm (DYNC1 and MAPRE1). Our study identifies new pathways involved in SARS-CoV-2 susceptibility and pathogenesis that will be useful to design strategies to combat the COVID-19 pandemic.

#### Abbreviations

ACN	Acetonitrile
ACE2	Angiotensin-converting enzyme 2

AGC	Automatic gain control
CAA	Chloroacetamide
DE	Differentially expressed
FA	Formic acid
FDR	False discovery rate
GO	Gene Ontology
HCD	Higher-energy collisional dissociation
IPA	Ingenuity Pathway Analysis
IS	Internal standard
KEGG	Kyoto Encyclopedia of Genes and Genomes
Kos	KEGG Ontologies
LC–ESI–MS/MS	Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometric
Mcov	Moderate COVID-19
NcNs	Non-COVID and nonsusceptible individuals
NcSus	Non-COVID and susceptible individuals
PSMs	Peptide spectral matches
ROS	Reactive oxygen species
Scov	Severe COVID-19
SDB-RPS	Reversed-phase Stage-Tips
TCEP	Tris(2–24 carboxyethyl) phosphine
TEAB	Triethylammonium bicarbonate
TMT	Tandem mass tags

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12014-024-09482-9>.

**Additional file 1: Figure S1.** Main interaction network from IPA analysis for DE proteins in saliva of the comparison of Moderate COVID (Mcov) vs Non-COVID nonsusceptible (NcNs). **Figure S2.** Main interaction networks from IPA analysis for DE proteins in saliva of the comparison of Moderate COVID (Mcov) vs Non-COVID susceptible (NcSus). **Figure S3.** Main interaction network from IPA analysis for DE proteins in saliva of the comparison of severe COVID (Scov) vs non-COVID susceptible (NcSus). **Figure S4.** Main interaction networks from IPA analysis for DE proteins in saliva of the comparison of severe COVID (Scov) vs non-COVID nonsusceptible (NcNs). **Figure S5.** Main interaction networks from IPA analysis for DE proteins in saliva of the comparison of severe COVID (Scov) vs moderate COVID (Mcov). **Figure S6.** Comparisons with other databases were performed to infer the main transcription factors, transcription factor targets, cell types, and tissues. **Table S1.** Quality-filtered nonredundant human proteins identified from saliva samples, their expression level, and their functional annotation. **Table S2.** Lists of DE human proteins identified from saliva samples that were used for functional analysis based on IPA. **Table S3.** Discriminatory pathways associated with DE human proteins herein identified in saliva samples of all conditions compared to those in all the -omic studies published related to COVID-19 by using Metascape. **Table S4.** Quality-filtered nonredundant microbial-produced proteins identified from the saliva samples and their expression levels.

## Acknowledgements

We thank all patients and healthcare workers who participated in the study.

## Author contributions

Conceptualizations: SSV, MF, FC, EM. Methodology: SSV, MF, FC, EM, SC, DJ, RB, SSC, SF. Patient recruitment: SSV, JMS, PV, RR, MSC. Laboratory measurements: EM, SC, DJ, SF, FC, MF. Funding acquisition: SSV, MF, FC. Supervision: SSV, MF, FC. Writing—original draft: EM. Writing—review and editing: EM, SC, DJ, JMS, PV, RR, MSC, RB, SSC, SM, SF, FC, MF, SSV.

## Funding

This work was supported by Instituto de Salud Carlos III (AC17/00019, PI18/00154, COV20/00349, ICI20/00058, PI21/00141), CRUE-Supera COVID, co-financed by the European Development Regional Fund “A way to achieve Europe” (ERDF), Merck, Sharp & Dohme Investigator Studies Program (code MISP# IIS 60257), Fondo Supera COVID-19 (2020-001), and a mobility grant from the Spanish Society of Infectious Diseases and Clinical Microbiology

(SEIMC). This work was also supported by Instituto de Salud Carlos III and Fundación Agencia Española contra el Cáncer (AC17/00022, within the ERA NET TRANSCAN-2) and by Consejo Superior de Investigaciones Científicas (CSIC) projects PIE 202020E108 and SEV 2017-0712. We also acknowledge financial support under Grants PID2020-112758RB-I00 and PDC2021-121534-I00 from the Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación (AEI) (Digital Object Identifier <https://doi.org/10.13039/501100011033>) and the European Union (“NextGenerationEU”).

## Availability of data and materials

The datasets supporting the conclusions of this article are included within the article (and its additional files). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [96] with the dataset identifiers PXD032333 and <https://doi.org/10.6019/PXD032333>.

## Declarations

### Ethics approval and consent to participate

The study was carried out at the Ramón and Cajal University Hospital in Madrid (Spain) and was approved by the local research ethics committee (Comite ético de Investigación clínica GAE Ramon y Cajal, ceic.hrc@salud.madrid.org, approval number 095/20). Informed consent was obtained from all subjects and/or their legal guardian(s). All methods were performed in accordance with the relevant guidelines and regulations.

### Consent for publication

Not applicable.

### Competing interests

Outside the submitted work, S. S.-V. Reports personal fees from Viiv Healthcare, Janssen Cilag, Gilead Sciences, and MSD as well as nonfinancial support from Viiv Healthcare and Gilead Sciences and research grants from MSD and Gilead Sciences. J.M.-S. nonfinancial support from Viiv Healthcare, nonfinancial support from Janssen Cilag, nonfinancial support from Gilead Sciences, outside the submitted work. E.M. has received mobility grants for conferences from Gilead and Viiv Healthcare; S.M. reports grants, personal fees and nonfinancial support from Viiv Healthcare, personal fees and nonfinancial support from Janssen, grants, personal fees and nonfinancial support from MSD, grants, personal fees and nonfinancial support from Gilead, outside the submitted work. There are no potential conflicts of interest.

### Author details

<sup>1</sup>Department of Infectious Diseases, Facultad de Medicina, Hospital Universitario Ramón y Cajal, IRYCIS, Carretera de Colmenar Viejo, Km 9.100, 28034 Madrid, Spain. <sup>2</sup>CIBERINFEC, Instituto de Salud Carlos III, 28029 Madrid, Spain. <sup>3</sup>Functional Proteomics Laboratory, Centro Nacional de Biotecnología (CNB), CSIC, 28049 Madrid, Spain. <sup>4</sup>Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor LL57 2UW, UK. <sup>5</sup>Instituto de Catalis y Petroquímica (ICP), CSIC, 28049 Madrid, Spain. <sup>6</sup>Facultad de Medicina, Universidad de Alcalá de Henares, 28801 Alcalá de Henares, Madrid, Spain. <sup>7</sup>Present Address: Centro de Biología Molecular Severo Ochoa (CBM), CSIC-UAM, 28049 Madrid, Spain.

Received: 13 October 2023 Accepted: 15 April 2024

Published online: 22 May 2024

## References

- Castagnola M, Scarano E, Passali GCC, Messana I, et al. Salivary biomarkers and proteomics: future diagnostic and clinical utilities. *Acta Otorinolaryngol Ital*. 2017;37:94–101.
- Lau WW, Hardt M, Zhang YH, Freire M, Ruhl S. The human salivary proteome wiki: a community-driven research platform. *J Dent Res*. 2021;100:1510–9.
- Keijser BJF, Zaura E, Huse SM, van der Vossen JMBM, Schuren FHJ, Montijn RC, ten Cate JM, Crielaard W, et al. Pyrosequencing analysis of the Oral Microflora of healthy adults. *J Dental Res*. 2008;87:1016–20.

4. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary microbiome. *Genome Res.* 2009;19:636–43.
5. Bik EM, Long CD, Armitage GC, Loomer P, et al. Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J.* 2010;4:962–74.
6. Zaura E, Keijsers B, Huse SM, Crielaard W. Defining the healthy “core microbiome” of oral microbial communities. *BMC Microbiol.* 2009;9:259.
7. Moussa DG, Ahmad P, Mansour TA, Siqueira WL. Current state and challenges of the global outcomes of dental caries research in the metagenomics era. *Front Cell Infect Microbiol.* 2022;12:887907.
8. Rudney JDD, Xie H, Rhodus NLL, Ondrey FGG, Griffin TJJ. A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. *Mol Oral Microbiol.* 2010;25:38–49.
9. Attaway AH, Scheraga RG, Bhimraj A, Biehl M, Hatipoğlu U. Severe COVID-19 pneumonia: pathogenesis and clinical management. *BMJ.* 2021. <https://doi.org/10.1136/bmj.n436>.
10. Shakaib B, Zohra T, Ikram A, Shakaib MB, et al. A comprehensive review on clinical and mechanistic pathophysiological aspects of COVID-19 Malady: how far have we come? *Virology.* 2021;18:120.
11. Mun D-G, Vanderboom PM, Madugundu AK, Garapati K, et al. DIA-based proteome profiling of nasopharyngeal swabs from COVID-19 patients. *J Proteome Res.* 2021;20:4165–75.
12. Zhou Y, Zhou B, Pache L, Chang M, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019. <https://doi.org/10.1038/s41467-019-09234-6>.
13. Hotter D, Bosso M, Jönsson KL, Krapp C, et al. IFI16 targets the transcription factor Sp1 to suppress HIV-1 transcription and latency reactivation. *Cell Host Microbe.* 2019;25:858–872.e13.
14. Ren F, Ren J-H, Song C-L, Tan M, et al. LncRNA HOTAIR modulates hepatitis B virus transcription and replication by enhancing SP1 transcription factor. *Clin Sci.* 2020;134:3007–22.
15. Khalil MI, Hay J, Ruyechan WT. Cellular transcription factors Sp1 and Sp3 suppress varicella-zoster virus origin-dependent DNA replication. *J Virol.* 2008;82:11723–33.
16. Xu J, Mao J, Han X, Shi F, et al. Porcine epidemic diarrhea virus inhibits HDAC1 expression to facilitate its replication via binding of its nucleocapsid protein to host transcription factor Sp1. *J Virol.* 2021;95: e0085321.
17. Nam C-H, Rabbitts TH. The role of LMO2 in development and in T cell leukemia after chromosomal translocation or retroviral insertion. *Mol Ther.* 2006;13:15–25.
18. Pike-Overzet K, de Ridder D, Weerkamp F, Baert MRM, et al. Ectopic retroviral expression of LMO2, but not IL2R $\gamma$ , blocks human T-cell development from CD34+ cells: implications for leukemogenesis in gene therapy. *Leukemia.* 2007;21:754–63.
19. Hesari A, Ghasemi F, Salarinia R, Biglari H, Tabar Molla Hassan A, Abdoli V, Mirzaei H. *Effects of curcumin on NF- $\kappa$ B, AP-1.* 2018, 119: 7898
20. Cheng C-Y, Tseng H-H, Chiu H-C, Chang C-D, et al. Bovine ephemeral fever virus triggers autophagy enhancing virus replication via upregulation of the Src/JNK/AP1 and PI3K/Akt/NF- $\kappa$ B pathways and suppression of the PI3K/Akt/mTOR pathway. *Vet Res.* 2019;50:79.
21. Miller EH, Annavajhala MK, Chong AM, Park H, et al. Oral microbiome alterations and SARS-CoV-2 saliva viral load in patients with COVID-19. *Microbiol Spectr.* 2021;9:e00055-e121.
22. Yeoh YK, Zuo T, Lui GC-Y, Zhang F, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut.* 2021;70:698–706.
23. Armstrong AJS, Parmar V, Blaser MJ. Assessing saliva microbiome collection and processing methods. *npj Biofilms Microb.* 2021;7:81.
24. Zuo T, Zhang F, Lui GCY, Yeoh YK, et al. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. *Gastroenterology.* 2020;159:944–955.e8.
25. He F, Zhang T, Xue K, Fang Z, et al. Fecal multi-omics analysis reveals diverse molecular alterations of gut ecosystem in COVID-19 patients. *Anal Chim Acta.* 2021;1180:338881.
26. Zhong W, Altay O, Arif M, Edfors F, et al. Next generation plasma proteome profiling of COVID-19 patients with mild to moderate symptoms. *EBioMedicine.* 2021;74:103723.
27. Ayass MA, Cao W, Zhang J, Dai J, et al. Noninvasive nasopharyngeal proteomics of COVID-19 patient identify abnormalities related to complement and coagulation cascade and mucosal immune system. *PLoS ONE.* 2022;17: e0274228.
28. Banerjee A, Pai MGJ, Singh A, Nissa MU, Srivastava S. Mass spectrometry and proteome analysis to identify SARS-CoV-2 protein from COVID-19 patient swab samples. *STAR Protoc.* 2022;3:101177.
29. Gouveia D, Miotello G, Gallais F, Gaillard J-C, et al. Proteotyping SARS-CoV-2 virus from nasopharyngeal swabs: a proof-of-concept focused on a 3 min mass spectrometry window. *J Proteome Res.* 2020;19:4407–16.
30. Ghosh S, Parikh S, Nissa MU, Acharjee A, et al. Semen proteomics of COVID-19 convalescent men reveals disruption of key biological pathways relevant to male reproductive function. *ACS Omega.* 2022;7:8601–12.
31. Fanelli V, Vlachou A, Ghannadian S, Simonetti U, et al. Acute respiratory distress syndrome: new definition, current and future therapeutic options. *J Thorac Dis.* 2013;5:326–34.
32. Pagani L, Chinello C, Mahajneh A, Clerici F, et al. Untargeted mass spectrometry approach to study SARS-CoV-2 proteins in human plasma and saliva proteome. *BioChem.* 2022;2:64–83.
33. Pappa E, Vastardis H, Mermelekas G, Gerasimidi-Vazeou A, et al. Saliva Proteomics analysis offers insights on type 1 diabetes pathology in a pediatric population. *Front Physiol.* 2018;9:444.
34. Rao PV, Reddy AP, Lu X, Dasari S, et al. Proteomic identification of salivary biomarkers of type-2 diabetes. *J Proteome Res.* 2009;8:239–45.
35. Mata AD, Marques D, Rocha S, Francisco H, et al. Effects of diabetes mellitus on salivary secretion and its composition in the human. *Mol Cell Biochem.* 2004;261:137–42.
36. Chambon C, Neyraud E, Sayd T, Bros P, et al. The salivary proteome reflects some traits of dietary habits in diabetic and non-diabetic older adults. *Eur J Nutr.* 2021;60:4331–44.
37. Contini C, Serrao S, Manconi B, Olianis A, et al. Salivary proteomics reveals significant changes in relation to Alzheimer’s disease and aging. *JAD.* 2022;89:605–22.
38. Jia SY, Zhang YL, Sun XY, Yuan C, Zheng SG. Impact of the glycemic level on the salivary proteome of middle-aged and elderly people with type 2 diabetes mellitus: an observational study. *Front Mol Biosci.* 2021;8:790091.
39. François M, Bull CF, Fenech MF, Leifert WR. Current state of saliva biomarkers for aging and Alzheimer’s disease. *CAR.* 2018;16:56–66.
40. Zelcer N, Tontonoz P. Liver X receptors as integrators of metabolic and inflammatory signaling. *J Clin Invest.* 2006;116:607–14.
41. Zhu R, Ou Z, Ruan X, Gong J. Role of liver X receptors in cholesterol efflux and inflammatory signaling (review). *Mol Med Rep.* 2012;5:895–900.
42. Tang X, Sun R, Ge W, Mao T, et al. Enhanced inflammation and suppressed adaptive immunity in COVID-19 with prolonged RNA shedding. *Cell Discov.* 2022;8:70.
43. Karki R, Sharma BR, Tuladhar S, Williams EP, et al. Synergism of TNF- $\alpha$  and IFN- $\gamma$  triggers inflammatory cell death, tissue damage, and mortality in SARS-CoV-2 infection and cytokine shock syndromes. *Cell.* 2021;184:149–168.e17.
44. Saheb Sharif-Askari N, Saheb Sharif-Askari F, Mdkhana B, Hussain Alsayed HA, et al. Upregulation of oxidative stress gene markers during SARS-CoV-2 viral infection. *Free Radical Biol Med.* 2021;172:688–98.
45. Silvina A, Chapuis N, Dunsmore G, Goubet A-G, et al. Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell.* 2020;182:1401–1418.e18.
46. Udeh R, Advani S, de Guadiana Romualdo LG, Dolja-Gore X. Calprotectin, an emerging biomarker of interest in COVID-19: a systematic review and meta-analysis. *J Clin Med.* 2021;10:775.
47. Salmen S, Colmenares M, Peterson DL, Reyes E, et al. HIV-1 Nef associates with p22-phox, a component of the NADPH oxidase protein complex. *Cell Immunol.* 2010;263:166–71.
48. Wang S, Song R, Wang Z, Jing Z, et al. S100A8/A9 in inflammation. *Front Immunol.* 2018;9:1298.
49. Sauve AA. NAD $^{+}$  and vitamin B $_{3}$ : from metabolism to therapies. *J Pharmacol Exp Ther.* 2008;324:883–93.
50. Schultz IC, Bertoni APS, Wink MR. Purinergic signaling elements are correlated with coagulation players in peripheral blood and leukocyte samples from COVID-19 patients. *J Mol Med.* 2022;100:569–84.
51. Alaiya A, Alshukairi A, Shinwari Z, Al-Fares M, et al. Alterations in the plasma proteome induced by SARS-CoV-2 and MERS-CoV reveal biomarkers for disease outcomes for COVID-19 patients. *J Inflamm Res.* 2021;14:4313–28.
52. Xu J, Zhang X, Pelayo R, Monestier M, et al. Extracellular histones are major mediators of death in sepsis. *Nat Med.* 2009;15:1318–21.



53. Bedalov A, Hirao M, Posakony J, Nelson M, Simon JA. NAD<sup>+</sup>-dependent deacetylase Hst1p controls biosynthesis and cellular NAD<sup>+</sup> levels in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2003;23:7044–54.
54. Albóniga OE, Jiménez D, Sánchez-Conde M, Vizcarra P, et al. Metabolic snapshot of plasma samples reveals new pathways implicated in SARS-CoV-2 pathogenesis. *J Proteom Res*. 2022;21:623.
55. Kongsbak M, Levring TB, Geisler C, von Essen MR. The vitamin D receptor and T cell function. *Front Immunol*. 2013. <https://doi.org/10.3389/fimmu.2013.00148>.
56. Ohmori Y, Hamilton TA. IL-4-induced STAT6 suppresses IFN- $\gamma$ -stimulated STAT1-dependent transcription in mouse macrophages. *J Immunol*. 1997;159:5474–82.
57. Hendrayani S-F, Al-Harbi B, Al-Ansari MM, Silva G, Aboussekhra A. The inflammatory/cancer-related IL-6/STAT3/NF- $\kappa$ B positive feedback loop includes AUF1 and maintains the active state of breast myofibroblasts. *Oncotarget*. 2016;7:41974–85.
58. Salem F, Li XZ, Hindi J, Casablanca NM, et al. Activation of STAT3 signaling pathway in the kidney of COVID-19 patients. *J Nephrol*. 2022;35:735–43.
59. Shamir I, Abutbul-Amaiti M, Abbas-Egbariya H, Pismanik-Chor M, et al. STAT3 isoforms differentially affect ACE2 expression: a potential target for COVID-19 therapy. *J Cell Mol Med*. 2020;24:12864–8.
60. Musiu C, Caligola S, Fiore A, Lamolinara A, et al. Fatal cytokine release syndrome by an aberrant FLIP/STAT3 axis. *Cell Death Differ*. 2022;29:420–38.
61. Dhall A, Patiyal S, Sharma N, Devi NL, Raghava GPS. Computer-aided prediction of inhibitors against STAT3 for managing COVID-19 associated cytokine storm. *Comput Biol Med*. 2021;137:104780.
62. Jafarzadeh A, Nemati M, Jafarzadeh S. Contribution of STAT3 to the pathogenesis of COVID-19. *Microb Pathog*. 2021;154:104836.
63. Gajjala BK, Zhou M-M. Calming the cytokine storm of COVID-19 through inhibition of JAK2/STAT3 signaling. *Drug Discov Today*. 2022;27:390–400.
64. Yu H, Lu J, Zuo L, Yan Q, et al. Epstein-Barr virus downregulates microRNA 203 through the oncoprotein latent membrane protein 1: a contribution to increased tumor incidence in epithelial cells. *J Virol*. 2012;86:3088–99.
65. Hemmaid KZ, Awadalla A, Elsayy E, Hussein A-AM, et al. Impact of Hepatitis C Virus (HCV) infection on biomolecular markers influencing the pathogenesis of bladder cancer. *Infect Agents Cancer*. 2013;8:24.
66. Santerre M, Bagashev A, Gorecki L, Lysek KZ, et al. HIV-1 Tat protein promotes neuronal dysregulation by inhibiting E2F transcription factor 3 (E2F3). *J Biol Chem*. 2019;294:3618–33.
67. Geng D, Song X, Ning F, Song Q, Yin H. MiR-34a inhibits viability and invasion of human papillomavirus-positive cervical cancer cells by targeting E2F3 and regulating survivin. *Int J Gynecol Cancer*. 2015;25:707–13.
68. Park MC, Kim H, Choi H, Chang MS, Lee SK. Epstein-barr virus miR-BART1-3p regulates the miR-17–92 cluster by targeting E2F3. *Int J Mol Sci*. 2021;22:10936.
69. Pathirana WGW, Chubb SP, Gillett MJ, Vasikaran SD. Faecal Calprotectin. *Clin Biochem Rev*. 2018;39:77–90.
70. Shabani F, Mahdavi M, Imani M, Hosseinpour-Feizi MAA, Gheibi N. Calprotectin (S100A8/S100A9)-induced cytotoxicity and apoptosis in human gastric cancer AGS cells: alteration in expression levels of Bax, Bcl-2, and ERK2. *Hum Exp Toxicol*. 2020;39:1031–45.
71. Donato R, Sorci G, Giambanco I. S100A6 protein: functional roles. *Cell Mol Life Sci*. 2017;74:2749–60.
72. Kaouass M, Beaulieu R, Balicki D. Histonefection: novel and potent non-viral gene delivery. *J Control Release*. 2006;113:245–54.
73. Balbín M, Hall A, Grubb A, Mason RW, et al. Structural and functional characterization of two allelic variants of human cystatin D sharing a characteristic inhibition spectrum against mammalian cysteine proteinases. *J Biol Chem*. 1994;269:23156–62.
74. Zhou X, Wang X, Huang K, Liao X, et al. Investigation of the clinical significance and prospective molecular mechanisms of cystatin genes in patients with hepatitis B virus-related hepatocellular carcinoma. *Oncol Rep*. 2019;42:189–201.
75. Zinellu A, Mangoni AA. Cystatin C, COVID-19 severity and mortality: a systematic review and meta-analysis. *J Nephrol*. 2022;35:59–68.
76. Neeland MR, Bannister S, Clifford V, Dohle K, et al. Innate cell profiles during the acute and convalescent phase of SARS-CoV-2 infection in children. *Nat Commun*. 2021;12:1084.
77. Conzelmann C, Groß R, Zou M, Krüger F, et al. Salivary extracellular vesicles inhibit Zika virus but not SARS-CoV-2 infection. *J Extracell Vesicles*. 2020;9:1808281.
78. Willemsen MH, Vissers LEL, Willemsen MAA, van Bon BWM, et al. Mutations in DYNC1H1 cause severe intellectual disability with neuronal migration defects. *J Med Genet*. 2012;49:179–83.
79. Baig AM, Khaleeq A, Ali U, Syeda H. Evidence of the COVID-19 virus targeting the CNS: tissue distribution, host-virus interaction, and proposed neurotropic mechanisms. *ACS Chem Neurosci*. 2020;11:995–8.
80. Sun T, Guan J. Novel coronavirus and the central nervous system. *Eur J Neurol*. 2020;27:e52.
81. Wu Y, Xu X, Chen Z, Duan J, et al. Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain Behav Immun*. 2020;87:18–22.
82. Merino-Gracia J, García-Mayoral MF, Rodríguez-Crespo I. The association of viral proteins with host cell dynein components during virus infection. *FEBS J*. 2011;278:2997–3011.
83. Naghavi MH. HIV-1 capsid exploitation of the host microtubule cytoskeleton during early infection. *Retrovirology*. 2021;18:19.
84. Patel SK, Juno JA, Lee WS, Wragg KM, et al. Plasma ACE2 activity is persistently elevated following SARS-CoV-2 infection: implications for COVID-19 pathogenesis and consequences. *Eur Respir J*. 2021;57:2003730.
85. Doré E, Joly-Beauparlant C, Morozumi S, Mathieu A, et al. The interaction of secreted phospholipase A2-IIA with the microbiota alters its lipidome and promotes inflammation. *JCI Insight*. 2022. <https://doi.org/10.1172/jci.insight.152638>.
86. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol*. 2016;16:341–52.
87. Serrano-Villar S, Moreno S, Ferrer M. The functional consequences of the microbiome in HIV: insights from metabolomic studies. *Curr Opin HIV AIDS*. 2018;13:88–94.
88. Centers for Disease Control (CDC) P for Disease Control C., (CDC), P. Interim Operational Considerations for Public Health Management of Healthcare Workers Exposed to or with Suspected or Confirmed COVID-19: non-US Healthcare Settings. *Centers for disease control and prevention* 2020, 2019, 1–11
89. Ciordia S, Alvarez-Sola G, Rullán M, Urman JM, et al. Digging deeper into bile proteome. *J Proteomics*. 2021;230:103984.
90. Yang R, Meyer AS, Droujinine IA, Udeshi ND, et al. A genetic model for *in vivo* proximity labelling of the mammalian secretome. *Open Biol*. 2022;12:220149.
91. Myers SA, Rhoads A, Cocco AR, Peckner R, et al. Streamlined protocol for deep proteomic profiling of FAC-sorted cells and its application to freshly isolated murine immune cells. *Mol Cell Proteom*. 2019;18:995–1009.
92. Ciordia S, Alvarez-Sola G, Rullán M, Urman JM, et al. In: Corrales FJ, Paradelo A, Marcilla M, editors, et al., *Clinical Proteomics*, vol. 2420. US, New York, NY: Springer; 2022. p. 1–10.
93. Qiagen, Qiagen IPA analysis features. *Website* 2022.
94. Ruiz-Ruiz S, Sanchez-Carrillo S, Ciordia S, Mena MC, et al. Functional microbiome deficits associated with ageing: Chronological age threshold. *Aging Cell*. 2020;19:e13063.
95. Team, R C, R Core Team. R: A language and environment for statistical computing, Website. R Found Stat Comput: R Found Stat Comput; 2022.
96. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res*. 2022;50:D543–52.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.