medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Clinical and molecular characterization of COVID-19 hospitalized patients

Elisa Benetti[3], Annarita Giliberti[1], Arianna Emiliozzi[3,5], Floriana Valentino[1], Laura Bergantini[6], Chiara Fallerini[1], Federico Anedda[7], Sara Amitrano[2], Edoardo Conticini[9], Rossella Tita[2], Miriana d'Alessandro[6], Francesca Fava[1,2], Simona Marcantonio[7], Margherita Baldassarri[1], Mirella Bruttini[1,2], Maria Antonietta Mazzei[8], Francesca Montagnani[3,5], Marco Mandalà[4], Elena Bargagli[6], Simone Furini[3], GEN-COVID Multicenter Study, Alessandra Renieri[1,2*] and Francesca Mari[1,2]

1) Medical Genetics, University of Siena, Italy
2) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Italy
3) Department of Medical Biotechnologies, University of Siena, Italy
4) Otolaryngology Unit, University of Siena, Italy
5) Department of Specialized and Internal Medicine, Tropical and Infectious Diseases Unit, Azienda Ospedaliera Universitaria Senese, Italy
6) Unit of Respiratory Diseases and Lung Transplantation, Department of Internal and Specialist Medicine, University of Siena
7) Department of Emergency and Urgency, Medicine, Surgery and Neurosciences, Unit of Intensive Care Medicine, Siena University Hospital, Italy
8) Department of Medical, Surgical and Neuro Sciences and Radiological Sciences, Unit of Diagnostic Imaging, University of Siena, Azienda Ospedaliera Universitaria Senese, Italy
9) Rheumatology Unit, Department of Medicine, Surgery and Neurosciences, University of Siena, Policlinico Le Scotte, Italy

*** Corresponding author:**

Professor Alessandra Renieri
Medical Genetics Unit
University of Siena
Policlinico "Santa Maria alle Scotte"
Viale Bracci, 2 -53100 Siena, Italy
Phone: 39 0577 233303 - FAX 39 0577 233325
Email: alessandra.renieri@unisi.it

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ABSTRACT

Clinical and molecular characterization by Whole Exome Sequencing (WES) is reported in 35 COVID-19 patients attending the University Hospital in Siena, Italy, from April 7 to May 7, 2020. Eighty percent of patients required respiratory assistance, half of them being on mechanical ventilation. Fiftyone percent had hepatic involvement and hyposmia was ascertained in 3 patients. Searching for common genes by collapsing methods against 150 WES of controls of the Italian population failed to give straightforward statistically significant results with the exception of two genes. This result is not unexpected since we are facing the most challenging common disorder triggered by environmental factors with a strong underlying heritability (50%). The lesson learned from Autism-Spectrum-Disorders prompted us to re-analyse the cohort treating each patient as an independent case, following a Mendelian-like model. We identified for each patient an average of 2.5 pathogenic mutations involved in virus infection susceptibility and pinpointing to one or more rare disorder(s). To our knowledge, this is the first report on WES and COVID-19. Our results suggest a combined model for COVID-19 susceptibility with a number of common susceptibility genes which represent the favorite background in which additional host private mutations may determine disease progression.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

MAIN TEXT

INTRODUCTION

Italy has been the first European Country experiencing the epidemic wave of SARS-CoV-2 infection, with an apparently more severe clinical picture, compared to other countries. Indeed, the case fatality rate has peaked to 14% in Italy, while it remains stable around 5% in China. At the time of the study , 12 May 2020, SARS-CoV-2 positive subjects in Italy have reached the threshold of 200.000 cases [1]. Since the beginning of the epidemic wave, one of the first observations has been a highly heterogeneous phenotypic response to SARS-CoV-2 infection among individuals. Indeed, while most affected subjects show mild symptoms, a subset of patients develops severe pneumonia requiring mechanical ventilation with a 20% of cases requiring hospitalization; 5% of cases admitted to the Intensive Care Unit (ICU), and 6,1% requiring intensive support with ventilators or extracorporeal oxygenation (ECMO) machines [2]. Although patients undergoing ventilatory assistance are often older and are affected by other diseases, like diabetes [3], the existing comorbidities alone do not fully explain the differences in clinical severity. As demonstrated for other viral diseases, the basis of these different outcomes there are host predisposing genetic factors leading to different immunogenicity/cytokine responses as well as specific receptor permissiveness to virus and antiviral defence [4–6]. Similarly, during the study of host genetics in influenza disease, a pattern of genetic markers has been identified which underlies increased susceptibility to a more severe clinical outcome (as reviewed in [7]). This hypothesis is also supported by a recent work reporting 50% heritability of COVID-19 symptoms [8].

The identification of host genetic variants associated with disease severity is of utmost importance to develop both effective treatments, based on a personalized approach,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and novel diagnostics. Also, it is expected to be of high relevance in providing guidance for the health care systems and societal organizations. However, nowadays, little is known about the impact of host genome variability on COVID-19 susceptibility and severity.

On March 16th, 2020 the University Hospital in Siena launched a study named GEN-COVID with the aim to collect the genomic DNA of 2,000 COVID-19 patients for host genetic analysis. More than 30 different hospitals and community centers throughout Italy joined the study and are providing samples and clinical detailed information of COVID-19 patients. This study is aimed to identify common and rare genetic variants of SARS-CoV-2 infected individuals, using a whole exome sequencing (WES) analysis approach, in order to establish an association between host genetic variants and COVID-19 severity and prognosis.

RESULTS

Clinical data

The cohort consists of 35 COVID-19 patients (33 unrelated and 2 sisters) of Caucasian ethnicity, except for one North African and one Hispanic. The mean and median age is 64 years (range 31-98): 11 females (median age 66 years) and 24 males (median age 62 years)**.**

The population is clustered into four qualitative severity groups depending on the respiratory impairment and the need for ventilation (groups 1-4 in **Table 1** and different colors in **Fig. 1**) (see Methods section). In the two most severe groups (groups 1 and 2, including 13 patients) there are 11 males and 2 females, while in the two mildest groups (groups 3 and 4 including 22 patients) males are 13 while females are 9.

Patients were also assigned a lung imaging grading according to X-Rays and CT

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

scans. The mean value is 13 for high care intensity group, 12 for intermediate care intensity group, 8 for low care intensity group and 5 for very low care intensity group.

Regarding immunological findings, a decrease in the total number of peripheral $CD4^+$ T cells were identified in 13 subjects, while NK cells' count was impaired in 10 patients. Six patients showed a reduction of both parameters. IL-6 serum level was elevated in 13 patients.

Hyposmia was present in 3 out of 34 evaluated cases (8.8%), and hypogeusia was present in the same subjects plus another case. These four cases belong to the first three severity groups. Liver involvement was present in 7 cases (20%), while pancreas involvement in 4 cases (11%); 10 patients presented both (29%). Heart involvement was detected in 13 cases (37%). 9 patients (25%) showed kidney involvement. Fibrinogen values below 200 mg/dL were identified in 2 cases (6%), between 200 and 400 mg/dL in 7 cases (20%), and above 400mg/dL in 22 cases (63%). D-dimer value below 500 ng/mL was present in 1 case (3%), between 500 and 5000 ng/mL in 26 cases (74%), and in 7 cases (20%) was 10 times higher than the normal value (>5000 ng/mL) (**Table 1**).

Table 1. Clinical characteristics COVID19 patients admitted to the University Hospital of Siena (Italy)

Subject characteristics	**Group 1**	**Group 2**	**Group 3**	**Group 4**
No. of subjects (%)	6(17.1%)	7(20%)	15(42.9%)	7(20%)
Mean age (SD)	63 (6.2)	61.6 (12.3)	70 (14)	54 (15.7)
Gender				
Male [n (%)]	5(14%)	6(17%)	7 (20%)	6 (17.1%)
Female[n (%)]	1(2.8%)	1(2.8%)	8 (22.8%)	1(2.8%)
PaO_2/FiO_2 [median (IQR)]	94.5 (37.7)	156 (74)	279.5 (162)	304 (73.5)
Lung imaging grading (CXR score) [median (IQR)]	13 (3.7)	13 (3)	8 (4)	5(6)
Laboratory findings				

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

CD4$^+$ T cells count					
	[median (IQR)]	300 (330.7)	582 (661)	458 (906)	623 (360)
NK cells count					
	[median (IQR)]	79.5 (72.2)	73 (110)	112 (90)	204 (174)
IL-6 value					
	[median (IQR)]	*598 (777.7)*	567 (648.2)	14.9 (28.4)	19 (5.3)
Fibrinogen					
	[median (IQR)]	*406 (409.7)*	518 (296)	566 (209)	546 (239)
CRP					
	[median (IQR)]	*1.22 (24.54)*	0.43 (4.6)	0.36 (1.52)	3.14 (4.97)
LDH					
	[median (IQR)]	377 (217)	407 (319)	272 (121)	255 (81)
D-Dimer					
	[median (IQR)]	5069.5 (20183)	1526 (54221)	1167 (2022)	884.5 (786.3)
Hyposmia (VAS score) [n (%)]					
	<2 (normal)	4(11.3%)	6 (17.1%)	14 (40%)	7(20%)
	2-5 (intermediate)	1(2.8%)	0	0	0
	>5 (severe)	0	1(2.8%)	1(2.8%)	0
Hypogeusia (VAS score) [n (%)]					
	No	4(11.3%)	6(17.1%)	13(37.1%)	7(20%)
	Yes	1(2.8%)	1(2.8%)	2(5.7%)	0
Heart involvement [n (%)]					
	Yes	4(11.3%)	3(8.6%)	6(17.1%)	0
	T=T-Troponin >15 (ng/L); B= pro-BNP M > 88 (pg/ml); F > 153 (pg/ml); A=arrhythmia)	T/B 2(5.7%) B 2(5.7%)	T/B 1(2.8%) T 1(2.8%) A 1(2.8%)	T/B 2(5.7%) T/A 1(2.8%) B/A 1(2.8%) A 1(2.8%) B 1(2.8%)	
	No	2(5.7%)	4(11.3%)	9(25.7%)	7(20%)
	Unknown	0	0	0	0
Hepatic (H)/Pancreatic involvement (P)					
	[n (%)] *H and P*	2(5.7%)	5(14.3%)	6(17.1%)	1(2.8%)
	H only	3(8.6%)	0	1(2.8%)	1(2.8%)
	P only	0	0	1(2.8%)	1(2.8%)

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

None	1(2.8%)	2(5.7%)	7(20%)	4(11.3%)
Kidney involvement [n (%)]				
Yes	0	3(8.6%)	5(14.3%)	1(2.8%)
No	6(17.1%)	4(11.3%)	10 (28.6%)	6(17.1%)
Co-morbidities [n (%)]				
Cardiovascular disease	1(2.8%)	2(5.7%)	3(8.6%)	
Hypertension	2(5.7%)	2(5.7%)	8(22.8%)	
Tumor	2(5.7%)	1(2.8%)	2(5.7%)	1(2.8%)
Diabetes			4 (11.3%)	
Pulmonary disease			1(2.8%)	1(2.8%)

Unbiased collapsing gene analysis

At first, we tested the hypothesis that susceptibility could be due to one or more common factor(s) in the cohort of patients compared to controls. According to this idea, damaging variants of that/those gene(s) should be either over- or under- represented in patients vs controls. We used, as controls, individuals of the Italian population assuming that the majority of them, if infected, would have shown no severe symptoms. WES data of 35 patients were compared with those of 150 controls (the Siena cohort of the Network of Italian Genomes NIG: http://www.nig.cineca.it) using a gene burden test which compares the rate of disrupting mutations per gene. The variants were collapsed on a gene-by-gene basis, in order to identify genes with mutational burden statistically different between COVID-19 samples and controls. The analysis identified genes harboring deleterious mutations (according to the DANN score) with a statistically significant higher frequency in controls than in COVID-19 patients such as the olfactory receptor gene *OR4C5* (adjusted p-value of 1.5E-10), (**Fig. 2** and **S1 Table**) and *NDUFAF7,* although to a lesser extent (**Fig. 2** and **S1 Table**). For all these genes, the susceptibility factor is represented by the functioning (or more functioning) gene. We also identified two additional genes, *PRKRA* and *LAPTM4B*, for which the probability of observing a deleterious variant was computed higher in the COVID-19 samples compared to controls **(Fig. 2** and **S2 Table).** In these latter

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

cases, the functioning gene represents indeed a protective factor.

Gene analysis using the Mendelian-like model

We then tested the hypothesis that COVID-19 susceptibility is due to different variants in different individuals. A recently acquired knowledge on the genetic bases of Autism Spectrum Disorders suggests that a common disorder could be the sum of many different rare disorders and this genetic landscape can appear indistinguishable at the clinical level [9]. Therefore, we analyzed our cohort treating each patient as an independent case, following a Mendelian-like model. According to the "pathogenic" definition in ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/), for each patient, we identified an average of 1 mutated gene involved in viral infection susceptibility and pinpointing to one or more rare disorder(s) or a carrier status of rare disorders **(Fig. 1)**. Following the pipeline used in routine clinical practice for WES analysis in rare disorders we then moved forward checking for rare variants "predicted" to be relevant for infection by the means of common annotation tools. We thus identified an average of additional 1-5 variants per patient which summed up to the previous identified pathogenic variants **(Fig. 1**, **S3 Table)**.

Known common susceptibility/protective variants analysis

We then checked the cohort for known non rare variants classified as either "pathogenic" or "protective" in ClinVar database and related to viral infection. Variants in six different genes matched the term of "viral infection" and "pathogenic" according to ClinVar **(Fig. 1)**. Overall, a mean of 3 genes with "pathogenic" common variants involved in viral infection susceptibility were present **(Fig. 1)**.

Among the common protective variants, we list as example three variants which confer protection to Human Immunodeficiency Virus (HIV), the first two, and leprosy, the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

third one: a *CCR2* variant (rs1799864) identified in 8 patients, a *CCR5* (rs1800940) in one patient and a *TLR1* variant (rs5743618) in 26 patients (not shown). A *IL4R* variant (rs1805015) associated with HIV slow progression was present in 8 patients (not shown).

Candidate gene overview

Although not identified by unbiased collapsing gene analysis a number of obvious candidate genes were specifically analyzed. First, we noticed that SARS-CoV-2 receptor, ACE2 protein is preserved in the cohort, only a silent mutation V749V being present in 2 males and 2 heterozygous females. This is in line with our previous suggestion that either rare variants or polymorphisms may impact infectivity [10]. The *IFITM3* polymorphism (rs12252) was found in heterozygosity in 4 patients as expected by frequency. Eight patients had heterozygous missense mutations in *CFTR* gene reported as VUS/mild variants, 7 / 8 being among the more severely affected patients.

DISCUSSION

In this study, we present a cohort of 35 COVID-19 patients admitted between April and May 2020 to the University Hospital of Siena who were clinically characterized by a team of 29 MDs belonging to 7 different specialties. As expected, the majority of hospitalized patients are males, confirming previously published data reporting a predominance of males among the most severe COVID-19 affected patients [11]. Lung imaging involvement, evaluated through a modified lung imaging grading system, did not completely correlate with respiratory impairment since among the 13 patients who required mechanical ventilation (group 1 and 2), grading was either moderate (10) or mild (3). In line with our previous data, lymphocyte subset immunophenotyping revealed a decrease in the total number of CD4 and NK cells count, especially in the most severe patients [12].

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Laboratory tests revealed a multiple-organ involvement, confirming that COVID-19 is a systemic disease rather than just a lung disorder (**Fig. 1**). We thus propose that only a detailed clinical characterization can allow to disentangle the complex relationship between genes and signs/symptoms.

In order to test the hypothesis that the COVID-19 susceptibility is due to one or more genes in common among patients, we used the gene burden test to compare the rate of disrupting mutations per gene. This test has already been successfully applied to discover susceptibility genes for Respiratory Syncytial Virus infection [13]. We identified 2 genes whose damage represents a susceptibility factor. Mutations in *PRKRA* (protein kinase activator A, alias PACT; OMIM# *603424), a protein kinase activated by viral double-stranded RNA may impair the down-stream IFN-mediated immune response [14,15]. Mutations in *LAPTM4B* (Lysosomal Protein Transmembrane 4 Beta) gene, may impair endosomal network, eventually compromising productive viral infection [16,17].

We then identified 2 genes whose damage represents a protective factor: *OR4C5* and *NDUFAF7*. *OR4C5* is a "resurrected" pseudogene, known to be non functioning in half of the European population, with a frequency of inactive allele of 0.62 in Asians, 0.48 in Europeans and 0.16 in Africans [18,19]. Expression of the "resurrected" pseudogene *OR4C5* may help in triggering the natural immunity leading to virus and cell death [20,21]. It is interesting to note that protein atlas shows *OR4C5* protein expression in the liver without the corresponding mRNA expression (www.proteinatlas.org) suggesting that *OR4C5* reaches the liver through nerve terminals [22]. If this is the case, those individuals expressing the resurrected *OR4C5* gene may have more triggers of innate immunity and subsequently higher liver damage, in agreement with the putative expression of *OR4C5* (white boxes) in patients with liver impairment (**Fig. 1**).

Previous studies reported a prevalence of olfactory disorders in COVID-19 population ranging from 5% to 98%. A recent meta-analysis of 10 studies demonstrated a

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

52.73% prevalence for smell dysfunction in COVID-19 subjects [23]. In our population, only 3/35 (8.6%) subjects reported olfactory disorders. Both the limited sample size and the characteristic of the population (severely affected hospitalized subjects) could explain this result. However, a report focusing on smell dysfunction in severely affected hospitalized subjects reported a prevalence of 23.7% among 59 patients [24].

We explored the hypothesis that each patient could have one unique combination of rare pathogenic/highly relevant variants related for different reasons to infection susceptibility [9] (**Fig. 1**): G6PD-deficient cells are more susceptible to several viruses including coronavirus and have down-regulated innate immunity (in line with the observed very low levels of IL-6) (**Fig. 1**) [25]; *ZEB1*-linked corneal dystrophy, known to function in immune cells, and playing an important role in establishing both the effector response and future immunity in response to pathogens [26]; *TGFBI* mutations (associated with corneal dystrophy); *ABCC6* gene mutations (associated with pseudoxanthoma elasticum); likely hypomorphic mutations in *CHD7* or *COL5A1/2* variants, playing a role as modulators of immune cells activity and/or response to infections [27–34]; *ADAR*, involved in viral RNA editing; *CLEC4M*, an alternative receptor for SARS-CoV [35] *HCRTR1/2*, receptors of Hypocretin, important in the regulation of fatigue during infections [36]; *FURIN*, a serine protease that cleaves the SARS-Cov-2 minor capsid protein important for ACE2 contact and viral entry into the host cells [37,38].

Finally, interesting rare variants have been identified in NitricOxide synthase *NOS3* and Opioid receptor *OPRM1*. Opioid ligands may regulate the expression of chemokines and chemokine receptors [39]. NitricOxide (NO), mainly produced by epithelial and white blood cells (iNOS) and to a lesser extent by endothelial cells (eNOS), is able to significantly reduce viral infection and replication of SARS-CoV in normal condition through two distinct mechanisms: impairment of the fusion between the spike protein and its receptor ACE2, and reduction of viral RNA production [40]. Mutations in NO synthase may disrupt

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

one or both the above reported functions and clinical trials are ongoing to evaluate the effectiveness of inhaled NO in COVID-19 patients [41,42]**.**

Several rare variants in Interleukins (*ILs*) and Interleukins receptors (*ILRs*) are found. Interleukins are crucial in modulating immune response against all types of infective agents. The variants reported in this study include different interleukins that are not specifically involved in the defense against virus but are critical in balancing both innate and specific adaptive immune response (**Fig. 1**)

Furthemore, we identified common "pathogenic" variants in genes known to be linked to viral infection, such as *MBL2*, *IRGM* and *SAA1,* and/or specific organ damage as *PRSS1*. Polymorphisms in *PRSS1*, a serine protease secreted from the pancreas, are associated with autosomal dominant hereditary pancreatitis (OMIM#167800) [43]. Polymorphisms in *MBL2*, a mannose-binding lectin secreted by the liver, cause increased susceptibility to infections, possibly due to a negative impact on the ability to mount an immune response [44,45]. Polymorphisms in *IRGM* may lead to impairment of autophagy which in turn controls innate and adaptive immunity [46,47]. *SAA1*, encoding the serum amyloid A (SAA) protein, is an apolipoprotein reactant, mainly produced by hepatocytes and regulated from inflammatory cytokines. In patients with chronic inflammatory diseases, the SAA cleavage product, Amyloid protein A (AA), is deposited systemically in vital organs including liver, spleen and kidneys, causing amyloidosis [48].

For the last above reported genes and pathogenic variants or predicted variants relevant for infection, a statistically significant difference in variant's frequency was not found between cases and controls looking at either the single variant or the single gene, as a burden effect of variants. However, as depicted in the overall **Fig. 1,** we could hypothesize a combined model in which common susceptibility genes will sum to less common or private susceptibility variants. A specific combination of these 2 categories may determine type (organotropism) and severity of the disease.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Our observations related to the huge amount of data, both on phenome and genome sides, and represented in Figure 1, could also lay the bases for association rule mining approaches. Artificial intelligence techniques based on pattern recognition may discover an intelligible picture which appears blurred at present.

We know that a possible limitation of this study is the heterogeneity of patients and controls, which are not matched for gender, major comorbidities and other clinical characteristics. For this reason, further analyses in a larger cohort of samples are mandatory in order to test this hypothesis of a combined model for COVID-19 susceptibility with a number of common susceptibility genes which represent the fertile background in which additional private, rare or low frequency mutations confer to the host the most favorable environment for virus growth and organ damage.

METHODS

Patients clinical data and Samples collection

The GEN-COVID study was approved by the University Hospital of Siena Ethical Review Board (Prot n. 16929, dated March 16, 2020). Thirty-five patients were recruited. WES data of these 35 patients were compared with those of 150 controls (Italian Genomes NIG http://www.nig.cineca.it). Patients have a mean age of 64 years with a Standard Deviation (SD) of 14.3 while the controls have a mean age of 46 years with a SD of 9.5. The percentage of males (M) and females (F) in patients is 68.5% and 31.4% respectively, while in controls is 51% and 49% respectively. The patients are clustered into four qualitative severity groups depending on the respiratory impairment and the need for ventilation: high care intensity group (those requiring invasive ventilation), intermediate care intensity group (those requiring non invasive ventilation i.e. CPAP and BiPAP, and high-flows oxygen therapy), low care intensity group (those requiring conventional oxygen therapy) and very low care intensity group (those not requiring oxygen therapy) (groups 1-4 in **Table 1** and different colors in **Fig. 1**).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Peripheral blood samples in EDTA-containing tubes and detailed clinical data were collected. All these data were inserted in a section dedicated to COVID-19 of the established and certified Biobank and Registry of the Medical Genetics Unit of the Hospital. An example of the Clinical questionnaire is illustrated in **S1 Fig.** .

Each patient was assigned a continuous quantitative respiratory score, the PaO2/FiO2 ratio (normal values >300) (P/F), as the worst value during the hospitalization.

Patients were also assigned a lung imaging grading according to X-Rays and CT scans. In particular, lung involvement was scored through imaging at the time of admission and during hospitalization (worst score), annotating the chest X-Ray (CXR) score (in 34 patients) and CT score in 1 patient for whom X-Rays were not available. To obtain the score (from 0 to 28) each CXR was divided in four quadrant (right upper, right lower, left upper and left lower) and for each quadrant the presence of consolidation (0= no consolidation; 1 <50%, 2>50%), ground glass opacities (GGOs: 0= no GGOs, 1<50%, 2 >50%), reticulation (0= no GGOs, 1<50%, 2 >50%) and pleural effusion on left or right side (0= no, 1= minimal; 2= large) were recorded. The same score was applied for CT (1 patient).

For each patient, the presence of hyposmia and hypogeusia was also investigated through otolaryngology examination, Burghart sniffin' sticks [49] and a visual analog scale (VAS). Whenever the sign was present, a score ranging from 0 to 10 was assigned to each patient using VAS where 0 means the best sense of smell and 10 represents the absence of smell sensation [50].

The presence of hepatic involvement was defined on the basis of a clear hepatic enzymes elevation as glutamic pyruvic transaminase (ALT) and glutamic oxaloacetic transaminase (AST) both higher than 40 UI/l. Pancreatic involvement was considered on the basis of an increase of pancreatic enzymes as pancreatic amylase higher than 53 UI/l and lipase higher than 60UI/l. Heart involvement was defined on the basis of one or more of the following abnormal data: Troponin T (>15 ng/L), indicative of ischemic disorder; NT-proBNP (M >88; F >153 pg/ml), indicative of heart failure and arrhythmias (indicative of electric disorder). Kidney involvement was defined in the presence of a creatinine value higher than 1,20 mg/dl in males and higher than 1,10 mg/dl in females (**Fig. 1**).

Whole Exome Sequencing analysis

Genomic DNA was extracted from peripheral blood using the MagCore®Genomic DNA Whole Blood kit (RBC Biosciences) according to manufacturer's protocol. Whole exome sequencing analysis was performed on Illumina NovaSeq 6000 system (Illumina, San

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Diego, CA, USA). DNA fragments were hybridized and captured by Illumina Exome Panel (Illumina) according to manufacturer's protocol. The libraries were tested for enrichment by qPCR, and the size distribution and concentration were determined using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The Novaseq 6000 platform (Illumina), along with 150 bp paired-end reads, was used for sequencing of DNA.

Genetic data analysis

Reads were mapped to the hg19 reference genome by the Burrow-Wheeler aligner BWA [51]. Variants calling was performed according to the GATK4 best practice guidelines [52]. Namely, duplicates were first removed by *MarkDuplicates*, and base qualities were recalibrated using *BaseRecalibration* and *ApplyBQSR*. *HaplotypeCaller* was used to calculate Genomic VCF files for each sample, which were then used for multi-sample calling by *GenomicDBImport* and *GenotypeGVCF*. In order to improve the specificity-sensitivity balance, variants quality scores were calculated by *VariantRecalibrator* and *ApplyVQSR*. Variants were annotated by ANNOVAR [53], and with the number of articles answering the query "gene_name AND viral infection" in Pubmed, where gene_name is the name of the gene affected by the variant.

In order to identify candidate genes according to the Mendelian-like model, rare variants were filtered by a prioritization approach. We used the ExAC database (http://exac.broadinstitute.org/), in particular the ExAC_NFE reported frequency to filter variants according to a minor allele frequency < 0.01. Synonymous, intronic and non-coding variants were excluded from the analysis. Mutation disease database ClinVar (ncbi.nlm.nih.gov/clinvar/) was used to identify previous pathogenicity classifications and variants reported as likely benign/benign were discarded. Filtering and prioritization of variants was completed using the CADD_Phred pathogenicity prediction tool. Finally, we selected genes involved in infection susceptibility using the term "viral infection" as Pubmed database search.

In order to identify genes with a different prevalence of functionally relevant variants between COVID-19 patients and control samples, the following score was calculated:

$$x_j = \sum_{i=1}^{K} w_i x_{i,j} \qquad (1),$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Where w_i is a weight associated with the *i-th* variant; and $x_{i,j}$ is equal to 0 if the variant is not present in sample *j*, 1 if sample *j* has the variant in heterozygous state, and 2 if sample *j* has the variant is homozygous state. The weight w_i was assumed equal to the DANN score of the variant [54], which provides an estimate of the likelihood that the variant has deleterious functional effects (i.e. variants more likely to have a functional effect contribute more to the score). The sum in equation (1) was performed over all the variants in the gene where the DANN score was available. Genes with less than 5 annotated variants were discarded from the analysis. The scores calculated by equation (1) were ranked for all the samples, and the sum of the ranking for the COVID-19 samples, named r_{COVID}, was calculated. Then, sample labels were permuted 10.000 times, and these permutations were used to estimate the average value and the standard deviation of r_{COVID} under the null-hypothesis. The p-value was calculated assuming a normal distribution for the sum of the ranking [55]. Moreover, we performed an additional more stringent quality check of genetic variants in the selected genes in order to remove calling artifacts that skipped the previous quality control.

DATA AVAILABILITY

Data about the gene-based analyses and variants are available as Supplementary Material. Genetic data are available in public database of the Italian genetic variability NIG (http://nigdb.cineca.it/index.php#).

REFERENCES

1. Dennison Himmelfarb CR, Baptiste D. Coronavirus Disease (COVID-19). J Cardiovasc Nurs. 2020;Publish Ah. doi:10.1097/jcn.0000000000000710
2. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020. doi:10.1056/NEJMoa2002032
3. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. JAMA. 2020. doi:10.1001/jama.2020.2648
4. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

individuals to HIV-1 infection. Cell. 1996. doi:10.1016/S0092-8674(00)80110-5

5. Woziwodzka A, Rybicka M, Sznarkowska A, Romanowski T, Dręczewski M, Stalke P, et al. TNF-α polymorphisms affect persistence and progression of HBV infection. Mol Genet Genomic Med. 2019. doi:10.1002/mgg3.935
6. Tian T, Huang P, Wu J, Wang C, Fan H, Zhang Y, et al. CD40 polymorphisms were associated with HCV infection susceptibility among Chinese population. BMC Infect Dis. 2019. doi:10.1186/s12879-019-4482-5
7. Nogales A, Dediego ML. Host single nucleotide polymorphisms modulating influenza a virus disease in humans. Pathogens. 2019. doi:10.3390/pathogens8040168
8. Williams FMK, Freydin M, Mangino M, Couvreur S, Visconti A, Bowyer RCE, et al. Self-reported symptoms of covid-19 including symptoms most predictive of SARS-CoV-2 infection, are heritable. MedRxiv. 2020.
9. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020. doi:10.1016/j.cell.2019.12.036
10. Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. Eur J Hum Genet. 2020. doi:10.1038/s41431-020-0691-z
11. Cai H. Sex difference and smoking predisposition in patients with COVID-19. The Lancet Respiratory Medicine. 2020. doi:10.1016/S2213-2600(20)30117-X
12. D'alessandro M, Bennett D, Montagnani F, Cameli P, Perrone A, Bergantini L, et al. Peripheral lymphocyte subset monitoring in COVID19 patients: a prospective Italian real-life case series. Minerva Med. 2020. doi:10.23736/S0026-4806.20.06638-0
13. Salas A, Pardo-Seco J, Cebey-López M, Gómez-Carballa A, Obando-Pacheco P, Rivero-Calle I, et al. Whole Exome Sequencing reveals new candidate genes in host genomic susceptibility to Respiratory Syncytial Virus Disease. Sci Rep. 2017. doi:10.1038/s41598-017-15752-4
14. Chan CP, Yuen CK, Cheung PHH, Fung SY, Lui PY, Chen H, et al. Antiviral activity of double-stranded RNA-binding protein PACT against influenza A virus mediated via suppression of viral RNA polymerase. FASEB J. 2018;32: 4380–4393. doi:10.1096/fj.201701361R
15. Miyamoto M, Komuro A. PACT is required for MDA5-mediated immunoresponses triggered by Cardiovirus infection via interaction with LGP2. Biochem Biophys Res Commun. 2017. doi:10.1016/j.bbrc.2017.10.048

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

16. Iwamoto M, Saso W, Sugiyama R, Ishii K, Ohki M, Nagamori S, et al. Epidermal growth factor receptor is a host-entry cofactor triggering hepatitis B virus internalization. Proc Natl Acad Sci U S A. 2019. doi:10.1073/pnas.1811064116
17. Tan X, Sun Y, Thapa N, Liao Y, Hedman AC, Anderson RA. LAPTM4B is a PtdIns(4,5)P 2 effector that regulates EGFR signaling, lysosomal sorting, and degradation . EMBO J. 2015. doi:10.15252/embj.201489425
18. Olender T, Waszak SM, Viavant M, Khen M, Ben-Asher E, Reyes A, et al. Personal receptor repertoires: olfaction as a model. BMC Genomics. 2012. doi:10.1186/1471-2164-13-414
19. Waszak SM, Hasin Y, Zichner T, Olender T, Keydar I, Khen M, et al. Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory receptor gene content diversity. PLoS Comput Biol. 2010. doi:10.1371/journal.pcbi.1000988
20. Durrant DM, Ghosh S, Klein RS. The Olfactory Bulb: An Immunosensory Effector Organ during Neurotropic Viral Infections. ACS Chemical Neuroscience. 2016. doi:10.1021/acschemneuro.6b00043
21. Mori I, Goshima F, Imai Y, Kohsaka S, Sugiyama T, Yoshida T, et al. Olfactory receptor neurons prevent disseminations of neurovirulent influenza A virus into the brain by undergoing virus-induced apoptosis. J Gen Virol. 2002;83: 2109–2116. doi:10.1099/0022-1317-83-9-2109
22. Streba LAM, Vere CC, Ionescu AG, Streba CT, Rogoveanu I. Role of intrahepatic innervation in regulating the activity of liver cells. World Journal of Hepatology. 2014. doi:10.4254/wjh.v6.i3.137
23. Tong JY, Wong A, Zhu D, Fastenberg JH, Tham T. The Prevalence of Olfactory and Gustatory Dysfunction in COVID-19 Patients: A Systematic Review and Meta-analysis. Otolaryngology - Head and Neck Surgery (United States). 2020. doi:10.1177/0194599820926473
24. Giacomelli A, Pezzati L, Conti F, Bernacchia D, Siano M, Oreni L, et al. Self-reported olfactory and taste disorders in SARS-CoV-2 patients: a cross-sectional study. Clin Infect Dis. 2020. doi:10.1093/cid/ciaa330
25. Wu YH, Tseng CP, Cheng ML, Ho HY, Shih SR, Chiu DTY. Glucose-6-phosphate dehydrogenase deficiency enhances human coronavirus 229E infection. J Infect Dis. 2008. doi:10.1086/528377
26. Guan T, Dominguez CX, Amezquita RA, Laidlaw BJ, Cheng J, Henao-Mejia J, et al.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ZEB1, ZEB2, and the miR-200 family form a counterregulatory network to regulate CD8+ T cell fates. J Exp Med. 2018. doi:10.1084/jem.20171352

27. Klamer SE, Dorland YL, Kleijer M, Geerts D, Lento WE, Van Der Schoot CE, et al. TGFBI expressed by bone marrow niche cells and hematopoietic stem and progenitor cells regulates hematopoiesis. Stem Cells Dev. 2018. doi:10.1089/scd.2018.0124
28. Ebersole JL, Peyyala R, Gonzalez OA. Biofilm-induced profiles of immune response gene expression by oral epithelial cells. Mol Oral Microbiol. 2019. doi:10.1111/omi.12251
29. Marton J, Albert D, Wiltshire SA, Park R, Bergen A, Qureshi S, et al. Cyclosporine a treatment inhibits Abcc6-dependent cardiac necrosis and calcification following coxsackievirus B3 infection in mice. PLoS One. 2015. doi:10.1371/journal.pone.0138222
30. Janssen N, Bergman JEH, Swertz MA, Tranebjaerg L, Lodahl M, Schoots J, et al. Mutation update on the CHD7 gene involved in CHARGE syndrome. Human Mutation. 2012. doi:10.1002/humu.22086
31. Theodoropoulos DS, Theodoropoulos GA, Edwards BM, Kileny PR, Van Riper LA. Immune deficiency and hearing loss in CHARGE association [3]. Pediatrics. 2003. doi:10.1542/peds.111.3.711-a
32. Gennery AR, Slatter MA, Rice J, Hoefsloot LH, Barge D, McLean-Tooke A, et al. Mutations in CHD7 in patients with CHARGE syndrome cause T-B + natural killer cell + severe combined immune deficiency and may cause Omenn-like syndrome. Clin Exp Immunol. 2008. doi:10.1111/j.1365-2249.2008.03681.x
33. Randall V, McCue K, Roberts C, Kyriakopoulou V, Beddow S, Barrett AN, et al. Great vessel development requires biallelic expression of Chd7 and Tbx1 in pharyngeal ectoderm in mice. J Clin Invest. 2009. doi:10.1172/JCI37561
34. Zhetkenev S, Khassan A, Khamzina A, Issanov A, Crape B, Akilzhanova A, et al. Association of rs12722 COL5A1 with Pulmonary Tuberculosis infection: a preliminary case-control study in a Kazakhstani population. 2019;2017: 19008995. doi:10.1101/19008995
35. Chan VSF, Chan KYK, Chen Y, Poon LLM, Cheung ANY, Zheng B, et al. Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. Nat Genet. 2006. doi:10.1038/ng1698
36. Zhan S, Cai GQ, Zheng A, Wang Y, Jia J, Fang H, et al. Tumor necrosis factor-alpha regulates the Hypocretin system via mRNA degradation and ubiquitination. Biochim

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Biophys Acta - Mol Basis Dis. 2011. doi:10.1016/j.bbadis.2010.11.003

37. Braun E, Hotter D, Koepke L, Zech F, Groß R, Sparrer KMJ, et al. Guanylate-Binding Proteins 2 and 5 Exert Broad Antiviral Activity by Inhibiting Furin-Mediated Processing of Viral Envelope Proteins. Cell Rep. 2019. doi:10.1016/j.celrep.2019.04.063
38. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci U S A. 2020. doi:10.1073/pnas.2003138117
39. Finley MJ, Happel CM, Kaminsky DE, Rogers TJ. Opioid and nociceptin receptors regulate cytokine and cytokine receptor expression. Cellular Immunology. 2008. doi:10.1016/j.cellimm.2007.09.008
40. Åkerström S, Gunalan V, Keng CT, Tan YJ, Mirazimi A. Dual effect of nitric oxide on SARS-CoV replication: Viral RNA production and palmitoylation of the S protein are affected. Virology. 2009. doi:10.1016/j.virol.2009.09.007
41. Åkerström S, Mousavi-Jazi M, Klingström J, Leijon M, Lundkvist Å, Mirazimi A. Nitric Oxide Inhibits the Replication Cycle of Severe Acute Respiratory Syndrome Coronavirus. J Virol. 2005. doi:10.1128/jvi.79.3.1966-1969.2005
42. Zamanian RT, Pollack C V., Gentile MA, Rashid M, Fox JC, Mahaffey KW, et al. Outpatient inhaled nitric oxide in a patient with vasoreactive idiopathic pulmonary arterial hypertension and COVID-19 infection. American Journal of Respiratory and Critical Care Medicine. 2020. doi:10.1164/rccm.202004-0937LE
43. Teich N, Nemoda Z, Köhler H, Heinritz W, Mössner J, Keim V, et al. Gene conversion between functional trypsinogen genes PRSS1 and PRSS2 associated with chronic pancreatitis in a six-year-old girl. Hum Mutat. 2005. doi:10.1002/humu.20148
44. Thio CL, Mosbruger T, Astemborski J, Greer S, Kirk GD, O'Brien SJ, et al. Mannose Binding Lectin Genotypes Influence Recovery from Hepatitis B Virus Infection. J Virol. 2005. doi:10.1128/jvi.79.14.9192-9196.2005
45. Dean MM, Flower RL, Eisen DP, Minchinton RM, Hart DNJ, Vuckovic S. Mannose-binding lectin deficiency influences innate and antigen-presenting functions of blood myeloid dendritic cells. Immunology. 2011. doi:10.1111/j.1365-2567.2010.03365.x
46. Singh SB, Davis AS, Taylor GA, Deretic V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. Science (80-). 2006. doi:10.1126/science.1129577
47. Rufini S, Ciccacci C, Di Fusco D, Ruffa A, Pallone F, Novelli G, et al. Autophagy

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and inflammatory bowel disease: Association between variants of the autophagy-related IRGM gene and susceptibility to Crohn's disease. Dig Liver Dis. 2015. doi:10.1016/j.dld.2015.05.012

48. Zhang Y, Zhang J, Sheng H, Li H, Wang R. Acute phase reactant serum amyloid A in inflammation and other diseases. Advances in Clinical Chemistry. 2019. doi:10.1016/bs.acc.2019.01.002
49. Oleszkiewicz A, Schriever VA, Croy I, Hähner A, Hummel T. Updated Sniffin' Sticks normative data based on an extended sample of 9139 subjects. Eur Arch Oto-Rhino-Laryngology. 2019. doi:10.1007/s00405-018-5248-1
50. Klimek L, Bergmann KC, Biedermann T, Bousquet J, Hellings P, Jung K, et al. Visual analogue scales (VAS) - Measuring instruments for the documentation of symptoms and therapy monitoring in case of allergic rhinitis in everyday health care. Allergo J. 2017. doi:10.1007/s40629-016-0006-7
51. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010. doi:10.1093/bioinformatics/btp698
52. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GA Van der, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2017. doi:10.1101/201178
53. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010. doi:10.1093/nar/gkq603
54. Quang D, Chen Y, Xie X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics. 2015. doi:10.1093/bioinformatics/btu703
55. Dering C, Hemmelmann C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: An overview of collapsing methods. Genet Epidemiol. 2011. doi:10.1002/gepi.20643

ACKNOWLEDGEMENTS

This study is part of GEN-COVID, https://sites.google.com/dbm.unisi.it/gen-covid the Italian multicenter study aimed to identify the COVID-19 host genetic bases The *Genetic*

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and COVID-19 Biobank of Siena, member of BBMRI-IT, of Telethon Network of Genetic Biobanks (project no. GTB18001), of EuroBioBank, and of D-Connect, provided us with specimens. We thank the CINECA consortium for providing computational resources and Network for Italian Genomes NIG http://www.nig.cineca.it. We thank private donors' support to A.R. (Department of Medical Biotechnologies, University of Siena) for the COVID-19 host genetics research project (D.L n.18 of March 17th 2020).

GEN-COVID Multicenter Study (composition at May 22, 2020, the representative of the GEN-COVID multicenter study is Prof. Francesca Mari email: francesca.mari@unisi.it)

Gabriella Doddato[1], Susanna Croci[1], Laura Di Sarno[1], Andrea Tommasi[1,2], Sergio Daga[1], Maria Palmieri[1], Massimiliano Fabbiani[5], Barbara Rossetti[5], Giacomo Zanelli[3,5], Paolo Cameli[6], David Bennett[6], Simona Marcantonio[7], Sabino Scolletta[7], Federico Franchi[7], Luca Cantarini[9], Bruno Frediani[9], Danilo Tacconi[10], Chiara Spertilli[10], Marco Feri[11], Alice Donati[11], Raffaele Scala[12], Luca Guidelli[12], Agostino Ognibene[13], Genni Spargi[14], Marta Corridi[14], Cesira Nencioni[15], Leonardo Croci[15], Gian Piero Caldarelli[16], Maurizio Spagnesi[17], Paolo Piacentini[17], Anna Canaccini[18], Agnese Verzuri[18], Valentina Anemoli[18], Massimo Vaghi[21], Antonella D'Arminio Monforte[22], Esther Merlini[22], Mario Umberto Mondelli[23,24], Stefania Mantovani[23], Serena Ludovisi[24], Massimo Girardis[25], Sophie Venturelli[25], Andrea Cossarizza[26], Andrea Antinori[27], Alessandra Vergori[27], Stefano Rusconi[28,29], Matteo Siano[28,29], Arianna Gabrieli[29], Daniela Francisci[30,31], Elisabetta Schiaroli[30], Pier Giorgio Scotton[32], Francesca Andretta[32], Sandro Panese[33], Renzo Scaggiante[34], Saverio Giuseppe Parisi[35], Francesco Castelli[36], Maria Eugenia Quiros Roldan[36], Paola Magro[36], Cristina Minardi[36], Matteo Della Monica[37], Carmelo Piscopo[37], Mario Capasso[38,39,40], Massimo Carella[41], Marco Castori[41], Giuseppe Merla[41], Filippo Aucella[42], Pamela Raggi[43], Matteo Bassetti[44,45], Antonio Di Biagio[45], Maurizio Sanguinetti[46,47], Luca Masucci[46,47], Chiara Gabbi[19], Serafina Valente[18], Susanna Guerrini[8], Elisa Frullanti[1], Ilaria Meloni[1], Maria Antonietta Mencarelli[2], Caterina Lo Rizzo[2], Anna Maria Pinto[2]

10) Department of Specialized and Internal Medicine, Infectious Diseases Unit, San Donato Hospital Arezzo, Italy
11) Department of Emergency, Anesthesia Unit, San Donato Hospital, Arezzo, Italy
12) Department of Specialized and Internal Medicine, Pneumology Unit and UTIP, San Donato Hospital, Arezzo, Italy
13) Clinical Chemical Analysis Laboratory, San Donato Hospital, Arezzo, Italy
14) Department of Emergency, Anesthesia Unit, Misericordia Hospital, Grosseto, Italy
15) Department of Specialized and Internal Medicine, Infectious Diseases Unit, Misericordia Hospital, Grosseto, Italy
16) Clinical Chemical Analysis Laboratory, Misericordia Hospital, Grosseto, Italy
17) Department of Prevention, Azienda USL Toscana Sud Est, Italy
18) Territorial Scientific Technician Department, Azienda USL Toscana Sud Est, Italy
19) Independent Scientist, Milan, Italy
20) Department of Cardiovascular Diseases, University of Siena, Italy

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

21) Chirurgia Vascolare, Ospedale Maggiore di Crema, Italy
22) Department of Health Sciences, Clinic of Infectious Diseases, ASST Santi Paolo e Carlo, University of Milan, Italy
23) Division of Infectious Diseases and Immunology, Department of Medical Sciences and Infectious Diseases, Pavia, Italy.
24) Department of Internal Medicine and Therapeutics, University of Pavia, Italy
25) Department of Anesthesia and Intensive Care, University of Modena and Reggio Emilia, Modena, Italy
26) Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, Modena, Italy
27) HIV/AIDS Department, National Institute for Infectious Diseases, IRCCS, Lazzaro Spallanzani, Rome, Italy
28) III Infectious Diseases Unit, ASST-FBF-Sacco, Milan, Italy
29) Department of Biomedical and Clinical Sciences Luigi Sacco, University of Milan, Milan, Italy
30) Infectious Diseases Clinic, Department of Medicine 2, Azienda Ospedaliera di Perugia and University of Perugia, Santa Maria Hospital, Perugia, Italy
31) Infectious Diseases Clinic, "Santa Maria" Hospital, University of Perugia, Perugia, Italy
32) Department of Infectious Diseases, Treviso Hospital, Local Health Unit 2 Marca Trevigiana, Treviso, Italy
33) Infectious Diseases Department, Ospedale Civile "SS. Giovanni e Paolo" , Venice , Italy
34) Infectious Diseases Clinic, ULSS1, Belluno, Italy
35) Department of Molecular Medicine, University of Padova, Italy
36) Department of Infectious and Tropical Diseases, University of Brescia and ASST Spedali Civili Hospital, Brescia , Italy.
37) Medical Genetics and Laboratory of Medical Genetics Unit, A.O.R.N. "Antonio Cardarelli", Naples, Italy.
38) Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy.
39) CEINGE Biotecnologie Avanzate, Naples, Italy
40) IRCCS SDN, Naples, Italy.
41) Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy.
42) Department of Nephrology and Dialysis, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy.
43) Department of Medical Sciences, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy.
44) Department of Health Sciences, University of Genova, Genova, Italy.
45) Infectious Diseases Clinic, Policlinico San Martino Hospital, IRCCS for Cancer Research Genova, Italy.
46) Microbiology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Catholic University of Medicine, Rome, Italy.
47) Department of Laboratory Sciences and Infectious Diseases, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy.

AUTHOR CONTRIBUTIONS STATEMENT

A.R. and F.M. designed the project and experiments. F.F, M.B, A.E, F.A, E.C., M.dA., S.M,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

M.A.M., F.M., M.M., E.B., A.R. and F.M. performed clinical evaluations. A.G., F.V., S.A., L.B., and M.B. carried the experiments. E.B. and S.F performed bioinformatic analyses. R.T., C.F., and E.B. carried out statistical analysis and prepared the tables. E.B., A.R. and F.M. wrote the manuscript. E.B. submitted this paper. All authors reviewed the manuscript.

ADDITIONAL INFORMATION

The authors declare no competing interests.

FIGURE LEGENDS

Fig. 1. Clinical characteristic and mutated genes. The population is clustered into four qualitative severity groups indicated with different colors depending on the respiratory impairment and the need for ventilation. Red color is used for high care intensity group (those requiring invasive ventilation), orange for intermediate care intensity group (those requiring non invasive ventilation i.e. CPAP and BiPAP, and high-flows oxygen therapy), pink for low care intensity groups (those requiring conventional oxygen therapy) and light blue for very low care intensity groups (those not requiring oxygen therapy). Patients COV132-57 and COV133-58 are reported in grey because they are siblings. A detailed clinical characterization is provided (i.e multiple organs involvement, presence of comorbidity, clinical laboratory parameters, etc.) along with the genetic background for each patient. Liver and pancreas involvement is indicated in green color, heart involvement in magenta and kidney involvement in purple. The presence of hyposmia, classified from 0 to 10 using VAS score, is indicated in light blue. A darker shade of color represents a more severe organ damage. In-silico predicted deleterious variants in genes relevant for infection and pathogenic variants (both common and rare) reported in ClinVar Database are described and a further subdivision between genes involved in a mendelian disorder and/or viral infection susceptibility is provided. For all these gene categories, dark grey is used to identify the homozygous status of the variants while light grey for the heterozygous status.

In the end, statistically significant genes obtained after Gene Burden analysis are listed: *PRKRA* and *LAPTM4B* mutational burden resulted to be enriched in the 35 COVID-19 patients compared to the controls, while *OR4C5* and *NDUFAF7* have proven to have an opposite trend, having a mutational burden more enriched in controls. For this reason, for *NDUFAF7, OR4C5*, genes the grey color and the white color are inverted because having

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

less variants and by consequence a more functional gene represents a susceptibility factor. For this category, white color underlies a higher mutational burden while grey color indicates lower mutational burden. Clinical known co-morbidities: CHF (Congestive Heart Failure), Hy (Hypertension), CC (Colon Cancer), Le (Leukemia), DM (Diabetes mellitus), pHBV (Previous HBV infection), AMI (Acute myocardial infarction), BC (Breast Cancer), Ht (Hypothyroidism), IC (Ictus Cerebri), O (Obesity), Dy (Dyslipidemia), KF (Kidney Failure), CR (Congenital Rickets), AF (Atrial Fibrillation), Di (Diverticolosis), Br (Bronchiectasis), COPD (Chronic obstructive pulmonary disease), D (Depression), Tn (Thyroid Nodules), PC (Prostate Cancer), M (Melanoma), BCC (Basal cell carcinoma). Immunotherapy: T (Tocilizumab), R (Ruxolitinib), B (Baricitinib). Steroids: M (Methylprednisolone), D (Dexamethasone). Other anti-COVID19 drugs: HCQ (Hydroxicloroquine), A (Azithromycine), REM (Remdesevir), L/R (Lopinavir/Ritonavir), C (Cloroquine).

Fig. 2. Mutational burden at the gene level. The significance threshold including the Bonferroni correction is shown as a red line (□ = 0.05, number of test, m = 7196). P-values for genes below the significance threshold are shown as grey dots. Black circles are used for the 4 genes that were identified as statistically different between COVID-19 samples and controls.

Table 1. Clinical characteristics of Italian COVID19 patients. COVID-19 cohort is grouped in 4 qualitative severity groups depending on the respiratory impairment and the need of ventilation. Group 1 requires invasive ventilation. Group 2 requires CPAP/BiPAP/high-flows oxygen therapy. Group 3 requires conventional oxygen therapy. Group 4 does not require oxygen therapy. Clinical characteristics are listed and the number of patients are indicated for each of them.

S1 Table. List of genes conferring COVID-19 susceptibility identified with the gene burden test analysis. Genes harboring deleterious mutations with statistically significant higher frequency in control than in COVID-19 patients are ordered based on p-value deriving from gene burden test analysis. The p-value adjusted is provided after Bonferroni correction.

S2 Table. List of COVID-19 protective genes identified with the gene burden test analysis. Genes harboring deleterious mutations with statistically significant higher frequency in COVID-19 patients than in control are ordered based on p-value deriving from

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

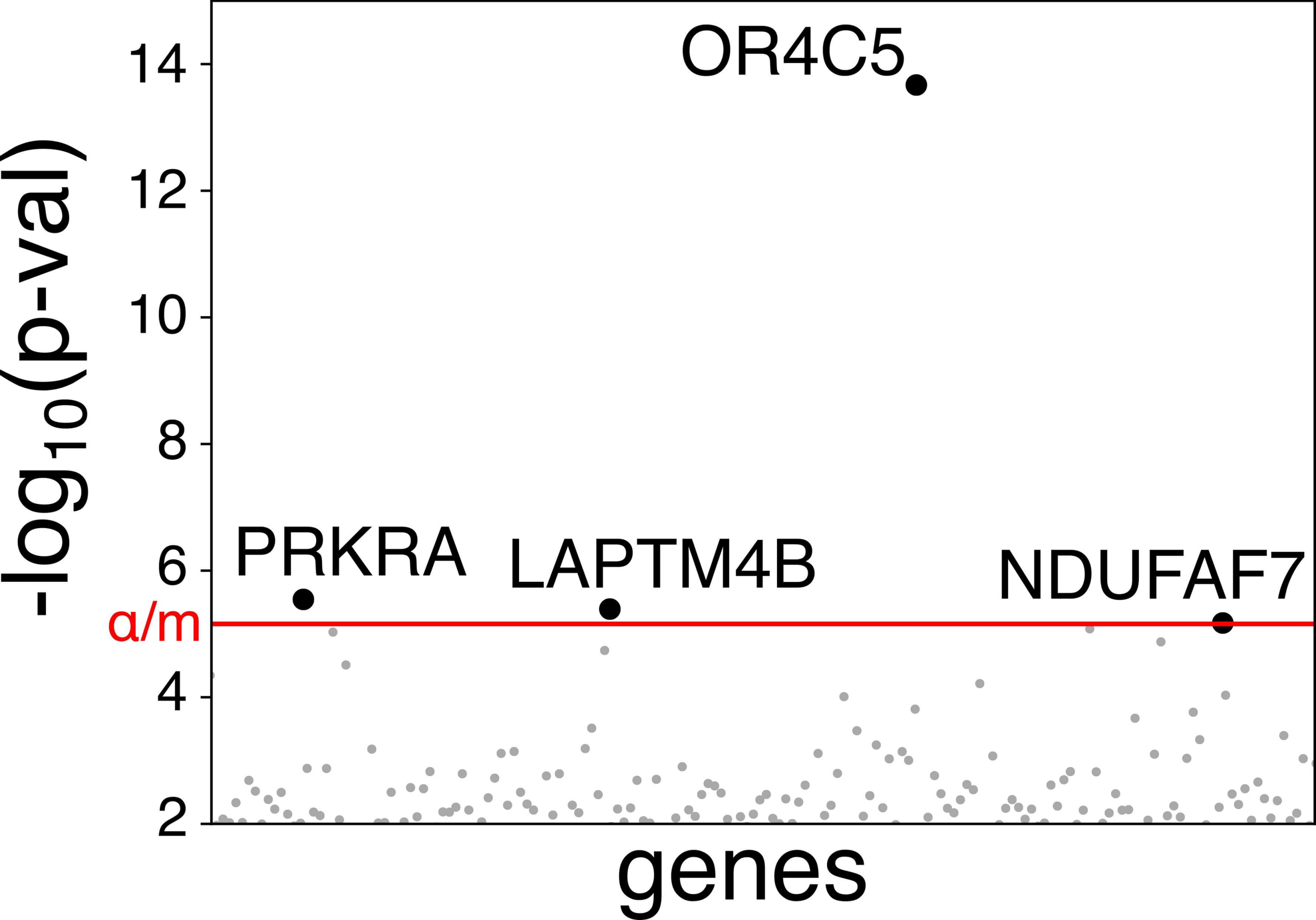
gene burden test analysis. The p-value adjusted is provided after Bonferroni correction.

S3 Table. Rare variants identified in patients cohort. Rare variants identified in COVID-19 patients according to the Mendelian-like model are reported (see *Methods section*)

S1 Fig. Clinical Data Questionnaire. The Questionnaire includes five different categories of data: Patient personal anamnesis and family history, Diagnostic Information, Laboratory Tests, Therapy and Complications. Clinical data were collected in detail for all COVID-19 patients.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20108845; this version posted November 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

	Sample ID	COV20-10	COV8-4	COV127-52	COV4-2	COV119-44	COV120-45	COV134-59	COV121-46	COV122-47	COV123-48	COV124-49	COV125-50	COV128-53	COV130-55	COV237-150	COV126-51	COV22-11	COV28-14	COV136-61	COV129-54	COV238-171
	Gender	F	M	F	M	F	F	M	M	M	F	F	M	M	F	M	M	M	F	M	M	M
	Respiratory Severity P/F score (worst)	293	318	140	285	368	100-200	264	280	Unknown	100-200	200-300	200-300	279	312	>300	304	347		>300	400	-
	Lung Imaging Grading (0-28) (worst)	6	7	14	5	8	11	10	10	3	9	7	14	8	6	4	8	9	9	3	5	0
	Hyposmia VAS Grading (0-10)	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0
	Hypogeusia	0	0	0	0	0	0	0	0	0	Yes	No	No	No	No	No	No	No	No	No	No	No
	Heart involvement	Yes T	Yes B/A	Yes B/A	No	No	No	Yes T/A	Yes T/B	Yes B	No	No	No	No	No	No	No	No	No	No	No	No
	Liver involvement	Yes	Yes	Yes	Borderline	No	No	No	Yes	No	No	Yes	No	No	No	Yes	Yes	Yes	No	Yes	No	No
	Pancreas involvement	Yes	No	Yes	Borderline	No	No	Yes	Yes	No	No	Yes	No	No	Yes	No	No	No	No	Yes	No	No
	Kidney involvement	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No	No	No	No
CLINICAL CHARACTERIZATION	**CD4 lymphocytes** (cell/ul (mm^3); n.v. 400-1500)(worst)	1493	458	185	2115	616	758	1705	361	591	194	623	226	485	1116	381	741	736				
	NK cells (cell/ul (mm ^3); n.v. 90-590)(worst)	37	120	128	112	85	147	180	305	98	101	334	40	141	204	291	117	252				
	Neutrophils (min value) **/Lymphocytes** (min value)	3.6	1.8	10.9	7.6	2.4	4.7	4.6	4.4	15.2	5.4	1.7	6.3	6.9	1.1	1.6	12.1	2.32				
	IL6 (pg/ml; n.v. <27)(basal value)	14.9	5.2	-	2.9	-	28.6	36.3	-	-	-	-	126.1	19	5.4	-	-	-				
	IL6 times the upper limit of normal	-	-	-	-	-	1.05	1.3	-	-	-	-	4.7	-	-	-	-	-				
	CRP (mg/dl; n.v. <0.5)	5.13	22.27	18.17	2.07	9.6	6.5	4.09	25.3	0.19	8.42	7.08	15.08	3.05	3.34	3.14	5.11	9.06	0.59	2.34	5.56	0.05
	Fibrinogen (mg/dl; n.v. Klauss 200-400)(min value)	566	701	575	772	680	282	486	859	672	391	471	660	373	531	427	660	546	378	-	623	-
	LDH (UI/l; n.v. M 135-225; F	593	371	350	255	528	264	250	272	198	264	358	412	220	335	213	309	292	232	294	255	143
	D-dimer (ng/ml; n.v. <500)	6765	903	2190	1306	2934	1749	790	1002	1007	1167	666	912	1055	23894	1148	1286	1509	599	-	621	425
	Immunotherapy	No	T	No	No	No	T	B	No	No	No	No	No	No	No	No	No	No	No	No	No	-
	Steroids	D	No	M	No	D	No	M	M	M	M	M	M	M	M	No	No	No	No	-	M	-
	Enoxaparin	No	Yes	Yes	No	Yes	Yes	No	Yes	no	Yes	Yes	No	Yes	Yes	-	Yes	Yes	Yes	-	Yes	-
	Other anti-COVID19 drugs	-	HCQ	HCQ, A	-	-	HCQ, A	HCQ	HCQ, A	-	A	HCQ	HCQ, A	HCQ, A	HCQ	-	HCQ	HCQ, L/R	-	HCQ, L/R	HCQ, A	-
	Clinical known co-morbidities	Ht, DM	Hy	BC, Ht, Hy	G6PD Deficiency, Hy	O, P	-	Hy AF, CHF	Ht	DM, pHBV, Ht, AMI	BC, DM, HT, O, Dy	Ht	AF, DM	Hy	Ht	Di	Br	-	D, TN	-	PC, M, BCC	-
RARE VARIANTS REPORTED AS PATHOGENIC IN ClinVar	**MENDELIAN DISORDER** - NO VIRAL INFECTION		PPP2R1B (G90D)													PRLR (I170L)				PRLR (I170L)		PPOX (P256R)
	MENDELIAN DISORDER - VIRAL INFECTION SUSCEPTIBILITY				G6PD/ ZEB1						PMS2											
	CARRIER of MEDELIAN DIS - VIRAL INFECTION			LPL				FCN3				MPO	FCN3	PMM2 (R141H)			ALG8 (R364*)	MPO		SLC36A2 (G87V)		ATP7B / OCA2
	CARRIER of MEDELIAN DIS - VIRAL INFECTION		UTR3'									Intronic		Intronic				R1304R	V400I			
RARE VARIANTS RELEVANT FOR INFECTION	*ADAR*									A935V												
	AIRE								H415R									R473T				
	CASP8											Y310*								I315V		
	CHD7					R947Q			G2786R													
	CLEC4M								V742M					Q10K								
	COL5A1/2				COL5A1 (P149L; E269K)											COL5A2 (I1279V)	COL5A1 (T915M)					
	FURIN									R483W												
	HCRTR1/2								V93I													
	ILs and ILRs				IL12RB1 (G441S)			IL13 (T41I)						IL16 (T439M)							IL13 (L72P)	
	Immunodeficiency genes								DCLRE1C (S520fs5*)		CORO1A (V174M); JAK3 (R840C)		FOXN1 (P475L); (S570L)	DOCK8 (H541Q)							CARD11 (I544L)	
	NOS3			V1073A													G173R					P1084S
	OPRM1			S437N	N386D									R260H								
COMMON VARIANTS REPORTED AS PATHOGENIC IN ClinVar	*MBL2*	G54D		G54D		G54D	G54D				G54D		G54D				G54D			G54D	G54D	
	IRGM (L105L)																					
	PTPRJ (Q276P)																					
	CST3 (A25T)																					
	SAA1 (A70V)																					
	PRSS1	N29I/ N54S	N29I/ N54S	N29I/ N54S	N29I/ N54S	N54S	N54S	N29I/ N54S	N29I	N54S	N29I	N29I/ N54S	N29I/ N54S	N29I/ N54S	N29I/ N54S	N29I/ N54S	N29I/ N54S	N54S	N29I/ N54S	N29I/ N54S	N54S	N29I/ N54S
STATISTICALLY SIGNIFICANT GENES by collapsing gene analysis	*PRKRA* (I226N)																					
	LAPTM4B					P50T													P219L			
	Mutations more common in controls:																					
	OR4C5																					
	NDUFAF7							E56D			R17H											


OR4C5
PRKRA
LAPTM4B
NDUFAF7
α/m
14
12
10
8
6
4
2
-log10(p-val)
genes