Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19

Immacolata Andolfo, Roberta Russo, Vito Alessandro Lasorsa, ..., Massimo Zollo, Achille Iolascon, Mario Capasso
mario.capasso@unina.it

Highlights
Genetic analysis was performed on 7,970 individuals hospitalized for COVID-19
Five SNPs within TMPRSS2/MX1 locus (chr.21) are associated with severe COVID-19
The minor alleles of the five SNPs correlated with high level of MX1 expression in blood
MX1 could be a potential therapeutic target in patients with COVID-19
A genome-wide association study (GWAS) (Ellinghaus et al., 2020) identified two susceptibility loci of severe COVID-19: the first locus on chromosome 3 harbors multiple genes (SLC6A20, LZFTL1, CCR9, CXCR6, XCR1, FYCO1) that could be functionally implicated in COVID-19 pathology; the second on chromosome 9 that defines the ABO blood groups (Ellinghaus et al., 2020). Other very recent papers reported the results from the analysis of two large independent GWASs that validated the two previous risk loci and found novel risk variants at chromosome 19p13.3, 12q24.13, and 21q22.1 associated with severe COVID-19 (Pairo-Castineira et al., 2021; Shelton et al., 2020). Two whole-exome sequencing studies showed that inactivating rare mutations in genes belonging to the type I interferon pathway predispose to life-threatening COVID-19 pneumonia (van der Made et al., 2020; Zhang et al., 2020). Additionally, preliminary results on a small set of Italian cases suggest that coding variants in TMPRSS2 and PCSK3 may contribute to the variability in infection susceptibility and severity (Latinis et al., 2020).
In our previous opinion article, based on the analysis of allele frequencies across different populations and expression quantitative trait loci (eQTLs) data, we hypothesized that common variants on chromosome 21 near TMPRSS2 and MX1 genes may be genetic risk factors associated with the COVID-19 different clinical manifestations (Russo et al., 2020). Both TMPRSS2 and MX1 are involved in the host response to SARS-CoV-2 infection. ACE2 is the main entry receptor for SARS-CoV-2 (Wang et al., 2020). Entry depends on the binding of the surface unit S1 of the spike (S) protein of the virus to the receptor. SARS-CoV-2 engages ACE2 as the entry receptor and employs the host cellular TMPRSS2 for S-protein priming (Hoffmann et al., 2020b; Matsuyama et al., 2020). Particularly, binding of SARS-CoV-2 S-protein with ACE2 receptor is then followed by host TMPRSS2-mediated cleavage of the viral S-protein. This process, defined as priming, involves cleavage of the S-protein at S1/S2 and S2 sites which is essential for the viral fusion with the host cell membrane before entry into the cell (Hoffmann et al., 2020b; Matsuyama et al., 2020). SARS-CoV-2 can use other proteases such as cathepsin B/L for S-protein in the absence of TMPRSS2 receptors. However, in the lungs (the primary organ for SARS-CoV-2 infection), cathepsin B/L cannot substitute for TMPRSS2 protease activity as the latter is indispensable for viral entry as observed for SARS-CoV and MERS-CoV (Hoffmann et al., 2020a). MX1 is an interferon-α/β inducible gene that encodes a guanosine triphosphate metabolizing protein involved in the cellular antiviral response (Ciancanelli et al., 2016).

In this study, to further support our hypothesis, we exploited GWAS meta-analysis data from the COVID-19 Host Genetics Initiative (COVID-19 Host Genetics Initiative, 2020) and performed an in-depth genetic analysis of chromosome 21 using summary statistics where common variants at this chromosome were associated with severe COVID-19 at the genome-wide significance level ($p \leq 5 \times 10^{-8}$). Using the cohort of 908,494 subjects with European origins, we found five SNPs at the TMPRSS2/MX1 locus showing suggestive association with the disease. All five SNPs replicated the association in two independent cohorts of Asian subjects, whereas two SNPs confirmed the association in African and one SNP in the Italian cohort. Significant eQTLs signals were found for the MX1 gene in blood.

RESULTS

**TMPRSS2/MX1 locus is associated with severe COVID-19**

To prove that common variants at TMPRSS2/MX1 (21q22.3) locus may affect the susceptibility to severe COVID-19 onset, we analyzed the summary statistics of a large available GWAS dataset released by the COVID-19 Host Genetics Initiative (COVID-19 Host Genetics Initiative, 2020). The data set includes 6,406 hospitalized cases and 902,088 controls with European ancestry (Table S1). A region on chromosome 21 appears to be significantly associated with severe COVID-19 at the genome-wide level (https://www.covid19hg.org/results/) as also demonstrated in a recently published GWAS study (Pairo-Castineira et al., 2021). To investigate whether more than one association signals may exist at chromosome 21, we selected 74 SNPs showing a $p \leq 1 \times 10^{-5}$ and we identified 3 independent loci among them (Table S2). The most significant signal was represented by rs13050728 ($p = 2.76 \times 10^{-12}$, OR = 0.83, Figure 1A) that maps within the INRA2 gene. The other two signals showed a suggestive significance level ($p \leq 1 \times 10^{-6}$ and were tagged by rs111783124 ($p = 2.39 \times 10^{-6}$, OR = 1.17, Figure 1B) and rs3787946 ($p = 2.73 \times 10^{-6}$, OR = 0.87, Figure 1C), respectively. The rs3787946 maps in an intronic region of TMPRSS2 and the closest locus was MX1 (Figure 1C); herein, we named this locus as “TMPRSS2/MX1”. An in-depth inspection of the TMPRSS2/MX1 locus showed that 13 SNPs were in linkage disequilibrium (LD) with the lead rs3787946 ($r^2 > 0.8$, Table 1) and that the 5 most significant SNPs ($p$ values ranging from $2.7 \times 10^{-8}$ to $5.8 \times 10^{-8}$, Table 1) were in strong LD with each other ($r^2 \geq 0.9$, Figure S1). The other 9 SNPs showed an LD with the lead SNP rs3787946 ranging from 0.8 to 0.9 and $p$ values ranging from $6 \times 10^{-4}$ to 0.04 (Table 1). We then sought to replicate the associations of the 14 SNPs in three independent cohorts of cases and controls of GenOMMIC GWAS (Pairo-Castineira et al., 2021) with non-European ancestry. All the 11 available SNPs replicated in the east asian (EAS) ancestry population, whereas two of five SNPs in the African (AFR) one (Table 1). By using the TaqMan assay, we typed the rs12329760 variant in samples from 226 hospitalized COVID-19 patients (Table S3) and 1848 controls from Southern Italy collected in our Institute. An additional Italian cohort of 1915 controls and 770 cases, typed for rs12329760 by whole-exome sequencing, was obtained from the Network for Italian Genomes (NIG) database (Daga et al., 2021). After combining the two cohorts, we confirmed the minor allele as a protective factor against the aggressive form of the disease (Table 2, $OR_{	ext{alle}le} = 0.89$, $P_{	ext{alle}le} = 0.07$; $OR_{	ext{dominant}} = 0.57$, $P = 0.01$; $OR_{	ext{CCvsTT}} = 0.57$, $P = 0.01$). The results of our case-control study suggest that the protective effect against the severity of COVID-19 is mainly due to the TT genotype.
SNPs at TMPRSS2/MX1 locus are enriched in regulatory regions active in the thymus

We tested if the 14 SNPs (Table 1) and their proxy SNPs ($r^2 > 0.8$) were significantly over-represented in active enhancers and promoters in multiple cell types and tissues by using HaploReg v4.1. These SNPs were enriched in the regulatory regions of several tissues (Table S4) but the best enrichment was found in induced pluripotent stem cells and thymus (Figure 2A).

Figure 1. Regional association plots of the SNPs at three independent association signals of chromosome 21

Plots were generated using LocusZoom. Y axes represent the significance of association ($-\log_{10}$ transformed $p$ values) and the recombination rate. SNPs are color-coded based on pairwise linkage disequilibrium ($r^2$) with indicated lead SNPs: rs13050728 (A), rs111783124 (B) and rs3787946 (C).
| RS number | EA | OA | MAF | r² | OR | P_EUR | OR | P_EAS | OR | P_SAS | OR | P_AFR | aRegion score | bTSS score | bPredicted function | Score | cCombined score |
|-----------|----|----|-----|----|----|--------|----|--------|----|--------|----|--------|-----------------|------------|-------------------|--------|-----------------|
| rs3787946 | C  | G  | 0.23 | 1.00 | 0.87 | 2.73 x 10⁻⁶ | 0.63 | 0.026 | 0.71 | 0.02 | 0.74 | 0.07 | 0.16 | 0.29 | INTRONIC | 2 | 6 |
| rs9983330 | G  | A  | 0.23 | 0.91 | 0.88 | 3.12 x 10⁻⁶ | 0.54 | 0.004 | 0.73 | 0.04 | 0.79 | 0.16 | 0.31 | 0.64 | REGULATORY | 4 | 26 |
| rs12329760 | T  | C  | 0.24 | 0.90 | 0.88 | 3.13 x 10⁻⁶ | 0.64 | 0.029 | 0.76 | 0.08 | 0.78 | 0.14 | 0.32 | 0.41 | MISSENSE | 7 | 23 |
| rs2298661 | A  | C  | 0.23 | 0.99 | 0.88 | 4.51 x 10⁻⁶ | 0.63 | 0.030 | 0.67 | 0.01 | 0.60 | 0.01 | 0.18 | 0.35 | INTRONIC | 2 | 9 |
| rs9985159 | T  | C  | 0.23 | 0.98 | 0.88 | 5.80 x 10⁻⁶ | 0.61 | 0.018 | 0.75 | 0.06 | 0.98 | 0.89 | 0.16 | 0.46 | INTRONIC | 2 | 15 |
| rs2298660 | T  | C  | 0.20 | 0.82 | 0.88 | 0.001 | NA | NA | NA | NA | NA | 0.12 | 0.28 | INTRONIC | 2 | 4 |
| rs7364088 | A  | G  | 0.26 | 0.84 | 0.91 | 0.002 | NA | NA | NA | NA | NA | 0.19 | 0.23 | INTRONIC | 2 | 6 |
| rs2298663 | T  | C  | 0.25 | 0.87 | 1.08 | 0.005 | 1.49 | 0.052 | 1.12 | 0.40 | 0.94 | 0.66 | 0.26 | 0.37 | REGULATORY | 4 | 15 |
| rs2094881 | C  | T  | 0.25 | 0.87 | 1.08 | 0.005 | 1.47 | 0.058 | 1.10 | 0.47 | 0.93 | 0.60 | 0.29 | 0.26 | REGULATORY | 4 | 13 |
| rs8131649 | T  | C  | 0.25 | 0.85 | 0.92 | 0.007 | 0.64 | 0.035 | 0.90 | 0.46 | 1.01 | 0.93 | 0.26 | 0.35 | REGULATORY | 4 | 12 |
| rs8134203 | T  | C  | 0.26 | 0.85 | 1.08 | 0.007 | 1.49 | 0.058 | 1.09 | 0.54 | 0.91 | 0.50 | 0.26 | 0.41 | REGULATORY | 4 | 17 |
| rs8134216 | T  | C  | 0.26 | 0.85 | 1.08 | 0.007 | 1.54 | 0.038 | 1.11 | 0.43 | 0.91 | 0.49 | 0.28 | 0.4 | REGULATORY | 4 | 19 |
| rs2104810 | A  | G  | 0.26 | 0.85 | 1.08 | 0.008 | 1.54 | 0.040 | 1.10 | 0.47 | 0.90 | 0.48 | 0.23 | 0.35 | REGULATORY | 4 | 11 |
| rs8131648 | C  | T  | 0.26 | 0.85 | 1.07 | 0.036 | NA | NA | NA | NA | NA | 0.33 | 0.42 | REGULATORY | 4 | 26 |

In bold the SNPs that replicated in at least one cohort.

EA: Effect Allele, OA: Other Allele; EUR: European; EAS: East Asian; SAS: South Asian; AFR: African; ITA: Italian; MAF: minor allele frequency; OR: odds ratio.

aScores from GWAVA predictor tool.
bScores from CADD predictor tool.
cGWAVA and CADD scores were ranked from the smallest to largest and the obtained values were summed.
Table 2. Association of rs12329760 SNP with severe COVID-19 in Italian population

| Genotype | SI cases n = 226 | SI controls n = 1848 | NIG cases n = 770 | NIG controls n = 1915 | All cases n = 996 | All controls n = 3763 |
|----------|-----------------|---------------------|-----------------|----------------------|-----------------|---------------------|
|          | Genotype | n | %  | n | %  | n | %  | n | %  | n | %  | n | %  | ns | P (CI: 95%) | OR (CI: 95%) | OR (CI: 95%) | P (CI: 95%) |
| CC       |          | 164 | 72.6 | 1274 | 68.9 | 532 | 69.1 | 1289 | 67.3 | 696 | 69.9 | 2563 | 68.1 | –     | –  | –  | –  | –  |
| CT       |          | 57  | 25.2 | 497  | 26.9 | 220  | 28.6 | 554  | 28.9 | 277  | 27.8 | 1051 | 27.9 | 0.47 | 0.89 (0.64–1.22) | 0.68 | 0.96 (0.79–1.15) | 0.71 | 0.97 (0.83–1.13) |
| TT       |          | 5   | 2.2  | 77   | 4.2  | 18   | 2.3  | 72   | 3.8  | 23   | 2.3  | 149   | 4.0  | 0.14 | 0.50 (0.20–1.26) | 0.06 | 0.60 (0.35–1.02) | 0.01 | 0.57 (0.36–0.89) |

| Allele | C | 385 | 85.2 | 3045 | 82.4 | 1284 | 83.4 | 3132 | 81.8 | 1669 | 83.8 | 6177 | 82.1 | –     | –  | –  | –  | –  |
|        | T | 67  | 14.8 | 651  | 17.6 | 256  | 16.6 | 698  | 18.2 | 323  | 16.2 | 1349 | 17.9 | 0.14 | 0.81 (0.62–1.07) | 0.16 | 0.89 (0.76–1.04) | 0.07 | 0.89 (0.78–1.01) |

| Dominant | CC/CT | 221 | 97.8 | 1771 | 95.8 | 752  | 97.7 | 1843 | 96.2 | 973  | 97.7 | 3614 | 96.0 | –     | –  | –  | –  | –  |
|          | TT   | 5   | 2.2  | 77   | 4.2  | 18   | 2.3  | 72   | 3.8  | 23   | 2.3  | 149   | 4.0  | 0.15 | 0.52 (0.20–1.30) | 0.06 | 0.61 (0.36–1.03) | 0.01 | 0.57 (0.37–0.89) |

| Recessive | CC | 159 | 70.4 | 1274 | 68.9 | 532  | 69.1 | 1289 | 67.3 | 691  | 69.4 | 2563 | 68.1 | –     | –  | –  | –  | –  |
|          | CT/TT | 62  | 27.4 | 574  | 31.1 | 238  | 30.9 | 626  | 32.7 | 300  | 30.1 | 1200 | 31.9 | 0.26 | 0.84 (0.61–1.14) | 0.37 | 0.92 (0.76–1.10) | 0.28 | 0.92 (0.79–1.07) |

NIG, Network for Italian Genomes; OR, odds ratio; CI, confidence interval; SI, Southern Italy.
In bold are highlighted the statistically significant results.
Functional role of the most significant SNPs at \textit{TMPRSS2}/MX1 locus

We then investigated the predicted functional role of the 14 SNPs by GWAVA and CADD tools. We found that two of the five most significant SNPs, i.e. rs9983330 and rs12329760, showed the first (combined score = 26) and second (combined score = 23) most significant score (Table 1). The rs12329760 was classified as a coding variant (p.Val197Met) localized in the exon 6 of the \textit{TMPRSS2} gene and was predicted to be pathogenic (PolyPhen-2 = probably damaging and SIFT = deleterious).

The most significant disease-associated SNPs are eQTLs for MX1 in blood

We verified if the top five SNPs (Table 1) might cause gene expression alterations interrogating the GTEx portal for all the common variants within \textit{TMPRSS2}/MX1 locus. We found that all the top five SNPs had eQTL signals for MX1 exclusively in blood tissue. Particularly, the minor alleles of these SNPs correlated with higher expression of MX1 compared to the major alleles (Figures 2B and S2A). Of note, all the other SNPs, except for rs2298660, did not have eQTL signals for MX1 in the blood (Table S5). The two SNPs rs12329760 and rs2298660 were confirmed as eQTLs for MX1 in the blood (p = 1.79 \times 10^{-6} and 2.8 \times 10^{-6}, minor alleles correlated with a higher expression compared to the major alleles) by interrogation of another independent publicly available data set (Westra et al., 2013). \textit{TMPRSS2} is highly expressed in lung (Russo et al., 2020), so we investigated if the top five SNPs were eQTLs for \textit{TMPRSS2} in lung tissues at a nominally statistically significant level (p \leq 0.05). We found that the minor alleles of four out of five SNPs correlated with lower expression of \textit{TMPRSS2} compared to the major alleles (Figures 2C and S2B). Notably, rs12329760 is also an eQTL for \textit{TMPRSS2} in osteoblasts treated with dexamethasone (Grundberg et al., 2011).

DISCUSSION

Despite the substantial advances made in recent months in the field of SARS-CoV-2 infection, the major question remains about the identification of the factors that modulate the variable clinical spectrum of COVID-19.

Host genetic risk factors are emerging as a potential explanation for the clinical heterogeneity of COVID-19 and are also crucial to find new druggable therapeutic targets (Asselta et al., 2020; Beck and Aksentijevich, 2020; Benetti et al., 2020; Pairo-Castineira et al., 2021; Singh et al., 2020). The main host cell entry factors of SARS-CoV-2 are ACE2 and TMPRSS2 (Asselta et al., 2020; Benetti et al., 2020). The spike (S) glycoprotein of the virus binds to the ACE2 making it essential for the entry of the virus into the host cell. S-protein priming by the serine protease TMPRSS2 allows the fusion of viral and cellular membranes, resulting in virus entry and replication in the host cells (Singh et al., 2020). TMPRSS2 is emerging as a host cell factor that is critical for SARS-CoV-2 infection (Hoffmann et al., 2020b).

In our previous study, we hypothesized that common variants at chromosome 21, driving \textit{TMPRSS2} and MX1 expression, might have a mild-to-moderate effect on the susceptibility to SARS-CoV-2 infection. Particularly, genetic variants associated with reduced \textit{TMPRSS2} and elevated MX1 expression might confer...
less individual susceptibility to SARS-CoV-2 infection and favor a better outcome (Russo et al., 2020). Here, to further support our hypothesis, we exploited GWAS data of a cohort of 908,494 subjects with European origins from the COVID-19 Host Genetics Initiative (COVID-19 Host Genetics Initiative, 2020) and performed an in-depth genetic analysis of chromosome 21. We identified five common variants (rs3787946, rs9983330, rs12329760, rs2298661, and rs9985159) at locus 21q22.3 within the TMPRSS2 gene that showed suggestive associations with severe COVID-19. In particular, we found that the alleles with minor frequency were less recurrent among hospitalized patients when compared to the control individuals, suggesting their protective role against the progression of the disease. Interestingly, all five SNPs were replicated in two cohorts of Asian origin, whereas two SNPs replicated in a case series of African ancestry. Additionally, we replicated the association of the rs12329760 SNP in an independent case-control cohort of Italian origin. As “proof of concept”, the rs12329760 SNP was also detected in recent studies (Hou et al., 2020; Vargas-Alarcon et al., 2020). It was demonstrated that the SNP, in addition to its eQTL role, decreased the stability of the protein, which might impede viral entry (Vishnubhotla et al., 2020); moreover, in silico analysis demonstrated that it created a de novo pocket protein (Paniri et al., 2020). These results confirm 21q22.3 as a novel susceptibility locus to unfavorable outcome of COVID-19. Furthermore, molecular mechanisms underlying this genetic predisposition may be common among individuals with different ethnicity.

The results from our enrichment analysis for regulatory genomic regions suggested that the identified SNPs and other proxy SNPs located at 21q22.3 locus can be associated with different outcomes of COVID-19 by altering DNA elements that regulate the transcription of MX1 and likely of other genes relevant to the thymus functions. The thymus plays a significant role in the regulation of adaptive immune responses. The effect of aging on the thymus and immune senescence is well established, and the resulting inflammaging is found to be implicated in the development of many chronic diseases (Gunes et al., 2020; Kellogg and Equils, 2020). Both aging and diseases of inflammaging are associated with severe COVID-19, and a dysfunctional thymus may be implicated in unfavorable outcome of disease (Gunes et al., 2020; Kellogg and Equils, 2020). Of note, MX1 plays an important role in the thymus as part of the innate antiviral immune response. Indeed, it is exclusively expressed after engagement of the type I interferon receptor by interferon-α/β in normal fetal and post-natal human thymus, but not in the periphery. The highest level of MX1 is properly found in mature thymocytes (Colantonio et al., 2011).

The five SNPs here identified had eQTL signals for MX1 exclusively in blood tissue. Particularly, the minor allele of these SNPs correlated with higher expression of MX1 and associated with a minor risk of developing severe COVID-19. These results support the evidence that MX1 can play a relevant role in determining less severe forms of disease and are in line with a recent study that suggests MX1 as an antiviral effector against SARS-CoV-2 (Bizzotto et al., 2020). Indeed, the expression of MX1 was found to be high in SARS-CoV-2 positive subjects, negatively correlated with age, and independently associated with increased viral load (Bizzotto et al., 2020). MX1 is part of the antiviral response induced by type I and III interferons (Zav’yalov et al., 2019). Inactivating mutations in genes belonging to type I interferon pathway and the consequently decreased levels of proteins have been shown to occur in patients with severe COVID-19 (Zhang et al., 2020).

Of note, within the region on chromosome 21, significantly associated with severe COVID-19 at the genome-wide level, the most significant signal was represented by rs13050728 that maps within the INFRA2 gene. Particularly, INFRA2 gene encodes for the type I membrane protein that forms the interferon-α/β receptor, involved in the canonical host antiviral signaling mediators (Duncan et al., 2015), so associated with interferon signaling like MX1. The SNP rs13050728 was previously identified as lead variant from the meta-analysis of overlapping SNPs between GenOMICC, The COVID-19 Host Genetics Initiative and 23andMe studies and its allele C was reported to reduce the odds of severe COVID-19 as associated with an increased expression of IFNAR2 (Pairo-Castineira et al., 2021). These findings, along with ours, further support the protective role of IFN pathway against severe COVID-19.

We also report that the minor allele of four of the top five SNPs might reduce the expression of TMPRSS2 in lung tissues. In particular, the rs12329760 coding variant (p.Val197Met) is predicted to decrease the TMPRSS2 protein stability and ACE2 binding, thus decreasing virus entry into the cells (Vishnubhotla et al., 2020). Of note, this variant was recently found to be less frequent among Chinese patients with critical COVID-19 disease (Wang et al., 2020). Additionally, it correlates with lower expression of TMPRSS2 in
osteoblast treated with dexamethasone (Grundberg et al., 2011), a drug currently used to inhibit an excessive inflammation response (Group et al., 2020). Together, these data suggest that even the functions of TMPRSS2 may be affected by the occurrence of protective variants against severe COVID-19.

Finally, we want to point out that our findings highlight the effectiveness of investigating other independent (putative) risk loci, when they do not pass genome-wide significance levels. These loci, usually overlooked in extensive meta-analysis and multi-cohorts efforts, might indeed contain important genetic variants associated with severe COVID-19 and map genes relevant to the pathogenesis of this disease. We then encourage post-GWAS genetic (re)analyses using multiple data sources to unravel novel COVID-19 risk loci and possible insights on the underlying biology.

In conclusion, our results provide evidence that common variants, regulating the expression of MX1, can predispose to the risk of developing severe COVID-19. Unraveling the role of regulatory variants at the TMPRSS2/MX1 locus could represent an important starting point for the treatment of COVID-19.

Limitations of the study
The data on eQTLs related to TMPRSS2 must be interpreted with caution as these eQTL signals in the lung (p = 0.019) do not pass the GTEx significance threshold adjusted for multiple comparisons (0.000055). Additional studies are required to further verify the role of genetic variants at TMPRSS2/MX1 locus in modulating the TMPRSS2 expression. Furthermore, the statistical approach adopted in this study did not include multivariate analyses to take into account confounding factors. Although this limitation does not affect the robustness of the presented genetic associations as replicated in multiple independent cohorts, we believe that future studies will help to better define the effect of genetic variants at TMPRSS2/MX1 locus on the clinical subgroups of COVID-19 disease; for instance, performing association analyses on patients stratified by disease aggressiveness or controlled for comorbidities in larger cohorts.

METHODS
All methods can be found in the accompanying transparent methods supplemental file.

Resource availability
Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Prof. Mario Capasso, mario.capasso@unina.it.

Material availability
This study did not generate nor use any new or unique reagents.

Data and code availability
Manhattan plot and QQ plot of the results from the large GWAS “The COVID-19 Host Genetics Initiative website” are available at the website (https://www.covid19hg.org/results/). The 770 hospitalized COVID-19 cases and 1915 controls typed for rs12329760 by whole-exome sequencing were retrieved from the web database Network for Italian Genomes (NIG) available at the website (http://nigdb.cineca.it/index.php).

Prediction of the functional impact of 14 SNPs at TMPRSS2/MX1 locus was assessed by Genome Wide Annotation of VAriants (GWAVA) tool available at the website (https://www.sanger.ac.uk/sanger/StatGen_Gwava) and by Combined Annotation Dependent Depletion (CADD) tool at (https://cadd.gs.washington.edu/).

The Blood eQTL Browser is available at (https://www.genenetwork.nl/bloodeqtlbrowser/).

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102322.
ACKNOWLEDGMENTS
This study was supported by the project “CEINGE TASK-FORCE COVID19”, code D64i200003800 by Regione Campania for the fight against Covid-19 (DGR n. 140 del 17 Marzo 2020). This manuscript has been released as a pre-print at https://www.medrxiv.org/content/10.1101/2020.12.18.20248470v1.

The authors thank Roberta Campochiaro for her useful help in content management and analysis.

AUTHOR CONTRIBUTIONS
I.A., R.R., and M.C. designed and conducted the study, and prepared the manuscript; M.C., V.A.L., and F.B. analyzed the data; B.E.R. sampled genomic DNA from COVID-19 patients; S.C. genotyped COVID-19 patients outside the Intensive Care Units (Ward- COVID). Ann. Am. Thorac. Soc. https://doi.org/10.1513/AnnalsATS.202008-1080OC.

DECLARATION OF INTERESTS
The authors declare that there are no competing interests.

REFERENCES
Anastassopoulou, C., Gkizarioti, Z., Patentos, G.P., and Tsakris, A. (2020). Human genetic factors associated with susceptibility to SARS-CoV-2 infection and COVID-19 disease severity. Hum. Genomics 14, 40.

Asselsta, R., Paraboschi, E.M., Mantovani, A., and Duga, S. (2020). ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. Aging (Albany NY) 12, 10087–10098.

Beck, D.B., and Aksentijevich, I. (2020). Susceptibility to severe COVID-19. Science 370, 404–405.

Belloni, G., Grasselli, G., Cecconi, M., Antonelli, L., Borelli, M., De Giacomi, F., Bosio, G., Latronico, N., Filippini, M., Gemma, M., et al. (2021). Noninvasive ventilatory support of COVID-19 patients outside the Intensive Care Units (Ward-COVID). Ann. Am. Thorac. Soc. https://doi.org/10.1513/AnnalsATS.202008-1080OC.

Benetti, E., Tita, R., Spiga, O., Ciolfi, A., Birolo, G., Bruselles, A., Doddato, G., Giliberti, A., Marconi, C., Musachia, F., et al. (2020). ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. Eur. J. Hum. Genet. 28, 1602–1614.

Bizzotto, J., Sanchis, P., Abbate, M., Lague-Vickers, S., Lavignolle, R., Toro, A., Olsonvecki, S., Sabater, A., Cascardo, F., Vazquez, E., et al. (2020). SARS-CoV-2 infection boosts MX1 antiviral effector in COVID-19 patients. iScience 23, 101585.

Cancanelli, M.J., Abel, L., Zhang, S.Y., and Casanova, J.L. (2016). Host genetics of severe influenza: from mouse Mx1 to human IRF7. Curr. Opin. Immunol. 38, 109–120.

Colantonio, A.D., Epeldegui, M., Jesiak, M., Jachimowski, L., Blom, B., and Uittenbogaart, C.H. (2011). IFN-alpha is constitutively expressed in the human thymus, but not in peripheral lymphoid organs. PLoS One 6, e24252.

COVID-19 Host Genetics Initiative (2020). The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur. J. Hum. Genet. 28, 715–718.

Daga, S., Fallerini, C., Baldassarri, M., Fava, F., Valentino, F., Doddato, G., Benetti, E., Furini, S., Giliberti, A., Tita, R., et al. (2021). Employing a systematic approach to biobanking and analyzing clinical and genetic data for advancing COVID-19 research. Eur. J. Hum. Genet. 1–15.

Duncan, C.J., Mohamad, S.M., Young, D.F., Skelton, A.J., Leachy, T.R., Munday, D.C., Butler, K.M., Morfopoulou, S., Brown, J.R., Hubank, M., et al. (2019). Human IFNAR2 deficiency: Lessons for antiviral immunity. Sci. Transl Med. 7, 307ra154.

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernandez, J., Prati, D., Baselli, G., Asseffa, R., et al. (2020). Genomewide association study of severe Covid-19 with respiratory failure. N. Engl. J. Med. 383, 1522–1534.

Grasselli, G., Cattaneo, E., and Scaravilli, V. (2021). Ventilation of coronavirus disease 2019 patients. Curr. Opin. Crit. Care 27, 6–12.

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., Cerdedo, D., Coluccello, A., Foti, G., Fumagalli, R., et al. (2020). Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region. Italy. JAMA 323, 1574–1581.

Group, R.C., Horby, P., Lim, W.S., Emberson, J.R., Mafham, M., Bell, J.L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., et al. (2020). Dexamethasone in hospitalized patients with COVID-19 - preliminary report. N. Engl. J. Med. 25, 693–704.

Grundberg, E., Adoue, V., Kwan, T., Ge, B., Duan, Q.L., Lam, K.C., Koka, V., Kindmark, A., Weiss, S.T., Tantisira, K., et al. (2011). Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. Plos Genet. 7, e1001279.

Gunes, H., Dincer, S., Acipayam, C., Yurttutan, S., and Ozkars, M.Y. (2020). What chances do children have against COVID-19? Is the answer hidden within the thymus? Eur. J. Pediatrics. 180, 983–986.

Hoffmann, M., Kleine-Weber, H., and Pohlmann, S. (2020a). A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol. Cell 78, 779–784.e5.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020b). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181, 271–280.e8.

Hou, Y., Zhao, J., Martin, W., Kallianpur, A., Chung, M.K., Jehi, L., Shanf, N., Erzurum, S., Eng, C., and Cheng, F. (2020). New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. BMC Med. 18, 216.

Kellogg, C., and Equils, O. (2020). The role of the thymus in COVID-19 disease severity: implications for antibody treatment and immunization. Hum. Vaccon. Immunother. 17, 638–643.
Latini, A., Agolini, E., Novelli, A., Borgiani, P., Gianninni, R., Gravina, P., Smarrazzo, A., Dauri, M., Andreoni, M., Rogliani, P., et al. (2020). COVID-19 and genetic variants of protein involved in the SARS-CoV-2 entry into the host cells. Genes (Basel) 11, 1010.

Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. J. Virol. 84, 12658–12664.

Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., Nagata, N., Sekizuka, T., Kato, H., Kato, F., et al. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc. Natl. Acad. Sci. U S A 117, 7001–7003.

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2021). Genetic mechanisms of critical illness in Covid-19. Nature 591, 92–98.

Paniri, A., Hosseini, M.M., and Akhavan-Niaki, H. (2020). First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. J. Biomol. Struct. Dyn. 1–18.

Richardson, S., Hirsch, J.S., Narasimhan, M., Crawford, J.M., McGinn, T., Davidson, K.W., the Northwell, C.-R.C., Barnaby, D.P., Becker, L.B., Chelico, J.D., et al. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA 323, 2052–2059.

Russo, R., Andolfi, I., Lasorsa, V.A., Iolascon, A., and Capasso, M. (2020). Genetic analysis of the coronavirus SARS-CoV-2 host protease TMPRSS2 in different populations. Front. Genet. 11, 872.

Shelton, J.F., Shastri, A.J., Ye, C., Weldon, C.H., Filshiein-Sommez, T., Coker, D., Symons, A., Esparza-Gordillo, J., The 23andMe COVID-19 Team, Aslibekyan, S., et al. (2020). Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity. medRxiv.

Singh, H., Choudhari, R., Nema, V., and Khan, A.A. (2020). ACE2 and TMPRSS2 polymorphisms in various diseases with special reference to its impact on COVID-19 disease. Microb. Pathog. 150, 104621.

van der Made, C.I., Simons, A., Schuurs-Hoeijmakers, C.J., van den Heuvel, G., Mantere, T., Kersten, S., van Deuren, R.C., Steehouwer, M., van Reijnemersdal, S.V., Jaeger, M., et al. (2020). Presence of genetic variants among young men with severe COVID-19. JAMA 324, 1–11.

Vargas-ALARCON, G., POSADAS-SANCHEZ, R., and Ramirez-Bello, J. (2020). Variability in genes related to SARS-CoV-2 entry into host cells (ACE2, TMPRSS2, TMPRSS11A, ELANE, and CTS1) and its potential use in association studies. Life Sci. 260, 118313.

Vishnubhotla, R., Vankadari, N., Ketavarapu, V., Amanchy, R., Ananthi, S., Bale, G., Reddy, D.N., and Sasikala, M. (2020). Genetic variants in TMPRSS2 and Structure of SARS-CoV-2 spike glycoprotein and TMPRSS2 complex. bioRxiv.

Wang, F., Huang, S., Gao, R., Zhou, Y., Lai, C., Li, Z., Xian, W., Qian, X., Li, Z., Huang, Y., et al. (2020). Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. Cell Discov. 6, 83.

Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243.

Zav’yalov, V.P., Hamalainen-Laanaya, H., Korpela, T.K., and Wahroos, T. (2019). Interferon-inducible myxovirus resistance proteins: potential biomarkers for differentiating viral from bacterial infections. Clin. Chem. 65, 739–750.

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I.K.D., Hodeib, S., Korol, C., et al. (2020). Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science 370, eabd4570.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 395, 1054–1062.
Supplemental information

Common variants at 21q22.3 locus influence

*MX1* and *TMPRSS2* gene expression

and susceptibility to severe COVID-19

Immacolata Andolfo, Roberta Russo, Vito Alessandro Lasorsa, Sueva Cantalupo, Barbara Eleni Rosato, Ferdinando Bonfiglio, Giulia Frisso, Pasquale Abete, Gian Marco Cassese, Giuseppe Servillo, Gabriella Esposito, Ivan Gentile, Carmelo Piscopo, Romolo Villani, Giuseppe Fiorentino, Pellegrino Cerino, Carlo Buonerba, Biancamaria Pierri, Massimo Zollo, Achille Iolascon, and Mario Capasso
Supplemental Information

Article title:
Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19

Authors:
Immacolata Andolfo, Roberta Russo, Alessandro Vito Lasorsa, Sueva Cantalupo, Barbara Eleni Rosato, Ferdinando Bonfiglio, Giulia Friso, Abete Pasquale, Gian Marco Cassese, Giuseppe Servillo, Gabriella Esposito, Ivan Gentile, Carmelo Piscopo, Romolo Villani, Giuseppe Fiorentino, Pellegrino Cerino, Carlo Buonerba, Biancamaria Pierri, Massimo Zollo, Achille Iolascon, Mario Capasso

1Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Napoli, Italy
2CEINGE Biotecnologie Avanzate, Napoli, Italy
3Dipartimento di Ingegneria chimica, dei Materiali e della Produzione industriale, Università degli Studi di Napoli Federico II, Napoli, Italy
4COVID Hospital, P.O.S. Anna e SS. Madonna della Neve di Boscotrecase, Ospedali Riuniti Area Vesuviana, Napoli, Italy
5Dipartimento di Neuroscienze e Scienze riproduttive ed odontostomatologiche, Università degli Studi di Napoli Federico II, Napoli, Italy
6Dipartimento di Medicina clinica e Chirurgia, Università degli Studi di Napoli Federico II, Napoli, Italy
7Medical and Laboratory Genetics Unit, A.O.R.N. ‘Antonio Cardarelli’, Napoli, Italy
8Poison Centre, A.O.R.N. ‘Antonio Cardarelli’, Napoli, Italy
9AORN dei Colli Presidio Ospedaliero Cotugno, Napoli, Italy
10Istituto Zooprofilattico Sperimentale del Mezzogiorno, Napoli, Italy
11Dipartimento di Medicina, Chirurgia e Odontoiatria "Scuola Medica Salernitana", Università di Salerno, Baronissi, Italy

Corresponding author and lead contact
Mario Capasso, Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Napoli, Italy; CEINGE Biotecnologie Avanzate, Napoli, Italy; mario.capasso@unina.it
Figure S1. Linkage disequilibrium block at **TMPRSS2/MXI** locus. Related to Figure 1.
Linkage disequilibrium of the 5 most significant SNPs (P-values ranged from $2.7 \times 10^{-6}$ to $5.8 \times 10^{-6}$) with the lead rs3787946 at **TMPRSS2/MXI** locus. The D’ and $r^2$ data are computed with the genetic information from the European population by using the web tool LD-link (https://ldlink.nci.nih.gov/?tab=home).
Figure S2. Analysis of the eQTL signals of the top four disease-associated SNPs in LD with the lead SNP rs3787946. Related to Figure 2.

Violin plots showing the eQTL signals for the rs9983330, rs12329760, rs2298661, and rs9985159 on MX1 expression in whole blood (a) and on TMPRSS2 expression in lung (b). The significance threshold adjusted for multiple comparisons is 0.000055.
### Table S1. Study groups that have contributed to GWAS meta-analyses of the COVID-19 Host Genetics Initiative. Related to Figure 1.

| Name                                      | n_cases | n_controls     |
|-------------------------------------------|---------|----------------|
| Amsterdam_UMC_COVID_study_group_EUR       | 108     | 1413           |
| DECODE_EUR                                | 89      | 274322         |
| BelCovid_EUR                              | 109     | 1484           |
| GENCovid_EUR                              | 571     | 2472           |
| FinnGen_FIN                               | 83      | 238628         |
| SPGRX_EUR                                 | 311     | 302            |
| HOSTAGE_EUR                               | 1610    | 2205           |
| BQC19_EUR                                 | 181     | 354            |
| UKBB_EUR                                  | 765     | 364341         |
| MVP_EUR                                   | 436     | 2180           |
| BoSCO_EUR                                 | 139     | 262            |
| Ancestry_EUR                              | 250     | 1967           |
| SweCovid_EUR                              | 78      | 3778           |
| genomicc_EUR                              | 1676    | 8380           |
|                                           | 6406    | 902088         |

*EUR: individuals have European origins*

*FIN: individuals with Finnish origins*
Table S2. Summary statistics at chromosome 21 from GWAS dataset (COVID-19 Host Genetics Initiative, "B2_ALL_vf_3r_23andme"). Related to Figure 1.

| CHR | POS   | REF | ALT | SNP               | all_meta_N | all_inv_var_meta_beta | all_inv_var_meta_s | all_inv_var_meta_s_p | all_inv_var_meta_s_p | all_meta_sample_N | all_meta_AF   | rsid   | OR   | CI95_L | CI95_U |
|-----|-------|-----|-----|-------------------|------------|-----------------------|-------------------|----------------------|---------------------|-------------------|-------------|--------|------|--------|--------|
| 21  | 346152| T   | C   | 21:346152:10      | T:C        | -1.82E-01             | 2.60E-02         | 2.76E-12             | 3.46E-02           | 905878            | 6.56E-01     | rs10050728  | 0.83 | 0.79 | 0.87   |
| 21  | 346148| G   | A   | 21:34614834:     | G:A        | -1.80E-01             | 2.60E-02         | 3.53E-02             | 905878            | 6.56E-01     | rs997682     | 0.83 | 0.79 | 0.87   |
| 21  | 346177| A   | G   | 21:34617729:     | A:G        | -1.80E-01             | 3.46E-02         | 4.03E-02             | 905878            | 6.75E-01     | rs225263     | 1.11 | 1.11 | 1.22   |
| 21  | 346204| A   | G   | 21:34620451:     | C:T        | -1.80E-01             | 2.60E-02         | 3.46E-02             | 905878            | 6.57E-01     | rs283416     | 0.83 | 0.79 | 0.87   |
| 21  | 346194| A   | G   | 21:34619445:     | A:G        | -1.80E-01             | 2.60E-02         | 3.53E-02             | 905878            | 6.57E-01     | rs283416     | 0.83 | 0.79 | 0.87   |
| 21  | 346202| C   | T   | 21:34620207:     | C:T        | -1.79E-01             | 2.60E-02         | 3.15E-02             | 905878            | 6.57E-01     | rs207336     | 1.11 | 1.11 | 1.23   |
| 21  | 346169| C   | A   | 21:34616923:     | C:A        | -1.79E-01             | 2.60E-02         | 5.90E-12             | 905878            | 6.56E-01     | rs225265     | 1.11 | 1.11 | 1.23   |
| 21  | 346179| G   | A   | 21:34617950:     | A:T        | -1.79E-01             | 2.60E-02         | 5.02E-12             | 905878            | 6.57E-01     | rs223675     | 0.83 | 0.79 | 0.88   |
| 21  | 346249| A   | G   | 21:34624917:     | G:T        | -1.79E-01             | 2.60E-02         | 3.69E-11             | 905878            | 6.93E-01     | rs228454     | 0.83 | 0.79 | 0.88   |
| 21  | 346180| A   | T   | 21:34618043:     | A:T        | -1.77E-01             | 2.66E-02         | 3.41E-12             | 905878            | 6.93E-01     | rs228455     | 0.83 | 0.79 | 0.88   |
| 21  | 346183| A   | G   | 21:34618331:     | A:G        | -1.76E-01             | 2.66E-02         | 3.53E-11             | 905878            | 6.93E-01     | rs283415     | 0.83 | 0.79 | 0.88   |
| 21  | 346165| A   | G   | 21:34616545:     | A:G        | -1.75E-01             | 2.66E-02         | 4.50E-11             | 905878            | 6.93E-01     | rs283415     | 0.83 | 0.79 | 0.88   |
| 21  | 346066| A   | C   | 21:34606634:     | A:C        | -1.65E-01             | 2.48E-02         | 3.25E-10             | 908494            | 3.43E-01     | rs963686     | 1.17 | 1.11 | 1.23   |
| 21  | 346099| A   | G   | 21:34609944:     | A:C        | -1.57E-01             | 2.33E-02         | 3.16E-10             | 908494            | 3.43E-01     | rs651715     | 1.17 | 1.11 | 1.23   |
| 21  | 346074| G   | A   | 21:34607436:     | G:A        | -1.65E-01             | 2.60E-02         | 3.15E-10             | 908494            | 3.43E-01     | rs17601680  | 1.17 | 1.11 | 1.23   |
| 21  | 346133| A   | G   | 21:34613301:     | A:G        | -1.57E-01             | 2.47E-02         | 4.39E-10             | 908494            | 3.43E-01     | 1.17 | 1.11 | 1.23   |
| 21  | 346115| C   | G   | 21:34611571:     | C:G        | -1.58E-01             | 2.50E-02         | 5.65E-02             | 634083            | 3.37E-01     | NA           | 1.17 | 1.11 | 1.23   |
| 21  | 346032| C   | G   | 21:34603249:     | C:G        | -1.64E-01             | 2.60E-02         | 5.72E-10             | 905878            | 3.43E-01     | NA           | 1.17 | 1.11 | 1.23   |
| 21  | 346045| G   | A   | 21:34604557:     | G:A        | -1.64E-01             | 2.60E-02         | 6.01E-10             | 905878            | 3.43E-01     | rs230037     | 1.17 | 1.11 | 1.23   |
| 21  | 346029| T   | C   | 21:34602934:     | T:C        | -1.64E-01             | 2.60E-02         | 5.30E-10             | 905878            | 3.43E-01     | rs124825     | 1.17 | 1.11 | 1.23   |
| 21  | 346023| C   | A   | 21:34602305:     | C:A        | -1.64E-01             | 2.60E-02         | 8.97E-10             | 905878            | 3.43E-01     | NA           | 1.17 | 1.11 | 1.23   |
| 21  | 346239| A   | G   | 21:34623919:     | A:G        | -1.64E-01             | 2.60E-02         | 8.97E-10             | 905878            | 3.43E-01     | rs176020     | 1.29 | 1.19 | 1.40   |
| 21  | 346142| T   | C   | 21:34614250:     | T:C        | -1.53E-01             | 4.08E-02         | 5.19E-10             | 908494            | 7.77E-02     | rs222920     | 1.28 | 1.18 | 1.39   |
| rsID       | Gene | Chromosome | Position | Effect | Odds Ratio (95% CI) |
|------------|------|------------|----------|--------|--------------------|
| rs207336   |      | 21         | 3462081  | A       | 1.13 (1.08, 1.18)  |
| rs178601   |      | 21         | 34614255 | T       | 1.17 (1.12, 1.23)  |
| rs224842   |      | 21         | 34605778 | G       | 1.17 (1.11, 1.23)  |
| rs120536   |      | 21         | 34618439 | A       | 1.17 (1.10, 1.22)  |
| rs113196   |      | 21         | 34611730 | T       | 1.17 (1.10, 1.23)  |
| rs622661   |      | 21         | 34593710 | T       | 1.17 (1.10, 1.23)  |
| rs520020   |      | 21         | 34599084 | G       | 1.17 (1.10, 1.23)  |
| rs124820   |      | 21         | 34605008 | G       | 1.17 (1.10, 1.23)  |
| rs124821   |      | 21         | 34611318 | C       | 1.17 (1.10, 1.23)  |
| rs113196   |      | 21         | 34611545 | T       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34609505 | A       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34601487 | T       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34622536 | A       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34621948 | A       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34618285 | G       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34611992 | A       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34605974 | C       | 1.17 (1.10, 1.23)  |
| rs1283416  |      | 21         | 34602246 | T       | 1.17 (1.10, 1.23)  |
| rs147641   |      | 21         | 34693885 | A       | 1.17 (1.10, 1.23)  |
| rs19095    |      | 21         | 34609596 | T       | 1.17 (1.10, 1.23)  |
| rs147641   |      | 21         | 34617213 | T       | 1.17 (1.10, 1.23)  |
| rs997553   |      | 21         | 34626854 | C       | 1.17 (1.10, 1.23)  |
| rs111783   |      | 21         | 35362848 | A       | 1.17 (1.10, 1.23)  |
| rs111783   |      | 21         | 3549337  | A       | 1.17 (1.10, 1.23)  |
| # | G/C | 21:42847735:G/C | 14 | -1.34E-01 | 2.86E-02 | 2.73E-06 | 8.59E-01 | 908494 | 2.64E-01 | rs378794 | 0.87 | 0.82 | 0.92 | 0.94 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 21 | 353637 | T/G | T/C:G | 14 | 1.55E-01 | 3.30E-02 | 2.79E-06 | 2.15E-01 | 908494 | 1.33E-01 | rs110882 | 68 | 1.16 | 1.09 | 1.24 |
| 21 | 346254 | A/G | A/G:G | 12 | -1.49E-01 | 3.19E-02 | 2.90E-06 | 5.10E-02 | 895822 | 6.64E-01 | rs223675 | 8 | 0.86 | 0.80 | 0.91 |
| 21 | 346323 | T/C | T/C:G | 13 | 1.40E-01 | 2.99E-02 | 2.90E-06 | 1.71E-01 | 898438 | 3.36E-01 | rs225022 | 6 | 1.15 | 1.08 | 1.21 |
| 21 | 428502 | A/G | A/G:G | 14 | -1.32E-01 | 2.84E-02 | 3.12E-06 | 9.08E-01 | 908494 | 2.76E-01 | rs998333 | 6 | 0.87 | 0.82 | 0.92 |
| 21 | 428524 | C/T | C/T:G | 14 | -1.32E-01 | 2.83E-02 | 3.13E-06 | 8.87E-01 | 908494 | 2.75E-01 | rs123297 | 17 | 2.08 | 1.44 | 2.73 |
| 21 | 385771 | A/C | A/C:G | 8 | 7.37E-01 | 1.58E-01 | 3.29E-06 | 2.81E-01 | 380965 | 1.53E-02 | rs563091 | 1 | 1.17 | 1.09 | 1.25 |
| 21 | 353822 | G/A | G/A:T | 14 | 1.59E-01 | 3.44E-02 | 3.85E-06 | 2.81E-01 | 908494 | 1.26E-01 | NA | 2 | 0.88 | 0.83 | 0.93 |
| 21 | 428640 | C/T | C/T:G | 14 | -1.23E-01 | 2.67E-02 | 4.25E-06 | 8.86E-01 | 908494 | 3.23E-01 | rs930574 | 5 | 1.14 | 1.08 | 1.21 |
| 21 | 345924 | T/C | T/C:G | 13 | 1.37E-01 | 2.98E-02 | 4.40E-06 | 2.11E-01 | 898438 | 3.36E-01 | rs112268 | 2 | 1.30 | 1.15 | 1.44 |
| 21 | 428456 | C/A | C/A:G | 14 | -1.32E-01 | 2.87E-02 | 4.51E-06 | 8.81E-01 | 908494 | 2.64E-01 | NA | 7 | 0.87 | 0.82 | 0.92 |
| 21 | 343111 | A/G | A/G:T | 13 | 1.36E-01 | 2.97E-02 | 4.62E-06 | 2.02E-01 | 898438 | 3.36E-01 | rs178602 | 41 | 1.14 | 1.07 | 1.21 |
| 21 | 342637 | T/A | T/A:G | 14 | -1.22E-01 | 2.67E-02 | 4.76E-06 | 9.23E-01 | 908494 | 3.22E-01 | rs101540 | 59 | 0.88 | 0.83 | 0.93 |
| 21 | 346340 | C/G | C/G:G | 12 | 1.47E-01 | 3.23E-02 | 4.94E-06 | 1.04E-01 | 895822 | 3.34E-01 | rs651715 | 5 | 1.15 | 1.08 | 1.23 |
| 21 | 428573 | C/G | C/G:G | 14 | -1.26E-01 | 2.75E-02 | 5.00E-06 | 9.18E-01 | 908494 | 3.10E-01 | rs998325 | 2 | 0.88 | 0.83 | 0.92 |
| 21 | 345025 | G/A | G/A:T | 14 | 2.60E-01 | 5.71E-02 | 5.25E-06 | 7.12E-01 | 908494 | 4.02E-02 | rs799978 | 10 | 1.29 | 1.15 | 1.44 |
| 21 | 353684 | G/T | G/T:T | 13 | 1.52E-01 | 3.35E-02 | 5.25E-06 | 1.42E-01 | 634083 | 1.25E-01 | rs126272 | 54 | 1.16 | 1.08 | 1.24 |
| 21 | 353659 | G/T | G/T:T | 13 | 1.48E-01 | 3.27E-02 | 5.57E-06 | 9.49E-02 | 621411 | 6.72E-01 | rs117014 | 54 | 0.86 | 0.80 | 0.91 |
| 21 | 428468 | C/T | C/T:G | 14 | -1.29E-01 | 2.85E-02 | 5.80E-06 | 9.14E-01 | 908494 | 2.64E-01 | NA | 2 | 0.87 | 0.82 | 0.92 |
| 21 | 353893 | A/G | A/G:T | 14 | 1.55E-01 | 3.43E-02 | 5.83E-06 | 3.00E-01 | 908494 | 1.27E-01 | rs110882 | 69 | 1.16 | 1.09 | 1.24 |
| 21 | 353954 | G/C | G/C:T | 14 | 1.55E-01 | 3.45E-02 | 7.06E-06 | 2.50E-01 | 908494 | 1.24E-01 | rs117024 | 97 | 8 | 1.16 | 1.08 | 1.24 |
| 21 | 428565 | T/C | T/C:G | 14 | -1.23E-01 | 2.75E-02 | 7.80E-06 | 9.20E-01 | 908494 | 3.11E-01 | rs283803 | 9 | 0.88 | 0.83 | 0.93 |
Table S3. Characteristics of Italian patients recruited by our research group. Related to Table 1.

| Characteristic                        | Severe cases | %     |
|---------------------------------------|--------------|-------|
|                                       | N=226        |       |
| **Age**                               |              |       |
| Years, mean (standard deviation)      | 62.3         | (16.6)|
| Unknown                               | 5            |       |
| **Sex - no. (%)**                     |              |       |
| Male                                  | 142          | 62.8  |
| Female                                | 74           | 32.8  |
| Unknown                               | 10           | 4.4   |
| **Previous coexisting disease - no. (%)** |          |       |
| 0-2                                   | 136          | 60.2  |
| >=3                                   | 41           | 18.1  |
| Unknown                               | 49           | 21.7  |
| **Oxigen Therapy**                    |              |       |
| No Mechanical ventilation or Intubation | 105          | 46.5  |
| Mechanical ventilation or Intubation  | 81           | 35.8  |
| Unknown                               | 40           | 17.7  |
### Table S4. Results of SNP enrichment analysis in regulatory elements in different tissues and cell types.
Related to Figure 2.

| Cell                                           | Observed | Expected | Fold | Binomial p | ^adjusted_p |
|------------------------------------------------|----------|----------|------|------------|-------------|
| E112 THYM (Thymus)                             | 6        | 0.2      | 30.0 | 0          | 0           |
| E021 IPSC.DF.6.9 (iPS DF 6.9 Cells)            | 6        | 0.2      | 30.0 | 0          | 0           |
| E012 ESDR.CD56.ECTO (hESC Derived CD56+ Ectoderm Cultured Cells) | 8        | 0.4      | 20.0 | 0          | 0           |
| E054 BRN.GANGEM.DR.NRSPHR (Ganglion Eminence derived primary cultured neurospheres) | 8        | 0.4      | 20.0 | 0          | 0           |
| E099 PLCNT.AMN (Placenta Amnion)               | 6        | 0.3      | 20.0 | 0          | 0           |
| E115 BLD.DND41.CNCR (Dnd41 TCell Leukemia Cell Line) | 6        | 0.3      | 20.0 | 0          | 0           |
| E121 MUS.HSMMT (HSMM cell derived Skeletal Muscle Myotubes Cells) | 8        | 0.4      | 20.0 | 0          | 0           |
| E024 ESC.4STAR (ES-UCSF4 Cells)                | 9        | 0.5      | 18.0 | 0          | 0           |
| E014 ESC.HUES48 (HUES48 Cells)                 | 8        | 0.5      | 16.0 | 0          | 0           |
| E003 ESC.H1 (H1 Cells)                         | 8        | 0.5      | 16.0 | 0          | 0           |
| E018 IPSC.15b (iPS-15b Cells)                  | 8        | 0.5      | 16.0 | 0          | 0           |
| E022 IPSC.DF.19.11 (iPS DF 19.11 Cells)        | 8        | 0.5      | 16.0 | 0          | 0           |
| E027 BRST.MYO (Breast Myoepithelial Primary Cells) | 11       | 0.7      | 15.7 | 0          | 0           |
| E120 MUS.HSM (HSMM Skeletal Muscle Myoblasts Cells) | 6        | 0.4      | 15.0 | 3.00E-06   | 0.000381    |
| E008 ESC.H9 (H9 Cells)                         | 3        | 0.2      | 15.0 | 0.001534   | 0.194818    |
| E016 ESC.HUES64 (HUES64 Cells)                 | 7        | 0.5      | 14.0 | 0          | 0           |
| E061 SKIN.PEN.FRSK.MEL.03 (Foreskin Melanocyte Primary Cells skin03) | 8        | 0.6      | 13.3 | 0          | 0           |
| E020 IPSC.20B (iPS-20b Cells)                  | 5        | 0.4      | 12.5 | 3.30E-05   | 0.0042      |
| E011 ESDR.CD184.ENDO (hESC Derived CD184+ Endoderm Cultured Cells) | 5        | 0.4      | 12.5 | 5.20E-05   | 0.0066      |
| E019 IPSC.18 (iPS-18 Cells)                    | 6        | 0.5      | 12.0 | 6.00E-06   | 0.0008      |
| E093 THYM.FET (Fetal Thymus)                   | 6        | 0.5      | 12.0 | 6.00E-06   | 0.0008      |
| E015 ESC.HUES6 (HUES6 Cells)                   | 6        | 0.6      | 10.0 | 9.00E-06   | 0.0011      |
| E077 GI.DUO.MUC (Duodenum Mucosa)              | 4        | 0.4      | 10.0 | 0.000381   | 0.0484      |
| E098 PANC (Pancreas)                           | 4        | 0.4      | 10.0 | 0.000738   | 0.0937      |
| E094 GI.STMC.GAST (Gastric)                    | 3        | 0.3      | 10.0 | 0.002056   | 0.2611      |
| E007 ESDR.H1.NEUR.PROG (H1 Derived Neuronal Progenitor Cultured Cells) | 3        | 0.3      | 10.0 | 0.002349   | 0.2983      |
| E075 GI.CLN.MUC (Colonic Mucosa)               | 2        | 0.2      | 10.0 | 0.01355    | 1.7209      |
| E101 GI.RECT.MUC.29 (Rectal Mucosa Donor 29)   | 2        | 0.2      | 10.0 | 0.021315   | 2.7070      |
| E118 LIV.HEPG2.CNCR (HepG2 Hepatocellular Carcinoma Cell Line) | 6        | 0.7      | 8.6  | 3.20E-05   | 0.0041      |
| E074 BRN.SUB.NIG (Brain Substantia Nigra)      | 3        | 0.4      | 7.5  | 0.008348   | 1.0602      |
| E059 SKIN.PEN.FRSK.MEL.01 (Foreskin Melanocyte Primary Cells skin01) | 2        | 0.3      | 6.7  | 0.044405   | 5.6394      |
| E090 MUS.LEG.FET (Fetal Muscle Leg)            | 5        | 0.8      | 6.3  | 0.000715   | 0.0908      |
| E071 BRN.HIPP.MID (Brain Hippocampus Middle)   | 3        | 0.5      | 6.0  | 0.011249   | 1.4286      |
| Code     | Description                                                                 | C1 | C2 | C3    | C4          | C5          |
|----------|-----------------------------------------------------------------------------|----|----|-------|-------------|-------------|
| E053     | BRN.CRTX.DR.NRSPHR (Cortex derived primary cultured neurospheres)           | 3  | 0.5| 6.0   | 0.014938    | 1.8971      |
| E001     | ESC.I3 (ES-I3 Cells)                                                         | 3  | 0.5| 6.0   | 0.014978    | 1.9022      |
| E088     | LNG.FET (Fetal Lung)                                                        | 3  | 0.6| 5.0   | 0.016702    | 2.1212      |
| E102     | GI.RECT.MUC.31 (Rectal Mucosa Donor 31)                                      | 2  | 0.4| 5.0   | 0.049562    | 6.2944      |
| E013     | ESDR.CD56.MESO (hESC Derived CD56+ Mesoderm Cultured Cells)                  | 2  | 0.4| 5.0   | 0.071638    | 9.0980      |
| E002     | ESC.WA7 (ES-WA7 Cells)                                                       | 1  | 0.2| 5.0   | 0.175795    | 22.3260     |
| E109     | GI.S.INT (Small Intestine)                                                   | 1  | 0.3| 4.0   | 0.099955    | 12.6943     |
| E089     | MUS.TRNK.FET (Fetal Muscle Trunk)                                            | 2  | 0.6| 3.3   | 0.119051    | 15.1195     |
| E070     | BRN.GRM.MTRX (Brain Germinal Matrix)                                         | 1  | 0.3| 3.3   | 0.28264     | 35.8953     |
| E072     | BRN.INF.TMP (Brain Inferior Temporal Lobe)                                   | 1  | 0.4| 2.5   | 0.322209    | 40.9205     |
| E068     | BRN.ANT.CAUD (Brain Anterior Caudate)                                        | 1  | 0.4| 2.5   | 0.335088    | 42.5562     |
| E069     | BRN.CING.GYR (Brain Cingulate Gyrus)                                         | 1  | 0.4| 2.5   | 0.337134    | 42.8160     |
| E116     | BLD.GM12878 (GM12878 Lymphoblastoid Cells)                                   | 1  | 0.4| 2.5   | 0.343155    | 43.5807     |
| E026     | STRM.MRW.MSC (Bone Marrow Derived Cultured Mesenchymal Stem Cells)           | 1  | 0.5| 2.0   | 0.394234    | 50.0677     |
| E005     | ESDR.H1.BMP4.TROP (H1 BMP4 Derived Trophoblast Cultured Cells)                | 1  | 0.5| 2.0   | 0.40053     | 50.8673     |
| E084     | GI.INT.FET (Fetal Intestine Large)                                           | 1  | 0.5| 2.0   | 0.404646    | 51.3900     |
| E129     | BONE.OSTEO (Osteoblast Primary Cells)                                        | 1  | 0.5| 2.0   | 0.41587     | 52.8155     |
| E085     | GI.INT.FET (Fetal Intestine Small)                                           | 1  | 0.5| 2.0   | 0.418745    | 53.1806     |
| E057     | SKIN.PEN.FRSK.KER.02 (Foreskin Keratinocyte Primary Cells skin02)             | 1  | 0.5| 2.0   | 0.425214    | 54.0022     |
| E006     | ESDR.H1.MSC (H1 Derived Mesenchymal Stem Cells)                              | 1  | 0.5| 2.0   | 0.426514    | 54.1673     |
| E119     | BRST.HMMEC (HMEC Mammary Mammary Primary Cells)                              | 1  | 0.6| 1.7   | 0.449466    | 57.0822     |
| E028     | BRST.HMEC.35 (Breast variant Human Mammary Epithelial Cells (vHMEC))         | 1  | 0.6| 1.7   | 0.458849    | 58.2738     |
| E091     | PLCNT.FET (Placenta)                                                        | 1  | 0.6| 1.7   | 0.481212    | 61.1139     |
| E017     | LNG.1M90 (IMR90 fetal lung fibroblasts Cell Line)                            | 0  | 0.6| 0.0   | 1           | 1           |
| E009     | ESDR.H9.NEUR.PROG (H9 Derived Neuronal Progenitor Cultured Cells)            | 0  | 0.4| 0.0   | 1           | 1           |
| E100     | ESDR.H9.NEUR (H9 Derived Neuron Cultured Cells)                              | 0  | 0.5| 0.0   | 1           | 1           |
| E004     | ESDR.H1.BMP4.MESO (H1 BMP4 Derived Mesoderm Cultured Cells)                   | 0  | 0.3| 0.0   | 1           | 1           |
| E062     | BLD.PER.MONUC.PC (Primary mononuclear cells from peripheral blood)           | 0  | 0.2| 0.0   | 1           | 1           |
| E034     | BLD.CD3.PPC (Primary T cells from peripheral blood)                          | 0  | 0.5| 0.0   | 1           | 1           |
| E045     | BLD.CD4.CD25I.CD127.TMEMPC (Primary T cells effector/memory enriched from peripheral blood) | 0  | 0.2| 0.0   | 1           | 1           |
| E033     | BLD.CD3.CPC (Primary T cells from cord blood)                                | 0  | 0.3| 0.0   | 1           | 1           |
| E044     | BLD.CD4.CD25.CD127M.TREGPC (Primary T regulatory cells from peripheral blood) | 0  | 0.3| 0.0   | 1           | 1           |
| E043     | BLD.CD4.CD25M.TPC (Primary T helper cells from peripheral blood)             | 0  | 0.5| 0.0   | 1           | 1           |
| E039     | BLD.CD4.CD25M.CD45RA.NPC (Primary T helper naive cells from peripheral blood) | 0  | 0.4| 0.0   | 1           | 1           |
| Code     | Description                                      | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 | Value 6 |
|----------|--------------------------------------------------|---------|---------|---------|---------|---------|---------|
| E078 GIDU.O.SM.MUS | (Duodenum Smooth Muscle)                          | 0       | 0.3     | 0.0     | 1       | 1       |
| E076 GI.CLN.SM.MUS | (Colon Smooth Muscle)                              | 0       | 0.4     | 0.0     | 1       | 1       |
| E103 GI.RECT.SM.MUS | (Rectal Smooth Muscle)                            | 0       | 0.3     | 0.0     | 1       | 1       |
| E111 GI.STMC.MUS | (Stomach Smooth Muscle)                           | 0       | 0.3     | 0.0     | 1       | 1       |
| E092 GI.STMC.FET | (Fetal Stomach)                                    | 0       | 0.5     | 0.0     | 1       | 1       |
| E106 GI.CLN.SIG | (Sigmoid Colon)                                    | 0       | 0.3     | 0.0     | 1       | 1       |
| E079 GI.ESO | (Esophagus)                                       | 0       | 0.3     | 0.0     | 1       | 1       |
| E086 KID.FET | (Fetal Kidney)                                    | 0       | 0.2     | 0.0     | 1       | 1       |
| E097 OVRY | (Ovary)                                           | 0       | 0.4     | 0.0     | 1       | 1       |
| E087 PANC.ISLT | (Pancreatic Islets)                               | 0       | 0.2     | 0.0     | 1       | 1       |
| E080 ADRL.GLND.FET | (Fetal Adrenal Gland)                              | 0       | 0.7     | 0.0     | 1       | 1       |
| E096 LNG | (Lung)                                            | 0       | 0.3     | 0.0     | 1       | 1       |
| E113 SPLN | (Spleen)                                          | 0       | 0.4     | 0.0     | 1       | 1       |
| E114 LNG.A549.ETOH002.CNCR | (A549 EtOH 0.02pct Lung Carcinoma Cell Line) | 0       | 0.4     | 0.0     | 1       | 1       |
| E117 CRVX.HELAS3.CNCR | (HeLa-S3 Cervical Carcinoma Cell Line)  | 0       | 0.4     | 0.0     | 1       | 1       |
| E122 VAS.HUVEC | (HUVEC Umbilical Vein Endothelial Primary Cells) | 0       | 0.5     | 0.0     | 1       | 1       |
| E123 BLD.K562.CNCR | (K562 Leukemia Cells)                             | 0       | 0.4     | 0.0     | 1       | 1       |
| E124 BLD.CD14.MONO | (Monocytes-CD14+ RO01746 Primary Cells)           | 0       | 0.4     | 0.0     | 1       | 1       |
| E125 BRN.NHA | (NH-A Astrocytes Primary Cells)                   | 0       | 0.4     | 0.0     | 1       | 1       |
| E126 SKIN.NHDFAD | (NHDF-Ad Adult Dermal Fibroblast Primary Cells)   | 0       | 0.6     | 0.0     | 1       | 1       |
| E127 SKIN.NHEK | (NHEK-Epidermal Keratinocyte Primary Cells)       | 0       | 0.5     | 0.0     | 1       | 1       |
| E128 LNG.NHLF | (NHLF Lung Fibroblast Primary Cells)              | 0       | 0.4     | 0.0     | 1       | 1       |

*P*-values corrected according to Bonferroni method
Table S5. Results of eQTL analysis for the common variants at **TMPRSS2/MX1** locus. Related to Figure 2.

| Gene | SNP      | GWAS_P   | eQTL_P  | eQTL_P Threshold | *Statistically significant eQTL | NES | T-statistic | Tissue       |
|------|----------|----------|---------|------------------|-------------------------------|-----|-------------|--------------|
| MX1  | rs3787946| 2.73E-06 | 0.0000011 | 0.000064         | YES                           | 0.17| 4.9         | Whole Blood  |
| MX1  | rs12329760| 3.13E-06 | 0.0000021 | 0.000064         | YES                           | 0.17| 4.8         | Whole Blood  |
| MX1  | rs2298661| 4.51E-06 | 0.0000022 | 0.000064         | YES                           | 0.17| 4.8         | Whole Blood  |
| MX1  | rs9983330| 3.12E-06 | 0.0000036 | 0.000064         | YES                           | 0.16| 4.7         | Whole Blood  |
| MX1  | rs2298660| 6.28E-04 | 0.0000140 | 0.000064         | YES                           | 0.15| 4.4         | Whole Blood  |
| MX1  | rs9985159| 5.80E-06 | 0.0000190 | 0.000064         | YES                           | 0.15| 4.3         | Whole Blood  |
| MX1  | rs2094881| 5.17E-03 | 0.0000660 | 0.000064         | 0                             |     | -4          | Whole Blood  |
| MX1  | rs7364088| 2.27E-03 | 0.0000760 | 0.000064         | 0                             | 0.13| 4           | Whole Blood  |
| MX1  | rs8131648| 3.58E-02 | 0.0001100 | 0.000064         | 0                             | 0.13| -3.9        | Whole Blood  |
| MX1  | rs8131649| 6.55E-03 | 0.0001100 | 0.000064         | 0                             | 0.13| -3.9        | Whole Blood  |
| MX1  | rs8134216| 7.14E-03 | 0.0001300 | 0.000064         | 0                             | 0.13| -3.9        | Whole Blood  |
| MX1  | rs8134203| 7.10E-03 | 0.0001400 | 0.000064         | 0                             | 0.13| -3.8        | Whole Blood  |
| MX1  | rs2298663| 4.65E-03 | 0.0001600 | 0.000064         | 0                             | 0.13| -3.8        | Whole Blood  |
| MX1  | rs2104810| 7.86E-03 | 0.0007000 | 0.000064         | 0                             | 0.12| -3.4        | Whole Blood  |

*Only SNPs with corrected P are considered statistically significant eQTLs*
Transparent methods supplemental file

Phenotype definition
Patients with severe COVID-19: laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based), hospitalization due to coronavirus-related symptoms.
Controls: Individuals from the general population not notified as cases.

GWAS
The summary statistics, P-value, odds ratio (OR), and 95% confidence interval (CI), of chromosome 21 were obtained from the GWAS dataset “B2_ALL_eur_leave_23andme” deposited in the COVID-19 Host Genetics Initiative website (COVID-19 Host Genetics Initiative, 2020). It includes 6,406 laboratory-confirmed SARS-CoV-2 infections and hospitalized for COVID-19 cases and 902,088 controls from the general population with European genetic ancestry (Table S2). Manhattan plot and QQ plot of the results from this large GWAS are available at the website (https://www.covid19hg.org/results/).

Replication
The summary statistics of the SNPs used for the replication study were retrieved from the GenOMICC study (Pairo-Castineira et al., 2020). Three independent cohorts of cases and controls with different ethnicity were available throughout GenOMMIC GWAS study (Pairo-Castineira et al., 2020): 182 individuals from African ancestry, 149 of East Asian ancestry (EAS), 237 of South-Asian ancestry (SAS). Moreover, 226 hospitalized COVID-19 cases and 1848 controls (Table S3) enrolled from public hospitals located in Campania (Southern Italy) were typed fort the rs12329760 variant by TaqMan® SNP Genotyping (Applied Biosystems by Thermo Fisher Scientific). We selected rs12329760 SNP as it appears to has the most relevant functional role among the others, indeed, it is predicted to damage TMPRSS2 protein and to be an eQTL for TMPRSS2 in osteoblasts treated with dexamethasone (Grundberg et al., 2011). Additionally, 770 hospitalized COVID-19 cases and 1915 controls typed for rs12329760 by whole-exome sequencing were retrieved from the web database Network for Italian Genomes (NIG) (http://nigdb.cineca.it/index.php) (Daga et al., 2021). The 1915 controls included 1685 unrelated Italian healthy controls and 230 unrelated individuals with asymptomatic SARS-CoV-2 infection who did not need hospitalization.
**Definition of independent genome-wide associated loci**

Using the 74 significant SNPs with \( P \leq 1 \times 10^{-5} \) of chromosome 21, we defined three independent associated loci by the following computational process. The SNPs were first sorted according to their association \( P \)-value. Then, the lead SNP, considered as the most significant SNP in a given genomic locus, was removed from this list and assigned to an independent locus together with all other SNPs which have an \( r^2 \) value less than or equal to 0.01 with this SNP. This procedure was recursively applied to the remaining SNPs in the list so that each SNP could be assigned to a locus and no SNPs were left in the original list.

**Assessment of the functional role of the SNPs**

Candidate regulatory SNPs were explored by HaploReg (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) (Ward and Kellis, 2012). In this analysis we included the 14 selected SNPs in addition to their proxy SNPs (\( r^2 > 0.8 \)). Prediction of the functional impact of 14 SNPs at \( TMPRSS2/MXI \) locus was assessed by Genome Wide Annotation of Variants (GWAVA) tool (http://www.sanger.ac.uk/sanger/StatGen_Gwava) (Ritchie et al., 2014) and by Combined Annotation Dependent Depletion (CADD) tool (https://cadd.gs.washington.edu/) (Rentzsch et al., 2019). The scores assigned to each variant by the two tools were combined as follows: GWAVA and CADD scores were ranked from the smallest to largest and the obtained values were summed. PolyPhen-2 (Adzhubei et al., 2010) and SIFT (Sim et al., 2012) scores were used to predict the impact of the missense rs12329760 variant on \( TMPRSS2 \) protein function.

We used published data on eQTL in relevant tissues to help explain how observed genetic associations may affect gene expression levels. In particular, the selected top 5 SNPs were examined for eQTLs by screening the GTEx database containing precomputed eQTL data for ~70M significant associations between SNP markers and 49 human tissues (Data Source: GTEx Analysis Release V8, dbGaP Accession phs000424.v8.p2) (Consortium et al., 2017). The significance threshold adjusted for multiple comparisons is equal to 0.000055. The Blood eQTL Browser (https://www.genenetwork.nl/bloodeqtlbrowser/) was also queried to confirm the signals for \( MXI \) in an independent dataset of eQTL from blood (Westra et al., 2013). eQTLs violin plots (Figure 2b-c) were obtained from the GTEx web portal.

**Statistical analysis**

Allele frequencies for rs12329760 SNP were compared using the Chi-square test. A two-sided \( P \leq 0.05 \) was considered statistically significant.
References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.

Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, et al. (2017). Genetic effects on gene expression across human tissues. Nature 550, 204-213.

COVID-19 Host Genetics Initiative. (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet 28, 715-718.

Daga, S., Fallerini, C., Baldassarri, M., Fava, F., Valentino, F., Doddato, G., Benetti, E., Furini, S., Giliberti, A., Tita, R., et al. (2021). Employing a systematic approach to biobanking and analyzing clinical and genetic data for advancing COVID-19 research. Eur J Hum Genet.

Grundberg, E., Adoue, V., Kwan, T., Ge, B., Duan, Q.L., Lam, K.C., Koka, V., Kindmark, A., Weiss, S.T., Tantisira, K., et al. (2011). Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. PLoS Genet 7, e1001279.

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2020). Genetic mechanisms of critical illness in Covid-19. Nature.

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47, D886-D894.

Ritchie, G.R., Dunham, I., Zeggini, E., and Flicek, P. (2014). Functional annotation of noncoding sequence variants. Nat Methods 11, 294-296.

Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40, W452-457.

Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40, D930-934.

Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 45, 1238-1243.