Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal Pre-proof

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, Mario Capasso

PII: S2589-0042(21)00290-X

DOI: https://doi.org/10.1016/j.isci.2021.102322

Reference: ISCI 102322

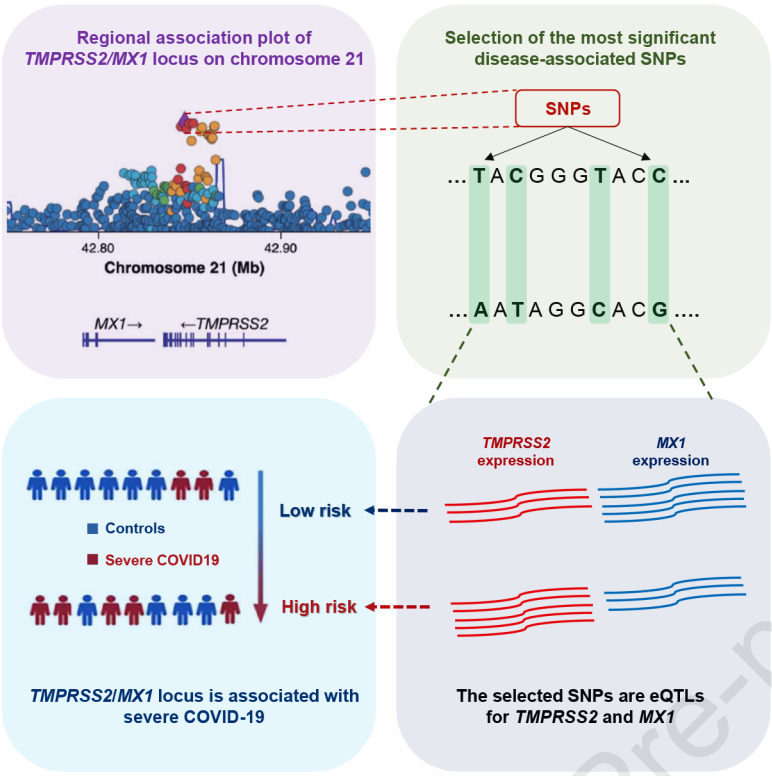
To appear in: *ISCIENCE*

Received Date: 8 January 2021

Revised Date: 15 February 2021

Accepted Date: 15 March 2021

Please cite this article as: Andolfo, I., Russo, R., Lasorsa, V.A., Cantalupo, S., Rosato, B.E., Bonfiglio, F., Frisso, G., Abete, P., Cassese, G.M., Servillo, G., Esposito, G., Gentile, I., Piscopo, C., Villani, R., Fiorentino, G., Cerino, P., Buonerba, C., Pierri, B., Zollo, M., Iolascon, A., Capasso, M., Common variants at 21q22.3 locus influence *MX1* and *TMPRSS2* gene expression and susceptibility to severe COVID-19, *ISCIENCE* (2021), doi: https://doi.org/10.1016/j.isci.2021.102322.



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

Immacolata Andolfo[1,2], Roberta Russo[1,2], Vito Alessandro Lasorsa[1,2], Sueva Cantalupo[1,2], Barbara Eleni Rosato[1,2], Ferdinando Bonfiglio[3], Giulia Frisso[1,2], Pasquale Abete[4], Gian Marco Cassese[4], Giuseppe Servillo[5], Gabriella Esposito[1,2], Ivan Gentile[6], Carmelo Piscopo[7], Romolo Villani[8], Giuseppe Fiorentino[9], Pellegrino Cerino[10], Carlo Buonerba[10], Biancamaria Pierri[10,11], Massimo Zollo[1,2], Achille Iolascon[1,2], Mario Capasso[1,2]

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

Keywords: COVID-19, SARS-CoV-2, *TMPRSS2, MX1*, SNP genotyping.

Running title: Analysis of *TMPRSS2/MX1* locus in severe COVID-19.

Corresponding author and lead contact:

Prof. Mario Capasso

Department of Molecular Medicine and Medical Biotechnologies

University of Naples, Federico II, 80145, Naples, Italy

CEINGE, Biotecnologie Avanzate,

Via Gaetano Salvatore, 486, 80145, Naples, Italy

Tel: +39-081-3737736

e-mail: mario.capasso@unina.it

Summary

The established risk factors of coronavirus disease 2019 (COVID-19) are advanced age, male sex and comorbidities, but they do not fully explain the wide spectrum of disease manifestations. Genetic factors implicated in the host antiviral response provide for novel insights into its pathogenesis.

We performed an in-depth genetic analysis of chromosome 21 exploiting the genome-wide association study data, including 6,406 individuals hospitalized for COVID-19 and 902,088 controls with European genetic ancestry from the COVID-19 Host Genetics Initiative. We found that five single nucleotide polymorphisms within *TMPRSS2* and near *MX1* gene show associations with severe COVID-19. The minor alleles of the five SNPs correlated with a reduced risk of developing severe COVID-19 and high level of *MX1* expression in blood.

Our findings demonstrate that host genetic factors can influence the different clinical presentations of COVID-19 and that MX1 could be a potential therapeutic target.

Introduction

The recent severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) pandemic has caused so far more than over 2.5 million deaths (https://covid19.who.int/). The coronavirus disease 2019 (COVID-19), caused by the SARS-Cov-2, is associated with diverse clinical presentations, ranging from asymptomatic or mildly symptomatic infections to respiratory failure and death (Bellani et al., 2021; Grasselli et al., 2021; Grasselli et al., 2020; Richardson et al., 2020). Advanced age is an established risk factor, as well as male sex and comorbidities such as hypertension and diabetes (Zhou et al., 2020). Since these risk factors do not fully explain the wide spectrum of disease manifestations, dissecting the genetics of the host response to SARS-CoV-2 infection may provide novel insights into its pathogenesis (Anastassopoulou et al., 2020).

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., 2020; 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). Addionally, 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.(Latini et al., 2020).

In our previous opinion article, based on the analysis of allele frequencies across different populations and expression quantitative triat 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., 2010). Particurarly, 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 single nucleotide polymorphisms (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 dataset includes 6,406 hospitalized cases and 902,088 controls with European ancestry ("**Table S1.** Study groups that have contributed to GWAS meta-analyses of the COVID-19 Host Genetics Initiative, Related to Figure 1"). 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., 2020). 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.** Summary statistics at chromosome 21 from GWAS dataset, Related to Figure 1). The most significant signal was represented by rs13050728 ($P = 2.76 \times 10^{-12}$, OR=0.83, **Figure 1a**) that maps within the *INFRA2* gene. The other two signals showed a suggestive significance level ($P \leq 1 \times 10^{-5}$) 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 first closest gene 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 2.7×10^{-6} to 5.8×10^{-6}, **Table 1**) were in strong LD with each other ($r^2 >= 0.90$, "**Figure S1.** Linkage disequilibrium block at TMPRSS2/MX1 locus, Related to Figure 1"). The other 9 SNPs showed an LD with the lead SNP rs3787946 ranging from 0.8 to 0.9 and P-values ranging from 6×10^{-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., 2020) with non-European ancestry. All the 11 available SNPs replicated in the east asian population (EAS) population; the top five SNPs replicated in the South Asian (SAS) ancestry population, whereas two out 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.** Characteristics of Italian patients recruited by our research group, Related to Table 1") 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_{allele}=0.89, P_{allele}=0.07; $OR_{dominant}$=0.57, P=0.01; OR_{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.** Results of SNP enrichment analysis in regulatory elements in different tissues and cell types, Related to Figure 2"), but the best enrichment was found in induced pluripotent stem cells and thymus (**Figure 2a**).

Functional role of the most significant SNPs at *TMPRSS2*/*MX1* locus

We then investigated the predicted functional role of the 14 SNPs by GWAVA and CADD tools. We found that two out 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 *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 *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 (**Figure 2b, "Figure S2a.** Results of SNP enrichment analysis in regulatory elements in different tissues and cell types, Related to Figure 2"). Of note, all the other SNPs, except for rs2298660, did not have eQTL signals for *MX1* in the blood ("**Table S5.** Results of eQTL analysis for the common variants at TMPRSS2/MX1 locus, Related to Figure 2). The two SNPs rs12329760 and rs2298660 were confirmed as eQTLs for *MX1* in the blood ($P=1.79\times10^{-6}$ and 2.8×10^{-6}, minor alleles correlated with a higher

178 expression compared to the major alleles) by interrogation of another independent publicly
179 available dataset (Westra et al., 2013). *TMPRSS2* is highly expressed in lung (Russo et al.,
180 2020), so we investigated if the top five SNPs were eQTLs for *TMPRSS2* in lung tissues at a
181 nominally statistically significant level ($P \leq 0.05$). We found that the minor alleles of four out
182 of five SNPs correlated with lower expression of *TMPRSS2* compared to the major alleles
183 (**Figure 2c and "Figure S2b.** Results of SNP enrichment analysis in regulatory elements in
184 different tissues and cell types, Related to Figure 2"). Notably, rs12329760 is also an eQTL
185 for *TMPRSS2* in osteoblasts treated with dexamethasone (Grundberg et al., 2011).
186

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., 2020; 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 *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 *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 *TMPRSS2* and near the *MX1* 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 the 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 interfern 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. Particurarly, *INFRA2* gene encodes for the type I membrane protein that forms the interferon-α/β receptor, involved in the canonical host antiviral signalling mediators (Duncan et al., 2015), so associated with interferon signlalling like MX1. The SNP rs13050728 was previously identified as lead variant from the meta-analysis

of overlapping SNPs between GenOMICC, The COVID-19 Host Gentics 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., 2020). These findings, along with ours, further strength 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 pathogensis 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 indipendent 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 aggressiviness or controlled for comorbidities in larger cohorts.

All methods can be found in the accompanying "Transparent methods supplemental file".

Resource availability

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 (htts://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/).

Author Contributions

IA, RR, and MC designed and conducted the study, and prepared the manuscript; MC, VAL, and FB analyzed the data; BER sampled genomic DNA from COVID-19 patients; SC genotyped COVID-19 patients and in-house controls; GF, AP, GMC, GS, GE, IG, CP, RV, GP, PC, CB, BP cared for COVID-19 patients; MZ and AI provided a critical review of the manuscript. All the authors read and approved the final manuscript.

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 data management and analysis.

Declaration of interests

The authors declare that there are no competing interests.

References

Anastassopoulou, C., Gkizarioti, Z., Patrinos, 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.

Asselta, 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.

Bellani, G., Grasselli, G., Cecconi, M., Antolini, 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.

Benetti, E., Tita, R., Spiga, O., Ciolfi, A., Birolo, G., Bruselles, A., Doddato, G., Giliberti, A., Marconi, C., Musacchia, 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., Lage-Vickers, S., Lavignolle, R., Toro, A., Olszevicki, 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.

Ciancanelli, 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.

Duncan, C.J., Mohamad, S.M., Young, D.F., Skelton, A.J., Leahy, T.R., Munday, D.C., Butler, K.M., Morfopoulou, S., Brown, J.R., Hubank, M., *et al.* (2015). 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., Asselta, 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., Cereda, 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.

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 Pediatr.

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 e775.

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 e278.

Hou, Y., Zhao, J., Martin, W., Kallianpur, A., Chung, M.K., Jehi, L., Sharifi, 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 Vaccin Immunother, 1-6.

Latini, A., Agolini, E., Novelli, A., Borgiani, P., Giannini, 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*.

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., Katoh, 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.* (2020). Genetic mechanisms of critical illness in Covid-19. Nature.

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., Andolfo, 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., Filshtein-Somnez, 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, J., van den Heuvel, G., Mantere, T., Kersten, S., van Deuren, R.C., Steehouwer, M., van Reijmersdal, S.V., Jaeger, M., *et al.* (2020). Presence of Genetic Variants Among Young Men With Severe COVID-19. JAMA.

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 CTSL) and its potential use in association studies. Life Sci *260*, 118313.

Vishnubhotla, R., Vankadari, N., Ketavarapu, V., Amanchy, R., Avanthi, 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 Wahlroos, 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*.

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

455

456

Figure legends

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 (−log10 transformed P values) and the recombination rate. SNPs are color□coded based on pair□wise linkage disequilibrium (r^2) with indicated lead SNPs: rs13050728 (panel a), rs111783124 (panel b) and rs3787946 (panel c).

Figure 2. Enrichment of SNPs in regulatory regions and eQTL analyses. The statistically significant fold enrichments (P<0.05 after Bonferroni correction) of SNPs in regulatory DNA regions active in different tissues are shown (a). eQTL violin plots between genotypes of rs3787946 (b) and rs3787946 (c) with *MX1* and *TMPRSS2* expression from the from the Genotype-Tissue Expression (GTEx). The significance threshold adjusted for multiple comparisons is equal to 0.000055.

Table 1. Associations of SNPs at *TMPRSS2/MX1* risk locus in linkage disequilibrium with the lead rs3787946 in different populations and prioritization scores

RS number	EA	OA	MAF	r^2	OR	P_EUR	OR	P_EAS	OR	P_SAS	OR	P_AFR	*Region score	*TSS score	^Predicted Function	^Score	°Combined score
rs3787946	C	G	0.23	1.00	0.87	**2.73E-06**	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	**3.12E-06**	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	**3.13E-06**	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	**4.51E-06**	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	**5.80E-06**	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	NA	0.12	0.28	INTRONIC	2	4
rs7364088	A	G	0.26	0.84	0.91	0.002	NA	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	NA	0.33	0.42	REGULATORY	4	26

*Scores from GWAVA predictor tool
^Scores from CADD predictor tool
°GWAVA and CADD scores were ranked from the smallest to largest and the obtained values were summed
In bold the SNPs that replicated in at least one cohort
EA: Effect Allele; OA: Other Allele
EUR: Europen; EAS: East Asian; SAS: South Asian; AFR: African; ITA: Italian
MAF: minor allele frequency
OR: Odds Ratio

Table 2. Association of rs12329760 SNP with severe COVID-19 in Italian population

	SI cases n=226		**SI controls** n=1848		**NIG cases** n=770		**NIG controls** n=1915		**All cases** n=996		**All controls** n=3763		**P_{SI}**	**OR (CI: 95%)**	**P_{NIG}**	**OR (CI: 95%)**	**P_{All}**	**OR (CI: 95%)**
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*						
Genotype																		
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

475

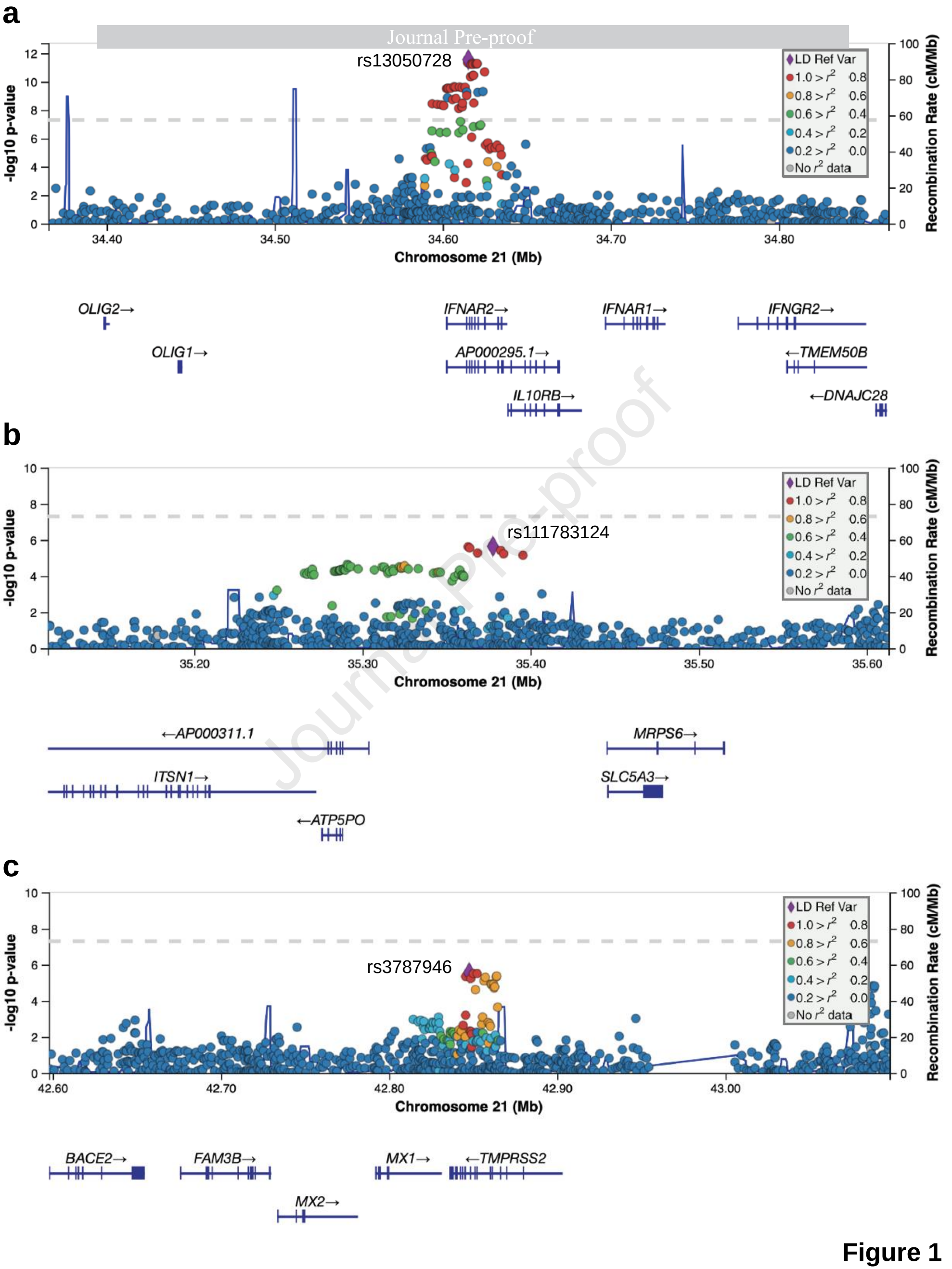


Figure 1

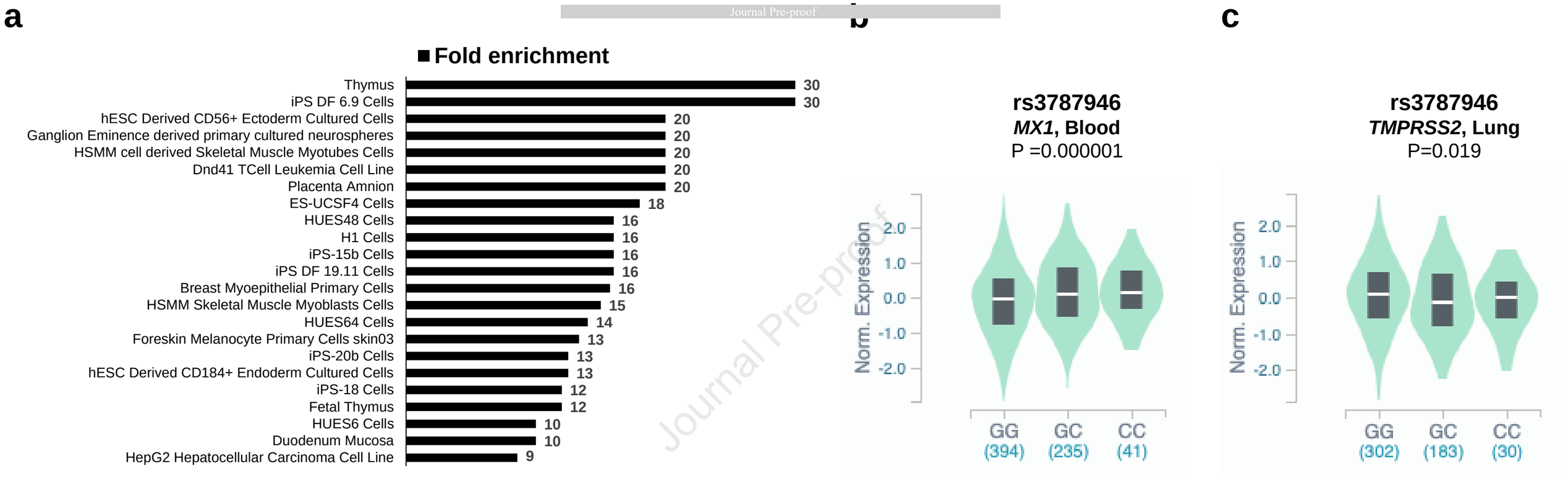


Figure 2

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

Immacolata Andolfo[1,2], Roberta Russo[1,2], Vito Alessandro Lasorsa[1,2], Sueva Cantalupo[1,2], Barbara Eleni Rosato[1,2], Ferdinando Bonfiglio[3], Giulia Frisso[1,2], Pasquale Abete[4], Gian Marco Cassese[4], Giuseppe Servillo[5], Gabriella Esposito[1,2], Ivan Gentile[6], Carmelo Piscopo[7], Romolo Villani[8], Giuseppe Fiorentino[9], Pellegrino Cerino[10], Carlo Buonerba[10], Biancamaria Pierri[10,11], Massimo Zollo[1,2], Achille Iolascon[1,2], Mario Capasso[1,2]

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

Keywords: COVID-19, SARS-CoV-2, *TMPRSS2*, *MX1*, SNP genotyping.

Running title: Analysis of *TMPRSS2/MX1* locus in severe COVID-19.

Corresponding author and lead contact:

Prof. Mario Capasso

Department of Molecular Medicine and Medical Biotechnologies

University of Naples, Federico II, 80145, Naples, Italy

CEINGE, Biotecnologie Avanzate,

Via Gaetano Salvatore, 486, 80145, Naples, Italy

Tel: +39-081-3737736

e-mail: 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.