

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

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

<sup>1</sup>Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Napoli, Italy

<sup>2</sup>CEINGE Biotecnologie Avanzate, Napoli, Italy

<sup>3</sup>Dipartimento di Ingegneria chimica, dei Materiali e della Produzione industriale, Università degli Studi di Napoli Federico II, Napoli, Italy

<sup>4</sup>COVID Hospital, P.O.S. Anna e SS. Madonna della Neve di Boscotrecase, Ospedali Riuniti Area Vesuviana, Napoli, Italy

<sup>5</sup>Dipartimento di Neuroscienze e Scienze riproduttive ed odontostomatologiche, Università degli Studi di Napoli Federico II, Napoli, Italy

<sup>6</sup>Dipartimento di Medicina clinica e Chirurgia, Università degli Studi di Napoli Federico II, Napoli, Italy

<sup>7</sup>Medical and Laboratory Genetics Unit, A.O.R.N. 'Antonio Cardarelli', Napoli, Italy

<sup>8</sup>Poison Centre, A.O.R.N. 'Antonio Cardarelli', Napoli, Italy

<sup>9</sup>AORN dei Colli Presidio Ospedaliero Cotugno, Napoli, Italy

<sup>10</sup>Istituto Zooprofilattico Sperimentale del Mezzogiorno, Napoli, Italy

<sup>11</sup>Dipartimento di Medicina, Chirurgia e Odontoiatria "Scuola Medica Salernitana", Università di Salerno, Baronissi, Italy

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

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

### Corresponding author:

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](mailto:mario.capasso@unina.it)

## Abstract

The COVID-19 disease, caused by the SARS-Cov-2, presents a heterogeneous clinical spectrum. The risk factors do not fully explain the wide spectrum of disease manifestations, so it is possible that genetic factors could account for novel insights into its pathogenesis.

In our previous study, we hypothesized that common variants on chromosome 21, near *TMPRSS2* and *MX1* genes, may be genetic risk factors associated to the different clinical manifestations of COVID-19. Here, 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 COVID-19 Host Genetics Initiative. We found that five single nucleotide polymorphisms (SNPs) within *TMPRSS2* and near *MX1* gene show suggestive associations ( $P \leq 1 \times 10^{-5}$ ) with severe COVID-19. All five SNPs replicated the association in two independent cohorts of Asian subjects while two and one out of the 5 SNPs replicated in African and Italian populations, respectively ( $P \leq 0.05$ ). The minor alleles of these five SNPs correlated with a reduced risk of developing severe COVID-19 and increased level of *MX1* expression in blood.

Our findings provide further evidence that host genetic factors can contribute to determine the different clinical presentations of COVID-19 and that *MX1*, an antiviral effector of type I and III interferon pathway, may be a potential therapeutic target.

## Introduction

The recent SARS-Cov-2 pandemic has caused so far more than one million of deaths (<https://covid19.who.int/>). The COVID-19, caused by the SARS-Cov-2, has several clinical conditions ranging from no or mild symptoms to rapid progression to respiratory failure (1, 2). Advanced age is a major risk factor, as well as male sex and some co-morbidities such as hypertension and diabetes (3). Since clinical risk factors do not fully explain the wide spectrum of disease manifestations, it is conceivable that dissecting the genetics of the disease may provide novel insights into its pathogenesis.

A recent genome-wide association study (GWAS) identified two susceptibility loci of severe COVID-19: the first locus on chromosome 3 harbors multiple genes (*SLC6A20*, *LZFTL1*, *CCR9*, *CXCR6*, *XCRI*, *FYCO1*) that could be functionally implicated in COVID-19 pathology; the second on chromosome 9 that defines the ABO blood groups (4). Another study showed that inactivating rare mutations in genes belonging to type I interferon pathway predispose to life-threatening COVID-19 pneumonia (5). 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 variants at chromosome 19p13.3, 12q24.13 and 21q22.1 associated with severe COVID-19 (6, 7).

In our previous opinion article, based on allele frequencies across different populations and expression quantitative loci (eQTLs) data, we hypothesized that common variants on chromosome 21 near *TMPRSS2* and *MXI* genes may be genetic risk factors associated to the COVID-19 different clinical manifestations (8). In this study, to further support our hypothesis, we exploited GWAS data from the COVID-19 Host Genetics Initiative (9) and performed an in-depth mapping of chromosome 21 using summary statistics where common variants at this chromosome were associated with severe COVID-19 at genome-wide significance level ( $P \leq 5 \times 10^{-8}$ ). Using the cohort of 908,494 subjects with European origins, we found 5 single nucleotide polymorphisms (SNPs) at the *TMPRSS2/MXI* locus showing suggestive association with the disease. All 5 SNPs replicated the association in two independent cohorts of Asian subjects whereas two SNPs confirmed the association in African and one SNP in Italian cohort. Significant eQTLs signals were found for *MXI* in blood.

## Methods

### GWAS

The summary statistics of chromosome 21 were obtained from the GWAS dataset “B2\_ALL\_eur\_leave\_23andme” deposited in the COVID-19 Host Genetics Initiative website (9). It includes 6,406 laboratory confirmed SARS-CoV-2 infection and hospitalized for COVID-19 cases and 902,088 controls from general population with European genetic ancestry.

### Replication

The summary statistics of the SNPs used for the replication study were retrieved from the GenOMICC study (6). Three independent cohorts of cases and controls with different ethnicity were available throughout GenOMICC GWAS study (REF 6): 182 individuals from African ancestry (AFR), 149 from East Asian ancestry (EAS), 237 from South Asian ancestry (SAS). Moreover, 157 hospitalized COVID-19 cases and 1388 controls with Italian origins were typed for the rs12329760 variant by TaqMan® SNP Genotyping (Applied Biosystems by Thermo Fisher Scientific).

### 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 a  $r^2$  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>) (10). Prediction of the functional impact of 14 SNPs at *TMPRSS2/MXI* locus was assessed by Genome Wide Annotation of VARIants (GWAVA) tool ([https://www.sanger.ac.uk/sanger/StatGen\\_Gwava](https://www.sanger.ac.uk/sanger/StatGen_Gwava)) (11) and by Combined Annotation Dependent Depletion (CADD) tool

(<https://cadd.gs.washington.edu/>) (12). The distribution of the eQTLs was explored by Genotype Tissue Expression (GTEx) database (<https://www.gtexportal.org/home>).

### **Statistical analysis**

Allele frequencies were compared using the Chi-square test. A two-sided  $P \leq .05$  was considered statistically significant.

## Results

### ***TMPRSS2/MXI* locus is associated with severe COVID-19**

To prove that common variants at *TMPRSS2/MXI* (21q22.3) locus may affect the susceptibility to severe COVID-19 onset, we analyzed the summary statistics of a large available GWAS dataset released from COVID-19 Host Genetics Initiative. The database includes 6,406 hospitalized cases and 902,088 controls with European ancestry. The region on chromosome 21 appears to be significantly associated with severe COVID-19 at the genome-wide level (<https://www.covid19hg.org/results/>) (6). To investigate whether more than one association signal may exist at chromosome 21, we selected 74 SNPs showing a  $P \leq 1 \times 10^{-5}$  and among these we identified 3 independent loci (**Supplementary Table 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 *MXI* (**Figure 1c**); herein we named this locus as “*TMPRSS2/MXI*”. An in-depth inspection of *TMPRSS2/MXI* locus showed that 13 SNPs were in LD with the lead rs3787946 ( $r^2 > 0.8$ , **Table 1**) and that the 5 most significant SNPs (P-values ranged from  $2.7 \times 10^{-6}$  to  $5.8 \times 10^{-6}$ , **Table 1**) were in strong LD with each other ( $r^2 \geq 0.90$ , **Supplementary 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  $6 \times 10^{-4}$  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 (6) with non-European ancestry. All the 11 available SNPs replicated in the EAS population; the top 5 SNPs replicated in SAS population whereas two out five SNPs in the AFR one (**Table 1**). By using TaqMan assay, we typed the rs12329760 variant in 157 hospitalized COVID-19 and 1388 controls with Italian origins (**Supplementary Table 2**) collected in our Institute and confirmed the minor allele as protective factor against the aggressive form of disease (**Table 1**, OR=0.72,  $P=0.05$ ).

### **SNPs at *TMPRSS2/MXI* locus are enriched in regulatory regions active in 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

(**Supplementary Table 3**), but the best enrichment was found in induced pluripotent stem cell and thymus (**Figure 2a**).

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

We then investigated the predicted functional role of the 14 SNPs by GWAVA and CADD tools. We found that 2 (rs12329760 and rs9983330) out of 5 of the most significant SNPs showed the second and third most significant score (**Table 1**). The rs12329760 was classified as coding variant (p.Val197Met) localized in the exon 6 of *TMPRSS2* gene and was predicted to be pathogenic (PolyPhen=probably damaging and SIFT=deleterious).

### **The most significant disease associated SNPs are eQTLs for *MXI* in blood**

We verified if the top 5 SNPs (**Table 1**) might cause gene expression alterations interrogating the GTEx portal. We found that all the top 5 SNPs had eQTL signals for *MXI* exclusively in blood tissue. Particularly, the minor alleles of these SNPs correlated with higher expression of *MXI* compared to the major alleles (**Figure 2b, Supplementary Figure 2a**). Of note, all the 9 other SNPs, except for rs2298660, did not have eQTL signals for *MXI* in blood (**Supplementary Table 4**). The two SNPs, rs12329760 and rs2298660, were confirmed as eQTLs for *MXI* in 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 dataset (13). *TMPRSS2* is highly expressed in lung (8), so we investigated if the top 5 SNPs were eQTL for *TMPRSS2* in lung tissues at nominally statistically significant level ( $P\leq 0.05$ ). We found that the minor alleles of 4 out of 5 SNPs correlated with lower expression of *TMPRSS2* compared to the major alleles (**Figure 2c and Supplementary Figure 2b**). Notably, rs12329760 is also an eQTL for *TMPRSS2* in osteoblasts treated with dexamethasone (14).

## Discussion

Despite the substantial advances made in the recent months in the field of the 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 clinical heterogeneity of COVID-19 and are also crucial to find new druggable therapeutic targets (15). The main host cell entry factors of SARS-CoV-2 are ACE2 and TMPRSS2 (16, 17). The transmembrane spike (S) glycoprotein of virus binds to the ACE2 making it essential for the invasion of the virus into the host cell, followed by attachment of the virus to the target cells. S- protein priming by TMPRSS2 allows the binding of viral and cellular membranes, resulting in virus entry and replication in the host cells (18).

In our previous study, we hypothesized that common variants at chromosome 21, driving *TMPRSS2* and *MXI* expression, might have a mild-to-moderate effect in the susceptibility to SARS-CoV-2 infection. Particularly, genetic variants associated with reduced *TMPRSS2* and elevated *MXI* expression might confer less individual susceptibility to SARS-CoV-2 infection and favor a better outcome (8). 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 (9) 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 *MXI* gene that showed suggestive associations ( $P \leq 1 \times 10^{-5}$ ) with severe COVID-19. In particular, we found that the alleles with minor frequency were less recurrent among the hospitalized patients when compared to the control individuals, suggesting their protective role against the progression of disease. Interestingly, all the five SNPs 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 origins. These results strongly indicate 21q22.3 as novel susceptibility locus to unfavorable outcome of COVID-19 and suggest that molecular mechanisms underlying this genetic predisposition may be common among individuals with different ethnicity.

Of note, the SNPs at this locus were enriched in the regulatory regions of induced pluripotent stem cell and thymus. 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. 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 (19, 20). Thus, our finding suggests that the identified SNPs could be severe COVID-19 risk factors also for their implications in the thymus function.

The five SNPs, here identified, had eQTL signals for *MXI* exclusively in blood tissue. Particularly, the minor allele of these SNPs correlated with higher expression of *MXI* and associated with a minor risk of developing severe COVID-19. These results support the evidence that *MXI* can play a relevant role in determining less severe forms of disease and are in line with a recent study that suggests *MXI* as antiviral effector against SARS-CoV-2 (21). Indeed, the expression of *MXI* was found to be higher in SARS-CoV-2 positive subjects, negatively correlated with age and independently associated with increased viral load (21). *MXI* is part of the antiviral response induced by type I and III interferons (IFN) (22). Inactivating mutations in genes belonging to type I IFN pathway and the consequent decreased levels of proteins have been shown to recur in patients with severe COVID-19 (5). Therefore, our results further support the evidence that the use of drugs, activating IFN signaling, could be an effective treatment to prevent the adverse outcome of disease. There are already several ongoing clinical trials for COVID-19 prevention and/or treatment using type I or III IFNs (NCT04343976 [Phase II], NCT04385095 [Phase II], NCT04354259 [Phase II], NCT04293887 [Early Phase I], NCT04344600 [Phase II], NCT04320238 [Phase III], NCT04388709 [Phase II]). On the contrary, it is important to note that IFN administration could enhance a “cytokine-storm” causing a hyper-inflammatory response and contributing to multiple organ failure (21). Therefore, it is highly relevant to spawn new drugs that have the capacity to boost the host immune response while controlling tissue damage as a consequence of these infections.

We also report that the minor allele of 4 of the top 5 SNPs might reduce the expression of *TMPRSS2* in lung tissues and that the rs12329760 coding variant (p.Val197Met) correlate with lower expression of *TMPRSS2* in osteoblast treated with dexamethasone (14), a drug currently used to inhibit an excessive inflammation response (23). Of note, the rs12329760

coding variant was recently found to be less frequent among Chinese patients with critical COVID-19 disease (24). This variant is predicted to decrease the *TMPRSS2* protein stability and ACE2 binding, thus decreasing the virus entry into the cells (25). Together, these data suggest that even the functions of *TMPRSS2* may be affected by risk variants of severe COVID-19. However, caution should be paid in considering these results as the eQTL signals in lung resulted to be not significant when corrected for multiple tests. Additional studies are needed to further verify the role of genetic variants *TMPRSS2/MXI* at locus in modulating the expression of *TMPRSS2*.

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

## References

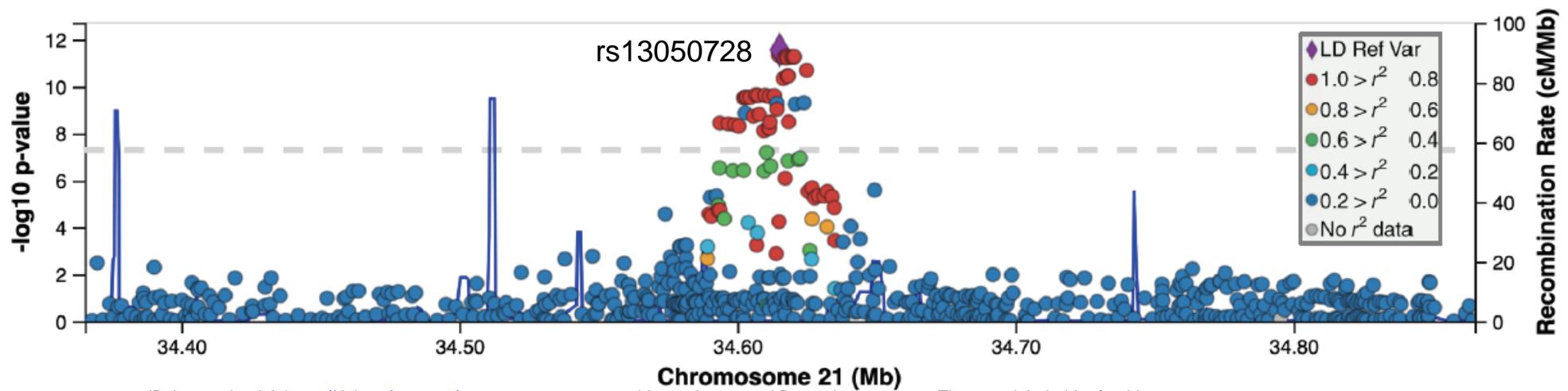
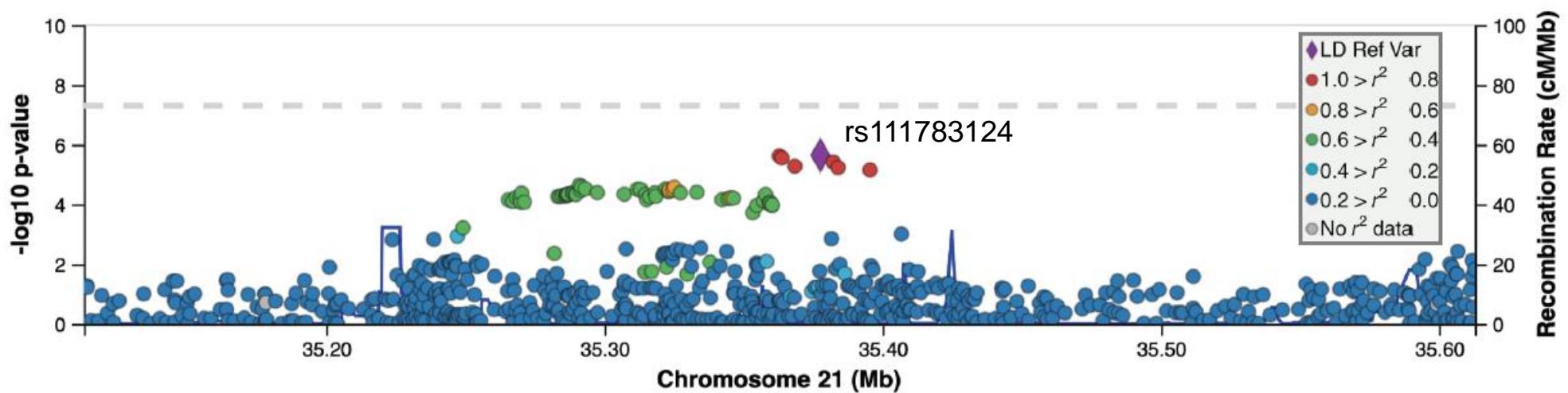
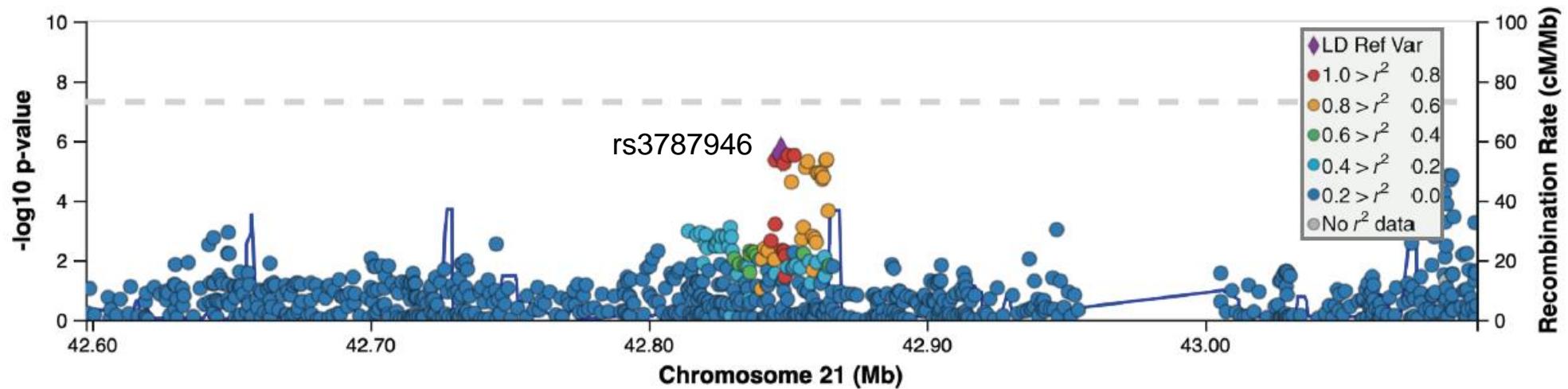
1. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*. 2020;323(20):2052-9.
2. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*. 2020;323(16):1574-81.
3. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-62.
4. Severe Covid GG, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med*. 2020;383(16):1522-34.
5. Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*. 2020;370(6515).
6. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in Covid-19. *Nature*. 2020.
7. Janie F, Shelton AJS, Chelsea Ye, Catherine H, Weldon, Teresa Filshtein-Somnez, Daniella Coker, Antony Symons, Jorge Esparza-Gordillo, The 23andMe COVID-19 Team, Stella Aslibekyan, Adam Auton. Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity. *medRxiv*. 2020.
8. Russo R, Andolfo I, Lasorsa VA, Iolascon A, Capasso M. Genetic Analysis of the Coronavirus SARS-CoV-2 Host Protease TMPRSS2 in Different Populations. *Front Genet*. 2020;11:872.
9. Initiative C-HG. 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*. 2020;28(6):715-8.
10. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40(Database issue):D930-4.
11. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat Methods*. 2014;11(3):294-6.
12. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D94.
13. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45(10):1238-43.
14. Grundberg E, Adoue V, Kwan T, Ge B, Duan QL, Lam KC, et al. Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet*. 2011;7(1):e1001279.

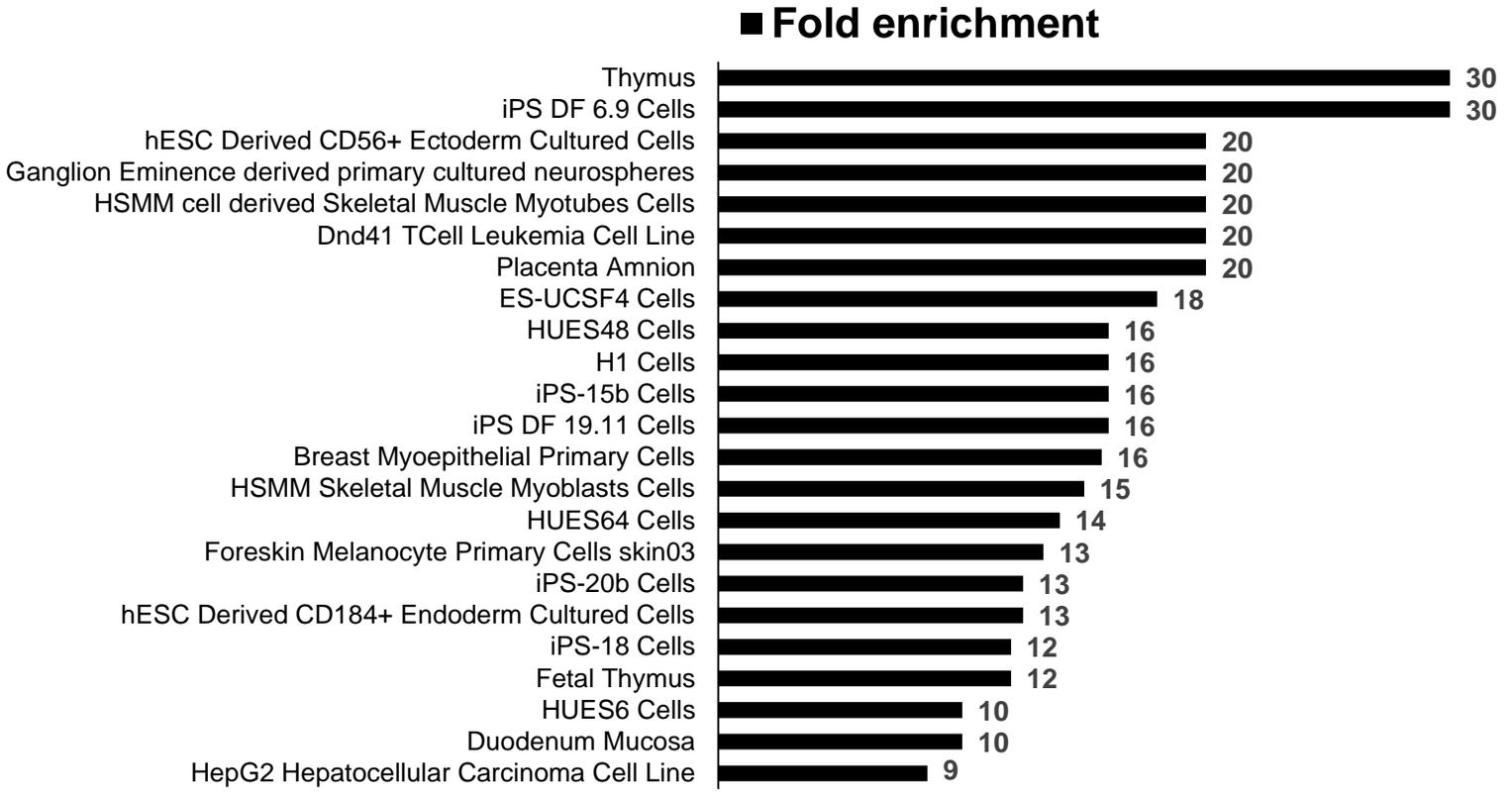
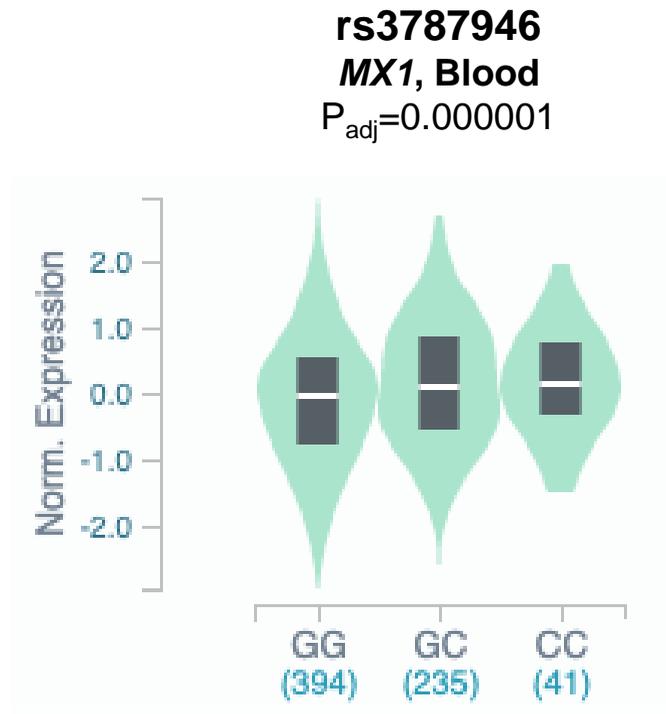
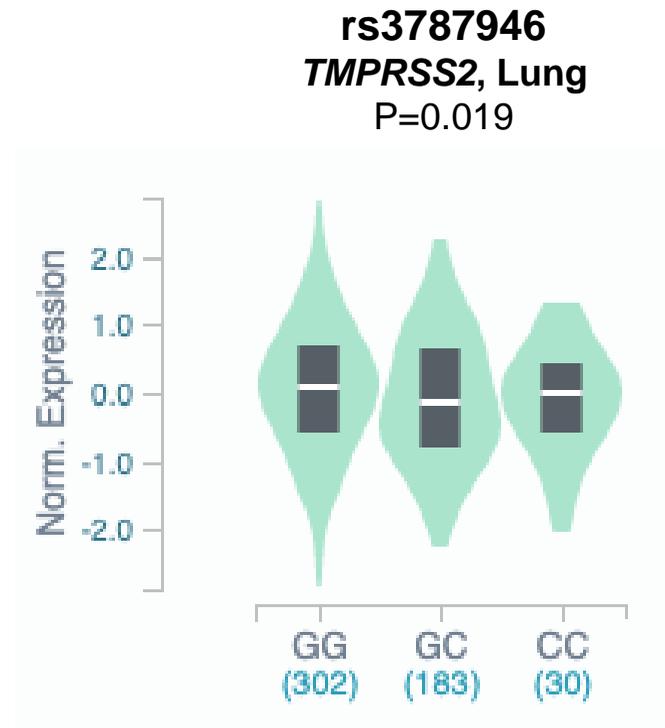
15. Beck DB, Aksentijevich I. Susceptibility to severe COVID-19. *Science*. 2020;370(6515):404-5.
16. Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *Eur J Hum Genet*. 2020.
17. Asselta R, Paraboschi EM, Mantovani A, Duga S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *Aging (Albany NY)*. 2020;12(11):10087-98.
18. Singh H, Choudhari R, Nema V, Khan AA. ACE2 and TMPRSS2 polymorphisms in various diseases with special reference to its impact on COVID-19 disease. *Microb Pathog*. 2020;150:104621.
19. Kellogg C, Equils O. The role of the thymus in COVID-19 disease severity: implications for antibody treatment and immunization. *Hum Vaccin Immunother*. 2020:1-6.
20. Gunes H, Dincer S, Acipayam C, Yurttutan S, Ozkars MY. What chances do children have against COVID-19? Is the answer hidden within the thymus? *Eur J Pediatr*. 2020.
21. Bizzotto J, Sanchis P, Abbate M, Lage-Vickers S, Lavignolle R, Toro A, et al. SARS-CoV-2 Infection Boosts MX1 Antiviral Effector in COVID-19 Patients. *iScience*. 2020;23(10):101585.
22. Zav'yalov VP, Hamalainen-Laanaya H, Korpela TK, Wahlroos T. Interferon-Inducible Myxovirus Resistance Proteins: Potential Biomarkers for Differentiating Viral from Bacterial Infections. *Clin Chem*. 2019;65(6):739-50.
23. Group RC, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*. 2020.
24. Wang F, Huang S, Gao R, Zhou Y, Lai C, Li Z, et al. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell Discov*. 2020;6(1):83.
25. Ravikanth Vishnubhotla NV, Vijayasathy Ketavarapu, Ramars Amanchy, Steffie Avanthi, Govardhan Bale, Duvvur Nageshwar Reddy, Mitnala Sasikala. Genetic variants in TMPRSS2 and Structure of SARS-CoV-2 spike glycoprotein and TMPRSS2 complex. *bioRxiv*. 2020.

## Figure legend

**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 (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 *MXI* and *TMPRSS2* expression from the Genotype-Tissue Expression (GTEx).

**a****b****c****Figure 1**

**a****b****c****Figure 2**

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

RS Number	EA	OA	MAF	r <sup>2</sup>	OR	P_EUR	OR	P_EAS	OR	P_SAS	OR	P_AFR	OR	P_ITA	*Region score	*TSS score	^Function	^Score
rs3787946	C	G	0.23	1.00	0.87	<b>2.73E-06</b>	0.63	<b>0.026</b>	0.71	<b>0.02</b>	0.74	<b>0.07</b>	NA	NA	0.16	0.29	INTRONIC	2
rs9983330	G	A	0.23	0.91	0.88	<b>3.12E-06</b>	0.54	<b>0.004</b>	0.73	<b>0.04</b>	0.79	0.16	NA	NA	0.31	0.64	REGULATORY	4
rs12329760	T	C	0.24	0.90	0.88	<b>3.13E-06</b>	0.64	<b>0.029</b>	0.76	<b>0.08</b>	0.78	0.14	0.72	<b>0.059</b>	<b>0.32</b>	<b>0.41</b>	<b>MISSENSE</b>	<b>7</b>
rs2298661	A	C	0.23	0.99	0.88	<b>4.51E-06</b>	0.63	<b>0.030</b>	0.67	<b>0.01</b>	0.60	<b>0.01</b>	NA	NA	0.18	0.35	INTRONIC	2
rs9985159	T	C	0.23	0.98	0.88	<b>5.80E-06</b>	0.61	<b>0.018</b>	0.75	<b>0.06</b>	0.98	0.89	NA	NA	0.16	0.46	INTRONIC	2
rs2298660	T	C	0.20	0.82	0.88	0.001	NA	NA	NA	NA	NA	NA	NA	NA	0.12	0.28	INTRONIC	2
rs7364088	A	G	0.26	0.84	0.91	0.002	NA	NA	NA	NA	NA	NA	NA	NA	0.19	0.23	INTRONIC	2
rs2298663	T	C	0.25	0.87	1.08	<b>0.005</b>	1.49	<b>0.052</b>	1.12	0.40	0.94	0.66	NA	NA	0.26	0.37	REGULATORY	4
rs2094881	C	T	0.25	0.87	1.08	<b>0.005</b>	1.47	<b>0.058</b>	1.10	0.47	0.93	0.60	NA	NA	0.29	0.26	REGULATORY	4
rs8131649	T	C	0.25	0.85	0.92	<b>0.007</b>	0.64	<b>0.035</b>	0.90	0.46	1.01	0.93	NA	NA	0.26	0.35	REGULATORY	4
rs8134203	T	C	0.26	0.85	1.08	<b>0.007</b>	1.49	<b>0.058</b>	1.09	0.54	0.91	0.50	NA	NA	0.26	0.41	REGULATORY	4
rs8134216	T	C	0.26	0.85	1.08	<b>0.007</b>	1.54	<b>0.038</b>	1.11	0.43	0.91	0.49	NA	NA	0.28	0.4	REGULATORY	4
rs2104810	A	G	0.26	0.85	1.08	<b>0.008</b>	1.54	<b>0.040</b>	1.10	0.47	0.90	0.48	NA	NA	0.23	0.35	REGULATORY	4
rs8131648	C	T	0.26	0.85	1.07	0.036	NA	NA	NA	NA	NA	NA	NA	NA	0.33	0.42	REGULATORY	4

\*Scores from GWAVA predictor tool

^Scores from CADD predictor tool

In bold the SNPs that replicated in at least one cohort

EA: Effect Allele; OA: Other Allele

EUR: Europe; EAS: Est Asia; SAS: South Asia; AFR: Africa; ITA: Italia

MAF: minor allele frequency

OR: Odds Ratio