



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A common TMPRSS2 variant has a protective effect against severe COVID-19

Citation for published version:

David, A, Parkinson, N, Peacock, TP, Pairo-Castineira, E, Khanna, T, Cobat, A, Tenesa, A, Sancho-Shimizu, V, Casanova, J-L, Abel, L, Barclay, WS, Baillie, JK & Sternberg, MJE 2022, 'A common TMPRSS2 variant has a protective effect against severe COVID-19', *Current Research in Translational Medicine*, vol. 70, no. 2, 103333. <https://doi.org/10.1016/j.retram.2022.103333>

Digital Object Identifier (DOI):

[10.1016/j.retram.2022.103333](https://doi.org/10.1016/j.retram.2022.103333)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Current Research in Translational Medicine

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Available online at
ScienceDirect
 www.sciencedirect.com

Elsevier Masson France
EM|consulte
 www.em-consulte.com



A common TMPRSS2 variant has a protective effect against severe COVID-19

Alessia David^{a,*}, Nicholas Parkinson^b, Thomas P Peacock^c, Erola Pairo-Castineira^b, Tarun Khanna^a, Aurelie Cobat^{d,e,f}, Albert Tenesa^b, Vanessa Sancho-Shimizu^{g,h,#}, Jean-Laurent Casanova^{d,e,f,i}, Laurent Abel^{d,e,f}, Wendy S. Barclay^c, J.Kenneth Baillie^{b,j}, Michael JE Sternberg^a

^a Centre for Integrative System Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London, SW7 2AZ, UK

^b Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, UK

^c Department of Infectious Diseases, Imperial College London, London, W2 1PG, UK

^d St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065, USA

^e Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Necker Hospital for Sick Children, Paris, EU France

^f University of Paris, Imagine Institute, Paris, EU France

^g Department of Paediatric Infectious Diseases & Virology, Imperial College London, London, UK

^h Centre for Paediatrics and Child Health, Faculty of Medicine, Imperial College London, London, UK

ⁱ Howard Hughes Medical Institute, New York, NY, USA

^j Intensive Care Unit, Royal Infirmary of Edinburgh, 54 Little France Drive, Edinburgh, EH16 5SA, UK

ARTICLE INFO

Article History:

Received 18 October 2021

Accepted 6 January 2022

Available online 10 January 2022

Keywords:

SARS-CoV-2

COVID-19

TMPRSS2

Targeting the host to prevent

COVID-19 severity

ABSTRACT

Background: The human protein transmembrane protease serine type 2 (TMPRSS2) plays a key role in SARS-CoV-2 infection, as it is required to activate the virus' spike protein, facilitating entry into target cells. We hypothesized that naturally-occurring TMPRSS2 human genetic variants affecting the structure and function of the TMPRSS2 protein may modulate the severity of SARS-CoV-2 infection.

Methods: We focused on the only common TMPRSS2 non-synonymous variant predicted to be damaging (rs12329760 C>T, p.V160M), which has a minor allele frequency ranging from 0.14 in Ashkenazi Jewish to 0.38 in East Asians. We analysed the association between the rs12329760 and COVID-19 severity in 2,244 critically ill patients with COVID-19 from 208 UK intensive care units recruited as part of the GenOMICC (Genetics Of Mortality In Critical Care) study. Logistic regression analyses were adjusted for sex, age and deprivation index. For in vitro studies, HEK293 cells were co-transfected with ACE2 and either TMPRSS2 wild type or mutant (TMPRSS2_{V160M}). A SARS-CoV-2 pseudovirus entry assay was used to investigate the ability of TMPRSS2_{V160M} to promote viral entry.

Results: We show that the T allele of rs12329760 is associated with a reduced likelihood of developing severe COVID-19 (OR 0.87, 95%CI:0.79–0.97, $p = 0.01$). This association was stronger in homozygous individuals when compared to the general population (OR 0.65, 95%CI:0.50–0.84, $p = 1.3 \times 10^{-3}$). We demonstrate in vitro that this variant, which causes the amino acid substitution valine to methionine, affects the catalytic activity of TMPRSS2 and is less able to support SARS-CoV-2 spike-mediated entry into cells.

Conclusion: TMPRSS2 rs12329760 is a common variant associated with a significantly decreased risk of severe COVID-19. Further studies are needed to assess the expression of TMPRSS2 across different age groups. Moreover, our results identify TMPRSS2 as a promising drug target, with a potential role for camostat mesilate, a drug approved for the treatment of chronic pancreatitis and postoperative reflux esophagitis, in the treatment of COVID-19. Clinical trials are needed to confirm this.

© 2022 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Introduction

The severe acute respiratory syndrome like coronavirus (SARS-CoV-2) has infected over 190 million individuals globally and has caused more than 4.2 million deaths as of August 2021 [1]. SARS-CoV-2 infection has a broad clinical spectrum, ranging from asymptomatic or

* Corresponding author.

E-mail address: alessia.david09@imperial.ac.uk (A. David).

GenOMICC Investigators, ISARIC4C Investigators

mildly symptomatic, to a life-threatening presentation requiring admission to intensive care. Age, and to a much lesser extent male gender and various underlying clinical conditions, such as cardiovascular disease, obesity and diabetes, are known risk factors associated with an increased COVID-19 morbidity and mortality [2,3]. The role of an individual's genetic background has recently emerged as an additional, yet not clearly understood, risk factor for COVID-19 [4–6]. Rare genetic variants in genes involved in the regulation of type I interferon (IFN) immunity, including autosomal recessive IRF7 and IFNAR1 deficiencies, have been identified in patients with life-threatening COVID-19 [6]. Autoantibodies to type I IFNs also account for at least 10% of cases of critical COVID-19 pneumonia [7]. Genome-wide association studies (GWAS) have discovered genetic haplotypes spanning several genes that are associated with COVID-19 severity [3,4,8].

The transmembrane protease serine type 2 (TMPRSS2) protein plays a key role in coronavirus infections [9–11], including SARS-CoV-2, as it is required for priming the virus' spike (S) glycoprotein through its cleavage, thus facilitating endosome-independent entry into target cells [12,13]. TMPRSS2, which is part of the type 2 transmembrane serine protease (TTSP) family, is characterized by androgen receptor elements located upstream to its transcription site [14]. As well as cleaving and activating viral glycoproteins of coronaviruses and influenza A and B viruses [15], TMPRSS2 is subject to autocleavage, which results in the liberation of its soluble catalytic domain [16]. The conditions under which autocleavage of TMPRSS2 and other members of the TTSPs family occurs are yet to be elucidated.

TMPRSS2 is expressed in lung and bronchial cells [17], but also in the colon, stomach, pancreas, salivary glands and numerous other tissues [18]. Moreover, it is co-expressed in bronchial and lung cells with the angiotensin-converting enzyme 2 (ACE2) [17], which is the best described SARS-CoV-2 cellular receptor [19]. In the olfactory epithelium of mice, the expression of TMPRSS2, but not ACE2, appears to be age-related and greater in older compared to younger animals [20]. Similarly, a recent study showed that expression of TMPRSS2 in mouse and human lung tissue is also age-related [21]. Studies in TMPRSS2 knock out (KO) mice reported reduced SARS-CoV and MERS-CoV replication in the lungs compared to wild-type mice, and a reduced proinflammatory viral response, especially cytokine and chemokine release via the Toll-like receptor 3 pathway [22,23]. We have recently shown that TMPRSS2 expression permits cell surface entry of SARS-CoV-2, allowing the virus to bypass potent endosomal restriction factors [24]. In vitro studies have shown that TMPRSS2 inhibitors prevent primary airway cell and organoid infection by SARS-CoV and SARS-CoV-2 [25,24,26]. In animal studies, mice infected with SARS-CoV and treated with the serine protease inhibitor camostat mesilate had a high survival rate [27]. Recently, camostat mesilate (which, in Japan, is already approved for patients with chronic pancreatitis and postoperative reflux esophagitis) was shown to block SARS-CoV-2 lung cell infection in vitro [12,24]. Furthermore, camostat mesilate and its metabolite GBPA have been shown to block SARS-CoV-2 spread in human lung tissue ex vivo [28]. Several clinical trials using camostat in COVID-19 patients are currently underway [29].

In view of the data from animal models and cell-based studies supporting a protective role of a knock out TMPRSS2 on coronavirus infection (including SARS and MERS), we hypothesized that naturally-occurring TMPRSS2 genetic variants affecting the structure and function of the TMPRSS2 protein may modulate the severity of SARS-CoV-2 infection.

Methods

TMPRSS2 three-dimensional structure and variant analysis

The recently released 3D structure of TMPRSS2 (PDB: 7meq) was used to assess the impact missense variants. The Phyre homology modelling algorithm [30] was used to resolve missing amino acid

regions in the SRSC domain that were not experimentally solved (described in Supplementary material). The FASTA sequence of TMPRSS2 was obtained from the UniProt protein knowledge database [31] (UniProt Id O15393, corresponding to 492 amino acid transcript Ensembl ID ENST00000332149.10). The recently released AlphaFold model [32] was compared to the Phyre model. The impact of each missense variant on the TMPRSS2 protein structure was assessed by analysing the following 16 features, using our in-house algorithm Missense3D [33]: breakage of a disulfide bond, hydrogen bond or salt bridge, introduction of a buried proline, clash, introduction of hydrophilic residue, introduction of a buried charged residue, charge switch in a buried residue, alteration in secondary structure, replacement of a charged with uncharged buried residue, introduction of a disallowed phi/psi region, replacement of a buried glycine with any other residue, alteration in a cavity, replacement of *cis* proline, buried to exposed residue switch, replacement of a glycine located in a bend. In addition, we used the SIFT [34] and Polyphen2 [35] variant predictors, which mainly use evolutionary conservation to assess the effect of each variant. The effect of variant rs12329760 was further assessed using: i. CONDEL [36], which reports a weighted average of the scores from fatHMM and MutationAssessor, and ii. FoldX5 force field [37], which calculates the stability of a protein based on the estimation of its free energy. A $\Delta\Delta G > 0.5$ kcal/mol (calculated as: $\Delta\Delta G = \Delta G_{wt} - \Delta G_{mut}$) was predicted to have a destabilizing effect.

Participants

Genetics Of Mortality In Critical Care (GenOMICC) and the International Severe Acute Respiratory Infection Consortium (ISARIC) Coronavirus Clinical Characterisation Consortium (4C) (ISARIC 4C)

Cases: this cohort was established between March 2020 and July 2020 (first COVID-19 wave) and comprises of 2244 critically ill, hospitalized COVID-19 positive patients from 208 UK intensive care units (ICUs): 2109, patients were recruited as part of the GenOMICC project, and an additional 135 cases as part of the International Severe Acute Respiratory Infection Consortium (ISARIC) Coronavirus Clinical Characterisation Consortium (4C) study. The clinical characteristics and comorbidities of these patients have been extensively reported in Pairo-Castineira et al. [8]. Only unrelated individuals (up to 3rd degree, based on kinship analysis (King 2.1)) were included. Samples were excluded if the genotype-based sex inference did not match the reported sex, or if a XXY karyotype was present. Moreover, patients of mixed genetic ancestry, and from ancestry groups with small numbers of cases (such as North American Indian, $n = 13$) defined using admixture supervised mode with 1000 genomes as reference, were excluded.

Controls: ancestry-matched controls (ratio 1 case to 5 controls) without a positive COVID-19 test were obtained from the UK BioBank population study. COVID-19 test results in BioBank are obtained from Public Health England, Public Health Scotland and SAIL for English, Scottish and Welsh data, respectively. The vast majority of results are from nose/throat swabs analysed by PCR. For patients admitted to hospital, results can also be from samples obtained from the lower respiratory tract. Only unrelated individuals (up to 3rd degree) were included. Individuals with sex mismatch were excluded. For validation, 45,875 unrelated individuals of European ancestry from the 100,000 Genomes Project were used as an alternative control group.

DNA extraction, genotyping and quality control have been described in detail previously [8]. Genetic ancestry was inferred using ADMIXTURE and reference individuals from the 1000 Genomes project. Imputation was performed using the TOPMed reference panel.

Cells, pseudovirus and plasmid

Human embryonic kidney 293T cells (293Ts; ATCC) were maintained in Dulbecco's modified Eagle's medium (DMEM), 10% foetal

calf serum (FCS), 1% non-essential amino acids (NEAA), 1% penicillin-streptomycin (P/S). Human Caco-2 (ATCC HTB-37) and Calu-3 (ATCC HTB-55) were maintained in DMEM, 20% FCS, 1% NEAA, 1% P/S. All cell lines were maintained at 37 °C, 5% CO₂.

Lentiviral pseudotype production was performed as previously described [24][38]. Briefly, pseudovirus was generated by co-transfecting 293Ts with lentiviral packaging constructs pCSFLW (minimal HIV genome with firefly luciferase reporter), pCAGGs-GAGPOL (HIV packing proteins) and the relevant viral glycoprotein in pcDNA3.1 – either the G glycoprotein from Indiana vesiculovirus (VSV-G) or SARS-CoV-2 spike protein. Co-transfections were performed at a plasmid ratio of 1.5:1:1 for pCSFLW:GAGPOL:glycoprotein. Pseudovirus was harvested at 48 and 72 h post-transfection, pooled, filtered, then frozen down. ACE2 FLAG was used as previously described [24]. TMPRSS2 expression plasmid was a kind gift from Roger Reeves (Addgene plasmid #53,887; <http://n2t.net/addgene:53887>; RRID: Addgene_53,887) [39]. Non-cleavable ACE2-FLAG and TMPRSS2 mutants were generated by overlap extension PCR or site-directed mutagenesis.

Phenotypic assays

293Ts were co-transfected with FLAG-tagged, non-cleavable ACE2 and TMPRSS2, as previously described [24]. Briefly, confluent 10cm² dishes of 293T cells were co-transfected with 1 µg each of TMPRSS2 and ACE2-FLAG. 24 h later, cells were resuspended in fresh media and either spun down for lysis and western blot or added to 96 well plates along with pseudovirus. 24 h later, media was refreshed and a further 24 h later, cells were lysed with reporter lysis buffer (Promega), and luminescence (measured as relative luminescence units, RLU) was read on a FLUOstar Omega plate reader (BMF Labtech) using the Luciferase Assay System (Promega).

Cell pellets for western blot were lysed in RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM TRIS, pH 7.4) supplemented with an EDTA-free protease inhibitor cocktail tablet (Roche). Cell lysates were combined with 4x Laemmli buffer (Bio-Rad) with 10% β-mercaptoethanol and boiled for 5 min. Membranes were probed with mouse anti-tubulin (abcam; ab7291), rabbit anti-TMPRSS2 (abcam; ab92323) and/or mouse anti-FLAG (F1804, Sigma). Near infra-red (NIR) secondary antibodies, IRDye® 680RD Goat anti-mouse (abcam; ab216776) and IRDye® 800CW Goat anti-rabbit (abcam; ab216773) were subsequently used. Blots were imaged using the Odyssey Imaging System (LI-COR Biosciences). Densitometry was performed using ImageJ.

Reagent, cell lines and antibody validation

All reagents, cell lines and antibodies used in this study are commercially available and validation data are available on the manufacturers' websites.

Statistical analysis

Sample size: Critically ill Covid-19 patients, $n = 2,244$; random controls matched by ancestry from UK Biobank, $n = 11,220$. The sample size was determined pragmatically by the number of cases recruited during the first wave of the outbreak in the UK (as described in [8]). No randomization was performed. Blinding was not used in this study because the exposure (genotype) and outcome (ICU admission) are objective. Confounding was controlled by the use of covariates: age, sex, deprivation score and genetic ancestry [8]. The association between the TMPRSS2 rs12329760 variant and COVID-19 severity was assessed using logistic regression. Genetic associations in the GenOMICC/ISARIC 4C cohort were analysed as previously described [8]. Briefly, logistic regression with additive and recessive models was performed in PLINKv1.9, adjusting for sex, age,

mean-centred age-squared, top 10 principal components (principal component analysis [PCA] performed to adjust for population stratification) and deprivation index decile based on UK postcode. Each major ancestry group alternative in the 100,000 Genomes control group was performed with mixed model association tests in SAIGE (v0.39) [40], including age, sex, age-squared, age-sex interaction and the first 20 principal components as covariates. Trans-ethnic meta-analysis of GenOMICC data for different ancestries was performed by METAL using an inverse-variance weighted method and the P-value for heterogeneity was calculated with Cochran's Q-test for heterogeneity implemented in the same software [41].

Additional publicly available genetic data were obtained from the COVID-19 Host Genetics Initiative meta-analyses, release 6 (June 15, 2021) [42]. The COVID-19 Host genetics initiative classifies COVID-19 severity according to the use of invasive and non-invasive ventilation during hospital admission. Here we report the four different phenotype comparisons:

- A2: 8,779 critically ill confirmed cases (inclusion criteria: hospitalized for COVID-19 and either death or on respiratory support including intubation, CPAP, BiPAP, continue external negative pressure, Optiflow/very high flow Positive End Expiratory Pressure Oxygen) versus 1,001,875 population controls,
- B1: 14,480 hospitalised cases versus 73,191 non-hospitalised cases,
- B2 24,274 hospitalised cases versus 2,061,529 population controls, and
- C2: 112,443 COVID-19 cases of unspecified severity versus 2,473,889 population controls.

Analyses used all data with the exclusion of the 23&Me study, for which full data were not publicly available. Meta-analysis in all cases was performed using a fixed effect, inverse variance-weighted model, either as a trans-ethnic meta-analysis or subsetted by ancestry group.

Data are presented as mean ± standard deviation. Log-normality was assessed using the Shapiro-Wilk test and QQ plot. A two-tailed Student's *t*-test was used to compare the means of two groups. One-way ANOVA was used to compare the means of more than two groups.

Colocalisation analysis

Colocalisation analysis for genetic associations was performed by an Approximate Bayes Factor approach using the package *coloc* version 5, in R 4.1.0 [43]. Summary statistics (beta and variance) were from GWAS data [8] and from lung eQTL data from GTEx v8 [44], in individuals of European ancestry. To reduce the likelihood of violation of the single causal variant assumption arising from multiple independent association signals, the analysis was restricted to a region extending to 5 kb upstream and downstream of the TMPRSS2 gene. With the assumption that exactly one measured SNP in the region was causative for each trait, SNP-level priors (p_1 and p_2) of $1/(n \text{ SNPs})$ were used for the probability of association with each individual trait, with an arbitrary prior of $0.1 \times p_1$ for p_2 , the SNP-level prior probability of association with both traits. Sensitivity analysis was performed to assess the impact of prior selection, comparing the selected priors to the more stringent default priors (10^{-4} for p_1 and p_2 , 10^{-5} for p_{12}), and varying the p_{12} range from p_1 to $p_1 \times p_2$.

Ethics

Research ethics committees (Scotland 15/SS/0110, England, Wales and Northern Ireland: 19/WM/0247). Current and previous versions of the study protocol are available at genomicc.org/protocol. All participants gave informed consent.

Role of funders

The Wellcome Trust, UKRI, MRC/UKRI, Howard Hughes Medical Institute, Rockefeller University, St. Giles Foundation, Fisher centre for Alzheimer's Research Foundation, Meyer Foundation, Square Foundation, Grandir Fonds de solidarité pour l'enfance, SCOR Corporate Foundation for Science, Institut National de la Santé et de la Recherche Médicale (INSERM), University of Paris, National Institutes of Health, French Foundation for Medical Research, FRM and French National Research Agency (ANR) GENCOVID, Agence Nationale de la Recherche, Health Data Research UK and BBSRC provided funding to support the salaries of the authors but had no role in the design, data collection, analysis, interpretation of the results or writing of the report. The content of this publication is solely the responsibility of the authors.

Results

We extracted 377 TMPRSS2 genetic variants reported as loss of function (LoF), missense or inframe and indel in the database of population genetic variations GnomAD (v2.1.1). All variants passed quality filters in GnomAD. Nine variants flagged by GnomAD as carrying dubious annotation or quality were excluded. Forty variants were loss of function or indel and the remaining 328 were missense. All variants except for one (rs12329760 C>T) were very rare (MAF <0.001). We focused on missense variants and studied the evolutionary conservation of TMPRSS2 amino acids and the impact of amino acid substitution on TMPRSS2 protein structure (described in Methods). The first experimental structure of TMPRSS2 was released in the public domain in June 2021. Although the trypsin domain was well resolved, several unstructured loops were present in the SRCR domain. We, therefore, used homology modelling and the Phyre2 server to model the missing regions of SRCR (Fig. 1). For completeness, a comparison between the Phyre2 model and the recently released (August 2021) AlphaFold [32] model is presented in Figure S1 (Phyre2 versus AlphaFold model: root mean square deviation [RMSD] 0.5 Å). We identified the chemical and physical bonds that stabilize the TMPRSS2 structure (i.e. hydrogen bonds, cysteine and salt bridges, as detailed in the Methods) and are affected by amino acid substitutions naturally occurring in the human population. A total of 137 variants were predicted damaging to the structure and/or function of TMPRSS2. Of these, 136 variants are extremely rare in the human population, with an average minor allele frequency (MAF) of 9.67×10^{-6} (cumulative MAF of 7.3×10^{-4}) and are, therefore, unlikely to be of use as a marker of COVID-19 infection severity in the general population. The remaining variant, rs12329760, (NC_000021.9:g.41480570 C>T [GRCh38.p13], p.Val160Met on Ensembl transcript ENST00000332149.5 and Val197Met on the Ensembl transcript ENST00000398585.3) is predicted damaging and causes the substitution of an evolutionary conserved valine to methionine (Figure S2). Overall, the minor allele frequency (MAF) of this variant is 0.25 in the human population, with 6.7% homozygous individuals (9,587 T/T homozygotes out of 141,456 individuals sequenced as part of the GnomAD project). Under Hardy-Weinberg equilibrium and a MAF of 0.25, it is expected that 37% of individuals will be heterozygous for this variant. The MAF of rs12329760 T varies according to ethnicities and ranges from 0.14 in Ashkenazi Jewish to 0.38 in East Asian populations (0.15 in Latino, 0.23 in non-Finnish Europeans, 0.25 in South Asians, 0.29 in African/African Americans and 0.37 in Finnish Europeans). This highly conserved valine occurs in the scavenger receptor cysteine-rich (SRCR) domain, whose function within TMPRSS2 is still not fully understood, although a role in ligand and/or protein interaction has been proposed [45]. Indeed, this domain is present in several proteins involved in host defence, such as CD5, CD6 and Complement factor I [46,47].

We first analysed the relation between TMPRSS2 rs12329760 and life-threatening SARS-CoV-2 infection in 2,244 critically ill, hospitalized, COVID-19 positive patients from 208 UK intensive care units (ICUs) (Table 1) recruited as part of the GenOMICC (genomicc.org) and ISARIC 4C (isaric4c.net) projects. These patients were representative of critically ill patients with COVID-19 in the UK population during the first Sars-Co-V2 outbreak of 2020[8]. Patients were treated in intensive care units (ICU/ITU) because of their propensity to critical respiratory failure due to COVID-19. Within the GenOMICC cohort ($n = 2,109$), mean age was 57.3 ± 12.1 , 624 (30%) patients were females, and 396(19%) had comorbidities; 1,557 (74%) required invasive ventilation and 459 (22%) died within 60 days. Within the ISARIC 4C cohort ($n = 135$), mean age was 57.3 ± 2.9 , 46 (34%) were females, and 40 (30%) had comorbidities; 25 (19%) required invasive ventilation and 22 (16%) died within 60 days, as described in[8]. 11,220 ancestry-matched individuals without a COVID-19-positive PCR test from the UK BioBank, acted as controls. Under an additive model, we found that the minor T allele of rs12329760 was significantly associated with a protective effect against severe COVID-19 in individuals of European ancestry (1,676 cases, 8,379 controls) with an OR of 0.87 (95%CI:0.79–0.97, $p = 0.01$). A protective effect was also observed in individuals of East Asian ancestry (149 cases, 745 controls; OR 0.64, 95%CI:0.43–0.95, $p = 0.03$). Similar effect sizes were observed in South Asians and Africans, but did not reach statistical significance, most likely as a result of the small sample size (Fig. 2). We further confirmed this protective effect on a trans-ethnic meta-analysis, using the entire cohort of 2,244 patients (OR 0.84, 95%CI:0.77–0.93, $p = 5.8 \times 10^{-4}$, Fig. 2, panel A). A heterogeneity analysis suggested that the T allele has a similar effect across different ethnicities ($p = 0.47$). To ascertain that this association was not an artefact due to population bias in the UK BioBank controls, the results from the European cohort were confirmed on an independent control population (45,875 unrelated individuals of European ancestry from 100 K Genomes [48]: OR 0.89, 95%CI:0.81–0.99, $p = 0.02$). Under a recessive model (i.e. individuals homozygous for the T allele), the trans-ethnic meta-analysis on 2,244 critically ill COVID-19 patients estimated an OR of 0.65 for TT homozygotes (95%CI:0.50–0.84, $p = 1.3 \times 10^{-3}$). In subset analyses, the OR was estimated at 0.70 (95%CI:0.52–0.95, $p = 0.024$) in Europeans, and 0.28 (95%CI:0.09–0.82, $p = 0.019$) in East Asians versus their corresponding ancestry-matched controls (Fig. 2, panel B).

To assess whether the rs12329760 could be a proxy for an association with a nearby expression quantitative trait locus (eQTL), colocalisation analysis was performed to compare the GWAS signal at the locus to eQTL associations for TMPRSS2 and neighbouring gene MX1 in GTex version 8 [44] (Figure S3 A-C), using an Approximate Bayes Factor approach [43]. Under an assumption of a single causal variant within the locus for each trait, the posterior probability of a common causal variant was 1.1% for TMPRSS2 expression and 2.0% for MX1 expression, compared to posterior probabilities of 67% and 42% respectively for independent associations. Sensitivity analysis showed that the analysis was robust to choice of prior probabilities: more stringent software-default single-trait priors increased the posterior probabilities of null or single-trait-only association hypotheses, but had little impact on the colocalisation probability (1.1% for TMPRSS2 and 0.2% for MX1); varying prior probability for colocalisation (Figure S3 D-F) had an impact only when approaching the prior for single-trait associations, and did not result in posterior probabilities for colocalisation exceeding those for separate associations. Although independent contributions from multiple variants towards the genetic association cannot be excluded, this indicates that any genetic association between rs12329760 and severe COVID-19 is unlikely to be attributable to linkage disequilibrium with an eQTL and, thus, modification of protein function is more likely.

For additional corroboration of the genetic signal, we investigated the results of large GWAS meta-analyses performed in the context of

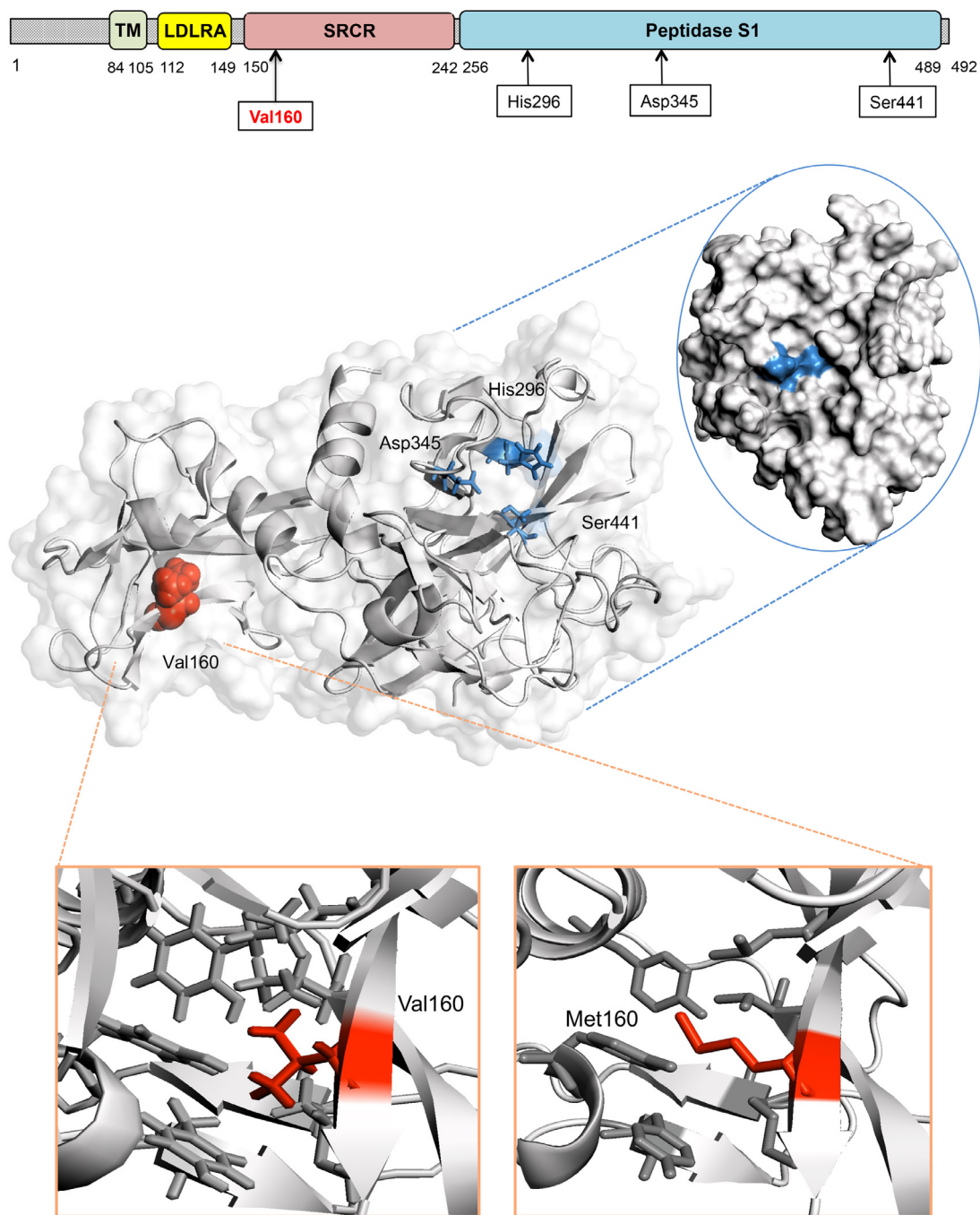


Fig. 1. The TMPRSS2 protein and the p.Val160Met variant

The TMPRSS2 protein is composed of a cytoplasmic region (residues 1–84), a transmembrane region (TM, residues 85–105) and an extracellular region (residues 106–492). The latter is composed of three domains: the LDLR class A (residues 112–149), the scavenger receptor cysteine-rich domain (SRCR) (residues 150–242) and the Peptidase S1 (residues 256–489), which contains the protease active site: residues His296, Asp345 and Ser441. The three-dimensional structure of the extracellular region residues 145–491 corresponding to domains SRCR-2 (in green) and Peptidase S1 (in blue) is presented. Valine 160 (Val 160, depicted as a red sphere on the cartoon), which harbours variant p.Val160Met, occurs in the SRCR domain and spatially far from the TMPRSS2 catalytic site (mapped onto the surface of TMPRSS2).

the COVID-19 Host Genetics Initiative (COVID-19hg, available at <https://www.COVID-19hg.org/>, release 6, June 2021) [42]. Compared with the general population, in a trans-ethnic meta-analysis, the minor T allele of rs12329760 was associated with a significantly protective effect against severe COVID-19 (patients requiring hospitalization for COVID-19): 24,274 cases versus 2,061,529 controls; OR 0.95, 95%CI:0.92–0.97, $p = 4.72 \times 10^{-6}$. In ancestry-specific subgroup analyses, this effect was significant for a European population (OR 0.94, 95%CI:0.91–0.96, $p = 5.66 \times 10^{-6}$), but not for individuals of African, Hispanic-American, or admixed African/Hispanic/American

ancestry; however, lower sample sizes for these groups limited study power, and subgroup analyses were not available for Asian populations. The protective effect was particularly evident in confirmed, critically ill cases (8,779 cases versus 1,001,875 population controls; OR 0.91, 95%CI:0.87–0.95, $p = 8.18 \times 10^{-6}$). Furthermore, the rs12329760 T allele was associated with reduced risk of hospitalisation after confirmed infection (14,480 hospitalised versus 73,191 cases not requiring hospitalization within 21 days after the test): OR 0.96, 95%CI:0.92–0.99, $p = 0.012$. Finally, there was no significant difference ($p = 0.056$) in the prevalence of the T allele between the

Table 1
Characteristics of 2,244 GenOMICC/ISARIC patients and 11,220 BioBank controls included in the study.

Patient characteristics	Cases (n = 2,244)	Controls (n = 11,220)
		missing data
Females, n. (%)	670[30]	6,075[54]
Age (years), mean ± SD	57.3 ± 11.6	66.1 ± 8.0
Invasive ventilation, n. (%)	1,582 (70.50)	66 (2.94)
Death, n. (%)	481 (21.44)	368 (16.40)
ancestry		
European, n. (MAF)	1,676 (0.20)	8,380 (0.23)
South Asian, n. (MAF)	237 (0.21)	1,185 (0.24)
African, n. (MAF)	182 (0.26)	910 (0.29)
East Asian, n. (MAF)	149 (0.28)	745 (0.38)

MAF, minor allele frequency; n.a., not available.

general population (n = 2,473,889) and pooled individuals with a laboratory-confirmed SARS-CoV-2 infection (including hospitalized and life-threatening COVID-19 cases from the meta-analyses previously described) or with a self-reported or physician-confirmed COVID diagnosis (total n = 112,443 cases, Fig. 2, panel A).

Although these meta-analyses include UK Biobank data and all except the hospitalised versus non-hospitalised comparison include data from the GenOMICC/ISARIC 4C cohort, and thus do not provide completely independent replication, these cohorts only comprise less than 25% of the total cases, limiting the impact of this single study on the overall results. These data therefore provide further support for our hypothesis that the TMPRSS2 rs12329760 variant has a protective effect against severe and/or life-threatening COVID-19. However, studies examining the prevalence of this variant in SARS-CoV-2 infected asymptomatic or pauci-symptomatic individuals are needed to ascertain its protective effect against mild viral infection.

To investigate the phenotypic effect of the TMPRSS2 V160M variant, we co-transfected 293Ts cells, which we have previously confirmed that they do not endogenously express ACE2 or TMPRSS2 [24], with ACE2 and either TMPRSS2 wild type (TMPRSS2_{WT}) or

V160M (TMPRSS2_{V160M}), as previously described [24]. We and others previously observed that co-expression of TMPRSS2 and ACE2 results in rapid cleavage of ACE2. We, therefore, used a mutant ACE2 that is more poorly degraded by TMPRSS2[49]. Two additional TMPRSS2 variants were included as controls: the catalytically inactive S441A (TMPRSS2_{S441A}) and the catalytically active R255Q (TMPRSS2_{R255Q}), that is unable to autocleave[16]. First, we investigated the autocleavage pattern of the different TMPRSS2 variants. The N-terminal membrane-bound part of TMPRSS2 can exist as different cleaved intermediates: a full-length uncleaved form of approximately 55 kDa, a partially cleaved form, and a fully cleaved form of 20 kDa. The latter is the product of TMPRSS2 autocleavage at arginine 255, which results in the liberation of the catalytically active protease domain in the extracellular space, leaving a small transmembrane N-terminal domain [16]. Wild type TMPRSS2 is expressed as roughly equal amounts of full-length and fully cleaved forms, with a small amount of partially cleaved product. As expected, the catalytically inactive TMPRSS2_{S441A} and the non-autocleavable TMPRSS2_{R255Q} resulted in only the full-length TMPRSS2 being expressed. However, TMPRSS2_{V160M} resulted in a significantly higher proportion of full-length (55 kDa), and a significantly lower proportion of fully cleaved protein (20 kDa) (p < 0.05, Student's t-test). This difference was clear across a range of TMPRSS2 concentrations, with TMPRSS2 showing a concentration-dependant autocleavage phenotype: the higher the concentration of TMPRSS2, the higher the amount of autocleavage. Overall, these data suggest the V160M substitution exerts a partial inhibitory effect on the proteolytic autocleavage of TMPRSS2 (see Fig. 3A-D, Supplementary Figure S4).

Subsequently, we investigated the effect of TMPRSS2_{V160M} on promoting viral entry, using a previously described SARS-CoV-2 pseudovirus entry assay [24]. Pseudovirus expressing the glycoprotein from the vesicular stomatitis virus (VSV-G) was used as a control, as this virus enters cells in a TMPRSS2-independent manner [24]. Briefly, cells co-transfected with ACE2 and TMPRSS2 wild type or variants were incubated with the pseudovirus (as described in [24,38]) and, after 48 h, luminescence was measured. TMPRSS2_{WT} enhanced viral

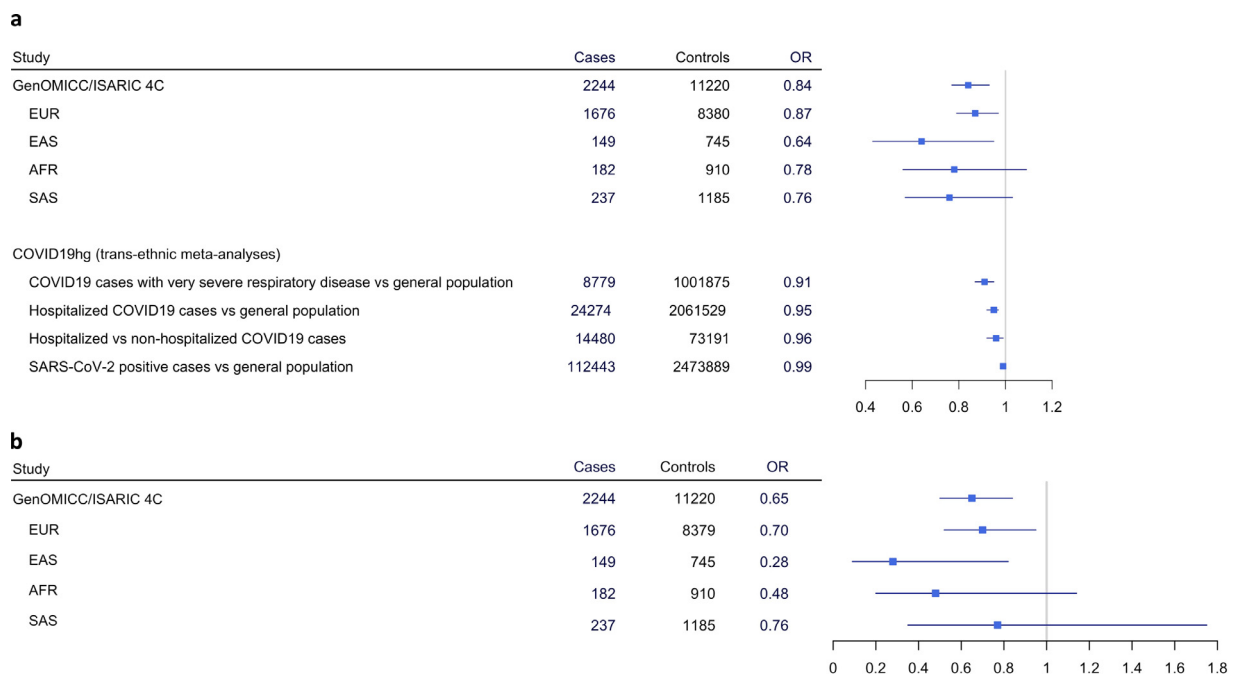


Fig. 2. Association of TMPRSS2 rs12329760 to COVID-19 severity

Results are presented for the additive (a) and the recessive (b) model using different COVID-19-positive patient cohorts. The results from large GWAS meta-analyses performed as part of the COVID-19hg initiative (1) are also shown.

OR, odds ratio; EUR, European; EAS, East Asian; AFR, African; SAS, South Asian.

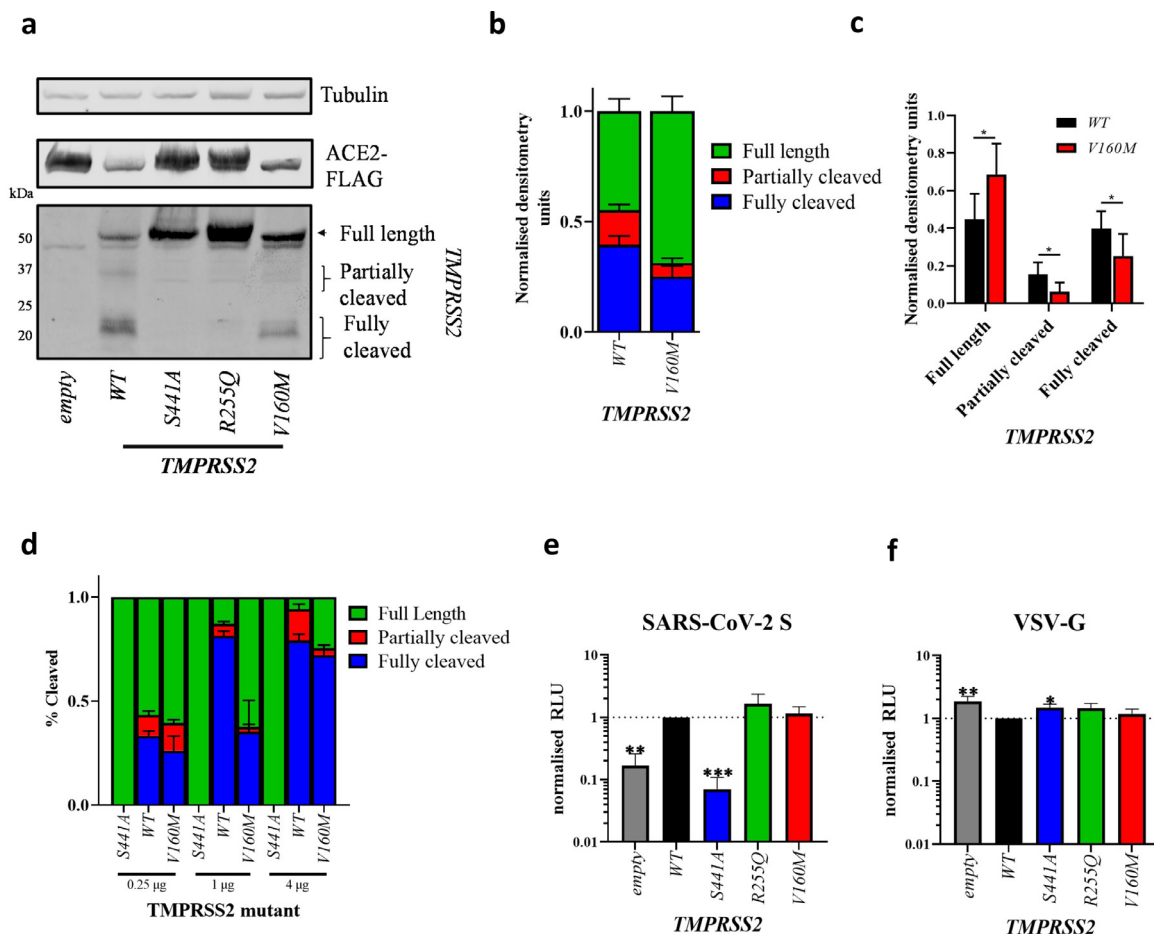


Fig. 3. Phenotypic impact of the TMPRSS2 V160M variant on autocleavage and SARS-CoV-2 spike-mediated entry; (A) Western blot analysis of TMPRSS2 autocleavage after expression in HEK 293Ts. (B,C, D) densitometry was determined in ImageJ and shows mean±standard deviation from N = 6 (B,C) or N = 3 (D) independent repeats. Statistics determined by two-tailed Student's *t*-test.

Entry of lentiviral pseudotypes expressing (E) SARS-CoV-2 spike glycoprotein or (F) Vesicular stomatitis virus glycoprotein (VSV-G) into HEK 293Ts co-expressing ACE2-FLAG and either empty vector or TMPRSS2 variants. Data shows mean±standard deviation of 3 independent repeats from different weeks, normalised to WT TMPRSS2. (E,F) Statistics determined by one-way ANOVA with multiple comparisons against WT on Log-transformed data (after determining log normality by the Shapiro-Wilk test and QQ plot). Values in μ g indicate the amount of TMPRSS2 plasmid added to each condition. RLU, relative luminescence units.

*, 0.05 \geq *P* > 0.01; **, 0.01 \geq *P* > 0.001, ***, 0.001 \geq *P*.

entry by ~5-fold compared to empty vector, while the catalytically dead TMPRSS2_{S441A} showed no enhancement (Fig. 3E). The non-autocleavable mutant TMPRSS2_{R255Q} showed similar enhancement, suggesting that autocleavage is dispensable for optimal TMPRSS2-mediated enhancement. TMPRSS2_{V160M} showed no significant difference in viral entry compared to the TMPRSS2_{WT}. Overall, the expression of catalytically active TMPRSS2 proteins only slightly inhibited VSV-G mediated entry (Fig. 3F).

The partial inhibitory effect exerted by the V160M variant on the proteolytic autocleavage of TMPRSS2 resulted in a far greater proportion of uncleaved, surface-expressed TMPRSS2_{V160M} compared to TMPRSS2_{WT}. We compared this autocleavage seen in transfected 293T cells, to that seen in several epithelial cell lines that naturally express endogenous ACE2 and TMPRSS2 [24]: the human lung cell line, Calu-3, and the human colorectal adenocarcinoma cell line, Caco-2, both of which are extensively used for SARS-CoV-2 research. Interestingly, no fully cleaved TMPRSS2 could be detected as opposed to 293T cells, while both cell lines expressed mostly full length or partially cleaved TMPRSS2. This again suggests that the high levels of autocleavage seen in 293T cells may be, in part, an artefact of overexpression (Supplementary Figure S5). Therefore, we re-assessed whether TMPRSS2_{V160M} affects SARS-CoV-2 S-expressing pseudovirus entry by using the double mutant TMPRSS2_{R255Q/V160M} (which does not autocleave and is, therefore, more similar to endogenous

TMPRSS2 in Calu-3 and Caco-2 cells), to control for protein cell-surface expression. Under these conditions, and across a range of plasmid titrations of both TMPRSS2 mutants and ACE2, TMPRSS2_{R255Q/V160M} showed a significantly reduced ability to promote SARS-CoV-2 S-expressing pseudovirus compared to TMPRSS2_{R255Q} alone, despite equal protein expression (Fig. 4A,C,D,F). Again, TMPRSS2_{R255Q/V160M} had no effect on VSV-G-mediated entry (Fig. 4B,E).

Discussion

Overall, our results suggest that the rs12329760 C>T variant results in a moderately less catalytically active TMPRSS2, which is less able to autocleave and prime the SARS-CoV-2 spike protein. This may explain the protective effect against life-threatening COVID-19 observed in our cohort of patients admitted to ICU, compared to the general population. Such an effect was more prominent in homozygotes (recessive model) for the rs12329760 C>T in whom a 30% (OR 0.70) risk reduction was observed. Unfortunately, we did not have samples from asymptomatic/pauci symptomatic patients, but data from COVID-19hg meta-analyses appear to suggest that the rs12329760 variant has no protective effect against SARS-CoV-2 infection per se.

The allele frequency of TMPRSS2 rs12329760 (data from GnomAD population database) varies across different populations and is higher

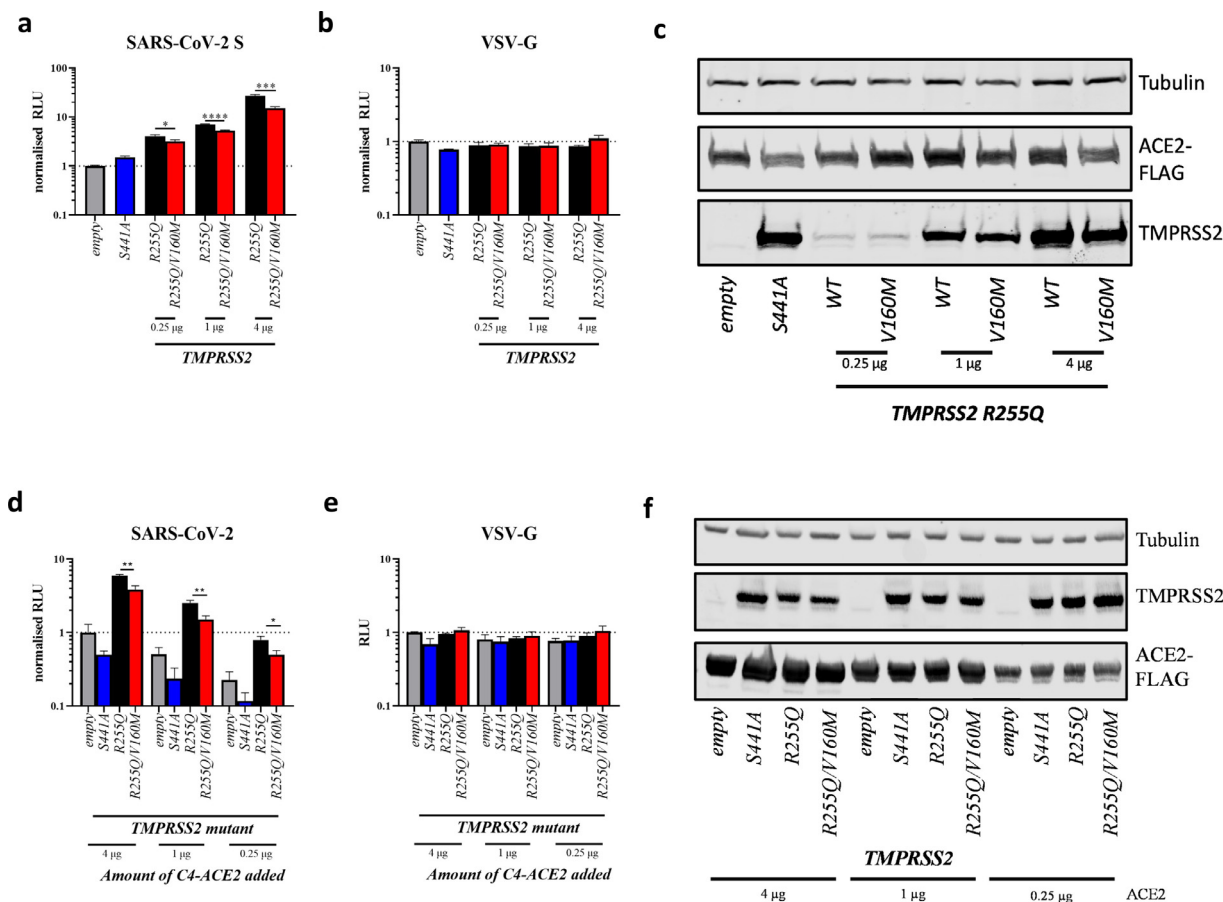


Fig. 4. Phenotypic impact of the TMPRSS2 non-autocleavable version of the V160M variant on SARS-CoV-2 spike-mediated entry

Entry of lentiviral pseudotypes expressing (A,D) SARS-CoV-2 spike glycoprotein or (B,E) vesicular stomatitis virus glycoprotein (VSV-G) into HEK 293Ts co-expressing ACE2-FLAG and either empty vector or TMPRSS2 variants. Data shows mean±standard deviation of 3 independent repeats with different weeks, normalised to empty vector. A-C shows titrations of mutant TMPRSS2 with constant ACE2 expression, while D-E show titrations of ACE2 with constant levels of TMPRSS2 expressed.

(A,B,D,E) statistics determined by two-tailed Student's *t*-test. (C,F) Western blot analysis of TMPRSS2 autocleavage mutant (R255Q) titration with or without the V160M substitution. µg values indicate the amount of TMPRSS2 or ACE2 plasmid added to each condition. RLU, relative luminescence units.

*, 0.05 ≥ *P* > 0.01; **, 0.01 ≥ *P* > 0.001, ***, 0.001 ≥ *P*.

in East Asian and Finnish individuals (MAF 0.38 and 0.37, respectively) compared to south Asians (MAF 0.25) and non-Finnish Europeans (MAF 0.23). The lowest frequency of the T allele is reported in Latino and Jewish-Ashkenazi individuals (MAF 0.15). However, in our study, the small sample size of populations of non-European ancestry does not allow conclusions on the effect size of TMPRSS2 rs12329760 in different ethnicities. Genotyping of the TMPRSS2 rs12329760 variant on large COVID19 cohorts of patients of non-European genetic ancestry is, therefore, needed to assess its role in determining the differences in the severity of COVID-19 across various populations (e.g. between East Asia and Europe [50]). Indeed, a recent study showed a lower T allele frequency in a small cohort of Chinese patients with life-threatening COVID-19 compared to the population frequency [51]. Although the differences in the proportion of SARS-CoV-2 patients who develop severe COVID-19 across different populations [50] are more likely to be explained by social behaviour, public health measures to curb outbreaks, exposure to other viruses and immunological factors, human genetic variation across different populations may also contribute to the observed differences.

The pharmacological inhibition of TMPRSS2 using serine protease inhibitors, such as camostat and nafamostat, has been proposed as a pharmacological treatment of COVID-19 patients. In vitro [12] and animal studies have demonstrated that camostat can block viral entry (reviewed in [52]), and initial reports on the repurposing of camostat in COVID-19 patients have provided promising results [53]. However, a recently completed clinical trial using camostat in patients

hospitalized for severe COVID-19, did not demonstrate a significant reduction in time-to-clinical improvement compared to placebo [54]. As the authors suggest, these patients were likely to have passed the most active stage of viral replication at the time of treatment and were in the hyper-inflammatory stage of the COVID-19, thus possibly explaining the lack of camostat efficacy. Several additional clinical trials of camostat in COVID-19 are currently underway [29]. Recently, the placebo-controlled phase III trial conducted in Japan on pauci symptomatic COVID-19 patients administering camostat mesilate 600 mg 4 times a day did not meet its primary end point of time to negative Sars-CoV-2 test [55], however data on secondary end points, such as progression to severe or life-threatening COVID-19, are still not publicly available.

Very little is known on TMPRSS2 and further extensive in vitro and in vivo studies on its pathophysiology are necessary. Since the beginning of the COVID-19 pandemic, the interest in TMPRSS2 has focused only on its role as a serine protease involved in the activation of the SARS-CoV-2 spike protein. However, as a soluble protease, TMPRSS2 may have additional substrates, and in vitro studies have demonstrated that PAR2 is one of these substrates [56,57]. PAR2 is expressed in several tissues, including lung, vascular endothelial and vascular smooth muscle cells [58,59] and its protease-mediated activation promotes inflammation by inducing prostaglandin synthesis and cytokine production in the lungs and other organs [60–64]. An intriguing hypothesis is that, similar to other soluble serine proteases, such as the human airway trypsin-like protease HAT (also known as

TMPRSS11D), the soluble wild type TMPRSS2 protease may also have a role in promoting inflammation in the lungs and other tissues.

Since May 2020, when we reported the TMPRSS2 variant rs12329760 as possibly damaging to protein structure/function and raised the possibility that it could partly explain host susceptibility to COVID-19 severity [65], several other studies have also supported this hypothesis [66–71]. In this study we have confirmed our initial hypothesis and provided a mechanistic effect to explain how this variant may contribute to the host susceptibility to severe COVID-19.

As previously discussed, one limitation of our study was the lack of access to a cohort of asymptomatic/pauci symptomatic COVID-19 patients. In the absence of such a cohort, we considered the general population as a good proxy and used this for comparison with COVID-19 severe cases. Indeed, a recent systematic review and meta-analysis shows that one third of COVID-19 positive cases do not develop symptoms [72]. When well-characterized cohorts of asymptomatic/pauci symptomatic COVID-19 patients become available, it will be possible to further investigate the role of TMPRSS2 variant rs12329760 on Sars-Co-V2 infection. Another limitation of this study is that we did not directly validate our results in endogenously expressing cell lines, such as like Calu-3, as this would require gene editing the endogenous TMPRSS2. Calu-3 cells are extremely slow growing and highly resistant to single cell cloning, thus making this cell line not particularly suitable for gene editing.

In conclusion, the T allele of the common TMPRSS2 variant rs12329760 confers a reduced risk of severe COVID-19. Similar to what observed in the TMPRSS2 KO mouse, the Val160Met substitution, which exerts a partial inhibitory effect on the proteolytic auto-cleavage of TMPRSS2 and the priming of the SARS-CoV-2 spike protein, is associated with a milder COVID-19 infection compared to the wild type. Differences in population frequency of this genetic variant may contribute to the reported variability in COVID-19 severity across various ethnicities and studies on large COVID-19 cohorts of patients of non-European genetic ancestry are needed to clarify this. Further studies are needed to assess the expression of TMPRSS2 across different age groups; indeed a reduced TMPRSS2 expression in younger compared to older individuals, as observed in mice and in preliminary human studies, could help explain age-related differences in COVID-19 morbidity. Moreover, TMPRSS2 could be a viable drug target in COVID-19 patients, and camostat mesilate, or other novel TMPRSS2 inhibitors, may have a role in the treatment of COVID-19. Clinical trials are needed to confirm this.

Funding

Wellcome Trust, BBSRC, UKRI Future Leader's Fellowship, Health Data Research UK

AD and NP were supported by the Wellcome Trust (grants 104,955/Z/14/Z, 218,242/Z/19/Z and 211,496/Z/18/Z) and TK by the BBSRC (grants BB/P011705/1 and BB/P023959/1), VSS is supported by UKRI Future Leader's Fellowship (MR/S032304/1), J-LC is supported by Howard Hughes Medical Institute, Rockefeller University, St. Giles Foundation, Fisher centre for Alzheimer's Research Foundation, Meyer Foundation, Square Foundation, Grandir - Fonds de solidarité pour l'enfance, SCOR Corporate Foundation for Science, Institut National de la Santé et de la Recherche Médicale (INSERM), University of Paris, National Institutes of Health (R01AI088364), French Foundation for Medical Research (EQU201903007798), FRM and French National Research Agency (ANR) GENCOVID project (ANR-20-COVI-0003); LA is supported by the Agence Nationale de la Recherche (ANR-10-IAHU-01, ANR-10-LABX-62-IBEID), TPP and WSB are supported by BBSRC grants BB/R013071/1 and BBSRC and the G2P-UK National Virology consortium (funded by MRC/UKRI, grant ref: MR/W005611/1), AT was supported by Roslin Institute Strategic Programme Grants from the BBSRC (BBS/E/D/10,002,070 and BBS/E/

D/30,002,275) and Health Data Research UK (references HDR-9004 and HDR-9003).

Author contributions

NP, EP-C, AT and JKB contributed to population data analysis. AD, TK and MJES contributed to 3D modelling and structural analysis. TPP and WSB contributed to laboratory work. NP, EP-C, TPP and AD contributed to data analysis. NP, TPP, WSB, AD contributed to study design. NP, TPP, EP-C, AC, VS-S, J-LC, LA, WSB, JKB, MJES and AD contributed to interpretation of findings and manuscript preparation. AD conceived the study, contributed to study coordination and wrote the first draft of the manuscript. All authors approved the final version of the manuscript.

Data availability

Full summary-level data in support of the findings of this study are available for download from <https://genomicc.org/data>. Individual level data can be analysed by qualified researchers in the ISARIC 4C/GenOMICC data analysis platform by application at <https://genomicc.org/data>. BioBank data and Genomics England data are available to registered researchers at <https://www.ukbiobank.ac.uk/> and <https://www.genomicsengland.co.uk/>. The COVID-19 Host Genetics Initiative2 (COVID-19hg) summary statistics are available at <https://www.COVID-19hg.org/>.

Supplementary Information is available for this paper.

Declaration of Competing Interest

Dr. David reports grants from Wellcome Trust during the conduct of the study; Dr. Parkinson reports grants from Wellcome Trust during the conduct of the study; Dr. Peacock reports grants from MRC/UKRI, grants from BBSRC during the conduct of the study; Dr. Pairo-Castineira has nothing to disclose. Dr. Khanna reports grants from BBSRC during the conduct of the study; Dr. Cobat has nothing to disclose. Dr. Tenesa reports grants from BBSRC, grants from Health Data Research UK during the conduct of the study. Dr. Sancho-Shimizu reports grants from UKRI Future Leader's Fellowship during the conduct of the study; Dr. Casanova reports other from Howard Hughes Medical Institute, other from Rockefeller University, other from St. Giles Foundation, other from Fisher centre for Alzheimer's Research Foundation, other from Meyer Foundation, other from Square Foundation, other from Grandir - Fonds de solidarité pour l'enfance, other from Institut National de la Santé et de la Recherche Médicale (INSERM), other from University of Paris, other from National Institutes of Health (NIH), other from French Foundation for Medical Research (FRM), other from FRM and French National Research Agency (ANR) GENCOVID project during the conduct of the study; Dr. Abel reports other from Agence Nationale de la Recherche during the conduct of the study; Dr. Barclay reports grants from BBSRC during the conduct of the study; Dr. Baillie has nothing to disclose. Dr. Sternberg reports grants from Wellcome Trust, grants from BBSRC, during the conduct of the study.

Acknowledgement

This research was conducted using the UK BioBank Resource under project 788

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.retram.2022.103333](https://doi.org/10.1016/j.retram.2022.103333).

References

- [1] WHO. WHO Coronavirus (COVID-19) Dashboard [Internet]. Available from: <https://covid19.who.int/>
- [2] Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J Infect* 2020;81(2):e16–25.
- [3] Zhang Q, Bastard P, Bolze A, Jouanguy E, Zhang S.-Y., C.O.V.I.D. Human Genetic Effort, et al. Life-threatening COVID-19: defective interferons unleash excessive inflammation. *Med (N Y)*. 2020;1(1):14–20.
- [4] Severe Covid-19 GWAS Group, D Ellinghaus, 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 15.
- [5] Wu BB, Gu D-Z, Yu JN, Yang J, Shen WQ. Association between ABO blood groups and COVID-19 infection, severity and demise: a systematic review and meta-analysis. *Infect Genet Evol* 2020;84:104485 Oct.
- [6] Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, et al. Inborn errors of type 1 IFN immunity in patients with life-threatening COVID-19. *Science* 2020;370(6515) Oct 23.
- [7] Bastard P, Rosen LB, Zhang Q, Michailidis E, Hoffmann H-H, Zhang Y, et al. Autoantibodies against type 1 IFNs in patients with life-threatening COVID-19. *Science* 2020;370(6515) Oct 23.
- [8] 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 Dec 11.
- [9] Glowacka I, Bertram S, Müller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol* 2011;85(9):4122–34 May.
- [10] Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M, Taguchi F. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J Virol* 2010;84(24):12658–64 Dec.
- [11] Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol* 2011;85(2):873–82 Jan.
- [12] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020;181(2):271–80 16e8.
- [13] Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci USA*. 2020 May 6.
- [14] Lin B, Ferguson C, White JT, Wang S, Vessella R, True LD, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. *Cancer Res* 1999;59(17):4180–4 Sep 1.
- [15] Limburg H, Harbig A, Bestle D, Stein DA, Moulton HM, Jaeger J, et al. TMPRSS2 is the major activating protease of influenza A virus in primary human airway cells and influenza B virus in human type II pneumocytes. *J Virol* 2019;93(21) Nov 1.
- [16] Afar DE, Vivanco I, Hubert RS, Kuo J, Chen E, Saffran DC, et al. Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. *Cancer Res* 2001;61(4):1686–92 Feb 15.
- [17] Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J* 2020:e105114 Apr 4.
- [18] Vaarala MH, Porvari KS, Kellokumpu S, Kyllönen AP, Vihko PT. Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues. *J Pathol* 2001;193(1):134–40 Jan.
- [19] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579(7798):270–3.
- [20] Bilinska K, Jakubowska P, Von Bartheld CS, Butowt R. Expression of the SARS-CoV-2 ENTRY PROTEINS, ACE2 and TMPRSS2, in cells of the olfactory epithelium: identification of cell types and trends with age. *ACS Chem Neurosci* 2020;11(11):1555–62 03.
- [21] Schuler BA, Habermann AC, Plosa EJ, Taylor CJ, Jetter C, Negretti NM, et al. Age-determined expression of priming protease TMPRSS2 and localization of SARS-CoV-2 in lung epithelium. *J Clin Invest* 2021;131(1) Jan 4.
- [22] Iwata-Yoshikawa N, Okamura T, Shimizu Y, Hasegawa H, Takeda M, Nagata N. TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J Virol* 2019;93(6):93.
- [23] Li F, Han M, Dai P, Xu W, He J, Tao X, et al. Distinct mechanisms for TMPRSS2 expression explain organ-specific inhibition of SARS-CoV-2 infection by enzalutamide. *Nat Commun* 2021;12(1):866. Feb 8.
- [24] Peacock TP, Goldhill DH, Zhou J, Bailion L, Frise R, Swann OC, et al. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* 2021 Apr 27.
- [25] Kawase M, Shirato K, van der Hoek L, Taguchi F, Matsuyama S. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *J Virol* 2012;86(12):6537–45 Jun.
- [26] Mykityn AZ, Breugem TI, Riesebosch S, Schipper D, van den Doel PB, Rottier RJ, et al. SARS-CoV-2 entry into human airway organoids is serine protease-mediated and facilitated by the multibasic cleavage site. *Elife* 2021;10 Jan 4.
- [27] Zhou Y, Vedantham P, Lu K, Agudelo J, Carrion R, Nunneley JW, et al. Protease inhibitors targeting coronavirus and filovirus entry. *Antiviral Res* 2015;116:76–84 Apr.
- [28] Hoffmann M, Hofmann-Winkler H, Smith JC, Krüger N, Arora P, Sørensen LK, et al. Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity. *EBioMedicine* 2021:103255 Mar 3.
- [29] ClinicalTrials.gov [Internet]. Available from: [https://clinicaltrials.gov/ct2/results?cond=&term=camostat&cntry=&state=&city=&dist=](https://clinicaltrials.gov/ct2/results?cond=&term=camostat&cntry=&state=&city=&dist=&rank=&term=camostat&cntry=&state=&city=&dist=)
- [30] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015;10(6):845–58 Jun.
- [31] The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45(D1):D158–69 Jan 4.
- [32] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 Jul 15.
- [33] Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J Mol Biol* 2019;431(11):2197–212 May 17.
- [34] Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc* 2016;11(1):1–9 Jan.
- [35] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7(4):248–9 Apr.
- [36] González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score. *Condel. Am J Hum Genet* 2011;88(4):440–9 Apr 8.
- [37] Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res* 2005;33:W382–8 Jul 1 Web Server issue.
- [38] McKay PF, Hu K, Blakney AK, Samnuan K, Brown JC, Penn R, et al. Self-amplifying RNA SARS-CoV-2 lipid nanoparticle vaccine candidate induces high neutralizing antibody titers in mice. *Nat Commun* 2020;11(1):3523. Jul 9.
- [39] Edie S, Zaghoul NA, Leitch CC, Klindedinst DK, Lebron J, Thole JF, et al. Survey of human chromosome 21 gene expression effects on early development in Danio rerio. *G3 (Bethesda)* 2018;8(7):2215–23 Jul 2.
- [40] Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* 2018;50(9):1335–41.
- [41] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010;26(17):2190–1 Sep 1.
- [42] COVID-19 host genetics initiative. Mapping the human genetic architecture of COVID-19. *Nature* 2021 Jul 8.
- [43] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10(5):e1004383 May.
- [44] GTEx Consortium. Laboratory, data analysis & coordinating center (LDACC)—analysis working group, statistical methods groups—analysis working group, enhancing GTEx (eGTEx) groups, NIH common fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–13 Oct 11.
- [45] Aruffo A, Bowen MA, Patel DD, Haynes BF, Starling GC, Gebe JA, et al. CD6-ligand interactions: a paradigm for SRCR domain function? *Immunol Today* 1997;18(10):498–504 Oct.
- [46] Freeman M, Ashkenas J, Rees DJ, Kingsley DM, Copeland NG, Jenkins NA, et al. An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. *Proc Natl Acad Sci USA* 1990;87(22):8810–4 Nov.
- [47] Resnick D, Pearson A, Krieger M. The SRCR superfamily: a family reminiscent of the Ig superfamily. *Trends Biochem Sci* 1994;19(1):5–8 Jan.
- [48] The National Genomic Research Library v5.1, Genomics England [Internet]. Available from: <https://doi.org/10.6084/m9.figshare.4530893.v6> 2020.
- [49] Heurich A, Hofmann-Winkler H, Gierer S, Liepold T, Jahn O, Pöhlmann S. TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J Virol* 2014;88(2):1293–307 Jan.
- [50] Yamamoto N, Bauer G. Apparent difference in fatalities between Central Europe and East Asia due to SARS-CoV-2 and COVID-19: four hypotheses for possible explanation. *Med Hypotheses* 2020;144:110160 Aug 5.
- [51] 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. Nov 10.
- [52] Breining P, Frølund AL, Højten JF, Gunst JD, Staerke NB, Saedder E, et al. Camostat mesylate against SARS-CoV-2 and COVID-19-Rationale, dosing and safety. *Basic Clin Pharmacol Toxicol* 2021;128(2):204–12 Feb.
- [53] Sakr Y, Bensasi H, Taha A, Bauer M, Ismail K, the UAE-Jena Research Group. Camostat mesylate therapy in critically ill patients with COVID-19 pneumonia. *Intensive Care Med* 2021;47(6):707–9 Jun.
- [54] Gunst JD, Staerke NB, Pahus MH, Kristensen LH, Bodilsen J, Lohse N, et al. Efficacy of the TMPRSS2 inhibitor camostat mesilate in patients hospitalized with COVID-19—a double-blind randomized controlled trial. *EclinicalMedicine* 2021;35:100849 May.
- [55] Foipan dropped as potential coronavirus treatment [Internet]. Available from: <https://www.thepharmaletter.com/in-brief/brief-foipan-dropped-as-potential-coronavirus-treatment>
- [56] Wilson S, Greer B, Hooper J, Zijlstra A, Walker B, Quigley J, et al. The membrane-anchored serine protease, TMPRSS2, activates PAR-2 in prostate cancer cells. *Biochem J* 2005;388(Pt 3):967–72.
- [57] Lam DK, Dang D, Flynn AN, Hardt M, Schmidt BL. TMPRSS2, a novel membrane-anchored mediator in cancer pain. *Pain* 2015;156(5):923–30 May.
- [58] D'Andrea MR, Derian CK, Leturcq D, Baker SM, Brunmark A, Ling P, et al. Characterization of protease-activated receptor-2 immunoreactivity in normal human tissues. *J Histochem Cytochem* 1998;46(2):157–64 Feb.

- [59] Cocks TM, Fong B, Chow JM, Anderson GP, Frauman AG, Goldie RG, et al. A protective role for protease-activated receptors in the airways. *Nature* 1999;398(6723):156–60 Mar 11.
- [60] Reed CE, Kita H. The role of protease activation of inflammation in allergic respiratory diseases. *J Allergy Clin Immunol* 2004;114(5):997–1008 Novquiz 1009.
- [61] Heuberger DM, Schuepbach RA. Protease-activated receptors (PARs): mechanisms of action and potential therapeutic modulators in PAR-driven inflammatory diseases. *Thromb J* [Internet]. 2019;17 Mar 29 [cited 2021 Feb 3] Available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6440139/>.
- [62] Su X, Camerer E, Hamilton JR, Coughlin SR, Matthay MA. Protease-activated receptor-2 activation induces acute lung inflammation by neuropeptide-dependent mechanisms. *J Immunol* 2005;175(4):2598–605 Aug 15.
- [63] Matsushima R, Takahashi A, Nakaya Y, Maezawa H, Miki M, Nakamura Y, et al. Human airway trypsin-like protease stimulates human bronchial fibroblast proliferation in a protease-activated receptor-2-dependent pathway. *Am J Physiol Lung Cell Mol Physiol* 2006;290(2):L385–95 Feb.
- [64] Chokki M, Yamamura S, Eguchi H, Masegi T, Horiuchi H, Tanabe H, et al. Human airway trypsin-like protease increases mucin gene expression in airway epithelial cells. *Am J Respir Cell Mol Biol* 2004;30(4):470–8 Apr.
- [65] David A, Khanna T, Beykou M, Hanna G, Sternberg MJ. Structure, function and variants analysis of the androgen-regulated TMPRSS2, a drug target candidate for COVID-19 infection. *BioRxiv* 2020 May.
- [66] Monticelli M, Mele BH, Andreotti G, Cubellis MV, Riccio G. Why does SARS-CoV-2 hit in different ways? Host genetic factors can influence the acquisition or the course of COVID-19. *Eur J Med Genet* 2021;64(6):104227 Jun.
- [67] Zarubin A, Stepanov V, Markov A, Kolesnikov N, Marusin A, Khitrinskaya I, et al. Structural variability, expression profile, and pharmacogenetic properties of TMPRSS2 gene as a potential target for COVID-19 therapy. *Genes (Basel)* 2020;12(1):E19. Dec 25.
- [68] Vargas-Alarcón G, Posadas-Sánchez R, Ramírez-Bello J. 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* 2020;260:118313 Nov 1.
- [69] Hou Y, Zhao J, Martin W, Kallianpur A, Chung MK, Jehi L, et al. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. *BMC Med* 2020;18(1):216. Jul 15.
- [70] Paniri A, Hosseini MM, Akhavan-Niaki H. First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. *J Biomol Struct Dyn* 2021;39(10):3576–93 Jul.
- [71] Senapati S, Kumar S, Singh AK, Banerjee P, Bhagavatula S. Assessment of risk conferred by coding and regulatory variations of TMPRSS2 and CD26 in susceptibility to SARS-CoV-2 infection in human. *J Genet* 2020;99(1):53.
- [72] Sah P, Fitzpatrick MC, Zimmer CF, Abdollahi E, Juden-Kelly L, Moghadas SM, et al. Asymptomatic SARS-CoV-2 infection: a systematic review and meta-analysis. *Proc Natl Acad Sci U S A*. 2021;118(34):e2109229118 Aug 24.