

# Variants in *ACE2* and *TMPRSS2* Genes Are Not Major Determinants of COVID-19 Severity in UK Biobank Subjects

David Curtis<sup>a, b</sup><sup>a</sup>UCL Genetics Institute, University College London, London, UK; <sup>b</sup>Centre for Psychiatry, Queen Mary University of London, London, UK

## Keywords

*ACE2* · *TMPRSS2* · COVID-19 · SARS-CoV-2

## Abstract

It is plausible that variants in the *ACE2* and *TMPRSS2* genes might contribute to variation in COVID-19 severity and that these could explain why some people become very unwell whereas most do not. Exome sequence data was obtained for 49,953 UK Biobank subjects, of whom 82 had tested positive for SARS-CoV-2 and could be presumed to have severe disease. A weighted burden analysis was carried out using SCOREASSOC to determine whether there were differences between these cases and the other sequenced subjects in the overall burden of rare, damaging variants in *ACE2* or *TMPRSS2*. There were no statistically significant differences in weighted burden scores between cases and controls for either gene. There were no individual DNA sequence variants with a markedly different frequency between cases and controls. Whether there are small effects on severity, or whether there might be rare variants with major effect sizes, would require studies in much larger samples. Genetic variants affecting the structure and function of the *ACE2* and *TMPRSS2* proteins are not the main explanation for why some people develop severe symptoms in response to infection with SARS-CoV-2. This research was conducted using the UK Biobank Resource.

© 2021 The Author(s)  
Published by S. Karger AG, Basel

## Introduction

There is wide variation in the severity of symptoms in patients infected with SARS-CoV-2, and there are reports in the UK that members of ethnic minorities are more severely affected. An obvious possible explanation for these findings would be that genetic polymorphisms affecting the structure or function of key proteins could influence host susceptibility and/or responses to infection. If these polymorphisms varied in frequency between different ethnic groups, this could contribute to differential outcomes.

Two key proteins involved in SARS-CoV-2 infective processes are *ACE2*, which is expressed on the cell surface and acts as a receptor for the viral S protein, and *TMPRSS2*, which cleaves the S protein to allow fusion of the viral and cellular membranes [1]. Variants in the genes coding for these proteins might contribute to different responses to infection.

A recent Italian study examining *ACE2* sequence variants in 131 COVID-19 patients and 258 controls reported that overall there was an excess of variants among controls ( $p = 0.029$ ) [2]. This result was partially driven by two common variants, Asn720Asp (rs41303171), which occurred in 2 cases and 11 controls, and Val749Val (rs35803318), which occurred in 5 cases and 25 controls. Another Italian study, using a different sample of 131 cases who tested positive for COVID-19, of whom 98 re-

quired ventilation, and 1,000 controls found that the cumulative frequency of variants was as expected from population frequencies and there was no association with severity [3].

Here, we present the results of a study comparing frequencies of variants in *ACE2* and *TMPRSS2* between cases with severe COVID-19 and controls.

## Methods

The COVID-19 results table was downloaded from UK Biobank on April 28, 2020. This contained results for 1,474 subjects who had undergone testing for SARS-CoV-2 infection between March 16 and April 14, 2020 [4]. During this period, testing in the UK was done almost exclusively on patients admitted to hospital with a clinical diagnosis of probable COVID-19, and thus patients testing positive can be assumed to have had severe disease because patients with milder symptoms were generally left at home. Of the subjects tested, 669 tested positive, meaning that they had at least one swab which demonstrated the presence of viral RNA at detectable levels, and of these 82 were exome sequenced. The proportion of infected subjects who require hospitalisation rises with age but is still only 0.18 for those aged 80 or over [5]. Thus, the subjects who tested positive could be regarded as cases with an unusually severe response to infection, whereas the subjects who tested negative or who were not tested could be regarded as unscreened controls, most of whom would not have severe symptoms even if infected. No attempt was made to discriminate between these subjects on other measures of severity, such as use of oxygen or admission to intensive care.

The exome sequence data consisted of the variant call files for 49,953 subjects who had undergone exome-sequencing and been genotyped using the GRCh38 assembly with coverage 20× at 94.6% of sites on average [6]. All variants were annotated using VEP, PolyPhen, and SIFT [7–9]. To obtain population principal components reflecting ancestry, version 1.90 beta of PLINK (<https://www.cog-genomics.org/plink2>) was run with the options `-- maf 0.1 -- pca header tabs -- make-rel` [10–12].

SCOREASSOC was then used to carry out a weighted burden analysis to test whether, in *ACE2* or *TMPRSS2*, sequence variants which were rarer and/or predicted to have more severe functional effects occurred more commonly in cases, that is, subjects who tested positive for SARS-CoV-2, than all the other sequenced subjects. All available variants in each gene were included in the analyses. As originally described, variants were weighted according to frequency so that rare variants were accorded 10 times the weight of common variants [13]. Variants were additionally weighted according to their functional annotation using the default weights provided with the GENEVARASSOC program, which was used to generate input files for weighted burden analysis by SCOREASSOC [13–15]. For example, a weight of 5 was assigned for a synonymous variant, 10 for a non-synonymous variant, and 20 for a stop-gained variant. Additionally, 10 was added to the weight if the PolyPhen annotation was possibly or probably damaging and also if the SIFT annotation was deleterious, meaning that a non-synonymous variant annotated as both damaging and deleterious would be assigned an overall weight of 30. *ACE2* is located on the X chro-

mosome and hemizygous males were treated as if they were homozygous for each variant, meaning that variant frequencies would be expected to be equal in males and females. Weighted burden testing using GENEVARASSOC and SCOREASSOC was carried out to see whether the overall burden of rare, functional variants differed between cases and controls using both *t* tests and likelihood ratio tests using ridge regression analysis incorporating the first 20 principal components, as described previously [15].

The two common variants referred to above, rs41303171 and rs35803318, had been genotyped in the whole UK Biobank sample, so their allele counts were compared between the 669 cases who had tested positive and all the remaining 487,708 subjects using the  $\chi^2$  test.

## Results

The genotype counts and frequencies of variants are presented in online supplementary Table 1 (see [www.karger.com/doi/10.1159/000515200](http://www.karger.com/doi/10.1159/000515200)), with variant positions and annotations redacted in order to preserve subject anonymity. There were 510 valid variants in *ACE2* and there was no tendency for the weighted burden scores to be different between cases (mean [SD] 24.4 [44.1]) and controls (22.6 [37.8]):  $t = 0.44$ , 49,951 df,  $p = 0.66$  and  $\chi^2 = 1.05$ , 1 df,  $p = 0.31$ . There were 658 valid variants in *TMPRSS2*, and although the weighted burden scores were lower in cases (65.9 [38.5]) than in controls (74.0 [48.9]), this difference was not statistically significant:  $t = -1.5$ , 49,951 df,  $p = 0.13$  and  $\chi^2 = 3.62$ , 1 df,  $p = 0.06$ . On visual inspection of the results there were no individual variants with markedly different frequencies between cases and controls. Of course, for both genes there were many rare variants which were observed in controls but not in cases, but this is as expected given the disparity in sample sizes.

With respect to the common variants which had been genotyped in the entire UK Biobank sample, the frequency of rs35803318 was 0.039 in cases and 0.044 in controls, and the frequency of rs41303171 was 0.025 in cases and 0.026 in controls. Neither of these differences was statistically significant.

## Discussion

Although the number of severely affected subjects who were sequenced is very small, it is nevertheless possible to draw some preliminary conclusions, and given the importance of the topic, it seems reasonable to communicate these findings. In general, the results are negative. It is not the case that a large proportion of severely affected subjects have a particular genetic variant in one of these genes which is relatively rare in the general population.

Nor is it the case that there is a common variant which confers strong protection against severe infection. It remains possible that there might be rare variants which have a major effect on risk in individual subjects, but such effects would only be detected with larger sample sizes.

The fact that the weighted burden scores were higher in controls than in cases is consistent with the hypothesis that rare genetic variants in *TMPRSS2* with functional effects disrupting functioning of the protein might be protective against severe infection. Although this is biologically plausible, it should be emphasised that the results obtained are not statistically significant. This could be investigated further by carrying out targeted sequencing of this gene in a sample of a few hundred severely affected subjects.

In conclusion, genetic variants affecting the structure and function of the ACE2 and *TMPRSS2* proteins are not the main explanation for why some people develop severe symptoms in response to infection with SARS-CoV-2.

## Acknowledgments

This research was conducted using the UK Biobank Resource. The author wishes to acknowledge the staff supporting the High Performance Computing Cluster, Computer Science Department, University College London.

## References

- 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 Apr;181(2):271–280.e8.
- Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, et al.; GEN-COVID Multicenter Study. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *Eur J Hum Genet*. 2020 Nov;28(11):1602–14.
- Novelli A, Biancolella M, Borgiani P, Cacciadiferro D, Colona VL, D'Apice MR, et al. Analysis of ACE2 genetic variants in 131 Italian SARS-CoV-2-positive patients. *Hum Genomics*. 2020 Sep;14(1):29.
- Armstrong J, Rudkin JK, Allen N, Crook DW, Wilson DJ, Wyllie DH, et al. Dynamic linkage of COVID-19 test results between Public Health England's Second Generation Surveillance System and UK Biobank. *Microb Genom*. 2020 Jul;6(7):mgen000397.
- Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. 2020 Jun;20(6):669–77.
- Van Hout CV, Tachmazidou I, Backman JD, Hoffman JX, Ye B, Pandey AK, et al. Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv*. 2019 Mar;572347.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016 Jun;17(1):122.
- Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013 Jan;Chapter 7:Unit7.20.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–81.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559–75.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015 Feb;4(1):7.
- Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al.; International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009 Aug;460(7256):748–52.
- Curtis D. A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. *Adv Appl Bioinform Chem*. 2012;5:1–9.
- Curtis D. Pathway analysis of whole exome sequence data provides further support for the involvement of histone modification in the aetiology of schizophrenia. *Psychiatr Genet*. 2016 Oct;26(5):223–7.
- Curtis D. A weighted burden test using logistic regression for integrated analysis of sequence variants, copy number variants and polygenic risk score. *Eur J Hum Genet*. 2019 Jan;27(1):114–24.

## Statement of Ethics

UK Biobank obtained ethics approval from the North West Multi-Centre Research Ethics Committee, which covers the UK (approval number: 11/NW/0382), and written informed consent from all participants. The UK Biobank approved application for use of the data (ID 51119). Analysis of the data was approved by the University College London Research Ethics Committee (approval number 11527/001).

## Conflict of Interest Statement

The author declares that he has no conflict of interest.

## Funding Sources

This work did not receive any external funding but was carried out in part using resources provided by BBSRC equipment grant BB/R01356X/1.

## Data Availability

The raw data is available on application from UK Biobank. Detailed results with unredacted variant counts cannot be made available because they might be used for subject identification.