

Outline

Abstract (submitted to conference)

SARS-CoV-2 is a highly infectious virus that is responsible for the COVID-19 global pandemic that swept the world in 2020. Disease outcomes range from asymptomatic to fatal. The virus initiates entry into host cells by the binding of its spike protein to the ACE2 receptor. Entry is finalized by the activation of spike glycoprotein by proteases including transmembrane protease, serine 2 (TMPRSS2) and FURIN which cleave the spike protein of the virus. Single nucleotide polymorphisms (SNPs) in TMPRSS2 may lead to functional changes which could underlie differences in disease severity. TMPRSS2 is also known to activate different respiratory illnesses including coronaviruses and influenza A (Shen et al., 2020). Previous studies have shown that knockout TMPRSS2 mice appeared healthy, experienced a decrease in viral spread within the respiratory system, and had a less severe immune response when infected with SARS-CoV and MERS-CoV (Baughn et al., 2020). Thus, we asked whether genetic variations in TMPRSS2 in humans lead to differences in infection rates or severity of disease symptoms of SARS-CoV-2. We examined the NCBI dbSNP database to identify SNPs in the TMPRSS2 gene. As of 10 December 2020, we found there were 11,023 intron variants, 393 missense variants, 186 synonymous variants, 3 in-frame insertion variants, 2 in-frame deletion variants, and 1 initiator codon variant reported. To narrow these down to 23 SNPs of interest, we first searched the ClinVar database to identify SNPs with general clinical significance, followed by searching the literature to determine SNPs specifically related to SARS-CoV-2 severity. One missense variant, rs12329760, results in an amino acid substitution, V160M, which has been predicted to alter TMPRSS2 function. A subset of these SNPs show differences in frequency in world populations, and we wondered if these SNPs had structural and functional consequences for the protein. A crystal structure of TMPRSS2 is not currently available. To visualize the structural consequences of amino acid substitutions, we performed homology modeling on TMPRSS2 (UniProt O15393) using the structure prediction software HHPred, RaptorX, and SwissModel based on the ~30% similarity to hepsin. The predicted structures of TMPRSS2 with various amino acid substitutions were then docked to the SARS-CoV-2 spike protein using I-TASSER and Haddock 2.4 to observe differences in binding interactions and therefore determine which sequence changes are predicted to alter binding interactions, potentially contributing to the wide variation of symptoms caused by COVID-19.

I. Introduction

1. Significance of SARS-CoV-2

- a. SARS-CoV-2 is a novel coronavirus that has infected over 100 million people and caused 2.1 million deaths worldwide as of January 2021 (Dong et al., 2020).
- b. There is currently no effective cure for SARS-CoV-2.

2. What is known about SARS-CoV-2

- a. Some patients present with severe symptoms that ultimately lead to death while others experience mild or no symptoms (Russo et al., 2020).
 - i. Those experiencing more severe symptoms are typically older or having pre-existing conditions. (Russo et al., 2020)
 - ii. Data from SARS-CoV-2 positive patients show that males are more likely to display severe symptoms, be hospitalized, and die from the disease than women (Mollica et al., 2020).
 - b. SARS-CoV-2 facilitates viral entry by first binding its spike protein to the ACE2 receptor, followed by the activation of the spike glycoprotein by proteases including TMPRSS2 (Hoffmann et al., 2020)
 - c. ACE2 is well understood to be important in healthy bodily functions, specifically through its role in the renin-angiotensin system (RAS) which regulates blood pressure (Baughn et al., 2020).
 - i. When knockout ACE2 mice were observed, they suffered lethal effects (Baughn et al., 2020)
3. What is known about TMPRSS2 that facilitates interest in its study
- d. TMPRSS2 facilitates the entry of H1N1 influenza virus, H7N9 influenza virus. Knockout TMPRSS2 mice were much less susceptible to infection by both of these viruses with no lethal side effects such as death, infertility, or sickness (Shen et al., 2020). TMPRSS2's protease activity is also crucial for activating SARS-CoV and MERS-CoV (Shen et al., 2020)
 - e. Heightened TMPRSS2 expression is linked to prostate cancer development and progression (Mollica et al., 2020).

- ii. Differences in docking interactions between TMPRSS2 and the spike protein of SARS-CoV-2 due to SNPs have been observed (Hussain et al., 2020)

II. Materials and Methods

1. Generating TMPRSS2 Gene Map and Model

- a. The NCBI dbSNP database and NCBI Genome data viewer were used to compose a gene map of TMPRSS2 detailing its location on the chromosome, the location of important transcriptional elements, and different isoforms.
- b. 3D models of TMPRSS2 were generated to understand the complete structure of the protease and which residues may be key to binding interactions with SARS-CoV-2 (Hussain et al., 2020).
- c. Modeling softwares I-Tasser, Swiss-Model, Raptor-X, HHPred, and PredMP were used to visualize TMPRSS2. FASTA sequence of TMPRSS2 was retrieved from Uniprot:O15393 and inputted into softwares (Ravikanth et al., 2020, Hussain et al., 2020).

2. Locating SNPs of interest

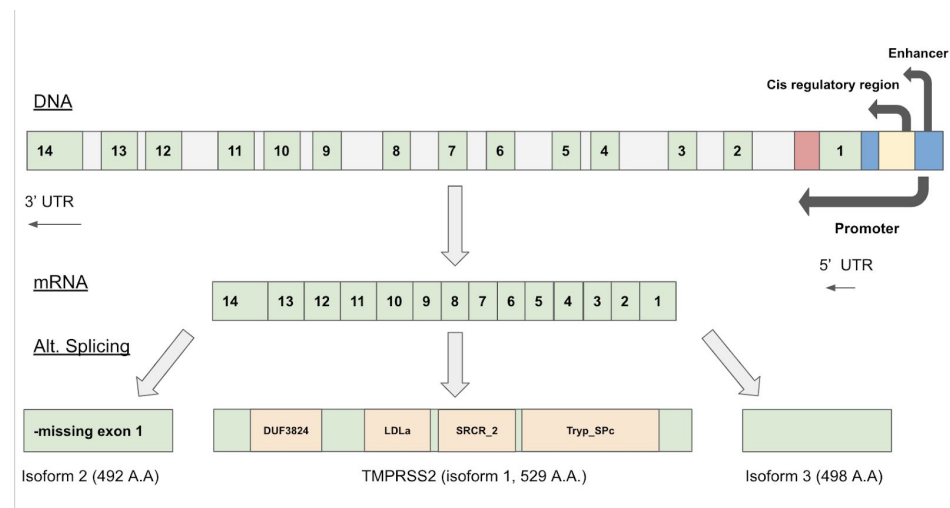
- a. NCBI dbSNP database and gnomAD browser were used to find all missense variants (323.2) of TMPRSS2.
- b. Missense SNPs were searched in the ClinVar database to identify those with published clinical significance.
- c. Literature search was conducted to confirm the clinical significance of SNPs identified through Clinvar and to identify other SNPs of interest.

- d. Population frequencies of SNPs were identified through the dbSNP database to discover which SNPs were relatively more prevalent (David et al., 2020).
3. Prediction of Effects of SNPs of interest
 - a. PolyPhen-2 and SIFT were used to predict the effects of SNPs on the structure of TMPRSS2 (Klaasen et al., 2020).
 - i. PolyPhen-2 predicts if amino acid substitutions will be damaging by scoring them on a scale of 0 to 1. Scores of 0-0.5= “Benign”, 0.5-0.9= “Possibly Damaging”, and 0.9-1.0= “Probably Damaging”.
 1. FASTA sequence of TMPRSS2 from Uniprot was inputted into software and amino acid substitution of each SNP was manually inputted
 - ii. SIFT predicts if amino acid substitutions will be damaging by scoring them on a scale to 0 to 1. Scores below 0.05 are marked “Deleterious” and scores above 0.05 are marked “Tolerated”.
 1. SNP identifiers were directly inputted into the software to retrieve results.
 4. Modelling SNPs of interest and their potential docking interactions.
 - a. Modeling software was used to generate a 3D visualization of the 18 variants of interest to observe any significant changes in structure.
 - i. Sequences of missense SNPS were generated to FASTA format to input into the software.

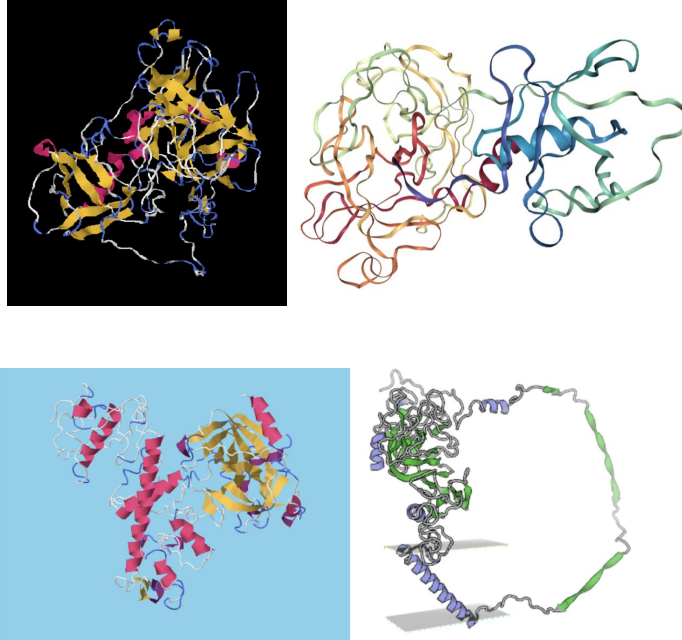
- b. Modeling software was used to visualize the docking interactions between each TMPRSS2 and the spike protein of SARS-CoV-2 using iC3nD.
 - i. Docking interactions were visualized between each SNP and the spike protein to observe any significant changes.
 - ii. TMPRSS2 multiple sequence alignment was generated.

III. Results

A. TMPRSS2 Gene Map and Model



1. Gene Map of TMPRSS2 located on 21q22.3 from 41,464,305bp - 41,508,158bp. There are 14 exons that can produce 3 different isoforms after splicing. Isoform 2 and 3 are not yet categorized, however isoform 1 is known to be expressed in viral target cells. The Domain of Unknown Function (DUF3824) is 48 amino acids long at 44-91 bp. The Low Density Lipoprotein Receptor Class A (LDLa) domain is 36 amino acids long at 150 bp-185 bp. The Scavenger receptor cysteine-rich (SRCR_2) domain is 90 amino acids long at 190-283 bp. The Trypsin-like serine protease (Tryp_SPC) domain is 232 amino acids long at 293-524 bp. The Cleavage site is 1 amino acid long at 293 bp.



2. 3D models of TMPRSS2 generated from I-Tasser, HHpred, RaptorX, and PredMP.

B. SNPs of Interest

1. Image of heat map of population frequencies for SNPs (to be constructed)
2. Graph plotting frequency vs. SNP # (to be constructed)

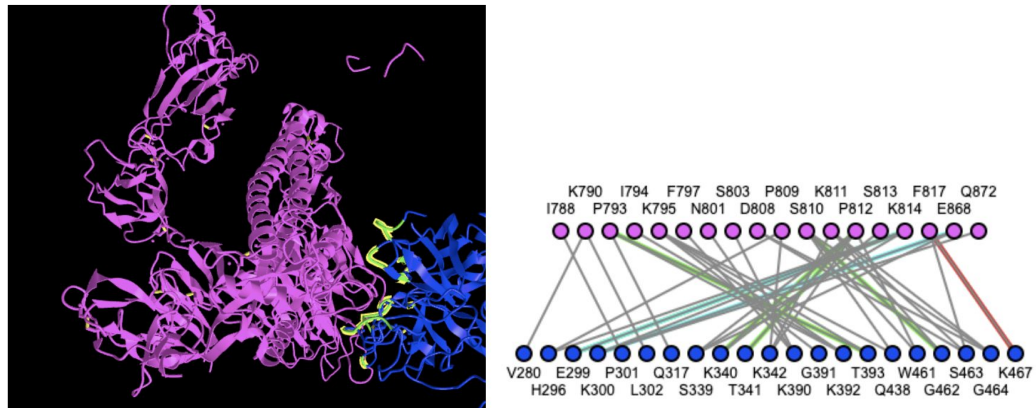
C. Predicting the Effect of SNPs

TMPRSS2 SNP Predictions				
rs Number	SIFT score	SIFT prediction	PolyPhen-2 score	PolyPhen-2 prediction
rs61735793	0.238	tolerated	0.015	Benign
rs75603675 G8V	0.201	tolerated	0.167	Benign
rs61735790	0.231	tolerated	0.033	Benign
rs12329760	0.009	deleterious	0.937	Probably Damaging
rs200291871	0.817	Tolerated	0.011	Benign
rs61735791	0.199	Tolerated	0.029	Benign
rs148125094	0.171	Tolerated	0.098	Benign
rs114363287	0.383	Tolerated	0.109	Benign
rs147711290 L128G	Not Found	-	0.920	Probably Damaging
rs147711290 L91P	0.005	Deleterious	1.000	Probably Damaging
rs147711290 L91R	Not Found	-	Not Found	-
rs150554820	0.004	Deleterious	0.549	Possibly Damaging
rs61735796	0.34	Tolerated	0.017	Benign
rs138651919	0.021	Deleterious	0.833	Possibly Damaging
rs61735795	0.551	Tolerated	0.086	Benign
rs142446494	0.015	Deleterious	0.294	Benign
rs201093031	1	Tolerated	0.00	Benign
rs768173297	Not Found	-	0.131	Benign

1. Table of TMPRSS2 SNP predictions using SIFT and PolyPhen-2 prediction softwares.

- a. 4 SNPs (rs12329760, rs147711290 L91P, rs150554820, rs138651919) were predicted to be damaging by both softwares. 1 SNP (rs147711290 L128G) was predicted to be damaging by PolyPhen-2 but was not found in SIFT. 1 SNP (rs142446494) was predicted to be deleterious by SIFT but benign by Poly-Phen2. 12 SNPs were predicted to be benign by both softwares and 1 SNP was not found in either software.

D. Modelling SNPs of interest and their potential docking interactions



1. Docking Interaction between TMPRSS2 and spike protein of SARS-CoV-2
2. Images of 3D models of SNPs (to be constructed)
3. Images of models of docking interactions that were shown to be potentially altered (to be constructed)

IV. Discussion

A. Final take away from models of SNPs and their potential docking interactions

1. Analyze the implications of the changes in binding interactions between TMPRSS2 and SARS-CoV-2 and how this may impact initial infection.
2. Discuss specific SNPs that were damaging and how they affected TMPRSS2's binding ability.
3. Comparison and evaluation of protein prediction and modeling softwares.

B. How findings relate to the literature

1. Aligns with current understanding that TMPRSS2 expression could play an important role in facilitating initial infection by SARS-CoV-2 (Shen et al., 2020, Paniri et al., 2020).
2. The importance of these docking interactions have been previously studied and align with our findings (Hussain et al., 2020).
 - a) rs12329760 was previously found to be damaging to TMPRSS2 and potentially disrupt binding to the spike protein (Ravikanth et al., 2020, Paniri et al., 2020).

C. How findings differ from the literature

1. Focused on prevalent SNPs rather than rare ones to overcome previous study's limitations (David et al., 2020).

D. Limitations

1. Accurate frequencies of SNPs are difficult to determine given that the whole population has not been sampled.
2. Accuracy of prediction models
3. Only missense SNPs were analyzed, yet other types of SNPs could have damaging effects on TMPRSS2.

E. Future directions

1. Looking at variants that could alter expression levels indirectly, such as intron or 3'UTR variants could offer a more complete picture of SNPs effect.
2. Exploring TMPRSS4 SNPs and their effect on binding interactions, as TMPRSS4 has been found to be very similar to TMPRSS2 and potentially increase viral entry (Zang et al., 2020).

References

- Baughn, L. B., Sharma, N., Elhaik, E., Sekulic, A., Bryce, A. H., & Fonseca, R. (2020). Targeting TMPRSS2 in SARS-CoV-2 Infection. *Mayo Clinic proceedings*, 95(9), 1989–1999. <https://doi.org/10.1016/j.mayocp.2020.06.018>
- Comparative Protein Structure Modeling Using MODELLER. Webb B, Sali A. *Curr Protoc Protein Sci*. 2016 Nov 1;86:2.9.1-2.9.37.
- David, A., Khanna, T., Beykou, M., Hanna, G., & Sternberg, M. J. (2020). Structure, function and variants analysis of the androgen-regulated TMPRSS2, a drug target candidate for COVID-19 infection. *bioRxiv*.
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*; published online Feb 19. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2), 271–280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>
- Hussain, M., Jabeen, N., Amanullah, A., Baig, A. A., Aziz, B., Shabbir, S., ... & Uddin, N. (2020). Molecular docking between human TMPRSS2 and SARS-CoV-2 spike protein: conformation and intermolecular interactions. *AIMS microbiology*, 6(3), 350.
- Klaassen, K., Stankovic, B., Zukic, B., Kotur, N., Gasic, V., Pavlovic, S., & Stojiljkovic, M.

- (2020). Functional prediction and comparative population analysis of variants in genes for proteases and innate immunity related to SARS-CoV-2 infection. *Infection, Genetics and Evolution*, 84, 104498.
- Mollica, V., Rizzo, A., & Massari, F. (2020). The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. *Future Medicine*, 16(27), 2029–2033.
<https://doi.org/10.2217/fon-2020-0571>
- Paniri, A., Hosseini, M. M., & Akhavan-Niaki, H. (2020). First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. *Journal of Biomolecular Structure and Dynamics*, (just-accepted), 1-18.
- Ravikanth V. et al. (2020) Genetic variants in TMPRSS2 and Structure of SARS-CoV-2 spike glycoprotein and TMPRSS2 complex. bioRxiv. doi:10.1101/2020.06.30.179663
- Russo, R., Andolfo, I., Lasorsa, V. A., Iolascon, A., & Capasso, M. (2020). Genetic analysis of the coronavirus SARS-CoV-2 host protease TMPRSS2 in different populations. *Frontiers in genetics*, 11, 872.
- Shen, L. W., Mao, H. J., Wu, Y. L., Tanaka, Y., & Zhang, W. (2017). TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. *Biochimie*, 142, 1-10.
- Zang, R., Gomez Castro, M. F., McCune, B. T., Zeng, Q., Rothlauf, P. W., Sonnek, N. M., Liu, Z., Brulois, K. F., Wang, X., Greenberg, H. B., Diamond, M. S., Ciorba, M. A., Whelan, S., & Ding, S. (2020). TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Science Immunology*, 5(47), eabc3582.

<https://doi.org/10.1126/sciimmunol.abc3582>

Zhou, L., Niu, Z., Jiang, X., Zhang, Z., Zheng, Y., Wang, Z., ... & Sun, Q. (2020). Systemic analysis of tissue cells potentially vulnerable to SARS-CoV-2 infection by the protein-validated single-cell RNA profiling of ACE2, TMPRSS2 and Furin proteases.

doi: <https://doi.org/10.1101/2020.04.06.028522>