Introduction

SARS-CoV-2 is a beta coronavirus responsible for the COVID-19 pandemic that has infected over 117 million people and resulted in over 2.6 million deaths as of March 2021 (Dong et al., 2020). While current vaccination efforts are becoming more successful in the prevention of initial infection, there is still a lack of therapeutics that efficiently cure active cases. Cases of COVID-19 have shown to differ drastically in their severity, with some patients appearing asymptomatic while others experience severe illness or even death (Russo et al., 2020). Epidemiologic studies have discovered numerous groups that are at-risk for developing severe cases of COVID-19, including those with older age or pre-existing conditions such as COPD or heart conditions (Baughn et al., 2020). There have also been numerous epidemiologic findings that males may present a heightened risk for severe disease and death from COVID-19, as data shows a higher percentage of men being hospitalized (men 60% and women 40%) and dying (men 62% and women 38%) of COVID-19 compared to their female counterparts (Mollica et al., 2020). While these observations have remained consistent throughout the pandemic, the underlying cause for the variation in pathogenicity among different populations remains unknown.

The SARS-CoV-2 virus facilitates entry into the host cell by use of its Spike (S) Protein. The S protein is a trimer that consists of two different subunits, S1 and S2. The S1 subunit binds to the host's angiotensin converting enzyme 2 (ACE2) receptor to enable the initial attachment of SARS-CoV-2 to the surface of the host cell (Hoffmann et al., 2020) . The S2 subunit acts to finalize the entry of the virus by fusing the viral and host cell membrane together. Importantly, this fusion requires priming of the S protein, which involves cleavage of its subunits by different proteases. The S1/S2 site is cleaved by the protease furin, while the S2' site is cleaved by the transmembrane protease serine 2 (TMPRSS2) (Hoffmann et al., 2020) . This cleavage is key to viral transmission, as it is what activates the S protein and allows it to finalize the entry of the virus into the host cell.

While ACE2 has garnered much interest as a potential therapeutic target against SARS-CoV-2 due to its critical role in initial infection, further understanding of ACE2's overall physiological role suggests that this is not a viable solution. The ACE2 receptor is required for the initial recognition and attachment of the virus onto the host cell, therefore many studies have looked into the potential role that ACE2 expression may play in the variation of disease severity of COVID-19 (Baughn et al., 2020). Specifically, researchers have focused on the overexpression of ACE2 in different body tissues as well as different populations to observe whether a correlation exists between heightened expression of ACE2 and an increase in COVID-19 severity (Baughn et al., 2020). While overexpression of ACE2 may play an important role in infection with the disease, ACE2 plays a critical role in the renin angiotensin system that renders it necessary for normal heart functioning (Baughn et al., 2020). A study conducted to determine the importance of ACE2 in heart health observed that when ACE2 expression was knocked out in mice, they suffered severe impairment of heart function and blood pressure regulation (Crackower et al., 2002). Therefore, directly targeting the expression of ACE2 is unlikely to be a feasible task due to the risk of adverse effects on heart health.

Similar studies investigating TMPRSS2's physiological function have shown that this protease is less likely to play a critical role in everyday functioning, making it a more viable potential target against SARS-CoV-2. Much of the physiological function of TMPRSS2 remains unknown, although its overexpression has been linked to the development and progression of prostate cancer and it has the highest expression in the luminal cells of the prostate epithelium (Mollica et al., 2020). It is also known to be involved in the spread of other respiratory viruses, including H1N1 and H7N9 influenza virus, as well as the previous coronaviruses, MERS-CoV and SARS-CoV (Zmora et al., 2015). A study conducted by Zmora et al. analyzed isoform 1 and 2 of TMPRSS2, finding that it activates these viruses through its proteolytic activity in a similar way to which it acts on the S protein of SARS-CoV-2. By using cell lines, researchers were able to infect cells with plasmids carrying SARS-CoV and influenza viruses and monitor the colocalization of TMPRSS2 with the viral strains and the resulting proteolytic cleavage that occured to activate the glycoproteins of the corresponding viruses (Zmora et al., 2015). Their findings align with what is known about TMPRSS2's mechanism of activating the S protein of the SARS-CoV-2 virus, and strengthens the interest in TMPRSS2 as a potential target against infection due to its role in infection with various viruses.

Unlike ACE2, knocking out TMPRSS2 expression in mice did not result in severely negative phenotypes. The mice presented no negative phenotypes, which included death, infertility, and sickness (Shen et al., 2020). This suggests that TMPRSS2 likely does not play a critical role in key processes of growth, development, or reproduction, making it a much safer therapeutic target than ACE2 (Baughn et al., 2020). Furthermore, when these knockout mice were infected with SARS-CoV and MERS-CoV, they experienced decreased viral spread and less severe immune responses (Baughn et al., 2020). This highlights the importance of TMPRSS2 expression for the infectivity and pathogenicity of coronaviruses and is cause for interest in differing expression levels of TMPRSS2 as means to prevent or lessen COVID-19 pathogenicity. Altering the expression of TMPRSS2 has been accomplished safely before, as TMPRSS2 inhibitors have been clinically approved and used in treatment for other diseases, such as prostate cancer, and have been shown to be effective in disrupting the viral entry of SARS-CoV-2 (Hoffmann et al., 2020). Therefore, alternate expression levels of TMPRSS2 may play a key role in pathogenicity of COVID-19, and its expression may be able to be safely manipulated as a means of treatment for the disease.

There is growing evidence that SNPs of TMPRSS2 may be involved in the development of certain diseases, such as prostate and breast cancer, by altering expression levels of the protease (Paniri et al., 2020). Given that TMPRSS2 expression has been linked to various diseases and is understood to play a critical role in initiating infection of SARS-CoV-2, this study was interested in the role that TMPRSS2 SNPs may play in altering the expression of TMPRSS2 to affect the development of COVID-19. Previous studies have identified SNPs that may be destabilizing to the protease and therefore hinder its ability to perform binding interactions important to initiating infection (David et al., 2020, Klassen et al., 2020, Paniri et al., 2020). Specifically, rs12329760 has been observed in multiple studies to destabilize the structure of TMPRSS2 and could potentially affect its ability to bind to the spike protein of SARS-CoV-2 (Painir et al., 2020). Missense SNPs have the potential to alter an amino acid residue that are on or near interaction sites between TMPRSS2 and SARS-CoV-2 S protein. Therefore, we hypothesize that missense SNPs have the potential to alter the form and function of TMPRSS2 in a way that affects its ability to bind the S protein of SARS-CoV-2.

Given that the crystallized structure of TMPRSS2 has not yet been uncovered, a bioinformatic analysis of TMPRSS2 and their SNPs using protein prediction software will provide the best knowledge of how variations on structure may alter binding ability. Such prediction software will allow us to predict how damaging SNPs may be to the structure of TMPRSS2 as well as visualize the different residues on the protein. Docking Software will allow us to physically visualize the interaction between the S protein and each SNP to better understand the structural implications on binding. To better understand the impact of these SNPs on the general population, SNP frequencies will be analyzed to determine if certain variations have higher prevalence in certain ethnic populations. Previous studies have been limited by their ability to identify SNPs that are not rare in the population. Therefore, this study aims to identify prevalent SNPs in order to determine whether structural variations in TMPRSS2 could potentially underlie variations in the pathogenicity of COVID-19.

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

Crackower, M. A., Sarao, R., Oudit, G. Y., Yagil, C., Kozieradzki, I., Scanga, S. E., ... & Penninger, J. M. (2002). Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature*, *417*(6891), 822-828.

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.

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.

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.

Zmora, P., Moldenhauer, A. S., Hofmann-Winkler, H., & Pöhlmann, S. (2015). TMPRSS2 Isoform 1 Activates Respiratory Viruses and Is Expressed in Viral Target Cells. *PloS one*, *10*(9), e0138380. https://doi.org/10.1371/journal.pone.0138380