Introduction

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) induced a global pandemic of the disease Covid-19. The virus belongs to the coronavirus family, characteristic for it's notable spike proteins along the surface of the virus. Previous coronavirus outbreaks include the 2002 SARS-CoV epidemic, which resulted in the deaths of around 700 people, and the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 which resulted in 886 deaths. However, SARS-CoV-2 is notable for its ability to produce a range of cases from asymptomatic to severe and its high infectivity rate. In the United States alone, over 500,000 deaths have been recorded and over 30 million cases have been documented. The spike protein is a trimer that contains two subunits: S1 and S2. The S1 subunit binds to host angiotensin converting enzyme 2 (ACE2), while the S2 site binds the viral and cell membranes together. Host proteases cleave the spike protein to allow entry of the virus into the host. Furin cleaves the spike (S) protein at the S1/S2 site, while transmembrane protease serine 2 (TMPRSS2) cleaves the S protein at the S2' site (Hoffman et al., 2020). This cleavage is important for the complete entry of SARS-CoV-2 into the host and unique to SARS-CoV-2.

TMPRSS2 is a serine protease that has no determined physiological function; it belongs to the family Type II transmembrane serine proteases (TTSPs) and more specifically, to the subfamily of hepsin/transmembrane protease/serine (Bugge, Antalis, & Wu, 2009). It has been expressed in luminal cells of the prostate epithelium and is positively regulated by androgens. Its enhancer is located 13 kb upstream of the transcription start site that is important for androgran regulation (Clinckemalie et al., 2013). Functional research also indicates that the expression of TMPRSS2 may increase the frequency of prostate cancer when fused with oncogene ERG (Mollica et al., 2020). TMPRSS2 is also known to activate different respiratory illnesses in

addition to coronaviruses, including influenza (Shen et al., 2020). This could imply that varying TMPRSS2 expression may be connected to Covid-19 symptom severity. TMPRSS2 is not a well studied gene, and also has no crystal structure. It is also believed that other TMPRSS genes (TMPRSS11a, TMPRSS4) may also have a role in influenza and other respiratory diseases. This research would provide additional information about TMPRSS2 and the TMPRSS family to the scientific community.

TMPRSS2 has no overt phenotype. In gene knockout studies, mice with TMPRSS2 were observed to have no lethal effects and showed decreased viral spread of the influenza virus, as well as a less severe immune response when infected with SARS-CoV and MERS-CoV (Shen et al., 2017). These mice appeared to be healthy. Conversely, ACE2 knockout mice showed lethal effects (Baughn et al., 2020). These results may indicate that TMPRSS2 could be a target for coronavirus therapeutics as opposed to ACE2. In addition, antiandrogens have been successful in reducing TMPRSS2 expression in prostate cancer cells, which increases the possibility of future interventions for Covid-19 disease (Strope and Chau, 2020).

TMPRSS2 is located on the 21st chromosome; specifically 21 q22.3. TMPRSS2 has 14 exons and it's promotor, regulatory regions, enhancer regions are all known. The terminator location has not been located on the gene. TMPRSS2 also has three known isoforms that are a direct result of alternative splicing. Isoform 1 has been the most studied due to its expression in viral target cells. Important areas within this isoform include the "low density lipoprotein receptor class a" (LDLa), which is a cysteine rich repeat that is important for mammalian cholesterol metabolism, and the "scavenger receptor cysteine rich domain" (SRCR_2) is a highly conserved domain. In addition, the "trypsin-like serine protease" (Tyrp_SPc) is the region that the active site is located, so conserving this region is also important. Both isoform 2 and isoform 3 have not been characterized yet; however both are truncated at around 490 amino acids. Active

sites of TMPRSS2 include His296, Asp345, and Ser441. The substrate binding sites include Asp435, Ser460, Gly462.

Due to its significance to Covid-19, single nucleotide polymorphisms/variations (SNPs) of TMPRSS2 could impact the structure and function of TMPRSS2 and alter binding to SARS-CoV-2. SNPs are changes in the nucleotide sequence that can impact the transcription of mRNA and protein translation. TMPRSS2 has 11,023 intron variants, 393 nonsynonymous variants, 186 synonymous variants, 3 in-frame insertion variants, 2 in-frame deletion variants and 1 initiator codon variant. Past research has determined some SNPs of potential interest. Variants found in ClinVar have been researched and published in the NCBI database. In the ClinVar database, 5 synonymous, 3 nonsynonymous, and 1 frameshift SNP were identified.

Nonsynonymous missense mutations are changes to an amino acid which can be impactful to the structure and function of a protein if it is located in interaction sites and conserved domains. More research has been done on SNP rs12329760, which changes valine 160 to methionine. It is located on the beta strand and may influence the secondary structure of the protein, and could explain differential susceptibility to SARS-CoV-2 infection and prostate cancer (Baughn et al., 2020). However, some research shows that it may not physically impact binding or interactions itself (Senpati et al., 2020). SNPs rs2070788 and rs383510 have been found to increase the risk for severe H1N1 A virus infection (Thunders and Delahunt, 2020). In particular, SNP rs2070788 is located in the regulatory region. SNP rs75603675 has been predicted to be associated with the interaction site of SARS-CoV-2 and forms a de novo proteolytic cleavage site (Baughn, 2020). SNP rs1475908 alternate allele has been linked to low TMPRSS2 expression while rs 74659079 and rs 283057 are associated with high TMPRSS2 infection.

Global SNP population data for TMPRSS2 remains sparse. While many have been identified in the NCBI database, only few have been cited, and even fewer have available population frequencies. Most of the frequency data come from European or white ancestry. SNP rs12329760 has been identified as one of the most common variants amongst the population, present in 22% of the available sample sizes. Across European ancestry, this SNP is present in 22% of the population, 29% in people with African ancestry, 38% in Asian ancestry, and 15% in Latin American ancestry. The remaining SNPs of interest were narrowed down by frequency according to NCBI ALFA aggregator.

No crystallized structure for TMPRSS2 has been finalized, so protein prediction software must be used to visualize the impacts of SNPs. Studies have shown that TMPRSS2 polymorphisms could likely influence the susceptibility and severity of Covid-19 disease.

The aim of this research is to determine the structural and functional implications of these SNPs and their relation to disease symptoms and pathogenicity. It is important to determine if these SNPs are impactful to the protein, and if they are damaging. Software exists that score and categorize the variants as benign/damaging and tolerated/deleterious (David et al., 2020). Using global data to determine the frequency of SNPS in the population and their potential effects, we anticipate that frequent SNPs may have more functional or structural impacts on the binding of TMPRSS2 to SARS-CoV-2. These functional or structural impacts may have implications for the severity of Covid-19 disease. Additionally, TMPRSS2 has been shown to be a potential target for drugs and therapeutics, so it is imperative to study the variants, structure, interaction sites, and frequency of TMPRSS2 in order to better understand its role in respiratory disease infection, and in particular SARS-CoV-2.

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

Bugge, T. H., Antalis, T. M., & Wu, Q. (2009). Type II transmembrane serine proteases. *Journal of Biological Chemistry, 284*(35), 23177-23181.

Clinckemalie, L., Spans, L., Dubois, V., Laurent, M., Helsen, C., Joniau, S., & Claessens, F. (2013). Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. *Molecular endocrinology, 27*(12), 2028-2040.

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

Senapati, S., Kumar, S., Singh, A. K., Banerjee, P., & Bhagavatula, S. (2020). Assessment of risk conferred by coding and regulatory variations of TMPRSS2 and CD26 in susceptibility to SARS-CoV-2 infection in human. *Journal of genetics, 99*, 1-5.

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. https://doi.org/10.1016/j.biochi.2017.07.016.

Strope, J. D., & Chau, C. H. (2020). TMPRSS2: Potential Biomarker for COVID‐19 Outcomes. *Journal of clinical pharmacology*.

Thunders, M., & Delahunt, B. (2020). Gene of the month: TMPRSS2 (transmembrane serine protease 2). *Journal of clinical pathology, 73(*12), 773-776.