Review Article

ACE2 and TMPRSS2 polymorphisms and the development of COVID-19: a review of the literature

Adrhyan Araújo da Silva Oliveira¹,², Ana Maisa Passos da Silva¹,², Jackson Alves da Silva Queiroz¹,², Paulo Ricardo Freitas de Souza¹,², Juan Miguel Villalobos Salcedo³ and Deusilene Souza Vieira¹-⁴*

¹Laboratório de Virologia Molecular, Fundação Oswaldo Cruz Rondônia, FIOCRUZ/RO, Porto Velho, RO, Brazil
²Programa de Pós-Graduação em Biologia Experimental, Universidade Federal de Rondônia, UNIR, Porto Velho, RO, Brazil
³Universidade Federal de Rondônia, UNIR, Porto Velho, RO, Brazil
⁴Centro de Pesquisa em Medicina Tropical, CEPEM, Porto Velho, RO, Brazil

Abstract

SARS-CoV-2 is a virus that has a positive-sense, single-stranded RNA genome that encodes 4 structural proteins, the main one being the S protein (Spike) responsible for mediating with ACE2 and TMPRSS2 for entry into the host cell. The study of single nucleotide polymorphisms (SNPs) of ACE2 and TMPRSS2 can elucidate their possible intervention in the action of the protein, its activity, and the gene expression of encoding these enzymes, which may increase susceptibility to viral infection. From this, literature searches were carried out until December 2021, listing 11,820 publications for literary analysis on the described genetic variations of these protein structures, as well as their relation and influence on the pathology. It was possible to conclude that there is a great influence exerted by genetic variability in ACE2 and TMPRSS2 increasing the ability of the virus to bind to the host cell and the development of COVID-19 with complications.

Introduction

In late 2019, in Wuhan in China, 27 cases of pneumonia of unknown cause were reported, subsequently concluding that the pathology was caused by a new type of Coronavirus with a high capacity for human-to-human transmission [1]. In February 2021, the International Committee on Viral Taxonomy (ICTV) named this virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), and the World Health Organization (WHO) identifies the disease it causes as COVID-19, elevating the disease status to a pandemic in March 2020 based on its high transmission rate and rapid global spread [1,2].

COVID-19 is an acute respiratory disease characterized by flu-like symptoms, where Severe Acute Respiratory Syndrome (SARS) is the aggravation of this pathology [3]. Since the beginning of the pandemic, a higher prevalence and development of the disease is observed in the elderly and people with comorbidities, however, there is the visualization of different clinical pictures among individuals who present similarities in age and sex, where some develop the symptoms of the disease, with worsening regardless of comorbidities, as well as cases of infected individuals with asymptomatic picture [4].

With the beginning of vaccination, it was possible to establish control over the incidence of the pathology in the world, and as a consequence, there were drops in its prevalence and mortality rates [5,6]. Cases of SARS-CoV-2 infection and the development of the disease are recurrently visualized even in immunized individuals, with the recommendation of the interposition of booster doses to ensure greater durability of the acquired immune response [6].

SARS-CoV-2 infection often results in the development of
COVID-19 and for the viral infection process to be successful there must be the interaction of the viral particle with specific human cellular receptors such as Angiotensin-Converting Enzyme 2 (ACE2) [7]. And in some cases, other proteins are involved, such as Transmembrane Protease Serine 2 (TMPRSS2) [3,8].

Disease susceptibility factors such as environment, climate, comorbidities, and especially the genetics of individuals are intrinsically linked [9]. Each and every structure of the body is encoded by specific genes present in the genetic material, each gene having combinations of alleles that configure different structures. Point mutations are known as Single Nucleotide Polymorphisms (SNPs) play a genotypic role in the phenotypic diversity of diseases [10].

Thus, the study of the SNPs of ACE2 and TMPRSS2 may be crucial for understanding the greater susceptibility to viral infection in some individuals and from this to elucidate possible causes of disease development and presentation of different symptoms even in similar individuals and in vaccinated individuals [3]. Therefore, understanding the functional and genetic role of ACE2 and TMPRSS2 and how their polymorphisms may influence the greater susceptibility to viral infection and the development of the pathology is of great importance.

Methods

Data source and strategy

Searches were conducted on the PubMed, PubMed Central, Gene, GEO DataSets, and Genome platforms of the National Center for Biotechnology and Information (NCBI) to identify relevant articles on the subject. With this, keywords related to COVID-19 and SARS-CoV-2 infection correlated to ACE2 and TMPRSS2 were used with an emphasis on their polymorphisms and their links to viral virus infection in humans. Clinical progression, severity, risk factors, and vaccination are research topics also addressed for inclusion criteria in this review.

Study selection and eligibility requirements

Observational studies (cross-sectional, cohort, and case-control studies) were selected with the exception of case reports and case series. From this, research reporting the importance of ACE2 and TMPRSS2 in SARS-CoV-2 viral infection was included, as well as research addressing polymorphisms in genes encoding these proteins involved in virus-infected individuals, the development of the disease, and increased susceptibility to infection as a consequence of genetic variability in the host.

Study identification and selection

Searches were conducted until December 2021, from this, 11,820 citations were identified (PubMed: 10,580; PubMed Central: 956; Gene: 99; Geo DataSets: 30; Genome: 155), from these studies articles were excluded based on their titles, abstract, relevance to the study, duplicates and study type (N=11,557), from the 263 remaining 224 did not meet the eligibility criteria. With this, 39 were selected for full-text evaluation. All studies comprised observational research and all were matched with the proposed inclusion criteria.

Results and discussion

Viral genomics of SARS-CoV-2 and its infectious process

SARS-CoV-2, belonging to the genus Betacoronavirus (group 2) and to the family Coronavidae [11,12], is an enveloped virus that has as genome a ≅ 30-kb positive-sense single-stranded RNA that encodes 4 structural proteins being N (nucleocapsid), M (Membrane), E (Envelope) and S (Spike) and 16 nonstructural proteins [2,4,12]. Being a novel beta coronavirus, SARS-CoV-2 has a genome with ≅ 80% similarity to the SARS-CoV genomic sequence and 50% to MERS-CoV [2,4,13], because of this, its structural proteins share more than 90% similarity with those of SARS-CoV [2,4] differing only from Protein S in total amino acid size, sequence similarity and receptor-binding domain (RBD) [2,4,5,13].

Like SARS-CoV, the viral infection process by SARS-CoV-2 depends on the interaction of the S protein in its “open” state with ACE2 [2,4,7,11]. After binding the RBD domain to ACE2, the Spike protein undergoes proteolytic activation mediated by host cell proteases such as FURIN or TMPRSS2, which cleave between the S1 and S2 subunits of the viral protein [3,12-15]. This process induces the fusion of the viral envelope with the cell membrane which results in the internalization of the genome into the cytosol [4,14,15]. Another method of entry of the viral particle into the cell is through receptor-mediated [14,15] where interaction with ACE2 induces an invagination of the cell membrane and the creation of an endosome[8]. Some proteases can act in this process such as Cathepsin B and L [8,14] inducing fusion of the viral envelope with the endosomal membrane and release of the viral genome into the cytoplasm [2,15].

In the cell cytoplasm, the Open Reading Phases (ORFs) 1a and ORF1ab of the viral genome will be translated by the host cell ribosomes into two polyproteins (pp1a and pp1ab) [4,15,16]. Translation products are cleaved by the viral proteases Plpro and Mpro into nonstructural proteins (Nsp) 1 to 16 [2,16], which will later assemble to form the Replication-Transcription Complex (RTC) in the host cell [2,4,16,17]. With the formation of the RTC, the ssRNA(+) strand will serve as a template for the synthesis of the negative-sense genomic RNA (ssRNA-) [2,17] and for the transcription of subgenomic mRNAs which will be translated into structural and accessory proteins. The ssRNA(−) in turn will be used for the synthesis of positive genomic RNA for the assembly of new viral particles [2,16,18]. The E, M, and S proteins are directed to the endoplasmic reticulum where they will be activated by proteolytic cleavage, except for the S protein [12,15-17], while
the N protein targets the positive RNA strand and binds to form the nucleoprotein complex [1,2,16,17]. From there, the nucleocapsid and structural proteins move to the Endoplasmic Reticulum-Golgi Intermediate Compartment (ERGIC) for the assembly of new virions [2,16]. After this process and its maturation, vesicles are formed that carry the viral particle to the extracellular environment and release it by exocytosis [2,14,16,17,19].

**Angiotensin-converting enzyme 2 (ACE2) a protein with vascular regulatory function**

Angiotensin-converting enzyme 2 (ACE2) is a protein responsible for cleaving angiotensin I into angiotensin 1-9, and angiotensin II into the vasodilator angiotensin 1-7, which is encoded by the ACE2 gene, located on sex chromosome Xp22.2 (Figure 1) [7,11]. The protein is part of the set of enzymes participating in the Renin-Angiotensin-Aldosterone System (RAAS), which plays a major role in the regulation of the cardiovascular system, affecting blood pressure, and sodium regulation, fluid balance, and vascular function [11].

The expression of this gene in the human body is mainly distributed in organs and cells that play roles in controlling cardiovascular, renal, and gastrointestinal function, but it can also be found in other tissues of the human body [11].

**ACE2 and its function in SARS-CoV-2 viral infection**

ACE2 also plays an associative role in the development of COVID-19, playing an essential role in the process of viral infection by the Severe Acute Respiratory Syndrome-causing Coronavirus (SARS-CoV-2), mediating the entry of the viral particle into the host cell through the interaction of the receptor-binding domain (RBD) of the surface glycoprotein Spike (S) (located on the viral envelope) with ACE2 [12]. The process triggered by the internalization of ACE2 leads to decreased physiological protein action, impairing the processes dependent on this enzyme [11].

The divergence of enzyme expression between individuals is a factor of irrefutable importance for understanding the risks of COVID-19, factors such as gender and age influence the higher or lower amount of ACE2 [20,21]. Gene expression assays with humans and mice have shown that there is a higher detection of the protein-coding mRNA in elderly individuals, especially males, where there is an excessive increase of the protein structure expression in the lungs and kidneys of these individuals [22]. In turn, females, despite having two X chromosomes each with a functional copy of the ACE2 gene, have not been detected to increase the protein in their organs due to a complex transcriptional and proteolytic control and hormonal factors that counteract the overexpression of the enzyme in women [20-23].

In theory, the low expression of this protein would be beneficial to combat SARS-CoV-2 by inhibiting its infection mechanism, but there is no data that proves the effectiveness of this measure without harm to the human organism [22]. After the viral particle binds to ACE2, there is the internalization of this protein, which leads to a decrease in its physiological action and damage to the RAAS (Figure 2) [11,12,22,24]. The absence of ACE2 in mice was associated with lung damage after viral infection [22], there is also a correlation between increased organ damage and the development of the severe condition of COVID-19 [5], in addition to the positive regulation of Ang II in humans due to ACE2 deficiency [11]. Therefore, decreased expression of ACE2 may contribute to the control of SARS-CoV-2 viral infection, however, as a consequence it will cause delayed conversion of Ang II to angiotensin 1-7 and thus increased interaction with ATR1 resulting in tissue damage [11,24].
ACE2 polymorphisms involved in increased susceptibility to COVID-19

Genetic variation influences increased susceptibility to developing COVID-19 [10,13], single nucleotide polymorphisms (SNPs) characterize allele changes that can configure differences in the regulation of the expression of genes encoding these structures and in their respective functions in individuals of a population [10]. The ACE2 gene has a high degree of genetic polymorphisms with more than 1700 variations identified, a portion is used for understanding the increased susceptibility to viral infection by SARS-CoV-2 in the molecular interaction with spike protein [3,12], among the associated SNPs, rs73635825 [3] and rs143936283 [25] have been shown to be variations of clinical importance.

The polymorphisms of ACE2 differ according to the allele frequency characterizing in some cases an influence on the RAAS, there is an association of SNPs of the protein with the development of several diseases such as hypertension, left ventricular hypertrophy, myocardial infarction [11,25]. With this, some described variations directly interfere with the surface interaction of ACE2 with angiotensinogen [25], it is important to note that dysfunction in this system can lead to cardiac and renal complications, since it covers vascular flow and sodium and hydric control [11], therefore, genetic variation is directly linked to increased susceptibility and to the development of the disorder COVID19 with complications.

Transmembrane Protease Serine 2 (TMPRSS2) a protease of importance

The process of protein degradation is necessary for the life-sustaining cellular proteins [14], for that, they are submitted to constant renewal procedures (apoptosis) that require new protein synthesis and subsequent degradation, this process requires the action of proteins called proteases that have as function the hydrolysis of peptide bonds such as TMPRSS2 [11,14].

Transmembrane protease serine 2 (TMPRSS2) is a protein encoded by the TMPRSS2 gene, located on autosomal chromosome 21q22.3 [14]. Just like other serine proteases their physiological function is not yet fully elucidated, but there is their possible participation in indigestion, coagulation, embryonic development, and spermatogenesis [14]. Its expression was initially associated with the small intestine, however, it was later found to be more expressed in the prostate epithelium [14,26], because of this, there is an association of protease in carcinogenesis due to overexpression in the development of prostate cancer by molecular alterations and the formation of the TMPRSS2-ERG arrangement, where one TMPRSS2 allele loses its promoter and one of the ERG alleles receives it [27].

TMPRSS2 has a type II transmembrane domain, A domain, scavenger receptor, and serine protease domain [14], possibly its functionality would be related to a cell-specific ligand mediating signal receptor [14,19]. It can be found in the nasal epithelium, trachea, epithelial cells of the bronchi, digestive tract, and prostate, its expression is developmentally regulated and increases by the age range of individuals [12,14,27,28].

TMPRSS2 and its role in SARS-CoV-2 viral infection

Of the 4 structural proteins encoded in the SARS-CoV-2 viral genome [19], the Spike protein is the main one for acting in the process of binding the viral particle to the host cell (Figure 3) [4,12,19]. The S protein consists of two subunits: S1 - acts as a binding region for ACE2; S2 - acts in promoting membrane fusion and proteolytic action [4,11]. At the end of the replicative cycle of the virus the new viral particles leave the host cell with the S protein not activated [4], therefore, the process of virus penetration is intrinsically dependent not only on the angiotensin-converting enzyme 2 but also on the participation of proteases such as TMPRSS2 [4,15,29].

Like other Coronaviruses, SARS-CoV-2 uses two steps for entry into target cells [4,11,12], first, the viral particle binds to ACE2 on the surface of the host cell membrane, then the S-protein is cleaved between the S1 and S2 subunits by endogenous proteases in the human body [15] as the transmembrane serine protease 2, it is important to emphasize that the cleavage must be performed after the binding of the virus to the surface protein, if it occurs earlier there is the possibility of inactivation of the Sprotein [14], thus, even though there is a large availability of proteases in the membrane, their mobility, and non-permanence in the vicinity of ACE2 besides the possibility of inactivation of the spike protein before the interaction process are factors that justify their disuse [15]. Hence, the location and time of action of proteases such as TMPRSS2 are important for viral infection and pathogenicity [14,15].

In addition to transmembrane serine protease 2, other proteases may be involved in cleavage and activation of the viral envelope.
protein S such as Cathepsin B and L in the case of receptor-mediated endocytosis [8,14], therefore, individuals who do not have large amounts of TMPRSS2 expression may use them as mediators of viral envelope protein activation and fusion with the endosome membrane [8]. Furin is another protease widely used for cleavage of the S protein of SARS-CoV-2 and is often associated with TMPRSS2 replacement in case of its absence [8,30].

**TMPRSS2 polymorphisms associated with increased susceptibility to COVID-19**

The genetic variability of TMPRSS2 is directly related to increased susceptibility to COVID-19 [9,31], the rs75603675 and rs12329760 (Val160Met) polymorphisms of TMPRSS2 (Figure 4) have great clinical importance, studies identify that the variation encoding Val160Met is identified in much of the world population with higher allele frequency ≅ 25% and especially for EAS (Overpopulation Code for East Asia) with an allele frequency of ≅ 40% [31-33]. These more prevalent SNPs on TMPRSS2 offer potential effects for greater differential genetic disposition on COVID-19, as well as, potentiating the development of risk factors such as prostate cancer in males, therefore, the study of oncogenic factors [9,26,32].

The potential action of TMPRSS2 on COVID-19 may be expressed in the rs2070788 and rs383510T polymorphisms, which are associated with higher protease expression in human lung tissue [31]. Similar to the above, most studies highlight a higher risk for males linked not only to polymorphisms but also to androgen disposition, an observation based on hair loss related to COVID19 male subjects admitted [15,34,35]. Interestingly, the TMPRSS2 promoter gene has a 15bp AR (Androgen Receptor) response element and the human protease appears to be regulated by these receptors, therefore indicating gender differences in SARS-CoV-2 infection [32].

Moreover, the gene encoding TMPRSS2, being located on autosomal chromosome 21q22.3, may in thesis verify a higher risk of infection by SARS-CoV-2 in individuals with trisomy 21 (Down Syndrome) [9]. Increased expression of protease may be potentiation of the infection factor [9,32,36]. Studies have identified an increase in the presence of the enzyme in both mice and humans as they age[4], therefore, identifying that possible regulation of TMPRSS2 development may result in increased protection in infants and children [27,37,38], thus, the study of age-related polymorphisms of this protein may be of great interest [31,35,39].

**Conclusion**

The ACE2 and TMPRSS2 polymorphisms were shown to be great indicators of greater susceptibility to COVID-19, since their variations not only affect the viral ability to infect the cell due to overexpression in some cases, but also the development of other pathologies due to the internalization of these proteins, as in the case of ACE2, causing a decrease in its action and, in some cases, due to allelic variants, even the loss of affinity in the catalytic function, impairing extremely important physiological processes, such as RAAS. Furthermore, there is a need for other studies that address these genetic variations in isolated groups such as children, infants, and people with Down syndrome to study the possible control of expression of genes encoding these structures in a non-harmful way.

The in-depth study of the genetic aspects involved may elucidate a new pathway for new therapies for COVID-19; the importance of the knowledge of the genetic profile of individuals infected by the virus indicates a high potential for the discovery of new mechanisms involved in the greater vulnerability to viral infection, such as the polymorphisms cited. Moreover, a more cautious look is needed regarding the studies related to the male population, once it is demonstrated that factors such as age, age range, and genetic variation can directly affect individuals, not only for COVID-19 but also with the development of other complications such as infertility, prostate cancer, and cardiac complications. Thus, genetic knowledge proves to be a strong ally in understanding and combating SARS-CoV-2.

**Authors’ contributions**

**Conceptualization:** Deusilene Souza Vieira.

**Data processing and analysis:** Adrhyan Araújo da Silva Oliveira; Ana Maísa Silva Passos.

**Methodology:** Adrhyan Araújo da Silva Oliveira; Ana Maísa Silva Passos.

**Writing–original draft:** Adrhyan Araújo da Silva Oliveira; Ana Maísa Silva Passos; Jackson Alves da Silva Queiroz.
**References**


ACE2 and TMPRSS2 polymorphisms and the development of COVID-19: a review of the literature


