The novel infection caused by the SARS-CoV-2 coronavirus has become one of the global challenges to mankind in the 21st century. Since mid-February 2020 the infection started spreading quickly across the globe. In March the WHO announced the COVID-19 outbreak a pandemic. By June (at the time of writing), over 13 million confirmed cases were recorded, more than half a million infected people died. Since the first SARS-CoV-2 genome sequencing until present, tens of thousands SARS-CoV-2 full-genome sequences have been obtained from different regions of the world. The data on the newly sequenced SARS-CoV-2 isolates is deposited in GenBank and other databases. The analysis of the genomes of COVID-19 patients with different clinical course and development on their basis of model for stratification of people according to the degree of susceptibility and severity of manifestation will make it possible to develop a personalized approach for the prevention and treatment of COVID-19. The heterogeneity of the COVID-19 clinical manifestation may be associated with the characteristics of the genome of both humans and the virus. A combination of allelic variants of genes associated with viral life cycle can determine susceptibility to SARS-CoV-2 infection. Allelic variants in genes ACE1, ACE2, TMPRSS2, IL6, SLC6A20, LZTFL1, CCR9, FYCO1, CXCR6, and XCR1 can determine the severity of COVID-19. Analysis of the genomes of COVID-19 patients with different clinical course and development on their basis of model for stratification of people according to the degree of susceptibility and severity of manifestation will make it possible to develop a personalized approach for the prevention and treatment of COVID-19.

Keywords: COVID-19, SARS-CoV-2, genomic determinants, genome, genomic variants, alleles, pathogenesis, SARS-COV-2

Received: 01.07.2020 Accepted: 17.07.2020 Published online: 27.07.2020
DOI: 10.47183/mes.2020.003

1. SARS-CoV-2 genomic structure

The SARS-CoV-2 coronavirus genome is ~30,000 nucleotides long. Unlike other highly pathogenic viruses causing severe disease in humans (SARS-CoV and MERS-CoV) this virus has a higher rate of transmission. The origin of the virus still remains unclear. The comparative data analysis has shown that SARS-CoV-2 could emerge due to recombination between the pangolin coronavirus Pangolin-CoV and the bat coronavirus RaTG13 [1]. The receptor binding domain of the Pangolin-CoV spike (S) protein has high sequence similarity with SARS-CoV-2: six inputs of the virus responsible for the major cell receptor binding have essentially identical sequences. The primary amino acid sequence of the novel coronavirus receptor binding domain is different from those of SARS-CoV, therefore, the receptor binding affinity of the SARS-CoV-S protein is 10 times higher compared to the SARS-CoV S protein [2].

2. SARS-CoV-2 life cycle

SARS-CoV-2 employs a number of receptors for cellular entry. The angiotensin-converting enzyme 2 (ACE2) is a major receptor for SARS-CoV-2 [3]. In addition to the major receptor, the virus may employ other cell proteins, such as CD147 and GRP78, as additional receptors. For efficient host cell penetration, the SARS-CoV-2 spike (S) protein should undergo the proteolytic activation by the following cell proteases: furin and cellular serine protease TMPRSS2 [4].

Apart from that, SARS-CoV-2 may use the endosomal way of cell entry involving the cathepsin L protease [5]. Once the virus gets inside the cell, it reprograms the host cell's biosynthetic pathways for its own use, exploiting various cell proteins and forming the interactome (whole set of molecular interactions in a particular cell) of viral proteins and RNA with host factors [6].

3. Polymorphisms in SARS-CoV-2 genome and their impact on biological properties of the virus

Since the first SARS-CoV-2 genome sequencing until present, tens of thousands SARS-CoV-2 full-genome sequences have been obtained from different regions of the world. The data on the newly sequenced SARS-CoV-2 isolates is deposited in
in various tissues and organs were carried out at single-cell resolution, which demonstrated that the described receptors expression could be observed not only in the cells of respiratory epithelium and lungs, but also in the intestinal epithelial cells, cardiomyocytes, hepatocytes and neurons [12]. Apparently, specific expression patterns of these receptors may determine the COVID-19 course.

The course of COVID-19 may be also defined by many other factors, such as comorbidities and environmental factors. It is hypothesised that genetic determinants (sets of gene variants responsible for body’s response to SARS-CoV-2 infection) play a vital part in susceptibility to the virus.

5. Genetic determinants of susceptibility to COVID-19

Expression of receptors and host factors essential to follow the basic viral life cycle stages is a major factor in the body’s susceptibility to coronavirus. The presence of receptor together with co-receptors is important for the effective viral penetration into the target cell. Thus, co-expression of ACE2 and TMPRSS2 may define the target cells for coronavirus. A number of studies using scRNA have demonstrated that different ACE2 and TMPRSS2 co-expression patterns are observed in various human cells, tissues and organs. This can explain the heterogeneity of the COVID-19 clinical manifestation, when the pathogenesis involves not only lungs, but also other organs: liver, kidney, intestines, blood vessels, myocardium and brain [13, 14].

The ACE2 and TMPRSS2 expression in normal cells is low. However, pulmonary disorders and exposure to pollutants and toxic chemicals are usually associated with increased expression of these receptors. Susceptibility of such cells to viral infection is higher compared to normal cells. That can explain the higher proportion of infection and more severe course of the disease in people with comorbidities.

One of the main properties of the virus is the ability to infect cells of the immune system and cause the immunodeficiency disorders [15]. Allelic variations determining the structure of proteins encoded by these genes, as well as the variants in the regulatory non-coding regions affect the expression, contribute to the body’s antiviral response and determine the severity of the disease.

The following international consortia facilitating the response to SARS-CoV-2 genetics basis research has been created for identification of such factors:
- COVID-19 Host Genetics Initiative [16];
- COVID-19 Genomics UK (COG-UK) Consortium [17], etc.

Genes and allelic variants most probably associated with COVID-19 susceptibility and severity are listed in the :tab_1:.

ACE2 gene variants and susceptibility to SARS-CoV-2

Carriers of different allelic variants in the ACE2 gene coding region demonstrate different viral spike (S) protein binding affinity. For example, alleles rs73635825 (S19P) and rs143936283 (E329G) significantly differ in the SARS-CoV-2 S protein binding affinity [17]. Both “dangerous” ACE2 alleles increasing the binding affinity of ACE2 to the S protein (S19P, I21V, E23K, K26R, T27A, N64K, T92I, Q102P and H378R) and “protective” ACE2 variants (K31R, N33I, H34R, E35K, E37K, D38V, Y50F, N51S, M62V, K68E, F72V, Y83H, G326E, G352V, D355N, Q388L and D509Y) decreasing the efficiency of receptor binding to the SARS-CoV-2 S protein have been reported [18]. The study of Italian population revealed some rare “protective” missense variants of ACE2 gene: p.Asn720Asp, p.Lys26Arg, p.Gly211Arg (MAF 0.002 to 0.015). These variants were able
to interfere with binding to the viral S protein. However, it should be remembered that recent analysis of 1000 genomes from the UK Biobank revealed no relationship between the COVID-19 severity and the variants in the ACE2 and TMPRSS2 genes [19].

**TMPRSS2 gene variants and susceptibility to SARS-CoV-2**

The TMPRSS2 cellular serine protease is essential for the SARS-CoV-2 S protein proteolytic activation needed for host cell entry. Differential expression of TMPRSS2 may determine the tissue specific virus-host interaction playing a vital part in susceptibility to viral infection. Thus, the lung-specific loci variants affecting the expression profiles (eQTL) associated with the TMPRSS2 expression may be responsible for different susceptibility and response to SARS-CoV-2 infection. It has been shown that the eQTL variant rs35074065 is associated with higher expression of TMPRSS2 and low expression of the interferon-induced MX1 gene [20]. A number of alleles associated with increased expression of TMPRSS2 (for example, rs2070788, rs9974589, rs7364083) are common in European population [21]. The eQTL most common in Europeans (rs8134378) located near the androgen-dependent enhancer TMPRSS2 may be associated with increased TMPRSS2 expression in men [22]. Despite the predicted associations between the ACE2 and TMPRSS2 allelic variants, the recently published report has confirmed no association between the described variants and the COVID-19 susceptibility [23].

**HLA and immune response features, immunodepletion in patients with coronavirus infection**

It has been reported that the ability to bind and present antigens is a key point of effective immune response mobilization. Various MHC molecules (HLA molecules) bind to the viral proteins’ antigenic peptides emerging in different way during the disease. This may explain the differences in the ability to develop an immune response between individuals.

Carriers of the HLA-B*46:01 variant had a few predicted binding peptides for SARS-CoV-2, suggesting that individuals with that allele might be particularly vulnerable to COVID-19. For example, it was shown that in people with such genotype the SARS-CoV course was more severe [24]. Among all HLA Class 1 alleles, HLA-A*02:02, HLA-B*15:03 and HLA-C*12:03 bind to the maximum range of the SARS-CoV-2 conserved antigenic peptides. On the contrary, alleles A*25:01, B*46:01, 150 C*01:02 HLA-A, -B, and -C are responsible for binding to the minimum range of antigenic peptides. It should be noted that when presenting the antigenic peptides of 8-13 amino acids in length, the 44 peptides are highly conserved and are found in all coronaviruses, including the common human coronaviruses (OC43, HKU1, NL63 and 229E) [24].


As mentioned above, the heterogeneity of the COVID-19 clinical manifestation may be associated with the differences between the allelic variants of genes not required for the viral life strategy implementation. The recent genome-wide association study (GWAS) aimed at the search for relationship between genes and COVID-19 severity in patients from Italy and Spain has revealed loci and genic variants responsible for discrimination of patients based on the disease severity. One of these is locus 3p21.31, which comprises genes SLC6A20, LZFNL1, CCR9, FYCO1, CXCR6 and XCR1. Genes of the locus 3p21.31 are associated with chemokines and the movement of immune cells toward sites of inflammation. It should be noted that gene SLC6A20 located within the described locus physically and functionally interacts with ACE2 and is able to modulate the properties of the receptor. The other detected locus (9q34) is associated with inheritance of ABO blood type antigens [26]. The relationship between the COVID-19 severity and the rs8176747, rs41302905 and rs8176719 alleles defining the blood type in Chinese population has been previously reported. It has been shown that patients with blood type O have a decreased risk of severe infection, and patients with blood type A are vulnerable to severe COVID-19 [27].

**Genetic variants involved in inflammatory response**

In patients with COVID-19, a complex of factors contributes to the excessive inflammatory response and the cytokine storm syndrome. That has been confirmed by recent histological analysis of samples acquired from the post-mortem examination of patients with fatal COVID-19. The study has demonstrated that excessive inflammatory response is the most common cause of death in COVID-19 patients [34].

One of the major clinical manifestations of pneumonia is high level of proinflammatory cytokines (IL-6, TNF-α and IL-1β) and acute phase proteins (APPs). Allelic variants in genes responsible for inflammatory response may affect the disease severity. For IL-6, the correlation with IL-6-174C allele cause of death in COVID-19 patients [34].

One of the major clinical manifestations of pneumonia is high level of proinflammatory cytokines (IL-6, TNF-α and IL-1β) and acute phase proteins (APPs). Allelic variants in genes responsible for inflammatory response may affect the disease severity. For IL-6, the correlation with IL-6-174C allele associated with high IL-6 level and severe course of pneumonia has been revealed in patients with severe COVID-19 [28]. Polymorphism of the C3 gene encoding the complement

<table>
<thead>
<tr>
<th>Molecular pathway</th>
<th>Genes/genomic variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell entry (receptor)</td>
<td>ACE2 rs73635625 (S19P) и rs143936283 (E329G)</td>
</tr>
<tr>
<td>Cell entry (receptor-protease)</td>
<td>TMPRSS2 rs2070788, rs9974589, rs7364083, eQTL (rs8134378),</td>
</tr>
<tr>
<td>Cell entry (protease)</td>
<td>TMPRSS4</td>
</tr>
<tr>
<td>Cell entry (protease)</td>
<td>Cathepsin L</td>
</tr>
<tr>
<td>Cell entry (protease)</td>
<td>Furin</td>
</tr>
<tr>
<td>Protease</td>
<td>PLG (plasmin)</td>
</tr>
<tr>
<td>Cell entry (co-receptor)</td>
<td>CD147</td>
</tr>
<tr>
<td>Cell entry (antiviral restriction factor)</td>
<td>GILT</td>
</tr>
<tr>
<td>Presentation of viral antigens</td>
<td>HLA- A<em>02:02, HLA- B</em>15:03, HLA- C<em>12:03, HLA- A</em>25:01, HLA- B<em>46:01, HLA- C</em>01:02</td>
</tr>
<tr>
<td>Inflammatory response</td>
<td>IL6</td>
</tr>
<tr>
<td>Inflammatory response</td>
<td>IL1B</td>
</tr>
</tbody>
</table>
activation of the interferon synthesis pathway. The secreted interferons induce the interferon-stimulated genes (ISG) activation via interaction with receptors. This confers resistance to viral infection. The described response involves more than a hundred factors, both sensors RIG-I, MDA5, MAVS, STING, cGAS, TLR3, TLR9, TRIM25, RNF186, and effectors IFNa, IFNβ, IFN-λ, OAS1, MX1 and IFITM3. The role of these genes’ allelic variants currently remains unknown. It has been shown that the variant rs12252 of the IFITM3 gene may be associated with excessive inflammatory response and severe COVID-19 [33].

CONCLUSION

The extensive genome-wide association studies launched by the international consortia are important steps in the process of the novel coronavirus infection pathogenesis investigation. Sample size increase together with various ethnic groups’ analysis will make it possible to identify the unique rare allelic variants responsible for susceptibility to COVID-19. Reconstructing the allelic variants’ cumulative contribution to complex gene networks regulating the antiviral response might shed some light on the COVID-19 pathogenesis and help to create a genetic risk prediction model allowing one to define the probability of severe COVID-19

References

17. https://www.cogconsortium.uk/
19. David Curtis Variants in ACE2 and TMPRSS2 genes are not major determinants of COVID-19 severity in UK Biobank subjects medRxiv 2020.05.01.20085860; doi: https://doi.org/10.1101/2020.05.01.20085860.
Zhang L, Jackson CB, Mou H, et al. The D614G mutation in the
Amin M, Sorour MK, Kasry A. Comparing the Binding Interactions
Gubernatorova EO, Gorshkova EA, Polinova AI, Drutskaya MS.
Wu Y, Feng Z, Li P, Yu Q. Relationship between ABO blood group
A Usme-Ciro, Nicolas David Franco-Sierra, Astrid C Florez-
SARS-CoV-2 spike protein reduces S1 shedding and increases
doi:10.1038/s41586-020-2355-0.
Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for
doi:10.1073/pnas.2003138117
cytogfr.2020.05.009.


34. Delanghe J.R., De Buyzere M.L., Speeckaert M.M. C3 and ACE1 polymorphisms are more important confounders in the spread and outcome of COVID-19 in comparison with ABO polymorphism. European Journal of Preventive Cardiology. 2020;2047487320931305.