COVID-19: Immunogenetics and Immuno-Epidemiological Parameters

Attapon Cheepsattayakorn¹ and Ruangrong Cheepsattayakorn²

¹110th Zonal Tuberculosis and Chest Disease Center, Chiang Mai, Thailand
²Department of Pathology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

ABSTRACT

As of February 15, 2020, 51,800 cases of COVID-19 disease, including more than 1,600 COVID-19-related deaths, had been laboratory-confirmed in mainland China, mainly in Hubei province. Additionally, 526 laboratory-confirmed cases have been reported across 25 other countries. Approximately, 15% of cases reported to the World Health Organization (WHO) are severe, 3% are critical, and 82% are mild clinical manifestations, whereas the estimated overall case fatality rate is approximately 2% or less, not different from the fatality identified in the seasonal influenza. The mean age of COVID-19 patients is 52.4 years, whereas children and adolescents are the least likely group to be infected with the COVID-19, occurring in only 2% of cases 19 years of age or younger. When the younger-age group get sick, they will get a mild form of COVID-19 without serious complications, with an average death rate of 0.2%. Men constitute more than two-thirds of the reported COVID-19 cases (73% vs. 27%) and are more than 1.5 times more likely to die from COVID-19 (death rate: 2.8% vs. 1.7%). This sexual distinction of the anti-viral immunity between men and women is due to the genetic factors, hormonal factors, and environmental factors. The unanswered questions include the pathophysiology of pulmonary clinical infection, influenza and other viral co-infection, and the rate of bacterial complications. SARS-CoV-2 (COVID-19) infection has evolved to become a pandemic, in contrast to infections with SARS and MERS, whereas SARS-CoV-2 (COVID-19) has demonstrated having the similarities of genome sequence, receptor affinity, pathogenesis, and disease manifestation.

In conclusion, although genomic evidence does not support the belief that COVID-19 is a laboratory construct, currently it is impossible to disprove or prove the theories of its origin. To identify the COVID-19 origin, obtaining virus sequences from immediate non-human animal sources would be the most definite method. In the absence of proper cure of COVID-19, it is necessary to identify the factors that may assist in assessment of the COVID-19 disease severity before rapid progression of the disease.

Keywords: COVID-19, SARS-Co-V-2, Immunogenetics, Immunopathology, Epidemiological Parameters, ABO Blood Type, Proximal Origin, Phylogenetic Analysis, Human Leukocyte Antigen

Abbreviations

ACE: Angiotensin-Converting Enzyme; ARDS: Acute Respiratory Distress Syndrome; BALF: Bronchoalveolar Lavage Fluid; CoV: Coronavirus; COVID-19: Coronavirus Disease 2019; 2019-nCoV: 2019-Novel Coronavirus; DAMPS: Damage-Associated Molecular Patterns; FCN: Ficolin; HLA: Human Leukocyte Antigen; ICU: Intensive Care Unit; IFITM1: Interferon-Induced Transmembrane Protein-1; IFN: Interferon; IL: Interleukin; ISGs: Interferon-Stimulated Genes; MERS: Middle-East Respiratory Syndrome; MRC: British Medical Research Centre; NK: Natural Killer; OR: Odd Ratio; PD-1: Programmed Death Ligand-1; PAMPs: Pathogen-Associated Molecular Patterns; PBMCs: Peripheral Blood Monocytes; pDC: Plasmacytoid Dendritic Cells; RBD: Receptor Binding Domain; RNA: Ribonucleic Acid; SARS: Severe Acute Respiratory Syndrome; Tim-3: T-cell Immunoglobulin and Mucin-3; TNF: Tumor Necrosis Factor; WHO: World Health Organization

Introduction

As of February 15, 2020, 51,800 cases of COVID-19 disease, including more than 1,600 COVID-19-related deaths, had been laboratory-confirmed in mainland China, mainly in Hubei province [1]. Additionally, 526 laboratory-confirmed cases have been reported across 25 other countries [1]. Approximately, 15% of cases reported to the World Health Organization (WHO) are severe, 3% are critical, and 82% are mild clinical manifestations, whereas the estimated overall case fatality rate is approximately 2% but the figure outside of Hubei province is approximately 0.05% or less, not different from the fatality identified in the seasonal influenza [2, 3]. There is no expected cross protection by a common human coronavirus infection and theoretically, COVID-19 can infect any one of the individuals [3]. COVID-19 complications target especially the elderly, but the complications...
can unpredictably occur in the younger age populations, as well as influenza [3]. Despite a low risk of COVID-19 complications, a mild clinical presentations of the disease allow a larger chain of transmission through various populations [3]. Critically, after first identified outbreak, understanding of epidemiology trajectory of infection still have a limitation. Although the WHO reported the dates of diagnosis of COVID-19 disease, but this is not enough information [3]. It is not yet clear why sustained chains of transmission have not been reported outside Asia. Genetic factors encouraging transmission within Asian populations, effective containment, inefficient transmission, poor reporting due to lacking molecular testing capacity in some low-income countries, and specific environmental conditions in Hubei province and mainland China [3].

Among the most severely affected patients, viral ribonucleic acid (RNA) has been detected in the plasma approximately 15% and viral detection in stool reveals possibility of fecal transmission [4, 5]. COVID-19 has been isolated in human saliva, nasopharynx and lower respiratory tract [6, 7]. Lacking lung biopsies or post-mortem sample investigations leads to an incomplete understanding of the pathogenesis of COVID-19 infection [3]. The innate immune cells are born capable of producing T regulatory cell cytokine (interleukin (IL)-10, T helper 17 (Th17) cell cytokines (IL-6 and IL-23), but inability of induction of T helper 1 (Th1) cell cytokines (type I interferons (IFNs), IFN-γ, and IL-12) [8]. A new lineage of oligoclonal T cells that express natural killer (NK)-related receptors (NKR) is formed, whereas the diversity of the T-cell receptor (TCR) repertoire decreased with age [9]. In addition to the experimental evidence regarding the immunological features in neonates and children that are more prominent than in adults. In consideration of the pro-inflammatory response to the SARS-CoV infection, an overwhelming inflammatory reaction in aging population is a logical possibility [10]. The mean age of COVID-19 patients is 52.4 years, whereas children and adolescents are the least likely group to be infected with the COVID-19, occurring in only 2 % of cases 19 years of age or younger [11]. When the younger-age group get sick, they will get a mild form of COVID-19 without serious complications, with an average death rate of 0.2 % [12]. Men constitute more than two-thirds of the reported COVID-19 cases (73 % vs. 27 %) and are more than 1.5 times more likely to die from COVID-19 (death rate : 2.8 % vs. 1.7 %) [13]. This sexual distinction of the anti-viral immunity between men and women is due to the genetic factors, hormonal factors, and environmental factors. The unanswered questions include the pathophysiology of pulmonary clinical infection, influenza and viral co-infection, and the rate of bacterial complications [3].

Human Leukocyte Antigen Map of COVID-19

HLA genotype plays a significant role in differential regulation and activation of T cells as well as disease duration and transmission [14]. A recent study on human leukocyte antigen (HLA) binding affinity of 48,395 unique peptides (possible 8- to 12-mers) from the SARS-CoV-2 (COVID-19) proteome for assessing the potential for cross-protective immunity conferred by previous exposures to common human coronaviruses (i.e. 229E, NL63, OC43, and HKU1) demonstrated that alleles HLA-A*02:01, HLA-B*07:01, and HLA-C*07:02 were the top presenters of conserved peptides. Fifty-six different HLA alleles, especially HLA-B*46:01 revealed a remarkable binding affinity (<500 nm) to all of the conserved SARS-CoV-2 (COVID-19) peptides, indicating a concomitant lack of potential for cross-protective immunity from other human coronaviruses. Considering the entire proteome of SARS-CoV-2 (COVID-19), HLA-A and HLA-C alleles expressed the relative largest and smallest capacity to present SARS-CoV-2 (COVID-19) antigens, respectively. No appreciable global correlation between conservation of the SARS-CoV-2 (COVID-19) proteome and its predicted MHC binding affinity, indicating a lack of selective pressure for the capacity to present coronavirus epitopes (p = 0.27, Fisher’s exact test). peptide presentation appears to be independent of estimated time of peptide production during SARS-CoV-2 (COVID-19) life cycle, with indistinguishable early and late SARS-CoV-2 (COVID-19) peptide presentation [14].

HLA binding assays carried out for 19 epitopes displaying positive T cell assays in all retrieved SARS-CoV-2 (SARS-CoV-2 (COVID-19) proteins in a recent study demonstrated that five distinct alleles, HLA-A*02:01, HLA-B*40:01, HLA-DRA*01:01, HLA-DRB1*07:01 and HLA-DRB1*04:01 were positive. The population coverage of these alleles in China was 32.36 % and 59.76 % globally [15]. Previous studies demonstrated an increased severity towards the closely related SARS-CoV infections in persons with HLA-B*46:01 [16]. Differences in HLA haplotype may influence the person’s response to SARS-CoV-2 (COVID-19) infection and some haplotypes may be associated with increased disease severity. Thus, HLA genotyping may help in identifying persons at risk. COVID-19 testing along with HLA genotyping is highly recommended to predict susceptibility to disease severity and assisting in future vaccination strategy plan.

Proximal Origin of Covid-19

COVID-19 (SARS-CoV-2) is the seventh member of the Coronaviridae known to infect humans [17]. Three of these viruses are: 1) SARS CoV-1, 2) MERS and 3) COVID-19 (SARS-CoV-2) can cause severe disease, and 4) HKU1, NL63, OC43 and 229E, are related to mild respiratory symptoms. The two notable characteristics of the COVID-19 genome are: 1) based on structural modelling and early biochemical experiments, COVID-19 is optimized for binding the human ACE2 receptor; and 2) the highly variable spike (S) protein of COVID-19 has a polybasic (furin) cleavage site at the S1 and S2 boundary via the insertion of twelve nucleotides. This event contributes to the acquisition of the three predicted O-linked glycans around the polybasic cleavage site [17].

Six residuals in the receptor binding domain (RBD) of the spike protein of SARS-CoV and SARS-related coronaviruses, the most variable part of the virus genome appear to be critical for binding to the human ACE2 receptor and the determining host range [18]. Five of these six residuals are mutated in COVID-19 compared to its most closely related virus, RaTG13 sampled from a Rhinolophus affinis bat to which it is approximately 96% identical [19]. COVID-19 seems to have an RBD that may bind with high affinity to ACE2 from human, non-human primate, cat, pig, and ferret [18]. COVID-19 may bind less efficiently to ACE2 in other species related to SARS-like viruses, such as rodents and civets [17]. Recent binding studies demonstrated that COVID-19 binds with high affinity to human ACE2 [20]. The COVID-19 spike appears to be the result of selection on human or human-like ACE2 permitting another suitable binding solution to occur and this strongly indicate that COVID-19 is not the genetic engineering product [17]. All COVID-19 sequenced genomes have the well adapted RBD and the polybasic cleavage site, and thus are derived from a common ancestor. Initial analyses demonstrated that Malayan pangolins (Manis javanica) illegally imported into Guangdong province, China contain a coronavirus (CoV) that is similar to COVID-19 [21, 22]. Nevertheless, no pangolin CoV has been identified to be sufficiently similar to COVID-19 across its entire genome for supporting direct human infection [17].
Phylogenetic Analysis of Covid-19
Volz, et al. demonstrated their study by analyzing 53 SARS-CoV-2 (COVID-19) whole genome sequences collected up to February 3, 2020. They found a strong association between the time of sample collection and accumulation of genetic diversity of COVID-19. By using Bayesian and maximum likelihood phylogenetic methods revealed that the COVID-19 was introduced into the human population in Wuhan, China in early December 2019 and has an epidemic doubling time of about 7 days. Precise estimated of epidemic size are not possible with current genetic data, the analyses demonstrated substantial heterogeneity in the number of secondary infections caused by each COVID-19-infected case that indicated by a high level of over-dispersion in the reproduction number [23]. Phylogenetic analysis demonstrates a common ancestor to SARS-CoV-2 (COVID-19), human SARS-CoV, and the bat SARS-CoV converge. The four structural viral proteins, envelope (E), membrane (M), nucleocapsid (N), and spike (S), imply a high degree of shared identity in range of 97.7-100 % between the SARS-CoV-2 (COVID-19) and bat coronaviruses that supports the descend of SARS-CoV-2 (V1D-19) from an animal [24].

ABO Blood Type Associated Covid-19
A recent study conducted by Zhao et al on the correlation compared between the ABO blood group among 1,775 COVID-19-infected patients and 3,694 normal individuals from Wuhan city, China and 23,386 normal persons from Shenzhen city, China revealed that blood group A had a specific significant higher risk for COVID-19 compared to non-blood group A groups (albeit modest effect size, odd ratio (OR) = 1.20, p = 0.02), while blood group O had a significant lower risk for COVID-19 compared to non-O blood groups (OR = 0.67, p < 0.001) [25].

Tanigawa et al compared blood group O frequencies between the Shenzhen controls and the UK biobank Chinese group and demonstrated that the study results were consistent with the study results conducted by Zhao et al. Nevertheless, Tanigawa et al found that the frequency of blood group O was different between the Wuhan controls and the UK Biobank Chinese group (p = 0.00121) indicating careful consideration of inferences regarding ABO blood group differences [26].

COVID-19 Induced Inflammation
Cytokine and chemokines play a significant role in anti-viral immunity. In COVID-19, the primary cause of acute respiratory distress syndrome (ARDS) and multiple organ failure is related to cytokine storm, a phenomenon of excessive inflammatory reaction mediated by the rapid production of large amounts of cytokines in response to infection [27, 28]. Initial delay in cytokine and chemokine secretion by innate immune cells with subsequent surge in pro-inflammatory cytokines and chemokines (CCL2, CCL5, IFNs, IL-1β, IL-6, IL-8, MCP-1) by the activated macrophages and other recruited lymphocytes is observed in COVID-19, contributing to the recruitment and activation of adaptive immune cells (neutrophils, NK cells, and T cells) along with further production of pro-inflammatory cytokines, and finally causing a cytokine storm and tissue damage [28]. In COVID-19 patients, elevated serum cytokines are G-CSF, GM-CSF, IFN-γ, IL-1β, IL-2, IL-7, IL-8, IL-9, IL-10, IL-17, IP10, MCP-1, MIP-1A, MIP-1B, and TNF-α [29]. Particularly, IL-1, IL-6, and Tumor Necrosis Factor (TNF-α) cytokines secreted by macrophages are significantly higher in severe patients compared to non-severe cases [29]. IL-6, a multifunctional cytokine involving the formation of follicular helper T cells, generation of plasma cells, and differentiation of Th 17 cell subsets plays a primary role in cytokine storm that occurs in patients with COVID-19 [30]. IL-6 also inhibit IFN-α, therefore, suppress CD8+ cytotoxic T cells [30]. As demonstrated by PD-1 and Tim-3 expressions, IL-6 induces T cell exhaustion, thus, T cell-mediated immune response might be suppressed during cytokine storm [31]. Role of IL-6 in COVID-19 disease severity has been demonstrated by increased IL-6 levels and its positive association with disease severity [31-38]. A recent report from Germany revealed that COVID-19 cases with IL-6 levels of at least 80 pg/ml had a 22-fold increased risk of respiratory failure with median time to mechanical ventilation of 1.5 days [39] and higher serum IL-6 levels were also reported even 24 hours before death [39-41]. Thus, IL-6 could be used for early detection of COVID-19 patients at risk for respiratory failure as a single parameter or in association with other parameters.

TNF-α, a pro-inflammatory or a pro-apoptotic cytokine may contribute to apoptosis of aged T cells that express high TNFR1 (receptor of TNF-α) [42]. Few previous studies among aging patients (> 60 years) revealed that TNF-α level were significantly higher along with decreased T cell counts and increased levels of PD1 and Tim-3 (T cell exhaustion markers). This study results indicate the role of TNF-α as a negative regulator of T cell proliferation or survival [31]. Nevertheless, few previous studies on COVID-19 reported a negative correlation of TNF-α with T cell count and reported no difference in TNF-α levels in patients with COVID-19 [29, 43-46]. In ARDS, IL-1β and its family (IL-18, IL-33) are significant players to increase the recruitment of immune cells subsequent production of cytokines [47]. IL-1β and TNF-α are required to develop Th17 cells and assist in Th17 mediated immune response and increased vascular permeability [47]. IL-17 and GM-CSF, cytokines of the Th17 pathway increasing in patients with severe COVID-19 have researchers urgently investigate the role of Th17 in severe COVID-19 cases [48, 49]. Th17 cell-increased expression in the peripheral blood of patients with COVID-19 indicates a player in the COVID-19 cytokine storm as reported in the patients with MERS and SARS [50]. Some Th17 pathway-specific cytokines, such as GM-CSF, IL-1β, IL-17, and TNF-α are elevated in severe COVID-19 patients A case study on severe COVID-19 demonstrated an elevated count of Th17 cells, activated CD4+, and CD8+ T cells, whereas another previous study revealed a decrease in Th17 subset indicated by low IL-17 secretion urges the need to investigate the role of Th17 specific response in COVID-19 [48, 51 & 52]. An increased IFN-γ, IL-1β, IP-10, and MCP-1 serum concentrations contribute to the activation of Th1 cell response and further aggravation the cytokine storm like the occurrence in MERS-CoV and SARS-CoV [53, 54]. A previous study on three COVID-19 patients demonstrated that IL-1 expressed significant expression changes prior to deterioration of the respiratory function, whereas the other pro-inflammatory cytokines were induced only after occurrence of respiratory symptoms. These IL-1 expression changes indicate the role of the IL-1 pathway in the initial progression of COVID-19-associated pulmonary immunopathology and the IL-1 receptor signaling in respiratory epithelium-inflammatory damage [55]. An increase in anti-inflammatory cytokines of Th2 cells (IL-4 and IL-10) were identified in SARS-CoV-2 (COVID-19) patients, in contrast to SARS without clarification [56].

In severely ill-COVID-19 patients (approximately 15 % of patients with COVID-19), there were significant elevated levels of serum G-CSF, IL-2, IL-7, IL-10, IL-17, MCP-1, MIP-1A, and TNF-α in comparison to non-severely ill cases, indicated diverse cytokine profile in the two groups and cytokine storm in disease progression and severity involvement [29]. Cytokine storm accompanies the transition from mild to severe form of
COVID-19. A recent study on immunophenotype the anti-viral response in 4 COVID-19 patients by using single-cell transcriptome sequencing of peripheral blood monocytes (PBMCs), collected at the periods of pre-Intensive Care Unit (ICU) stay, ICU stay, and post-ICU stay demonstrated that there was a significant increase in monocytes and plasmacytoid dendritic cell populations in the ICU-stay specimens compared to the pre- and post-ICU-stay samples [57]. A reported gene signature in the ICU specimens demonstrated elevation of expression of interferon-stimulated genes (ISGs), such as IFITM1 when compared to pre- and post-ICU specimens, indicating a significant viral load regulated type I interferon response (Type-I IFN response) in onset of ARDS and disease progression [57]. Thus, there is evidence of dampened or delayed type interferon response in the initial phase of COVID-19 infection with subsequent increase in active viral replication that is a part of SARS-CoV pathway [58, 59].

**Dysregulation of Immune Cell Subset**

COVID-19 viral replication induced pyroptosis contributes to release of Damage-Associated Molecular Pattern (DAMPs), IL-1, and Pathogen-Associated Molecular Pattern (PAMPS), such as ATP, ASC oligomers, and viral nucleic acid that induce the production of pro-inflammatory cytokines and chemokines, such as IL-6, IL-10, MCP-1, MIP-1α, MIP-1β from alveolar macrophages and adjacent epithelial cells [29, 54]. Therefore, the resultant microenvironment attracts both innate and adaptive immunity cells and pro-inflammatory cascade sets [29, 54]. Pulmonary macrophages are the primary players in both uncontrolled immunopathology of COVID-19 and effective host immunity against COVID-19 that contribute to cytokine storm by secretion of IFN-α, IFN-γ, IL-12, IL-23, IL-1β, and TNF-α. This critical patients, characterized by reduced levels of IFN-α and IFN-β accompanying with high levels of IL-6 and TNF-α. This study also revealed a significant downregulation of 6 ISGs that specify type-I IFN response in severe COVID-19 patients and reduction of pDC population compared to healthy subjects [63].

**Conclusion**

SARS-CoV-2 (COVID-19) infection has evolved to become a pandemic, in contrast to infections with SARS and MERS, whereas SARS-CoV-2 (COVID-19) has demonstrated having the similarities of genome sequence, receptor affinity, pathogenesis, and disease manifestation. Although genomic evidence does not support the belief that COVID-19 is a laboratory construct, currently it is impossible to disprove or prove the theories of its origin. To identify the COVID-19 origin, obtaining virus sequences from immediate non-human animal sources would be the most definite method. Additionally, experimental studies of the role of the polybasic cleavage site and predicted O-linked glycans and receptor binding would be more helpful to obtain more viral genetic data. For substantially refining of phylogenetic estimates of epidemic size and growth rate of COVID-19 in Wuhan, Hubei province and mainland China, larger numbers of more systematically sampled sequences from across China are needed. Only a small proportion of COVID-19-infected patients progress to severe COVID-19 requiring critical care. In the absence of proper cure of COVID-19, it is necessary to identify the factors that may assist in assessment of the COVID-19 disease severity before rapid progression of the disease. Nevertheless, our knowledge of SARS or MERS has not enough to restrain current COVID-19 pandemic.

**References**

coronavirus infection in hospitalized infants under 1 year of age in China. JAMA 323: 1313-1314.


