Host Genetic and Epigenetic Factors in Determining Clinical Outcome of Coronavirus Disease-2019

Tanveer Khalid¹, Ali Amar² and Shagufta Khaliq³

ABSTRACT
The infection caused by Severe Acute Respiratory Syndrome-Corona Virus 2 (SARS-CoV-2) has rapidly emerged as a serious pandemic, causing substantial morbidity and sometimes mortality with a significant healthcare burden. Unfortunately, Pakistan is among top twenty countries of the world affected by COVID-19. The clinical spectrum in COVID-19 is diverse, ranging from mild disease having flu-like symptoms to potentially fatal ARDS, cytokine storm, multiple organ failure and death. Common risk factors associated with severe outcome in COVID-19 infection include male gender, older age and presence of comorbidities such as hypertension, diabetes and cardiovascular disease. Here we reviewed the available literature and report that the underlying mechanisms that account for a broad range of symptoms during respiratory viral infections, that have been well studied in the case of influenza viruses, adenoviruses, SARS-CoV and MERS-CoV, suggest that host genetic and epigenetic factors may also play a significant role in determining susceptibility and clinical outcome in COVID-19 infection. In this review we discuss the potential roles of host genetic factors including cellular receptors for COVID-19, HLA and inflammatory cytokine genes. Based on the SARS-CoV-2 genome and protein-protein interactions map between host and viral proteins we also describe the potential roles of several viral proteins in epigenetic modulation of host inflammatory innate immune response by targeting different cellular pathways particularly NF-κB activation, which may lead to the inflammatory cytokine storm and a severe COVID-19 disease. Investigations of these genetic and epigenetic mechanisms during COVID-19, especially in local settings, will be helpful in management of patients with higher risks and in the development of novel antiviral therapeutics.

KEYWORDS: SARS-CoV-2, COVID-19, Host genetics, Host epigenetics.

How to Cite This:

INTRODUCTION
Severe Acute Respiratory Syndrome-Corona Virus 2 (SARS-CoV-2), the novel Coronavirus, which causes Coronavirus Disease 2019 (COVID-19), a respiratory airway and pulmonary infection, is an enveloped positive-sense, single-stranded RNA beta Coronavirus of the family Coronaviridae.¹,² As of May 19, 2020, over 4.6 million confirmed cases of COVID-19 have been reported from 213 different countries/territories including more than 0.3 million confirmed deaths³ which caused global social and economic disruption. Pakistan is the 18th most affected country from COVID-19 in the world, as during fourth week of May 2020 > 64,000 cases have been reported along with > 1300 deaths.⁴ The current estimate for global mortality rate is
4.7%, however, this can vary widely by location from a high of 10.8% in Italy to a low of 0.7% in Germany.\textsuperscript{5} causes Coronavirus Disease 2019 (COVID-19), a The clinical spectrum in COVID-19 is diverse, where most patients present with only a mild disease having flu-like symptoms (\textapprox 80\%), whereas others experience severe (\textapprox 14\%) and potentially fatal (\textapprox 5\%) phenotypes including severe interstitial pneumonia, Acute Respiratory Distress Syndrome (ARDS), cytokine storm, cardiomyopathy and death. Still, about 1\% COVID-19 cases remain asymptomatic.\textsuperscript{6}

The risk factors associated with severe outcome in COVID-19 infection include male gender, older age and presence of comorbidities such as hypertension, diabetes and cardiovascular disease.\textsuperscript{7,8} However, presence of significant number of COVID-19 patients that developed the severe disease with fatal outcome despite absence of any of the above mentioned risk factors and lessons from previous outbreaks of respiratory viral infections, especially severe acute respiratory syndrome (SARS), Middle East Respiratory Syndrome (MERS) and influenza, suggest that host genetic and epigenetic factors may play a significant role in determining susceptibility and clinical outcome in COVID-19 infection.\textsuperscript{9}

Here we present a brief account of the SARS-CoV-2 viral genome and also of host genetic and epigenetic factors that can influence clinical outcome in COVID-19 disease, also referencing Pakistani COVID-19 scenario in this context.

\textbf{Lessons from Previous Outbreaks of SARS-CoV and MERS-CoV}

During last two decades, the highly pathogenic human Coronaviruses including SARS-CoV in 2002 infected about 8,000 people worldwide with \textapprox 10\% mortality rate and MERS-CoV during 2012 infected around 2,500 people with a much higher (36\%) mortality rate. ARDS and long-term reduction in lung function, arrhythmia or even death can result due to infection with these highly pathogenic respiratory viruses. As compared to MERS-CoV and SARS-CoV, the SARS-CoV-2 has lower mortality rate but its efficient spread made it difficult to contain.\textsuperscript{10} For the development of advanced therapeutic approaches against SARS-CoV-2 infection and the associated COVID-19 pathology, it is important to gain insight into the viral, host genetic and epigenetic mechanisms at molecular level which are utilized by the Coronaviruses to hijack the host immune system.

To date, no clinically available antiviral drugs have been developed for SARS-CoV-2, SARS-CoV, or MERS-CoV. Nucleoside analogue, RNA-dependent RNA Polymerase (RdRP) inhibitor Remdesivir is in clinical trials against COVID-19.\textsuperscript{11} and more recently a laboratory animal trial data suggests a new nucleoside analogue may be effective against SARS-CoV-2 infection.\textsuperscript{12} Clinical trial of repurposed human transmembrane protease serine 2 (encoded by the gene TMPRSS2) inhibitors and several vaccine candidates are also underway.\textsuperscript{13,14} Therapeutic strategies targeting the interaction of viral proteins with host genetic and epigenetic factors with least chances of viral evasion by mutations, could potentially present a long term and broad range of treatment modalities,\textsuperscript{15} but, lack of detailed knowledge of the molecular mechanisms of SARS-CoV-2 infection is a limiting factor.

\textbf{SARS-CoV-2 Genome Sequence Analysis}

SARS-CoV-2 sequence analysis reveals that 30kb genome encodes 14 open reading frames (Orfs). The Orf1a/Orf1ab encodes a polyprotein, which is auto-proteolytically cleaved into 16 non-structural proteins (Nsp1-16) that form the replicate/transcriptase complex (RTC). The RTC is a multiple enzyme complex, including Nsp3 (papain-like protease), Nsp5 (main protease), the Nsp7-Nsp8 primase complex, Nsp12 (primary RNA-dependent RNA polymerase), Nsp13 (a helicase/triphosphatase), Nsp14 (exoribonuclease), Nsp15 (endonuclease), and Nsp10/Nsp16 (N7- and 2'-O-methyltransferases).\textsuperscript{1,15,17} At the 3’ end of the SARS-CoV-2 genome, 13 Orfs are expressed from nine predicted sub-genomic RNAs, which express four structural proteins, Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N), and nine putative accessory factors.

\textbf{SARS-CoV-2 and Human Protein-Protein Interaction Map}

The protein-protein interactions between SARS-
CoV-2 and human proteins may involve several biological pathways and complexes, including Nsp1 in DNA replication, Nsp5, Nsp8, Nsp13 and E in epigenetics and gene expression regulation, Nsp2, Nsp6, Nsp7, Nsp10, Nsp13, Nsp15, Orf3a, Orf8, E and M in trafficking of vesicles, N and Nsp8 in regulation and RNA processing and modifications of lipid (Spike), Orf10 ubiquitin ligases, Orf9b, N, Nsp8 and Nsp13 in signaling, Orf6, Nsp9 and Nsp15 in nuclear transport machinery, Nsp1 and Nsp13 with cytoskeleton, Orf9c, Nsp4 and Nsp8 mitochondria, and Nsp9 with extracellular matrix. Nearly 40% of SARS-CoV-2 interacting proteins are linked with trafficking of vesicles and endomembrane compartments.\[18\]

Potential Role of Host Genetic Factors in COVID-19

Host genetic factors that can influence viral pathogenicity in COVID-19 infection may include cellular receptors, such as angiotensin converting enzyme II (ACE2)\[19\] and other viral entry facilitating host proteins for example transmembrane protease serine 2 (TMPRSS2) and furin endoprotease encoded by the FUR gene (FES upstream region)\[20\] for SARS-CoV-2. Similarly, major histocompatibility-complex antigen loci (HLA) and other markers of immune response in COVID-19 infection such as inflammatory cytokine genes including interleukin 1, interleukin 6 and tumor necrosis factors (IL-1, IL-6 and TNFs) are also prototypical candidates for genetic susceptibility to severe COVID-19 infection.\[21\] In addition, unexplored novel loci specific for SARS-CoV-2 may also contribute towards individual variability in COVID-19 clinical response. Functional genetic variants (both common and rare), epigenetic changes and differential expression profile of these candidate COVID-19 pathogenesis loci, as analysed by multi-OMICS tools including genome-wide association study (GWAS), whole exome/genome sequencing (WES/WGS), epigenome and transcriptome, may help predict the disease course in COVID-19 infection.

Understanding the desperate need to gain insights into COVID-19 pathogenesis and potential prophylactic/treatment options, the international human genetics community has launched “COVID-19 host genetics initiative” available at https://COVID19hg.netlify.com/. This collaborative effort provides a suitable platform to share resources and data and involves multi-OMICS to identify individuals at unusually high or low risk, and contribute to global knowledge of the biology of SARS-CoV-2 infection and disease. More than 50 studies from different genetic research groups across the globe have already registered their studies with this initiative, demonstrating importance of COVID-19 host genetics research and its potential implications in combating COVID-19 pandemic. A couple of preliminary studies that analysed ACE2 genetic variance and expression patterns already show promising results in this regard.\[20,22\] Since differences in genetic landscape between different populations may change relative applicability and usefulness of host genetic determinants of COVID-19 pandemic, therefore, studies from diverse ethnic backgrounds, including that of Pakistani or Indian sub-continent, are needed.

Host Epigenetics in COVID-19 and Innate Immune Response

Environmental cues induce changes in chromatin which results in regulation of gene expression, which plays a crucial role in multiple biological and pathological pathways. Epigenetic regulation involves several coordinated mechanisms, predominately DNA modifications (methylation), histone modifications (acetylation and methylation), remodelling of chromatin structure and non-coding RNAs (ncRNAs).\[23\] During any microbial infection, epigenetic regulation not only involves programming and reprogramming of host chromatin states and gene expression for inducing innate immune responses, but also facilitates immune response circumvention by pathogen.\[24\] Coronavirus Disease 2019 (COVID-19), a During Coronavirus infection innate immunity is the earliest protective barrier. Following the viral infection pathogen-associated molecular patterns (PAMPs) of viral structure proteins and viral genome are recognized by pathogen recognition receptors (PRRs) including Toll-like receptors (e.g., TLR 3, 4 and 7), retinoic acid-inducible gene 1 (RIG-I), melanoma differentiation-associated gene 5 (MDA5) and other recognizing molecules. In vivo,
numerous other key players of innate immune response for example MyD88 (myeloid differentiation primary response gene 88), TRIF (TLR4, TLR7 and TLR3/TIR-domain-containing adapter inducing interferon-β) and STAT1 (signal transducer and activator of transcription 1) also perform their roles to reduce severity of Coronavirus infection. Moreover, interferons (IFN-α (α), IFN-beta (β), and IFN-gamma (γ)) also play important roles in controlling SARS-CoV both in vitro and in vivo.\textsuperscript{25} Studies during the 2002/2003 outbreak revealed the correlation between the outcomes of SARS disease and differential IFN and interferon-stimulated gene (ISG) expression level in patients. The protective functions of MyD88, select ISGs and TLRs against SARS-CoV pathogenesis have been demonstrated in different studies involving mouse models.\textsuperscript{26,27,28,29}

But the Coronavirus, for example MERS-CoV and SARS-CoV, have evolved certain genetic mechanisms which antagonize and/or delay the ISG effector functions and viral recognition. Several proteins encoded by SARS-CoV antagonize the interferon induction, thus modulating the innate immune response in host. Many non-structural proteins encoded by ORF1a/b including Nsp1, Nsp3, Nsp14 and Nsp16 oppose NF-κB, different signalling and sensing pathways, evade interferon-induced protein with tetratrico peptide repeats (IFIT) 1-3 ISGs and play role in capping of viral messenger RNAs (mRNAs).\textsuperscript{30,31,32}

Studies conducted during the past few decades provide evidences to support the hypothesis that cytoplasmic replicating lytic RNA viruses, including Coronavirus and influenza viruses have evolved and developed the intricate and highly coordinated processes which control the host epigenetics, consequently taking over the control of host antiviral innate immune defence. Which results in robust viral replication and sever pathogenesis.\textsuperscript{25}

**NF-κB Mediated Epigenetic Modulation of Innate Immunity by SARS-CoV-2**

During replication in cytoplasm Coronavirus elicit expression of inflammatory genes by several mechanisms. Transcriptomics analysis of HCoV-229E (human Coronavirus 229E) infected HuH7 and A549 cells demonstrated a specific set of upregulated genes. In comparison with stimulation of cells with interleukin (IL)-1 (inflammatory cytokine). HCoV-229E replication resulted in attenuation of inducible activity of transcription factor NF-κB. Moreover, nuclear concentration of NF-κB restricted by partial degradation of NEMO, IKKβ, IκBα and by upregulation of tumor necrosis factor α-induced protein 3 (TNFAIP3/A20). Although for efficient replication of virus, constitutive activity of IKK and basal expression levels of TNFAIP3 is required. In addition, characterization of enhancers and actively transcribed genomic regions in HCoV-229E infected cells, demonstrated the correlation between genome-wide expression changes and histone modifications (H3K4me1, H4K5ac, H3K9ac, H3K27ac, H3K36ac) and Ser5-phosphorylated RNA polymerase II recruitment. Thus, providing evidence that, human Coronavirus infection can result in fine-tuning of NF-κB signalling pathway at multiple levels and regulation of host chromatin landscape.\textsuperscript{33}

Although there are several pathways activated during inflammatory immune response but NF-κB is considered as principal factor in asthmatic inflammation, as it is activated by different extracellular stimuli, for example viruses, interleukin-1 β (IL-1 β), tumor necrosis factor α (TNF α) and various immune challenges. NF-κB expressed ubiquitously within cells, it enhances the activity of other cell and signal specific transcription factors in addition to its role in regulating the expression of inflammatory genes. Following activation NF-κB translocate into cell nucleus and binds with sequence specific DNA binding factors in the promoters of responsive genes.\textsuperscript{34,35} In a temporal manner NF-κB can induce histone modifications including histone acetylation\textsuperscript{36} which results in expression of inflammatory genes by recruitment of co-activator and remodelling complexes.\textsuperscript{37}

In a cell specific manner NF-κB-induced histone acetylation has been reported in human lung epithelial cells, at NF-κB responsive regulatory regions. Moreover, histone acetyltransferases (HATs) association with NF-κB, including members of the p160 family and transcriptional intermediary factor-2 (TIF-2) and steroid receptor coactivator-1 (SRC-1) have been reported.\textsuperscript{38} It has been quite
evident that NF-κB after activation by diverse cellular stimuli regulates the expression of different gene patterns because of temporal NF-κB activation, concentration in nucleus and variable cofactor usage.\textsuperscript{39,40,41}

These results predict that slight alterations in activation profile of NF-κB may have enormous effect on co-factor recruitment and consequent gene induction. For example, p65 (subunit of NF-κB) is inactivated when non-phosphorylated and frequently associated with HDAC1, but phosphorylated p65 following IKK-2 stimulation is activated and can bind to coactivators.

Histone deacetylase inhibitor trichostatin A has been reported to upregulate the NF-κB mediated transcription of inflammatory genes in different cell lines.\textsuperscript{42} Two main pathways for this outcome have been suggested. In the first case it has been described that a NF-κB associated HDAC acts as break against the ability of NF-κB to activate local HAT activity, when bound to DNA. Inhibition of associated HDAC results in elevated local HAT activity and subsequent upregulation of inflammatory gene expression.\textsuperscript{42,43} In an alternate mechanism it is proposed that HDAC3 mediated modification of NF-κB nuclear-cytoplasmic transport and association with IkBα results in higher nuclear retention of activated p65 which is insensitive to inactivation by IkBα. Recent evidences suggested that IkBα mediated sequestering of both HDAC1 and HDAC3 in the cytoplasm results in enhanced NF-κB activity.\textsuperscript{42}

Consistent upregulated expression levels of proinflammatory cytokines including IL-1, IL-6 and TNF, chemokines including IL-8(CXCL8), MCP-1(CCL2) and IP-10(CXCL10) and several other NF-κB target genes have been reported in biopsies of lungs obtained from Coronavirus infected humans and primates, which suggest that CoV infection elicit typical innate immune response involving elevated expression of inflammatory genes. However, some studies also reported a more variable pattern of SARS-CoV, MERS-CoV and HCoV-229E, infected cells gene expression.\textsuperscript{44,45} NF-κB upregulate the expression of proinflammatory genes by engaging transcriptional co-activator proteins having inherent HAT activity.\textsuperscript{46}

The protein-protein interactions map of SARS-CoV-2 proteins with human proteins predicts that Nsp5 the main protease of SARS-CoV-2 impacts the trafficking into mitochondria and endoplasmic reticulum in addition to its interaction with the epigenetic regulator histone deacetylase 2 (HDAC2). The predicted cleavage site between the HDAC domain and the nuclear localization sequence suggests that Nsp5 may inhibit HDAC2 transport into the nucleus and potentially affecting the HDAC2 mediated inflammation and interferon response.\textsuperscript{10,47,48} This interaction of Nsp5 with HDAC2 predicts the virus mediated modification of host epigenetic marks, which results in inappropriate inflammatory innate immune response and leads to severity of symptoms in infected individuals.

Similarly it has been predicted that the transmembrane E protein, of SARS-CoV-2 binds to BRD2 and BRD4, belong to bromodomain and extra-terminal (BET) domain family of epigenetic readers that bind with acetylated histones to regulate gene expression.\textsuperscript{49} The C-terminal region of E resembles the N-terminal region of histone H3, a well-known interacting partner of the bromodomains. Notably, C-terminal region of E is greatly conserved in SARS and bat Coronaviruses, suggesting its conserved essential function during pathogenesis. Interestingly in the NS1 protein of influenza A H3N2 a similar short peptide motif has also been reported, which interferes with host antiviral response supporting transcriptional processes.\textsuperscript{50,51} The interaction between BRDs and protein E can be disrupted by Bromodomain inhibitors (iBETs), which might be helpful in future in the development of novel therapeutic antiviral drug against SARS-CoV-2 infection.

Moreover, numerous host proteins involved in innate immune response signalling are also found to be targeted by SARS-CoV-2 proteins. The TBK1 and TBKBP1 of IFN pathway are targeted by Nsp13, Nsp15 interacts with RNF41/Nrdp1 and Orf9b interact with TOMM70; while the TLE1, 3, and 5 of NF-κB pathway are targeted by Nsp13, Orf9c interacts with NDFIP2, NLRX1 and F2RL1. Also, the two more, E3 ubiquitin ligases, MIB1 and TRIM59 which regulate the antiviral innate immune response signalling, are targeted by Nsp9 and Orf3a, respectively.\textsuperscript{18,52}
COVID-19 Host Genetics and Epigenetics: Pakistani Context

COVID-19 is becoming a serious threat to already under-resourced healthcare system of Pakistan with COVID-19 cases on a rising curve. Pakistani COVID-19 patients have also displayed same clinical variability in disease outcome; however, no studies or data are available on host genetic determinants of COVID-19 infection. A preliminary study using different bioinformatics tools and online databases suggested potential of ACE2 genetic variability based modulation of COVID-19 susceptibility in Pakistani population, warranting detailed studies using wet lab approach in local COVID-19 patients that may enable prediction and stratification of COVID-19 patients that may experience severe disease for informed management and utilization of limited healthcare resources in local clinical settings.

Moreover, urgent need of effective, available, and affordable drugs to control and diminish the COVID-19 pandemic has led to initiation of clinical trials evaluating use of different repurposed drugs, such as anti-malarial drug hydroxychloroquine, in treatment of COVID-19. However, there are concerns regarding efficacy and potential adverse effects of such drugs in some groups of COVID-19 patients, which may be predictable using a pharmacogenetics approach analyzing genetic variants of genes involved in the metabolism of these drugs.

Recently the role of epigenetic dysregulation by hypomethylation of ACE2 and demethylation of interferon-regulated gene, major cytokine genes and NF-κB in increased COVID-19 susceptibility and severity has been reported in lupus patients. Nsp5 the main protease of SARS-CoV-2 interact with the epigenetic regulator histone deacetylase 2 (HDAC2) and affect the HDAC2 mediated inflammation and interferon response. As described previously another putative epigenetic player is transmembrane E protein of SARS-CoV-2 which binds to BRD2 and BRD4 epigenetic readers that in turn interact with acetylated histones to regulate gene expression. These interactions may have a contributing role in elevated inflammatory response and consequent severe variable symptoms in infected individuals, as observed in large number of COVID-19 patients in Pakistan. To date there is no systematic study available describing the underlying host epigenetic mechanisms during COVID-19. Understanding and disruption of these epigenetic pathways will be very helpful in future for the development of novel therapeutic antiviral drug against SARS-CoV-2 infection.

CONCLUSION AND FUTURE RECOMMENDATIONS

Being highly diffusible, SARS-CoV-2 spreads through air droplets, direct contact with infected (symptomatic and asymptomatic) individual and contact with infected objects, having incubation period ranging from 1-14 days. A significant number of infected patients show mild to moderate respiratory symptoms but many infected individuals may remain asymptomatic. Among the infected individuals only 5 – 10% suffer from complete severe respiratory syndrome know as Corona Virus Disease (COVID-19). Although the mortality rate is low (0.2%) in young healthy subjects, but it increases with age and comorbidities, highest mortality rate have been observed in patients older than 80 years with previous history of heart disease. According to published statistics in late March in the CDC’s Morbidity and Mortality Weekly Report (MMWR), approximately 40% of the hospitalized patients suffering from COVID-19 during the period of mid-February and mid-March were between the ages of 20 and 54 years. Among these patients 20% hospitalized and 12% patients in ICUs were between the ages of 20 to 44 years. Among the 44 cases with known consequence, 15 (34%) deaths were reported among adults aged ≥85 years, 20 (46%) among adults aged 65 – 84 years, and nine (20%) among adults aged 20 – 64 years. Fatality percentages were directly proportional to age, no deaths reported among patients aged ≤ 19 years. These statistics indicate that although mortality rate is higher in old age patients but there is significant number of young individuals are also affected with variable severity of disease.

The underlying genetic and epigenetic mechanisms account for a broad range of symptoms during respiratory viral infections have been well studied as in the case of influenza viruses, adenoviruses, SARS-CoV and MERS-CoV.
host and viral proteins interactions and potential of ACE2 genetic variability based modulation of COVID-19 susceptibility and pathogenesis in Pakistani population have been studied by in silico means recently, but wet lab investigations of these and roles of viral factors in regulating the host innate immune response by expression of inflammatory cytokines and regulation of expression of these cytokines by transcription factors particularly the NF-kB pathways in association with epigenetic mechanisms have not been explored yet.

Further investigations in local COVID-19 patients with broad range of disease symptoms will enable prediction and stratification of patients, which will be helpful in management and treatment. Moreover, efficacy and adverse reactions of potential drugs against SARS-CoV-2 infection can be predicted by pharmacogenetics approach.

Reversible epigenetic variations can result rapid phenotypic changes both in pathogen and host infection, there is altered transcription profile of pathogen and host genes, which may lead to efficient clearance of the pathogen by host or immune evasion by pathogen with severity of disease symptoms. As described earlier through protein interaction maps there are several SARS-CoV-2 proteins including Nsp5, Nsp13 and transmembrane protein E which involve in epigenetic regulation of host innate immune response. In order to understand potential roles of chromatin modifiers in innate immune response and inflammation during infection several important questions should be answered including how pathogen-based factors and innate immune factors target the particular chromatin modifiers to induce sequence specific epigenetic regulation.

Moreover, advance epigenetic analysis techniques including epigenome microarray and chromatin immunoprecipitation followed by sequencing (ChIP–Seq) could be utilised to decipher the multiprotein-DNA chromatin complexes mediating epigenetic regulations of immune response during viral infections. Altogether, a new class of therapeutics is required to actively fight SARS-CoV-2 and other viral infections and drugs targeting the epigenetic factors could lead to significant therapeutic innovations in future.

**LIMITATIONS OF STUDY**

As narrative review, in general, does not follow any specified protocol for retrieval and selection of articles, hence bias of the author’s interpretation and conclusions may have occurred.

**CONFLICT OF INTEREST**

None to declare.

**FINANCIAL DISCLOSURE**

None to disclose.

**REFERENCES**


**Author’s Contribution**

**TK, AA:** Conception of study, acquisition of data, drafting of manuscript.

**SK:** Conception of study, critical review with intellectual input, approval of the final version to be published.