From bedside to bench: regulation of host factors in SARS-CoV-2 infection

Samantha Y. Q. Ong¹, Ilya M. Flyamer², Wendy A. Bickmore² and Simon C. Biddie¹,²

Abstract
The zoonotic coronavirus SARS-CoV-2 (severe acute respiratory syndrome coronavirus-2), which causes COVID-19 (coronavirus disease-2019), has resulted in a pandemic. This has led to an urgent need to understand the molecular determinants of SARS-CoV-2 infection, factors associated with COVID-19 heterogeneity and severity, and therapeutic options for these patients. In this review, we discuss the role of host factors in SARS-CoV-2 infection and describe variations in host factor expression as mechanisms underlying the symptoms and severity of COVID-19. We focus on two host factors, angiotensin-converting enzyme 2 (ACE2) and transmembrane serine protease 2 (TMPRSS2), implicated in SARS-CoV-2 infection. We also discuss genetic variants associated with COVID-19 severity revealed in selected patients and based on genome-wide association studies (GWASs). Furthermore, we highlight important advances in cell and chromatin biology, such as single-cell RNA and chromatin sequencing and chromosomal conformation assays, as methods that may aid in the discovery of viral–host interactions in COVID-19. Understanding how regulation of host factor genes varies in physiological and pathological states might explain the heterogeneity observed in SARS-CoV-2 infection, help identify pathways for therapeutic development, and identify patients most likely to progress to severe COVID-19.

Introduction
The emergence of the novel coronavirus (CoV) severe acute respiratory syndrome (SARS) coronavirus 2 (SARS-CoV-2) has resulted in a pandemic. In general, the heterogeneous symptomatology following SARS-CoV-2 infection, which ranges from asymptomatic to severe life-threatening COVID-19, suggests the involvement of host factors. Epidemiological studies to date have implicated several host factors in severe COVID-19, including male sex, older age, cardiovascular disease, obesity, and diabetes¹,². These host factors have behavioral, environmental, and genetic contributions³.

Sequencing of the SARS-CoV-2 genome has enabled comparative studies with other beta-coronaviruses to identify pathogenic features. For example, the structural similarity of the SARS-CoV-2 and SARS-CoV spike (S) proteins has accelerated the identification of angiotensin-converting enzyme 2 (ACE2) as a key host factor for viral entry⁴. The endogenous role of ACE2 is to hydrolyse the vasoconstrictor peptide angiotensin. Attachment of SARS-CoV-2 to host cells is mediated by binding of the viral S1 subunit receptor-binding domain to ACE2, permitting entry and replication⁵. Following ACE2 engagement, proteolytic cleavage of S2 by host proteases, such as plasma membrane-associated type II transmembrane serine protease (TMPRSS2), mediates membrane fusion and cytosolic release of the viral RNA. TMPRSS2 is essential for the establishment of airway infection by SARS-CoV-2 and other zoonotic and human CoVs⁶.

In general, understanding host factor regulation through chromatin structure and DNA-binding factor analyses in various cell types and pathophysiological states is crucial. In this review, we discuss the transcriptional regulation of host factors associated with SARS-CoV-2 and severe COVID-19. We focus on key host factors for SARS-CoV-2 infection, namely, ACE2 and TMPRSS2,
and examine their expression in cell types linked to symptoms. We also explore regulatory control and context-dependent changes in host factor expression associated with infection and risk phenotypes. Furthermore, we review genetic variants that might confer an increased risk of severe COVID-19 and outline methodologies to guide genotype-phenotype correlation for functional validation of host factors.

ACE2 and TMPRSS2 tissue distribution and correlation with clinical presentation

High-throughput sequencing of RNA (RNA-seq) has shown ubiquitous expression of ACE2 and TMPRSS2. ACE2 expression is particularly enriched in the lungs, heart, kidneys, intestine, brain, skin, oral mucosa, and endothelial cells (Fig. 1a). TMPRSS2 is mainly expressed in the prostate, salivary glands, and luminal epithelia of the gastrointestinal, urogenital and respiratory tracts (Fig. 1b). Patterns of ACE2 expression correlate with susceptibility to infection with SARS-CoV-2, leading to the hypothesis of higher SARS-CoV-2 tropism for tissues with higher ACE2 expression and consequent symptomatology.

ACE2 and TMPRSS2 expression and pulmonary pathology

Mainly transmitted through droplets, aerosols, and direct contact, the predominant targets of SARS-CoV-2 are the upper respiratory epithelium and lungs. COVID-19 typically presents with constitutional respiratory symptoms, with the majority of patients having pyrexia and dry cough. The central role of the lung in COVID-19 is supported by a significant rise in SARS-CoV-2 viral load in these organs in more severe disease, which does not occur in stool or serum. In 10-20% of patients, the disease progresses in severity with bilateral ground-glass opacities on chest computed tomography. In the first wave of COVID-19, approximately 5-20% of patients became critically ill, requiring invasive mechanical ventilation due to acute respiratory distress syndrome (ARDS) with bilateral diffuse alveolar damage and cellular exudates on postmortem analysis.

Given their role in SARS-CoV-2 infection, expression of the ACE2 and TMPRSS2 genes in the lung may contribute to the observed pathology. Immunohistochemistry demonstrates the highest expression of the ACE2 protein within the sinonasal cavity and pulmonary alveoli, which are presumptive sites of viral transmission and disease development. Single-cell RNA-seq (scRNA-seq) of healthy adult lung tissue from eight donors showed that <1% of cells in the lung parenchyma expressed ACE2 mRNA, 83% of which were surfactant-producing alveolar type 2 (AT2) cells. A low proportion of ACE2+ cells has been observed in other scRNA-seq datasets, with proportionally higher ACE2+ AT2 cells than other cell types. Consistent with immunohistochemical evidence of ACE2 expression on apical surfaces, this provides compelling evidence for that the virus targets AT2 cells. scRNA-seq studies have also found that a subset of ACE2+ cells strongly coexpress TMPRSS2 and other viral process-related genes, suggesting enhanced vulnerability of AT2 cells to SARS-CoV-2. AT2 cells may thus represent a potential viral reservoir for replication or alternatively undergo viral-induced cell death to initiate barrier damage and inflammation.

The proinflammatory state of COVID-19 may also contribute to ACE2 expression through proinflammatory transcription factor (TF) binding. scRNA-seq of the human lung revealed that ACE2 expression is associated with expression of interferon-stimulated genes (ISGs) and components of the type 1 interferon (IFN) signaling pathway, which is central to the immediate antiviral response. In primary human upper airway epithelial cells, stimulation with IFN-γ and IFN-α2 leads to ACE2 upregulation, possibly through signal transducer and activator of transcription-1 (STAT-1) binding sites in the ACE2 promoter. Similarly, human goblet secretory cells of the nasal epithelia exhibit ACE2 upregulation in response to influenza viruses, likely via an IFN pathway. IFN-induced ACE2 expression may confer susceptibility to severe disease: local induction of IFN and ISG have been observed in the bronchoalveolar lavage (BAL) of critically ill patients, and high circulating IFN levels have been shown to persist in later stages of severe COVID-19. Nevertheless, IFN-induced expression has been shown to represent a non-functional truncated ACE2 isoform, suggesting that conditions of high IFN levels are unlikely to increase viral S-protein-mediated cellular entry to promote infection. Furthermore, the paucity of IFN production and high viral loads observed in the incubation phase suggest a protective role for IFN in early disease, which is currently being evaluated in clinical trials. The relationship between spatiotemporal kinetics and tissue specificity of IFN responses and host factor expression during SARS-COV-2 infection requires further study.

Extrapulmonary involvement and host factor expression

Several extrapulmonary manifestations of COVID-19 have been observed, from benign symptoms involving the eyes and gastrointestinal and nasal systems to severe organ failure involving the neurological, cardiac, and renal systems. Diarrhea occurs in approximately 5% of cases, while headache and anosmia have been described as common early symptoms. Regarding severe cases, acute cardiac injury was seen in 12% of patients in a Chinese cohort. Severe acute kidney injury (AKI) requiring renal replacement therapy is common among critically ill patients, affecting 20-40% of those in the ICU.
Fig. 1 ACE2 and TMPRSS2 mRNA human tissue expression. Representation of RNA-seq-derived median protein-transcripts per million (pTPM) levels of ACE2 (a) and TMPRSS2 (b) genes in 54 nondiseased tissue sites in females (left) and males (right). Organs not labeled include the brain, aorta, coronary artery, heart, lungs, liver, esophagus, stomach, terminal ileum, colon (transverse, sigmoid), adrenal glands, kidney, and bladder. Data were obtained from Genotype-Tissue Expression (GTEx) Portal V8 [67].
The severity and multisystem involvement observed in a subset of COVID-19 patients suggest a possible role for viral interaction with extrapulmonary host factors. Whole-tissue ACE2 mRNA expression and recent scRNA-seq studies reported high ACE2 expression in cardiac pericytes and cardiomyocytes, epithelial cells of the oral mucosa, ileum, stomach, colon and rectum, proximal renal tubules, and endothelial cells. Similar to the lung, coexpression of ACE2 and TMPRSS2 in corneal, intestinal, and nasal epithelial cells is associated with activation of innate immune pathways, high expression in the oral mucosa, ileum, stomach, colon and rectum, cardiac pericytes and cardiomyocytes, epithelial cells of the renal tubules, liver, heart, brain and blood, and smooth muscle cells. TMPRSS2 expression facilitates SARS-CoV-2 S protein-mediated entry in human ACE2 mature enterocytes in human small intestinal enteroids. In the eye, high coexpression of ACE2 and TMPRSS2 has been observed in mouse cornea and human pterygium cell lines, which may explain SARS-CoV-2 detection in the tears and conjunctival secretions of COVID-19 patients with conjunctivitis. Thus, organ systems expressing host entry factors can explain extrapulmonary symptoms such as diarrhea and conjunctivitis.

Given the greater risk of mortality in COVID-19 patients with cardiovascular disease, regulation of ACE2 in the cardiovascular system is a mechanism that may be involved. For instance, upregulated ACE2 expression has been observed in the myocardium of COVID-19 patients with heart failure, and single-nucleus RNA-seq (snRNA-seq) analysis has revealed higher ACE2 expression in myocytes from patients with cardiomyopathy than in healthy controls. Under cardiac stress conditions, ACE2 is repressed while ACE is induced, with consequent increased angiotensin II-mediated cardiac hypertrophy and fibrosis. This response to cardiac stress in endothelial cells is the result of forkhead box M1 (FOXM1) promoter binding. In addition, regulation of ACE2 in the cardiovascular system appears to be cell-specific, as ACE2 was found to be expressed at low levels in cardiac fibroblasts but highly expressed in myocytes, macrophages and endothelial, and smooth muscle cells.

Postmortem studies have provided direct evidence of viral RNA in the lungs and pharynx and, to a lesser extent, in the renal tubules, liver, heart, brain, and blood. Similarly, endothelial dysfunction in postmortem lung and kidney tissue is associated with viral inclusion bodies, suggesting direct viral infection. Hence, differential tissue expression of ACE2 and TMPRSS2 might link viral injury to organ systems and may explain the observed correlation between high respiratory and plasma viral loads and severe COVID-19 and mortality. Other studies, however, have reported no association between viral load and disease severity. This is supported by a systematic meta-analysis of viral load dynamics in COVID-19 whereby no viable virus was detected beyond day nine of the disease, despite prolonged SARS-CoV-2 RNA shedding in respiratory and stool samples. Furthermore, a postmortem study demonstrated disconnect between SARS-CoV-2 detection and organ inflammation and dysfunction. Together, these results suggest that disease progression and mortality may not be directly virus dependent but rather driven by aberrant host immune responses.

**Host factor expression associated with COVID-19 risk factors**

Male sex has been independently associated with severe COVID-19 disease at presentation, the need for intensive care unit admission, and higher case fatality rates. Differential host factor expression may explain the sex difference in COVID-19. A recent meta-analysis of scRNA-seq studies showed increased overall ACE2 and TMPRSS2 lung expression in males; in contrast, others have found no significant sex difference with regard to ACE2 expression. Given the relatively low expression of ACE2 and predominant AT2 cell expression, data for whole tissues compared to single-cell studies may also mask subtle differences. It is also noteworthy that while ACE2 is located on the X chromosome, analysis of GTex data suggests that ACE2 escapes X chromosome inactivation, resulting in sex-biased gene expression. Sex differences in ACE2 expression may also be due to regulation by hormone-dependent TFs. In female mice, ACE2 basal activity in the kidney was repressed by estrogen, suggesting estrogen receptor (ER) response elements at the ACE2 locus. In the elderly human male atrial myocardium, ex vivo administration of estrogen induced ACE2 mRNA through an ERα-dependent mechanism. Although these findings support ACE2 control via an ER response element, differences in the response might vary by tissue type, sex, age, and pathological state.

Dnase-seq has revealed differential shared and tissue-specific regulatory element activation for the ACE2 and TMPRSS2 loci (Fig. 2a), likely regulated by cell-specific TF binding at regulatory elements. ENCODE datasets show binding of the insulator CCCTC-binding factor (CTCF) at shared sites for ACE2 and TMPRSS2, whereas activator protein-1 (AP-1) subunits bind to the tissue-specific 5' upstream element of ACE2, suggesting that variable host factor gene expression may be explained by differential engagement of regulatory elements and selective TF binding.

Regulation of TMPRSS2 has primarily been explored in prostate cancer. In this context, TMPRSS2 is induced by androgens through a distal androgen receptor (AR) binding enhancer, resulting in AR-induced expression of a TMPRSS2-ERG (ETS-related gene) fusion gene conferring prostate cancer pathogenicity. Mapping of an AR
Fig. 2 (See legend on next page.)
response element to the induction of TMPRSS2 expression has led to the hypothesis that androgenic phenotypes are risk factors for COVID-19. Epidemiological studies have also suggested androgenetic alopecia as a separate risk profile for COVID-19. However, AR blockade in mice did not alter lung TMPRSS2 expression, suggesting that lung TMPRSS2 may not be androgen regulated, possibly through tissue-specific enhancer inaccessibility. As TMPRSS2 promotores in mammary epithelial cells have been found to be responsive to estrogens, the role of TMPRSS2 in conferring risk in men through sex-specific expression in relevant tissues remains unclear.

Moreover, sex-biased behavior may contribute to male predominance of COVID-19. Higher rates of smoking among men have been hypothesized to contribute to male predisposition towards the disease, and active smoking is implicated as a negative prognostic indicator for COVID-19. This would be consistent with increased ACE2 in the bronchial epithelium and small airway epithelium of chronic smokers. scRNA-seq data showed dose-dependent upregulation of ACE2 by cigarette smoke in rodent and human lungs, with chronic smoke exposure triggering an expansion of ACE2+ mucus-secreting goblet cells.

Age-related dynamics in host factor expression may similarly explain the greater severity and mortality that occurs among older patients. Indeed, older age is associated with higher viral load in oropharyngeal salivary and respiratory samples. Lung ACE2 expression has been shown to increase ~1.2-fold with every 10-year increase in age. scRNA-seq of developing mouse lungs and temporally resolved RNA-in situ hybridization (ISH) analysis also found an age-related increase in TMPRSS2 expression. Single-nucleus ATAC-seq (snATAC-seq) in human AT2 cells demonstrated age-associated changes in accessible chromatin around TMPRSS2 at sites harboring sequence motifs for multiple TFs, including forkhead box A (FOXA), C/EBPα, the proinflammatory factors STAT, IFN-regulatory factor (IRF), and AP-1, suggesting age-dependent inflammatory regulation of TMPRSS2. High ACE2 and TMPRSS2 levels in the elderly may therefore be associated with increased viral infection, although altered immune response and physiology may also explain divergence in severity with age extremes.

Although lower nasal epithelial ACE2 expression reported in children has been suggested to explain the milder disease observed in this population, others have found that pulmonary ACE2 is not reduced in children. milder disease observed in this population, others have found that pulmonary ACE2 is not reduced in children.

This contradictory evidence suggests a complex relationship between pulmonary ACE2 expression and COVID-19 as both a viral receptor and protective factor, implicating other host factors in regulating disease severity and susceptibility.

Host genetic variants and the spectrum of COVID-19

Host genetic factors have been implicated in influenza and zoonotic diseases such as SARS and MERS. Genetic variants may alter disease through viral infection, replication or immune function by altering host gene expression or protein function, leading to heterogeneous severity.

Genetic variants of ACE2 and TMPRSS2

Human variants of ACE2 could alter the protein interface with the SARS-CoV-2 S protein. Moreover, as ACE2 is located on the X chromosome, hemizygosity for harmful variants might contribute to poor outcomes in men with COVID-19. Genomic variant analysis from Genome Aggregation Database (gnomAD) reveals a very low frequency of non-synonymous ACE2 variants, including those predicted to be loss-of-function (Fig. 3).
Furthermore, no correlation between rare deleterious ACE2 exonic and splice junction variants with COVID-19 was observed in an Italian population\textsuperscript{70}. Given the role of ACE2 in cardiovascular disease and lung disease, including ARDS, ACE2 polymorphisms are of interest beyond COVID-19. For example, a common exonic SNP in ACE2, rs2285666 (Fig. 2a), has been implicated in the risk of cardiovascular disease and type 2 diabetes mellitus, with the rare A allele found at a frequency of 0.2\% in European populations and 0.55\% in East Asians\textsuperscript{70}. Two ACE2 intronic SNPs, rs2074192 and rs2106809, associated with left ventricular hypertension are expression quantitative trait loci (eQTLs) for ACE2 and may therefore link ACE2 cardiac expression with high penetrance of alleles, increasing ACE2 expression in East Asians\textsuperscript{61}. ACE2 variants could therefore be linked to risk factors for COVID-19, but further clarification within comorbid subgroups is required.

For TMPRSS2, the frequency of rare deleterious and common exonic variants speculated to alter protein levels or activity varies between populations. For common variants, two haplotypes have been identified. One is linked to an eQTL harboring rs8134378 and mapping to an AR regulatory element 15 kb upstream of the TMPRSS2 CpG island\textsuperscript{70}. The second includes the intronic SNP rs2070788 (Fig. 2a), which has been found in small-scale genome-wide association studies (GWASs) to be linked to an increased risk of influenza\textsuperscript{71}. Less is known about TMPRSS2 variants, warranting further investigation. However, in the first two published GWASs for severe COVID-19, there was no significant association with ACE2 or TMPRSS2 variants, suggesting that variants of these factors do not contribute to COVID-19 severity\textsuperscript{3,72}.

**Genetic variants in intracellular RNA sensing and type I interferon signaling**

Based upon the gene mutations implicated in life-threatening influenza pneumonia, rare but predicted deleterious mutations have been examined in patients with severe COVID-19\textsuperscript{72}. Mutations in the genes encoding double-stranded RNA sensor TLR3 and in downstream components of type I IFN cell-intrinsic immunity have been detected, including UNC93B1, which forms a complex with and is involved in intracellular trafficking of TLR3\textsuperscript{73} and other downstream components, such as TRIF and TBK1, a key kinase linking pattern recognition receptor activation and activation of ISGs. Mutations in the type I IFN amplification pathway were also found, including in the genes coding for IRF7 and the cell surface receptors IFNAR1 and IFNAR2. In addition, common variants at IFNAR2 were found in a GWAS for severe COVID-19\textsuperscript{3}. Although no rare variants in the JAK/STAT pathway downstream of IFNAR1/2 have been reported in severe COVID-19, variants at TYK2 on chromosome 19 were reported in the GWAS of a UK intensive care cohort of COVID-19 patients with critical respiratory failure\textsuperscript{3}. Of
note, patients with complete TYK2 deficiency are susceptible to bacterial and viral infection. The contribution of genetic variants to reduced IFN signaling in severe COVID-19 is further supported by observations of impaired IFN responsiveness in neutrophils from patients with severe COVID-19. Nonetheless, the WHO Solidarity Trial has shown that treatment of COVID-19 patients with IFN-β1a did not decrease mortality or the need for mechanical ventilation.

Some critical COVID-19 patients with mutations in IFN pathways exhibit reduced serum IFN-α, though other studies have reported higher IFN-γ in helper T cells of patients who died from COVID-19. IFN-γ is linked to IL-6 signaling through the IL-6 receptor (IL6R), suggesting interplay between IFN and interleukin signaling in acute inflammation. Genetic variants of IL6R are associated with a lower risk of hospitalization following SARS-CoV-2 infection, yet these variants were not significant in the COVID-19 GWASs above. Nevertheless, modulation of the IL6R axis using inhibitors, such as tocilizumab and sarilumab, was shown in the REMAP-CAP and EMPACTA trials to reduce the need for mechanical ventilation and reduce mortality in patients with severe COVID-19.

In the first published GWAS for severe COVID-19, a significant cross-replicating locus at 3p21.31 was found in Italian and Spanish populations, spanning six genes: LC6A20, LZTFL1, CCR9, FYCO1, CXCR6, and XCR1. The insertion–deletion GA risk variant of rs11385942 in an intron of LZTFL1 is associated with reduced CXCR6 but increased SLCA20 expression. Interestingly, the risk allele was detected frequently in patients who required mechanical ventilation, and furthermore, the few individuals homozygous for the risk allele were younger, suggesting an association with COVID-19 severity. The frequency of risk allele heterozygosity varies between populations, occurring in less than 0.5% in China and 37.8% in Bangladesh. A GWAS of the UK intensive care cohort of COVID-19 patients with critical respiratory failure replicated the significance of the chromosome 3 locus.

The genes at the 3p21.31 locus have plausible roles in COVID-19 and lung injury. SLCA20 encodes a sodium-dependent amino transporter that has been shown to interact with ACE2 in a Xenopus oocyte expression system. CXCR6 is a marker of lung-resident memory CD8+ T cells, producing cytokines in response to stimuli. Its ligand, CXCL16, is elevated in the serum of patients with acute lung injury, and both ligand CXCL16-CXCR6 receptor genes are upregulated in models of lung injury. CXCL16-CXCR6 acts to disrupt epithelial integrity through downregulation of E-cadherin in lung injury, which might represent the fibrosis observed in ARDS. CCR9, as part of a CCL25-CCR9 chemokine axis, has been shown to mediate gut-specific recruitment of lung IFN-γ-producing CCR9+ CD4+ T-cells in influenza, contributing to the gut injury observed in viral infection. FYCO1, encoding an autophagy adaptor protein, was found to be upregulated in intensive care patients with sepsis.

The second locus uncovered by Ellinghaus et al. maps to the ABO blood group gene, whereby group A confers a higher risk of severe COVID-19 and group O demonstrates a protective effect. The plausibility of an association is supported by an epidemiological study in a Chinese population, showing increased COVID-19 risk with blood group A and reduced risk with blood group O. However, the association of ABO blood groups with COVID-19 severity was not replicated by the GenOMICC study. Additional loci associated with COVID-associated ICU admission include the OAS gene cluster on chromosome 12 and DPP9 on chromosome 19. OAS genes are activated by IFN and are involved in double-stranded RNA degradation pathways in the early innate antiviral response. DPP9 encodes a serine protease implicated in lung fibrosis. These pathways may therefore represent...
therapeutic targets to mitigate progression to critical respiratory failure in COVID-19.

Mechanisms of genetic variants driving increased disease risk

Most GWAS disease-risk loci are in noncoding regions of the genome. The cataloging of human regulatory elements on a genome-scale across multiple cell and tissue types has uncovered a correlation between accessible chromatin, DNA-binding factors and SNPs. These large-scale studies implicate altered TF binding as a core feature of noncoding genetic variants. For example, in lymphoblastoid cell lines, interindividual variation in NF-kB binding has been associated with genetic variants and alterations in gene expression. Genetic variation could, therefore, alter responses to stimuli, such as infection, between individuals and thus influence severity.

To infer phenotypic consequences of genetic variants, functional assignment of genetic variants to gene regulation is required but remains challenging. Chromatin structure can be used to aid functional characterization of variants. Chromatin higher-order structures have been found to form topologically-associated domains (TADs), regions of chromatin that show high interaction frequencies separated by boundary regions. At the megabase scale, TADs are generally stable between cell types and tissues. TAD boundaries are marked by binding sites for CTCF, which impedes cohesin-driven loop extrusion, the process responsible for TAD formation. For the most part, enhancers are generally found in the same TAD as the gene(s) they regulate. Therefore, incorporating information on chromatin structure, including smaller scale accessibility information with larger scale TADs, may aid in the discovery of associations between noncoding variants and genes to guide fine-mapping experiments and has allowed functional refinement of SNPs and eQTLs.

ACE2 and TMPRSS2 are located within tissue-conserved TADs. However, eQTL analysis of SNPs at the ACE2 locus did not reveal an association with gene expression in lung tissue. Within the ACE2 TAD (hg38 chrX:15,300,000–15,600,000), long-distance interactions were observed between multiple SNPs and other genes associated with eQTLs in the lung, such as PIR, a modulator of NF-kB signaling, and CASB, which is involved in acid-base metabolism.

Incorporating open chromatin might also contribute to the identification of functional genetic variants. By applying snATAC-seq to healthy human lungs, age-dependent accessible regions in TMPRSS2-expressing AT2 cells were found to be enriched for genes associated with inflammatory signals. In AT2 cells, age-dependent accessible chromatin near TMPRSS2 correlated with sequence variants, including common variants with a minor allele frequency >1%. These variants are predicted to disrupt sequence motifs of DNA-binding factors and thus alter expression and are associated with GWAS phenotypes for emphysema, asthma, and bacterial pneumonia.

Conclusion

The sequalae following SARS-CoV-2 infection are highly heterogeneous, from asymptomatic to multiorgan failure, and host factors are likely variables involved in COVID-19 severity. Risk factors for COVID-19 contribute to modified disease risk through changes in viral–host interactions, most likely through host factor expression. ACE2 and TMPRSS2 have been identified as SARS-CoV-2 tropic factors. However, ACE2 was found to be protective against pulmonary injury through maintenance of vascular competence in mouse models of acute lung injury. The balance between viral exploitation of host factors and the role of that factor in COVID-19 pathophysiology is not clear. Furthermore, ACE2 and TMPRSS2 genetic variants are not convincingly associated with COVID-19 severity. Other important host factors involved in viral–host interactions, viral protein processing, and immune function are more likely to contribute to severity. The therapeutic effect of glucocorticoids, such as dexamethasone or hydrocortisone, in decreasing COVID-19 mortality or the need for organ support strongly suggests an inflammatory process in severe disease. IFN pathways are also likely to play key roles in COVID-19, and their impairment might explain progression to critical illness. While the WHO Solidarity trial does not support the use of IFN as a treatment modality, questions on the timing and regime remain while highlighting a need for alternate targets. Exploring the expression variation and regulatory control of other host factors, such as the 3p21.31 locus identified from COVID-19 GWASs, will aid in our understanding of the disease spectrum and uncover therapeutic targets. It will also be important to integrate comorbidities with the genetic variants implicated in COVID-19 to understand the contributions of risk factors to COVID-19 severity.

The identification of functional regulatory elements in health and pathological states is a crucial pursuit, with many approaches being developed following the mapping of noncoding genomic regions. We have highlighted a number of methods that have been employed to identify regulatory elements, but their use in conglomerate is likely required, taking a multiomics approach to identify functional consequences. This is particularly important for understanding dynamic changes in regulatory control, for instance, during inflammation. The development of single-cell methodologies, as an example, has also revolutionised our discovery and understanding of cell types and heterogeneous RNA expression in complex tissues.
Nevertheless, functional studies of complex tissues have been largely limited to model organisms, with questionable translational potential. The development of human organoid systems may allow for a more relevant understanding of the molecular processes that drive variation in disease presentation and progression. Combining these approaches with functional assays for regulatory control of gene expression by using, for example, high-throughput CRISPR-based perturbation methods\textsuperscript{105} or a high-throughput self-expressing reporter system to identify functional enhancer elements in cell-specific control\textsuperscript{106}, would allow dissection of the roles of regulatory elements, genetic variants or biological pathways in disease processes such as viral infection. Such approaches will allow the identification of variables to allow risk stratification for public health initiatives, resource allocation for at-risk groups and the development of novel therapeutics.

Acknowledgements

The views expressed are those of the authors and independent from the NHS. S.C.B. is supported by the Scottish Clinical Research Excellence Development Scheme (SCREDS) and the NHS Education for Scotland (NES)/Chief Science Officer (CSO), Scotland, UK. IMF. and W.A.B. are supported by MRC University Unit grant (MC_UU_00007/2).

Conflict of interest

The authors declare no competing interests.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 30 November 2020 Revised: 1 February 2021 Accepted: 15 February 2021.

Published online: 7 April 2021

References

1. Cummings, M. J. et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. \textit{Lancet} \textbf{395}, 1763–1770 (2020).

2. Grasselli, G. et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. \textit{JAMA} \textbf{323}, 1574–1581 (2020).

3. Paro-Castanier, E. et al. Genetic mechanisms of critical illness in Covid-19. Nature https://doi.org/10.1038/s41586-020-03065-y (2020).

4. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural maps and institutional affiliations.

5. Of COVID-19 patients. Cell Host Microbe \textbf{27}, 885–890.e2 (2020).

6. Lukassen, S. et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. \textit{EMBO J} \textbf{39}, e10511 (2020).

7. Ziegler, C. G. K. et al. SARS-CoV-2 receptor ACE2 Is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. Cell \textbf{181}, 1016–1035.e19 (2020).

8. Zhou, Z. et al. Heightened innate immune responses in the respiratory tract of COVID-19 patients. Cell Host Microbe \textbf{27}, 885–890.e2 (2020).

9. Lu, J. D. et al. Expression of SARS-CoV-2 receptor ACE2 and TMPRSS2 in human primary conjunctival and pterygium cell lines and in mouse cornea. \textit{Eye} \textbf{34}, 1212–1219 (2020).

10. Chen, L. et al. Ocular manifestations of a hospitalised patient with confirmed 2019 novel coronavirus disease. \textit{Br. J. Ophthalmol.} \textbf{104}, 748–751 (2020).

11. Chen, L., Li, X., Chen, M., Feng, Y. & Xiong, C. The ACE2 expression in human heart indicates new potential mechanism of heart injury among patients infected with SARS-CoV-2. Cardiovasc. Res. \textbf{116}, 1097–1100 (2020).

12. Yang, J. et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir. Med. \textbf{8}, 420–422 (2020).

13. Ortiz, M. E. et al. Heterogeneous expression of the SARS-CoV-2 receptor ACE2 in the human respiratory tract. \textit{EBioMedicine} \textbf{60}, 102576 (2020).

14. Zhao, Y. et al. Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. \textit{Am. J. Respir. Crit. Care Med.} \textbf{202}, 756–759 (2020).

15. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature \textbf{587}, 619–625 (2020).

16. Lukassen, S. et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. \textit{EMBO J} \textbf{39}, e10511 (2020).

17. Ziegler, C. G. K. et al. SARS-CoV-2 receptor ACE2 Is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. Cell \textbf{181}, 1016–1035.e19 (2020).

18. Zhou, Z. et al. Heightened innate immune responses in the respiratory tract of COVID-19 patients. Cell Host Microbe \textbf{27}, 885–890.e2 (2020).

19. Lucas, C. et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. Nature \textbf{584}, 463–469 (2020).

20. Onabajo, O. O. et al. Interferons and viruses induce a novel truncated ACE2 isoform and not the full-length SARS-CoV-2 receptor. Nat. Genet. \textbf{52}, 1283–1293 (2020).

21. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Sci. \textit{Adv.} \textbf{6}, 719–724 (2020).

22. Blanco-Melo, D. et al. IMbalanced host responses to SARS-CoV-2 drives development of COVID-19. Cell \textbf{181}, 1036–1045.e9 (2020).

23. Hung, I. F. N. et al. Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. Lancet \textbf{395}, 1695–1704 (2020).

24. Guan, W. et al. Clinical characteristics of coronavirus disease 2019 in China. N. Engl. J. Med. \textbf{382}, 1708–1720 (2020).

25. Giaconelli, A. et al. Self-reported olfactory and taste disorders in SARS-CoV-2 patients: a cross-sectional study. Clin. Infect. Dis. \textbf{71}, 889–890 (2020).

26. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet \textbf{395}, 497–506 (2020).

27. Qi, F., Qian, S., Zhang, S. & Zhang, Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. Biochem. Biophys. Res. Commun. \textbf{526}, 135–140 (2020).

28. Zou, X. et al. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Front. Med. \textbf{14}, 185–192 (2020).

29. Sungnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat. Med. \textbf{26}, 681–687 (2020).

30. Zang, R. et al. TMPRSS2 and TMPRSS4 promote SARS-COV-2 infection of human small intestinal enterocytes. \textit{Sci. Immunol.} \textbf{5}, eabc3582 (2020).

31. Ma, D. et al. Expression of SARS-CoV-2 receptor ACE2 and TMPRSS2 in human primary conjunctival and pterygium cell lines and in mouse cornea. \textit{Eye} \textbf{34}, 1212–1219 (2020).

32. Chen, L. et al. Ocular manifestations of a hospitalised patient with confirmed 2019 novel coronavirus disease. \textit{Br. J. Ophthalmol.} \textbf{104}, 748–751 (2020).

33. Chen, L., Li, X., Chen, M., Feng, Y. & Xiong, C. The ACE2 expression in human heart indicates new potential mechanism of heart injury among patients infected with SARS-CoV-2. Cardiovasc. Res. \textbf{116}, 1097–1100 (2020).

34. Yang, J. et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir. Med. \textbf{8}, 420–422 (2020).

35. Grobe, J. L. et al. ACE2 overexpression inhibits hypoxia-induced collagen production by cardiac fibroblasts. Clin. Sci. \textbf{137}, 357–364 (2007).

36. Puelles, V. G. et al. Multigain and renal tropism of SARS-CoV-2. N. Engl. J. Med. \textbf{383}, 590–592 (2020).

Official journal of the Korean Society for Biochemistry and Molecular Biology
37. Varga, Z. et al. Endothelial cell infection and endothelitis in COVID-19. Lancet 395, 1417–1418 (2020).
38. Pujadas, E. et al. SARS-CoV-2 viral load predicts COVID-19 mortality. Lancet Respir. Med. 8, e70 (2020).
39. Fajgowasser, J. et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. Nat. Commun. 11, 1–9 (2020).
40. Argyropoulos, K. V. et al. Association of initial viral load in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) patients with outcome and symptoms. Am. J. Pathol. 190, 1881–1887 (2020).
41. Dorward, D. A. et al. Tissue-specific ImmunePathology in Fatal COVID-19. Am. J. Respir. Crit. Care Med. https://doi.org/10.1164/rcrm.202008-3265oc (2020).
42. Jin, J. M. et al. Gender differences in patients with COVID-19: focus on severity and mortality. Front. Public Health 8, 152 (2020).
43. Muus, C. et al. Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells. Preprint at https://doi.org/10.1101/2020.04.19.049254 (2020).
44. Chen, Y., Shan, K. & Qian, W. Asians do not exhibit elevated expression or functional activity of the SARS-CoV-2 receptor ACE2. Preprint at https://www.preprints.org/manuscript/202002.0528/v2 (2020).
45. Li, G. et al. Assessing ACE2 expression patterns in lung tissues in the pattern of leukocyte infiltration seen during acute in Inflammation and COPD patients: implications for COVID-19. Eur. Respir. J. 61, 1522–1528 (2020).
46. Tukiainen, T. et al. Landscape of X chromosome inactivation across human biosamples. Genome Biol. 21, e145 (2020).
47. Liu, J. et al. Sex differences in renal angiotensin converting enzyme 2 (ACE2) expression and localization of SARS-CoV-2 in lung epithelium. J. Exp. Med. 211, 2397–2410 (2019).
48. Bukowska, A. et al. Protective regulation of the ACE2/ACE gene expression by estrogen in human atrial tissue from elderly men. Exp. Mol. Med. 242, 1424–1423 (2017).
49. Schreiber, J., Bimes, J. & Noble, W. S. Completing the ENCODE3 compendium yields accurate imputations across a variety of assays and human biosamples. Genome Biol. 21, 82 (2020).
50. Biddle, S. C. et al. Transcription factor AP1 potentiates chromatin accessibility and expression of the SARS-CoV-2 receptor ACE2. Preprint at https://doi.org/10.1101/6616961, molcell201106016 (2011).
51. Lin, B. et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res. 59, 4180–4184 (1999).
52. Tomlins, S. A. et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Sci. (80-59) 310, 644–648 (2005).
53. Lam, P. J. R., Rébillard, X., Vacherot, F. & de la Taille, A. Androgenic hormones and the excess male mortality observed in COVID-19 patients: new considerations. J. Exp. Med. 211, 1522–1528 (2020).
54. Rey, D. et al. Orhanic acidosis and mortality in severe COVID-19. Lancet. Respir. Med. 8, e3 (2020).
55. Combes, A. J. et al. Global absence and targeting of protective immune states in severe COVID-19. Nature https://doi.org/10.1038/s41591-020-02334-7 (2021).
56. WHO Solidarity Trial Consortium, Pan, H., Petros, R., AM, H.-R. & Preziosi, M. Repurposed antiviral treatments for COVID-19— interim WHO solidarity trial results. N. Engl. J. Med. 383, 497–511 (2020).
57. Gardotti, A. C. et al. IFN-y is an independent risk factor associated with mortality in patients with moderate and severe COVID-19 infection. Virus Res. 289, 198171 (2020).
58. Hurst, S. M. et al. IL-6 and its soluble receptor orchestrate a temporal switch in the pattern of leukocyte recruitment seen during acute inflammation. Immunity 14, 709–714 (2007).
59. Gacott, A. C. et al. Androgenetic alopecia present in the majority of patients hospitalized with COVID-19: the “Gabin sign.” J. Am. Acad. Dermatol. 83, 680–682 (2020).
60. Baratchian, M. et al. Sex, androgens and regulation of pulmonary AR, TMPRSS2 and ACE2. bioRxiv Prepr. Serv. Biol. https://doi.org/10.1101/2020.04.21.051201 (2020).
61. Baena, E. et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. Genes Dev. 27, 683–698 (2013).
62. Parascandola, M. & Xiao, L. Tobacco and the lung cancer epidemic in China. Transl. Lung Cancer Res. 8, S21–S30 (2019).
63. Breidenbach, J. D. et al. Impact of comorbidities on SARS-CoV-2 viral entry-related genes. J. Pers. Med. 10, 1–11 (2020).
64. Leung, J. M. et al. ACE-2 expression in the small airway epithelia of smokers and COPD patients: implications for COVID-19. Eur. Respir. J. 55, 2000688 (2020).
65. Leung, J. M. et al. ACE-2 expression in the small airway epithelia of smokers and COPD patients: implications for COVID-19. Eur. Respir. J. 55, 2000688 (2020).
66. Smith, J. C. et al. Cigarette smoke exposure and inflammatory signaling increase the expression of the SARS-CoV-2 receptor ACE2 in the respiratory tract. Dev. Cell 53, 514–523.e3 (2020).
67. Chen, J. et al. Individual variation of the SARS-CoV-2 receptor ACE2 gene expression and regulation. Aging Cell 19, e13168 (2020).
68. To, K. K. W. et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect. Dis. 20, 565–574 (2020).
69. Schuler, B. A. et al. Age-determined expression of priming protease TMPRSS2 and localization of SARS-CoV-2 in lung epithelium. J. Clin. Invest. 131, e140766 (2020).
70. Wang, J. et al. Respiratory influenza virus infection induces intestinal immune injury via microbiomodulated Th17 cell-dependent inflammation. J. Exp. Med. 211, 2397–2410 (2014).
71. Bunyavichit, S., Do, A. & Vicenço, A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. JAMA 363, 2427–2429 (2020).
72. Dong, Y. et al. Epidemiology of COVID-19 among children in China. Pediatrics 145, 20200702 (2020).
73. Lee, N. et al. IFITM3, TLR3, and CD55 Gene SNPs and cumulative genetic risks for severe outcomes in chinese patients With H1N1/H1N1pdm09 influenza. J. Infect. Dis. 216, 97–104 (2017).
74. Tu, X. et al. Functional polymorphisms of the CCL2 and MBL2 genes cumulatively increase susceptibility to severe acute respiratory syndrome coronavirus infection. J. Infect. 71, 101–109 (2015).
75. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 14,154 humans. Nature 581, 433–443 (2020).
76. Astelita, R., Paraboshi, E., Montovani, A. & Duga, S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. Aging (Albany N.Y.) 12, 10087–10098 (2020).
77. Sheppard, K. M. et al. IFNAR2 is required for anti-influenza immunity and alters susceptibility to post-influenza bacterial superinfections. Front. Immunol. 9, 2589 (2018).
78. Fingerlin, T. E. et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat. Genet. 45, 613–620 (2013).
79. Maunder, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195 (2012).
91. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
92. Kasowski, M. et al. Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010).
93. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* https://doi.org/10.1038/nature11082 (2012).
94. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
95. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
96. Szabo, Q.; Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
97. Kumasaka, N.; Knights, A. J. & Gaffney, D. J. High-resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nat. Genet.* **51**, 128–137 (2019).
98. Chen, C. H. et al. Determinants of transcription factor regulatory range. *Nat. Commun.* **11**, 2472 (2020).
99. Fadason, T. et al. A transcription regulatory network within the ACE2 locus may promote a pro-viral environment for SARS-CoV-2 by modulating expression of host factors. Preprint at https://doi.org/10.1101/2020.04.14.042002 (2020).
100. Wang, A. et al. Single cell multiomic profiling of human lung reveals cell type-specific and age-dynamic control of SARS-CoV2 host genes. *Elife* **9**, e62522 (2020).
101. Zhang, X. et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583**, 437–440 (2020).
102. Imai, Y. et al. Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436**, 112–116 (2005).
103. RECOVERY Collaborative Group et al. Dexamethasone in hospitalized patients with COVID-19—preliminary report. *N. Engl. J. Med.* **384**, 693–704 (2020).
104. Angus, D. C. et al. Effect of hydrocortisone on mortality and organ support in patients with severe COVID-19: The REMAP-CAP COVID-19 corticosteroid domain randomized clinical. *JAMA* **324**, 1317–1329 (2020).
105. Liu, B. et al. Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nat. Commun.* **11**, 164 (2020).
106. Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
107. GTEx Consortium, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
108. Leung, D. et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518**, 350–354 (2015).
109. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).