Received: 14 December 2020 | Revised: 10 February 2021 | Accepted: 11 February 2021

DOI: 10.1111/dgd.12717

REVIEW ARTICLE

Development, Growth & Differentiation WILEY

Genetic variation analyses indicate conserved SARS-CoV-2–host interaction and varied genetic adaptation in immune response factors in modern human evolution

Ji-Won Lee[1] | In-Hee Lee[2] | Takanori Sato[1] | Sek Won Kong[2,3] | Tadahiro Iimura[1]

[1]Department of Pharmacology, Faculty and Graduate School of Dental Medicine, Hokkaido University, Sapporo, Japan

[2]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA

[3]Department of Pediatrics, Harvard Medical School, Boston, MA, USA

Correspondence
Tadahiro Iimura, Department of Pharmacology, Faculty and Graduate School of Dental Medicine, Hokkaido University, N13W7, Kita-Ku, Sapporo 060-8586, Japan.
Email: iimura@den.hokudai.ac.jp

Funding information
Grant-in-Aids for Scientific Research from the Japan Society for the Promotion of Science (JSPS KAKENHI), Grant/Award Number: 18H02983, 18K19649 and 19K10044; National Institution of Health, Grant/Award Number: R01MH107205, R24OD024622 and U01TR002623; Takeda Science Foundation, Grant/Award Number: PK43200005

Communicating Editor: Harukazu Nakamura

Abstract

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a pandemic as of early 2020. Upon infection, SARS-CoV-2 attaches to its receptor, that is, angiotensin-converting enzyme 2 (ACE2), on the surface of host cells and is then internalized into host cells via enzymatic machineries. This subsequently stimulates immune response factors. Since the host immune response and severity of COVID-19 vary among individuals, genetic risk factors for severe COVID-19 cases have been investigated. Our research group recently conducted a survey of genetic variants among SARS-CoV-2-interacting molecules across populations, noting near absence of difference in allele frequency spectrum between populations in these genes. Recent genome-wide association studies have identified genetic risk factors for severe COVID-19 cases in a segment of chromosome 3 that involves six genes encoding three immune-regulatory chemokine receptors and another three molecules. The risk haplotype seemed to be inherited from Neanderthals, suggesting genetic adaptation against pathogens in modern human evolution. Therefore, SARS-CoV-2 uses highly conserved molecules as its virion interaction, whereas its immune response appears to be genetically biased in individuals to some extent. We herein review the molecular process of SARS-CoV-2 infection as well as our further survey of genetic variants of its related immune effectors. We also discuss aspects of modern human evolution.

KEYWORDS
COVID-19, genetic variant, human evolution

1 | INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by SARS-CoV-2, which is the third coronavirus to cause severe disease in human populations. As with other coronaviruses, SARS-CoV-2 infection is initialized by the spike (S) protein of the virus attaching to the host cell surface molecule angiotensin-converting enzyme 2 (ACE2). It then achieves cellular internalization by activating enzymatic machineries in the host cells. The immune response after infection varies among individuals, and some cases have shown acute respiratory syndrome with cytokine storm. Therefore, attention has been focused on the investigation of genetic risk factors.

Our research group recently surveyed genetic variants among SARS-CoV-2-interacting domains, including ACE2, across populations

(Lee et al., 2020). This survey concluded no obvious genetic biases in SARS-CoV-2-interacting domains across global populations.

Recent genome-wide association studies have identified genetic risk factors for severe COVID-19 with acute respiratory syndrome in a segment of chromosome 3, and the risk variants in this segment were shown to be inherited from Neanderthals (Severe Covid et al., 2020; Zeberg & Paabo, 2020). These findings have suggested the notion of genetic adaptation against viral pathogens in modern human evolution, concentrated in genes that encode molecules involved in the immune response process (Enard & Petrov, 2020).

In this short review, we will describe the molecular process of SARS-CoV-2 infection with current vaccine development, and recent genetic findings, including the results of a further survey of genetic variants of related immune response factors.

2 | PHYSIOLOGY AND PATHOLOGY OF HUMAN ACE2 IN SARS-CoV-2 INFECTION

Angiotensin-converting enzyme (ACE or ACE1) is a component of the renin–angiotensin system, which converts the hormone angiotensin 1 (Ang1) to the active vasoconstrictor angiotensin 2 (Ang2). ACE also has kininase activity, degrading bradykinin. Therefore, ACE indirectly increases blood pressure by regulating the body fluid and vasculature. ACE inhibitors are currently used as major pharmaceutical drugs for the treatment of cardiovascular diseases. In contrast, while ACE2 is closely related to ACE1 in structure, it downregulates blood pressure by catalyzing Ang2, thus functionally counteracting ACE1 (Hirano & Murakami, 2020).

ACE2 is also the entry point for SARS-CoV and SARS-CoV-2 into cells. SARS-CoV-2 can efficiently use ACE2 as a receptor for cellular entry, with an estimated 10- to 20-fold higher affinity to ACE2 than SARS-CoV (Wrapp et al., 2020; Wu et al., 2020; Zhao et al., 2020), which underscores the ancestral ACE2-using viral lineage (Wells et al., 2020). Upon infection, ACE2 is occupied and endocytosed with virus particles, resulting in a reduction in the ACE2 expression on cells. This functional downregulation of ACE2 induces an increased level of Ang2. While the physiological function of Ang2 causes vasoconstriction and an increase in blood pressure, it also acts as a proinflammatory cytokine through binding to angiotensin 2 receptor type 1 (ATR1) (Eguchi et al., 2018). Therefore, this increased level of Ang2 due to the infection activates NF-κB, disintegrin, and metalloprotease domain 17 (ADAM17) and eventually triggers viscous amplification of interleukin-6 (IL-6) signaling, known as cytokine release syndrome (CRS) (Hirano & Murakami, 2020).

3 | BASIC SCENARIOS OF VIRUS–HOST INTERACTION OF SARS-CoV-2, AND STRATEGIES FOR VACCINE DEVELOPMENT

SARS-CoV-2 is a single-stranded RNA (ssRNA) virus whose genomic RNA is encased by its viral envelope with radially projecting S proteins (Figure 1). The virus–host interaction of SARS-CoV-2 is initiated by the binding of the S proteins to ACE2 receptors on host cells (Zhou et al., 2020). Cell entry by the coronavirus requires S protein priming by host cell proteases, which cleave the S protein at the boundary of the S1 and S2 subunits (Du et al., 2009). The S protein priming of SARS-CoV-2 is processed by type 2 transmembrane serine protease (TMPRSS2) and endosomal cysteine proteases cathepsin B and L (CatB/L) (Hoffmann et al., 2020). The fusion process of the viral and cellular membranes enables viral genomic RNA to be released into the host cell. Toll-like receptors, such as TLR3, TLR7, and TLR8, recognize the viral RNA and trigger innate immune responses through the production of type I interferons and proinflammatory cytokines (Iwasaki & Pillai, 2014; Iwasaki & Yang, 2020).

Currently, more than 90 vaccines are being developed against SARS-CoV-2 across the world with different vaccine designs (Callaway, 2020; Haynes, 2021). Some vaccines, like many existing vaccines, are made using a weakened or inactivated form of the virus itself, but they require extensive safety testing. Nucleic acid vaccines have been also under development, and some have been licensed and are currently clinically available, in which DNA- or RNA-encoding S protein is introduced to human cells by electroporation, lipofection, or viral vector adenovirus. As these approaches are believed to prompt an immune response, they are being taken against SARS-CoV-2, although no licensed vaccines previously used this technology. Direct introductions of a receptor-binding domain (RBD) peptide of SARS-CoV-2 to ACE2 or virus-like particles to the body are also being tried. A collection of resources on COVID-19 vaccines (https://www.nejm.org/covid-vaccine) will be useful to update the current status.

4 | RECENT GENETIC FINDINGS IN POPULATIONS

4.1 | Evolutionary constraint of SARS-CoV-2-interacting proteins revealed by a survey of genetic variants across human populations

The expression of human ACE2 might be a critical factor associated with the susceptibility, symptoms, and outcome of virus infection. A positive correlation between the ACE2 expression and infection with SARS-CoV has been reported in vitro (Hofmann et al., 2004). A recent analysis of single-cell RNA sequencing (RNA-seq) demonstrated that ACE2 expression is much higher in lung cells from Asian donors than in those from Caucasian or African American donors (Zhao et al., 2020). In contrast, microarray has not shown higher ACE2 expression in lung tissues from Asian donors . One of the reasons for this discrepancy might be that the lung is a complex organ with multiple types of cells, and ACE2 expressed type II alveolar (AT2) and type I alveolar (AT1) epithelial cells were employed in the analysis. Thus, these results indicate that there were no significant differences between Asians and Caucasians or males and females (Cai et al., 2020).

We recently surveyed the genetic variants of ACE2 to investigate the allele spectrum of the key genes implicated in SARS-CoV-2 entry

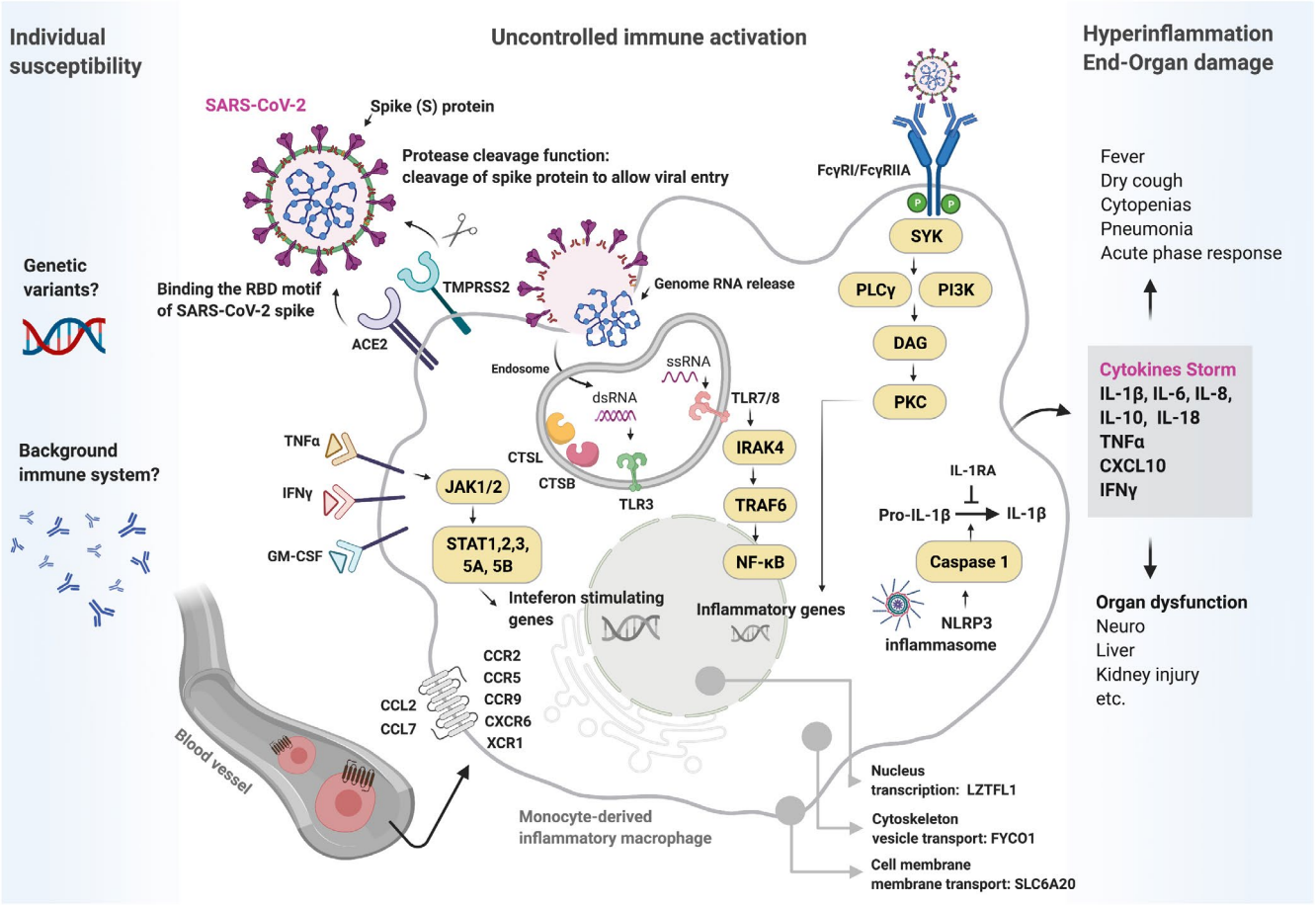

FIGURE 1 Proposed host immune responses to SARS-CoV-2. The spike (S) protein of virus binds to the ACE2 receptor in a host cell. Viral RNA is recognized by innate immune receptors, such as TLRs and the inflammasome sensor NLRP3. This process is involved in the activation of transcription factor NF-κB and subsequent signals for proinflammatory cytokines (for example, IL-6, CXCL10, etc.). Cytokines and chemokines released from infected cells regulate the adaptive immune response by recruiting immune cells, such as macrophages, T cells, and B cells, to eliminate the virus. A dysregulated immune response can lead to hyperinflammation, causing severe clinical symptoms of COVID-19

and sensing of viral RNAs across populations (Lee et al., 2020). We compiled the *ACE2* residues critical for interacting with the SARS-CoV-2 RBD (Hussain et al., 2020; Lan et al., 2020; Li et al., 2005; Shang et al., 2020; Wrapp et al., 2020; Yan et al., 2020). For the genes critical for proteolytic activity and sensing viral RNA (i.e., *TMPRSS2*, *CatB/L*, and *TLR3/7/8*), we collected the residues with negative functional impact according to previous studies (Afar et al., 2001; Bell et al., 2006; de Bouteiller et al., 2005; Sarkar et al., 2007; Tanji et al., 2015; Zhang et al., 2016). Then, for each gene, we surveyed multiple population-scale genetic variant databases, e.g., the Genome Aggregation Database (gnomAD) (Karczewski et al., 2020), the Korean Reference Genome Database (KRGD) (Jung et al., 2020), and TogoVar (a Japanese genetic variation database, https://togovar.biosciencedbc.jp/), to find all reported genetic variants that result in amino acid changes for those functionally important residues. Our analysis showed that variants are extremely rare on genomic loci influencing the binding affinity between SARS-CoV-2 and ACE2, thus implying little difference in affinity between populations. We also found that genetic variants in *TMPRSS2*, *CatB/L*, and *TLR3/7/8* were very rare, that is, allele frequency (AF) <0.01%, across all populations. Consistently, a recent study identified a few functional extremely rare variants in *TMPRSS2*, all of them with AFs <0.001 (Gupta et al., 2020; Russo et al., 2020). While our survey was conducted to investigate as diverse a population as possible, the variant databases still consisted of predominantly European populations, requiring caution when investigating rare variants among underrepresented populations. Also, based on the published evidence of functional consequences, we focused on variants at critical protein residues. The variants on regulatory elements of SARS-CoV-2-interacting genes may be worthy of further investigation. As such, the epigenetic regulation of key genes of the SARS-CoV-2-interacting domains or nongenetic factors, such as preexisting medical conditions and environmental risks, is suspected to be a major determinant of individual susceptibility.

The critical loci of the SARS-CoV-2-interacting domains for host–virus interaction and viral genomic RNA sensing are highly conserved in all populations with very rare variation. A very recent study determined that the crystal structure of ACE2 is highly conserved

when interacting with the RBD of SARS-CoV-2 (Lan et al., 2020). Our analysis also showed markedly fewer loss-of-function variants in *ACE2* and *TLR7* than expected in large variant databases, such as gnomAD (https://gnomad.broadinstitute.org/) (Lee et al., 2020), suggesting the presence of strong selection pressure on these genes in human evolution. Given the above, SARS-CoV-2 targets a highly conserved molecular domain in human populations as its interacting domain (Lee et al., 2020; Zhang et al., 2020).

4.2 | Genetic variability of immune response factors related to the severity of COVID-19

Several studies have now established that hyperinflammatory responses, such as the cytokine release syndrome (CRS) induced by SARS-CoV-2, are a major cause of disease severity and death in infected patients (Gubernatorova et al., 2020; Hirano & Murakami, 2020). We assessed the degrees of genetic constraint of SARS-CoV-2 susceptibility-related and cytokine regulation-related genes in immune cells based on the scores calculated from gnomAD (v2.1.1; https://gnomad.broadinstitute.org) (Figure 1 and Table 1). Compared with highly constrained genes, such as *ACE2* and *TLR7*, genes encoding key immune regulatory molecules, such as cytokines and chemokines, demonstrated a broadly low level of genetic constraint (Table 1). The probability of being loss-of-function intolerant (pLI) score represents the tolerance of a single gene to protein-truncating variants (e.g., frameshift, splice donor and acceptor, and stop-gain) in population-scale genome sequencing databases. Genes with a pLI score ≥0.9 can be highly constrained, such that protein-truncating variants are not frequently observed in the population, as is seen in the pLI scores of *ACE2* and *TLR7* (Table 1). Surprisingly, the genes involved in cytokine storm and chemokines are much less constrained. (highlighted in green in Table 1). Therefore, diverse locus and allelic heterogeneities exist in the genes involved in the immune response to SARS-CoV-2 infection. Furthermore, the clustered genes of *LZTFL1*, *SLC6A20*, *FYCO1*, *CCR9*, *CXCR6*, and *XCR1* in 3p21.31, which are significantly associated with severe COVID-19 cases, also showed extremely low pLI scores (Severe Covid et al., 2020) (highlighted in green with superscript b in Table 1). In contrast to the genes encoding SARS-CoV-2-interacting proteins, such as ACE2 and TLR7, the genes encoding immune effectors as listed here are genetically variable among individuals and thus interpretated to be evolutionarily less constrained.

Further genome-wide association studies combined with detailed phenotypic manifestation and clinical courses would uncover the genetic underpinnings of host responses to SARS-CoV-2 infection. As a next step, with surveys of additional population-scale genomic databases for diverse populations, it will be possible to identify the genetic predisposition to cytokine storm that causes an acute progression of illness. Accumulated survey of juvenile severe cases would especially provide further useful genetic information. Besides, systematic mutagenesis analysis of the structural interaction between SARS-CoV-2 receptor-binding motif (RBM) and ACE2 will help to understand host–virus interactions across populations (Lan et al., 2020).

4.3 | Genetic adaptation against viral pathogens

As briefly mentioned above, the gene cluster on chromosome 3 (3p21.31) has been identified as the major genetic risk locus for severe symptoms after SARS-CoV2 infection, being a genomic segment of 50 kb in size and comprising genes such as *LZTFL1*, *SLC6A20*, *FYCO1*, *CCR9*, *CXCR6*, and *XCR1* (Severe Covid et al., 2020; Zeberg & Paabo, 2020). Surprisingly, Zeberg and Pääbo reported that genetic variants in this segment associated with severe COVID-19 were inherited from Neanderthals, carried by around 50% and 16% of people in South Asia and Europe, respectively (Zeberg & Paabo, 2020). Therefore, their relatively high rate of interindividual genetic variants as scored here (Table 1) can be attributed to genetic sequences inherited from Neanderthals, at least to some extent.

Another recent report by Enard and Petrov proposed that Neanderthals and modern humans interbred and exchanged their own pathogenic viral environment, which drove adaptive DNA introgression against viruses (Enard & Petrov, 2020). Regulatory variants of these genes that might increase viral resistance were observed, especially in Europeans. Zeberg and Pääbo also noted that the Neanderthal-derived risk variants were highly concentrated in South Asia but almost absent in East Asia. A possible argument they made for this is that the Neanderthal haplotype had been positively selected in this regional area, possibly carrying protective effects against other pathogens. The other argument is that the risk variants have decreased in frequency in East Asia because of negative selection of coronavirus-related or other pathogens. Therefore, such adaptive introgression from Neanderthals to modern human has carried risk variants for SARS-CoV2 but also resistant variants against other RNA viruses, such as hepatitis C virus (HCV) and dengue viruses. Furthermore, genetic variants related to viral-interacting proteins, immune effector processes, and viral genome replication have reportedly been under directional (either positive or negative) selection pressure during human evolution before and after the splitting of the human population, thus enabling the detection of ancient epidemics in human history.

Based on our previous genetic survey of SARS-CoV2-interacting domains, *ACE2* and *TLR7* appeared to be under strong selection pressure, as evidenced by their lower number of loss-of-function variants than expected by large variant database analyses (Karczewski et al., 2020; Lee et al., 2020). We observed only 3 variants out of 31 expected ones for ACE2 and 2 variants out of 20.6 expected ones for TLR7 (Table 1), and we detected nonsynonymous variants in these genes at very low frequencies. These findings suggest that functional alteration of these genes would have been unlikely. Therefore, SARS-CoV-2 uses a highly conserved receptor in populations around the world, which may be a causative factor for its reaching pandemic status.

TABLE 1 Measures of intolerance to loss-of-function variants for genes related to SARS-CoV-2 susceptibility, hyperinflammation, and severity of COVID-19

Signal	Gene name	Isoforms	Official symbol	Expected LoF SNVs[a] (E)	Observed LoF SNVs[a] (O)	O/E[a] (90% CI)	pLI[a]
Receptor	ACE2		ACE2	31	3	0.1 [0.04–0.25]	1
Sensing viral RNA	TLR7		TLR7	20.6	2	0.1 [0.04–0.3]	0.98
Intermediate	SYK		SYK	34.1	2	0.06 [0.02–0.18]	1
	PLC-γ		PLCG1	82.1	24	0.29 [0.21–0.41]	0
	IRAK4		IRAK4	23.1	16	0.69 [0.47–1.05]	0
	TRAF6		TRAF6	21.9	1	0.05 [0.01–0.22]	1
	NF-kB	NFKB1	NFKB1	45.6	2	0.04 [0.02–0.14]	1
		NFKB2	NFKB2	47.7	3	0.06 [0.03–0.16]	1
	JAK	JAK1	JAK1	61.4	7	0.11 [0.06–0.21]	1
		JAK2	JAK2	60.6	13	0.21 [0.14–0.34]	0.65
	STAT	STAT1	STAT1	48.9	4	0.08 [0.04–0.19]	1
		STAT2	STAT2	57	11	0.19 [0.12–0.32]	0.93
		STAT3	STAT3	50.1	1	0.02 [0.01–0.1]	1
		STAT5A	STAT5A	45.4	4	0.09 [0.04–0.2]	1
		STAT5B	STAT5B	47.8	4	0.1 [0.05–0.22]	1
	PI3K	PIK3CA	PIK3CA	66.4	3	0.05 [0.02–0.12]	1
		PIK3CB	PIK3CB	62.4	9	0.14 [0.09–0.25]	1
		PIK3CD	PIK3CD	53.6	5	0.09 [0.05–0.2]	1
		PIK3CG	PIK3CG	44.9	18	0.4 [0.28–0.59]	0
	M-CSF		CSF1	21	1	0.05 [0.02–0.23]	1
	GM-CSF		CSF2	5.9	0	0 [0–0.51]	0.83
	LZTFL1[b]		LZTFL1	20.6	6	0.29 [0.16–0.57]	0.06
	SLC6A20[b]		SLC6A20	29.6	27	0.91 [0.67–1.26]	0
	FYCO1[b]		FYCO1	71.9	52	0.72 [0.58–0.91]	0
Chemokine	CCL2		CCL2	2.953	0	0.0 [0.0 – 1.0]	0.608
	CCL7		CCL7	3.141	4	1.27 [0.58–1.91]	0.001
	CCR2		CCR2	9.4	4	0.43 [0.21–0.98]	0.02
	CCR5		CCR5	7.9	12	1.51 [0.93–1.93]	0
	CCR9[b]		CCR9	9.3	5	0.54 [0.28–1.13]	0
	CXCR6[b]		CXCR6	8.3	7	0.84 [0.48–1.55]	0
	XCR1[b]		XCR1	8.4	4	0.47 [0.23–1.08]	0.02
	CXCL10		CXCL10	4.7	1	0.21 [0.07–1]	0.37
Cytokine storm	IL-1β		IL-1β	14.1	4	0.28 [0.14–0.65]	0.13
	IL-6		IL-6	8.4	2	0.24 [0.1–0.75]	0.32
	IL-8		CXCL8	4.5	5	1.12 [0.56–1.86]	0
	IL-10		IL-10	10.1	5	0.49 [0.26–1.04]	0.01
	IL-18		IL-18	6	3	0.5 [0.23–1.28]	0.03
	IL-1RA		IL-1RA	10.1	4	0.4 [0.19–0.91]	0.03
	TNF		TNF	9.736	1	0.103 [0.04–0.49]	0.803
	IFN-γ		IFNG	5.768	1	0.17 [0.06–0.82]	0.472

Note: The column "Expected LoF SNVs" shows expected variant counts using mutation model based on sequence context, coverage, and methylation, while "Observed LoF SNVs" shows counts of QC-passed variants that occurred in gnomAD dataset (release 2.1.1). For genes intolerant to loss-of-function variants, the ratio between these two (the column "O/E") would be close to 0. The column "pLI" (the probability of being loss-of-function intolerant) denotes the probability of a gene being haploinsufficient (heterozygous variants are not tolerated) based on the deviation between observed and expected variant counts. The pLI values close to 1 indicate genes intolerant to loss-of-function variants. Highlighted in green with pLI score ≤ 0.1.

[a]Values taken from gnomAD v2.1.1 (https://gnomad.broadinstitute.org, last access: 2020-11-01).

[b]Genes located in the susceptibility loci (3p21.31) associated with severe COVID-19 with respiratory failure

In this report, we have listed the ratios of actually observed loss-of-function variants to theoretically expected ones for genes associated with the susceptibility, hyperinflammation, and severity of COVID-19. We observed relatively high rates of genetic variants of intracellular molecules and chemokines, including *LTTFL1*, *SLC6A20*, *FYCO1*, *CCR9*, *CXCR6*, and *XCR1*, which may reflect a distinct history of genetic adaptation of individuals. These genetic survey analyses of the SARS-CoV-2-interacting domains and the genetic factors do not fully explain current geographic and regional distribution of severe cases. Therefore, it should still be emphasized that preexisting medical conditions and environmental risks could be major determinants of individual susceptibility.

Risk factor genes of *CCR9*, *CXCR6*, *XCR1*, *LZTFL1*, *SLC6A20*, and *FYCO1*, whose functions in viral infection and the immune response remain to be well elucidated, definitively require further functional investigation. CCR9 and its specific ligand CCL25 are categorized as a mucosal chemokine system, which is responsible for recruiting a subset of T cells to the lamina propria of the small intestine (Kunkel et al., 2000; Norment et al., 2000; Vicari et al., 1997; Zlotnik & Yoshie, 2012). The axis of CCL-25-CCR9 has been investigated as a therapeutic target of Crohn's disease, tumors, and autoimmunity (Biswas et al., 2019; Korbecki et al., 2020; Pathak & Lal, 2020; Xu et al., 2020). CXCR6 regulates localization of tissue-resident memory CD8 T cells to the airways throughout the sustained immune response to airway pathogens such as influenza viruses (Wein et al., 2019). CXCR6 and its ligand CXCL16 axis, like most of other chemokine axes, also plays roles in lymphocyte migration, tumor progression, and metastasis (Audsley et al., 2020; Groblewska et al., 2020; Jovanovic et al., 2015; La Porta, 2012). XCR1-positive dendritic cells (DCs) have the unique ability to cross-present cell-associated antigens and MHC class I to CD8(+) T cells (Audsley et al., 2020; Gutierrez-Martinez et al., 2015; Shi et al., 2020). Therefore, these chemokine receptors are likely to be involved in immune response after SARS-CoV2 infection, and their functional changes possibly contribute to distinct output in individuals. Notably, deficiency of another chemokine receptor, CCR1, increases susceptibility to coronavirus infection (Hickey et al., 2007). *LZTFL1* encodes leucine zipper transcription factor-like 1, which associates with the immune synapse of the interface between an antigen-presenting cell or target cell and a lymphocyte such as a T/B cell or natural killer cell (Jiang et al., 2016). *SLC6A20* encodes sodium–imino acid (proline) transporter 1 (SIT1), which reportedly interacts with ACE2 (Kuba et al., 2010; Vuille-dit-Bille et al., 2015). *FYCO1* encodes FYVE and coiled-coil domain containing 1 (FYCO1), an autophagy adaptor protein, is suggested to be a key mediator linking ER-derived double membrane vesicles, the primary replication site for coronaviruses, with the microtubule network in host cells (Parkinson et al., 2020; Reggiori et al., 2011). Further functional studies on how these genetic variants affect the epigenetic, transcriptional, and functional control of each gene and eventually contribute to the immune response of individuals should be conducted, as these findings will be necessary for the development of pharmacological agents to combat COVID-19.

5 | CONCLUSION AND FUTURE PERSPECTIVE

Our survey of genetic variations showed that SARS-CoV-2-interacting domains are highly conserved across human populations. The genetic risk segment includes genes of several chemokine receptors, such as *CCR9*, *CXCR6*, and *XCR1*, which play critical roles in immune responses and pathogenesis (Zlotnik & Yoshie, 2012). It is of note that the repeated genetic duplication of chemokines and their receptors in vertebrate evolution has been proposed to be driven by the coevolution between host and pathogens (Nomiyama et al., 2013; Zlotnik & Yoshie, 2012). The number of chemokine and chemokine receptor genes has been increased during vertebrate evolution (Bajoghli, 2013). These genes are clustered onto several chromosomal locations, which highlights the relevance of genetic duplication of *Hox* genes in vertebrate evolution (Duboule, 2007; Nomiyama et al., 2013). In fact, some genetic loci including chemokine receptor genes appear to accompany *Hox* cluster duplication (DeVries et al., 2006). Orthologous relationship analyses across species have demonstrated that the chemokine system is highly evolved, especially in mammals possibly because of increased exposure to pathogens in their intimate life environment, and also have experienced difficulty in determination of orthologous genes across mammalian species, suggesting genetic diversity due to selective pressures imposed by distinct pathogens (Nomiyama et al., 2013). In modern human genetics, a loss-of-function variant of *CCR5 delta 32* is well known to be concentrated in Northern Europe and has long been investigated in terms of resistance to human immunodeficiency virus (HIV) infection and age-related disorders such as osteoporosis (Lee et al., 2017; Xie et al., 2019). Together, genetic variations in immune response genes discussed here support the idea that the selective pressure imposed by pathogens during human evolution is a potential cause of genetic diversity, and shapes the repertoire of host defense systems across populations and individuals. The variants of risk factor genes such as *CCR9*, *CXCR6*, *XCR1*, *LZTFL1*, *SLC6A20*, and *FYCO1* definitively require further functional investigation, for the development of better medical treatment to combat COVID-19.

ACKNOWLEDGMENTS

The authors express sincere thanks to Prof. Akira Hasebe (Dept. Oral Molecular Microbiology, Faculty and Graduate School of Dental Medicine, Hokkaido University) for his helpful comments on the current vaccine development. This work was partly supported by Grant-in-Aids for Scientific Research from the Japan Society for the Promotion of Science (JSPS KAKENHI) grant numbers 18H02983 and 18K19649 to T. I. SWK was supported in part by the grants from the National Institution of Health (R01MH107205, R24OD024622 and U01TR002623). JWL was supported by Grant-in-Aids for Scientific Research from the Japan Society for the Promotion of Science (19K10044) and the grant from Takeda Science Foundation (PK43200005).

CONFLICTS OF INTEREST

On behalf of all authors, the corresponding author states that there are no conflicts of interest.

ORCID

In-Hee Lee https://orcid.org/0000-0002-8857-1355
Tadahiro Iimura https://orcid.org/0000-0002-1431-0689

REFERENCES

Afar, D. E., Vivanco, I., Hubert, R. S., Kuo, J., Chen, E., Saffran, D. C., Raitano, A. B., Jakobovits, A. (2001). Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. *Cancer Res*, *61*, 1686–1692.

Audsley, K. M., McDonnell, A. M., & Waithman, J. (2020). Cross-presenting XCR1(+) dendritic cells as targets for cancer immunotherapy. *Cells*, *9*, 565. https://doi.org/10.3390/cells9030565

Bajoghli, B. (2013). Evolution and function of chemokine receptors in the immune system of lower vertebrates. *Eur J Immunol*, *43*, 1686–1692. https://doi.org/10.1002/eji.201343557

Bell, J. K., Askins, J., Hall, P. R., Davies, D. R., & Segal, D. M. (2006). The dsRNA binding site of human Toll-like receptor 3. *Proc Natl Acad Sci U S A*, *103*, 8792–8797. https://doi.org/10.1073/pnas.0603245103

Biswas, S., Bryant, R. V., & Travis, S. (2019). Interfering with leukocyte trafficking in Crohn's disease. *Best Pract Res Clin Gastroenterol*, *38–39*, 101617. https://doi.org/10.1016/j.bpg.2019.05.004

Cai, G., Bosse, Y., Xiao, F., Kheradmand, F., & Amos, C. I. (2020). Tobacco smoking increases the lung gene expression of ACE2, the receptor of SARS-CoV-2. *Am J Resp Crit Care Med*, *201*, 1557–1559. https://doi.org/10.1164/rccm.202003-0693LE

Callaway, E. (2020). The race for coronavirus vaccines: A graphical guide. *Nature*, *580*, 576–577. https://doi.org/10.1038/d41586-020-01221-y

de Bouteiller, O., Merck, E., Hasan, U. A., Hubac, S., Benguigui, B., Trinchieri, G., Bates, E. E. M., & Caux, C. (2005). Recognition of double-stranded RNA by human toll-like receptor 3 and downstream receptor signaling requires multimerization and an acidic pH. *J Biol Chem*, *280*, 38133–38145. https://doi.org/10.1074/jbc.M507163200

DeVries, M. E., Kelvin, A. A., Xu, L., Ran, L., Robinson, J., & Kelvin, D. J. (2006). Defining the origins and evolution of the chemokine/chemokine receptor system. *J Immunol*, *176*, 401–415. https://doi.org/10.4049/jimmunol.176.1.401

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat Rev Microbiol*, *7*, 226–236. https://doi.org/10.1038/nrmicro2090

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development*, *134*, 2549–2560. https://doi.org/10.1242/dev.001065

Eguchi, S., Kawai, T., Scalia, R., & Rizzo, V. (2018). Understanding angiotensin II type 1 receptor signaling in vascular pathophysiology. *Hypertension*, *71*, 804–810. https://doi.org/10.1161/HYPERTENSIONAHA.118.10266

Enard, D., & Petrov, D. A. (2020). Ancient RNA virus epidemics through the lens of recent adaptation in human genomes. *Philos Trans R Soc B Biol Sci*, *375*, 20190575.–https://doi.org/10.1098/rstb.2019.0575

Groblewska, M., Litman-Zawadzka, A., & Mroczko, B. (2020). The role of selected chemokines and their receptors in the development of gliomas. *Int J Mol Sci*, *21*, 3704. https://doi.org/10.3390/ijms21103704

Gubernatorova, E. O., Gorshkova, E. A., Polinova, A. I., & Drutskaya, M. S. (2020). IL-6: Relevance for immunopathology of SARS-CoV-2. *Cytokine Growth Factor Rev*, *53*, 13–24. https://doi.org/10.1016/j.cytogfr.2020.05.009

Gupta, R., Charron, J., Stenger, C. L., Painter, J., Steward, H., Cook, T. W., Faber, W., Frisch, A., Lind, E., Bauss, J., Li, X., Sirpilla, O., Soehnlen, X., Underwood, A., Hinds, D., Morris, M., Lamb, N., Carcillo, J. A., Bupp, C., … Prokop, J. W. (2020). SARS-CoV2 (COVID-19) structural/evolution dynamicome: Insights into functional evolution and human genomics. *bioRxiv*.

Gutierrez-Martinez, E., Planes, R., Anselmi, G., Reynolds, M., Menezes, S., Adiko, A. C., Saveanu, L., & Guermonprez, P. (2015). Cross-presentation of cell-associated antigens by MHC class I in dendritic cell subsets. *Front Immunol*, *6*, 363. https://doi.org/10.3389/fimmu.2015.00363

Haynes, B. F. (2021). A new vaccine to battle Covid-19. *N Engl J Med*, *384*, 470–471. https://doi.org/10.1056/NEJMe2035557

Hickey, M. J., Held, K. S., Baum, E., Gao, J. L., Murphy, P. M., & Lane, T. E. (2007). CCR1 deficiency increases susceptibility to fatal coronavirus infection of the central nervous system. *Viral Immunol*, *20*, 599–608. https://doi.org/10.1089/vim.2007.0056

Hirano, T., & Murakami, M. (2020). COVID-19: A new virus, but a familiar receptor and cytokine release syndrome. *Immunity*, *52*, 731–733. https://doi.org/10.1016/j.immuni.2020.04.003

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, *181*, 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

Hofmann, H., Geier, M., Marzi, A., Krumbiegel, M., Peipp, M., Fey, G. H., Gramberg, T., & Pöhlmann, S. (2004). Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. *Biochem Biophys Res Commun*, *319*, 1216–1221. https://doi.org/10.1016/j.bbrc.2004.05.114

Hussain, M., Jabeen, N., Raza, F., Shabbir, S., Baig, A. A., Amanullah, A., & Aziz, B. (2020). Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J Med Virol*, *92*, 1580–1586. https://doi.org/10.1002/jmv.25832

Iwasaki, A., & Pillai, P. S. (2014). Innate immunity to influenza virus infection. *Nat Rev Immunol*, *14*, 315–328. https://doi.org/10.1038/nri3665

Iwasaki, A., & Yang, Y. (2020). The potential danger of suboptimal antibody responses in COVID-19. *Nat Rev Immunol*, *20*, 339–341. https://doi.org/10.1038/s41577-020-0321-6

Jiang, H., Promchan, K., Lin, B. R., Lockett, S., Chen, D., Marshall, H., & Natarajan, V. (2016). LZTFL1 upregulated by all-trans retinoic acid during CD4+ T cell activation enhances IL-5 production. *J Immunol*, *196*, 1081–1090.

Jovanovic, I., Zivkovic, M., Djuric, T., Popovic, M., Alavantic, D., & Stankovic, A. (2015). CXCL16 in vascular pathology research: From macro effects to microRNAs. *J Atheroscler Thromb*, *22*, 1012–1024. https://doi.org/10.5551/jat.29942

Jung, K. S., Hong, K. W., Jo, H. Y., Choi, J., Ban, H. J., Cho, S. B., & Chung, M. (2020). KRGDB: The large-scale variant database of 1722 Koreans based on whole genome sequencing. *Database (Oxford)*.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7

Korbecki, J., Grochans, S., Gutowska, I., Barczak, K., & Baranowska-Bosiacka, I. (2020). CC chemokines in a tumor: A review of pro-cancer and anti-cancer properties of receptors CCR5, CCR6, CCR7, CCR8, CCR9, and CCR10 ligands. *Int J Mol Sci*, *21*, 7619.

Kuba, K., Imai, Y., Ohto-Nakanishi, T., & Penninger, J. M. (2010). Trilogy of ACE2: A peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. *Pharmacol Therape*, *128*, 119–128. https://doi.org/10.1016/j.pharmthera.2010.06.003

Kunkel, E. J., Campbell, J. J., Haraldsen, G., Pan, J., Boisvert, J., Roberts, A. I., Ebert, E. C., Vierra, M. A., Goodman, S. B., Genovese, M. C., Wardlaw, A. J., Greenberg, H. B., Parker, C. M., Butcher, E. C., Andrew, D. P., & Agace, W. W. (2000). Lymphocyte CC chemokine receptor 9 and epithelial thymus-expressed chemokine (TECK) expression distinguish the small intestinal immune compartment: Epithelial expression of tissue-specific chemokines as an organizing principle in regional immunity. *J Exp Med*, *192*, 761–768. https://doi.org/10.1084/jem.192.5.761

La Porta, C. A. (2012). CXCR6: The role of environment in tumor progression. Challenges for therapy. *Stem Cell Rev Rep*, *8*, 1282–1285. https://doi.org/10.1007/s12015-012-9383-6

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q. I., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*, 215–220. https://doi.org/10.1038/s41586-020-2180-5

Lee, I. H., Lee, J. W., & Kong, S. W. (2020). A survey of genetic variants in SARS-CoV-2 interacting domains of ACE2, TMPRSS2 and TLR3/7/8 across populations. *Infect Genet Evol*, *85*, 104507. https://doi.org/10.1016/j.meegid.2020.104507

Lee, J.-W., Hoshino, A., Inoue, K., Saitou, T., Uehara, S., Kobayashi, Y., Ueha, S., Matsushima, K., Yamaguchi, A., Imai, Y., & Iimura, T. (2017). The HIV co-receptor CCR5 regulates osteoclast function. *Nat Commun*, *8*, 2226. https://doi.org/10.1038/s41467-017-02368-5

Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y. I., Choe, H., & Farzan, M. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J*, *24*, 1634–1643. https://doi.org/10.1038/sj.emboj.7600640

Nomiyama, H., Osada, N., & Yoshie, O. (2013). Systematic classification of vertebrate chemokines based on conserved synteny and evolutionary history. *Genes Cells*, *18*, 1–16. https://doi.org/10.1111/gtc.12013

Norment, A. M., Bogatzki, L. Y., Gantner, B. N., & Bevan, M. J. (2000). Murine CCR9, a chemokine receptor for thymus-expressed chemokine that is up-regulated following pre-TCR signaling. *J Immunol*, *164*, 639–648. https://doi.org/10.4049/jimmunol.164.2.639

Parkinson, N., Rodgers, N., Head Fourman, M., Wang, B. O., Zechner, M., Swets, M. C., Millar, J. E., Law, A., Russell, C. D., Baillie, J. K., & Clohisey, S. (2020). Dynamic data-driven meta-analysis for prioritisation of host genes implicated in COVID-19. *Sci Rep*, *10*, 22303. https://doi.org/10.1038/s41598-020-79033-3

Pathak, M., & Lal, G. (2020). The regulatory function of CCR9(+) dendritic cells in inflammation and autoimmunity. *Front Immunol*, *11*, 536326. https://doi.org/10.3389/fimmu.2020.536326

Reggiori, F., de Haan, C. A., & Molinari, M. (2011). Unconventional use of LC3 by coronaviruses through the alleged subversion of the ERAD tuning pathway. *Viruses*, *3*, 1610–1623. https://doi.org/10.3390/v3091610

Russo, R., Andolfo, I., Lasorsa, V. A., Iolascon, A., & Capasso, M. (2020). Genetic analysis of the coronavirus SARS-CoV-2 host protease TMPRSS2 in different populations. *Front Genet*, *11*, 872. https://doi.org/10.3389/fgene.2020.00872

Sarkar, S. N., Elco, C. P., Peters, K. L., Chattopadhyay, S., & Sen, G. C. (2007). Two tyrosine residues of Toll-like receptor 3 trigger different steps of NF-kappa B activation. *J Biol Chem*, *282*, 3423–3427.

Severe Covid-19 GWAS Group, Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M. M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F. … Karlsen, T. H. (2020). Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med*, *383*, 1522–1534.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*, 221–224. https://doi.org/10.1038/s41586-020-2179-y

Shi, J.-W., Yang, H.-L., Fan, D.-X., Yang, S.-L., Qiu, X.-M., Wang, Y., Lai, Z.-Z., Ha, S.-Y., Ruan, L.-Y., Shen, H.-H., Zhou, W.-J., & Li, M.-Q. (2020). The role of CXC chemokine ligand 16 in physiological and pathological pregnancies. *Am J Reprod Immunol*, *83*, e13223. https://doi.org/10.1111/aji.13223

Tanji, H., Ohto, U., Shibata, T., Taoka, M., Yamauchi, Y., Isobe, T., Miyake, K., & Shimizu, T. (2015). Toll-like receptor 8 senses degradation products of single-stranded RNA. *Nat Struct Mol Biol*, *22*, 109–115. https://doi.org/10.1038/nsmb.2943

Vicari, A. P., Figueroa, D. J., Hedrick, J. A., Foster, J. S., Singh, K. P., Menon, S., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Bacon, K. B., & Zlotnik, A. (1997). TECK: A novel CC chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development. *Immunity*, *7*, 291–301. https://doi.org/10.1016/S1074-7613(00)80531-2

Vuille-dit-Bille, R. N., Camargo, S. M., Emmenegger, L., Sasse, T., Kummer, E., Jando, J., Hamie, Q. M., Meier, C. F., Hunziker, S., Forras-Kaufmann, Z., Kuyumcu, S., Fox, M., Schwizer, W., Fried, M., Lindenmeyer, M., Götze, O., & Verrey, F. (2015). Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors. *Amino Acids*, *47*, 693–705. https://doi.org/10.1007/s00726-014-1889-6

Wein, A. N., McMaster, S. R., Takamura, S., Dunbar, P. R., Cartwright, E. K., Hayward, S. L., McManus, D. T., Shimaoka, T., Ueha, S., Tsukui, T., Masumoto, T., Kurachi, M., Matsushima, K., & Kohlmeier, J. E. (2019). CXCR6 regulates localization of tissue-resident memory CD8 T cells to the airways. *J Exp Med*, *216*, 2748–2762. https://doi.org/10.1084/jem.20181308

Wells, H. L., Letko, M., Lasso, G., Ssebide, B., Nziza, J., Byarugaba, D. K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B. A., Tingley, M. W., Diuk-Wasser, M., Goldstein, T., Johnson, C. K., Mazet, J., Chandran, K., Munster, V. J., Gilardi, K., & Anthony, S. J. (2020). The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. *bioRxiv*.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*, 1260–1263. https://doi.org/10.1126/science.abb2507

Wu, F., Zhao, S. U., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y. I., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y. I., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*, 265–269. https://doi.org/10.1038/s41586-020-2008-3

Xie, Y., Zhan, S., Ge, W., & Tang, P. (2019). The potential risks of C-C chemokine receptor 5-edited babies in bone development. *Bone Res*, *7*, 4. https://doi.org/10.1038/s41413-019-0044-0

Xu, B., Deng, C., Wu, X., Ji, T., Zhao, L., Han, Y., Yang, W., Qi, Y., Wang, Z., Yang, Z. & Yang, Y. (2020). CCR9 and CCL25: A review of their roles in tumor promotion. *J Cell Physiol*, *235*, 9121–9132.

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, *367*, 1444–1448. https://doi.org/10.1126/science.abb2762

Zeberg, H., & Paabo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*, *587*, 610–612. https://doi.org/10.1038/s41586-020-2818-3

Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol*, *30*, 1346–1351.e42.

Zhang, Z., Ohto, U., Shibata, T., Krayukhina, E., Taoka, M., Yamauchi, Y., Tanji, H., Isobe, T., Uchiyama, S., Miyake, K., & Shimizu, T. (2016). Structural analysis reveals that Toll-like Receptor 7 is a dual receptor for guanosine and single-stranded RNA. *Immunity*, *45*, 737–748. https://doi.org/10.1016/j.immuni.2016.09.011

Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., & Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *Am J Resp Crit Care Med*, *202*, 756–759. https://doi.org/10.1164/rccm.202001-0179LE

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X. I., … Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*, 270–273. https://doi.org/10.1038/s41586-020-2012-7

Zlotnik, A., & Yoshie, O. (2012). The chemokine superfamily revisited. *Immunity*, *36*, 705–716. https://doi.org/10.1016/j.immuni.2012.05.008

How to cite this article: Lee J-W, Lee I-H, Sato T, Kong SW, Iimura T. Genetic variation analyses indicate conserved SARS-CoV-2–host interaction and varied genetic adaptation in immune response factors in modern human evolution. *Develop Growth Differ*. 2021;63:219–227. https://doi.org/10.1111/dgd.12717