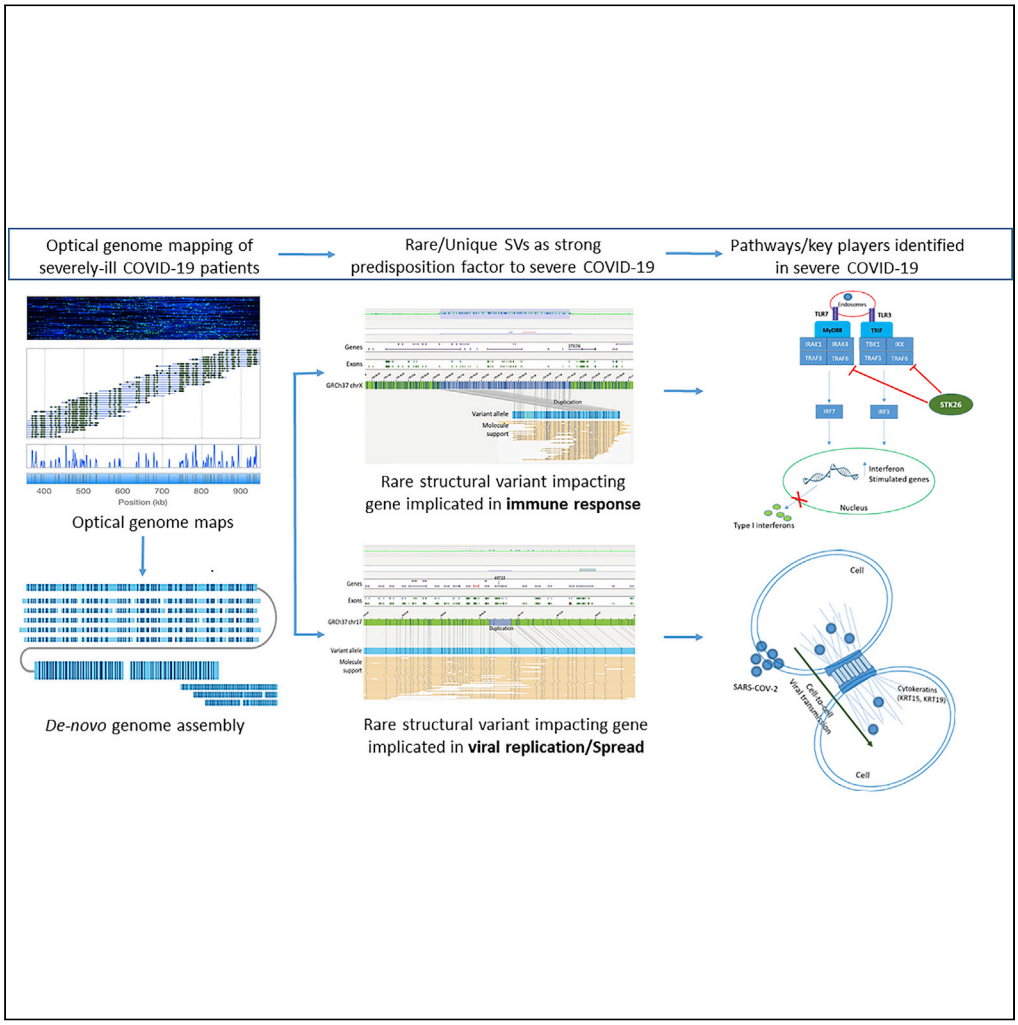
iScience

CellPress
OPEN ACCESS

Article

# Optical genome mapping identifies rare structural variations as predisposition factors associated with severe COVID-19




Nikhil Shri Sahajpal, Chi-Yu Jill Lai, Alex Hastie, ..., Alexander Hoischen, Alka Chaubey, Ravindra Kolhe

rkolhe@augusta.edu

## Highlights

Genetic predisposition studies remain limited to investigation of sequencing variants

We used optical genome mapping to investigate large SVs in severe COVID-19 patients

Identified seven rare SVs, with SVs in *STK26* and *DPP4* as the most promising candidates

Study highlights the potential role of SVs in the pathogenesis of COVID-19 severity

Sahajpal et al., iScience *25*, 103760
February 18, 2022 
https://doi.org/10.1016/j.isci.2022.103760

iScience

CellPress
OPEN ACCESS

Article

# Optical genome mapping identifies rare structural variations as predisposition factors associated with severe COVID-19

Nikhil Shri Sahajpal,[1] Chi-Yu Jill Lai,[2] Alex Hastie,[2] Ashis K. Mondal,[1] Siavash Raeisi Dehkordi,[3] Caspar I. van der Made,[4] Olivier Fedrigo,[5] Farooq Al-Ajli,[5] Sawan Jalnapurkar,[6] Marta Byrska-Bishop,[7] Rashmi Kanagal-Shamanna,[8] Brynn Levy,[9] Maximilian Schieck,[10] Thomas Illig,[10,11] Silviu-Alin Bacanu,[12] Janet S. Chou,[13] Adrienne G. Randolph,[14,15] Amyn M. Rojiani,[1] Michael C. Zody,[7] Catherine A. Brownstein,[16] Alan H. Beggs,[16] Vineet Bafna,[3] Erich D. Jarvis,[5,17,18] Alexander Hoischen,[4] Alka Chaubey,[1,2] Ravindra Kolhe,[1,20,*] and The COVID19hostgenomesv Consortium[19]

## SUMMARY

**Impressive global efforts have identified both rare and common gene variants associated with severe COVID-19 using sequencing technologies. However, these studies lack the sensitivity to accurately detect several classes of variants, especially large structural variants (SVs), which account for a substantial proportion of genetic diversity including clinically relevant variation. We performed optical genome mapping on 52 severely ill COVID-19 patients to identify rare/unique SVs as decisive predisposition factors associated with COVID-19. We identified 7 SVs involving genes implicated in two key host-viral interaction pathways: innate immunity and inflammatory response, and viral replication and spread in nine patients, of which SVs in *STK26* and *DPP4* genes are the most intriguing candidates. This study is the first to systematically assess the potential role of SVs in the pathogenesis of COVID-19 severity and highlights the need to evaluate SVs along with sequencing variants to comprehensively associate genomic information with interindividual variability in COVID-19 phenotypes.**

## INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) first reported in the city of Wuhan, China in December 2019 has led to an ongoing global pandemic of coronavirus disease, COVID-19. Since then, there have been almost 241 million COVID-19 cases reported worldwide and over four million COVID-19-related deaths (https://coronavirus.jhu.edu/map.html, last accessed October 20, 2021). The clinical manifestations of SARS-CoV-2 infected patients vary from asymptomatic or mild symptoms, including low-grade fever and flu-like symptoms, to COVID-19 pneumonia, and life-threatening acute respiratory distress syndrome (ARDS) (Berlin et al., 2020; Zhu et al., 2020; Zhou et al., 2020; Li et al., 2020a, 2020b; Williamson et al., 2020; Wu and McGoogan, 2020). Clinical studies have identified advanced age, male gender, hypertension, diabetes, diseases related to metabolic syndrome, and other chronic diseases, as risk factors associated with COVID-19-related severe illness (Zhou et al., 2020; Li et al., 2020a, 2020b; Docherty et al., 2020). However, variable clinical manifestations within these subsets and poor clinical outcomes in individuals without associated comorbidities or risk factors clearly implicate the role of host genetics in the predisposition of SARS-CoV-2 infected individuals to severe COVID-19 (Richardson et al., 2020; Abbasi, 2020). A recent meta-analysis of host genetic factors in coronaviruses infection identified 1,832 relevant research publications with 105 of high significance. Of the 105 articles, seventy-five investigated human host genetic factors, identifying multiple significant loci, including 16 related to susceptibility (seven of which identified protective alleles) and 16 related to outcomes (three of which identified protective alleles). In addition, 30 articles investigated inter-species differences in disease susceptibility and pathogenesis by studying both human and nonhuman host genetic factors (LoPresti et al., 2020).

To date, two distinct genomic approaches have been used to identify likely disease susceptibility and severity gene variants in COVID-19 patients. The first approach utilizes relatively unbiased

[1]Department of Pathology, Medical College of Georgia, Augusta University, 1120 15th Street, BF-207, Augusta, GA 30912, USA

[2]Bionano Genomics, Inc., San Diego, CA 92121, USA

[3]Department of Computer Science and Engineering, University of California at San Diego, San Diego, CA 92093, USA

[4]Department of Human Genetics, Radboud University Medical Center for Infectious Diseases (RCI), Department of Internal Medicine, Radboud Institute for Molecular Life Sciences, and Radboud Expertise Center for Immunodeficiency and Autoinflammation, Radboud University Medical Center, 6525 Nijmegen, the Netherlands

[5]Vertebrate Genome Lab, The Rockefeller University, New York, NY 10065, USA

[6]Department of Medicine, Medical College of Georgia, Augusta University, GA 30912, USA

[7]New York Genome Center, New York, NY 10013, USA

[8]Hematopathology, University of Texas M.D. Anderson Cancer Center, Houston 77030, USA

[9]Department of Pathology and Cell Biology, Columbia University Medical Center, New York 10032, USA

[10]Department of Human Genetics, Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Hannover 30625, Germany

*Continued*

genome-wide association studies (GWAS) to understand risk factors at the population level, which provides insights into the pathophysiological mechanisms and biology, but this approach cannot be used to investigate individual patients. The second approach involves short-read gene panel, whole exome, or genome sequencing to identify rare variants in known genes whose biological functions suggest plausible models by which they may function as severe risk factors, such as those that contribute to inborn errors of immunity (IEIs) which remain unremarkable until infection, or completely novel IEIs/primary immunodeficiencies (PIDs) that predispose to severe COVID-19. Using a GWAS approach, several groups have found blood group A of the ABO groups to be an independent risk factor for COVID-19 related disease susceptibility (Zhao et al., 2020a, 2020b). In additional studies, chr3p21.31, chr12q24.13, chr19p13.2, chr19p13.3, and chr21q22.1 loci were associated with severe COVID-19, and blood group A was confirmed to be the risk factor for disease severity (Ellinghaus et al., 2020; Pairo-Castineira et al., 2021). Reports utilizing a targeted variant approach demonstrated the importance of rare/unique events that cause immunodeficiencies and predispose them to severe COVID-19. For example, putative loss of function variants in *TLR7* in two different pairs of young and otherwise healthy brothers was associated with severe COVID-19 (Van der Made et al., 2020), this finding was recently replicated (Fallerini et al., 2021). Further, the rare variants in 13 genes in the TLR3 and IRF7 pathway have been implicated as risk factors in patients with severe COVID-19 (Zhang et al., 2020; Hadjadj et al., 2020), although a recent study reported inconsistent findings (Kosmicki et al., 2021).

In a continued effort to understand the varied host response, the COVID-19 Host Genetics Initiative has been established to encourage data generation, sharing, and meta-analysis of the GWAS statistics from around the world (COVID-19 Host Genetics Initiative, 2020). Although these studies have identified certain genetic loci associated with disease severity, they remain limited to short or single nucleotide variants, which explain only a small portion of the heritability of complex traits. Importantly, structural variants (SVs) that comprise a substantial proportion of genomic variation among individuals, and which can drive evolutionary processes are beyond the purview of these studies as a result of the need for new technologies for sensitive detection of SVs (Eichler, 2019; Levy-Sakin et al., 2019). SVs are defined as genetic variants involving larger regions of an individual genome (>50bp) (Eichler, 2019), and are a diverse group of genetic variants that include copy-number changes, insertions, deletions, inversions, repeat contraction, repeat expansions, and translocations. As these remain difficult to detect by routine short-read next-generation sequencing (Mahmoud et al., 2019), these categories of SVs have not been investigated with respect to their relationship with COVID-19 predisposition. According to the Human Gene Mutation Database (HGMD), more than 34% of all known disease-causing variations are larger than a single base pair change, i.e., single nucleotide variation (SNVs) - a trend that has been slowly edging upward over the past decade (Eichler, 2019). Several studies have demonstrated the importance of large SVs in the characterization of human immunity profiles (Watson et al., 2013; Isa et al., 2020). Despite the host genome investigation initiatives across the globe to understand the genetic predisposition to disease severity in COVID-19, a substantial portion of the genome and variant classes remains intractable because of the technical limitations of applied genomic sequencing technologies (Monlong et al., 2018; Brandt et al., 2015). The major limitation for SV detection using short-read next-generation sequencing is inherent to the data type and alignment algorithms that are effective for unique sequences but break down within repetitive regions, which are highly enriched with SVs (Chaisson et al., 2019).

To address the limited SV analysis related to COVID-19 symptom severity, we formed the COVID-19 Host Genome Structural Variation Consortium and performed a preliminary study using optical genome mapping (OGM) to identify SVs (>500 bp) in an unbiased fashion in severely ill patients with COVID-19. OGM is an emerging technology that images long DNA molecules (>150kb), labeled at specific sequence motifs (CTTAAG) that span the entire genome, enabling genome-wide coverage. The images are converted to molecules to generate *de novo* genome assembly, which can be compared to a reference genome to identify SVs and using a separate coverage based algorithm to identify CNVs. OGM has demonstrated its ability to accurately detect all classes of SVs including copy number variations, balanced translocations, inversions, repeat expansions, and contractions (Mantere et al., 2021; Neveling et al., 2021). Recently, a side-by-side comparison study of OGM with long-read sequencing technology (PacBio HiFi reads) by the Human Genome Structural Variation Consortium (HGSVC) revealed a higher sensitivity of OGM for detecting large SVs. We hypothesize that structural variants in genes implicated in the viral response pathway(s), intractable by other genomic profiling techniques, may predispose some individuals to severe COVID-19.

[11]Hannover Unified Biobank (HUB), Hannover 30625, Germany

[12]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA 23284, USA

[13]Division of Immunology, Boston Children's Hospital, Boston, MA 02115, USA

[14]Department of Anesthesiology, Critical Care, and Pain Medicine, Boston Children's Hospital, Boston, MA 02115, USA

[15]Departments of Anesthesia and Pediatrics, Harvard Medical School, Boston, MA 02115, USA

[16]Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

[17]Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY 10065, USA

[18]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

[19]COVID19hostgenomesv consortium https://www.covid19hostgenomesv.org/consortium.html

[20]Lead contact

*Correspondence: rkolhe@augusta.edu

https://doi.org/10.1016/j.isci.2022.103760

**Table 1. Clinical characteristics of patients included in the study**

| Patient Characteristics | Categories | Numbers/Mean ± SD |
|---|---|---|
| Patients | | 52 |
| Age | | 60.9 ± 13.9 (19–83) |
| Sex | Male | 27 |
| | Female | 25 |
| Ethnicity | African American | 34 |
| | Caucasian | 16 |
| | Hispanic | 2 |
| BMI | | 34.3 ± 7.9 (19.7–50.9) |
| Comorbidities | Diabetes | 22 |
| | Hypertension | 32 |
| | Chronic Kidney disease | 10 |
| | Asthma | 3 |
| ICU parameters | Requiring Mechanical Ventilation | 32 |
| | Mean duration of Intubation | 12 ± 5.5 (3–20) |
| | Intubation PTA | 13 |
| | Proned | 7 |
| Outcome | Dead | 13 |
| | Recovered | 39 |

## RESULTS

### Patient characteristics

A total of 52 severely ill COVID-19 patients that met inclusion criteria were prospectively included from March 2020 to February 2021. The demographics and clinical characteristics of the patients (Table 1) were a mean age of 60.9 ± 13.9 SD (range 19-83), 48% (25/52) female, 65.3% (34/52) African American, 30% (16/52) Caucasian, and two Hispanic, which is similar to the demographics of COVID-19 patients in the State of Georgia, USA (Gold et al., 2020). Chronic preexisting medical conditions in these patients included diabetes mellitus type 2 (42%), hypertension (61%), chronic kidney disease (19%), and asthma (5.7%). Six patients had no known comorbidities at the time of admission to the ICU. On ICU admission, 32 patients needed mechanical ventilation, with a mean intubation duration of 12 ± 5.5 days, of which, 25% (13/52) were intubated before admission (PTA), and 13.4% (7/52) were ventilated in prone-position. Of the 52 patients, 39 recovered and 13 died during ICU stay.

### Rare/unique SVs as strong predisposing factors associated with severe COVID-19

An unbiased whole-genome mapping investigation of the severely ill patients was performed using a filter for SVs that did not appear in the Bionano controls (Figure 1). The complete list of rare/unique SVs is provided in Data S1. On average, each of the 52 individuals with severe COVID-19 carried 40 rare SVs, i.e., a total sum of 2062 rare SVs in 52 cases. This number is very similar to recent manuscripts (using similar controls to define the rare nature of events) (Mantere et al., 2021). SVs changing the copy number or disrupting the gene coding regions, and implicated in viral host response pathways led to the identification of seven rare/unique SVs, impacting 31 genes, identified in nine patients that shared common chronic preexisting conditions. These included genes implicated in innate immunity and inflammatory response (*EDARADD, DPP4, ZDHHC1, STK26, and CD300* gene clusters), and viral replication and spread (*KRT15, KRT19, ZNF443)* (Table 2, Figures 2, 3, 4, 5, 6, 7, and 8). All these SVs were confirmed by qPCR (Data S2). In addition, a complete list of CNVs is provided as Data S3.

### Confirmation of Saphyr SVs by FaNDOM and qPCR

FaNDOM and qPCR dosage experiments confirmed all the 7 SV calls made by Bionano Access Software (Table 2).

### Expression analysis

We next evaluated the impact of a rare X-chromosomal SV that partially duplicates the 5′UTR and coding exon one of the *STK26* gene (Figure 8). *STK26* (also known as MST4) is widely expressed, with robust

**Table 2. List of unique structural variants overlapping coding region(s) of the gene(s) identified in patients with severe COVID-19**

| S.No | Chr | RefStartPos | RefEndPos | Size (bp) | Zygosity | SV | Gene(s) | Samples | Present in % of BNGO controls | gnomAD (Allele frequency) | qPCR | FaNDOM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 236,604,233 | 236,766,495 | 162,263 | _ | Duplication | *EDARADD, HEATR1* | 22,26 | 0 | 0.001059 | Yes | Yes |
| 2 | 2 | 162,887,379 | 162,911,439 | 10,027 | heterozygous | Deletion | *DPP4* | 38 | 0 | 0 | Yes | Yes |
| 3 | 16 | 67,308,198 | 67,455,019 | 146,821 | Heterozygous | Deletion | *PLEKHG4, KCTD19, LRRC36, TPPP3, U1, ZDHHC1* | 39 | 0 | 0 | Yes | Yes |
| 4 | 17 | 39,662,399 | 39,690,882 | 28,482 | _ | Duplication | *KRT15, KRT19* | 2 | 0 | 0.001637 | Yes | Yes |
| 5 | 17 | 71,844,581 | 72,678,517 | 833,937 | _ | Duplication_Split | *RPL38, MGC16275, TTYH2, Z49982, DNAI2, CD300E, CD300LD,CD300C, CD300LB, CD300A, GPRC5C, GPR142, BTBD17, KIF19, RAB37* | 19 | 0 | 0.0006741 | Yes | Yes |
| 6 | 19 | 12,512,276 | 12,552,113 | 39,838 | _ | Duplication | *ZNF443* | 13,18 | 0 | 0.008378 | Yes | Yes |
| 7 | X | 130,629,618 | 131,164,603 | 534,985 | _ | Duplication_Split | *OR13H1,LOC286467,5S_rRNA, STK26* | 44 | 0 | 0 | Yes | Yes |

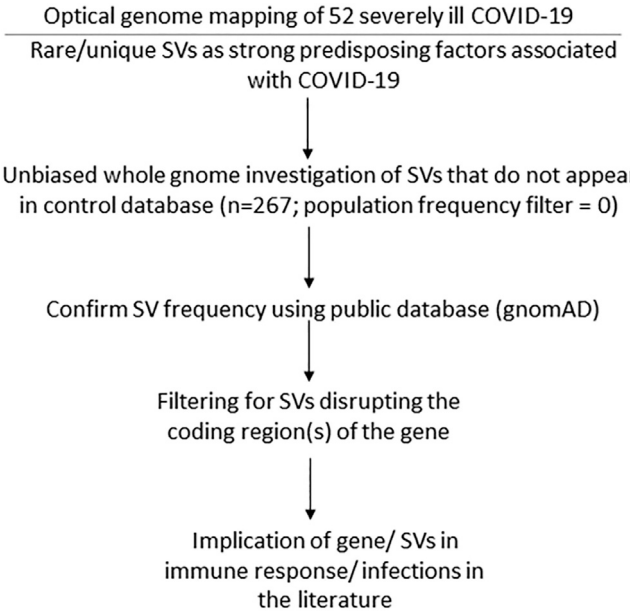



**Figure 1. Workflow diagram for analysis of COVID-19 patient SVs**

expression in monocytes as well as T, B, and NK cells, and has previously been investigated in the context of bacterial, although not viral, infections. It inhibits Toll-like receptor (TLR) signaling by phosphorylating the adaptor protein TRAF6, thereby impairing the assembly of TRAF6-mediated complexes needed for downstream TLR signaling (Jiao et al., 2015). Gene transcripts of *STK26* were quantified from the blood of 15 asymptomatic and 13 severely ill patients with COVID-19, normalized against 18S rRNA. *STK26* transcripts were found to be significantly ($7.2 \pm 2.3$ vs. $1.1 \pm 0.5$; $p < 0.001$) increased in severely ill patients compared to asymptomatic controls, with the highest expression observed in patient-44 (11.7-fold), who was the patient that harbored the SV partially duplicating the 5′UTR and coding exon one of the *STK26*. Interestingly, the disease-associated elevation of *STK26* expression was higher in females ($9.0 \pm 2.6$ vs. $1.1 \pm 0.4$; $p < 0.001$) compared to males ($4.8 \pm 1.3$ vs. $1.2 \pm 0.6$; $p < 0.001$) when severely ill patients were compared to asymptomatic controls. Gene transcripts for *TRAF6* ($29.7 \pm 11.8$ vs. $1.6 \pm 1.8$; $p < 0.001$) and *IRF3* ($2.8 \pm 1.2$ vs. $1.3 \pm 0.9$; $p < 0.001$) were significantly increased, whereas, the gene transcripts of interferon-stimulated genes, *ISG15* ($0.9 \pm 1.5$ vs. $2.7 \pm 4.1$; $p < 0.05$) and *IFIT1* ($0.5 \pm 0.7$ vs. $1.7 \pm 2.0$; $p < 0.001$) were significantly decreased in severely ill patients compared to asymptomatic controls. Expression of *ISG15* and *IFIT1* in patient-44 were amongs the lowest in the cohort (Figures 9A–9G).

### sDPP4 measurements

The ~24.1 kb heterozygous deletion on chr2 identified in patient 38 is an intragenic copy number loss of the *DPP4* gene deleting exons 3–5. This multi-exonic heterozygous copy number loss results most likely in an out-of-frame deletion (p.Thr32Thrfs*11). *DPP4* encodes dipeptidylpeptidase 4 (DPP4), a ubiquitously expressed enzyme that cleaves N-terminal proline-containing or alanine-containing dipeptides. DPP4 exists in two forms: 1) a membrane-bound form and 2) a cleaved product of the receptor (sDPP4) lacking the intracellular and transmembrane domain present in the circulation (Trzaskalski et al., 2020). sDPP4 interacts with GLP-1 and cytokines and regulates the T-cell receptor-mediated T cell activation. The sDPP4 serum concentration was significantly decreased in severely ill patients compared to asymptomatic controls ($1732.4 \pm 1049.6$ vs. $3015.3 \pm 853.0$; $p < 0.05$) ng/mL, with the lowest concentration observed in patient 38 (117 ng/mL) (Figure 9H), who harbored the SV of the *DPP4* gene (Figure 3).

## DISCUSSION

This study addressed the current technology bias of short-read sequencing that has confined most host genome studies to the interrogation of only SNVs and small indels associated with COVID-19 severity (Zhao et al., 2020a, 2020b; Ellinghaus et al., 2020; Pairo-Castineira et al., 2021; Van der Made et al., 2020; Fallerini et al., 2021; Zhang et al., 2020). A comparison of traditional cytogenetic technologies,

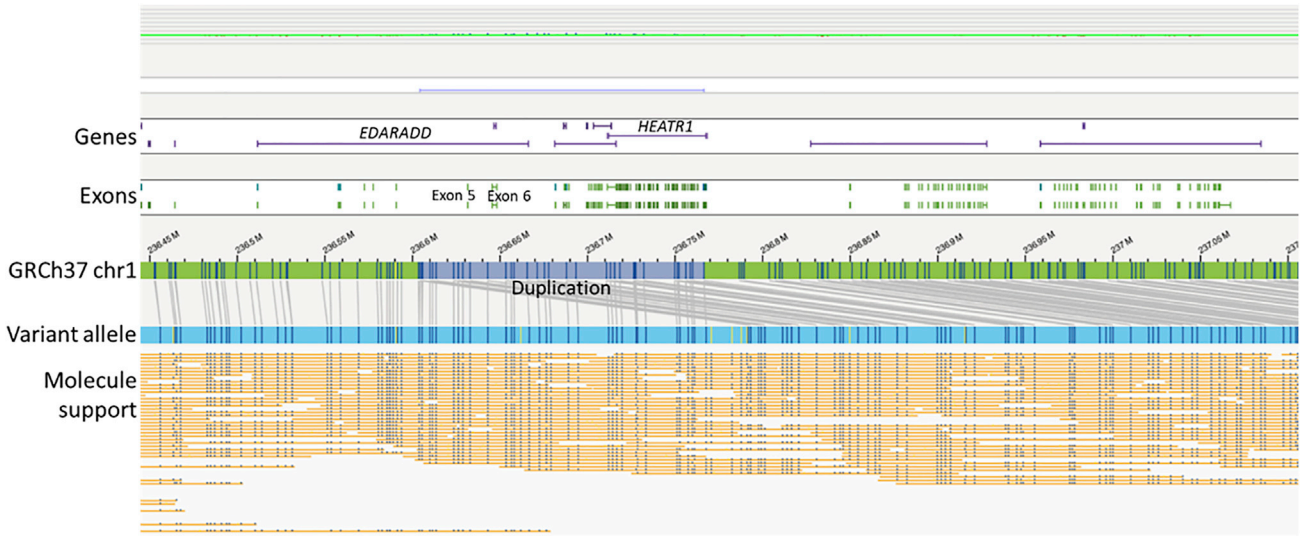



**Figure 2. A ~162.2 kb duplication of chr 1 (chr1:236,604,233-236,766,495) partially disrupting the *EDARADD* (exon five and six; NM_145861.3) and *HEATR1* (exons 3–45; NM_018072.6) genes in two patients (cases 22 and 26)**

short-read sequencing with OGM is provided as Data S4. Although these approaches have been successful in associating certain genetic loci, large gaps remain in explaining the wide diversity of clinical responses. Our data show that large SVs identified using optical genome mapping might further explain the interindividual clinical variability in response to COVID-19. Among 52 patients with severe COVID-19, we identified seven rare/unique SVs (not found in 267 Bionano controls) impacting 31 genes. Although individual effects of any of these SVs remain difficult to predict, several of these may be considered candidate loci as strongly predisposing factors associated with severe COVID-19. These SVs were implicated in three key host-viral interaction pathways: 1) innate immunity and inflammatory response (*EDARADD, DPP4, ZDHHC1, STK26, and CD300* gene cluster); and 2) viral replication and spread (*KRT15, KRT19, ZNF443*) (Figure 10).

## SVs affecting genes implicated in immune and inflammatory response

In this investigation, five rare/unique SVs impacting key immune genes (*STK26, ZDHHC1, DPP4, EDARADD,* and *CD300* genes cluster) were identified in six patients. Variability in the host's innate and adaptive immune responses to viral infections clearly can influence clinical manifestations and outcomes (Akira et al., 2006; Brubaker et al., 2015; Channappanavar et al., 2019). The innate immune response is the first line of defense against the invading pathogen and utilizes multiple pattern recognition receptors of which, Toll-like receptors (TLRs) have been implicated as key receptors in the recognition of ssRNA of MERS-CoV and SARS-CoV in murine models (Channappanavar et al., 2019; Choudhury et al., 2020). Independent studies have identified rare predicted loss of function (LOF) variants at 13 loci governing TLR3 and TLR7 pathway genes and affecting the type I interferon response (Van der Made et al., 2020; Fallerini et al., 2021; Zhang et al., 2020). In addition, several groups have identified impaired type I interferon response in severe COVID-19 pathogenesis (Hadjadj et al., 2020; Blanco-Melo et al., 2020). In this study, a rare/unique SV, ~534.9 kb tandem duplication of chr X was identified in one female patient (case 44). The SV partially disrupts *STK26*, duplicating the 5'UTR and the coding exon one region of the gene. In synthetic biology, dual 5'UTR constructs have shown increased transcription and translation of the gene (Balzer et al., 2020). Consistent with this hypothesis, this patient had the highest fold increase of mRNA transcripts as compared to asymptomatic COVID-19 patients, possibly exacerbating the underlying pathogenic mechanism (Figure 9A). *STK26* encodes for serine/threonine kinase 26, a negative regulator of TLR signaling (Konno et al., 2009), directly binds and phosphorylates TRAF6 at two threonine residues (T463 and T486) in the C-terminal TRAF domain, and inhibits the oligomerization and autoubiquitination activity of TRAF6 (Jiao et al., 2015). TRAF6 is a downstream signaling molecule in the type I-IFN pathways that mediates the NF-kB and type I interferon response (Van der Made et al., 2020). The mRNA transcripts of a selection of type I IFN pathway genes were measured to investigate whether the *STK26* overexpression inhibits TLR signaling and impairs the type I interferon response, putatively predisposing to severe

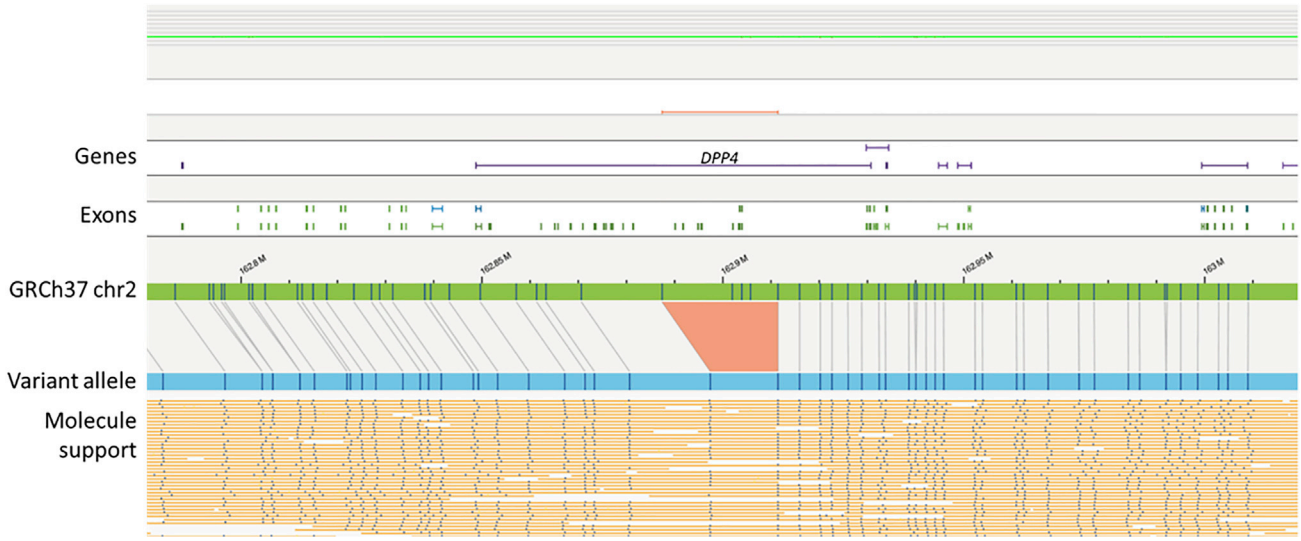



**Figure 3. A ~10.0 kb heterozygous deletion of chr 2 (chr2:162,887,379-162,911,439) partially deleting the *DPP4* gene (exons three and four; NM_001935.4) in one patient (case 38)**

COVID-19. Interestingly, the *TRAF6* and *IRF3* genes were significantly overexpressed in severely ill patients compared to asymptomatic controls, which is consistent with published reports (Figures 9D and 9E), (Sohn et al., 2020) whereas, the interferon-stimulated genes (ISGs), *ISG15,* and *IFIT1* were both significantly down-regulated in severely ill patients compared to asymptomatic controls, with the expression of ISGs in case-44 (harboring the *STK26* SV) amongst the lowest in the cohort (Figures 4F and 4G). The upregulation of *STK26* detected coincidentally in all severe COVID-19 patients, most notably in females, implicates this gene/pathway in the pathogenesis of severe COVID-19. Recently, Picchiotti et al. reported rare loss-of-function variants of *STK26* that were shown to be protective of COVID-19 in men in a GWAS study using 1,318 subjects (Picchiotti et al., 2021). These findings support the hypothesis that a variant leading to overexpression of *STK26* such as established in case-44 conversely leads to an elevated risk for severe disease, possibly resulting from the inhibition of TLR signaling, consequently impairing the type I interferon response leading to reduced viral clearance. It should be further investigated whether this is the result of an effort to

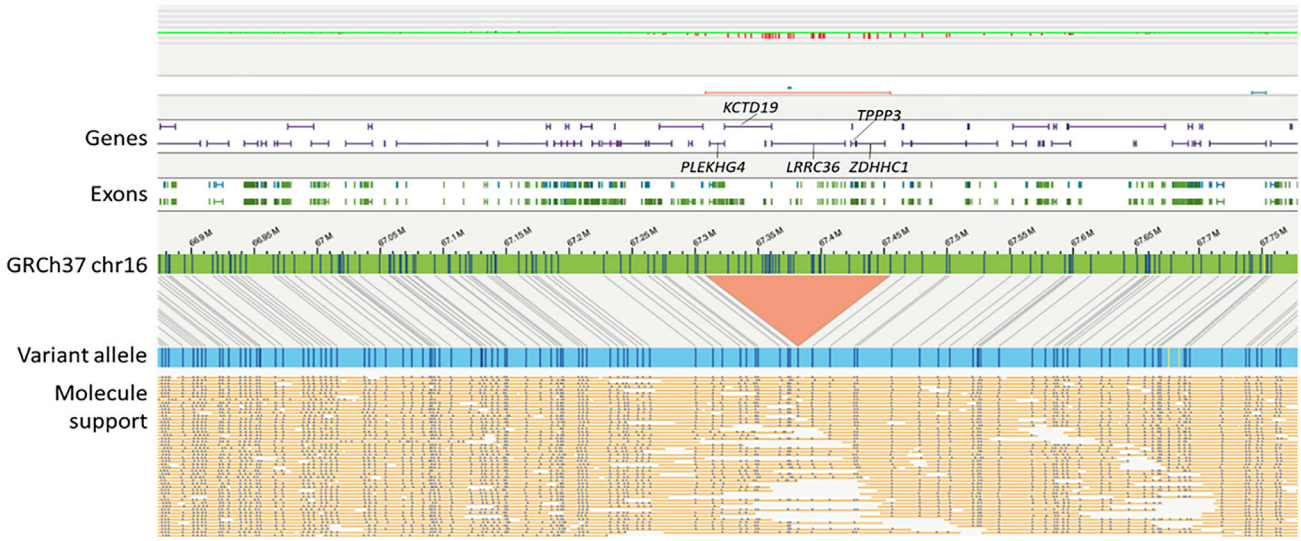



**Figure 4. A ~146.8 kb heterozygous deletion of chr 16 (chr16:67,308,198-67,455,019) that contained six genes (*PLEKHG4, KCTD19, LRRC36, TPPP3, U1, ZDHHC1*) was identified in one patient (case 39)**

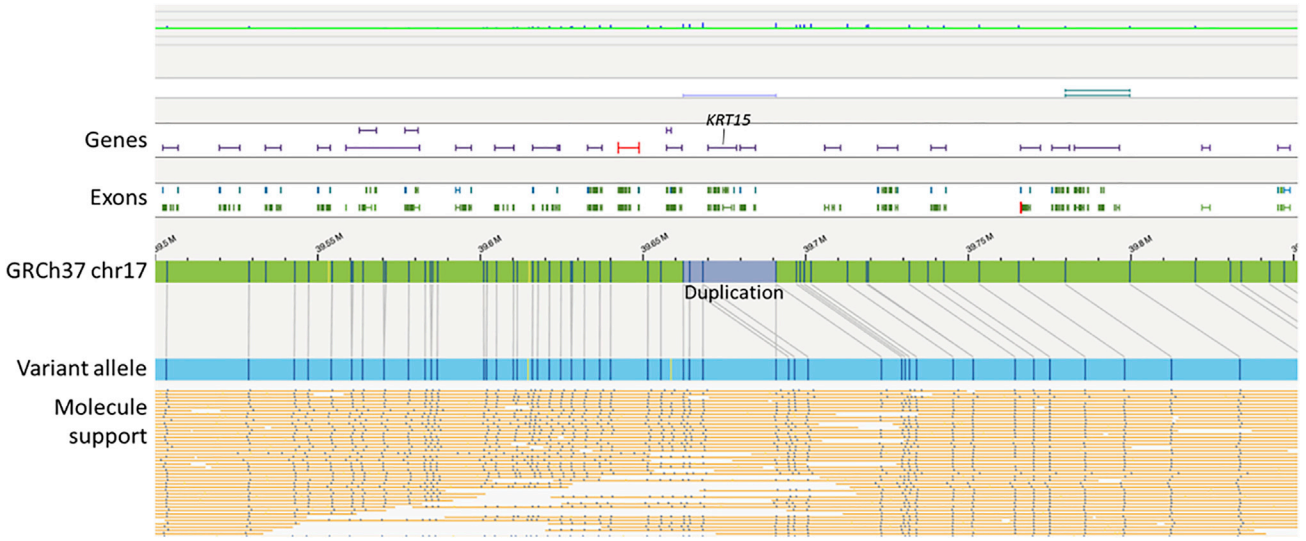



**Figure 5. A ~ 28.4 kb duplication of chr 17 (chr17:39,662,399-39,690,882) including two genes (*KRT15* and *KRT19*) in one patient (case 2)**

counter regulate the overproduction of type I IFNs during the disease course of COVID-19, or whether this repression of type I IFN signaling precedes severe disease.

The *ZDHHC1* gene deleted in a ~146.8 kb heterozygous region in one patient (case 39), is also a gene implicated in the type I interferon response. *ZDHHC1* expression has been shown to mount a type I interferon response against DNA and RNA viruses via different downstream players (Xu et al., 2017). *ZDHHC1* encodes an endoplasmic reticulum-associated palmitoyltransferase that mediates MITA-dependent IRF3 activation and type I interferon response against DNA viruses (Figure 10) (Wang et al., 2020). Predominantly, in RNA virus infection, ZDHHC1 mediates IFITM3 palmitoylation, which ensures the antiviral activity of IFITM3. IFITM3, localized in the endosomal and endolysosomal compartments of cells, prevents the viral entry through the lipid bilayer into the cytoplasm, preventing viral fusion with cholesterol-depleted endosomes against multiple viruses including influenza, SARS-CoV, and HIV-1 (Wang et al., 2020). Knockdown of

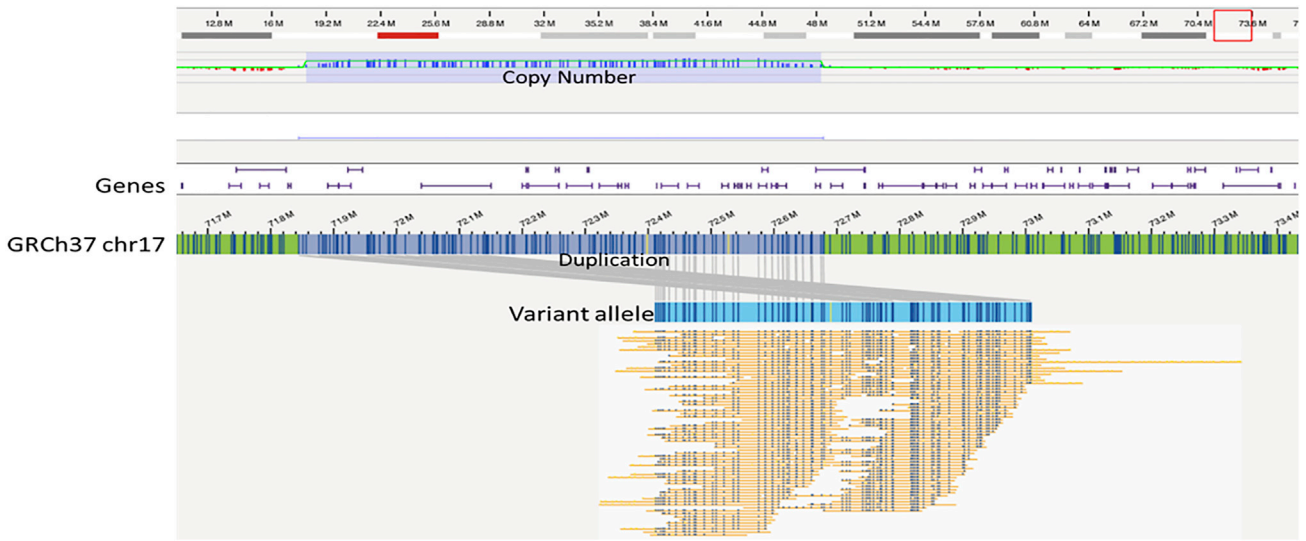



**Figure 6. A ~833 kb tandem duplication of chr 17 (chr17:71,844,581-72,678,517) including 15 genes (*RPL38, MGC16275, TTYH2,* Z49982, *DNAI2, CD300E, CD300LD, CD300C, CD300LB, CD300A, GPRC5C, GPR142, BTBD17, and KIF19*), and partially disrupting *RAB37* (exon one; NM_175738.5) in one patient (case 19)**

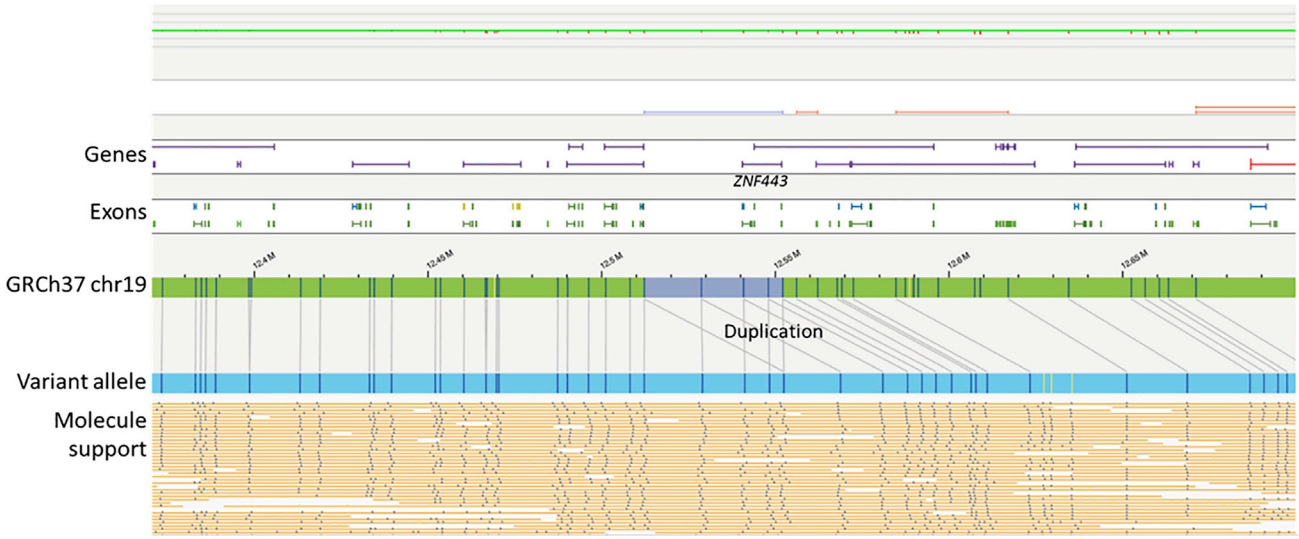



**Figure 7. A ~39.8 kb duplication of chr 19 (chr19:12,512,276- 12,552,113) including the entire *ZNF443* gene (NM_005815.5) in two patients (cases 13 and 18)**

ZDHHC1 leads to compromised ZDHHC1/IFITM3 antiviral activity against the Japanese encephalitis virus (Wang et al., 2020). Additionally, IFITM3 expression was found to be upregulated early in epithelial lung cells in response to SARS-CoV-2 (Bozzo et al., 2020), and that SNPs in IFITM3 are associated with more severe disease (Nikoloudis et al., 2020). We hypothesize that the heterozygous deletion of the *ZDHHC1* gene could compromise the type I interferon response, especially at the level of IFITM3, which might have led to the severe COVID-19 symptoms in this patient.

The rare/unique SV involving the *DPP4* gene that encodes for T cell activation antigen CD26 in a severely ill patient (case 38) is consistent with prior studies implicating dipeptidylpeptidase 4 (DPP4) activity in MERS-CoV (Reinhold and Brocke, 2014; Inn et al., 2018) and severe COVID-19 (Schlicht et al., 2020). Recently, *DPP9*, a family member of this gene, has been associated with severe COVID-19 in a GWAS study (Pairo-Castineira et al., 2021). Although bioinformatics models predicted human DPP4-SARS-CoV-2 protein interactions, and protein docking studies show SARS-CoV-2 virus utilizes DPP4 as a co-receptor (not as strong as with ACE-2) (Li et al., 2020a, 2020b), *in vitro* studies found that DPP4 serves as a viral entry receptor for MERS-CoV, but not for SARS-CoV-2 (Letko et al., 2020). Apart from the membrane-bound receptor, the circulating sDPP4 enzyme interacts with GLP-1 and cytokines and regulates T-cell receptor-mediated T cell activation. The ~24.1 kb heterozygous deletion on chr2 identified in patient 38 is an intragenic copy number loss of the *DPP4* gene deleting exons 3–5. This multi-exonic heterozygous copy number loss results most likely in an out-of-frame deletion (p.Thr32Thrfs*11) (Klemann et al., 2016). Notably, the lowest circulating sDPP4 level was observed in this patient, but in addition, sDDP4 levels were significantly decreased in severely ill patients compared to asymptomatic controls. Circulating sDPP4 levels have been reported to be reduced in both MERS-CoV and SARS-CoV-2 infection but to contrasting effects (Inn et al., 2018; Schlicht et al., 2020). In MERS-CoV infection, the sDPP4 levels were below the threshold needed to exert an antiviral effect, as sufficient viral particles were available for entry via the membrane-bound receptor. The reduced sDPP4 levels possibly reflect a compromised immune response and may not signify competition for SARS-CoV-2 entry via the DPP4 receptor. Several case reports and the Japanese Adverse Drug Event Report (JADER) have highlighted the risk and incidence of interstitial pneumonia with DPP4 inhibitor (vildagliptin) (Japanese Adverse Drug Event Report database, 2004; Sada et al., 2014; Tanaka et al., 2020). Further, loss of DPP4 activity is associated with a prothrombotic state in myocardial microvessels because of the upregulation of the procoagulant tissue factor (Krijnen et al., 2012), and reduced postoperative DPP4 activity has been associated with worse patient outcomes after cardiac surgery because of paradoxical impairment of angiogenesis and endothelial function (Sun et al., 2013; Noels et al., 2018). Taken together, the out-of-frame deletion of *DPP4* likely leads to significantly reduced sDPP4 levels that might have contributed toward compromised immune response predisposing to severe COVID-19.

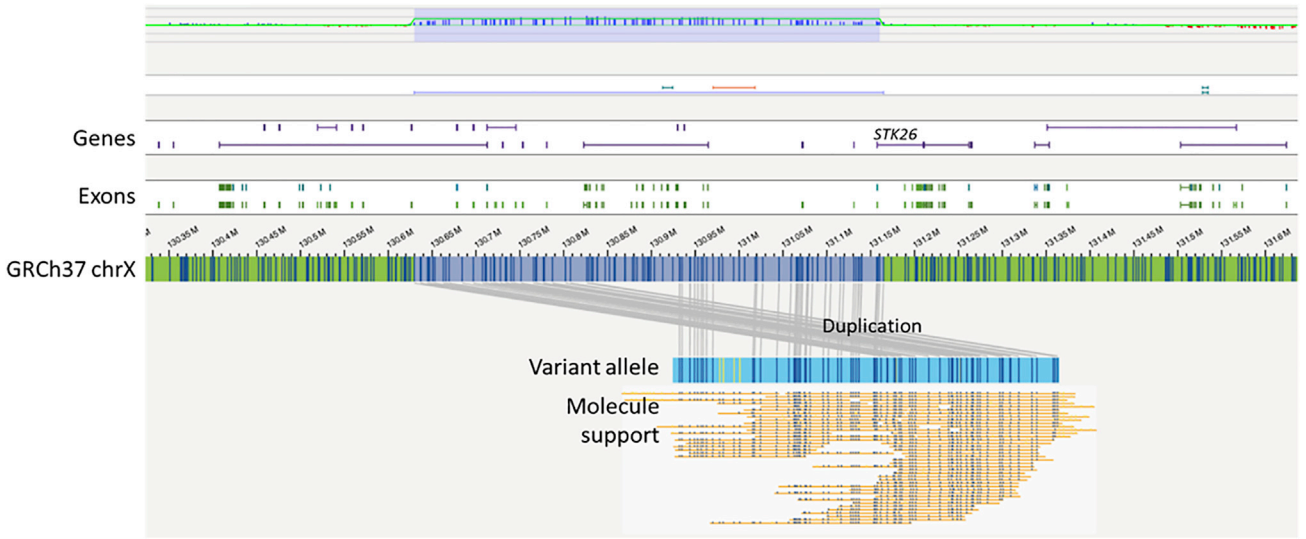



**Figure 8. A ~ 534.9 kb tandem duplication of chr X (chrX: 130,629,618–131,164,603) including four genes (*OR13H1*, LOC286467, *5S_rRNA*, *STK26*) in one patient (case 44)**

The rare/unique 833 kb tandem duplication of 14 genes (*RPL38, MGC16275, TTYH2, Z49982, DNAI2, CD300E, CD300LD, CD300C, CD300LB, CD300A, GPRC5C, GPR142, BTBD17, and KIF19*) in one patient (case 19) includes the cluster of CD300 genes that constitute an important family of receptors expressed on immune cells (myeloid and lymphoid). CD300 molecules have been identified to participate in mechanisms employed by viruses to escape the immune response and infect host cells (Carnec et al., 2015; Vitallé et al., 2019). Further, CD300 molecules downregulate the cytolytic activity of natural killer (NK) cells against infected cells (Vitallé et al., 2019). Duplication of the region involving CD300 genes in this patient may lead to overexpression of these genes that help the virus to evade the immune response leading to disease severity upon infection.

The 162.2 kb duplication partially duplicating the *EDARADD* gene is among the three genes identified to cause ectodermal dysplasia (ED) and is involved in NF-kB activation (Martínez-Romero et al., 2019; Smahi et al., 2002). The two patients with this SV did not have any form of ectodermal dysplasia and the SV was not found in any public database (DGV and gnomAD, last accessed $1^{1}/_{8}$/2020). To our knowledge, nine *EDARADD* pathogenic sequencing variants have been described, three leading to an autosomal recessive inheritance and six to an autosomal dominant mode of inheritance (Cluzeau et al., 2019). ED caused by variants in the *EDARADD* gene is associated with recurrent pulmonary infections, accompanied by bronchospasm that requires steroidal treatment, and has been found to impair NF-κB signaling (Martínez-Romero et al., 2019). Cluzeau et al., have identified a homozygous deletion of exon four in a Tunisian family with severe ED that resulted in the complete abolition of NF-KB signaling (Cluzeau et al., 2019). In addition, several reports have shown that a heterozygous mutation in *EDARADD* in patients with ECTD11A leads to severe impairment of NF-κB activation (Bal et al., 2007; Wohlfart et al., 2016). Further, Wohlfart et al. reported a heterozygous missense variant in the *EDARADD* gene that did not impact the interaction between EDAR and EDARADD proteins but led to an impaired ability of the gene to activate NF-κB signaling (Wohlfart et al., 2016). Notably, this patient did not exhibit a phenotype consistent with autosomal dominant ED. Further investigation of the functional consequences of these variants is needed for a more informed interpretation.

Although these findings are preliminary in the abovementioned immune genes, some candidate genes/loci suggest interesting follow-up opportunities. Generally, some observations seem to be in line with individual variants affecting key immune response genes - collectively they may well explain a few percent of all severe/life-threatening COVID-19 cases (Van der Made et al., 2020; Zhang et al., 2020). The highlighted rare/unique SVs all deserve further follow-up before any conclusions can be drawn. These may include segregation analyses in the respective families and additional downstream expression analysis or protein levels in patient plasma as well as functional studies. It would also be of interest to learn

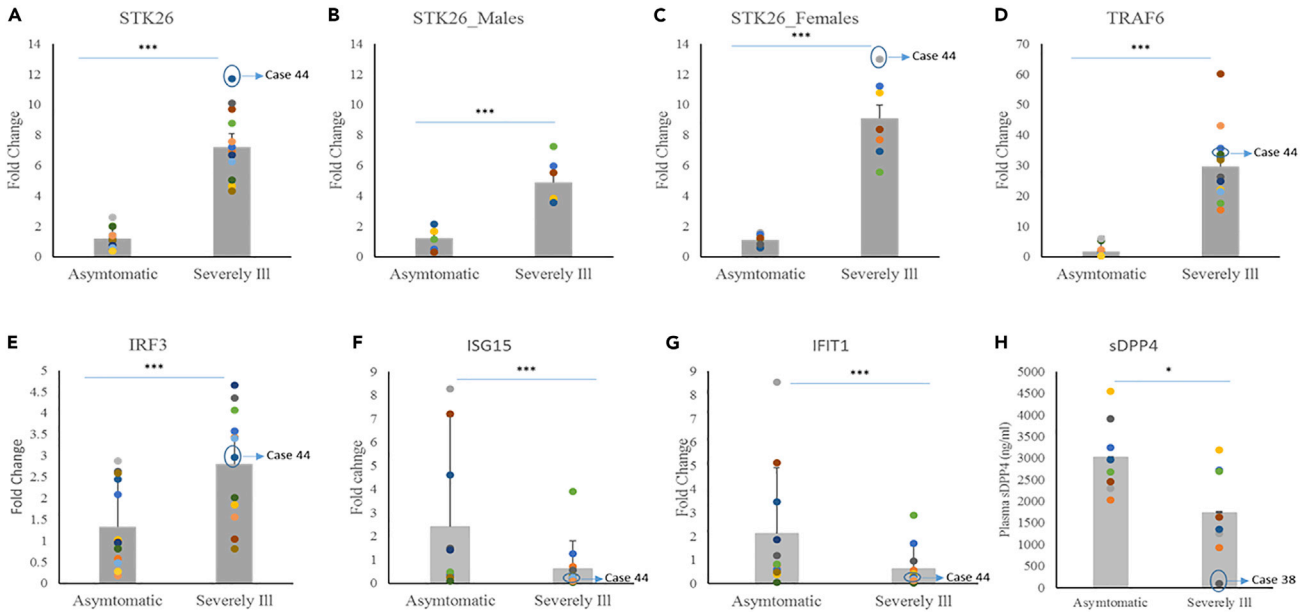



**Figure 9. Expression and protein analysis of select markers**

(A) *STK26* transcripts were found to be significantly (7.2 ± 2.3 vs. 1.1 ± 0.5; $p < 0.001$) increased in severely ill patients compared to asymptomatic controls, with the highest expression observed in patient-44 (11.7-fold), who was the patient that harbored the SV partially duplicating the 5'UTR and coding exon one of the STK26 gene.

(B) The gene transcripts of STK26 were found to be significantly increased in males (4.8 ± 1.3 vs. 1.2 ± 0.6; $p < 0.001$), and(C) females (9.0 ± 2.6 vs. 1.1 ± 0.4; $p < 0.001$) in severely ill patients compared to asymptomatic controls.

(D) The gene transcripts for TRAF6 (29.7 ± 11.8 vs. 1.6 ± 1.8; $p < 0.001$) and (E) IRF3 (2.8 ± 1.2 vs. 1.3 ± 0.9; $p < 0.001$) were significantly increased, whereas, the gene transcripts of interferon-stimulated genes, (F) ISG15 (0.9 ± 1.5 vs. 2.7 ± 4.1; $p < 0.05$) and (G) IFIT1 (0.5 ± 0.7 vs. 1.7 ± 2.0; $p < 0.001$) was significantly decreased in severely ill patients compared to asymptomatic controls.

(H) The sDPP4 serum concentration was significantly decreased in severely ill patients compared to asymptomatic controls (1732.4 ± 1049.6 vs. 3015.3 ± 853.0; $p < 0.05$) ng/mL, with the lowest concentration observed in patient 38 (117 ng/mL), who harbored the SV partially deleting exon three to five of the DPP4 gene

whether any of the identified candidate genes harbor rare/unique point mutations (SNVs/InDels) in WES/WGS data of severe COVID-19 cases or show altered expression levels in larger cohorts of severe COVID-19 cases.

## Rare/unique SVs affecting genes implicated in viral spread and replication

Two rare/unique SVs impacting the keratin genes (*KRT15* and *KRT19)* and *ZNF443* genes were found in three patients. A rare/unique SV, with a ~28.4 kb duplication including two genes (*KRT15* and *KRT19*) was identified in one patient (case 2). Keratins are intermediate filament proteins responsible for the structural integrity of epithelial cells and are part of the cytoskeletal structure of host cells. The keratins, including *KRT15,* have been identified to play a major role in the cell-to-cell transmission (i.e., spread) of influenza and SARS-CoV-2 viruses, and are upregulated in the respiratory tract during infection (Pociask et al., 2017; Basu et al., 2020). Another rare/unique duplication was found in the *ZNF443* gene, which has sequence homology to the SARS-CoV-2 virus and might aid in the replication process (Pasquier and Robichon, 2020a; 2020b). Duplication of these genes could lead to overexpression and it is hypothesized that this might assist viral replication and spread leading to increased disease severity.

The current study provides a powerful framework for investigating structural variations in host genomes that might predispose individuals to severe COVID-19. To our knowledge, this is the first study investigating structural variation using optical genome mapping in host-virus interactions in SARS-CoV-2 infected patients. The diverse interindividual variability observed in the human population after SARS-CoV-2 infection attributed to the disease severity, transmissibility, viral titers, and immune response highlights that numerous pathways/genes are involved in this complex host-viral

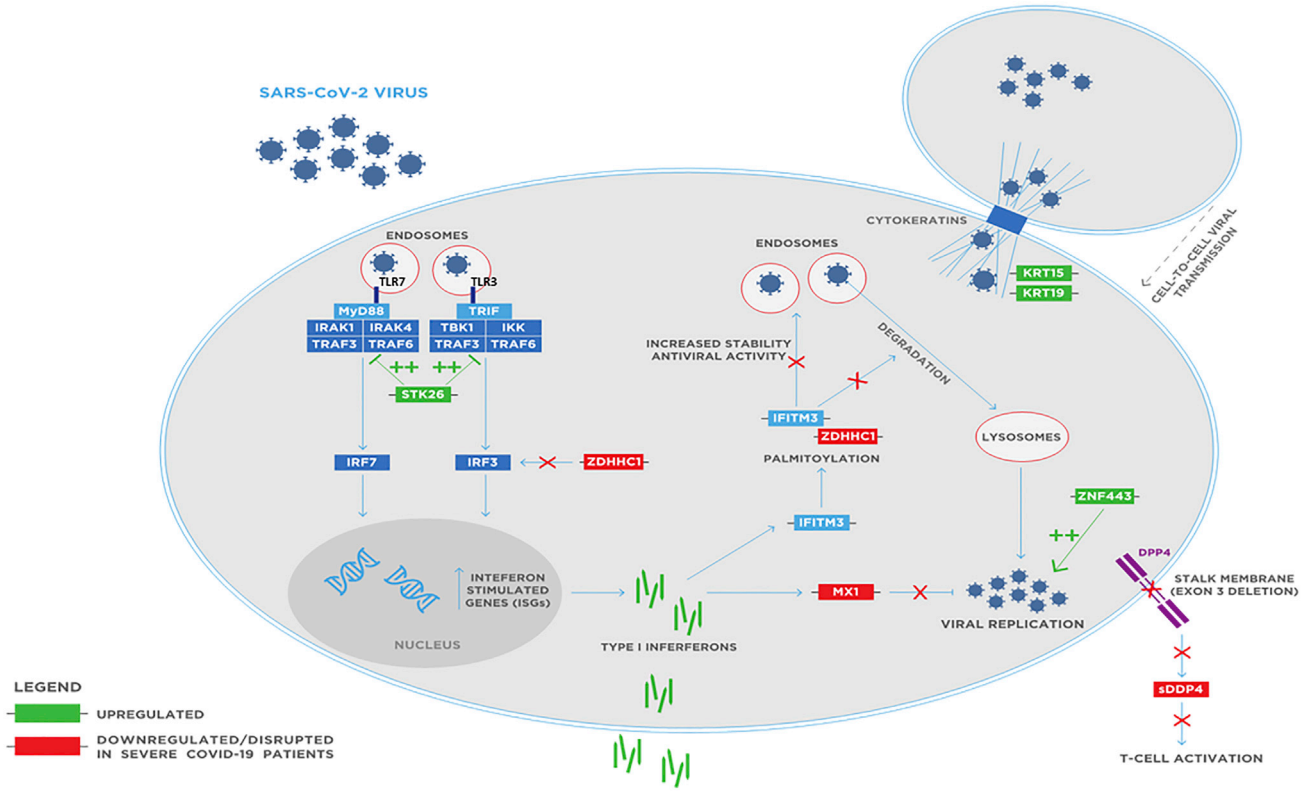



**Figure 10. A diagram of the key viral host interaction pathways highlighting genes (in green and red) found to contain structural variation in COVID-19 patients**

interaction. It is insufficient to consider only genomic variation at the level of SNVs/small indels to understand host variability in response to COVID-19 exposure. The present investigation highlights rare SVs in genes implicated in distinct pathways of host-viral interaction disrupted by structural variations in patients with severe COVID-19. The events affecting these genes might have predisposed these patients toward a severe disease course. Further, several findings in this study are consistent with and support previous reports implicating the type I interferon pathway and type I interferon response in the pathogenesis of COVID-19. One finding, a rare SV leading to partial duplication of the TLR negative regulator, *STK26*, and the coincident upregulation of *STK26* is especially intriguing. Remarkably, this finding led us to discover that, relative to levels in asymptomatic COVID-19 cases, *STK26* was upregulated in all 13 tested patients with severe COVID-19, with the highest expression in the patient with the partial duplication. This overexpression may be part of a more widespread severe response profile, but direct involvement of the SV identified in the respective individual in disease severity seems possible. Another rare SV leads to a partial deletion of *DPP4*, associated with very low sDPP4 levels in the respective individual. Intriguingly, we identified low sDPP4 levels in all severely affected COVID-19 patients in line with previous literature. In summary, we propose that genomic analyses to understand the human variation that impacts the predisposition to severe COVID-19 should include systematic SV assessment to fully understand the host genome. Here we exemplify that rare SVs may influence the host immune response, and highlight two rare SVs in two severely affected individuals that may add to a better understanding of COVID-19 host genetics.

## Limitations of the study

There are some limitations in the present investigation, which will benefit from additional follow-up studies. Although the study is limited to a small patient population, rare SVs were significantly enriched for immune genes, so the findings would be made stronger by validation in a larger cohort. Second, the study focused on large SVs (>500 bp) and did not investigate smaller indels that might

also be implicated as predisposition factors associated with severe COVID-19. Third, better-matched controls for the genomic analysis, such as infected close relatives with distinct clinical manifestations (mild/asymptomatic), would help control for the confounding influences of common environmental factors. Gaining access to more ideal control samples and performing optical mapping in larger cohorts was beyond the scope of this initial study. However, to address these limitations, the following measures were implemented: comparing the frequency of SVs to Bionano control dataset and public databases (DGV and gnomAD), and identifying unique SVs that do not appear in controls and SVs that overlap/disrupt the coding region of the genes implicated in host-viral interaction. Fourth, the samples of the patients were not drawn during the same point in their disease course; however, the functional data shows a clear distinction between asymptomatic patients and those with severe disease. Nevertheless, the whole genome investigation of 52 severely ill patients with COVID-19 identified candidate loci that may provide novel insights into the pathogenesis of severe COVID-19 in some cases.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study participants
- METHOD DETAILS
  - Optical genome mapping
  - Data analysis
  - Quantitative PCR (qPCR) confirmation of OGM SVs
  - Independent assessment of Saphyr SVs by FaNDOM
  - Expression analysis
  - Soluble DPP4 (sDPP4) measurements
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.103760.

## ACKNOWLEDGMENTS

The authors thank Christoph Lange for providing valuable feedback for the manuscript and Ben Clifford for providing technical support.

## AUTHOR CONTRIBUTIONS

Conceptualization, N.S.S., A.C., and R.K.; Data curation, N.S.S., S.J., and A.K.M.; Formal analysis, N.S.S., C.J.L., A.H., S.R.D., M.B.B., M.C.Z., F.A.A., V.B., A.C., and R.K.; Funding acquisition, R.K.; Investigation: N.S.S., C.J.L., A.H., A.C., and R.K.; Methodology: N.S.S., A.H., C.I.V.D.M., M.C.Z., C.A.B., A.H.B., V.B., E.D.J., A.H., A.C., and R.K.;. Visualization: N.S.S., C.J.L., A.H., A.C., and R.K.;. Writing—original draft: N.S.S.; Writing—review & editing: N.S.S., C.J.L., A.H., S.R.D., C.I.V.D.M., O.F., R.K.S., B.L., M.S., T.I., S.A.B., J.S.C., A.G.R., A.M.R., M.C.Z., C.A.B., A.H.B., V.B., E.D.J., A.H., A.C., and R.K. All authors have read and agreed to the published version of the manuscript.

## DECLARATION OF INTEREST

J.L., A.H., and A.C. are salaried employees of Bionano Genomics Inc. R.K. has received honoraria, travel funding, and research support from Illumina, Asuragen, QIAGEN, Perkin Elmer Inc, Bionano Genomics, and BMS. A.H.B. has received funding from the NIH, MDA (USA), AFM Telethon, Alexion Pharmaceuticals Inc., Audentes Therapeutics Inc., Dynacure SAS, and Pfizer Inc. He has consulted and received compensation or honoraria from Asklepios BioPharmaceutical Inc, Audentes Therapeutics, Biogen, F. Hoffman-La

Roche AG, GLG Inc, Guidepoint Global, and Kate Therapeutics, and holds equity in Ballard Biologics and Kate Therapeutics.

Received: August 22, 2021
Revised: November 4, 2021
Accepted: January 7, 2022
Published: February 18, 2022

## REFERENCES

Abbasi, J. (2020). Younger adults caught in COVID-19 crosshairs as demographics shift. JAMA *324*, 2141–2143.

Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. Cell *124*, 783–801.

Bal, E., Baala, L., Cluzeau, C., El Kerch, F., Ouldim, K., Hadj-Rabia, S., Bodemer, C., Munnich, A., Courtois, G., Sefiani, A., and Smahi, A. (2007). Autosomal dominant anhidrotic ectodermal dysplasias at the EDARADD locus. Hum. Mutat. *28*, 703–709.

Balzer, L.S., Onsager, I., Lorentzen, J.A., and Lale, R. (2020). Dual UTR-A novel 5′ untranslated region design for synthetic biology applications. Synth. Biol. *5*, ysaa006.

Basu, A., Sarkar, A., and Maulik, U. (2020). Study of cell to cell transmission of SARS CoV 2 virus particle using gene network from micro array data. bioRxiv. https://doi.org/10.1101/2020.05.26.116780.

Berlin, D.A., Gulick, R.M., and Martinez, F.J. (2020). Severe covid-19. N. Engl. J. Med. *383*, 2451–2460.

Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.C., Uhl, S., Hoagland, D., Møller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D., et al. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell *181*, 1036–1045.e9.

Bozzo, C.P., Nchioua, R., Volcic, M., Krüger, J., Heller, S., Stürzel, C.M., Kmiec, D., Conzelmann, C., Müller, J., Zech, F., and Schütz, D. (2020). IFITM proteins promote SARS-CoV-2 infection and are targets for virus inhibition. bioRxiv. https://doi.org/10.1101/2020.08.18.255935.

Brandt, D.Y., Aguiar, V.R., Bitarello, B.D., Nunes, K., Goudet, J., and Meyer, D. (2015). Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. G3 (Bethesda) *5*, 931–941.

Brubaker, S.W., Bonham, K.S., Zanoni, I., and Kagan, J.C. (2015). Innate immune pattern recognition: a cell biological perspective. Annu. Rev. Immunol. *33*, 257–290.

Carnec, X., Meertens, L., Dejarnac, O., Perera-Lecoin, M., Hafirassou, M.L., Kitaura, J., Ramdasi, R., Schwartz, O., and Amara, A. (2015). The phosphatidylserine and phosphatidylethanolamine receptor CD300a binds dengue virus and enhances infection. J. Virol. *90*, 92–102.

Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. *10*, 1784.

Channappanavar, R., Fehr, A.R., Zheng, J., Wohlford-Lenane, C., Abrahante, J.E., Mack, M., Sompallae, R., McCray, P.B., Jr., Meyerholz, D.K., and Perlman, S. (2019). IFN-I response timing relative to virus replication determines MERS coronavirus infection outcomes. J. Clin. Invest. *129*, 3625–3639.

Choudhury, A., Das, N.C., Patra, R., Bhattacharya, M., and Mukherjee, S. (2020). In silico analyses on the comparative sensing of SARS-CoV-2 mRNA by intracellular TLRs of human. bioRxiv. https://doi.org/10.1101/2020.11.11.377713.

Cluzeau, C., Marrakchi, S., Picard, C., Munnich, A., Smahi, A., and Turki, H. (2019). First homozygous large deletion in EDARADD gene associated with a severe form of anhidrotic ectodermal dysplasia. J. Eur. Acad. Dermatol. Venereol. *33*, e55–e57.

COVID-19 Host Genetics Initiative (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur. J. Hum. Genet. *28*, 715–718.

Dehkordi, R.S., Luebeck, J., and Bafna, V. (2021). FaNDOM: fast nested distance-based seeding of optical maps. Patterns *2*, 100248.

Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., et al. (2020). Features of 20133UK patients in hospital with covid -19using the ISARIC WHO Clinical Characterisation Protocol :prospective observational cohort study. BMJ *369*, m1985.

Ebert, P., Audano, P.A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M.J., Sulovari, A., Ebler, J., Zhou, W., Mari, R.S., et al. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science *372*, eabf7117.

Eichler, E.E. (2019). Genetic variation, comparative genomics, and the diagnosis of disease. N. Engl. J. Med. *381*, 64–74.

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., et al. (2020). Genomewide association study of severe covid-19 with respiratory failure. N. Engl. J. Med. *383*, 1522–1534.

Fallerini, C., Daga, S., Mantovani, S., Benetti, E., Picchiotti, N., Francisci, D., Paciosi, F., Schiaroli, E., Baldassarri, M., Fava, F., et al. (2021). Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: findings from a nested case-control study. Elife *10*, e67569.

Gold, J.A.W., Wong, K.K., Szablewski, C.M., Patel, P.R., Rossow, J., da Silva, J., Natarajan, P., Morris, S.B., Fanfair, R.N., Rogers-Brown, J., et al. (2020). Characteristics and clinical outcomes of adult patients hospitalized with COVID-19 - Georgia. MMWR Morb. Mortal. Wkly. Rep. *69*, 545–550.

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., et al. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science *369*, 718–724.

Inn, K.S., Kim, Y., Aigerim, A., Park, U., Hwang, E.S., Choi, M.S., Kim, Y.S., and Cho, N.H. (2018). Reduction of soluble dipeptidyl peptidase 4 levels in plasma of patients infected with Middle East respiratory syndrome coronavirus. Virology *518*, 324–327.

Isa, I.I., Jamaluddin, J., Achim, N.H., and Abubakar, S. (2020). Population-specific profiling of CCL3L1 copy number of the three major ethnic groups in Malaysia and the implication on HIV susceptibility. Gene *754*, 144821.

Izzy, S., Tahir, Z., Cote, D.J., Al Jarrah, A., Roberts, M.B., Turbett, S., Kadar, A., Smirnakis, S.M., Feske, S.K., Zafonte, R., et al. (2020). Characteristics and outcomes of latinx patients with COVID-19 in comparison with other ethnic and racial groups. Open Forum Infect. Dis. *7*, ofaa401.

Japanese adverse drug event report database (2004), Pharmaceutical and Medical Device Agency. https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0005.html.

Jiao, S., Zhang, Z., Li, C., Huang, M., Shi, Z., Wang, Y., Song, X., Liu, H., Li, C., Chen, M., et al. (2015). The kinase MST4 limits inflammatory responses through direct phosphorylation of the adaptor TRAF6. Nat. Immunol. *16*, 246–257.

Klemann, C., Wagner, L., Stephan, M., and von Hörsten, S. (2016). Cut to the chase: a review of CD26/dipeptidyl peptidase-4's (DPP4) entanglement in the immune system. Clin. Exp. Immunol. *185*, 1–21.

Konno, H., Yamamoto, T., Yamazaki, K., Gohda, J., Akiyama, T., Semba, K., Goto, H., Kato, A., Yujiri, T., Imai, T., and Kawaguchi, Y. (2009). TRAF6 establishes innate immune responses by activating NF-κB and IRF7 upon sensing cytosolic viral RNA and DNA. PLoS One *4*, e5674.

Kosmicki, J.A., Horowitz, J.E., Banerjee, N., Lanche, R., Marcketta, A., Maxwell, E., Bai, X., Sun, D., Backman, J.D., Sharma, D., et al. (2021). Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. Am. J. Hum. Genet. *S0002-9297*, 00222–00226.

Krijnen, P.A., Hahn, N.E., Kholová, I., Baylan, U., Sipkens, J.A., van Alphen, F.P., Vonk, A.B., Simsek, S., Meischl, C., Schalkwijk, C.G., and van Buul, J.D. (2012). Loss of DPP4 activity is related to a prothrombogenic status of endothelial cells: implications for the coronary microvasculature of myocardial infarction patients. Basic Res. Cardiol. *107*, 233.

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat. Microbiol. *5*, 562–569.

Levy-Sakin, M., Pastor, S., Mostovoy, Y., Li, L., Leung, A.K., McCaffrey, J., Young, E., Lam, E.T., Hastie, A.R., Wong, K.H., et al. (2019). Genome maps across 26 human populations reveal population-specific patterns of structural variation. Nat. Commun. *10*, 1–4.

Li, X., Xu, S., Yu, M., Wang, K., Tao, Y., Zhou, Y., Shi, J., Zhou, M., Wu, B., Yang, Z., et al. (2020a). Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J. Allergy Clin. Immunol. *146*, 110–118.

Li, Y., Zhang, Z., Yang, L., Lian, X., Xie, Y., Li, S., Xin, S., Cao, P., and Lu, J. (2020b). The MERS-CoV receptor DPP4 as a candidate binding target of the SARS-CoV-2 spike. iScience *23*, 101160.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods *4*, 402–408.

LoPresti, M., Beck, D.B., Duggal, P., Cummings, D.A.T., and Solomonm B, D. (2020). The role of host genetic factors in coronavirus susceptibility: review of animal and systematic review of human literature. Am. J. Hum. Genet. *107*, 381–402.

Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., and Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it. Genome Biol. *20*, 246.

Mantere, T., Neveling, K., Pebrel-Richard, C., Benoist, M., van der Zande, G., Kater-Baats, E., Baatout, I., van Beek, R., Yammine, T., Oorsprong, M., et al. (2021). Optical genome mapping enables constitutional chromosomal aberration detection. Am. J. Hum. Genet. *108*, 1409–1422.

Martínez-Romero, M.C., Ballesta-Martínez, M.J., López-González, V., Sánchez-Soler, M.J., Serrano-Antón, A.T., Barreda-Sánchez, M., Rodriguez-Peña, L., Martínez-Menchon, M.T., Frías-Iniesta, J., Sánchez-Pedreño, P., et al. (2019). EDA, EDAR, EDARADD and WNT10A allelic variants in patients with ectodermal derivative impairment in the Spanish population. Orphanet J. Rare. Dis. *14*, 281.

Monlong, J., Cossette, P., Meloche, C., Rouleau, G., Girard, S.L., and Bourque, G. (2018). Human copy number variants are enriched in regions of low mappability. Nucleic Acid. Res. *46*, 7236–7249.

Neveling, K., Mantere, T., Vermeulen, S., Oorsprong, M., van Beek, R., Kater-Baats, E., Pauper, M., van der Zande, G., Smeets, D., Weghuis, D.O., et al. (2021). Next-generation cytogenetics: comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping. Am. J. Hum. Genet. *108*, 1423–1435.

Nikoloudis, D., Kountouras, D., and Hiona, A. (2020). The frequency of combined IFITM3 haplotype involving the reference alleles of both rs12252 and rs34481144 is in line with COVID-19 standardized mortality ratio of ethnic groups in England. PeerJ *12*, e10402.

Noels, H., Theelen, W., Sternkopf, M., Jankowski, V., Moellmann, J., Kraemer, S., Lehrke, M., Marx, N., Martin, L., Marx, G., et al. (2018). Reduced post-operative DPP4 activity associated with worse patient outcome after cardiac surgery. Sci. Rep. *8*, 1–2.

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2021). Genetic mechanisms of critical illness in Covid-19. Nature *591*, 92–98. https://doi.org/10.1038/s41586-020-03065-y.

Pasquier, C., and Robichon, A. (2020a). Computational search of hybrid human/SARS-CoV-2 dsRNA reveals unique viral sequences that diverge from other coronavirus strains. bioRxiv. https://doi.org/10.1101/2020.04.08.031856.

Pasquier, C., and Robichon, A. (2020b). SARS-CoV-2 might manipulate against its host the immunity RNAi/Dicer/Ago system. bioRxiv. https://doi.org/10.1101/2020.04.08.031856.

Picchiotti, N., Benetti, E., Fallerini, C., Daga, S., Baldassarri, M., Fava, F., Zguro, K., Valentino, F., Doddato, G., Giliberti, A., et al. (2021). Post-Mendelian genetic model in COVID-19. medRxiv. https://doi.org/10.1101/2021.01.27.21250593.

Pociask, D.A., Robinson, K.M., Chen, K., McHugh, K.J., Clay, M.E., Huang, G.T., Benos, P.V., Janssen-Heininger, Y.M., Kolls, J.K., Anathy, V., and Alcorn, J.F. (2017). Epigenetic and transcriptomic regulation of lung repair during recovery from influenza infection. Am. J. Pathol. *187*, 851–863.

Reinhold, D., and Brocke, S. (2014). DPP4-directed therapeutic strategies for MERS-CoV. Lancet Infect. Dis. *14*, 100–101.

Richardson, S., Hirsch, J.S., Narasimhan, M., Crawford, J.M., McGinn, T., Davidson, K.W., Barnaby, D.P., Becker, L.B., Chelico, J.D., Cohen, S.L., et al. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. JAMA *323*, 2052–2059.

Sada, K.E., Wada, J., Morinaga, H., Tuchimochi, S., Uka, M., and Makino, H. (2014). Sarcoid-like lung granulomas in a hemodialysis patient treated with a dipeptidyl peptidase-4 inhibitor. Clin. Kidney J. *7*, 182–185.

Schlicht, K., Rohmann, N., Geisler, C., Hollstein, T., Knappe, C., Hartmann, K., Schwarz, J., Tran, F., Schunk, D., Junker, R., and Bahmer, T. (2020). Circulating levels of soluble Dipeptidylpeptidase-4 are reduced in human subjects hospitalized for severe COVID-19 infections. Int. J. Obes. *44*, 2335–2338.

Smahi, A., Courtois, G., Rabia, S.H., Döffinger, R., Bodemer, C., Munnich, A., Casanova, J.L., and Israël, A. (2002). The NF-kappaB signalling pathway in human diseases: from incontinentia pigmenti to ectodermal dysplasias and immune-deficiency syndromes. Hum. Mol. Genet. *11*, 2371–2375.

Sohn, K.M., Lee, S.G., Kim, H.J., Cheon, S., Jeong, H., Lee, J., Kim, I.S., Silwal, P., Kim, Y.J., Paik, S., and Chung, C. (2020). COVID-19 patients upregulate toll-like receptor 4-mediated inflammatory signaling that mimics bacterial sepsis. J. Korean Med. Sci. *35*, e343.

Sun, C.K., Leu, S., Sheu, J.J., Tsai, T.H., Sung, H.C., Chen, Y.L., Chung, S.Y., Ko, S.F., Chang, H.W., and Yip, H.K. (2013). Paradoxical impairment of angiogenesis, endothelial function and circulating number of endothelial progenitor cells in DPP4-deficient rat after critical limb ischemia. Stem Cell Res. Ther. *4*, 1–7.

Tanaka, Y., Soda, H., Fukuda, Y., Nio, K., Ono, S., Tomono, H., Shimada, M., Yoshida, M., Harada, T., Umemura, A., and Iwasaki, K. (2020). Vildagliptin-induced ground-glass nodules mimicking lung metastases in a cancer patient receiving Lactobacillus probiotic supplementation. Thorac. Cancer *11*, 470–474.

Trzaskalski, N.A., Fadzeyeva, E., and Mulvihill, E.E. (2020). Dipeptidyl peptidase-4 at the interface between inflammation and metabolism. Clin. Med. Insights. Endocrinol. Diabetes *13*, 1179551420912972.

Van der Made, C.I., Simons, A., Schuurs-Hoeijmakers, J., van den Heuvel, G., Mantere, T., Kersten, S., van Deuren, R.C., Steehouwer, M., van Reijmersdal, S.V., Jaeger, M., et al. (2020). Presence of genetic variants among young men with severe COVID-19. JAMA *324*, 1–11.

Vitallé, J., Terrén, I., Orrantia, A., Zenarruzabeitia, O., and Borrego, F. (2019). CD300 receptor family in viral infections. Eur. J. Immunol. *49*, 364–374.

Wang, X., Wu, Z., Li, Y., Yang, Y., Xiao, C., Liu, X., Xiang, X., Wei, J., Shao, D., Liu, K., and Deng, X. (2020). p53 promotes ZDHHC1-mediated IFITM3 palmitoylation to inhibit Japanese encephalitis virus replication. PLoS Pathog. *16*, e1009035.

Watson, C.T., Steinberg, K.M., Huddleston, J., Warren, R.L., Malig, M., Schein, J., Willsey, A.J., Joy, J.B., Scott, J.K., Graves, T.A., et al. (2013). Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. Am. J. Hum. Genet. *92*, 530–546.

Williamson, E.J., Walker, A.J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C.E., Curtis, H.J., Mehrkar, A., Evans, D., Inglesby, P., and Cockburn, J. (2020). Factors associated with COVID-19-related death using OpenSAFELY. Nature *584*, 430–436.

Wohlfart, S., Söder, S., Smahi, A., and Schneider, H. (2016). A novel missense mutation in the gene EDARADD associated with an unusual phenotype of hypohidrotic ectodermal dysplasia. Am. J. Med. Genet. A *170A*, 249–253.

Wu, Z., and McGoogan, J.M. (2020). Outbreak in China: summary of a report of 72314 cases from the Chinese center for disease control and prevention. JAMA *323*, 1239.

Xu, X., Li, M., Wu, Z., Wang, H., Wang, L., Huang, K., Liu, X., Hou, Q., Lin, G., and Hu, C. (2017). Endoplasmic reticulum transmembrane proteins ZDHHC1 and STING both act as direct adaptors for IRF3 activation in teleost. J. Immunol. *199*, 3623–3633.

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I.K.D., Hodeib, S., Korol, C., et al. (2020). Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science *370*, eabd4570.

Zhao, F., Fan, Q., Zhang, W., Li, B., and Zhang, J. (2020a). Association between ABO blood group system and COVID-19 susceptibility in Wuhan. Front. Cell. Infect. Microbiol. *10*, 404.

Zhao, J., Yang, Y., Huang, H.P., Li, D., Gu, D.F., Lu, X.F., Zhang, Z., Liu, L., Liu, T., Liu, Y.K., et al. (2020b). Relationship between the ABO blood group and the COVID-19 susceptibility. Clin. Infect. Dis. *73*, 328–331.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet *395*, 1054–1062.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in China. N. Engl. J. Med. *382*, 727–733.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Critical commercial assays | | |
| Bionano Genomics DNA Isolation Kit | Bionano Genomics | 90057 |
| DLS Labeling Kit | Bionano Genomics | 80005 |
| Saphyr Chip G2.3 (FC) | Bionano Genomics | 20366 |
| Qubit® BR (Broad Range) dsDNA Assay Kit | Thermo Fisher | Q32850 |
| Qubit® HS (High Sensitivity) dsDNA Assay Kit | Thermo Fisher | Q32851 |
| HemoCue Microcuvettes | Fisher Scientific | 22-601-017 |
| HemoCue WBC Analyzer | Fisher Scientific | 22-601-018 |
| Fluorimeter, Qubit | Thermo Fisher | Q33226 |
| sDPP4 ELISA Assay | ThermoFisher Scientific | EHDPP4 |
| Deposited data | | |
| All suuplimentary files | Mendeley Data | doi: https://doi.org/10.17632/fmvxxbdtpz.1 |
| Oligonucleotides | | |
| Primers for dosage and expression analysis (supplemental information) | Integrated DNA Technologies | NA |
| Software and algorithms | | |
| Bionano Access (v.1.5)/Bionano Solve (v.3.5) software | Bionano Genomics | NA |
| Other | | |
| Saphyr Instrument, ICS PC & Monitor, Access Server, Accessory Kit | Bionano Genomics | 90001 |
| QuantStudio 3 real-time cycler | Thermo Fisher | A28136 |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ravindra Kolhe (rkolhe@augusta.edu).

#### Material availability

This study did not generate new unique reagents.

#### Data and code availability

Additional supplemental items are available from Mendeley Data: https://doi.org/10.17632/fmvxxbdtpz.1. DOI is also included in key resources table.

Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Study participants

During the period from March 2020 to February 2021, 52 severely ill patients were identified for this study. The criteria for inclusion included patients with a confirmed SARS-CoV-2 RT-PCR test (from nasopharyngeal swabs or other biological fluids), and a requirement for mechanical ventilation or a fraction of inspired oxygen (FiO2) of at least 60% or more. The samples were collected under an approved HAC by the IRB Committee A (IRB REGISTRATION # 1597188-2), Augusta University, GA. Based on the IRB approval, all PHI was removed, and all data was anonymized before accessing for the study.

## METHOD DETAILS

### Optical genome mapping

Peripheral blood from critically ill COVID-19 patients was drawn and used to isolate ultra-high molecular weight (UHMW) DNA prepared for analysis on the Bionano Genomics Saphyr® platform (Bionano Genomics, San Diego, USA, #90001) following the manufacturer's protocols (Bionano Genomics, San Diego, USA, #90057)). Briefly, a frozen blood aliquot (650µl) was thawed and cells were counted using HemoCue (Microcuvettes and WBC analyzer, HemoCue Holding AB, Ängelholm, Sweden). Subsequently, a blood aliquot comprising approximately 1.5 million nucleated white blood cells was centrifuged and digested with Proteinase K. DNA was precipitated using isopropanol and washed using buffers (buffer A and B), while the DNA remained adhered to the nanobind magnetic disk. The UHMW bound DNA was resuspended in elution buffer and quantified with Qubit dsDNA BR assay kits (ThermoFisher Scientific, San Francisco, USA, #Q32850) using Qubit fluorimeter (ThermoFisher Scientific, San Francisco, USA, #Q33226).

DNA labeling was performed following manufacturer's protocols (Bionano Genomics, USA, #80005) in which Direct Labeling Enzyme 1 (DLE-1) reactions to a specific 6-base sequence motif (CTTAAG) in the genome were carried out using 750 ng of purified high molecular weight DNA. Labeled DNA was quantified using Qubit dsDNA HS assay kits (ThermoFisher Scientific, San Francisco, USA, #Q32851), and loaded onto flow cells of Saphyr chips (Bionano Genomics, USA, #20366) for optical imaging. The fluorescently labeled DNA molecules were imaged on the Saphyr platform after the labeled DNA molecules were electrophoretically linearized in the nanochannel arrays. Effective genome coverage of approximately 100X was achieved for tested samples after evaluating the molecule quality metrics. The quality control metrics for each sample achieved the recommended molecule map rates of greater than 70% and molecule N50 values greater than 250kb in size.

Genome analyses were performed using Bionano Access (v.1.5)/Bionano Solve (v.3.5) software, a *de novo* assembly analysis was performed on all the samples to assess and interrogate all germline SVs. Briefly, molecules of a given sample dataset were first *de novo* assembled into consensus genome maps, the genome maps were aligned to the GRCH37 reference human genome assembly. SVs were identified where *de novo* assemblies differed from the GRCH37 reference genome, insertion, duplications, deletions, inversions, and translocations could be called based on this alignment. SVs generated by the *de novo* assembly pipeline were then annotated with known canonical gene sets extracted from the reference genome assembly and compared to a control dataset to estimate the population frequency of SVs.

### Data analysis

Rare/unique SVs (>500bp) were investigated to determine candidate gene/loci as potentially strong predisposing factors associated with severe COVID-19. Unbiased whole-genome analysis of SVs was performed to identify unique/rare genic SVs in these patients that did not appear in the population dataset (Bionano controls comprising 267 individuals). Additionally, only SVs disrupting the coding region(s) of the gene(s) were selected and reviewed for relevance to response to COVID-19 infection. The Bionano control dataset includes 267 healthy individuals, of which 45 were African, 16 were Admixed American, 17 were East Asian, 44 were European, 15 were South Asian and 180 were of unknown origin (Levy-Sakin et al., 2019; Izzy et al., 2020; Ebert et al., 2021). The frequency of the SVs were also determined using the public database (gnomAD) (Figure 1).

### Quantitative PCR (qPCR) confirmation of OGM SVs

Quantitative PCR (qPCR) was used to confirm the selected SVs (copy number gains and losses). Relative quantitation of genomic dosage was determined using the QuantStudio 3 real-time cycler (Thermo Fisher Scientific, San Francisco, USA, # A28136) and calculated by the relative threshold cycle (ddCt) method (Livak and Schmittgen, 2001). PCR amplicons were generated in triplicate for each individual and four controls. Genomic dosage for the selected SVs (*EDARADD, DPP4, ZDHHC1, KRT15, CD300A, ZNF443, STK26,*) were determined by SYBR green incorporation using RNAseP as the reference (BioRad, USA). Three primer pairs were designed for each SV using the Primer 3.0 software (List of primer sequences and targets is available in supplemental information). Relative genomic dosage was calculated as $2^{-\Delta\Delta CT}$ where $\Delta CT$=(mean $Ct_{Target}$)-(mean $Ct_{Reference}$) and $\Delta\Delta CT = \Delta CT_{patient} - \Delta CT_{control}$.

### Independent assessment of Saphyr SVs by FaNDOM

To provide an independent assessment of the SVs called by the Bionano Solve pipelines developed by Bionano Genomics, we used a novel method, FaNDOM (Fast Nested Distance-based seeding of Optical Maps) (https://github.com/jluebeck/FaNDOM) (Dehkordi et al., 2021). FaNDOM maps optical map fragments to an *in silico* digested reference genome, relying on a fast filtering strategy to provide an order of magnitude speedup while still maintaining high sensitivity over 95% for concordant reads, and 79% for SVs on the benchmark NA12878 human genome. FaNDOM was used on the COVID-19 patient data to call breakpoints (two distinct reference positions that are adjacent in the donor), multi-chromosomal translocation, and insertion events. SVs were called on breakpoints based on the orientation of the fragments mapping to the breakpoint, with a (+,+) orientation suggestive of a deletion, (-,-) orientation suggestive of a tandem duplication, and opposing orientation suggesting an inversion or translocation. The SVs called by FaNDOM were used to independently confirm SVs called by Bionano Access Software.

### Expression analysis

Total RNA was isolated from the peripheral blood of 15 asymptomatic patients [7 male, 8 female, age range 25-59 years [mean 40.8 $\pm$ 11.5 (SD)] and 13 severely ill COVID-19 patients [6 male and 7 female, age range 19-81 years [mean 60.9 $\pm$ 15.1 (SD)], using mRNeasy mini kit (QIAGEN, Germany). Peripheral blood was collected from the asymptomatic controls during active infection (positive for SARS-CoV-2 by RT-PCR at time of blood collection), and during ICU treatment for the severely ill patients. The quantity of total RNA from the samples was determined by an ultraviolet spectrophotometer (Nanodrop, Thermo Fisher Scientific, and Pittsburgh, PA). Total RNA (500 ng) was reverse transcribed using the iScript cDNA synthesis kit (170–8891) from BioRad Laboratories (Hercules, CA). Quantitative real-time PCR (q-RT-PCR) was performed using gene-specific primers, and an SYBR Green assay on the QuantStudio 3 system (Thermo Fisher Scientific, CA) (supplemental information). The specific products were confirmed by SYBR green single melt curve analysis. The results were normalized to the expression of the 18S rRNA housekeeping gene and the relative fold change was calculated using the delta-delta Ct method.

### Soluble DPP4 (sDPP4) measurements

sDPP4 serum concentrations were measured in 8 asymptomatic and severely ill patients (4 male, 4 female) in each group, respectively, using the Enzyme-linked immunosorbent assay (ELISA) (ThermoFisher Scientific, USA, Cat# EHDPP4) following the manufacturer's instructions.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses was performed for qPCR expression and sDPP4 serum concentration experiments. Comparison between groups were made using unpaired T-test and a p value $<0.05$ was considered statistically significant.