

Host genome analysis of structural variations by Optical Genome Mapping provides clinically valuable insights into genes implicated in critical immune, viral infection, and viral replication pathways in patients with severe COVID-19.

Nikhil Shri Sahajpal¹, Chi-Yu Jill Lai², Alex Hastie², Ashis K Mondal¹, Siavash Raeisi Dehkordi³, Cas van der Made⁴, Olivier Fedrigo⁵, Farooq Al-Ajli⁵, Sawan Jalnapurkar⁶, Rashmi Kanagal-Shamanna⁷, Brynn Levy⁸, Silviu-Alin Bacanu⁹, Michael C Zody¹⁰, Catherine A. Brownstein¹¹, Aryn M. Rojiani¹, Alan H. Beggs¹¹, Vineet Bafna³, Alexander Hoischen⁴, Erich D. Jarvis^{5,12,13}, Alka Chaubey^{1,2}, Ravindra Kolhe^{1*} and the COVID19hostgenomesv consortium[†].

¹ Department of Pathology, Medical College of Georgia, Augusta University, GA, U.S.A.

² Bionano Genomics, Inc., San Diego, CA, U.S.A.

³ Department of Computer Science and Engineering, University of California at San Diego, CA, U.S.A.

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

⁵ Vertebrate Genome Lab, The Rockefeller University, New York, NY, USA.

⁶ Department of Medicine, Medical College of Georgia, Augusta University, GA, U.S.A

⁷ Hematopathology, University of Texas M.D. Anderson Cancer Center, Houston, U.S.A

⁸ Department of Pathology and Cell Biology, Columbia University Medical Center, New York, U.S.A

⁹ Department of Psychiatry, Virginia Commonwealth University, VA, U.S.A

¹⁰ New York Genome Center, New York, NY 10013, U.S.A.

¹¹ Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Harvard Medical School, Boston, MA, U.S.A

¹² Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY, USA

¹³ Howard Hughes Medical Institute, Chevy Chase, MD, USA

†COVID19hostgenomesv consortium

<https://www.covid19hostgenomesv.org/consortium.html>

***Corresponding author**

Ravindra Kolhe MD, PhD.

Medical College of Georgia| Augusta University.

1120 15th Street| Augusta, GA 30912, BF-207

Email: rkolhe@augusta.edu

P: (706)-721-2771, F: (706)-434-6053

rkolhe@augusta.edu

Abstract

Background: The varied clinical manifestations and outcomes in patients with SARS-CoV-2 infections implicate a role of host-genetics in the predisposition to disease severity. This is supported by evidence that is now emerging, where initial reports identify common risk factors and rare genetic variants associated with high risk for severe/ life-threatening COVID-19. Impressive global efforts have focused on either identifying common genetic factors utilizing short-read sequencing data in Genome-Wide Association Studies (GWAS) or whole-exome and genome studies to interrogate the human genome at the level of detecting single nucleotide variants (SNVs) and short indels. However, these studies lack the sensitivity to accurately detect several classes of variants, especially large structural variants (SVs) including copy number variants (CNVs), which account for a substantial proportion of variation among individuals. Thus, we investigated the host genomes of individuals with severe/life-threatening COVID-19 at the level of large SVs (500bp-Mb level) to identify events that might provide insight into the inter-individual clinical variability in clinical course and outcomes of COVID-19 patients.

Methods: Optical genome mapping using Bionano's Saphyr® system was performed on thirty-seven severely ill COVID-19 patients admitted to intensive care units (ICU). To extract candidate SVs, three distinct analyses were undertaken. First, an unbiased whole-genome analysis of SVs was performed to identify rare/unique genic SVs in these patients that did not appear in population datasets to determine candidate loci as decisive predisposing factors associated with severe COVID-19. Second, common SVs with a population frequency filter was interrogated for possible association with severe COVID-19 based on literature surveys. Third, genome-wide SV enrichment in severely ill patients versus the general population was investigated by calculating odds ratios to identify top-ranked genes/loci. Candidate SVs were confirmed using qPCR and an independent bioinformatics tool (FaNDOM).

Results: Our patient-centric investigation identified 11 SVs involving 38 genes implicated in three key host-viral interaction pathways: (1) innate immunity and inflammatory response, (2) airway resistance to pathogens, and (3) viral replication, spread, and RNA editing. These included seven rare/unique SVs (not present in the control dataset), identified in 24.3% (9/37) of patients, impacting up to 31 genes, of which *STK26* and *DPP4* are the most promising candidates. A duplication partially overlapping *STK26* was corroborated with data showing upregulation of this gene in severely ill patients. Further, using a population frequency filter of less than 20% in the Bionano control dataset, four SVs involving seven genes were identified in 56.7% (21/37) of patients.

Conclusion: This study is the first to systematically assess and highlight SVs' potential role in the pathogenesis of COVID-19 severity. The genes implicated here identify novel SVs, especially *STK26*, and extend previous reports involving innate immunity and type I interferon response in the pathogenesis of COVID-19. Our study also shows that optical genome mapping can be a powerful tool to identify large SVs impacting disease outcomes with split survival and add valuable genomic information to the existing sequencing-based technology databases to understand the inter-individual variability associated with SARS-CoV-2 infections and COVID-19 mortality.

Introduction

The emergence of COVID-19 in the city of Wuhan, China in December 2019 led to an ongoing global pandemic. Since then, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected more than 80,721,623 individuals across the globe, with at least 1,763,690 COVID-19 related deaths (<https://coronavirus.jhu.edu/map.html>, last accessed December 27, 2020). The clinical manifestations of SARS-CoV-2 infected patients vary from asymptomatic or mild, including low-grade fever and flu-like symptoms, to severe symptoms, including acute respiratory distress syndrome (ARDS), pneumonia and death [1-6]. Clinical studies have associated advanced age, male gender, hypertension, diabetes, and other obesity-related diseases as risk factors predisposing to COVID-19 related severe illness [7-9]. However, variable clinical manifestations within these sub-sets and poor clinical outcome in individuals with no associated co-morbidities or risk factors clearly implicate the role of host genetics in the predisposition of SARS-CoV-2 infected individuals to disease severity [10-11]. A recent meta-analysis on host genetic factors in coronavirus infection by LoPresti et al. [12] 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 non-human host genetic factors [12].

In principle, there have been two distinct approaches to investigate the host genome to identify genetic loci associated with disease susceptibility and severity in COVID-19 patients. The first approach utilizes relatively unbiased genome-wide association studies (GWAS) to understand risk factors at the population level, which provides insights into the pathophysiological mechanisms and biology, but rarely at an individual level. The second approach involves short-read 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 that blood group A of the ABO groups is an independent risk factor for COVID-19 related disease susceptibility [13, 14]. 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 [15, 16]. Reports utilizing a targeted variant approach demonstrated the importance of rare/unique events that cause immunodeficiencies and predispose 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 [17]. Further, the 13 genes in the TLR3/IRF3 pathway have been implicated in patients with severe COVID-19 [18, 19].

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 around the world [20]. Although these studies have identified certain genetic loci

associated with disease severity, they remain limited to nucleotide variants, which explains only a small portion of the heritability of complex traits. Importantly, structural variants (SVs) are beyond the purview of these studies as a result of technical limitations, and they comprise a substantial proportion of genomic variation among individuals, which can drive evolutionary processes [21-23]. SVs are diverse, and include large copy-number changes, insertions, deletions, inversions, and translocations. SVs involve larger regions of an individual genome (500 bp and up) than small variants (up to several hundred basepairs). As most of these are not detected by routine short-read next generation sequencing and current analytical pipelines these categories of SVs remain undocumented 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 variation are larger than a single base-pair substitution, i.e. single nucleotide variation (SNVs) [20]. Several studies have demonstrated the importance of large SVs in the characterization of human immunity profiles [24-26]. 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 [26-28].

To address this issue, we formed the COVID-19 Host Genome Structural Variation Consortium, and performed a preliminary study aimed at evaluating optical genome mapping to identify SVs in an unbiased fashion in severely ill patients with COVID-19. 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.

Materials and Methods

Study Participants:

A total of 37 severely ill COVID-19 patients defined as those admitted to the intensive care unit (ICU) who required mechanical ventilation or had a fraction of inspired oxygen (FiO₂) of at least 60% or more, and with a confirmed SARS-CoV-2 RT-PCR test (from nasopharyngeal swabs or other biological fluids) were included in the study. 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.

Optical Genome Mapping:

Peripheral blood from critically ill COVID-19 patients was used to isolate ultra-high molecular weight (UHMW) DNA following the manufacturer's protocols (Bionano Genomics, San Diego, USA). Briefly, a frozen blood aliquot (650µl) was thawed and cells were counted using HemoCue (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 using Qubit dsDNA assay kits (ThermoFisher Scientific, San Francisco, USA).

DNA labeling was performed at the specific 6-base sequence motifs following manufacturer's protocols (Bionano Genomics, USA) in which Direct Labeling Enzyme 1 (DLE-1) reactions to a known repeated sequence in the genome were carried out using 750 ng of purified high molecular weight DNA. Labeled DNA was loaded on to flow cells of Saphyr chips for optical imaging. The fluorescently labeled DNA molecules were imaged after the labelled DNA molecules were electrophoretically linearized, using the Bionano Genomics Saphyr® platform. Effective genome coverage of approximately 100X was achieved for tested samples after evaluating the molecule quality metrics. The quality control metric for each sample achieved the recommended molecule map rates of greater than 70% and molecule N50 values greater than 250kb.

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 germ line SVs. Briefly, molecules of a given sample dataset were first *de novo* assembled into consensus genome maps, the genome maps were aligned to the hg19 reference human genome assembly. SVs were identified where *de novo* assemblies differed from the hg19 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:

To identify SVs associated with disease severity, three distinct analyses were undertaken. **In the first analysis**, rare/unique SVs were investigated to determine candidate gene/loci as potentially strong predisposing factors associated with severe COVID-19. An 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. Additionally, only the SVs disrupting the coding region(s) of the gene(s) were selected and reviewed for relevance in the COVID-19 infection. **In the second analysis**, common SVs were interrogated for possible association with severe COVID-19 based on a literature survey. A thorough literature review was performed with the combination of the following key words that included “SARS”, “Viral infection”, “COVID-19”, “host genetics”, “host susceptibility”, “immune response”, and “immune genes”. The literature included electronic, ahead-of-print, and pre-print articles which investigated DNA-based genetic variations, proteomics, transcriptomic, and bio-informatics modelling to identify candidate pathways/ genes. A gene list comprising of 881 genes was manually curated after searching the articles (last accessed November 15, 2020) for relevance pertaining to a general method, genes, variants, statistical results, associations, and p-value (**Supplementary file 1**). The SVs overlapping the genes of interest in severely ill patients were compared with a population dataset [Bionano Genomics (Bionano) controls] with <20% frequency filter. The SVs were further interrogated against the African American ethnicity in the Bionano control dataset [29-31]. Subsequently, only variants

disrupting the coding regions of genes of interest were selected for further analysis. **In the third analysis**, genome-wide SV enrichment in severely ill patients versus a general population cohort (Bionano controls comprising 267 individuals, including 150 assayed with DLE-1 enzyme) was investigated by calculating odds ratio to identify top ranked genes/loci, with the following filtering criteria that includes, SVs observed in <20% of Bionano controls. The odds ratio of a gene was considered equal to the proportion of total severely ill patients having at least a SV on the gene, divided by proportion of total Bionano controls (150 individuals labeled with DLE-1 enzyme; **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) and calculated by the relative threshold cycle (ddCt) method [32]. PCR amplicons were generated in triplicate for each individual and four controls. Genomic dosage for the selected genes (*EDARADD*, *APOBEC3B*, *ZNF443*, *CCL4L2*) was 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 **Supplementary file 2**). Relative genomic dosage was calculated as $2^{-\Delta\Delta CT}$ where $\Delta CT = (\text{mean } Ct_{\text{Target}}) - (\text{mean } Ct_{\text{Reference}})$ and $\Delta\Delta CT = \Delta CT_{\text{patient}} - \Delta CT_{\text{control}}$.

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>). 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 in excess of 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 peripheral blood of 11 asymptomatic patients [(3 Male, 8 female) age range 25-59 years [43.2 (mean) \pm 11.3 (Stdev)] and 12 severely ill COVID-19 patients [(6 Male and 6 Female) age range 18-81 years [(61.8 (mean) \pm 15.4 (Stdev)], 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 the total RNA from the samples was

determined by 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 a SYBR Green assay on the QuantStudio 3 system (Thermo Fisher Scientific, CA). 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 delta-delta Ct method.

Results

Patient Characteristics

During the period from March 30 to August 13, 2020, 37 severely ill patients were identified for this study. The criteria for inclusion included 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 demographics and clinical characteristics of the patients (**Table 1**) were a mean age of 61.6 ±12.4 SD (range 19-83), 49% (18/37) female, and 70.2% (26/37) African American, 27% (10/37) Caucasian, and one Hispanic, which is similar to the demographics of COVID-19 patients in the State of Georgia, USA [33]. Chronic preexisting medical conditions in these patients included diabetes (43.2%), hypertension (59%), chronic kidney disease (24%), and asthma (5.4%). Four patients had no known comorbidities at the time of admission to the ICU. On ICU admission, 30 patients needed mechanical ventilation, with mean intubation duration of 12 ± 5.5 days, of which, 35% (13/37) were intubated prior to admission (PTA), and 16% (6/37) were ventilated in prone-position. Of the 37 patients, 25 recovered and 12 died.

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. The complete list of rare/unique SVs is provided a **Supplementary file 3**. Private SVs disrupting the gene coding regions led to the identification of seven rare/unique SVs, possibly impacting 31 genes, identified in 24.3% (9/37) patients. These included: a ~162.2 kb duplication of chr 1 (chr1:236604233-236766495) partially disrupting the *EDARADD* (exon 5 and 6; NM_145861.3) and *HEATR1* (exons 3-45; NM_018072.6) genes in two patients (cases 22 and 26); a ~24.06 kb heterozygous deletion of chr 2 (chr2:162887379-162911439) partially deleting the *DPP4* gene (exons 3 and 4; NM_001935.4) in one patient (case 38); a ~146.8 kb heterozygous deletion of chr 16 (chr16:67308198-67455019) that contained six genes (*PLEKHG4*, *KCTD19*, *LRRC36*, *TPPP3*, *U1*, *ZDHC1*) was identified in one patient (case 39); a ~ 28.4 kb duplication of chr 17 (chr17:39662399-39690882) including two genes (*KRT15* and *KRT19*) in one patient (case 2); a ~833 kb **tandem** duplication of chr 17 (chr17:71844581-72678517) including 15 genes (*RPL38*, *MGC16275*, *TTYH2*, *Z49982*, *DNAI2*, *CD300E*, *CD300LD*, *CD300C*, *CD300LB*, *CD300A*, *GPRC5C*, *GPR142*, *BTBD17*, *KIF19*), and partially disrupting *RAB37* (exon 1; NM_175738.5) in one patient (case 19); a ~39.8 kb duplication of chr 19 (chr19:12512276- 12552113) including the entire *ZNF443* gene (NM_005815.5) in two

patients (cases 13 and 18); and a ~ 534.9 kb tandem duplication of chr X (chrX: 130629618-131164603) including four genes (*OR13H1*, *LOC286467*, *5S_rRNA*, *STK26*) in one patient (case 44) (**Table 2**, **Figure 2a-g**). All these SVs were confirmed by qPCR (**Supplementary file 4**).

Common SVs in candidate loci previously associated with severe COVID-19

We next searched for common SVs that overlap with previously suggested candidate loci/genes implicated in COVID-19. This literature-based approach involved investigating SVs overlapping coding regions of the 881 genes identified in literature survey. These SVs were further filtered by considering only those at frequency of less than 20% in Bionano control data sets (267 individuals). This filtering resulted in identification of four SVs impacting seven genes in 56.7% (21/37) severely ill patients. Based on the OGM data and the minimum genomic coordinates identified by the Bionano Access software, the four SVs identified include: a ~248.1 kb heterozygous inversion of chr 3 (chr3:195725611-195477445) disrupting the *MUC4* gene (exons 1-23; NM_018406.7) in five patients of African American (AA) ethnicity (cases- 2, 3, 8, 26, 29); A ~29.8 kb duplication of chr 17 (chr17:34437711- 34571682) including 3 genes (*TBC1D3B*, *CCL3L1*, *CCL4L2*) in ten patient (cases 12, 20, 28, 29, 6, 8, 13, 19, 21, 22); a ~28.9 kb heterozygous deletion of chr 22 (chr22:39347202-39391164) disrupting 2 genes (*APOBEC3B* and *APOBEC3A*) in three patients; and a ~1.5 kb insertion in chr 21 (chr21:42784603-42793518) overlapping *MXI* gene (exon 1; NM_001144925.2) in 11 patient (cases 13, 17, 18, 20, 21, 24, 25, 28, 4, 8, 38) (**Table 3**, **Figure 3a-d**). All SVs, except the *MUC4* inversions and *MXI* insertion were confirmed by qPCR (**Supplementary file 5**).

Genome-wide SV enrichment in severely ill vs. population

The odds ratio was calculated comparing SVs identified in the 37 severely ill patients with the population dataset (150 individuals). The data was enriched by directly filtering with numbers of cases on each gene (greater than 2 in the severely ill group) and filtering for SVs that appear in less than 20% frequency in the Bionano controls. The top three gene were found to be *TRPV1*, *MIR4710*, and *SUMO3*, with a total of 242 genes enriched with log odds ratio greater than three (**Supplementary file 6**).

Confirmation of Saphyr SVs by FaNDOM and qPCR

FaNDOM confirmed all the 11 SV calls (seven rare/unique + four common) made by Bionano Access Software. Of the 11 SVs, nine events (**Table 2**: all events, **Table 3**: events 2 and 3) were confirmed by qPCR.

Expression Analysis

To evaluate the impact of X-chromosomal duplication involving the *STK26* gene, *STK26* transcripts were quantified from blood of 11 asymptomatic and 12 severely ill patients with COVID-19, normalized against 18S rRNA. *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 (11.4 fold) (**Figure 4**), who was the one patient

that harbored the SV partially duplicating the 5'UTR and coding exon 1 of the *STK26* gene (**Figure 2**).

Discussion

This study addressed the current technology bias of short read sequencing that has confined most host genome studies to interrogation of only SNVs and small indels associated with COVID-19 severity [13-18]. Although these approaches have been successful in associating certain genetic loci, but large gaps remain in explaining the wide diversity of clinical responses. We aimed to investigate large SVs to identify events that might further explain the inter-individual clinical variability in response to COVID-19, using optical genome mapping. Among 37 patients with severe COVID-19, we found seven rare/private SVs (not found in 267 Bionano controls) potentially impacting 31 genes, and four common SVs impacting seven genes. While individual effect of any of these SVs remains difficult to predict, several of these may be considered candidate loci as strongly predisposing factors associated with severe COVID-19. Altogether, these SVs were implicated in three key host-viral interaction pathways: 1) innate immunity and inflammatory response (*EDARADD*, *DPP4*, *ZDHHC1*, *STK26*, *MX1*, *TBC1D3B*, *CCL3L1*, *CCL4L2*, *CD300* gene cluster); 2) airway resistance to pathogens (airway mucous secretion - *MUC4*); and 3) viral replication, spread and RNA editing (*KRT15*, *KRT19*, *ZNF443*, *APOBEC3A*, *APOBEC3B*) (**Figure 5**).

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. The variability in the host innate and adaptive immune responses to viral infections influence clinical manifestations and outcomes [34, 35]. 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 [36, 37]. Independent studies have identified rare genetic variants predicted to be loss of function (LOF) at 13 loci governing TLR3 and TLR7 pathway genes and affecting the type I interferon response [17, 18]. In addition, several groups have identified impaired type I interferon response in severe COVID-19 pathogenesis [19, 38]. 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 1 region of the gene. In synthetic biology, dual 5'UTR constructs have shown increased transcription and translation of the gene [39]. 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. *STK26* encodes for serine/threonine kinase 26, a negative regulator of the TLR signaling [40], 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 [41]. TRAF6 is a downstream signaling molecule in the TLR3/TLR7 pathways that mediates the NF- κ B and type I interferon response [17, 42]. It is

intriguing to speculate that the overexpression of *STK26* in this patient may inhibit the TLR signaling, and via this route impairing the type I interferon response leading to reduced viral clearance in the early phase of the infection and therefore increasing the COVID-19 severity. The upregulation of *STK26* detected coincidentally in all severe COVID-19 patients implicates this gene/pathway in the pathogenesis of severe COVID-19.

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* has been shown to mount a type I interferon response against DNA and RNA viruses via different downstream players [43-45]. *ZDHHC1* is an endoplasmic reticulum-associated protein that mediates MITA-dependent IRF3 activation and type I interferon response against DNA viruses (Figure 5) [43, 44]. 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 [43-45]. Knockdown of *ZDHHC1* leads to compromised *ZDHHC1*/IFITM3 anti-viral activity against Japanese encephalitis virus [45]. Additionally, IFITM3 expression was found to be upregulated early in epithelial lung cells in response to SARS-CoV-2, [46] and that SNPs in IFITM3 are associated with more severe disease [47]. We hypothesize that the heterozygous deletion of the *ZDHHC1* gene could possibly 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 *DPP4* gene in a severely ill patient (case 38) is consistent with prior studies implicating DPP4 (protein) in MERS-CoV [48-50] and severe COVID-19 [51]. *DPP4* exists in two forms: 1) a membrane bound glycoprotein receptor; and 2) a cleaved product of the receptor (sDPP4) that lacks the intracellular and transmembrane domain present in the circulation (Figure 5) [52]. The ~24.06 kb heterozygous deletion on chr2 is an intragenic copy number loss of the *DPP4* gene (exon3 and 4). Exons 3 and 4 encode the flexible stalk and α/β hydrolase domain of the receptor. The flexible stalk serves as the cleavage site recognized by matrix metalloproteinase cleavage enzymes [53]. The deletion of the two domains is hypothesized to lead to loss of function and significantly low sDPP4 levels. Although bioinformatic models predicted human DPP4-SARS-CoV-2 protein interactions and protein docking studies show SARS-CoV-2 virus with DPP4 as a co-receptor (not as strong as with ACE-2) [54], *in vitro* studies found that DPP4 serves as a viral entry receptor for MERS-CoV, and not for coronaviruses including SARS-CoV-2 [55]. Apart from the membrane bound receptor, the circulating sDPP4 enzyme interacts with GLP-1 and cytokines and regulates the T-cell receptor mediated T-cell activation. The circulating sDPP4 levels have been found to be reduced in both MERS-CoV and SARS-CoV-2 infection, but to contrasting effects [50, 51]. 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 demonstrate a compromised immune response and does not represent a competition for SARS-CoV-2 entry via the DPP4 receptor. Several case reports and Japanese Adverse Drug Event Report

(JADER) has highlighted the risk and incidence of interstitial pneumonia with DPP4 inhibitor (vildagliptin) [56-58]. Further, loss of DPP4 activity is associated with a prothrombotic state in myocardial microvessels due to the upregulation of the procoagulant tissue factor [59], and reduced post-operative DPP4 activity has been associated with worse patient outcome after cardiac surgery due to paradoxical impairment of angiogenesis and endothelial function [60, 61]. Taken together, the partial deletion of *DPP4* in this patient might further contribute to the reduced sDPP4 levels and the compromised immune response predisposing to severe COVID-19.

The rare/unique 833 kb **tandem** duplication of 14 genes (*RPL38*, *MGC16275*, *TTYH2*, *Z49982*, *DNAI2*, *CD300E*, *CD300LD*, *CD300C*, *CD300LB*, *CD300A*, *GPRC5C*, *GPR142*, *BTBD17*, *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). The *CD300* molecules have been identified to participate in mechanisms employed by viruses to escape the immune response and infect host cells [62, 63]. Further, *CD300* molecules downregulate the cytolytic activity of natural killer (NK) cells against infected cells [63]. The 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-κB activation [64, 65]. 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 11/8/2020). To our knowledge, nine *EDARADD* pathogenic sequencing variants have been described, three leading to an autosomal recessive inheritance and six to autosomal dominant mode of inheritance [66]. ED caused by variants in the *EDARADD* gene is associated with recurrent pulmonary infections, accompanied by bronchospasm that requires steroidal treatment, and have been found to impair NF-κB signaling [66]. Cluzeau *et al.*, have identified a homozygous deletion of exon 4 in a Tunisian family with severe ED that resulted in complete abolition of NF-κB signaling [66]. In addition, several reports have shown that a heterozygous mutation in *EDARADD* in patients with ECTD11A leads to severe impairment of NF-κB activation [67, 68]. 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 [68]. 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.

In addition to these rare/private events, two common SVs impact a cluster of chemokine genes (*TBC1D3B*, *CCL3L1*, *CCL4L2*) in 10 of the 37 patients. The *CCL3L-CCL4L* locus is a copy number variable region (CNVR) with inter-individual variability in breakpoints, and has been extensively associated with HIV-related outcomes [69]. A corollary of our findings, single-cell RNA sequencing analysis has found significantly upregulated *CCL4L2* in the alveolar macrophages and lung epithelial cells in severely ill compared to mild patients with COVID-19,

respectively [70, 71]. The duplication of CCL4L2 might explain the upregulation of this chemokine observed in the severely ill patients with COVID-19. Considering chemokine storm to be an important clinical manifestation in severely ill patients, this CNVR region seems to be a critical locus that might influence disease severity by enhancing mobilization and activation of inflammatory cells in these patients. Notably, this SV has a higher frequency in African Americans in our cohort and might be an independent association that could help explain the higher predisposition of the African American population to severe COVID-19. The insertion in the enhancer/promoter region of the *MX1* gene observed in eleven patients could affect regulation of this gene. Several studies have found reduced *MX1* expression downstream of a diminished interferon type I and III pathway in patients with severe COVID-19 [17, 18, 38].

While these findings are preliminary in the above-mentioned immune genes, some candidate genes/loci suggests 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 (Zhang et al Science, vd Made JAMA). 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 whether any of the identified candidate genes harbor rare/private 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.

Common SVs disrupting genes of mucociliary axis

The ~248.1 kb heterozygous inversion disrupting the *MUC4* gene identified in five patients suggests it as a candidate to play a role in increasing disease severity. The mucins have been found to play a critical role in host defense against several respiratory pathogens that include *Pseudomonas aeruginosa* (PA), respiratory syncytial virus (RSV), influenza viruses, and SARS-CoV-2 [72, 73]. Respiratory mucous functions as the primary barrier against the inhaled pathogen(s) [75]. The epithelium of the respiratory tract is coated with mucous and is the site of primary interaction of the inhaled pathogen with the host. Mucous maintains the hydration of the respiratory tract, traps particulate matter including pathogens, and regulates the immune response [76]. Of the 15 mucins in the human respiratory tract, *MUC4* is among the five major mucins and belongs to the membrane-tethered class (*MUC1*, *MUC4*, and *MUC16*) of mucins [76]. Tethered mucins play a number of roles, including the activation of intracellular signal transduction pathways, regulation of the immune response, cell differentiation and proliferation [77]. In human studies using single-cell RNA sequencing of bronchoalveolar lavage fluid (BALF) cells from patients infected with SARS-CoV-2, upregulation of all the five major mucins was observed [74]. Interestingly, the overexpression of *Muc5ac* in mice was protective against influenza virus infection [78], whereas, *Muc5b*^{-/-} mice showed a number of severe deficiencies in airway function, reduced mucociliary clearance and severe infections [79]. *Muc1*^{-/-} mice have been found to be more susceptible to greater lung injury due to inflammation following PA challenge [80]. The anti-inflammatory role of *MUC1* was also observed in an RSV in vitro challenge model [81]. Further,

female *Muc4*^{-/-} mice develop severe illness that include weight loss, difficulty in breathing, and increased inflammatory cytokines following infection with SARS-CoV-2 [82]. Based on the literature evidence and the SVs detected in our cohort, we hypothesize that the disruption of this gene of the mucociliary axis could play a meaningful role in predisposing patients to severe illness following SARS-CoV-2 infection.

Rare and common SVs affecting genes implicated in viral spread and replication

Two rare/unique SVs impacting the keratin genes (*KRT15* and *KRT19*), and *ZNF443*, and one common SV impacting *APOBEC3A* and *APOBEC3B* genes were found in six patients. A rare/unique SV, with a ~ 28.4 kb duplication including two genes (*KRT15* and *KRT19*) was identified in one patient (case 2). The 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 have been found to be upregulated in the respiratory tract during infection [83, 84]. 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 [85, 86]. 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 28.9 kb heterozygous deletion disrupting *APOBEC3B* and *APOBEC3A* is a known polymorphism that results in a hybrid allele that has the promoter and coding region of *APOBEC3A* but the 3' UTR of *APOBEC3B*. The apolipoprotein B messenger RNA-editing, enzyme-catalytic, polypeptide-like 3 (APOBEC3) family of cytidine deaminases plays an important role in the innate immune response to viral infections by editing viral genomes [87, 88]. The deletion polymorphism has been found to result in decreased expression of both genes [89], and is associated with susceptibility to HBV, HIV and HPV infections [90-92]. A recent study analyzing the RNA sequences from BALF demonstrated RNA editing marks typical of the ADAR and APOBEC deaminases, suggesting a role of these deaminases in restricting viral propagation [93]. The RNA editing deaminases are critical in determining the fate of both the host and the virus, and the finding of this polymorphism that occurred in three severely ill patients (4.5% in DGV, 11% in gnomAD, and 20% in Bionano controls) warrant further analysis of this key polymorphism which might be related to the variable clinical outcomes in patients with COVID-19.

Conclusion

The current study provides a powerful framework for investigating structural variations in host genomes that might predispose 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 inter-individual 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 interaction. It is insufficient to consider only genomic variation at the level of

SNVs/ small indels to understand host variability in response COVID-19 exposure. The present investigation highlights that genes implicated in distinct pathways of host-viral interaction were disrupted by structural variations in patients with severe COVID-19. The events affecting these genes might have predisposed these patients towards severe disease course. Further, several findings in this study are consistent with and support the previous reports implicating the TLR3/TLR7 pathway and type I interferon response in the pathogenesis of COVID-19. One finding, a partial duplication of the TLR negative regulator, *STK26*, and the coincident upregulation of *STK26* is especially intriguing. Incredibly, this finding led us to discover that, relative to levels in asymptomatic COVID-19 cases, *STK26* was upregulated in all 12 tested patients with severe COVID-19. This overexpression may be part of a more widespread severe response profile.

There are some limitations in the present investigation, which will benefit from additional follow-up studies. First, the study is limited to a small patient population, and the findings would be made stronger by validation in a larger cohort. Second, 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. Thus, 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 coding region of the genes implicated in host-viral interaction. Thirdly, functional experiments were not performed to confirm downstream effect of all the SVs in respective genes, because of limited volume of blood available from these patients and no access to their post-mortem lung biopsies. Nevertheless, the whole genome investigation of 37 severely ill patients with COVID-19 identified candidate loci that may provide novel insights into the pathogenesis of severe COVID-19 in some cases.

References

1. Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. *N Engl J Med*. 2020 doi:10.1056/nejmcp2009575.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P. China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China. *N Engl J Med*. 2020, 382(8), 727-33.
3. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020, 395(10229), 1054-1062.
4. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, Shi J, Zhou M, Wu B, Yang Z, Zhang C. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *Journal of Allergy and Clinical Immunology*. 2020, 146(1), 110-118. doi:10.1016/j.jaci.2020.04.006
5. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020, 584(7821), 430-6.
6. Wu Z, McGoogan JM. Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention. *JAMA*. 2020, 323(13), 1239.
7. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B.. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020, 395, 1054-62.
8. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, Shi J, Zhou M, Wu B, Yang Z, Zhang C, Yue J, Zhang Z, Renz H, Liu X, Xie J, Xie M, Zhao J. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *J Allergy Clin Immunol*. 2020. 146(1), 110-118.
9. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, Holden KA, Read JM, Dondelinger F, Carson G, Merson L, Lee J, Plotkin D, Sigfrid L, Halpin S, Jackson C, Gamble C, Horby PW, Nguyen-Van-Tam JS, Ho A, Russell CD, Dunning J, Openshaw PJ, Baillie JK, Semple MG; ISARIC4C investigators. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020, 369, m1985.
10. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW; the Northwell COVID-19 Research Consortium, Barnaby DP, Becker LB, Chelico JD, Cohen SL, Cookingham J, Coppa K, Diefenbach MA, Dominello AJ, Duer-Hefele J, Falzon L, Gitlin J, Hajizadeh N, Harvin TG, Hirschwerk DA, Kim EJ, Kozel ZM, Marrast LM, Mogavero JN, Osorio GA, Qiu M, Zanos TP. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020, 323(20), 2052-9.
11. Abbasi J. Younger Adults Caught in COVID-19 Crosshairs as Demographics Shift. *JAMA*. 2020, 324(21), 2141-2143.

12. LoPresti M, Beck DB, Duggal P, Cummings DAT, Solomon BD. The Role of Host Genetic Factors in Coronavirus Susceptibility: Review of Animal and Systematic Review of Human Literature. *Am J Hum Genet.* 2020, 107(3), 381-402.
13. Zhao F, Fan Q, Zhang W, Li B, Zhang J. Association between ABO blood group system and COVID-19 susceptibility in Wuhan. *Front Cell Infect Microbiol.* 2020, 10, 404.
14. Zhao J, Yang Y, Huang HP, Li D, Gu DF, Lu XF, Zhang Z, Liu L, Liu T, Liu YK, He YJ. Relationship between the ABO Blood Group and the COVID-19 Susceptibility. *MedRxiv.* 2020.
15. 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 MM, Milani C, Aziz F, Kässens J, May S, Wendorff M, Wienbrandt L, Uellendahl-Werth F, Zheng T, Yi X, de Pablo R, Chercoles AG, Palom A, Garcia-Fernandez AE, Rodriguez-Frias F, Zanella A, Bandera A, Protti A, Aghemo A, Lleo A, Biondi A, Caballero-Garralda A, Gori A, Tanck A, Carreras Nolla A, Latiano A, Fracanzani AL, Peschuck A, Julià A, Pesenti A, Voza A, Jiménez D, Mateos B, Nafria Jimenez B, Quereda C, Paccapelo C, Gassner C, Angelini C, Cea C, Solier A, Pestaña D, Muñoz-Diaz E, Sandoval E, Paraboschi EM, Navas E, García Sánchez F, Ceriotti F, Martinelli-Boneschi F, Peyvandi F, Blasi F, Téllez L, Blanco-Grau A, Hemmrich-Stanisak G, Grasselli G, Costantino G, Cardamone G, Foti G, Aneli S, Kurihara H, ElAbd H, My I, Galván-Femenia I, Martín J, Erdmann J, Ferrusquía-Acosta J, Garcia-Etxebarria K, Izquierdo-Sanchez L, Bettini LR, Sumoy L, Terranova L, Moreira L, Santoro L, Scudeller L, Mesonero F, Roade L, Rühlemann MC, Schaefer M, Carrabba M, Riveiro-Barciela M, Figuera Basso ME, Valsecchi MG, Hernandez-Tejero M, Acosta-Herrera M, D'Angiò M, Baldini M, Cazzaniga M, Schulzky M, Cecconi M, Wittig M, Ciccarelli M, Rodríguez-Gandía M, Bocciolone M, Miozzo M, Montano N, Braun N, Sacchi N, Martínez N, Özer O, Palmieri O, Faverio P, Preatoni P, Bonfanti P, Omodei P, Tentorio P, Castro P, Rodrigues PM, Blandino Ortiz A, de Cid R, Ferrer R, Gualtierotti R, Nieto R, Goerg S, Badalamenti S, Marsal S, Matullo G, Pelusi S, Juzenas S, Aliberti S, Monzani V, Moreno V, Wesse T, Lenz TL, Pumarola T, Rimoldi V, Bosari S, Albrecht W, Peter W, Romero-Gómez M, D'Amato M, Duga S, Banales JM, Hov JR, Folseraas T, Valenti L, Franke A, Karlsen TH. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med.* 2020, 383(16), 1522-34.
16. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD, Furniss J, Richmond A, Gountouna E, Wrobel N, Harrison D, Wang B, Wu Y, Meynert A, Griffiths F, Oosthuyzen W, Kousathanas A, Moutsianas L, Yang Z, Zhai R, Zheng C, Grimes G, Beale R, Millar J, Shih B, Keating S, Zechner M, Haley C, Porteous DJ, Hayward C, Yang J, Knight J, Summers C, Shankar-Hari M, Klenerman P, Turtle L, Ho A, Moore SC, Hinds C, Horby P, Nichol A, Maslove D, Ling L, McAuley D, Montgomery H, Walsh T, Pereira A, Renieri A; GenOMICC Investigators; ISARICC Investigators; COVID-19 Human Genetics Initiative; 23andMe Investigators; BRACOVIC Investigators; Gen-COVID Investigators, Shen X, Ponting CP, Fawkes A, Tenesa A, Caulfield M, Scott R, Rowan K, Murphy L, Openshaw PJM, Semple MG, Law A, Vitart V, Wilson JF, Baillie JK. Genetic mechanisms of critical illness in Covid-19. *Nature.* 2020, doi: 10.1038/s41586-020-03065-y.

17. Van der Made CI, Simons A, Schuurs-Hoeijmakers J, van den Heuvel G, Mantere T, Kersten S, van Deuren RC, Steehouwer M, van Reijmersdal SV, Jaeger M, Hofste T, Astuti G, Corominas Galbany J, van der Schoot V, van der Hoeven H, Hagmolen Of Ten Have W, Klijn E, van den Meer C, Fiddelaers J, de Mast Q, Bleeker-Rovers CP, Joosten LAB, Yntema HG, Gilissen C, Nelen M, van der Meer JWM, Brunner HG, Netea MG, van de Veerdonk FL, Hoischen A. Presence of Genetic Variants Among Young Men With Severe COVID-19. *JAMA*. 2020, ;324(7), 1–11.
18. Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, Ogishi M, Sabli IKD, Hodeib S, Korol C, Rosain J, Bilguvar K, Ye J, Bolze A, Bigio B, Yang R, Arias AA, Zhou Q, Zhang Y, Onodi F, Korniotis S, Karpf L, Philippot Q, Chbihi M, Bonnet-Madin L, Dorgham K, Smith N, Schneider WM, Razoooky BS, Hoffmann HH, Michailidis E, Moens L, Han JE, Lorenzo L, Bizien L, Meade P, Neehus AL, Ugurbil AC, Corneau A, Kerner G, Zhang P, Rapaport F, Seeleuthner Y, Manry J, Masson C, Schmitt Y, Schlüter A, Le Voyer T, Khan T, Li J, Fellay J, Roussel L, Shahrooei M, Alosaimi MF, Mansouri D, Al-Saud H, Al-Mulla F, Almourfi F, Al-Muhsen SZ, Alshime F, Al Turki S, Hasanato R, van de Beek D, Biondi A, Bettini LR, D'Angio' M, Bonfanti P, Imberti L, Sottini A, Paghera S, Quiros-Roldan E, Rossi C, Oler AJ, Tompkins MF, Alba C, Vandernoot I, Goffard JC, Smits G, Migeotte I, Haerynck F, Soler-Palacin P, Martin-Nalda A, Colobran R, Morange PE, Keles S, Çölkesen F, Ozcelik T, Yasar KK, Senoglu S, Karabela ŞN, Rodríguez-Gallego C, Novelli G, Hraiech S, Tandjaoui-Lambiotte Y, Duval X, Laouénan C; COVID-STORM Clinicians; COVID Clinicians; Imagine COVID Group; French COVID Cohort Study Group; CoV-Contact Cohort; Amsterdam UMC Covid-19 Biobank; COVID Human Genetic Effort; NIAID-USUHS/TAGC COVID Immunity Group, Snow AL, Dalgard CL, Milner JD, Vinh DC, Mogensen TH, Marr N, Spaan AN, Boisson B, Boisson-Dupuis S, Bustamante J, Puel A, Ciancanelli MJ, Meyts I, Maniatis T, Soumelis V, Amara A, Nussenzweig M, García-Sastre A, Krammer F, Pujol A, Duffy D, Lifton RP, Zhang SY, Gorochov G, Béziat V, Jouanguy E, Sancho-Shimizu V, Rice CM, Abel L, Notarangelo LD, Cobat A, Su HC, Casanova JL. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*. 2020, 370(6515). eabd4570.
19. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Smith N, Péré H, Charbit B, Bondet V, Chenevier-Gobeaux C, Breillat P, Carlier N, Gauzit R, Morbieu C, Pène F, Marin N, Roche N, Szwebel TA, Merklings SH, Treluyer JM, Veyer D, Mouthon L, Blanc C, Tharaux PL, Rozenberg F, Fischer A, Duffy D, Rieux-Laucat F, Kernéis S, Terrier B. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*. 2020, 369(6504), 718-24.
20. COVID-19 Host Genetics Initiative. 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*. 2020, 28(6), 715-18.
21. Eichler EE. Genetic variation, comparative genomics, and the diagnosis of disease. *N Engl J Med*. 2019, 381(1), 64-74.

22. Levy-Sakin M, Pastor S, Mostovoy Y, Li L, Leung AK, McCaffrey J, Young E, Lam ET, Hastie AR, Wong KH, Chung CY. Genome maps across 26 human populations reveal population-specific patterns of structural variation. *Nat Commun.* 2019, 10(1), 1-4.
23. Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Mari RS, Yilmaz F. De novo assembly of 64 haplotype-resolved human genomes of diverse ancestry and integrated analysis of structural variation. *bioRxiv.* 2020. doi: <https://doi.org/10.1101/2020.12.16.423102>.
24. Mikocziova I, Gidoni M, Lindeman I, Peres A, Snir O, Yaari G, Sollid LM. Polymorphisms in human immunoglobulin heavy chain variable genes and their upstream regions. *Nucleic Acids Res.* 2020, 48(10), 5499-510.
25. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, Schein J, Willsey AJ, Joy JB, Scott JK, Graves TA, Wilson RK. 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.* 2013, 92(4), 530-46.
26. Isa II, Jamaluddin J, Achim NH, Abubakar S. Population-specific profiling of CCL3L1 copy number of the three major ethnic groups in Malaysia and the implication on HIV susceptibility. *Gene.* 2020, 144821.
27. Monlong J, Cossette P, Meloche C, Rouleau G, Girard SL, Bourque G. Human copy number variants are enriched in regions of low mappability. *Nucleic Acids Res.* 2018, 46(14), 7236-49.
28. Brandt DY, Aguiar VR, Bitarello BD, Nunes K, Goudet J, Meyer D. Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. *G3: Genes, Genomes, Genetics.* 2015, 5(5), 931-41.
29. Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, Ivison GT, Ranganath T, Vergara R, Hollis T, Simpson LJ. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat Med.* 2020, 1-7.
30. Poulson M, Geary A, Annesi C, Allee L, Kenzik K, Sanchez S, Tseng J, Dechert T. National Disparities in COVID-19 Outcomes between Black and White Americans. *J Natl Med Assoc.* 2020 S0027-9684(20), 30149-8.
31. Izzy S, Tahir Z, Cote DJ, Al Jarrah A, Roberts MB, Turbett S, Kadar A, Smirnakis SM, Feske SK, Zafonte R, Fishman JA, El Khoury J. Characteristics and Outcomes of Latinx Patients With COVID-19 in Comparison With Other Ethnic and Racial Groups. *Open Forum Infect Dis.* 2020, 7(10), ofaa401.
32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001, 4, 402-8.
33. Gold JAW, Wong KK, Szablewski CM, Patel PR, Rossow J, da Silva J, Natarajan P, Morris SB, Fanfair RN, Rogers-Brown J, Bruce BB, Browning SD, Hernandez-Romieu AC, Furukawa NW, Kang M, Evans ME, Oosmanally N, Tobin-D'Angelo M, Drenzek C, Murphy DJ, Hollberg J, Blum JM, Jansen R, Wright DW, Sewell WM 3rd, Owens JD, Lefkove B, Brown FW, Burton DC, Uyeki TM, Bialek SR, Jackson BR. Characteristics and Clinical Outcomes of Adult Patients Hospitalized with COVID-19 - Georgia. *MMWR Morb Mortal Wkly Rep* 2020, 69, 545–50.

34. Brubaker SW, Bonham KS, Zanoni I, Kagan JC. Innate immune pattern recognition: a cell biological perspective. *Annu Rev Immunol*, 2015, 33, 257-90.
35. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. *Cell*, 2006, 124(4), 783-801.
36. Channappanavar R, Fehr AR, Zheng J, Wohlford-Lenane C, Abrahante JE, Mack M, Sompallae R, McCray PB Jr, Meyerholz DK, Perlman S. IFN-I response timing relative to virus replication determines MERS coronavirus infection outcomes. *J Clin Invest*. 2019, 129(9), 3625-3639.
37. Choudhury A, Das NC, Patra R, Bhattacharya M, Mukherjee S. In silico analyses on the comparative sensing of SARS-CoV-2 mRNA by intracellular TLRs of human. *bioRxiv*. 2020. doi: <https://doi.org/10.1101/2020.11.11.377713>.
38. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, Jordan TX, Oishi K, Panis M, Sachs D, Wang TT, Schwartz RE, Lim JK, Albrecht RA, tenOever BR. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. 2020, 181(5), 1036-1045.e9.
39. Balzer Le S, Onsager I, Lorentzen JA, Lale R. Dual UTR-A novel 5' untranslated region design for synthetic biology applications. *Synthetic Biology*. 2020; 5(1), ysa006.
40. Shi Z, Zhou Z. MST kinases in innate immune signaling. *Cell Stress*. 2018, 1, 4.
41. Jiao S, Zhang Z, Li C, Huang M, Shi Z, Wang Y, Song X, Liu H, Li C, Chen M, Wang W. The kinase MST4 limits inflammatory responses through direct phosphorylation of the adaptor TRAF6. *Nat Immunol*. 2015, 16(3), 246-57.
42. Konno H, Yamamoto T, Yamazaki K, Gohda J, Akiyama T, Semba K, Goto H, Kato A, Yujiri T, Imai T, Kawaguchi Y. TRAF6 establishes innate immune responses by activating NF- κ B and IRF7 upon sensing cytosolic viral RNA and DNA. *PloS one*. 2009, 4(5), e5674.
43. Zhou Q, Lin H, Wang S, Wang S, Ran Y, Liu Y, Ye W, Xiong X, Zhong B, Shu HB, Wang YY. The ER-associated protein ZDHHC1 is a positive regulator of DNA virus-triggered, MITA/STING-dependent innate immune signaling. *Cell Host Microbe*. 2014, 16(4), 450-61.
44. Xu X, Li M, Wu Z, Wang H, Wang L, Huang K, Liu X, Hou Q, Lin G, Hu C. Endoplasmic reticulum transmembrane proteins ZDHHC1 and STING both act as direct adaptors for IRF3 activation in teleost. *J Immunol*. 2017, 199(10), 3623-33.
45. Wang X, Wu Z, Li Y, Yang Y, Xiao C, Liu X, Xiang X, Wei J, Shao D, Liu K, Deng X. p53 promotes ZDHHC1-mediated IFITM3 palmitoylation to inhibit Japanese encephalitis virus replication. *PLoS Pathog*. 2020, 16(10), e1009035.
46. Bozzo CP, Nchioua R, Volcic M, Krüger J, Heller S, Stürzel CM, Kmiec D, Conzelmann C, Müller J, Zech F, Schütz D. IFITM proteins promote SARS-CoV-2 infection and are targets for virus inhibition. *bioRxiv*. 2020, doi: <https://doi.org/10.1101/2020.08.18.255935>.
47. Nikoloudis D, Kountouras D, Hiona A. 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*. 2020, 12;8, e10402.
48. Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P, Guo D, Fu L, Cui Y, Liu X, Arledge KC. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res*. 2013, 23(8), 986-93.

49. Reinhold D, Brocke S. DPP4-directed therapeutic strategies for MERS-CoV. *Lancet Infect Dis.* 2014, 14(2), 100-1.
50. Inn KS, Kim Y, Aigerim A, Park U, Hwang ES, Choi MS, Kim YS, Cho NH. Reduction of soluble dipeptidyl peptidase 4 levels in plasma of patients infected with Middle East respiratory syndrome coronavirus. *Virology.* 2018 May 1;518:324-7.
51. Schlicht K, Rohmann N, Geisler C, Hollstein T, Knappe C, Hartmann K, Schwarz J, Tran F, Schunk D, Junker R, Bahmer T. Circulating levels of soluble Dipeptidylpeptidase-4 are reduced in human subjects hospitalized for severe COVID-19 infections. *Int J Obes.* 2020, 44(11), 2335-8.
52. Trzaskalski NA, Fadzeyeva E, Mulvihill EE. Dipeptidyl peptidase-4 at the interface between inflammation and metabolism. *Clin Med Insights Endocrinol Diabetes.* 2020, 13, 1179551420912972.
53. Klemann C, Wagner L, Stephan M, von Hörsten S. Cut to the chase: a review of CD26/dipeptidyl peptidase-4's (DPP4) entanglement in the immune system. *Clin Expe Immunol.* 2016, 185(1), 1-21.
54. Li Y, Zhang Z, Yang L, Lian X, Xie Y, Li S, Xin S, Cao P, Lu J. The MERS-CoV receptor DPP4 as a candidate binding target of the SARS-CoV-2 spike. *Iscience.* 2020, 101160.
55. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol.* 2020, 5(4), 562-9.
56. Pharmaceutical and Medical Device Agency. Japanese Adverse Drug Event Report database (since April, 1, 2004). Available from URL: <https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0005.html>.
57. Sada KE, Wada J, Morinaga H, Tuchimochi S, Uka M, Makino H. Sarcoid-like lung granulomas in a hemodialysis patient treated with a dipeptidyl peptidase-4 inhibitor. *Clin Kidney J.* 2014, 7(2), 182-5.
58. Tanaka Y, Soda H, Fukuda Y, Nio K, Ono S, Tomono H, Shimada M, Yoshida M, Harada T, Umemura A, Iwasaki K. Vildagliptin-induced ground-glass nodules mimicking lung metastases in a cancer patient receiving Lactobacillus probiotic supplementation. *Thorac Cancer.* 2020, 11(2), 470-4.
59. Krijnen PA, Hahn NE, Kholová I, Baylan U, Sipkens JA, van Alphen FP, Vonk AB, Simsek S, Meischl C, Schalkwijk CG, van Buul JD. Loss of DPP4 activity is related to a prothrombotic status of endothelial cells: implications for the coronary microvasculature of myocardial infarction patients. *Basic Res Cardiology.* 2012, 107(1), 233.
60. Sun CK, Leu S, Sheu JJ, Tsai TH, Sung HC, Chen YL, Chung SY, Ko SF, Chang HW, Yip HK. 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.* 2013, 4(2), 1-7.
61. Noels H, Theelen W, Sternkopf M, Jankowski V, Moellmann J, Kraemer S, Lehrke M, Marx N, Martin L, Marx G, Jankowski J. Reduced post-operative DPP4 activity associated with worse patient outcome after cardiac surgery. *Sci Rep.* 2018, 8(1), 1-2.

62. Carnec X, Meertens L, Dejarnac O, Perera-Lecoin M, Hafirassou ML, Kitaura J, Ramdasi R, Schwartz O, Amara A. The Phosphatidylserine and Phosphatidylethanolamine Receptor CD300a Binds Dengue Virus and Enhances Infection. *J Virol*. 2015, 90(1), 92-102.
63. Vitallé J, Terrén I, Orrantia A, Zenarruzabeitia O, Borrego F. CD300 receptor family in viral infections. *Eur J Immunol*. 2019, 49(3), 364-374.
64. Martínez-Romero MC, Ballesta-Martínez MJ, López-González V, Sánchez-Soler MJ, Serrano-Antón AT, Barreda-Sánchez M, Rodríguez-Peña L, Martínez-Menchon MT, Frías-Iniesta J, Sánchez-Pedreño P, Carbonell-Meseguer P, Glover-López G, Guillén-Navarro E; GIEDE (Spanish multidisciplinary research group for ectodermal dysplasia). EDA, EDAR, EDARADD and WNT10A allelic variants in patients with ectodermal derivative impairment in the Spanish population. *Orphanet J Rare Dis*. 2019, 14(1), 281.
65. Smahi A, Courtois G, Rabia SH, Döffinger R, Bodemer C, Munnich A, Casanova JL, Israël A. The NF-kappaB signalling pathway in human diseases: from incontinentia pigmenti to ectodermal dysplasias and immune-deficiency syndromes. *Hum Mol Genet*. 2002, 11(20), 2371-5.
66. Cluzeau C, Marrakchi S, Picard C, Munnich A, Smahi A, Turki H. First homozygous large deletion in EDARADD gene associated with a severe form of anhidrotic ectodermal dysplasia. *J Eur Acad Dermatol Venereol*. 2019, 33(2), e55-e57.
67. Bal E, Baala L, Cluzeau C, El Kerch F, Ouldin K, Hadj-Rabia S, Bodemer C, Munnich A, Courtois G, Sefiani A, Smahi A. Autosomal dominant anhidrotic ectodermal dysplasias at the EDARADD locus. *Hum Mutat*. 2007, 28(7), 703-9.
68. Wohlfart S, Söder S, Smahi A, Schneider H. A novel missense mutation in the gene EDARADD associated with an unusual phenotype of hypohidrotic ectodermal dysplasia. *Am J Med Genet A*. 2016, 170A(1), 249-53.
69. Colobran R, Pedrosa E, Carretero-Iglesia L, Juan M. Copy number variation in chemokine superfamily: the complex scene of CCL3L-CCL4L genes in health and disease. *Clin Exp Immunol*. 2010, 162(1), 41-52.
70. Bost P, Giladi A, Liu Y, Bendjelal Y, Xu G, David E, Blecher-Gonen R, Cohen M, Medaglia C, Li H, Deczkowska A, Zhang S, Schwikowski B, Zhang Z, Amit I. Host-Viral Infection Maps Reveal Signatures of Severe COVID-19 Patients. *Cell*. 2020, 181(7), 1475-1488.
71. Chen H, Liu W, Liu D, Zhao L, Yu J. SARS-CoV-2 activates lung epithelia cell proinflammatory signaling and leads to immune dysregulation in COVID-19 patients by single-cell sequencing. *medRxiv*. 2020. doi: <https://doi.org/10.1101/2020.05.08.20096024>
72. Zanin M, Baviskar P, Webster R, Webby R. The interaction between respiratory pathogens and mucus. *Cell Host Microbe*. 2016, 19(2), 159-68.
73. Chatterjee M, van Putten JP, Strijbis K. Defensive Properties of Mucin Glycoproteins during Respiratory Infections—Relevance for SARS-CoV-2. *Mbio*. 2020, 22, 11(6), :e02374-20.
74. He J, Cai S, Feng H, Cai B, Lin L, Mai Y, Fan Y, Zhu A, Huang H, Shi J, Li D. Single-cell analysis reveals bronchoalveolar epithelial dysfunction in COVID-19 patients. *Protein Cell*. 2020, 11(9), 680-7.

75. Lund Håheim L. Epithelial cilia is the first line of defence against Coronavirus; addressing the observed age-gradient in the COVID-19 infection. *Med Hypotheses*. 2020, 143, 110064.
76. Zanin M, Baviskar P, Webster R, Webby R. The Interaction between Respiratory Pathogens and Mucus. *Cell Host Microbe*. 2016, 19(2), 159-68.
77. Lillehoj EP, Hyun SW, Feng C, Zhang L, Liu A, Guang W, Nguyen C, Luzina IG, Atamas SP, Passaniti A, Twaddell WS, Puché AC, Wang LX, Cross AS, Goldblum SE. NEU1 sialidase expressed in human airway epithelia regulates epidermal growth factor receptor (EGFR) and MUC1 protein signaling. *J Biol Chem*. 2012, 287(11), 8214-31.
78. Ehre C, Worthington EN, Liesman RM, Grubb BR, Barbier D, O'Neal WK, Sallenave JM, Pickles RJ, Boucher RC. Overexpressing mouse model demonstrates the protective role of Muc5ac in the lungs. *Proc Natl Acad Sci U S A*. 2012, 109(41), 16528-33.
79. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, Alexander SN, Bellinghausen LK, Song AS, Petrova YM, Tuvim MJ, Adachi R, Romo I, Bordt AS, Bowden MG, Sisson JH, Woodruff PG, Thornton DJ, Rousseau K, De la Garza MM, Moghaddam SJ, Karmouty-Quintana H, Blackburn MR, Drouin SM, Davis CW, Terrell KA, Grubb BR, O'Neal WK, Flores SC, Cota-Gomez A, Lozupone CA, Donnelly JM, Watson AM, Hennessy CE, Keith RC, Yang IV, Barthel L, Henson PM, Janssen WJ, Schwartz DA, Boucher RC, Dickey BF, Evans CM. Muc5b is required for airway defence. *Nature*. 2014, 505(7483), 412-6.
80. Lu W, Hisatsune A, Koga T, Kato K, Kuwahara I, Lillehoj EP, Chen W, Cross AS, Gendler SJ, Gewirtz AT, et al. Cutting edge: enhanced pulmonary clearance of *Pseudomonas aeruginosa* by Muc1 knockout mice. *J Immunol*. 2006, 176, 3890-94.
81. Li Y, Dinwiddie DL, Harrod KS, Jiang Y, Kim KC. Anti-inflammatory effect of MUC1 during respiratory syncytial virus infection of lung epithelial cells in vitro. *Am J Physiol Lung Cell Mol* 2010, 298, L558–L563.
82. Plante JA, Plante K, Gralinski L, Beall A, Ferris MT, Bottomly D, Green RR, McWeeney S, Heise MT, Baric RS, Menachery VD. Mucin 4 Protects Female Mice from Coronavirus Pathogenesis. *bioRxiv*. 2020, doi: <https://doi.org/10.1101/2020.02.19.957118>
83. Pociask DA, Robinson KM, Chen K, McHugh KJ, Clay ME, Huang GT, Benos PV, Janssen-Heininger YM, Kolls JK, Anathy V, Alcorn JF. Epigenetic and transcriptomic regulation of lung repair during recovery from influenza infection. *Am J Pathol*. 2017, 187(4), 851-63.
84. Basu A, Sarkar A, Maulik U. Study of cell to cell transmission of SARS CoV 2 virus particle using gene network from micro array data. *bioRxiv*. 2020, doi: <https://doi.org/10.1101/2020.05.26.116780>
85. Pasquier C, Robichon A. Computational search of hybrid human/SARS-CoV-2 dsRNA reveals unique viral sequences that diverge from other coronavirus strains. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.04.08.031856>
86. Pasquier C, Robichon A. SARS-CoV-2 might manipulate against its host the immunity RNAi/Dicer/Ago system. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.04.08.031856>
87. Stavrou S, Ross SR. APOBEC3 Proteins in Viral Immunity. *J Immunol*. 2015, 195(10), 4565-70.
88. Harris RS, Dudley JP. APOBECs and virus restriction. *Virology*. 2015, 479-480, 131-45.

89. Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J, Kozlowski P. The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population. *Oncotarget*. 2017, 8(44), 76357-76374.
90. Zhu YP, Peng ZG, Wu ZY, Li JR, Huang MH, Si SY, Jiang JD. Host APOBEC3G protein inhibits HCV replication through direct binding at NS3. *PLoS One*. 2015, 10(3), e0121608.
91. Janahi EM, McGarvey MJ. The inhibition of hepatitis B virus by APOBEC cytidine deaminases. *J Viral Hepat*. 2013, 20(12), 821-8.
92. Yang H, Ito F, Wolfe AD, Li S, Mohammadzadeh N, Love RP, Yan M, Zirkle B, Gaba A, Chelico L, Chen XS. Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. *Nat Commun*. 2020, 11(1). 632.
93. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv*. 2020, 6(25), eabb5813.

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

Patient Characteristics	Categories	Numbers / Mean \pm SD
Patients		37
Age		61.6 \pm 12.4 (19-83)
Sex	Male	19
	Female	18
Ethnicity	African American	26
	Caucasian	10
	Hispanic	1
BMI		34.5 \pm 7.1 (19.7-50.9)
Co-morbidities		
	Diabetes	16
	Hypertension	22
	Chronic Kidney disease	9
	Asthma	2
ICU parameters		
	Requiring mechanical Ventilation	30
	Mean duration of Intubation	12 \pm 5.5 (3-20)
	Intubation PTA	13
	Proned	6
Outcome	Dead	12
	Recovered	25

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

S.No	Chr	RefStartPos	RefEndPos	Size (bp)	Zygoty	SV	Gene(s)	Samples	Present in % of new BNGO controls with same enzyme	qPCR	FaNDOM
1	1	236604233	236766495	162263	_	Duplication	<i>EDARADD, HEATR1</i>	22,26	0	Yes	Yes
2	2	162887379	162911439		heterozygous	Deletion	<i>DPP4</i>	38	0	Yes	Yes
3	16	67308198	67455019		Heterozygous	Deletion	<i>PLEKHG4, KCTD19, LRRC36, TPPP3, U1, ZDHHC1</i>	39	0	Yes	Yes
4	17	39662399	39690882		_	Duplication	<i>KRT15, KRT19</i>	2	0	Yes	Yes
5	17	71844581	72678517	833937	_	Duplication_Split	<i>RPL38, MGC16275, TTYH2, Z49982, DNAI2, CD300E, CD300LD, CD300C, CD300LB, CD300A, GPRC5C, GPR142, BTBD17, KIF19, RAB37</i>	19	0	Yes	Yes
6	19	12512276	12552113	39838	_	Duplication	<i>ZNF443</i>	13,18	0	Yes	Yes
7	23	130629618	131164603		_	Duplication_Split	<i>OR13H1, LOC286467, 5S_rRNA, STK26</i>	44	0	Yes	Yes

Table 3. List of structural variants overlapping coding region(s) of gene(s) previously associated with COVID-19/ viral infections/ immune response identified in severe COVID-19 patients.

S.No	Chr	RefStartPos	RefEndPos	Size (bp)	Zygoty	SV	Gene(s)	Samples	Ethnicity (AA/ C)	Present in % of BNGO AFR controls with same enzyme	Present in % of new BNGO controls with same enzyme	qPCR	FaNDOM
1	3	195725611	195477445	248166	heterozygous	Inversion	<i>MUC4</i>	2,3,8,26,29	5 AA	35	9	_	Yes
2	17	34437711	34571682	133972	_	Duplication_inverted	<i>TBC1D3B, CCL3L1, CCL4L2</i>	12, 20, 28, 29, 6, 8, 13,19,21,22	9 AA, 1 C	45	6.9	Yes	Yes
3	22	39347202	39391164	29851	heterozygous	Deletion	<i>APOBEC3A, APOBEC3B</i>	15,29; 44	1 C, 1 AA, 1 H	15	20	Yes	Yes
4	21	42784603	42793518	1591	heterozygous	Insertion	<i>MXI</i>	13, 17, 18, 20, 21, 24, 25, 28, 4, 8, 38	10 AA, 1 C	65	21.4	_	Yes

AA: African American, C: Caucasian, H: Hispanic

Optical genome mapping on 37 severely ill patients SV analyses

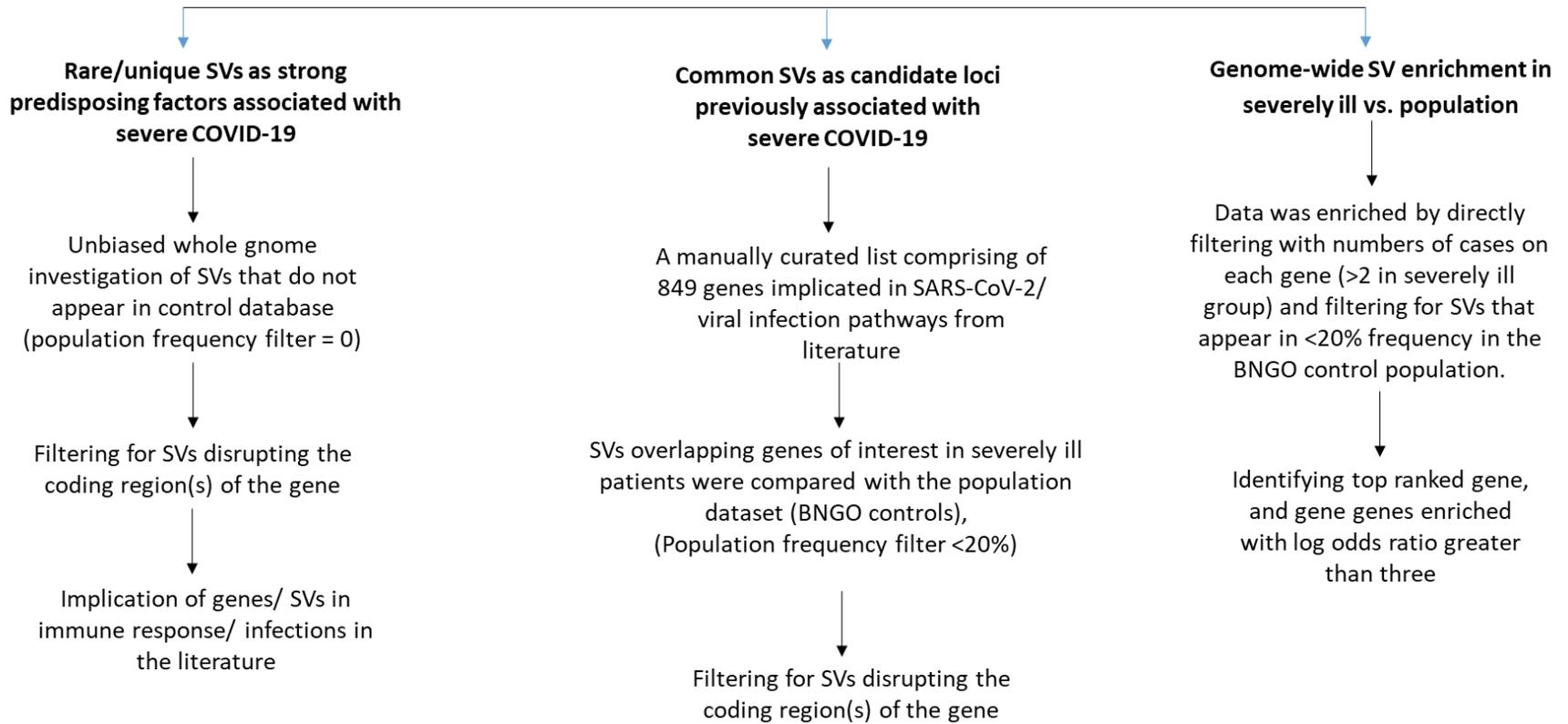


Figure 1: Workflow diagram for analysis of COVID-19 patient SVs.

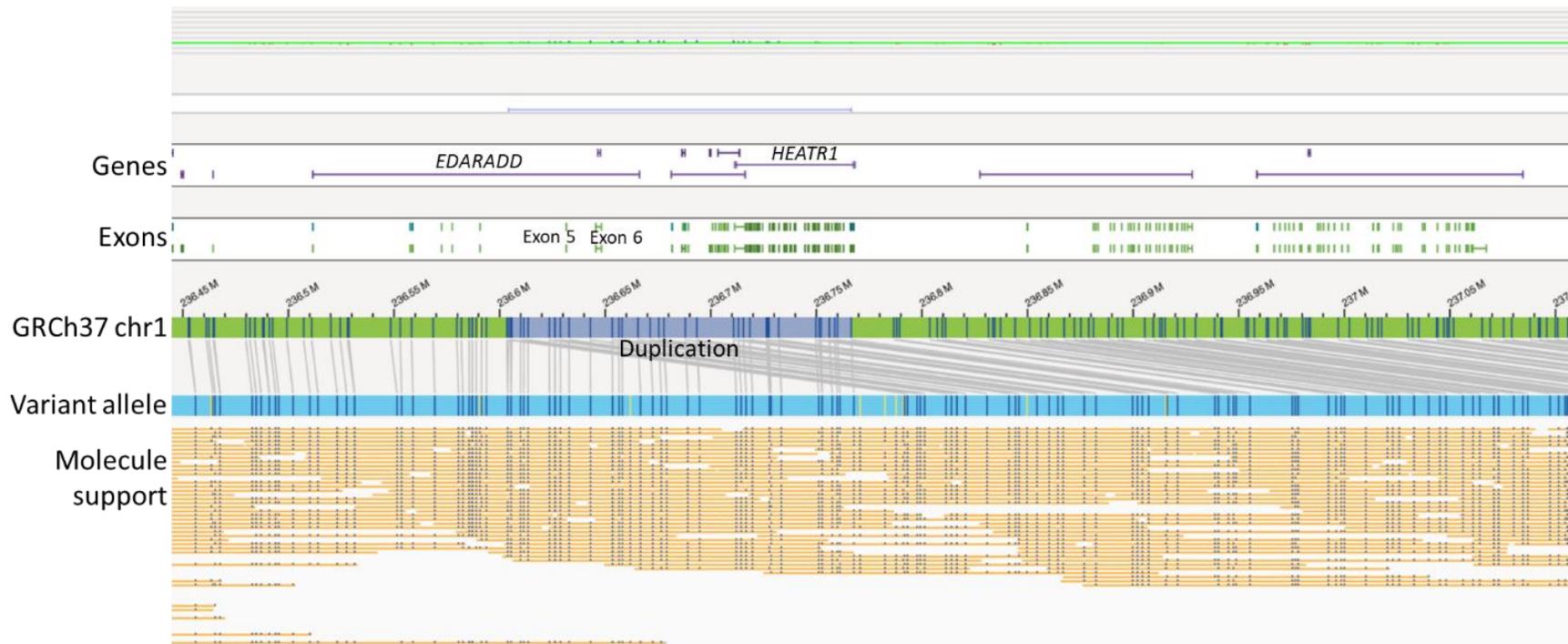


Figure 2a: A ~162.2 kb duplication of chr 1 (chr1:236604233-236766495) partially disrupting the *EDARADD* (exon 5 and 6; NM_145861.3) and *HEATR1* (exons 3-45; NM_018072.6) genes in two patients (cases 22 and 26).

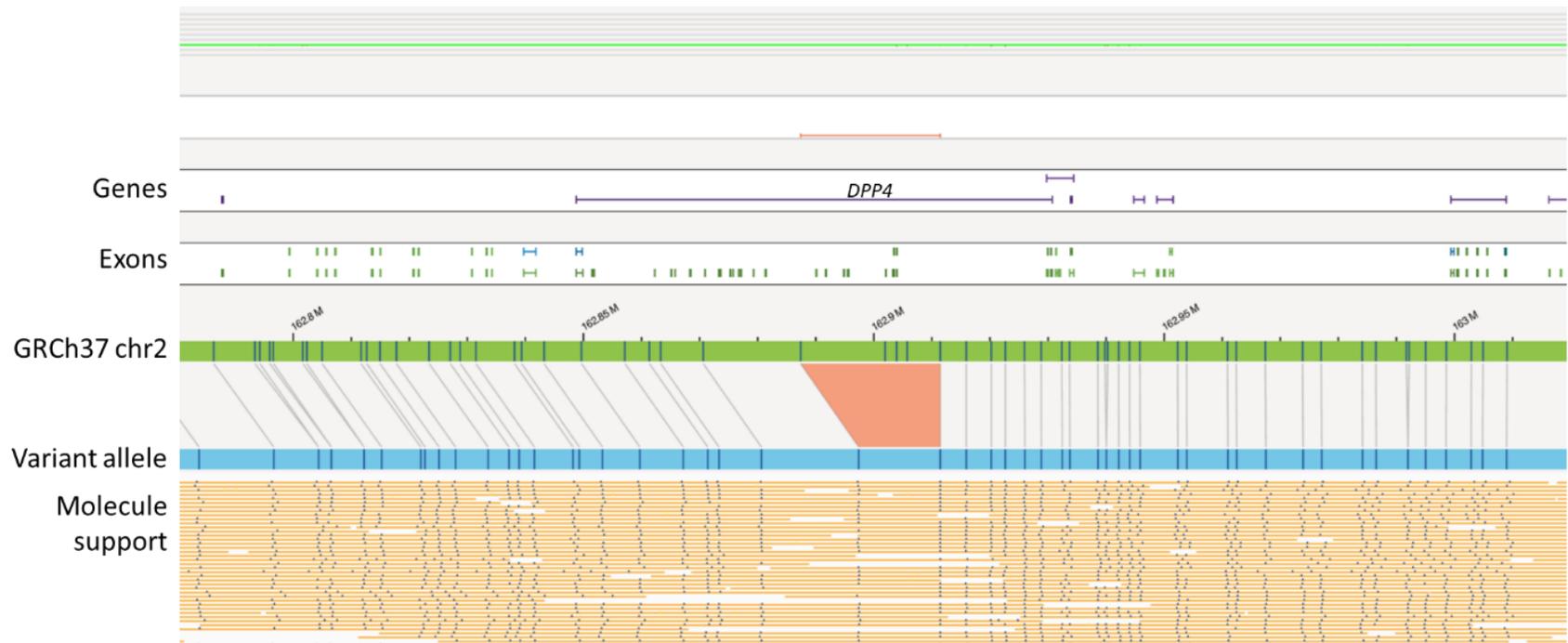


Figure 2b: A ~24.06 kb heterozygous deletion of chr 2 (chr2:162887379-162911439) partially deleting the *DPP4* gene (exons 3 and 4; NM_001935.4) in one patient (case 38).



Figure 2c: A ~146.8 kb heterozygous deletion of chr 16 (chr16:67308198-67455019) that contained six genes (*PLEKHG4*, *KCTD19*, *LRRC36*, *TPPP3*, *U1*, *ZDHHC1*) was identified in one patient (case 39)

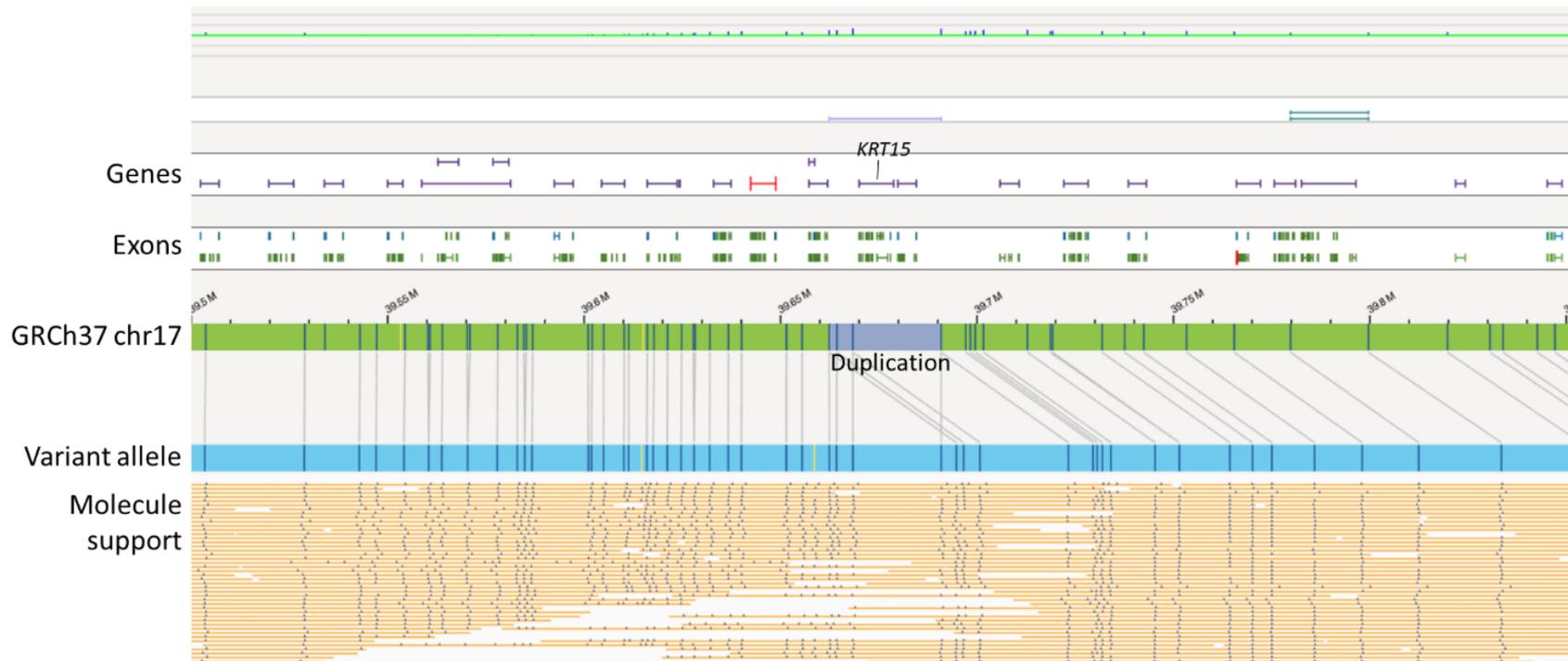


Figure 2d: A ~ 28.4 kb duplication of chr 17 (chr17:39662399-39690882) including two genes (*KRT15* and *KRT19*) in one patient (case 2).

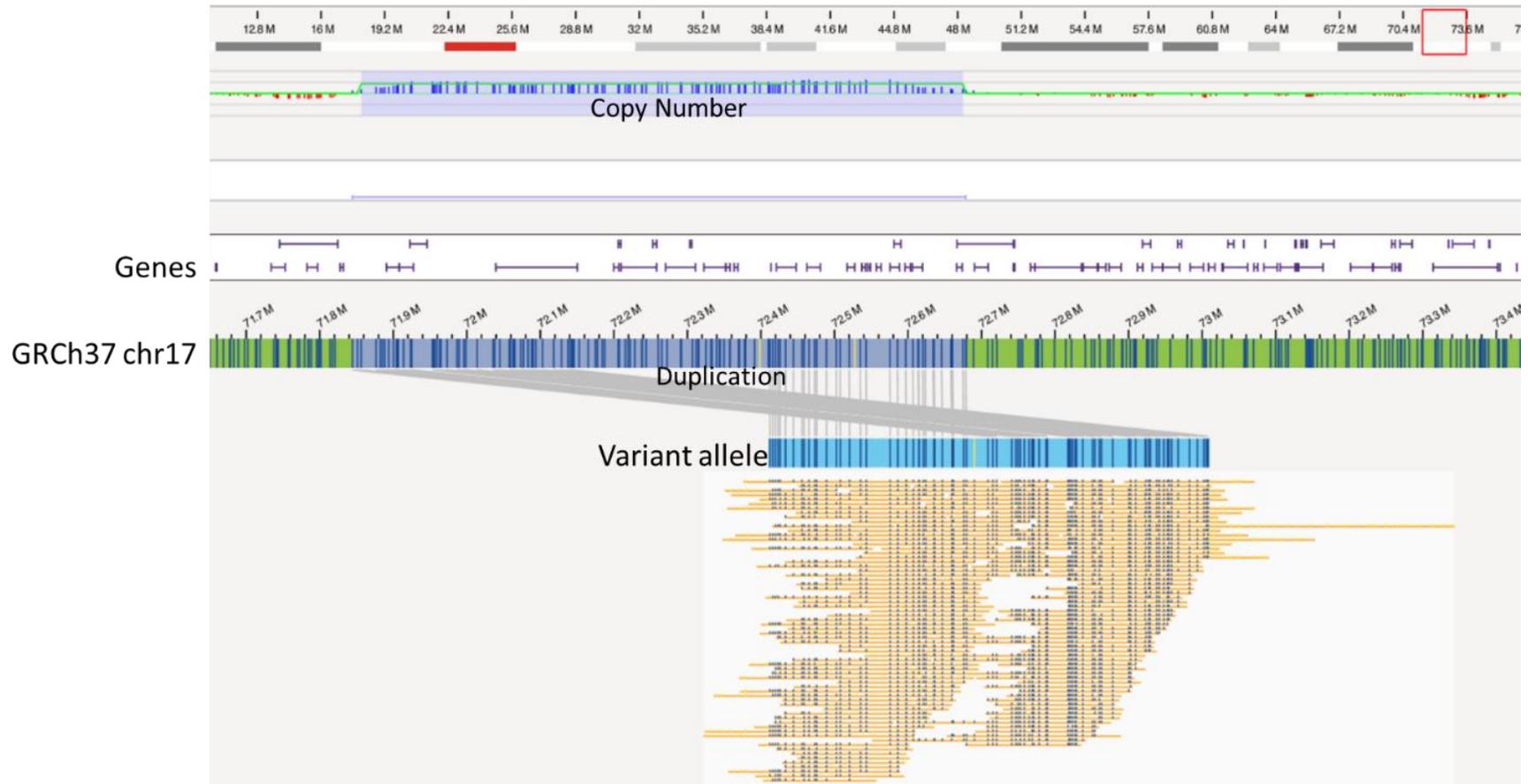


Figure 2e: A ~833 kb **tandem** duplication of chr 17 (chr17:71844581-72678517) including 15 genes (*RPL38*, *MGC16275*, *TTYH2*, *Z49982*, *DNAI2*, *CD300E*, *CD300LD*, *CD300C*, *CD300LB*, *CD300A*, *GPRC5C*, *GPR142*, *BTBD17*, *KIF19*), and partially disrupting *RAB37* (exon 1; NM_175738.5) in one patient (case 19).

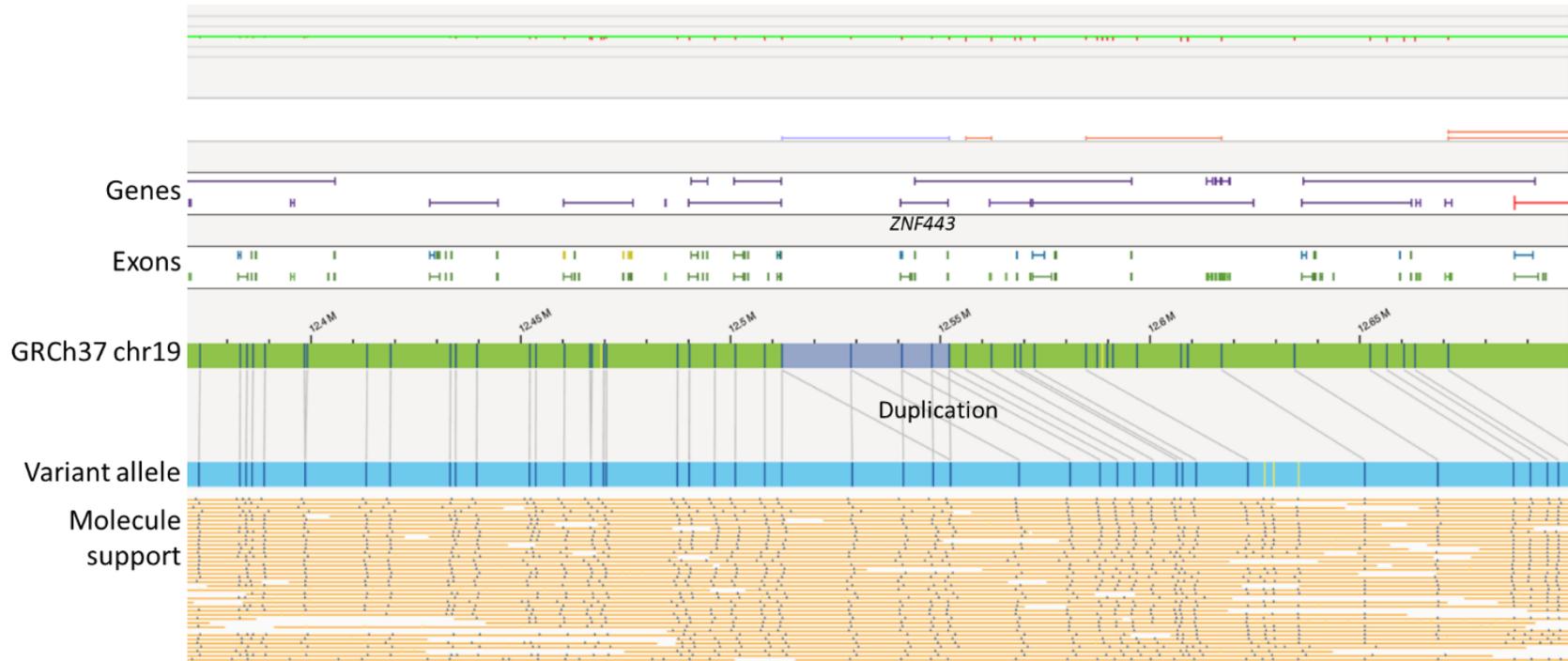


Figure 2f: A ~39.8 kb duplication of chr 19 (chr19:12512276- 12552113) including the entire *ZNF443* gene (NM_005815.5) in two patients (cases 13 and 18).

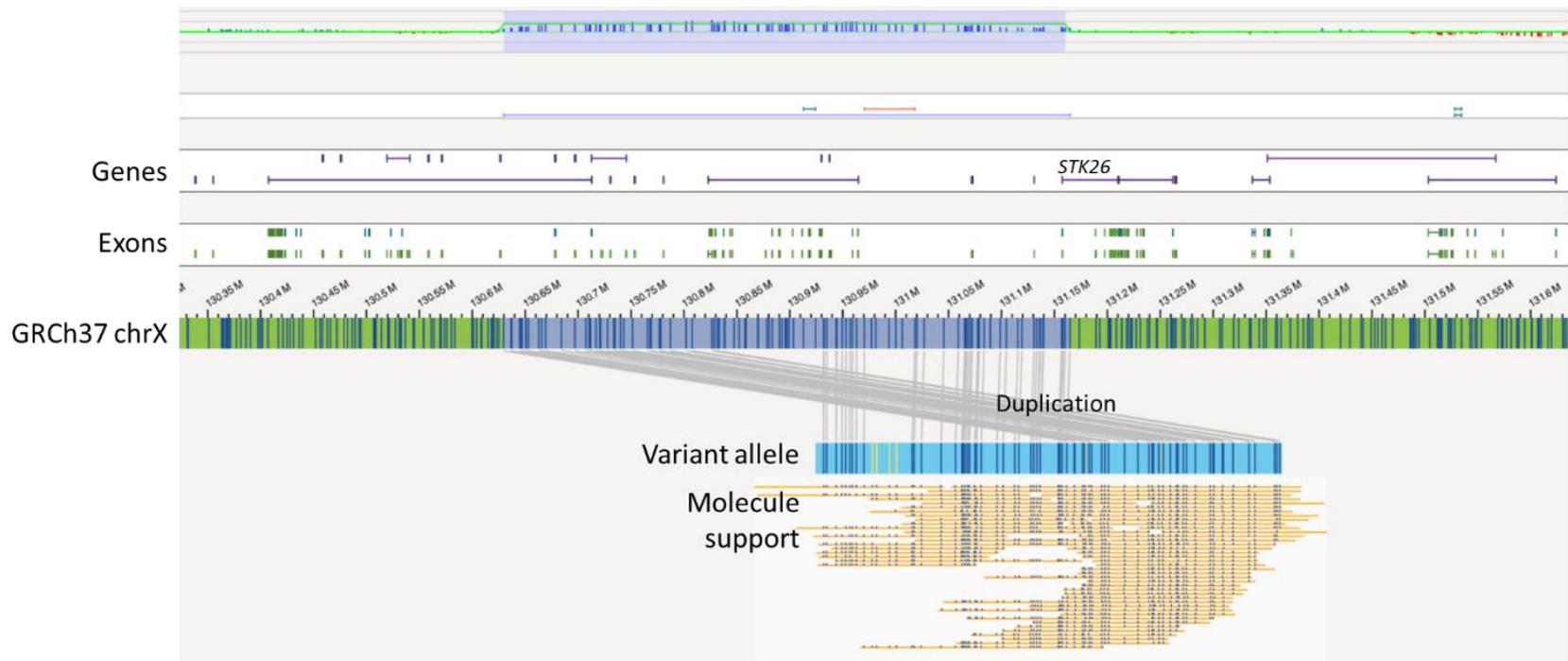


Figure 2g: A ~ 534.9 kb tandem duplication of chr X (chrX: 130629618-131164603) including four genes (*OR13H1*, *LOC286467*, *5S_rRNA*, *STK26*) in one patient (case 44).



Figure 3a: A ~248.1 kb heterozygous inversion of chr 3 (chr3:195725611-195477445) disrupting the *MUC4* gene (exons 1-23; NM_018406.7) in five patients of African American (AA) ethnicity (cases- 2, 3, 8, 26, 29).

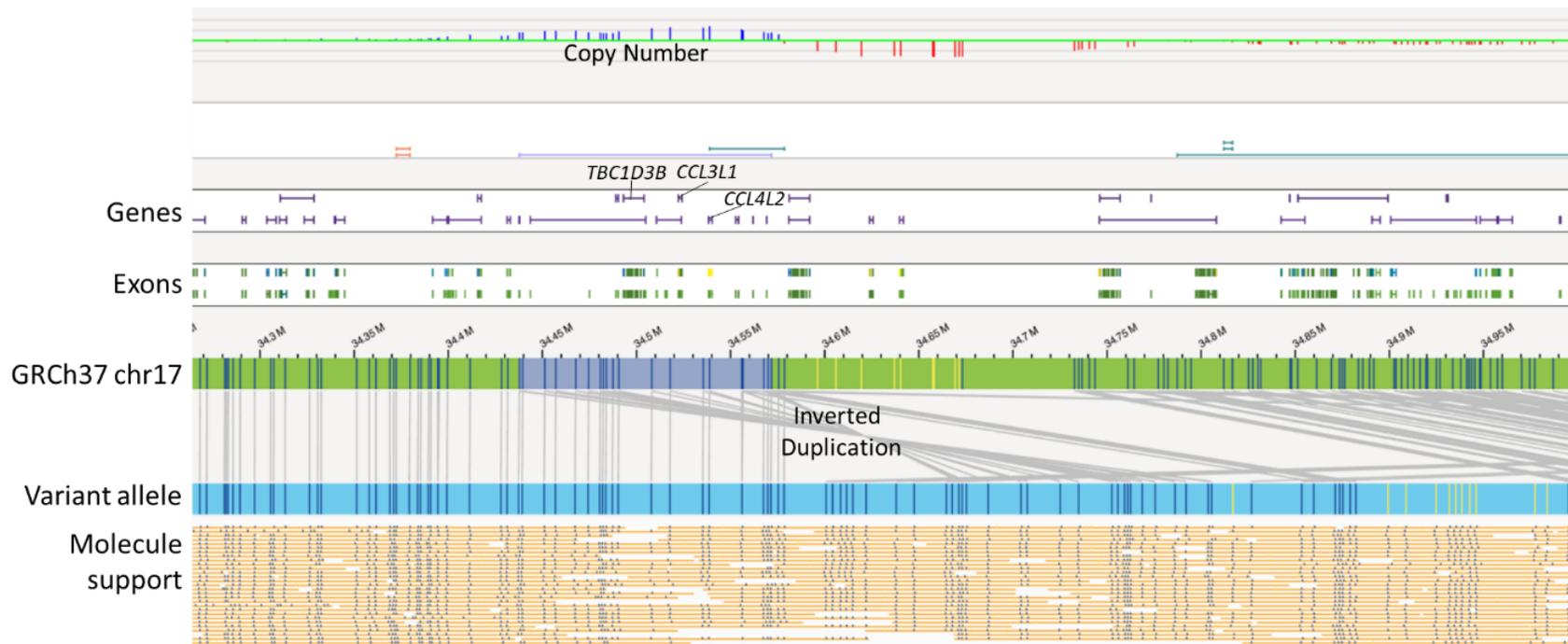


Figure 3b: A ~29.8 kb duplication of chr 17 (chr17:34437711- 34571682) including 3 genes (*TBC1D3B*, *CCL3L1*, *CCL4L2*) in ten patient (cases 12, 20, 28, 29, 6, 8, 13, 19, 21, 22).

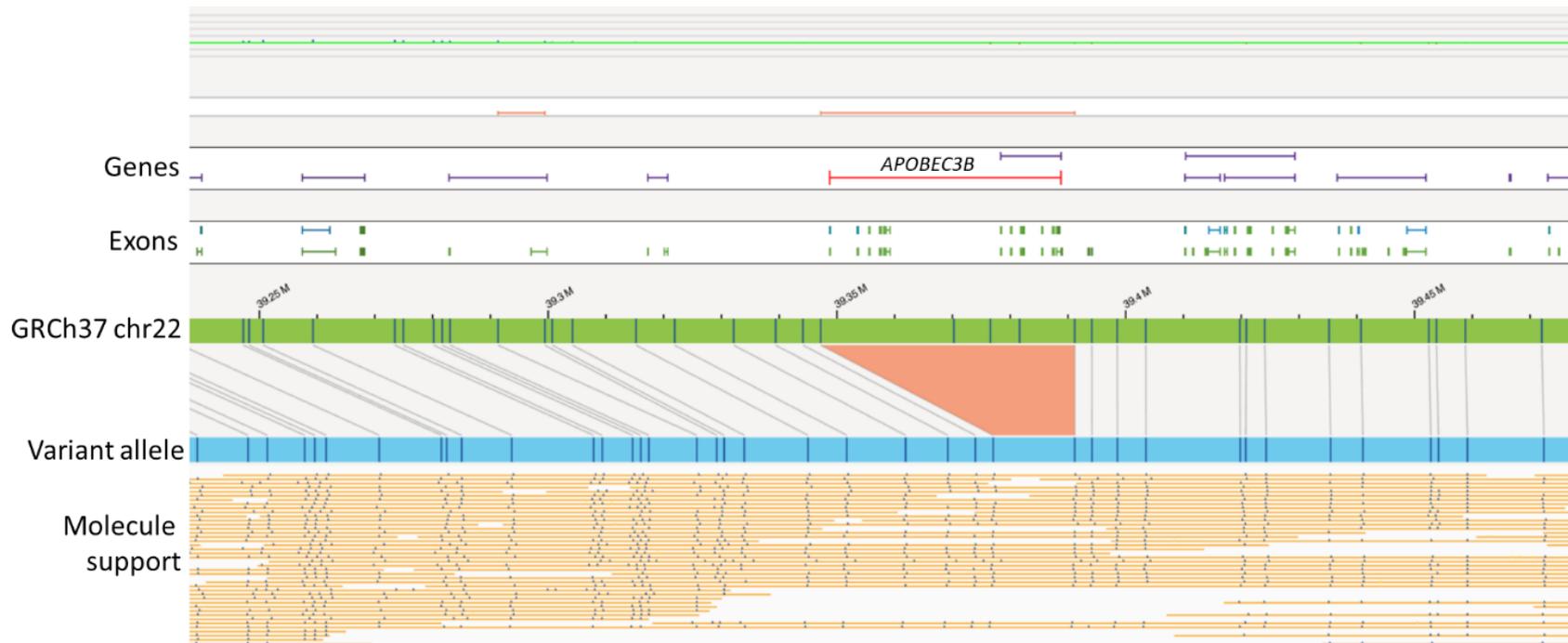


Figure 3c: A ~28.9 kb heterozygous deletion of chr 22 (chr22:39347202-39391164) disrupting 2 genes (*APOBEC3B* and *APOBEC3A*) in three patients.

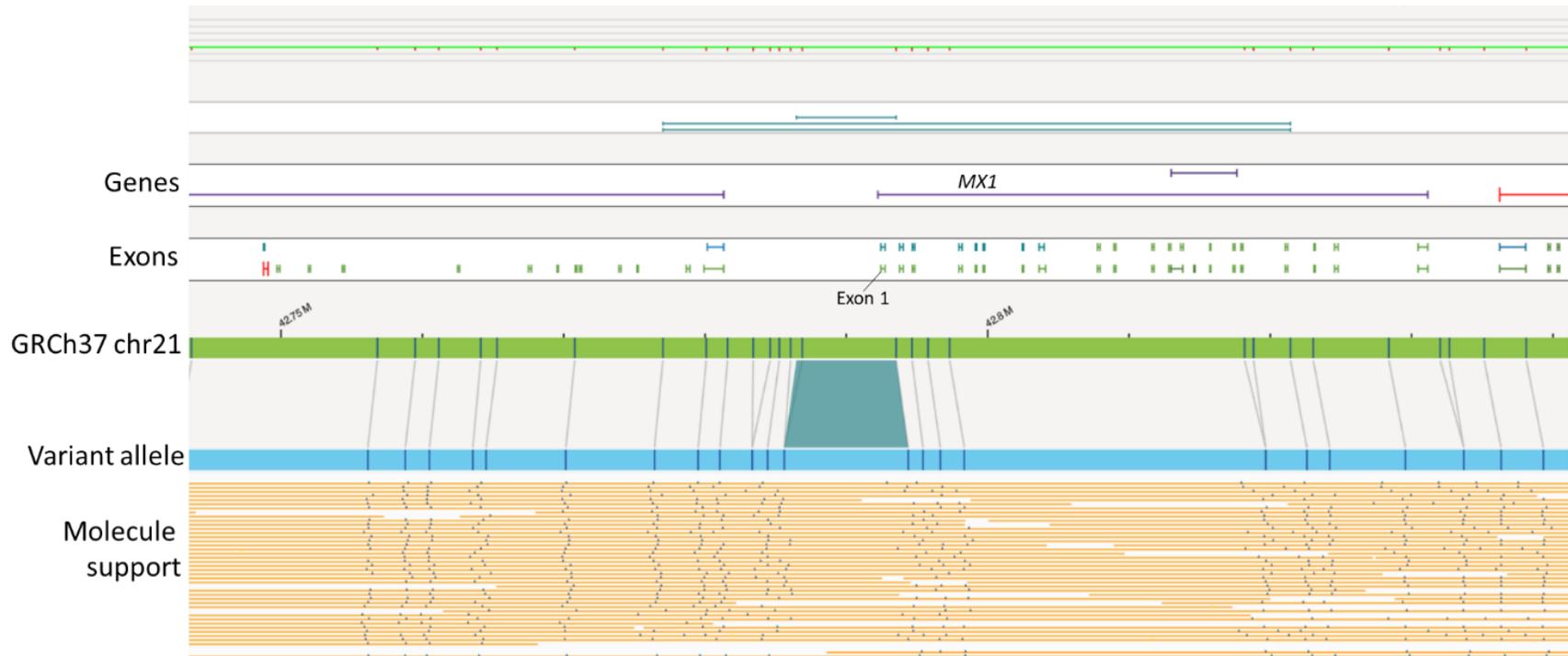


Figure 3d: A ~1.5 kb insertion in chr 21 (chr21:42784603-42793518) overlapping *MX1* gene (exon 1; NM_001144925.2) in 11 patient (cases 13, 17, 18, 20, 21, 24, 25, 28, 4, 8, 38).

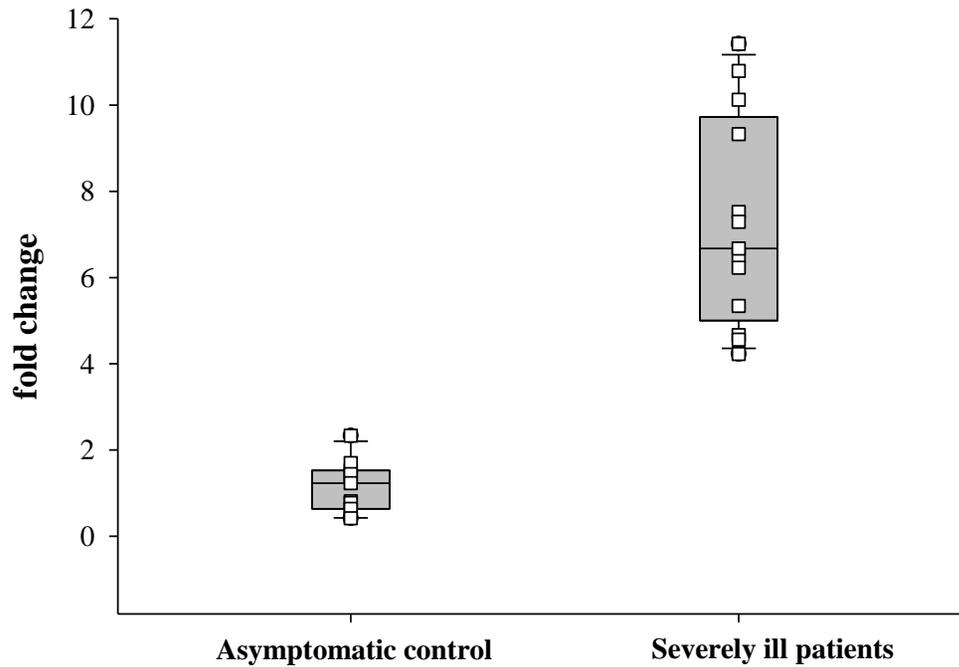


Figure 4: Expression analysis of *STK26*. The *STK26* gene transcripts were quantified in 11 asymptomatic (3 Male and 8 Female) and 12 severely ill patients (6 Male and 6 Female) with COVID-19, normalized against 18S RNA. The *STK26* transcripts were found to be significantly ($p < 0.001$) increased in severely ill patients compared to asymptomatic controls, with the highest expression observed in patient 44, which harbored the SV partially duplicating the *STK26* gene.

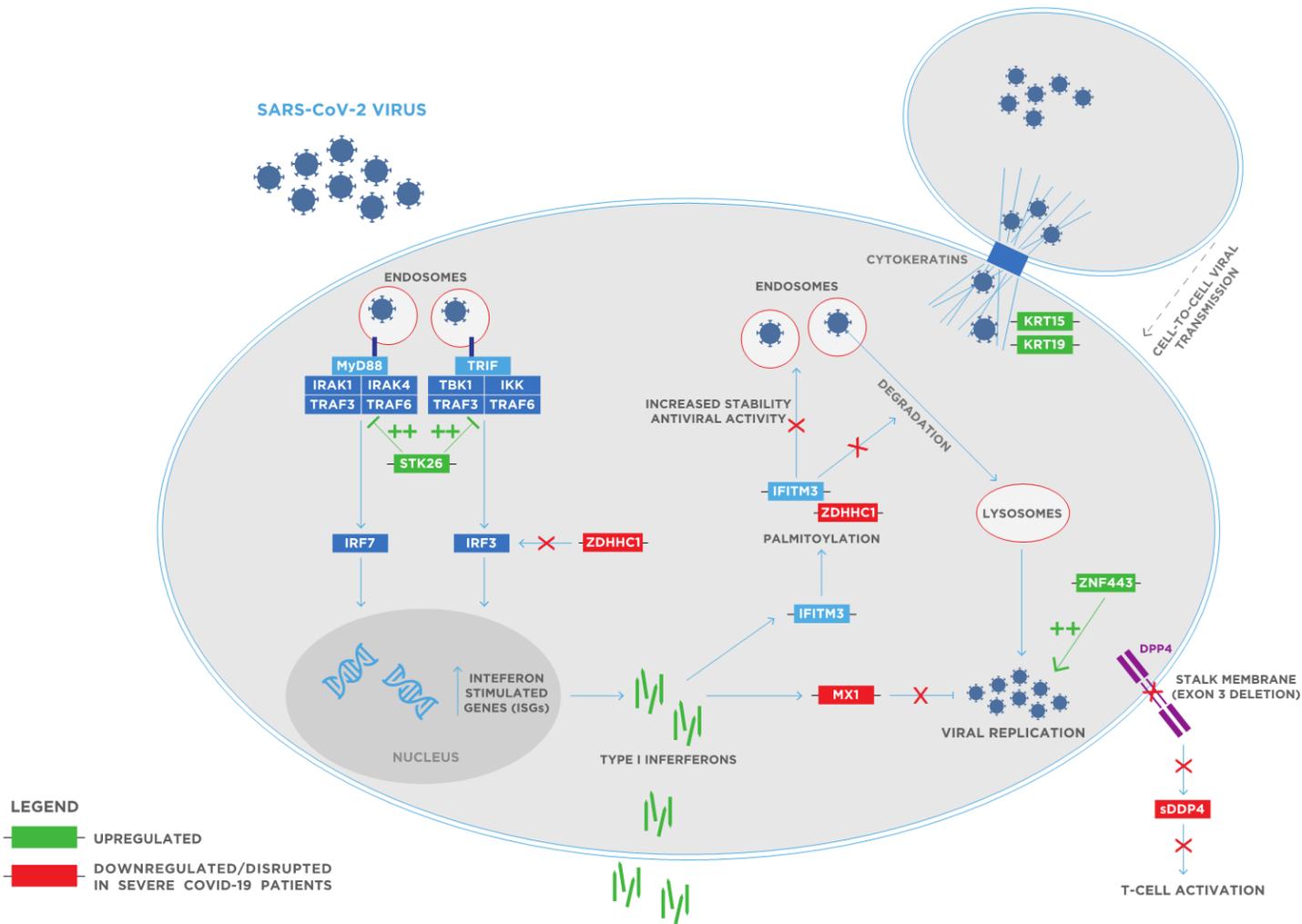


Figure 5: A diagram of the key viral host interaction pathways highlighting genes found to contain structural variation in COVID-19 patients.