



A host-based whole genome sequencing study reveals novel risk loci associated with severity of influenza A(H1N1)pdm09 infection

Mo Li^{a*}, Yongkun Chen^{b*}, Tao Chen^c, Shixiong Hu^d, Luan Chen^a, Lu Shen^a, Fangcai Li^d, Jing Yang^c, Yan Sun^e, Dayan Wang^c, Lin He^a, Shengying Qin^{a*} and Yuelong Shu^{b,c*}

^aBio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai, People's Republic of China; ^bSchool of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, People's Republic of China; ^cNational Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, People's Republic of China; ^dHunan Provincial Center for Disease Control and Prevention, Changsha, People's Republic of China; ^eChangsha Central Hospital, Changsha 410004, People's Republic of China

ABSTRACT

Influenza A(H1N1)pdm09 virus has remained in a seasonal circulation since being recognized in 2009. Although it followed a mild course in most patients, in others it caused a series of severe clinical illnesses. Epidemiologic studies have implicated that host factors have a major influence on the disease severity of influenza A(H1N1)pdm09 infection. However, an understanding of relevant genetic variations and the underlying mechanisms is still limited. In this present study, we used a host-based whole genome sequencing (WGS) method to comprehensively explore the genetic risk loci associated with severity of influenza A(H1N1)pdm09 infection. From the common single-nucleotide variants (SNVs) analysis, we identified the abnormal nominally significant ($P < 1 \times 10^{-4}$) common SNVs enriched in *PTBP3* gene. The results of rare functional SNVs analysis supported that there were several novel candidate genes might confer risk of severe influenza A(H1N1)pdm09 diseases, such as *FTSJ3*, *CPVL*, *BST2*, *NOD2* and *MAVS*. Moreover, our results of gene set based analysis indicated that the HIF-1 transcription factor and IFN- γ pathway might play an important role in the underlying mechanism of severe influenza A(H1N1)pdm09. These findings will increase our knowledge about biological mechanism underlying the severe influenza A(H1N1)pdm09 and facilitate to design novel personalized treatments.

ARTICLE HISTORY Received 2 October 2020; Revised 1 December 2020; Accepted 24 December 2020

KEYWORDS Influenza A(H1N1)pdm09; host disease severity; whole genome sequencing; hypoxia inducible factor-1; interferon gamma

Introduction

In April 2009, a novel influenza A(H1N1) virus [A(H1N1)pdm09] was recognized and quickly spread around the world. By June 2009, the World Health Organization (WHO) declared the start of the first twenty-first century influenza pandemic [1]. It was estimated that there were 123,000~203,000 pandemic respiratory deaths globally in the last 9 months of 2009, which was 10-fold higher than the laboratory-confirmed mortality count from WHO [2]. In contrast to seasonal influenza virus, influenza A(H1N1)pdm09 virus sustained a substantially higher human-to-human transmissibility [3] and caused more severe morbidity and mortality in children and young adults [1]. Some influenza A(H1N1)pdm09 infected patients required hospitalization because of the unusually severe pneumonia or acute respiratory distress

syndrome (ARDS), which was different from typical influenza symptoms [4]. To date, influenza A(H1N1)pdm09 virus still posed a threat to public health. Moreover, disease severity has the potential to alter, especially if virus mutation occur [5,6].

Thorough understanding of contributions to severe influenza A(H1N1)pdm09 pathogenesis is essential to advance the treatments and design of new vaccines. Earlier researches on the severity of influenza mainly focused on the viral virulence and properties [7]. It has been noticed that the host genetic variants may be related to the susceptibility of influenza A(H1N1)pdm09 virus infection among different populations [8–11]. Nevertheless, the host genetic background responsible for the development of severe influenza A(H1N1)pdm09 diseases have not been well defined. About one-quarter to one-half of patients

CONTACT Yuelong Shu shuylong@mail.sysu.edu.cn National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Prevention and Control, Beijing 102206, People's Republic of China; School of Public Health (Shenzhen), Sun Yat-sen University, Guangdong 510275, People's Republic of China; Shengying Qin chinsir@sjtu.edu.cn Bio-X Institutes, Shanghai Jiao Tong University, 209 Little White House, No.1954 Hua Shan Road, Shanghai 200030, People's Republic of China

*These authors contributed equally.

Supplemental data for this article can be accessed at <https://doi.org/10.1080/22221751.2020.1870412>

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

with severe influenza A(H1N1)pdm09 were previously healthy and did not have comorbid conditions, implicating that the inter-individual genetic variation may have accounted for the distinct disease severity of influenza A(H1N1)pdm09 infection [12,13].

Multiple immunity-related candidate genes have been identified to be risk factors of severe influenza A(H1N1)pdm09 infection, including *CCR5*, *KIR*, *IFITM3*, *IGHG2* [13]. The most widely studied is the *IFITM3* gene, which has been shown to play a role in restricting influenza virus replication in vitro and in vivo experiments [14,15]. Polymorphisms on *IL-6*, *TLR3*, *SFTPB*, *TNF*, *ST3GAL1* and *GLDC* genes have also been found to be associated with increased risk of influenza A(H1N1)pdm09 disease severity [16–21]. In addition, several small-scale genome wide association studies have reported a number of novel susceptibility genes of severe influenza A(H1N1)pdm09 outcomes [12,22–24]. Unfortunately, most of these results of the association between specific genetic loci and influenza A(H1N1)pdm09 disease severity are yet to be validated with enough power or repeatable evidences [25]. In general, an understanding of the host genetic contributions on severe influenza A(H1N1)pdm09 is still limited. A more comprehensive efficient study about host genetic background is required to detect the severity of influenza A(H1N1)pdm09 infection.

Recent development of next generation sequencing (NGS) technology has enabled scientists to made progress in research on the severity of infectious diseases. Whole genome sequencing (WGS) provides a great opportunity to deepen our understanding of the relationship between the influenza genome and disease severity [26]. However, few study was conducted to study the host-related genomic factors of severity of influenza A(H1N1)pdm09 infection using NGS.

In this study, we used an influenza A(H1N1)pdm09 host-based WGS approach to identify novel genetic risk loci associated with severity of influenza A(H1N1)pdm09 infection and explore its hidden mechanisms. We separately excavated the contributions of common SNVs and rare functional SNVs in the process of severe influenza A(H1N1)pdm09 infection. To the best of our knowledge, this is the first attempt to analyse genetic factors associated with severe influenza A(H1N1)pdm09 infection using a host-based WGS method. Our study will be helpful to improve the treatment of severe influenza A(H1N1)pdm09 infected patients and develop new vaccines.

Materials and methods

Study participants

We collected samples from 165 patients with influenza A(H1N1)pdm09 viral infection, which was confirmed

by positive results of real-time RT PCR. Mildly symptomatic patients were defined as outpatients who were not admitted to hospitals. Patients with severe disease were defined as hospitalized patients who met at least one of the criteria described in the “Protocol for diagnosis and treatment of influenza (2019 version)” [27]. Basic demographics information was retrieved from the clinical management system. The study was approved by the Ethics Committee of the National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (CDC), and patients enrolled in this study provided written informed consent to participate.

Whole genome sequencing

Whole-blood-derived genomic DNAs were extracted from all individuals using the QIAamp DNA blood Midi Kit (QIAGEN GmbH, Hilden, Germany) with standard protocols. The DNA purity was detected using NanoPhotometer spectrophotometer (IMPLEN, CA, USA). The DNA quantity was assessed using the Qubit Fluorometer (Life Technologies, CA, USA). Next, 1 ug of high quantity genomic DNA was fragmented. The fragments were then end repaired, poly-A tailed and adapter ligated using TruSeq DNA Sample Preparation Kit (Illumina, 15026486 Rev.C), according to the manufacturer’s instructions. Adapter-ligated libraries were amplified by 6 cycles of PCR. After that, libraries were assessed for sequencing using Agilent 2100 Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, Santa Clara, CA, USA). Next, the qualified libraries were sequenced using Illumina HiSeq X Ten sequencer from Illumina (Illumina, San Diego, CA, USA) with 2 × 150 bp paired-end reads.

Data processing

The sequencing reads of the 165 samples were mapped to the human reference genome (hg19) using the Burrows–Wheeler Aligner (version 0.7.15, <http://bio-bwa.sourceforge.net/>). After the alignment, duplicate reads were marked using Picard (version 2.15.0, <http://picard.sourceforge.net/>). Insertion/deletion (indel) realignment, base quality score recalibration, and variant calling were performed using Genome Analysis Tool Kit (GATK, version 3.8, <https://software.broadinstitute.org/gatk/>) software. According to the GATK best-practices protocol, we merged individual gVCF files into a combined variant calling format (VCF) and ran variant filtering using variant quality score recalibration. Then, we selected “PASS” single-nucleotide variants (SNV) and further strictly controlled the quality of these SNVs to reach the following criteria: genotype quality >10.0 and approximate read depth >20.0. Finally, we used VCFtools (<http://vcftools.sourceforge.net/>) software to convert VCF

format file with high quality SNVs into PLINK format files for further analyses. All subsequent analyses in this study were performed only with biallelic SNVs.

Single-nucleotide variant association analyses

The main SNV association analyses consists of common variants and several rare variants analyses for a connection with severity of influenza A (H1N1)pdm09 infection risk. Quality control (QC) procedures for samples and SNVs were performed as described in previous study [28]. We also excluded close relative samples by detecting the kinship of the samples used PLINK software (v1.9, <http://zzz.bwh.harvard.edu/plink/plink2.shtml>) [29]. In the end, we prepared a set of clean SNVs for subsequent analyses.

PLINK software was first used to perform a logistic regression association analysis of all observed common SNVs (minor allele frequency $\geq 1\%$), correcting for the first 5 principal components (PCs) of genetic ancestry. The genome-wide significance threshold for multiple test correction was set at 5×10^{-8} . Furthermore, Genome-wide Complex Trait Analysis (GCTA, v1.91.6beta, <https://cns.genomics.com/software/gcta/#Overview>) software [30] was used to perform the chi-square tests of subgroup common SNVs on a gene ± 50 kb of UTRs or a 100Kb segment region. A standard Bonferroni correction would yield a significance level of $\sim 2 \times 10^{-6}$ for the common SNVs set based tests.

For rare functional variants analysis, we defined rare SNVs as minor allele frequency $< 1\%$ in East Asian population from multiple public databases, including 1000 Genomes Project (release in 2013 May), Exome Aggregation Consortium (ExAC r0.3.1) and Genome Aggregation Database (exome samples data). Then, we classified those coding rare SNVs based on their functional annotation: loss of function (LOF, including start loss, stop loss and splicing site) and missense SNVs. From these annotations, we created two rare variants sets for single gene testing: (1) all functional SNVs (both LOF and missense), (2) only LOF SNVs. Both analyses mentioned above were used optimal sequencing kernel association test (SKAT-O, with parameter: number permutation = 10,000, $\alpha = 0.05$, $\beta_1 = 1$, $\beta_2 = 20$) in KGGseq software (v1.0) [31]. We defined the genes with $P < 0.05$ statistically significance as candidate loci. In order to identify most potential risk loci, we used influenza A pathway genes from KEGG database (hsa05164) as training gene set and performed a further ranking analysis of all the candidate genes using the ToppGene software [32]. ToppGene is a one-stop portal for gene list enrichment analysis and candidate gene prioritization based on their functional annotations and protein interactions network. In addition, we used STRING software (v 11.0, <https://string-db.org/>) [33]

to build the protein interaction networks (combined score > 0.4) of these candidate genes. Then, Cytoscape (v3.5.1) software [34] was used to visualize the largest network diagram and speculate the hub genes through the cytoHubba plug-in.

At lastly, we also conducted gene set based testing of pathways using all rare functional SNVs. The used gene sets were downloaded from Molecular Signatures Database (MSigDB) (v7.1, <https://www.gsea-msigdb.org/gsea/downloads.jsp>) [35]: (1) canonical pathways collection contain 2232 gene sets that are canonical representations of a biological process compiled by domain experts; (2) gene ontology collection generated with 10192 gene sets that are derived from Gene Ontology (GO) annotations including molecular function (MF), cellular component (CC) and biological process (BP). The analyses strategy, as described above, used the same method as that single gene testing.

Statistical analyses

The collected demographic information was evaluated by Mann–Whitney U test or Fisher exact test using SPSS 17.0, and differences were considered statistically significant at $P \leq 0.05$. We performed a principal component analysis (PCA) using PLINK (v1.9). To determine which common SNVs were independently associated with severe influenza A(H1N1)pdm09 disease, a logistics regression analysis was utilized to control for the first 5 PCs of ancestry using PLINK (v1.9). To detect the joint effect of common SNVs, a chi-square test was implemented to incorporate the subgroup common SNVs within one gene or segment region using GCTA (v1.91.6beta). The relationships between the rare functional SNVs in single gene or gene set and severe influenza A(H1N1)pdm09 disease were specifically assessed with SKAT-O models using KGGseq (v1.0).

Results

Descriptive characteristics of patients and SNVs

Basic characteristics of patients are shown in Table 1 and the study design is shown in Figure 1. In short, 70 severe and 95 mild influenza A(H1N1)pdm09 infected patients were enrolled. The mean age of the

Table 1. Basic characteristics of patients.

Variable	Disease classification		P-value ^a
	Severe (n = 70)	Mild (n = 95)	
Age, median years (interquartile range)	3 (1–8.75)	4.5 (3–7)	0.053
Age ≤ 13 years	55 (78.57%)	82 (86.32%)	0.212
Age ≥ 65 years	6 (8.57%)	2 (2.11%)	0.072
Sex male	51 (72.86%)	67 (70.53%)	0.862

^aThe Fisher exact test and Mann–Whitney U test were used for categorical and continuous variables, respectively.

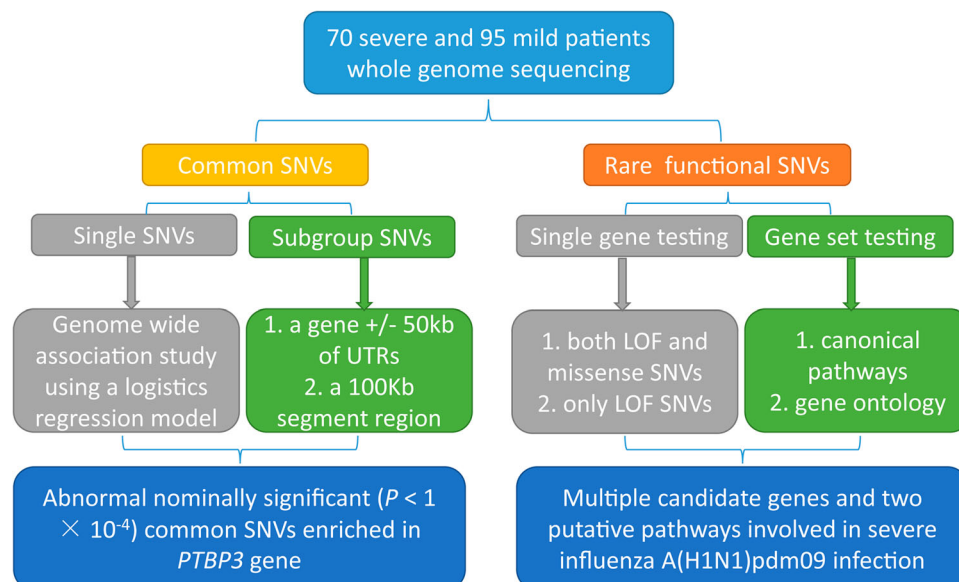


Figure 1. The study design.

In this study, we enrolled and sequenced 70 severe and 95 mild influenza A(H1N1)pdm09 virus infected patients. Then, we systematically evaluated host-related genomic differences between the severe patient group and mild patient group using common single-nucleotide variants (SNVs) and rare functional SNVs. Through a series of association analyses and bioinformatics interpretation, we have identified multiple candidate genes and two putative pathways involved in severe influenza A(H1N1)pdm09 infection.

all patients was 12 years; 71.5% (118/165) were male. No significant differences were detected in the demographic variables (ages, age groups and sex proportion) between the severe and mild patient groups.

These sequence data had a 99.9% median alignment rate, 12.4% median duplication rate, and $34.9 \times$ median coverage per individual. After variant calling and QC, only PCA removed 1 severe patient, and there were about 12.38 million SNVs retained. Among these SNVs, there were about 7.18 million common SNVs.

Abnormal nominally significant common SNVs enriched in *PTBP3* gene

To determine whether common SNVs were independently associated with severe influenza A (H1N1)pdm09 disease, a genome wide association study was carried out using a logistics regression model. Figure 2 displayed the Manhattan and quantile–quantile plots of this analysis. Although no genome-wide significant associations were identified from the common SNVs analysis, we tried to seek

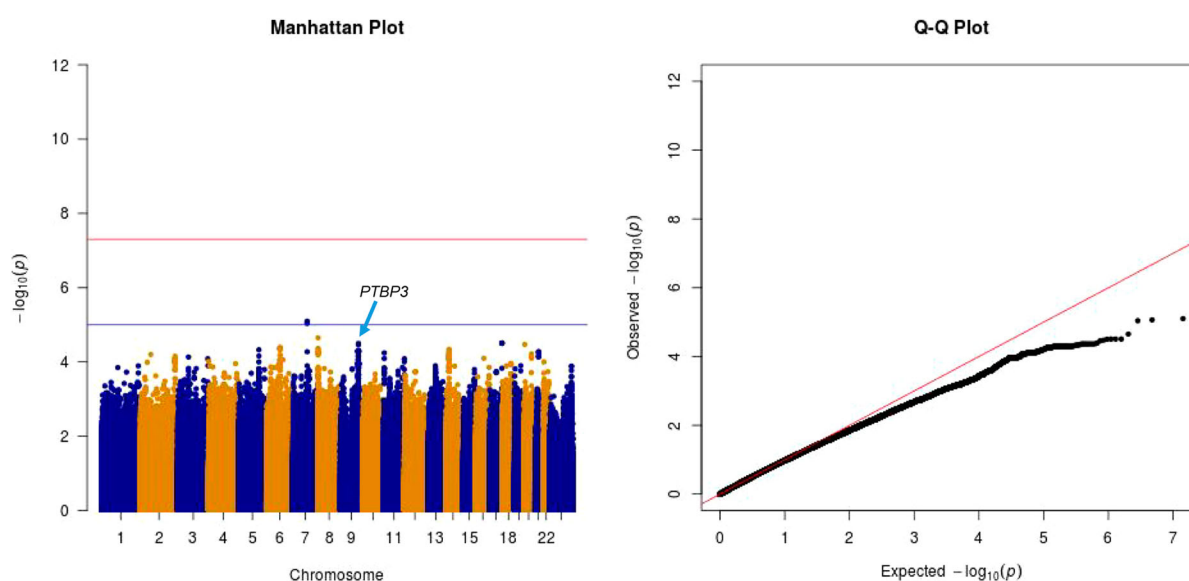


Figure 2. The Manhattan and quantile–quantile plots from the common SNVs analysis.

The genome wide association study was carried out with a logistics regression model for common SNVs. In the Manhattan plot, the blue line means P value is 1×10^{-5} and red line means P value is 5×10^{-8} . The Q-Q plot didn't show any signs of global inflation of test statistics. Although no genome-wide significant associations were identified from the common SNVs analysis, using a less stringent level of 1×10^{-4} , 27 SNPs (represented by an intron variant rs4634725 with $P = 3.16 \times 10^{-5}$) locating on *PTBP3* gene were identified.

some potential risk SNVs. Using a less stringent level ($P < 1 \times 10^{-4}$) [22], there are 164 SNVs nominally significant associated with severe influenza A (H1N1)pdm09 risk (Table S1). Among them, an intron variant rs4634725 ($P = 3.16 \times 10^{-5}$) was the most significant SNV locating on a protein coding gene, that is polypyrimidine tract-binding protein 3 (*PTBP3*). Additionally, another 26 nominally significant associated SNVs locating on *PTBP3* gene were also identified, and they have a strong linkage relationship, as shown in Fig. S1.

In order to understand the joint effect of common SNVs, we performed gene and segment based analyses between severe and mild patient groups. Because our sample size was limited, the Bonferroni correction caused very conservative results in this test. So, a less stringent significant threshold was set at $P < 1 \times 10^{-4}$. The gene based analysis showed that only liver expressed antimicrobial peptide 2 (*LEAP2*) gene suggestively associated ($P = 5.26 \times 10^{-5}$) with severe influenza A(H1N1)pdm09 disease. In addition, the segment based analysis indicated that the nearby region of *PTBP3* gene appeared to have the statistically significant ($P = 3.92 \times 10^{-5}$) association signal.

Association of candidate genes with severity of influenza A(H1N1)pdm09 infection

To evaluate the contribution of rare SNVs to the severity of influenza A(H1N1)pdm09 infection on a single gene level, we tested which genes associated with severity of influenza A(H1N1)pdm09 infection. Two kinds of tests were conducted using the all rare functional SNVs or only rare LOF SNVs. The rare functional SNVs test identified a total of 125 genes as significant candidate loci ($P < 0.05$) being presented in Table S2. Among them, FtsJ RNA 2'-O-methyltransferase 3 (*FTSJ3*) gene appeared to be the most significant ($P = 3.61 \times 10^{-3}$) risk gene of the severity of influenza A(H1N1)pdm09 infection. On the other hand, the rare LOF SNVs test discovered 12 genes that were significantly ($P < 0.05$) associated with severe influenza A(H1N1)pdm09 disease as shown in Table S2. Carboxypeptidase vitellogenic like (*CPVL*) and chimerin 2 (*CHN2*) gene seemed to be the most significant risk gene ($P = 7.62 \times 10^{-3}$) in this test. It was noted that 6 genes were found to be significant in both two tests as mentioned above, including *CPVL*.

Bioinformatics interpretations of candidate genes identified by rare functional SNVs

The 131 candidate genes were priority ranked based on their annotations and networks through ToppGene software so that we could dive deep into the biological function and interrelation of these candidate genes identified by single gene testing. Because the candidate

gene mitochondrial antiviral signaling protein (*MAVS*) was included in the training gene set, it was automatically excluded from the priority analysis. The ranked result (Table S3) showed that the nucleotide-binding oligomerization domain 2 (*NOD2*) gene had the highest priority, followed by unc-51 like autophagy activating kinase 1 (*ULK1*), interleukin-10 (*IL-10*) and bone marrow stromal cell antigen 2 (*BST2*) etc.

Besides, we speculated on the protein interaction networks of the 131 candidate genes through STRING software and described the largest independent protein network comprising of 31 candidate genes using Cytoscape software (Figure 3). Then, we used 12 topology analysis methods to detect the top 5 hub genes of this largest network by cytoHubba plug-in respectively. We found that both *NOD2* and *MAVS* gene appeared in 8 methods with most frequently (Table S4).

Putative pathways involved in severity of influenza A(H1N1)pdm09 infection

To illustrate which specific pathways may involve in severity of influenza A(H1N1)pdm09 infection, we downloaded two classic collections of gene sets from MSigDB. Overall, we detected some pathways statistically significantly ($P < 0.05$) related with severity of influenza A(H1N1)pdm09 infection. The results of canonical pathways testing suggested that hypoxia inducible factor-1 transcription factor (HIF-1 TF) pathway was the most significant ($P = 1.89 \times 10^{-3}$), indicating it might play an important role in severe influenza A(H1N1)pdm09 disease (Table 2). The GO testing revealed that positive regulation of response to interferon gamma (IFN- γ) and dendritic cell cytokine production terms were significantly ($P = 5.4 \times 10^{-4}$ and $P = 7.75 \times 10^{-4}$ respectively) associated with severity of influenza A(H1N1)pdm09 infection (Table 3).

Discussion

The seasonal spread of influenza A(H1N1)pdm09 virus remains the main subtype in intensive care patients [36], which alerted us to speed up research on severe viral infection. In this study, we systematically studied the molecular mechanism underlying the disease severity of influenza A(H1N1)pdm09 infection using a host-based WGS method. The results revealed several novel host-related candidate genes and putative pathways associated with severity of influenza A(H1N1)pdm09 infection from different categories, including common SNVs and rare functional SNVs.

Genome wide association study of common SNVs discovered that multiple common SNVs locating on

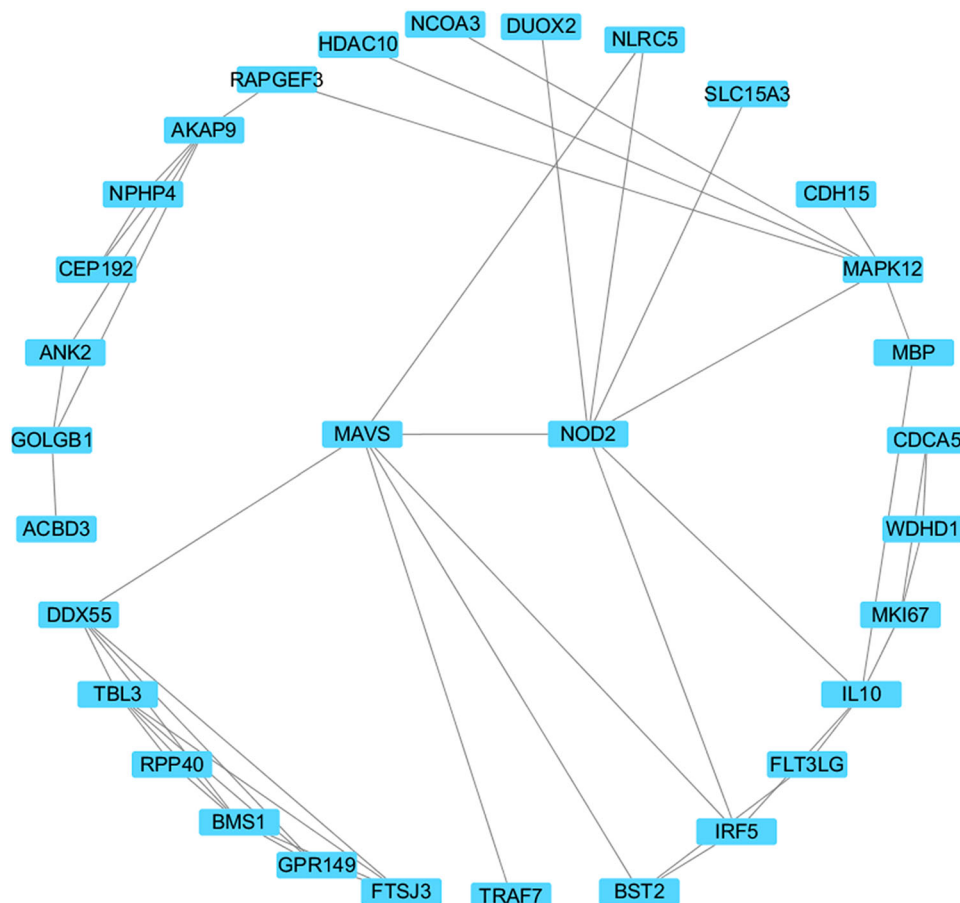


Figure 3. The largest independent protein interaction network of candidate genes.

We founded 131 candidate genes by single gene based testing of rare functional SNVs and speculated protein interaction networks through STRING software (v 11.0, <https://string-db.org/>). Cytoscape software (v3.5.1) was used to describe the largest independent protein interaction network comprising of 31 candidate genes. NOD2 and MAVS were identified as hub genes with the highest frequency through the cytoHubba plug-in.

PTBP3 gene had significant differences in allele frequency between severe and mild influenza A (H1N1)pdm09 infected patient groups. The gene and segment based analysis of the joint effect of common SNVs emphasized the importance of the *PTBP3* and *LEAP2* gene. *PTBP3* belongs to the PTB family of RNA binding protein and is a regulator of cell differentiation. Interestingly, *PTBP3* protein has been found to be highly restricted to B cells where it played pivotal role in regulating antibody diversification [37]. An earlier study has shown that *PTBP3*

protein expression was increased in human colorectal cancer through directly binding to 5'UTR HIF-1 α mRNA and enhanced HIF-1 α protein expression [38]. On the other hand, *LEAP2* encodes a cysteine-rich cationic antimicrobial peptide that is expressed predominantly in the liver and plays a vital role in host immunity. *LEAP2* has been suggested to be related to a potential mechanism of microbial translocation and subsequent immune activation in HIV-

Table 2. The top 10 significant pathways of gene set analysis using rare functional SNVs.

Gene set	P-value ^a
PID: HIF1 TF PATHWAY	0.00189
KEGG: LEISHMANIA INFECTION	0.00443
Biocarta: MPR PATHWAY	0.00677
Reactome: SIGNALING BY EGFR	0.00759
KEGG: ECM RECEPTOR INTERACTION	0.00846
Reactome: TRANSCRIPTIONAL REGULATION BY THE AP 2 TFAP2 FAMILY OF TRANSCRIPTION FACTORS	0.00904
Biocarta: LIS1 PATHWAY	0.0104
Biocarta: MELANOCYTE PATHWAY	0.0122
Reactome: NR1H2 NR1H3 REGULATE GENE EXPRESSION LINKED TO GLUCONEOGENESIS	0.0122
Reactome: TRAFFICKING OF AMPA RECEPTORS	0.0136

^aThe optimal sequencing kernel association test (SKAT-O) was used for this gene set analysis.

Table 3. The top 10 significant GO terms of gene set analysis using rare functional SNVs.

Gene set	P-value ^a
GO_MF: JUN KINASE BINDING	0.000358
GO_BP: POSITIVE REGULATION OF RESPONSE TO INTERFERON GAMMA	0.000540
GO_BP: DENDRITIC CELL CYTOKINE PRODUCTION	0.000775
GO_BP: POSITIVE REGULATION OF METALLOENDOPEPTIDASE ACTIVITY	0.00142
GO_BP: RELEASE OF SEQUESTERED CALCIUM ION INTO CYTOSOL BY ENDOPLASMIC RETICULUM	0.00164
GO_MF: RRNA GUANINE METHYLTRANSFERASE ACTIVITY	0.00174
GO_BP: NEGATIVE REGULATION OF DNA TEMPLATED TRANSCRIPTION ELONGATION	0.00210
GO_MF: 3',5'-CYCLIC-GMP PHOSPHODIESTERASE ACTIVITY	0.00216
GO_BP: AMINE METABOLIC PROCESS	0.00249
GO_BP: POSITIVE REGULATION OF DENDRITIC CELL CYTOKINE PRODUCTION	0.00259

^aThe optimal sequencing kernel association test (SKAT-O) was used for this gene set analysis.

infected and HCV-infected patients [39]. In general, these discoveries showed that common SNVs might play an important role that could not be ignored in severe influenza A(H1N1)pdm09.

Analysis of rare functional SNVs showed that the *FTSJ3* gene was most significantly associated with severe influenza A(H1N1)pdm09 disease. *FTSJ3* protein is an RNA 2'-O-methyltransferase recruited by HIV-1 to evade innate immune recognition [40]. In addition, Ribose 2'-O-methylation provides a molecular signature by which the cellular innate immune system distinguishes self from non-self mRNA [41]. In the rare SNVs analysis that only considered LOF SNVs, we found that the *CPVL* gene may contribute to the emergence of severe influenza A(H1N1)pdm09 disease. *CPVL* protein has been found to be involved in an inflammatory protease cascade, and trimming of peptides for antigen presentation [42]. The results of the candidate genes priority analysis suggested that *NOD2* was the most significant risk gene, followed by *ULK1*, *IL10* and *BST2* etc. In the protein interaction analysis, we further analysed these candidate genes mutual relationship and discovered that *NOD2* and *MAVS* were most valuable hub genes. From the perspective of virus life cycle, we further investigated the role of these candidate genes in the antiviral process. The *ULK1* complex initiates autophagosome formation, linking cellular nutrient status to downstream events in autophagy [43]. An earlier study suggested that SNPs in the *IL10* gene might be associated with disease severity in influenza A(H1N1)pdm09 infected patients [44]. *BST2* protein has been described to inhibit the virus replication by tethering the virus release from the cell surface [45] or by inducing the NF- κ B dependent antiviral immune response [46] or by targeting and degrading virus nucleocapsid protein [47]. As a cytosolic sensor, *NOD2* contribute to host defense against microbial infection in human through an IL-32-dependent dendritic cells differentiation pathway [48]. A recent study has shown that influenza A M2 protein colocalized and interacted with *MAVS* on mitochondria, and positively regulated *MAVS*-mediated innate immunity [49]. Taken together, we successfully identified several novel candidate genes might confer risk for severe influenza A(H1N1)pdm09 infection through changing the activity of host immunity or interfering the survival of influenza viruses. Functional experiments are worthy to validate these findings in the future.

Canonical pathway gene sets analysis of all rare functional SNVs revealed that the HIF-1 TF pathway had a significant correlation with severe influenza A(H1N1)pdm09 disease. A recent study implied that in vitro influenza A(H1N1)pdm09 induced nuclear translocation of HIF-1 α without altering the expression of HIF-1 α , which may promote the

secretion of proinflammatory cytokines during influenza A(H1N1)pdm09 infection [50]. It is well-known that the over-secretion of inflammatory cytokines is considered to be a key contributor to the severity of influenza A(H1N1)pdm09 infection [51–53]. More importantly, another study has confirmed that deficiency of HIF-1 α enhanced influenza A virus replication *in vitro* by promoting autophagy [54]. Therefore, further study is needed be conducted to compare host HIF-1 TF related variants and study the true influence of nuclear translocation of HIF-1 α .

GO gene sets analysis results suggested the impact of positive regulation of response to IFN- γ on severe influenza A(H1N1)pdm09 disease. There is a study supported the statement that IFN- γ have a detrimental role in the pathogenesis of influenza A through a restriction in lymphoid cell group II activity [55]. Additionally, another study has been demonstrated that thymic innate CD8(+)CD44(hi) single-positive T-cells have critical roles in severe influenza A(H1N1)pdm09 infection induced thymic atrophy through secreting IFN- γ [56]. Surprisingly, rare putative LOF variants of X-chromosomal *TLR7* were identified that were associated with impaired IFN- γ responses in a case study of 4 young male patients with severe coronavirus disease 2019 (COVID-19) [57]. Therefore, the differential levels of IFN- γ expression derived from its correlative genomic variants may influence specific immune cells viability and eventually dictate the disease outcome. In consistent with our findings, one recent study has suggested that IFN- γ played an important role in acute lung injury induced by severe influenza A(H1N1)pdm09 infection, and monoclonal antibodies against IFN- γ could be useful as a potential therapeutic remedy for future influenza pandemics [58].

There exist limitations in our study. First, the sample size we used in this study was relatively small. The number of samples limited the statistical power of the analysis performed in this study. Another limitation is that the age distribution was not normal. Most patients were aged under 13 years old, but there was no distinct bias between the severe and mild patient groups. In fact, this is also consistent with the feature that the influenza A(H1N1)pdm09 caused more severe morbidity and mortality in children and young adults. Therefore, the range of our results may be limited in explaining the severe influenza A(H1N1)pdm09 infection mechanism.

In summary, we used an efficient host-based WGS method to comprehensively evaluate the host genetic contributions on the occurrence mechanism of severe influenza A(H1N1)pdm09. Through such strategy, we revealed several novel candidate genes and pathways might play an important role in the process of severe influenza A(H1N1)pdm09 infection. These findings will facilitate us to expand our knowledge about the

mechanisms of severe influenza A(H1N1)pdm09 development and provide a theoretical basis for the design of new therapeutic strategy for patients with severe influenza. Further functional studies on these candidate genes are worthy to identify susceptibility marker for the severe influenza A(H1N1)pdm09.

Acknowledgements

The authors thank the patients who participated in this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Shenzhen Science and Technology Program (KQTD20180411143323605), National Science Fund for Distinguished Young Scholars (81525017), National key research and development program (2016YFC1200202, 2016YFC1200203), National Mega-projects for Infectious Diseases (2018ZX10305409-004-003), the 4th Three-year Action Plan for Public Health of Shanghai (15GWZK0101).

ORCID

Mo Li  <http://orcid.org/0000-0002-9915-1348>

Shengying Qin  <http://orcid.org/0000-0002-8458-5960>

References

- [1] Josset L, Engelmann F, Haberthur K, et al. Increased viral loads and exacerbated innate host responses in aged macaques infected with the 2009 pandemic H1N1 influenza A virus. *J Virol*. 2012;86(20):11115–11127.
- [2] Simonsen L, Spreeuwenberg P, Lustig R, et al. Global mortality estimates for the 2009 influenza pandemic from the GLaMOR project: a modeling study. *PLoS Med*. 2013;10(11):e1001558.
- [3] Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* (New York, NY). 2009;324(5934):1557–1561.
- [4] Munster VJ, de Wit E, van den Brand JM, et al. Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. *Science* (New York, NY). 2009;325(5939):481–483.
- [5] Patel M, Dennis A, Flutter C, et al. Pandemic (H1N1) 2009 influenza. *Br J Anaesth*. 2010;104(2):128–142.
- [6] Chen Y, Zhu W, Bai T, et al. The substitution V379I in PA protein attenuates the pathogenicity of influenza A (H1N1) pdm09 viruses in mice. *Sci China Life Sci*. 2017;60(9):1044–1046.
- [7] Trammell RA, Liberati TA, Toth LA. Host genetic background and the innate inflammatory response of lung to influenza virus. *Microbes Infect*. 2012;14(1):50–58.
- [8] Morales-García G, Falfán-Valencia R, García-Ramírez RA, et al. Pandemic influenza A/H1N1 virus infection and TNF, LTA, IL1B, IL6, IL8, and CCL polymorphisms in Mexican population: a case-control study. *BMC Infect Dis*. 2012;12:1–9.
- [9] Liu Y, Li S, Zhang G, et al. Genetic variants in IL1A and IL1B contribute to the susceptibility to 2009 pandemic H1N1 influenza A virus. *BMC Immunol*. 2013;14:1–10.
- [10] Falfán-Valencia R, Narayanankutty A, Reséndiz-Hernández JM, et al. An increased frequency in HLA class I alleles and haplotypes suggests genetic susceptibility to influenza A (H1N1) 2009 pandemic: a case-control study. *J Immunol Res*. 2018;2018:1–12.
- [11] Elsayed SM, Hassanein OM, Hassan NHA. Influenza A (H1N1) virus infection and TNF-308, IL6, and IL8 polymorphisms in Egyptian population: a case-control study. *J Basic Appl Zool*. 2019;80(1):1–6.
- [12] Zhou J, To KK, Dong H, et al. A functional variation in CD55 increases the severity of 2009 pandemic H1N1 influenza A virus infection. *J Infect Dis*. 2012;206(4):495–503.
- [13] Keynan Y, Malik S, Fowke KR. The role of polymorphisms in host immune genes in determining the severity of respiratory illness caused by pandemic H1N1 influenza. *Public Health Genomics*. 2013;16(1-2):9–16.
- [14] Everitt AR, Clare S, Pertel T, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature*. 2012;484(7395):519–523.
- [15] Allen EK, Randolph AG, Bhangale T, et al. SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nat Med*. 2017;23(8):975–983.
- [16] Paquette SG, Banner D, Zhao Z, et al. Interleukin-6 is a potential biomarker for severe pandemic H1N1 influenza A infection. *PloS one*. 2012;7(6):e38214.
- [17] Esposito S, Molteni CG, Giliani S, et al. Toll-like receptor 3 gene polymorphisms and severity of pandemic A/H1N1/2009 influenza in otherwise healthy children. *Virol J*. 2012;9:1–8.
- [18] To KKW, Zhou J, Song YQ, et al. Surfactant protein B gene polymorphism is associated with severe influenza. *Chest*. 2014;145(6):1237–1243.
- [19] García-Ramírez RA, Ramírez-Venegas A, Quintana-Carrillo R, et al. IL6, and IL1B Polymorphisms Are associated with severe influenza A (H1N1) virus infection in the Mexican population. *PloS one*. 2015;10(12):e0144832.
- [20] Maestri A, Sortica VA, Tovo-Rodrigues L, et al. Sial2-3Galβ1-receptor genetic variants Are associated with influenza A(H1N1)pdm09 severity. *PloS one*. 2015;10(10):e0139681.
- [21] Zhou J, Wang D, Wong BH, et al. Identification and characterization of GLDC as host susceptibility gene to severe influenza. *EMBO Mol Med*. 2019;11(1):1–14.
- [22] Zúñiga J, Buendía-Roldán I, Zhao Y, et al. Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. *Eur Respir J*. 2012;39(3):604–610.
- [23] Cheng Z, Zhou J, To KK, et al. Identification of TMPRSS2 as a susceptibility gene for severe 2009 pandemic A(H1N1) influenza and A(H7N9) influenza. *J Infect Dis*. 2015;212(8):1214–1221.
- [24] Garcia-Etxebarria K, Bracho MA, Galán JC, et al. No major host genetic risk factor Contributed to A (H1N1)2009 influenza severity. *PloS one*. 2015;10(9):e0135983.
- [25] To KK, Zhou J, Chan JF, et al. Host genes and influenza pathogenesis in humans: an emerging paradigm. *Curr Opin Virol*. 2015;14:7–15.

- [26] Goldstein EJ, Harvey WT, Wilkie GS, et al. Integrating patient and whole-genome sequencing data to provide insights into the epidemiology of seasonal influenza A (H3N2) viruses. *Microbial Genomics*. 2018;4(1):e000137.
- [27] National Health Commission of the People's Republic of China. Protocol for diagnosis and treatment of influenza (2019 version). *Chin J Clin Infectious Diseases*. 2019;12(6):451–455.
- [28] Li M, Shen L, Chen L, et al. Novel genetic susceptibility loci identified by family based whole exome sequencing in Han Chinese schizophrenia patients. *Transl Psychiatry*. 2020;10(1):1–10.
- [29] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–575.
- [30] Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
- [31] Li M, Li J, Li MJ, et al. Robust and rapid algorithms facilitate large-scale whole genome sequencing downstream analysis in an integrative framework. *Nucleic Acids Res*. 2017;45(9):e75.
- [32] Chen J, Bardes EE, Aronow BJ, et al. Toppgene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009;37:W305–W311.
- [33] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–D613.
- [34] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504.
- [35] Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739–1740.
- [36] Sarda C, Palma P, Rello J. Severe influenza: overview in critically ill patients. *Curr Opin Crit Care*. 2019;25(5):449–457.
- [37] Chen J, Cai Z, Bai M, et al. The RNA-binding protein ROD1/PTBP3 cotranscriptionally defines AID-loading sites to mediate antibody class switch in mammalian genomes. *Cell Res*. 2018;28(10):981–995.
- [38] Hou P, Chen F, Yong H, et al. PTBP3 contributes to colorectal cancer growth and metastasis via translational activation of HIF-1 α . *J Exp Clin Cancer Res*. 2019;38(1):1–14.
- [39] Shata MT, Abdel-Hameed EA, Hetta HF, et al. Immune activation in HIV/HCV-infected patients is associated with low-level expression of liver expressed antimicrobial peptide-2 (LEAP-2). *J Clin Pathol*. 2013;66(11):967–975.
- [40] Ringgaard M, Marchand V, Decroly E, et al. FTSJ3 is an RNA 2'-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature*. 2019;565(7740):500–504.
- [41] Züst R, Cervantes-Barragan L, Habjan M, et al. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat Immunol*. 2011;12(2):137–143.
- [42] Mahoney JA, Ntlosi B, DaSilva RP, et al. Cloning and characterization of CPVL, a novel serine carboxypeptidase, from human macrophages. *Genomics*. 2001;72(3):243–251.
- [43] Lin MG, Hurley JH. Structure and function of the ULK1 complex in autophagy. *Curr Opin Cell Biol*. 2016;39:61–68.
- [44] Choudhary ML, Alagarasu K, Chaudhary U, et al. Association of single nucleotide polymorphisms in TNFA and IL10 genes with disease severity in influenza A/H1N1pdm09 virus infections: a study from Western India. *Viral Immunol*. 2018;31(10):683–688.
- [45] Van Damme N, Goff D, Katsura C, et al. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe*. 2008;3(4):245–252.
- [46] Tokarev A, Suarez M, Kwan W, et al. Stimulation of NF- κ B activity by the HIV restriction factor BST2. *J Virol*. 2013;87(4):2046–2057.
- [47] Kong N, Shan T, Wang H, et al. BST2 suppresses porcine epidemic diarrhea virus replication by targeting and degrading virus nucleocapsid protein with selective autophagy. *Autophagy*. 2020;16(10):1737–1752.
- [48] Schenk M, Krutzik SR, Sieling PA, et al. NOD2 triggers an interleukin-32-dependent human dendritic cell program in leprosy. *Nat Med*. 2012;18(4):555–563.
- [49] Wang R, Zhu Y, Lin X, et al. Influenza M2 protein regulates MAVS-mediated signaling pathway through interacting with MAVS and increasing ROS production. *Autophagy*. 2019;15(7):1163–1181.
- [50] Guo X, Zhu Z, Zhang W, et al. Nuclear translocation of HIF-1 α induced by influenza A (H1N1) infection is critical to the production of proinflammatory cytokines. *Emerg Microbes Infect*. 2017;6(5):e39.
- [51] Thomas M, Mani RS, Philip M, et al. Proinflammatory chemokines are major mediators of exuberant immune response associated with influenza A (H1N1) pdm09 virus infection. *J Med Virol*. 2017;89(8):1373–1381.
- [52] Bradley-Stewart A, Jolly L, Adamson W, et al. Cytokine responses in patients with mild or severe influenza A(H1N1)pdm09. *J Clin Virol Off Publ Pan Am Soc Clin Virol*. 2013;58(1):100–107.
- [53] Wang Q, Fang P, He R, et al. O-GlcNAc transferase promotes influenza A virus-induced cytokine storm by targeting interferon regulatory factor-5. *Sci Adv*. 2020;6(16):eaaz7086.
- [54] Zhao C, Chen J, Cheng L, et al. Deficiency of HIF-1 α enhances influenza A virus replication by promoting autophagy in alveolar type II epithelial cells. *Emerg Microbes Infect*. 2020;9(1):691–706.
- [55] Califano D, Furuya Y, Roberts S, et al. IFN- γ increases susceptibility to influenza A infection through suppression of group II innate lymphoid cells. *Mucosal Immunol*. 2018;11(1):209–219.
- [56] Liu B, Zhang X, Deng W, et al. Severe influenza A (H1N1)pdm09 infection induces thymic atrophy through activating innate CD8(+)CD44(hi) T cells by upregulating IFN- γ . *Cell Death Dis*. 2014;5(10):e1440.
- [57] van der Made CI, Simons A, Schuurs-Hoeijmakers J, et al. Presence of genetic variants Among young Men With severe COVID-19. *JAMA*. 2020;324(7):1–11.
- [58] Liu B, Bao L, Wang L, et al. Anti-IFN- γ therapy alleviates acute lung injury induced by severe influenza A (H1N1) pdm09 infection in mice. *J Microbiol Immunol Infect*. 2019;S1684-1182(18):1–8.