

# Identification of *TMPRSS2* as a Susceptibility Gene for Severe 2009 Pandemic A(H1N1) Influenza and A(H7N9) Influenza

Zhongsan Cheng,<sup>1,a</sup> Jie Zhou,<sup>1,2,3,a</sup> Kelvin Kai-Wang To,<sup>1,2,3,4</sup> Hin Chu,<sup>1,2,3</sup> Cun Li,<sup>1</sup> Dong Wang,<sup>1</sup> Dong Yang,<sup>1</sup> Shufa Zheng,<sup>6</sup> Ke Hao,<sup>7</sup> Yohan Bossé,<sup>8</sup> Ma'en Obeidat,<sup>9</sup> Corry-Anke Brandsma,<sup>10</sup> You-Qiang Song,<sup>5</sup> Yu Chen,<sup>6</sup> Bo-Jian Zheng,<sup>1,2,3</sup> Lanjuan Li,<sup>6</sup> and Kwok-Yung Yuen<sup>1,2,3,4</sup>

<sup>1</sup>Department of Microbiology, <sup>2</sup>Research Centre of Infection and Immunology, <sup>3</sup>State Key Laboratory of Emerging Infectious Diseases, <sup>4</sup>Carol Yu Centre for Infection, and <sup>5</sup>Department of Biochemistry, The University of Hong Kong, Pok Fu Lam; <sup>6</sup>State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, First Affiliated Hospital, College of Medicine, Zhejiang University, China; <sup>7</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York; <sup>8</sup>Department of Molecular Medicine, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Laval University, and <sup>9</sup>University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada; and <sup>10</sup>Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, The Netherlands

**The genetic predisposition to severe A(H1N1)2009 (A[H1N1]pdm09) influenza was evaluated in 409 patients, including 162 cases with severe infection and 247 controls with mild infection. We prioritized candidate variants based on the result of a pilot genome-wide association study and a lung expression quantitative trait locus data set. The GG genotype of rs2070788, a higher-expression variant of *TMPRSS2*, was a risk variant (odds ratio, 2.11; 95% confidence interval, 1.18–3.77;  $P = .01$ ) to severe A(H1N1)pdm09 influenza. A potentially functional single-nucleotide polymorphism, rs383510, accommodated in a putative regulatory region was identified to tag rs2070788. Luciferase assay results showed the putative regulatory region was a functional element, in which rs383510 regulated *TMPRSS2* expression in a genotype-specific manner. Notably, rs2070788 and rs383510 were significantly associated with the susceptibility to A(H7N9) influenza in 102 patients with A(H7N9) influenza and 106 healthy controls. Therefore, we demonstrate that genetic variants with higher *TMPRSS2* expression confer higher risk to severe A(H1N1)pdm09 influenza. The same variants also increase susceptibility to human A(H7N9) influenza.**

**Keywords.** *TMPRSS2*; genetic variation; lung eQTL; A(H1N1)pdm09 influenza; A(H7N9) influenza.

The emergence of influenza virus A(H1N1)2009 (A[H1N1]pdm09) in early 2009 caused the first influenza pandemic of the 21st century. The global mortality of A(H1N1)pdm09 infection has been estimated to be about 0.2 million [1]. In Hong Kong, as of October 2010, a total of 299 severe cases were documented [2].

After the 2009 pandemic, A(H1N1)pdm09 virus continued to circulate and caused major epidemics in Northern America and Eurasia [3, 4]. About one-quarter to one-half of patients with severe A(H1N1)pdm09 influenza were previously healthy, without preexisting medical conditions that could enhance the risk, implicating the presence of host genetic predisposition for severe A(H1N1)pdm09 infection [5]. The World Health Organization, therefore, identified the study of host genetic factors on susceptibility and severity of influenza infection as a priority [6]. Genetic predisposition to severe A(H1N1)pdm09 influenza has been extensively studied in our laboratory and others [7–9]. We have demonstrated that genetic polymorphisms of surfactant protein B (*SFTPB*) and *CD55* contribute to the severe illness of A(H1N1)pdm09 infection [7, 8].

Received 3 February 2015; accepted 27 March 2015; electronically published 22 April 2015.

<sup>a</sup>Z. C. and J. Z. contributed equally to this work.

Correspondence: Kwok-Yung Yuen, Department of Microbiology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong Special Administrative Region, China (kyuen@hkucc.hku.hk).

**The Journal of Infectious Diseases**® 2015;212:1214–21

© The Author 2015. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.

DOI: 10.1093/infdis/jiv246

Four years after the 2009 pandemic A(H1N1) influenza, a novel reassortant avian influenza virus A(H7N9) was identified to cause human infection; which has affected 500 patients, with the case fatality rate of >30% as of 19 January 2015 [10]. Moreover, the virus continues to be detected in animals and the environment in China [11], leading to the second and third waves of human A(H7N9) influenza outbreak in the winters of 2013–2014 and 2014–2015, respectively. Most patients with A(H7N9) influenza had significant exposure to infected poultry or contaminated environment such as the live poultry market [12]. In China, a high proportion of the common population are assumed to be exposed to live poultry in both urban and rural settings. Therefore, we speculate that individuals with certain genetic predisposition are more likely to develop symptomatic A(H7N9) influenza.

We have conducted a small-scale genome-wide association study (GWAS), comparing the distribution of genetic variants in severe and mild cases of patients with A(H1N1)pdm09 influenza, aiming to identify the risk variants or susceptible genes predisposing to severe A(H1N1)pdm09 influenza. Instead of prioritizing the candidate variants with the association *P* value, we used the human lung expression quantitative trait locus (eQTL) data set based on a rational assumption that the genes differentially expressed in lung tissues would be more important in determining the severity of A(H1N1)pdm09 influenza in affected individuals.

We chose rs2070788, an intronic single nucleotide polymorphism (SNP) in transmembrane protease S2 (*TMPRSS2*) and a lung cis-eQTL for *TMPRSS2*, from the pilot GWAS. We verified

the genetic association of rs2070788 with severe infection in a cohort of 409 patients with A(H1N1)pdm09 influenza, including 162 cases with severe illness and 247 controls with mild illness. We also replicated the genetic association of rs2070788 with the susceptibility to A(H7N9) influenza in a cohort of 102 patients with A(H7N9) influenza and 108 healthy controls. In addition, we performed functional studies to validate the expression variation underlying the genetic association.

## MATERIALS AND METHODS

### Study Participants

The study participants in the A(H1N1)pdm09 cohort were described elsewhere [7, 8]. Their clinical characteristics are specified in Table 1. The patients and controls in A(H7N9) cohort have also been described previously [13]. This study was approved by the institutional review board of The University of Hong Kong/Hospital Authority of Hong Kong and the First Affiliated Hospital, College of Medicine, Zhejiang University.

### Genotyping Methods and Analysis

The genomic DNA samples from 42 patients with severe and 42 with mild A(H1N1)pdm09 influenza were genotyped using the Genome-Wide Human SNP Array 6.0 (Affymetrix), as described elsewhere [8]. A total of 561 844 SNPs were left after exclusion of SNPs with allele frequency <5% and genotyping missing rate >10%. Genomic DNA samples from additional 325 patients with A(H1N1)pdm09 influenza were genotyped using the SEQUENOM MassArray Genotyping platform (available at Center

**Table 1. Demographic and Clinical Profiles of Patients With Severe Cases and Controls With Mild A(H1N1)pdm09 Infection**

Patient Profile	Patients, No. (%) <sup>a</sup>		<i>P</i> Value <sup>c</sup>
	Severe Infection (n = 162) <sup>b</sup>	Mild Infection (n = 247)	
Demographic characteristic			
Age, median years (interquartile range)	56 (45–65)	40 (25–55)	<.001
Female	64 (39.5)	142 (57.5)	<.001
Risk conditions			
Age ≥65 y	43 (26.5)	35 (14.2)	.003
Pregnancy	1 (0.6)	8 (3.2)	.094
Chronic pulmonary disease	49 (30.2)	24 (9.7)	<.001
Chronic cardiac disease	36 (22.2)	26 (10.5)	.002
Metabolic disorders	47 (29.0)	35 (14.2)	<.001
Chronic renal disease	18 (11.1)	13 (5.3)	.035
Chronic hepatic disease	5 (3.1)	8 (3.2)	>.99
Neurological conditions	23 (14.2)	17 (6.9)	.017
Hemoglobinopathy	1 (0.6)	2 (0.8)	>.99
Immunosuppression	23 (14.2)	40 (16.2)	.675
Obesity	27 (16.7)	3 (1.2)	<.001

<sup>a</sup> Data represent No. (%) of patients unless otherwise stated.

<sup>b</sup> Among patients with severe illness, 26 died.

<sup>c</sup> The Fisher exact test and Mann–Whitney *U* test were used for categorical and continuous variables, respectively.

for Genomic Sciences, The University of Hong Kong). The association *P* values of SNPs were generated using PLINK v1.07 software [14]. The rs2070788 genotyping data of patients with A(H7N9) influenza and controls were obtained from the previous study [13]. The genotyping results of rs383510 were obtained by local imputation using MACH v1.0 software [15] with genotype data from 286 East Asian individuals (97 Chinese Han in Beijing, 100 Southern Han Chinese, and 89 Japanese in Tokyo) from the 1000 Genomes Project, released in June 2011 as a reference panel.

Haploview software was used to analyze and visualize the linkage disequilibrium (LD) pattern of the interested variants [16]. The in-house generated program indelLDplot (<https://sourceforge.net/projects/indelldplot/files>) was used to search for functional variants in high LD with the user-interested variant through mining the publically available 1000 Genomes and HapMap databases. IndelLDplot software can also annotate these high-LD variants with biological information, such as genomic regulatory features from the Ensembl and ENCODE database.

### Reporter Vector Construction, Transfection and Luciferase Assay

A putative regulatory element in *TMPRSS2* intronic region (chr21: 42 856 950–42 858 646; human genome assembly 19) accommodating rs383510 was amplified by polymerase chain reaction with primer pair GCCGAGGTACCCCAAGAAATGT and CGGTAAGGTACCCCTCCTGCC (KpnI digestion site underlined), cloned into pGL3-Basic vector (Promega) to generate 2 luciferase vectors with genotype TT or CC at locus rs383510. The correctness of the constructs was verified by direct sequencing. The 2 reporter vectors and unmodified pGL3-Basic vector were transfected into A549 cells, the cell line commonly used to model human respiratory epithelial cells, for luciferase assay, as described elsewhere [17].

### Statistical Analysis

The linear regression analysis incorporated in PLINK was used to examine the correlation of a specific variant to the quantitative expression in human lung tissues, using an additive model to estimate the effect of one copy increment of the variant. The luciferase assay results were analyzed using Student *t* test. Stepwise multivariate analysis and  $\chi^2$ , Fisher exact, and Mann-Whitney *U* tests were used for data analysis. Differences were considered statistically significant at  $P \leq .05$ .

## RESULTS

### Genetic Variation of *TMPRSS2* Associated With Disease Severity of A(H1N1)pdm09 Infection

We performed a small-scale GWAS in 42 patients with severe and 42 with mild A(H1N1)pdm09 influenza. A total of 28 566 SNPs showed differential allele distribution between the 2 groups ( $P < .05$ ). These SNPs were intersected with lung

**Table 2. Genetic Association of rs2070788 With Severe A(H1N1)pdm09 Infection**

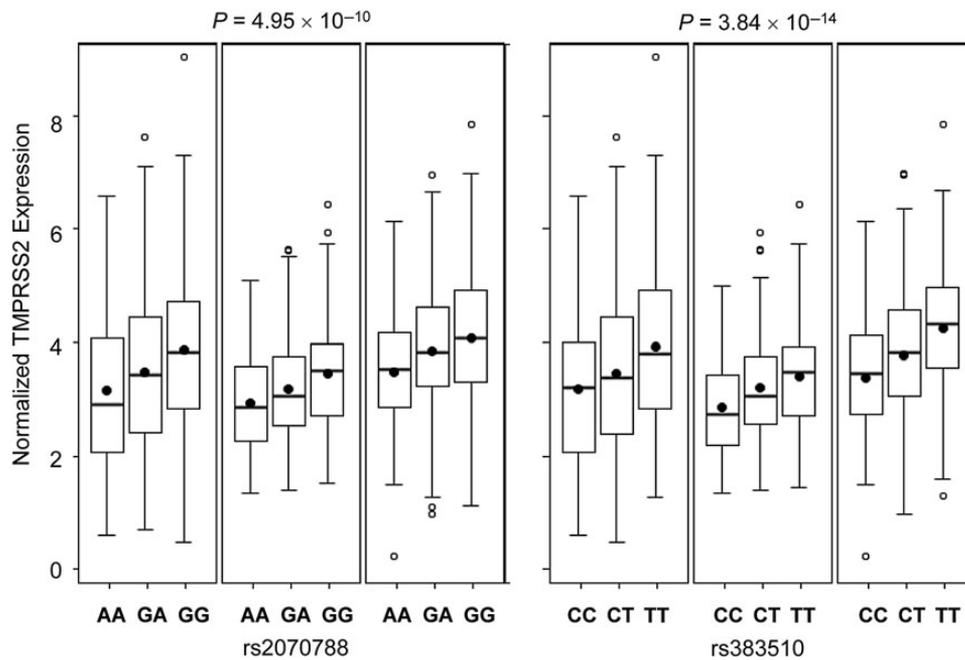
Variant Distribution and Analysis	Severe Patients (n = 162)	Mild Controls (n = 247)
Genotypic distribution, n (%)		
GG	30 (18.5)	24 (9.7)
AG	77 (47.5)	112 (45.3)
AA	55 (34.0)	111 (45.0)
Allelic analysis (G)		
OR (95% CI)	1.54 (1.14–2.06)	
<i>P</i> value	.004	
Genotype analysis (GG)		
OR (95% CI)	2.11 (1.18–3.77)	
<i>P</i> value	.01	
Multivariate analysis (GG)		
OR (95% CI)	2.15 (1.13–4.09)	
<i>P</i> value	.02	

Abbreviations: CI, confidence interval; OR, odds ratio.

cis-eQTLs at  $P < 10^{-8}$ , generating a total of 1114 SNPs linked to the expression of 445 unique genes. Based on the mounting recognition of *TMPRSS2* for influenza virus propagation, we chose an intronic SNP in *TMPRSS2*, rs2070788, which was mapped as a lung cis-eQTL ( $P = 4.95 \times 10^{-10}$ ) for *TMPRSS2*. We evaluated its genetic association with severe A(H1N1)pdm09 influenza in an extended cohort of 409 patients with A(H1N1)pdm09. The clinical characteristics of the patients are shown in Table 1.

It was demonstrated that the rs2070788 G allele was significantly overrepresented in severe compared with mild cases ( $P = .004$ ; Table 2). The GG genotype conferred >2-fold higher vulnerability to severe infection under the recessive model (odds ratio, 2.11;  $P = .01$ ). Because patient age, sex, and the presence of some predisposition conditions differed significantly between the severe and mild infection groups (Table 1), we performed multivariate regression analysis to exclude the effect of the confounding factors. We demonstrated that the rs2070788 GG genotype remained as an independent significant association signal for severe A(H1N1)pdm09 influenza (odds ratio, 2.15;  $P = .02$ ).

We profiled the *TMPRSS2* expression in correlation to rs2070788 genotypes from the lung eQTLs data set [18]. The profile showed that individuals with GG genotype have the highest expression, GA heterozygotes have intermediate expression and AA homozygotes have the lowest expression of *TMPRSS2* (Figure 1, left). Therefore, our results demonstrated that genetic variations of *TMPRSS2* were associated with the predisposition to severe A(H1N1)pdm09 infection. The higher *TMPRSS2* expression variant, rs2070788 GG genotype, was associated with higher susceptibility to severe illness in patients with A(H1N1)pdm09 influenza.



**Figure 1.** Genotype-specific *TMPRSS2* expression in human lung tissues. Boxplots show expression of *TMPRSS2* corresponding to rs2070788 (*left*) and rs383510 (*right*) genotypes. Boxes denote interquartile ranges; thick lines, medians; black dots, means; and open dots, possible outliers. Expression data from the 3 participating centers are shown.

### Identification of Putative Functional Variant

Based on lung eQTL analysis, rs2070788 genotypes were correlated with the differential expression level of *TMPRSS2*. However, there is currently little evidence implicating the functional basis of rs2070788 for *TMPRSS2* expressional regulation. As shown in Figure 2A, rs2070788 does not reside in any known regulatory region, which may affect the expression of *TMPRSS2*. Therefore, we proceeded to seek the functional variants in high LD with rs2070788 that caused the differential expression of this gene and generated the association signal of rs2070788 with severe A(H1N1)pdm09 infection.

We identified 9 potentially functional variants in a 20-kb range upstream and downstream of *TMPRSS2* using our in-house generated program IndelLDplot. We plotted the LD pattern of these potentially functional variants with rs2070788 using genotyping data of Chinese (Chinese Han in Beijing) and Asian populations from the 1000 Genomes Project. Four of 9 variants in the vicinity of *TMPRSS2* (rs461194, rs2298662, rs4816720, and rs383510) showed high LD with rs2070788 ( $r^2 > 0.80$ ) in the Chinese and Asian populations (Figure 2B).

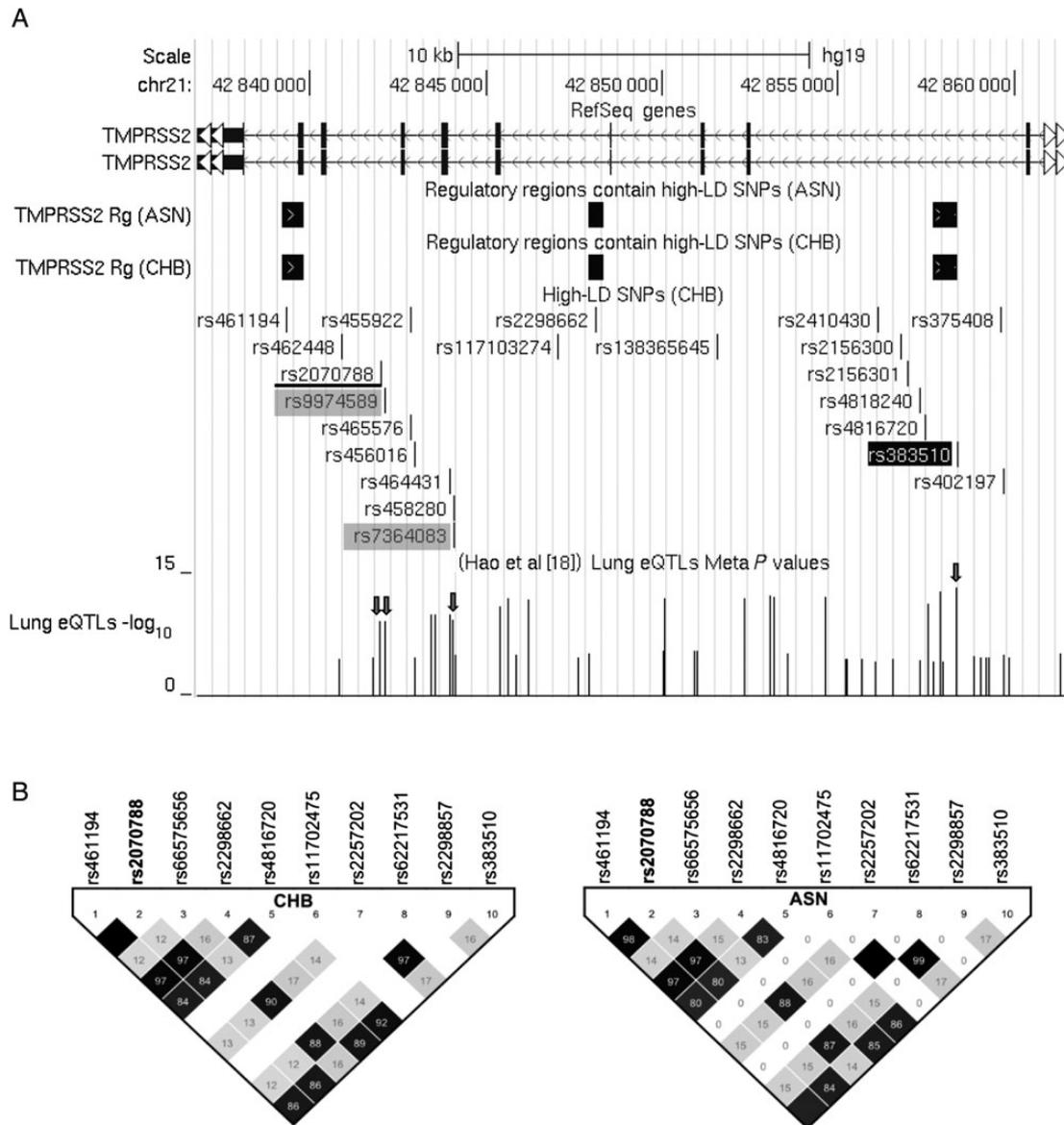
Based on the discovery of ENCODE and Roadmap experiments, 2 international collaborative projects to functionally annotate human genome, rs383510 was situated in a regulatory region predicted as a putative enhancer (Figure 2A). Moreover, among all the high-LD variants with rs2070788, rs383510 emerged as a lung cis-eQTL ( $P = 3.84 \times 10^{-14}$ ; Figure 1, *right*), the strongest among all the eQTLs for *TMPRSS2* gene in human

lung tissues. Therefore, we inferred that rs383510 might be the functional variant that dictated the differential *TMPRSS2* expression and underlay the genetic association of rs2070788 with severe A(H1N1)pdm09 infection.

### Rs383510 T Allele Increasing Enhancer Activity in Luciferase Assay

Despite its identity as a lung cis-eQTL, we cannot directly extrapolate rs383510 as a functional regulatory variant for *TMPRSS2* expression. According to the data of ENCODE and Roadmap, the DNA element harboring rs383510 is predicted to be a putative regulatory region with enhancer activity. To verify the prediction, we cloned the element into a luciferase reporter vector and transfected the reporter gene constructs into A549 cells for luciferase assay.

We demonstrated that the putative regulatory region was functional because the reporter vectors with the insertions displayed much higher luciferase activity than the unmodified vector ( $P = .006$ ; Figure 3). Furthermore, the luciferase activity of rs383510 TT plasmid was significantly greater than that of CC plasmid ( $P = .02$ ). Therefore, using a reporter gene assay, we identified a regulatory DNA element that can potentially govern *TMPRSS2* expression, in which the rs383510 T variant exhibited a significantly higher transcriptional level than the C variant. The luciferase assay results suggested that rs383510 was not only correlated with differential *TMPRSS2* expression but also functionally regulated its expression.

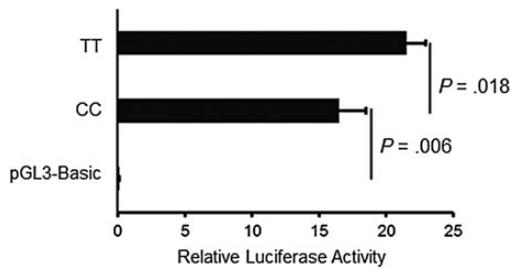


**Figure 2.** Prioritization of rs383510 as the functional variant for the genetic association. *A*, The genetic architecture of *TMPRSS2* gene. Upper panel shows the chromosomal region that accommodates *TMPRSS2* gene in human reference genome assembly 19; middle panel, the regulatory regions (*TMPRSS2* Rg) accommodating high-linkage disequilibrium (LD) single-nucleotide polymorphisms (SNPs) in Chinese (Chinese Han in Beijing [CHB]) and Asian (ASN) populations; lower panel, high-LD variants with rs2070788 (*underlined*). The variants as lung expression quantitative trait loci (eQTLs) are highlighted. Lung eQTL panel shows the locations of eQTLs and  $P$  values in  $-\log_{10}$  scale; horizontal bar represents the  $-\log_{10}$   $P$  value of 15. Among all the high-LD SNPs with rs2070788, rs383510 presents as both a potentially regulatory SNP and the strongest lung eQTL. *B*, A total of 9 putatively functional variants in vicinity of *TMPRSS2* gene were applied to LD analysis with association SNP rs2070788, based on genotype data for CHB and ASN populations from the 1000 Genomes Project.

### Association of rs2070788 and rs383510 With Susceptibility to A(H7N9) Infection

Influenza A(H7N9) virus, similar to the A(H1N1)pdm09 virus, has a monobasic cleavage site in the viral hemagglutinin (HA) protein, which is activated by *TMPRSS2* for viral replication and virus pathogenicity in vitro and in vivo [19, 20]. Therefore, we assessed the association of rs2070788 and rs383510 with

susceptibility to human A(H7N9) influenza in patients with A(H7N9) influenza and healthy controls. Remarkably, the association of rs2070788 was replicated in the A(H7N9) cohort (Tables 3 and 4). The carriers of rs2070788 genotype GG and GA were conferred an approximately 2-fold higher susceptibility to A(H7N9) infection ( $P = .02$ ). As expected, rs383510 generated a similar association signal. Collectively,



**Figure 3.** Luciferase activity of the putative enhancer. The putative enhancer region containing rs383510 was inserted in a luciferase reporter vector. Two constructs representing rs383510 genotype TT and genotype CC, as well as unmodified vector pGL3-Basic, were transfected into A549 cells. A *Renilla* luciferase vector pGL4.70[hRluc] was cotransfected as an internal control, and luciferase assay was performed 32 hours after transfection. The relative luciferase activity (ratio of firefly to *Renilla* luminescence) is presented as the mean and SD of triplicated transfection in a representative experiment.

we demonstrated the genetic association of rs2070788 and rs383510 with susceptibility to human A(H7N9) infection.

## DISCUSSION

Although GWASs have been widely conducted to understand the genetic basis underlying human diseases in an unbiased manner, the functional basis of the genetic variations leading to disease susceptibility cannot be directly inferred from these studies. On the other hand, genetic predisposition of most human diseases involves many genes, each of which may make a modest contribution. In human infectious diseases, such as avian influenza, the limited sample size of affected patients represents an additional hurdle for the discovery of

**Table 3. Genotypic Distributions of rs2070788 and rs383510 in Patients With Human A(H7N9) Influenza and Healthy Controls**

Variant	Distribution in Patients or Controls, No. (%)	
	Patients With H7N9 Influenza (n = 102)	Healthy Controls (n = 106)
<b>rs2070788</b>		
GG	16 (15.7)	9 (8.5)
GA	51 (50.0)	43 (40.6)
AA	35 (34.3)	54 (50.9)
Allele frequency (G), %	40.7	28.8
<b>rs383510</b>		
TT	14 (13.7)	7 (6.6)
TC	55 (53.9)	47 (44.3)
CC	33 (32.4)	52 (49.1)
Allele frequency (T), %	40.7	28.8

**Table 4. Genetic Associations of rs2070788 and rs383510 With Susceptibility to Human A(H7N9) Influenza**

Variant and Analysis	OR (95% CI)	P Value
<b>rs2070788</b>		
Allelic model (G)	1.70 (1.13–2.55)	.01
Dominant model	1.99 (1.14–3.48)	.02
<b>rs383510</b>		
Allelic model (T)	1.70 (1.13–2.55)	.01
Dominant model	2.01 (1.15–3.54)	.01

Abbreviations: CI, confidence interval; OR, odds ratio.

susceptible/resistance genes via GWASs. The eQTLs are currently the most abundant and systematically surveyed class of functional consequence of genetic variation [18, 21]. The local eQTLs (cis-eQTLs) tend to have larger effect on gene expression than distant eQTLs (trans-eQTLs) [22]. The availability of systematically generated eQTLs prioritized the discovery of GWAS and facilitated the identification of causal genes [23].

In the current study, we initiated a small-scale GWAS to investigate the role of genetic variation in the susceptibility to severe A(H1N1)pdm09 influenza. To avoid possible false-positive findings due to the small sample size, we prioritized the candidate variants by their status as a lung cis-eQTL instead of their association P values. We focused on rs2070788 based on its role as a lung cis-eQTL. The genetic association was evaluated in an extended cohort of 409 patients with A(H1N1)pdm09 influenza. The rs2070788 GG genotype, corresponding to higher *TMPRSS2* expression in human lung tissues, conferred >2-fold higher risk to severe A(H1N1)pdm09 influenza.

Based on its high LD with rs2070788, a SNP in a putative enhancer of *TMPRSS2* and an even stronger lung cis-eQTL, rs383510 was identified as the potentially functional variant for the association with severe A(H1N1)pdm09 influenza. Because of the shared usage of *TMPRSS2* for replication and pathogenicity of A(H1N1)pdm09 and A(H7N9) viruses, we assessed the genetic association of rs2070788 and rs383510 in another cohort of patients with A(H7N9) influenza and controls. Collectively, we demonstrated that genetic variants with higher *TMPRSS2* expression in the lung tissue conferred a higher risk to severe A(H1N1)pdm09 infection. Notably, the same variants also increased susceptibility to A(H7N9) infection.

Influenza virus infection is initiated with the attachment of virus HA to the cellular receptor and the subsequent fusion of viral envelope and cellular endosomal membrane. Posttranslational cleavage of HA by host protease is a prerequisite for the fusion and hence for virus infectivity, tissue tropism, and virus pathogenicity. Most human influenza A virus strains, including H1N1 and H3N2 have a single arginine at the HA cleavage site by trypsin-like proteases, which are mainly present in the respiratory or gastrointestinal tract [24]. *TMPRSS2* is the

major host protease that cleaves and activates the HA of influenza viruses in the human respiratory tract [25]. The highly pathogenic avian influenza viruses, such as subtypes H5 and H7, are cleaved at the multibasic motif by ubiquitous cellular proteases and cause lethal systemic infection [26]. Despite their avian virus origin, the novel A(H7N9) viruses possess a monobasic HA cleavage site [27], consistent with their low pathogenicity in poultry. *TMPRSS2*-deficient mice were highly tolerant to the lethal challenge of A(H1N1)pdm09 and A(H7N9) viruses, demonstrating the essential role of *TMPRSS2* in the pathogenesis of A(H1N1)pdm09 and A(H7N9) virus infections [20].

In this study, we found that the same variant of *TMPRSS2* increased the risk to human A(H7N9) influenza and severe A(H1N1)pdm09 influenza. Clinically, the patients with severe A(H1N1)pdm09 influenza and most patients with A(H7N9) influenza displayed similar symptoms. Human A(H1N1)pdm09 influenza primarily presents as an upper respiratory tract infection [28], whereas most patients with A(H7N9) influenza had rapidly progressive pneumonia, manifesting as a lower respiratory tract infection [29]. The receptor-binding specificity may contribute to the distinct clinical presentations of A(H1N1)pdm09 and A(H7N9) infections.

Despite the dual receptor-binding specificity, A(H7N9) HA retains a preference for  $\alpha$ 2,3-linked sialic acid ( $\alpha$ 2,3-SA) receptor over  $\alpha$ 2,6-SA receptor [30, 31]. In humans,  $\alpha$ 2,6-SA receptors are more abundantly expressed in the upper respiratory tract than  $\alpha$ 2,3-SA receptors, which are predominantly distributed in the lower respiratory tract [32]. Consistent with the receptor-binding pattern, A(H7N9) replication was more efficient in lung tissue than in the trachea [33]. Therefore, the preference of A(H7N9) virus binding to  $\alpha$ 2,3-SA receptor over the  $\alpha$ 2,6-SA counterpart and the robust viral replication in lung tissue may constitute the biological basis for the clinical manifestation of human A(H7N9) influenza as primarily a lower respiratory tract infection. Any risk factor, such as a high expression level of endogenous *TMPRSS2*, may robustly propagate the invading A(H7N9) virus and facilitate its spread to the lower respiratory tract, leading to symptomatic infection.

On the other hand, A(H1N1)pdm09 virus basically binds  $\alpha$ 2,6-SA receptor; and A(H1N1)pdm09 influenza is basically manifested as an upper respiratory tract infection in most patients. Given the existence of factors that facilitate viral growth and access to the lower respiratory tract, A(H1N1)pdm09 infection from the upper respiratory tract will spread downward to cause viral pneumonia and other complications, a severe outcome of A(H1N1)pdm09 influenza. A recent study clearly demonstrated that the volume of inoculum can affect disease severity of the A(H1N1)pdm09 infection in ferrets. The same dose of virus in a greater inoculation volume can optimally deliver the viruses to the lower respiratory tract and caused more severe diseases in ferrets [34]. Therefore, the viral access to the lower respiratory

tract and replication therein constitute the pathological basis of severe A(H1N1)pdm09 infection. Taken together, regardless of viral subtype, both severe A(H1N1)pdm09 influenza and A(H7N9) influenza are lower respiratory tract infections. In this study, we demonstrated that the higher *TMPRSS2* expression variant, the rs2070788 G allele, conferred a higher risk of severe A(H1N1)pdm09 infection as well as higher susceptibility to A(H7N9) influenza.

Interestingly, in A(H7N9) influenza, the proportion of male patients was more than double that of female patients [35]. It has been known that androgen can positively regulate *TMPRSS2* expression [36]. Therefore, further study might be conducted to compare *TMPRSS2* levels in females and males and determine whether androgen-enhanced *TMPRSS2* may increase the risk of A(H7N9) influenza in the latter. Apart from influenza viruses, *TMPRSS2* is an important host protease for the entry of several coronaviruses causing human infections, such as human coronavirus 229E [37], severe acute respiratory syndrome coronavirus [38], and Middle East respiratory syndrome coronavirus [39] with ongoing outbreaks in the Middle East. It would be interesting to learn whether the interindividual heterogeneity of *TMPRSS2* expression may also affect susceptibility to infections with these coronaviruses.

The role of *TMPRSS2* in the pathogenesis of A(H1N1)pdm09 and A(H7N9) has been clearly elucidated in mouse studies. In this study, we further demonstrated that the differential *TMPRSS2* expression derived from genetic variation may substantially affect disease outcome in affected patients on exposure of the 2 subtypes of virus. The discovery of *TMPRSS2* as a susceptibility marker for the severe A(H1N1)pdm09 influenza and human A(H7N9) influenza will help identify high-risk individuals for better prophylactic and therapeutic intervention.

## Notes

**Acknowledgments.** We thank our colleagues in the Centre for Genomic Sciences, The University of Hong Kong, for providing the excellent genotyping service for this study.

**Financial support.** This work is partially supported by the Providence Foundation Limited in memory of the late Lui Hac Minh and the Health and Medical Research Fund (grant nos. 12111412, RRG-05 and HKM-15-M03), Food and Health Bureau, Hong Kong Special Administrative Region.

**Potential conflicts of interest.** All authors: No potential conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

1. Widdowson MA, Iuliano AD, Dawood FS. Challenges to global pandemic mortality estimation. *Lancet Infect Dis* 2014; 14:670–2.
2. Swine and Seasonal Flu Monitor: Centre for Health Protection, Hong Kong, 2010. [http://www.chp.gov.hk/files/pdf/revised\\_SSFM\\_7\\_10\\_2010.pdf](http://www.chp.gov.hk/files/pdf/revised_SSFM_7_10_2010.pdf). Accessed 12 May 2015.
3. Napolitano LM, Angus DC, Uyeki TM. Critically ill patients with influenza A(H1N1)pdm09 virus infection in 2014. *JAMA* 2014; 311:1289–90.

4. Sherbany H, McCauley J, Meninger T, et al. Return of pandemic H1N1 influenza virus. *BMC Infect Dis* **2014**; 14:3848.
5. 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**; 16:9–16.
6. World Health Organization. WHO public health research agenda for influenza. version 1, **2009**. [http://www.who.int/influenza/resources/research/2010\\_04\\_29\\_global\\_influenza\\_research\\_agenda\\_version\\_01\\_en.pdf](http://www.who.int/influenza/resources/research/2010_04_29_global_influenza_research_agenda_version_01_en.pdf). Accessed 12 May 2015.
7. 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**; 206:495–503.
8. To KK, Zhou J, Song YQ, et al. Surfactant protein B gene polymorphism is associated with severe influenza. *Chest* **2014**; 145:13.
9. Everitt AR, Clare S, Pertel T, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* **2012**; 484:519–23.
10. Avian influenza report. Volume 11, No. 03. [http://www.chp.gov.hk/files/pdf/2015\\_avian\\_influenza\\_report\\_vol11\\_wk03.pdf](http://www.chp.gov.hk/files/pdf/2015_avian_influenza_report_vol11_wk03.pdf). Accessed 12 May 2015.
11. Wang C, Wang J, Su W, et al. Relationship between domestic and wild birds in live poultry market and a novel human H7N9 virus in China. *J Infect Dis* **2014**; 209:34–7.
12. Cowling BJ, Jin L, Lau EH, et al. Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: a population-based study of laboratory-confirmed cases. *Lancet* **2013**; 382:129–37.
13. Chen Y, Zhou J, Cheng Z, et al. Functional variants regulating LGALS1 (Galectin 1) expression affect human susceptibility to influenza A (H7N9). *Sci Rep* **2015**; 5:8517.
14. 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**; 81:559–75.
15. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **2010**; 34:816–34.
16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**; 21:263–5.
17. Zhou J, Tan T, Tian Y, et al. Kruppel-like factor 15 activates hepatitis B virus gene expression and replication. *Hepatology* **2011**; 54:109–21.
18. Hao K, Bosse Y, Nickle DC, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **2012**; 8:e1003029.
19. Tarnow C, Engels G, Arendt A, et al. TMPRSS2 is a host factor that is essential for pneumotropism and pathogenicity of H7N9 influenza A virus in mice. *J Virol* **2014**; 88:4744–51.
20. Sakai K, Ami Y, Tahara M, et al. The host protease TMPRSS2 plays a major role in in vivo replication of emerging H7N9 and seasonal influenza viruses. *J Virol* **2014**; 88:5608–16.
21. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. *Nat Genet* **2007**; 39:1202–7.
22. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. *Nature* **2008**; 452:423–8.
23. Battle A, Montgomery SB. Determining causality and consequence of expression quantitative trait loci. *Hum Genet* **2014**; 133:727–35.
24. Kido H, Okumura Y, Takahashi E, et al. Role of host cellular proteases in the pathogenesis of influenza and influenza-induced multiple organ failure. *Biochim Biophys Acta* **2012**; 1824:186–94.
25. Bertram S, Heurich A, Lavender H, et al. Influenza and SARS-coronavirus activating proteases TMPRSS2 and HAT are expressed at multiple sites in human respiratory and gastrointestinal tracts. *PLoS One* **2012**; 7:e35876.
26. Bottcher E, Matrosovich T, Beyerle M, Klenk HD, Garten W, Matrosovich M. Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium. *J Virol* **2006**; 80:9896–8.
27. Gao R, Cao B, Hu Y, et al. Human infection with a novel avian-origin influenza A (H7N9) virus. *N Engl J Med* **2013**; 368:1888–97.
28. Riley S, Kwok KO, Wu KM, et al. Epidemiological characteristics of 2009 (H1N1) pandemic influenza based on paired sera from a longitudinal community cohort study. *PLoS Med* **2011**; 8:e1000442.
29. To KK, Chan JF, Chen H, Li L, Yuen KY. The emergence of influenza A H7N9 in human beings 16 years after influenza A H5N1: a tale of two cities. *Lancet Infect Dis* **2013**; 13:809–21.
30. Xu R, de Vries RP, Zhu X, et al. Preferential recognition of avian-like receptors in human influenza A H7N9 viruses. *Science* **2013**; 342:1230–5.
31. Xiong X, Martin SR, Haire LF, et al. Receptor binding by an H7N9 influenza virus from humans. *Nature* **2013**; 499:496–9.
32. de Graaf M, Fouchier RA. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO J* **2014**; 33:823–41.
33. Zhou J, Wang D, Gao R, et al. Biological features of novel avian influenza A (H7N9) virus. *Nature* **2013**; 499:500–3.
34. Moore IN, Lamirande EW, Paskel M, Donahue D, Qin J, Subbarao K. Severity of clinical disease and pathology in ferrets experimentally infected with influenza viruses is influenced by inoculum volume. *J Virol* **2014**; 88:13879–91.
35. Jernigan DB, Cox NJ. H7N9: preparing for the unexpected in influenza. *Ann Rev Med* **2015**; 66:361–71.
36. Lin B, Ferguson C, White JT, et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. *Cancer Res* **1999**; 59:4180–4.
37. Bertram S, Dijkman R, Habjan M, et al. TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *J Virol* **2013**; 87:6150–60.
38. Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol* **2011**; 85:873–82.
39. Shirato K, Kawase M, Matsuyama S. Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *J Virol* **2013**; 87:12552–61.