Polymorphisms in the C-type lectin genes cluster in chromosome 19 and predisposition to severe acute respiratory syndrome coronavirus (SARS-CoV) infection

H Li,[1,2,3,4] N L-S Tang,[4,5,6] P K-S Chan,[7,8] C-Y Wang,[1,2,3,4] D S-C Hui,[8,9] C Luk,[6] R Kwok,[6] W Huang,[4,5,6] J J-Y Sung,[7,8] Q-P Kong,[1,2,4] Y-P Zhang[1,2,4]

► Additional tables and figure are published online only at http://jmg.bmj.com/content/vol45/issue11

[1] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China; [2] Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China; [3] Graduate University of the Chinese Academy of Sciences, Beijing, China; [4] KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, China; [5] Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China; [6] Departments of Chemical Pathology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China; [7] Departments of Microbiology, The Chinese University of Hong Kong, Hong Kong, China; [8] Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Hong Kong, China; [9] Departments of Medicine & Therapeutics, The Chinese University of Hong Kong, Hong Kong, China

Correspondence to: Professor Y-P Zhang, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; zhangyp1@263.net.cn

HL and NLS-T contributed equally to this work

Received 12 March 2008
Revised 8 May 2008
Accepted 16 June 2008
Published Online First 12 August 2008

ABSTRACT

Background: Polymorphisms of *CLEC4M* have been associated with predisposition for infection by the severe acute respiratory syndrome coronavirus (SARS-CoV). DC-SIGNR, a C-type lectin encoded by *CLEC4M*, is a receptor for the virus. A variable number tandem repeat (VNTR) polymorphism in its neck region was recently associated with susceptibility to SARS infection. However, this association was controversial and was not supported by subsequent studies. Two explanations may account for this discrepancy: (1) there may be an unknown predisposition polymorphism located in the proximity which is linked to the VNTR; or (2) it was a spurious association due to unrecognised population structure in the VNTR.

Methods: We performed a comprehensively genetic association study on this C-type lectin gene cluster (*FCER2, CLEC4G, CD209, and CLEC4M*) at 19p13.3 by a tagging single nucleotide polymorphisms (SNPs) approach.

Results: 23 tagSNPs were genotyped in 181 SARS patients and 172 population controls. No significant association with disease predisposition was detected. Genetic variations in this cluster also did not predict disease prognosis. However, we detected a population stratification of the VNTR alleles in a sample of 1145 Han Chinese collected from different parts of China.

Conclusion: The results indicated that the genetic predisposition allele was not found in this lectin gene cluster and population stratification might cause the previous positive association.

Severe acute respiratory syndrome (SARS) is a human infectious disease caused by a new coronavirus, SARS-CoV. A major outbreak in China and Asian countries occurred in 2003 and infected more than 8000 people worldwide.[1 2] A spectrum of the disease severity among infected patients was found ranging from a mild febrile illness to severe respiratory distress requiring assisted ventilation.[3 4] On average, 20–30% of SARS patients had a severe disease who required admission to intensive care and/or died of respiratory failure or other complications.[4]

The pronounced heterogeneity in disease outcome suggested an underlying genetic predisposition factor that might determine susceptibility to infection or disease progression/prognosis. Several studies reported an association between host genetic factors with the susceptibility to SARS. For example, *HLA-B**4601 and *HLA-B**0703 alleles have been associated with the susceptibility to SARS[5 6] but these alleles are rare and could not account for predisposition in the majority of patients. On the other hand, we and others studied the polymorphisms in the angiotensin converting enzyme-2 (*ACE2*), the major functional receptor for SARS-CoV infection, but found no association with the susceptibility and outcome of SARS-CoV infection.[7–9] Our previous studies also observed a difference in the extent of chemokine response among SARS patients, and patients with intense IP-10 expression after infection were more likely to suffer from adverse outcomes.[10 11]

Recently, Chan *et al* studied the VNTR polymorphism of the neck region of *CLEC4M* (also known as L-SIGN/CD-209L).[12] Heterozygotes of the VNTR polymorphism were more susceptible to SARS-CoV infection. *CLEC4M* is a highly plausible susceptibility gene as it is a co-receptor for the virus. In vitro functional study also showed a lower binding affinity towards SAR-CoV in a cell line expressing the neck region in a heterozygote manner. We and another research group in China cannot replicate this association and the trend of association was reversed, though not significantly, in a Beijing sample (that is, homozygotes were more frequent in the infection group).[13 14] There are two potential explanations for this controversy. First, the VNTR is not the genuine functional polymorphism determining susceptibility to SARS infection but is merely a marker indirectly linked with another functional variant in a nearby locus. Second, the association is a spurious one due to unrecognised population stratification, and we presented some data supporting a possible population structure in this VNTR among Chinese in Beijing and Hong Kong.[13] However, it is uncertain if other polymorphisms in this cluster of four C-type lectin genes are associated with predisposition for SARS infection. Therefore, we performed this association study using tagging single nucleotide polymorphisms (SNPs) to give a more comprehensive analysis of genetic variations in this chromosome 19 region.

Dendritic cell specific intracellular adhesion molecular-3 grabbing non-integrin (*DC-SIGN*, encoded by *CD209)* is a prototype of C-type lectin and express primarily on phagocytic cells, such as dendritic cells and macrophages. *DC-SIGNR* (dendritic cell specific ICAM grabbing non-integrin

related, encoded by *CLEC4M*) is a homologue of *DC-SIGN*. It is also known as *L-SIGN* or *CD-209L* as it is preferentially expressed by endothelial cells in liver and lymph nodes. They share 77% amino acid and 73% nucleotide identity with identical exon–intron organisation. Both of them can bind pseudovirus expressing the SARS-CoV spike gene[15 16] and *DC-SIGNR* also facilitates viral entry. *CLEC4G* (liver and lymph node sinusoidal endothelial cell C-type lectin) is co-expressed with *DC-SIGNR* on sinusoidal endothelial cells in liver and lymph nodes and can interact with surface protein of a number of viruses including the spike protein of SARS-CoV.[17] *FCER2* (also known as *CLEC4J*) is another lectin gene with a high level of homology to *CLEC4G*, which plays a key role in B cell activation. All these four genes are mapped within a region of 81 kb on 19p13.3 and form a C-type lectin gene cluster. The C-type lectins share a common structure of an attachment factor which is formed by a C-terminal carbohydrate recognition domain (CRD). It binds high mannose oligosaccharides or N-linked glycosylated protein.[18] The CRD is connected to a neck region composed of repeat units and the number of repeats may be variable in the population, which is represented by the variable number tandem repeat (VNTR). The VNTR in the neck region of *CLEC4M* is the only one showing a high degree of polymorphism in the population among the four lectin genes.

In this study, we performed a comprehensive study of the lectin cluster by using 23 tagging SNPs to investigate the association between genetic polymorphisms in the C-type lectin gene cluster and susceptibility to SARS-CoV infection and clinical outcomes. In order to characterise further the population structure of genetic polymorphism in this region, we examined the population stratification of the VNTR polymorphism among 1145 healthy Han Chinese and 742 Chinese ethnic minority samples collected from different parts of China.

PATIENTS AND METHODS

Patients

All patients (n = 181) had a confirmed diagnosis of SARS by at least one of the following laboratory procedures—serological conversion to SARS-CoV or a positive viral culture or a positive SARS-CoV detection by reverse transcriptase-polymerase chain reaction (RT-PCR)—and were recruited in 2003. Details of disease course, results of biochemical and haematological investigations, and co-morbidity (including history of diabetes, chronic lung diseases, hypertension, cerebrovascular accident, cancer, ischaemic heart disease, chronic renal failure and chronic liver disease) were examined for association with clinical outcome. All patients had been treated according to a standard protocol as previously detailed elsewhere.[19 20] Adverse disease outcome was defined as either admission to the intensive care unit (ICU) or death due to SARS-CoV infection. This study has been approved by our institutional and hospital research ethics committees.

Control groups

There are two control groups. The first one was an ethnically matched healthy population control (n = 172) which was recruited from local university students in Hong Kong to represent the genetic variation among Southern Chinese in Hong Kong. The second one was an expanded sample from our previous effort to genotype the exon 4 VNTR of *CLEC4M*. Altogether, there were 1145 Han Chinese samples collected, including five new collections of urban community samples of Han population from four provinces of China, including Zhanjiang (n = 194) and Meizhou (n = 156) of Guangdong province, Shandong (n = 268), Liaoning (n = 262) and Sichuan (n = 265). Meanwhile, seven ethnic minority populations from different regions were also analysed to examine the distribution of this VNTR (total n = 742), including Miao (n = 74), Yao (n = 128), Zhuang (n = 170), Dai (n = 117), Dongxiang (n = 48), Uzbek (n = 57), and Uygur (n = 148). Informed consent was obtained from all subjects.

Selection of tagSNPs and genotyping methods

Based on HapMap phase I and dbSNP/Perlegene data of Han Chinese, we identified 23 tagging SNPs by haplotype based algorithm to represent genetic variation in the four genes *FCER2, CLEC4G, CD209* and *CLEC4M* genes (primers shown in supplemental table 1). Haploblock Finder program (http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi) was used to determine the structure of haploblock and then tag SNPs were defined in each of the haploblocks (supplemental fig 1).[21] Firstly, the *Haploblock Finder* program defines the location of haploblocks in the target region. Then it selects sufficient SNPs within each of the haploblocks to define all or most of the haplotypes found in the population. This method is different from the pairwise LD based tagSNP selection using the pairwise r^2 in that it considers tagSNP along a consecutive series of SNPs which essentially also considered the order and spatial relationship between SNPs. All SNPs with minor allele frequencies of at least 5% in Asians were used by the algorithm and we showed that the haploblock based approach was appropriate for tagSNP identification in expanded genetic loci containing multiple genes.[22]

Genomic DNA was extracted from peripheral blood using DNA extraction kit according to the manufacturer's instruction (Roche, USA). Genotyping was performed by PCR-restriction fragment length polymorphism (PCR-RFLP) or mismatched PCR-RFLP. PCR was performed in 25 μl reactions comprising 0.25 μM of each primer pair, 2 mM $MgCl_2$, 0.6 U of Ampli Taq Gold Polymerase (Applied Biosystems, Foster City, California, USA) and PCR buffer (10 mM Tris–HCl, pH 8.3; 50 mM KCl). Reaction cycle was started at 96°C for 15 min to activate the polymerase and amplification was achieved by 35 cycles of 96°C for 30 s, annealing temperatures for 45 s and 72°C for 45 s. The final elongation step was 72°C for 7 min. For restriction enzyme digestion, 7 ul of the PCR product was digested by 5 U of the required enzyme overnight. The genotype call was made by separating the DNA in a 4% agarose gel and stained with ethidium bromide. To validate the genotyping results, 10% of the samples were re-genotyped by either duplicated genotyping experiments or direct DNA sequencing. Genotyping the VNTR in exon 4 of the *CLEC4M* gene used the same protocol as described previously.[12]

Statistical analysis

Statistical analysis of genotype distribution and allele frequencies was performed by a χ^2 or Fisher exact test (SPSS for windows 11.5). Hardy–Weinberg equilibrium (HWE) test for VNTR in the patients and control populations was performed by GENEPOP software (http://wbiomed.curtin.edu.au/genepop/index.html). For SNPs, HWE was also determined by either χ^2 or Fisher exact tests.

Univariate association between risk factors (categorical variables) and adverse outcome in the patient groups was performed by χ^2 or Fisher's exact test. Conditional logistic regression was used to identify independent predictors for

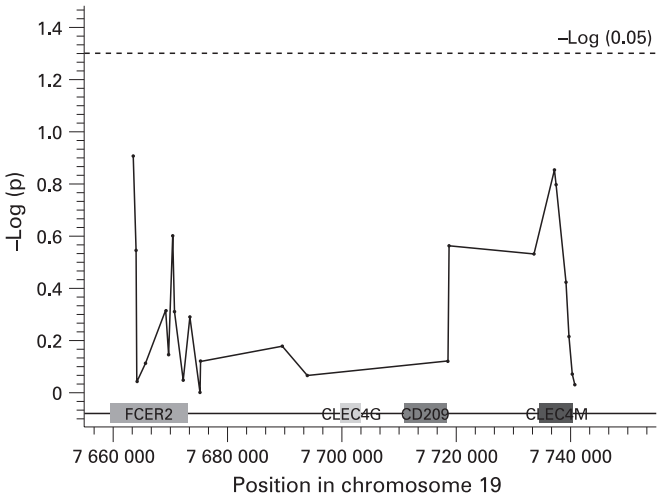

Figure 1 Association between 23 tagging single nucleotide polymorphisms (SNPs) and susceptibility to severe acute respiratory syndrome coronavirus (SARS-CoV). Plot of –log p values of SNP association by χ^2 statistics along the C-type lectin family. A value above 1.30 represents a potential significant association at $p<0.05$. None of the SNPs showed significant association.

adverse disease outcome and determine the adjusted odds ratio and 95% confident intervals (CI) of these risk factors.

RESULTS

Demographic data

The demographic parameters of the SARS patients and control group are listed in supplemental table 2. Among the 181 patients studied, 131 had an uneventful recovery while another 50 patients were classified as having an adverse outcome. Patients with adverse outcomes included 27 patients who recovered after admission to the ICU for mechanical ventilation, and 23 who died as a result of SARS.

Genotypic and allelic frequencies of tagging SNPs

The genotype and allele frequencies of the 23 tagSNPs in *FCER2*, *CLEC4G, CD209* and *CLEC4M* genes are shown in supplemental tables 3 and 4. The genotype frequencies of SARS patients and control population did not deviate from the HWE except for two SNPs (by both χ^2 and a Markov chain method in GENEPOP) ($p>0.05$). Our results indicated that there was no significant difference in the genotype distribution and allele frequencies of the two groups for all 23 polymorphism sites (supplemental tables 3 and 4, fig 1).

Among the SARS patients, we further compared the genotype and allele frequencies of 23 tagSNPs between patients with different outcomes by univariate analyses. There was also no significant difference between the two groups (supplemental tables 5 and 6). The other clinical risk factors associated with adverse outcome were similar to those reported in previous studies (supplemental table 2). In brief, older age, male sex, high plasma lactate dehydrogenase (LDH) values and high white cell counts were significant risk factors for adverse outcome. Logistic regression was performed to define the independent risk factors for adverse outcome and found that only age, sex and LDH values were the independent risk factors. Furthermore, it confirmed that none of the 23 tagSNPs were associated with disease outcome after controlling for other risk factors.

Population structure of VNTR in CLEC4M among Chinese from different geographic regions

Genotype frequencies and homozygote proportions of VNTR in exon 4 of the *CLEC4M* gene of these patient and control groups have been reported before.[13] As described previously, VNTR was not associated with predisposition for SARS infection but there was a difference in allele frequencies between Northern Chinese (Beijing) and Southern Chinese (Hong Kong).

The population structure of this VNTR in Han Chinese was investigated in five new collections of urban Han community samples (total $n=1145$) from four additional provinces of China. The homozygote proportion varied from the lowest 46.0% to the highest 54.5% (table 1, fig 2). Interestingly, the two extreme homozygote frequencies were found in two samples from the same Guangdong province (46.0% in Hong Kong in the south of Guangdong province and 54.5% in Meizhou of North Guangdong). This difference of homozygote proportion fell just short of statistical significance ($p=0.06$). On the other hand, the difference in both genotype and allele distributions are significant ($p<0.05$ by GENEPOP) between pooled Guangdong (Zhanjiang and Hong Kong) Chinese and Sichuan Chinese. Together with a generally higher homozygote proportion due to a higher frequency of 7-repeat allele in Northern and Southwest China, this suggested the presence of population structure for this VNTR in *CLEC4M* gene in Han Chinese. Among all our Han Chinese samples, the genotype distributions followed HWE equilibrium except Shandong ($p=0.042$, calculated with POPGENE), which may be due to a small sample size and a possible population admixture in this important commercial province.

Population structure of VNTR in CLEC4M among minority populations

In order to understand better the genetic background among the early settlements and the extent of cultural isolation on allele frequencies, seven ethnic minority populations from different regions were also analysed for the allelic distribution of this VNTR (total $n=742$). At first glance, the homozygote proportion varied from the lowest 41.2% in Zhuang to the highest 64.8% in Miao, though both minorities reside in southwest China (fig 3). As expected, the homozygote proportion highly correlated with the frequencies of 7-repeats ($r^2=0.87$), which was the predominant allele across all samples. Among the four minorities (Miao, Yao, Dai and Zhuang) clustered within the three southwest provinces of China, Miao and Yao had comparable and high levels of homozygosity at 64.8% and 59.4%, respectively. These are the highest levels of homozygosity found in this study. Actually, Miao and Yao share a common language lineage and previous phylogeographic studies indicated that they were the descendents of northern inhabitants. On the contrary, Dai and Zhuang, who were the local southern residents and share a common southern language lineage, had the lowest homozygosity at 42.0% and 41.2%. The profiles of allelic distribution of Dai and Zhuang were similar to other southeast Asian and Han in Zhanjiang and Hong Kong, which suggested a low homozygosity in the ancient southern settlements. Three northwest minorities (Uygur, Ozbek and Dongxiang) were all located in the ''Silk road'' region where there is a significant admixture with the European gene pool.[23 24] Interestingly, they all had a similarly low level of homozygosity (Uygur 44.6%, Ozbek 42.1% and Dongxiang 43.8%) (fig 3), similar to what had been reported in Europeans.[25]

Table 1 Genotype distribution of encoding VNTR in exon 4 of *CLEC4M* gene in Chinese population samples collected from different parts of China

	Southern China						Northern China				SW China	
	Zhanjiang, Guangdong (n = 194)		Meizhou, Guangdong (n = 156)		Hong Kong* (n = 463)		Shandong† (n = 268)		Liaoning (n = 262)		Sichuan (n = 265)	
Genotypes§												
4/7	0	0.0%	0	0.0%	1	0.2%	0	0.0%	0	0.0%	0	0.0%
5/5	4	2.1%	9	5.8%	17	3.7%	4	1.5%	10	3.8%	4	1.5%
5/6	3	1.5%	0	0.0%	8	1.7%	0	0.0%	2	0.8%	1	0.4%
5/7	44	22.7%	25	16.0%	94	20.3%	60	22.4%	54	20.6%	47	17.7%
5/9	8	4.1%	8	5.1%	20	4.3%	14	5.2%	10	3.8%	13	4.9%
6/6	1	0.5%	0	0.0%	0	0%	1	0.4%	0	0.0%	2	0.8%
6/7	15	7.8%	12	7.7%	46	10.0%	14	5.2%	10	3.8%	13	4.9%
6/9	2	1.0%	1	0.7%	5	1.1%	1	0.4%	6	2.3%	2	0.8%
7/7	78	40.2%	74	47.4%	189	40.8%	123	45.9%	111	42.4%	125	47.2%
7/8	1	0.5%	0	0.0%	0	0%	1	0.4%	0	0.0%	0	0.0%
7/9	31	16.0%	25	16.0%	75	16.2%	40	14.9%	49	18.7%	54	20.3%
8/9	0	0.0%	0	0.0%	1	0.2%	0	0.0%	0	0.0%	0	0.0%
9/9	7	3.6%	2	1.3%	7	1.5%	10	3.7%	10	3.8%	4	1.5%
Homozygotes‡	90	46.4%	85	54.5%	213	46.0%	138	51.5%	131	50.0%	135	50.9%
Heterozygotes	104	53.6%	71	45.5%	250	54.0%	130	48.5%	131	50.0%	130	49.1%
Alleles												
5	63	16.2%	51	16.4%	156	16.9%	82	15.3%	86	16.4%	69	13.0%
6	22	5.7%	13	4.2%	59	6.4%	17	3.2%	18	3.5%	20	3.8%
7	247	63.7%	210	67.3%	594	64.2%	361	67.3%	335	63.9%	364	68.7%
8	1	0.2%	0	0.0%	1	0.1%	1	0.2%	0	0.0%	0	0.0%
9	55	14.2%	38	12.1%	115	12.4%	75	14.0%	85	16.2%	77	14.5%

*Data from Hong Kong population control is cited from Tang *et al*.[13]
†Genotypes deviated from Hardy–Weinberg equilibrium (p = 0.042).
‡The difference of the homozygote proportion was marginally significant between Hong Kong and Meizhou (p = 0.06)
§There were significant differences in genotype and allele distributions between pooled Guangdong Zhanjiang/Hong Kong sample (n = 657) and Sichuan sample (n = 265).

DISCUSSION

Our results showed that there was no association between tagSNPs in the four genes of the C-type lectin gene cluster in chromosome 19 (*FCER2, CLEC4G, CD209, CLEC4M)* and susceptibility to SARS infection or its clinical outcome. We used a tagSNP approach to provide a comprehensive coverage of the genetic variations in this lectin gene cluster. In addition to those SNPs we have genotyped, we should be able to detect any association signal due to other functional SNPs that were not directly genotyped in this project, through linkage disequilibrium with one or more genotyped tagSNPs. Therefore, our results indicated that there was no significant association between genetic variations in this locus and SARS infection.

In a previous study, Chan *et al* reported homozygotes of *CLEC4M* exon 4 VNTR were protected from SARS-CoV infection, as the homozygote proportion was 46.0% in SARS patients and 56.0–58.7% in various control groups. But this association was not replicated in two subsequent association studies, one carried out in Hong Kong[13] and another in Northern China.[14] The subsequent two replication reports found that the homozygote proportions were similar in both patients and control group.[13 14] In fact, the presumably protective homozygotes were more frequent in the patient groups collected in Northern China,[14] which was in contrast with the hypothesis of Chan *et al*. Here, our results from tagging SNPs again confirmed that there was no association between *CLEC4M* and its neighbouring C type-lectin genes with susceptibility to SARS infection.

A potential role of host genetic factors in the predisposition to SARS-CoV infection has been suggested by various groups.[5 6 10–12 26 27] More recently, Chan *et al* re-examined additional non-synonymous coding SNPs of *FCER2* and one SNP in *ICAM3*.[28] Their data were consistent with our findings in that there was no association between the SNPs in *FCER2* and SARS. On the other hand, they reported an association between an SNP in *ICAM3* gene which was located more than 2.5 million basepairs away from the C-type lectin gene cluster. Interestingly, three out of five SNPs in *FCER2* showed only a borderline HWE in the control subjects (p values between 0.055 and 0.062), which suggested the possibility of an unrecognised population structure in the control group studied by Chan *et al*.[12]

When the data of VNTR of the *CLEC4M* gene were examined in greater details, it was interesting to note that the difference of results across the three Chinese studies was largely attributed to the different homozygote proportions of VNTR among control groups (46.0% in Tang *et al* and 51.5% in Zhi *et al vs* 56.0–58.7% in Chan *et al*), rather than difference among patient groups. Our data strongly suggested that a previous unrecognised population stratification, represented by a geographic difference in genotype proportions and allele frequencies, existed among Chinese from different geographic areas. Our previous paper[13] provided preliminary evidence of this genetic structure of the *CLEC4M* gene by revealing differences in genotype proportions between two control samples collected from Hong Kong (Southern Chinese) and Beijing (Northern Chinese). In this study, we expanded the samples of control to over 1000 subjects and covered a wider geographic area of China. Meanwhile, it supported our previous observation of genetic structure and the homozygote proportions in controls varied from 44.0–55.0% in the Chinese Han population. This level of difference in the healthy population would lead to a significant association if the sample sizes were more than 180 in each group.

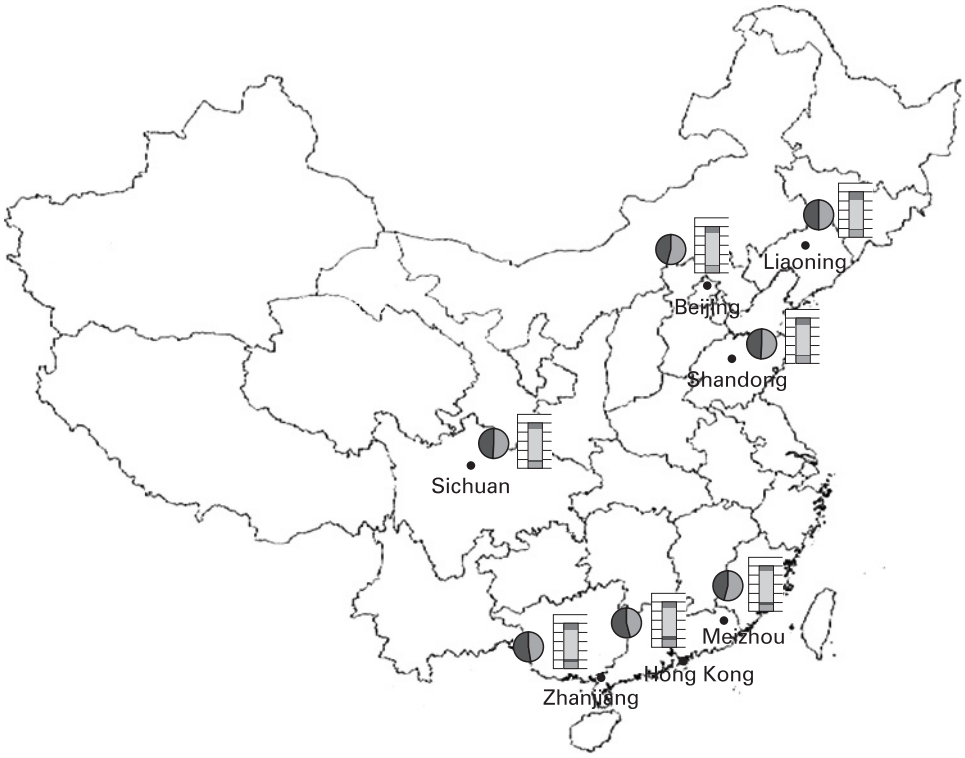

Figure 2 VNTR allele frequency in Han Chinese. Homozygosity and allelic frequencies of samples of Han Chinese collected from different region of China. The pie chart is the homozygote proportion (light grey) *vs* heterozygote (dark grey). The stacked bar shows the allele distribution, with 5-repeat allele at the bottom and 9-repeat allele at the top.

To gain more insight into the origin of this population structure, we genotyped 2290 alleles of this VNTR among Han Chinese sampled across different regions of China. Beside a general decreasing trend of the homozygote frequency from north to south (fig 2), a few localised areas in Southern China also showed a higher homozygote frequencies than the rest of Southern China such as in Guangdong (Meizhou) and Sichuan. Guangdong province in Southern China received several waves of northern immigrants in the past. Meizhou of Guangdong is a city populated by northern immigrants (Hakka people) who moved to Guangdong over the past 1000 years. We found that the Meizhou samples had a high homozygosity (54.5%), which was similar to other northern populations. The current Meizhou population with a majority of Hakka people is the

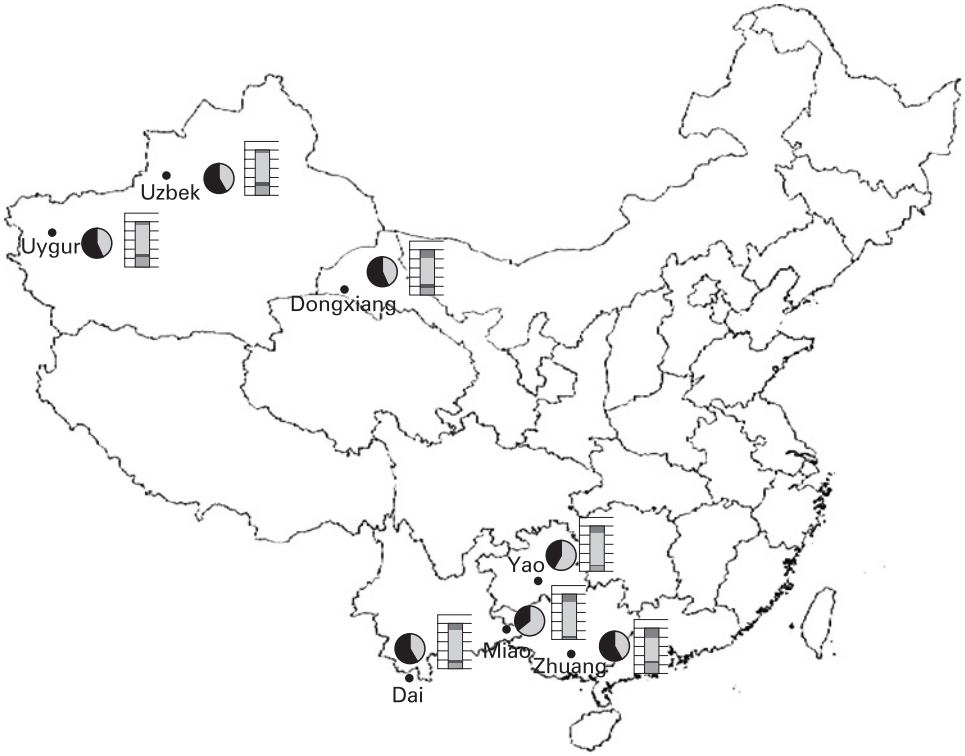

Figure 3 VNTR allele frequency in Chinese ethnic minorities. Homozygosity and allelic frequencies of samples of Chinese ethnic minorities collected from different regions of China. The pie chart is the homozygote proportion (grey) *vs* heterozygote (black). The stacked bar shows the allele distribution, the bottom one representing 5-repeat and the top one 9-repeat.

Key points

- *CLEC4M* is one member of the C-type lectin gene cluster in chromosome 19. Recently it has been shown to be a receptor for the severe acute respiratory syndrome (SARS) virus and a VNTR polymorphism in its neck region has been associated with susceptibility to infection. However, this association was controversial and not supported by subsequent studies.
- A comprehensive genetic association study on the whole C-type lectin gene cluster indicated that no significant association with disease predisposition was detected.
- On the other hand, we detected a population structure of the VNTR alleles in a large Chinese sample collected from different regions of China, which suggested that a previous positive association with *CLEC4M* was likely confounded by population stratification.

most well recorded Northern Han immigrants in Guangdong and this sample revealed the huge effect of immigrants on homozygosity of this VNTR. If we excluded the Meizhou sample, it is apparent that the north–south differentiation of homozygosity could be partially accounted for by a higher frequency of 6-repeat allele among the Southern Han (~6% *vs* ~3%, $p<0.05$ after Bonferroni correction).

These results are consistent with the hypothesis that a higher degree of admixture is present in Southern Han due to historical migration events.[29] When compared to published data, the distribution of genotypes and alleles in the Zhanjiang sample basically followed what had been reported in Hong Kong by Tang *et al*.[13] On the other hand, the homozygote proportions reported in the three sets of controls by Chan *et al* were all significantly higher than our Guangdong/Zhanjiang sample (by χ^2: blood donor control: $p = 0.05$, outpatient control: $p = 0.03$, health care worker control: $p = 0.02$). However, it was similar to our Guangdong/Meizhou sample, which suggested that the control samples of Chan *et al* might be biased towards Han of northern origin and might well be a stratified sample.

Furthermore, our data from ethnic minorities demonstrated that population stratification also exists among these ethnic minorities, and different origin and cultural isolation are important contributing factors toward allelic and homozygosity differentiation. Although we do not rule out the role of selection in this process, the contrasting difference of allelic distribution among minorities living in close proximity in Southwest China suggested that isolation played a major role in shaping the allelic distribution. Meanwhile, the subpopulation structure can also be affected by the other factors, like migration and admixture, such as the northwest minorities and Southern Han in our studies.

Population structure has also been shown in this VNTR polymorphism across ethnic groups in a global perspective. Barreiro *et al*[25] investigated the population diversity of *CLEC4M* VNTR in the Centre d'Etudes du Polymorphism Human panel (CEPH), which included 1064 individuals from 52 worldwide populations. Their data showed the allelic distributions and therefore homozygote proportions were different across ethnic groups. From their results, highest homozygosity (54.63%) was found among the Native Americans while the lowest value was in the Oceanian populations (28.21%), and the difference was significant ($p<0.05$ by GENEPOP). Based on additional SNP data, they concluded that a balancing selection sharpened the variation in this VNTR in the non-African populations and contributed to the high level of variability in *CLEC4M* when compared to *CD209*.[30]

Finally, we concluded that VNTR of *CLEC4M* and other genetic variations in the C-type lectin gene cluster of chromosome 19 did not predispose or affect the prognosis after SARS-CoV infection in the Chinese population. This study suggests that population stratification is an important factor which should be taken into consideration in genetic association studies.

Acknowledgements: We are grateful to the members for participating in this research. We thank Drs Wen-Zhi Wang, Liang Xie and Mr Wai Chiu Suen for their help in preparing the manuscript.

Funding: This work was supported by grants the National Natural Science Foundation of China (30621092), Bureau of Science and Technology of Yunnan Province and a Direct grant from the Faculty of Medicine, The Chinese University of Hong Kong (2007.1.019).

Competing interests: None declared.

Patient consent: Obtained

REFERENCES

1. **Tsang KW,** Ho PL, Ooi GC, Yee WK, Wang T, Chan-Yeung M, Lam WK, Seto WH, Yam LY, Cheung TM, Wong PC, Lam B, Ip MS, Chan J, Yuen KY, Lai KN. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 2003;**348**:1977–85.
2. **Peiris JS,** Yuen KY, Osterhaus AD, Stohr K. The severe acute respiratory syndrome. *N Engl J Med* 2003;**349**:2431–41.
3. **Lee N,** Hui D, Wu A, Chan P, Cameron P, Joynt GM, Ahuja A, Yung MY, Leung CB, To KF, Lui SF, Szeto CC, Chung S, Sung JJ. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 2003;**348**:1986–94.
4. **World Health Organization**. Cumulative number of reported probable cases of SARS. 2004. http://www.who.int/csr/sars/country/2003_07_11/en/ (Accessed May 2004).
5. **Lin M,** Tseng HK, Trejaut JA, Lee HL, Loo JH, Chu CC, Chen PJ, Su YW, Lim KH, Tsai ZU, Lin RY, Lin RS, Huang CH. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet* 2003;**4**:9.
6. **Ng MH,** Lau KM, Li L, Cheng SH, Chan WY, Hui PK, Zee B, Leung CB, Sung JJ. Association of human-leukocyte-antigen class I (B*0703) and class II (DRB1*0301) genotypes with susceptibility and resistance to the development of severe acute respiratory syndrome. *J Infect Dis* 2004;**190**:515–8.
7. **Li W,** Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;**426**:450–4.
8. **Chiu RW,** Tang NL, Hui DS, Chung GT, Chim SS, Chan KC, Sung YM, Chan LY, Tong YK, Lee WS, Chan PK, Lo YM. ACE2 gene polymorphisms do not affect outcome of severe acute respiratory syndrome. *Clin Chem* 2004;**50**:1683–6.
9. **Chan KC,** Tang NL, Hui DS, Chung GT, Wu AK, Chim SS, Chiu RW, Lee N, Choi KW, Sung YM, Chan PK, Tong YK, Lai ST, Yu WC, Tsang O, Lo YM. Absence of association between angiotensin converting enzyme polymorphism and development of adult respiratory distress syndrome in patients with severe acute respiratory syndrome: a case control study. *BMC Infect Dis* 2005;**5**:26.
10. **Tang NL,** Chan PK, Wong CK, To KF, Wu AK, Sung YM, Hui DS, Sung JJ, Lam CW. Early enhanced expression of interferon-inducible protein-10 (CXCL-10) and other chemokines predicts adverse outcome in severe acute respiratory syndrome. *Clin Chem* 2005;**51**:2333–40.
11. **Lo AW,** Tang NL, To KF. How the SARS coronavirus causes disease: host or organism? *J Pathol* 2006;**208**:142–51.
12. **Chan VS,** Chan KY, Chen Y, Poon LL, Cheung AN, Zheng B, Chan KH, Mak W, Ngan HY, Xu X, Screaton G, Tam PK, Austyn JM, Chan LC, Yip SP, Peiris M, Khoo US, Lin CL. Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nat Genet* 2006;**38**:38–46.
13. **Tang NL,** Chan PK, Hui DS, To KF, Zhang W, Chan FK, Sung JJ, Lo YM. Lack of support for an association between CLEC4M homozygosity and protection against SARS coronavirus infection. *Nat Genet* 2007;**39**:691–2.
14. **Zhi L,** Zhou G, Zhang H, Zhai Y, Yang H, Zhang F, Wang S, Wei M, Cao W, He F. Lack of support for an association between CLEC4M homozygosity and protection against SARS coronavirus infection. *Nat Genet* 2007;**39**:692–4.
15. **Marzi A,** Gramberg T, Simmons G, Moller P, Rennekamp AJ, Krumbiegel M, Geier M, Eisemann J, Turza N, Saunier B, Steinkasserer A, Becker S, Bates P, Hofmann H, Pohlmann S. DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *J Virol* 2004;**78**:12090–5.
16. **Yang ZY,** Huang Y, Ganesh L, Leung K, Kong WP, Schwartz O, Subbarao K, Nabel GJ. pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. *J Virol* 2004;**78**:5642–50.

17. **Gramberg T,** Hofmann H, Moller P, Lalor PF, Marzi A, Geier M, Krumbiegel M, Winkler T, Kirchhoff F, Adams DH, Becker S, Munch J, Pohlmann S. LSECtin interacts with filovirus glycoproteins and the spike protein of SARS coronavirus. *Virology* 2005;**340**:224–36.
18. **Soilleux EJ,** Barten R, Trowsdale J. DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. *J Immunol* 2000;**165**:2937–42.
19. **Hui DS,** Sung JJ. Treatment of severe acute respiratory syndrome. *Chest* 2004;**126**:670–4.
20. **Sung JJ,** Wu A, Joynt GM, Yuen KY, Lee N, Chan PK, Cockram CS, Ahuja AT, Yu LM, Wong VW, Hui DS. Severe acute respiratory syndrome: report of treatment and outcome after a major outbreak. *Thorax* 2004;**59**:414–20.
21. **Zhang K,** Jin L. HaploBlockFinder: haplotype block analyses. *Bioinformatics* 2003;**19**:1300–1.
22. **Tang NL,** Pharoah PD, Ma SL, Easton DF. Evaluation of an algorithm of tagging SNPs selection by linkage disequilibrium. *Clin Biochem* 2006;**39**:240–3.
23. **Yao YG,** Lu XM, Luo HR, Li WH, Zhang YP. Gene admixture in the silk road region of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes Genet Syst* 2000;**75**:173–8.
24. **Yao YG,** Kong QP, Wang CY, Zhu CL, Zhang YP. Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. *Mol Biol Evol* 2004;**21**:2265–80.
25. **Barreiro LB,** Quintana-Murci L. DC-SIGNR neck-region polymorphisms and HIV-1 susceptibility: from population stratification to a possible advantage of the 7/5 heterozygous genotype. *J Infect Dis* 2006;**194**:1184–5; author reply 5–7.
26. **Itoyama S,** Keicho N, Quy T, Phi NC, Long HT, Ha le D, Ban VV, Ohashi J, Hijikata M, Matsushita I, Kawana A, Yanai H, Kirikae T, Kuratsuji T, Sasazuki T. ACE1 polymorphism and progression of SARS. *Biochem Biophys Res Commun* 2004;**323**:1124–9.
27. **Yuan FF,** Boehm I, Chan PK, Marks K, Tang JW, Hui DS, Sung JJ, Dyer WB, Geczy AF, Sullivan JS. High prevalence of the CD14-159CC genotype in patients infected with severe acute respiratory syndrome-associated coronavirus. *Clin Vaccine Immunol* 2007;**14**:1644-5.
28. **Chan KY,** Ching JC, Xu MS, Cheung AN, Yip SP, Yam LY, Lai ST, Chu CM, Wong AT, Song YQ, Huang FP, Liu W, Chung PH, Leung GM, Chow EY, Chan EY, Chan JC, Ngan HY, Tam P, Chan LC, Sham P, Chan VS, Peiris M, Lin SC, Khoo US. Association of ICAM3 genetic variant with severe acute respiratory syndrome. *J Infect Dis* 2007;**196**:271–80.
29. **Yao YG,** Kong QP, Bandelt HJ, Kivisild T, Zhang YP. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 2002;**70**:635–51.
30. **Barreiro LB,** Patin E, Neyrolles O, Cann HM, Gicquel B, Quintana-Murci L. The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. *Am J Hum Genet* 2005;**77**:869–86.

Polymorphisms in the C-type lectin genes cluster in chromosome 19 and predisposition to severe acute respiratory syndrome coronavirus (SARS-CoV) infection

H Li, N L-S Tang, P K-S Chan, C-Y Wang, D S-C Hui, C Luk, R Kwok, W Huang, J J-Y Sung, Q-P Kong and Y-P Zhang

J Med Genet 2008 45: 752-758 originally published online August 12, 2008
doi: 10.1136/jmg.2008.058966