

ORIGINAL ARTICLE

DEFA gene variants associated with IgA nephropathy in a Chinese population

YY Qi^{1,2,3,4}, XJ Zhou^{1,2,3,4}, FJ Cheng^{1,2,3,4}, P Hou^{1,2,3,4}, L Zhu^{1,2,3,4}, SF Shi^{1,2,3,4}, LJ Liu^{1,2,3,4}, JC Lv^{1,2,3,4} and H Zhang^{1,2,3,4}

IgA nephropathy (IgAN) is a complex syndrome with high genetic heterogeneity. More recently, a genome-wide association study (GWAS) from Southern Han population revealed that variants within 8p23.1, where the *DEFA* genes encoding α -defensins assembled, were associated with susceptibility to IgAN. To replicate the association and fine-map the genetic variants, a case-control genetic study from an independent Northern Han cohort was conducted. A total of 60 single-nucleotide polymorphisms in a region spanning 350 kb encompassing the *DEFA* genes cluster were analyzed in 2096 individuals. Copy number variations of *DEFA1A3* within the loci were also checked for the independent association. Functional significance of the associated variants was further examined by the *in silico* method as well as by *cis*-acting expression quantitative trait loci analysis with mRNA. It showed that 17 out of 60 (28.3%) variants were associated with susceptibility to IgAN. Two independent signals with functional potentials were discovered (rs2738058, $P = 4.64 \times 10^{-5}$, odds ratio (OR) = 0.76, 95% confidence interval (CI) 0.66–0.87 and rs9644778, $P = 4.78 \times 10^{-3}$, OR = 1.21, 95% CI 1.06–1.39). Besides, marginally significant association of rs9644778 risk genotype with lower proportion of gross hematuria (CC+CA vs AA 35.2% vs 30.2%, $P = 0.073$) was observed. In conclusion, *DEFA* gene polymorphisms have potentially pathogenic roles in IgAN, and the role of mucosal immunity in the pathogenesis of IgAN has to be emphasized.

Genes and Immunity (2015) 16, 231–237; doi:10.1038/gene.2015.1; published online 12 February 2015

INTRODUCTION

IgA nephropathy (IgAN), the most common cause of end-stage renal disease in Asian population, is a genetically complex and clinically heterogeneous renal disease.¹ It is characterized as immune complexes containing IgA deposited in the glomerular mesangium, in renal biopsy. Although the pathogenesis of IgAN is not clear, the four-hit model (glycosylation deficiency of IgA1, synthesis of antiglycan IgG antibodies, formation of IgA1-containing immune complex and mesangial deposition of immune complex) suggesting abnormality in the innate and adaptive immunity may be involved in its pathogenesis.² The familial aggregation, different prevalence among different ethnic populations and variable disease courses in individual patients support that genetic predisposing factors contribute to IgAN susceptibility and development.^{3–8}

The IgAN susceptibility loci discovered in the genome-wide association studies (GWASs) were mainly from eight genomic regions, including the *MHC* region at 6p21, *DEFA* locus at 8p23, *TNFSF13* locus at 17p23, *HORMAD2* locus at 22q12, *CFH/CFHR* locus at 1q32, *ITGAM-ITGAX* locus at 16p11, *VAV3* locus at 1p13 and *CARD9* locus at 9q34.^{5,6,9} Genetic studies have provided a glimpse into IgAN pathogenesis and identified the involvement of defects in adaptive immunity, innate immunity, mucosal immunity and the alternative complement pathway.² Although the role of mucosal immunity in IgAN is still controversial, the loci including *TNFSF13*, *DEFA* and *HORMAD2* discovered in the

GWAS of IgAN were all likely implicated in the response of mucosal immunity, suggesting a deregulation role due to genetic defects in mucosal immunity. More recently, GWAS conducted in 4 discovery cohorts and 10 replication cohorts covering European and East Asian, the two major ancestries, confirmed that *DEFA* locus is strongly associated with IgAN susceptibility.⁹ *DEFA* locus, in which six human α -defensin peptides are encoded by only five *DEFA* genes, encodes α -defensins, which are antimicrobial peptides involved in maintenance of the intestinal mucosal barrier or regulation of the mucosal immune response.¹⁰ The *DEFA1* and *DEFA3* genes are interchangeable occupants of a 19-kb copy-variable unit. Both single-nucleotide polymorphisms (SNPs) and copy number variations (CNVs) can associate with diseases and influence gene expression. It has been reported that they have an effect on several diseases.^{11–18} Two GWASs discovered variants in the association of *DEFA* locus with IgAN susceptibility. However, the top SNPs were inconsistent.^{6,9} Replication by others can ensure that the original findings are robust and can provide a more accurate estimate of the likely effect size. Besides, variants discovered by previous GWASs were within the noncoding sequences of a gene or intergenic regions. To integrate expression data from diverse tissue types, genome and experimental-based annotation data covering noncoding and clinical phenotype association analysis are prominent. From a

¹Renal Division, Peking University First Hospital, Beijing, People's Republic of China; ²Peking University Institute of Nephrology, Beijing, People's Republic of China; ³Key Laboratory of Renal Disease, Ministry of Health of China, Beijing, People's Republic of China and ⁴Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Education, Beijing, People's Republic of China. Correspondence: Professor H Zhang, Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Key Laboratory of Renal Disease, Ministry of Health of China, Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Education, No. 8, Xishiku Street, Xicheng District, Beijing 100034, People's Republic of China.

E-mail: hongzh@bjmu.edu.cn

Received 19 August 2014; revised 23 December 2014; accepted 29 December 2014; published online 12 February 2015

more important aspect for disease pathogenesis, the underlying functional significance needs to be further addressed.

In this study, we focus on exploring independent susceptibility variants in the *DEFA* region in a bid to provide evidence *DEFA* gene involvement in the pathogenesis of IgAN.

RESULTS

The significant SNPs in the *DEFA* region associated with IgAN susceptibility

A total of 60 SNPs in the *DEFA* region were analyzed (Figure 1). Seventeen (28.3%) of them were significantly associated with the susceptibility to IgAN ($P < 0.05$; Table 1). Among them, rs2738058 located in 14 kb 3' of the *DEFA1* intergenic region (GRCh37/hg19) showed the strongest association signal ($P = 4.64 \times 10^{-5}$, odds ratio (OR) = 0.76, 95% confidence interval (CI) 0.66–0.87). The association of top SNP rs2738048 identified in the previous GWAS with susceptibility to IgAN was also confirmed ($P = 2.13 \times 10^{-3}$, OR = 0.81, 95% CI 0.71–0.93).⁶ Power calculations indicated that we had at least 98.6% power to detect loci with allelic frequencies > 0.10 and relative risk > 1.5 assuming an α -level of 0.05 ($P < 0.05$) in the current study. Detailed power calculation data for the identified 17 variants are shown in Supplementary Table 1.

Further linkage disequilibrium (LD) analysis in the control groups revealed that these 17 SNPs can be grouped into four clusters (Figure 2). The most strongly associated SNPs from each block were chosen as tag SNPs, including rs2738058 (a variant at 14 kb in 3' of *DEFA1*), rs2702910 (a variant 11 kb 3' of *DEFA1*), rs4300027 (a variant 4.8 kb 5' of *DEFA3*) and rs9644778 (a variant 3.2 kb 3' of *DEFA5*), and together with rs2738048 (a variant 12 kb 3' of *DEFA1*), which was the top signal from the GWAS from Southern Han, for identifying the independent susceptibility variant associated with IgAN.

To test the independence of the identified five SNPs, a stepwise logistic regression analysis was performed using the forward method. rs2738058 showed the strongest association with IgAN (OR = 1.25(1.08–1.43); $P = 5.82 \times 10^{-5}$). The second independent locus was rs9644778 (OR = 1.15(1.00–1.32); $P = 0.015$; Supplementary Table 2). Conditional logistic regression analysis

was also performed to investigate the independent effects of the identified five SNPs, which showed the same results (Supplementary Table 3). Accordant with the data from the stepwise logistic regression analysis and conditional logistic regression analysis, rs2738058 and rs9644778 were in low LD ($r^2 = 0.16$). These suggest that rs2738058 and rs9644778 were independent variants associated with susceptibility to IgAN.

Association analysis between *DEFA1A3* CNVs and SNPs

To determine whether the association of the SNPs within *DEFA* loci with IgAN is independent or is due to their LD with *DEFA1A3* CNVs, we conducted a two-stage association analysis considering *DEFA1A3* copy numbers (CNs).

In the first stage, we correlated the two different variants (SNP and CNV) in 45 HapMap Chinese Han Beijing (CHB) individuals by χ^2 test and calculated the Pearson's correlation coefficients for LD estimation. The previously GWAS-reported variant rs2738048, which was located at 12 kb in 3' of *DEFA1*, was significantly associated with CNs of *DEFA1A3*, with the risk allele correlating with lower *DEFA1A3* CNs ($P = 0.012$, $r = -0.370$). Another marginal significance was observed between rs2702910 risk genotypes and lower CNs ($P = 0.124$, $r = -0.233$). For validation, we further checked *DEFA1A3* CNs of 70 healthy controls with Northern Chinese ethnicity and their associations with the two tag SNPs. Similar to data from HapMap CHB, the correlation coefficients observed in the current Beijing controls were also approximately -0.3 ($P = 0.012$ and $r = -0.298$ for rs2738048, $P = 0.033$ and $r = -0.255$ for rs2702910; Figure 3). However, none of rs2738058, rs9644778 and rs4300027 genotypes was observed to correlate with CNs of *DEFA1A3* in any of the cohorts (Supplementary Figure 1).

Regulatory variant annotation and *cis*-eQTL (expression quantitative trait loci) analysis exploring possible functional loci
As the majority of associated variants mapped to noncoding intergenic regions, we first checked the functionality of lead SNPs, including rs2738048, rs2738058 and rs9644778, by rSNPBase. In the rSNPBase, all SNPs were annotated with references to

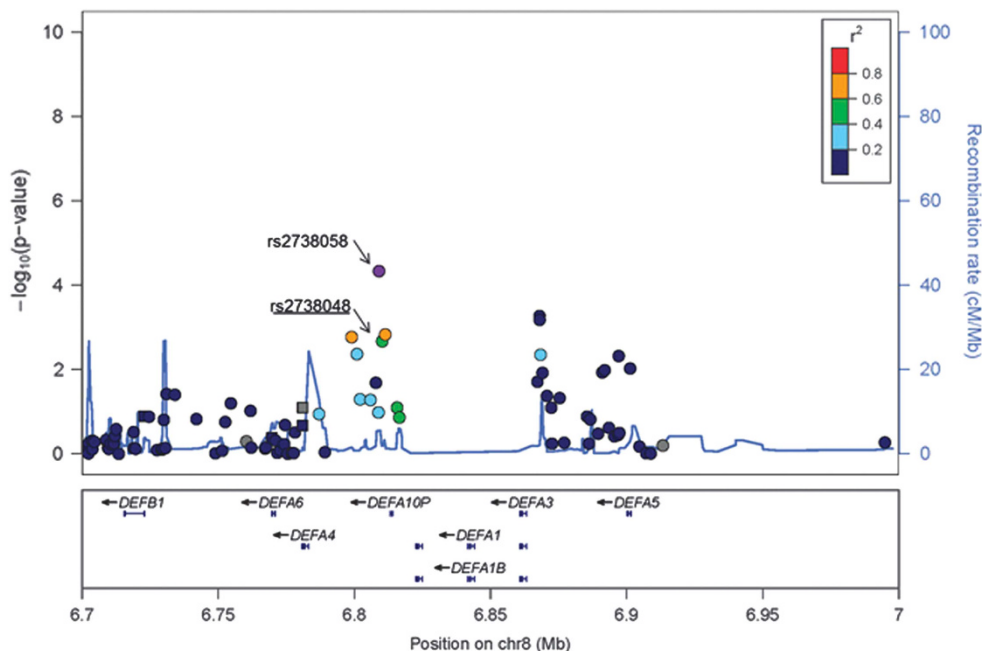


Figure 1. Regional association plots for SNPs in the *DEFA* gene cluster. *x* axis: chromosomal position (GRCh37/hg19). *y* axis: association with IgAN from our current analysis ($-\log_{10} P$ -value). The figure was generated using Locus Zoom.

Table 1. The association of *DEFA* region polymorphisms with IgAN

CHR	SNP	BP	Location	Minor allele	MAF in case	MAF in control	Trend test P-value	OR
8	rs2741658	6798969	Intergenic 16 kb 5' of <i>DEFA4</i>	A	0.22	0.27	1.70×10^{-3}	0.80 (0.65–1.86)
8	rs2702852	6800911	Intergenic 18 kb 5' of <i>DEFA4</i>	G	0.21	0.25	4.29×10^{-3}	0.81 (0.70–0.94)
8	rs12680018	6807874	Intergenic 15 kb 3' of <i>DEFA1</i>	G	0.18	0.15	2.06×10^{-2}	1.22 (1.03–1.44)
8	rs2738058	6809027	Intergenic 14 kb 3' of <i>DEFA1</i>	G	0.27	0.33	4.64×10^{-5}	0.76 (0.66–0.87)
8	rs2738048	6810195	Intergenic 12 kb 3' of <i>DEFA1</i>	C	0.27	0.31	2.13×10^{-3}	0.81 (0.71–0.93)
8	rs2702910	6811301	Intergenic 11 kb 3' of <i>DEFA1</i>	T	0.24	0.28	1.48×10^{-3}	0.80 (0.69–0.92)
8	rs7825750	6867213	Intergenic 4 kb 5' of <i>DEFA3</i>	C	0.39	0.35	1.95×10^{-2}	1.16 (1.03–1.32)
8	rs4300027	6867985	Intergenic 4.8 kb 5' of <i>DEFA3</i>	T	0.32	0.37	5.37×10^{-4}	0.80 (0.70–0.91)
8	rs4512398	6868004	Intergenic 4.8 kb 5' of <i>DEFA3</i>	C	0.32	0.37	6.65×10^{-4}	0.80 (0.70–0.91)
8	rs7826487	6868335	Intergenic 5.1 kb 5' of <i>DEFA3</i>	G	0.05	0.07	4.45×10^{-3}	0.69 (0.54–0.89)
8	rs4288398	6869048	Intergenic 5.8 kb 3' of <i>DEFA3</i>	A	0.31	0.35	1.19×10^{-2}	0.85 (0.74–0.96)
8	rs9774483	6870701	Intergenic 7.5 kb 5' of <i>DEFA3</i>	A	0.21	0.18	4.23×10^{-2}	1.18 (1.01–1.37)
8	rs6982814	6875366	Intergenic 12 kb 5' of <i>DEFA3</i>	G	0.21	0.18	4.75×10^{-2}	1.17 (1.00–1.37)
8	rs12682030	6891017	Intergenic 9.2 kb 3' of <i>DEFA5</i>	C	0.08	0.10	1.17×10^{-2}	0.76 (0.61–0.94)
8	rs6999181	6891919	Intergenic 8.3 kb 3' of <i>DEFA5</i>	A	0.08	0.10	1.04×10^{-2}	0.75 (0.61–0.94)
8	rs9644778	6897064	Intergenic 3.2 kb 3' of <i>DEFA5</i>	C	0.32	0.28	4.78×10^{-3}	1.21 (1.06–1.39)
8	rs10095331	6901260	Intron <i>DEFA5</i>	A	0.08	0.10	9.45×10^{-3}	0.75 (0.61–0.93)

Abbreviations: BP, base pair; CHR, chromosome; IgAN, IgA nephropathy; MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.

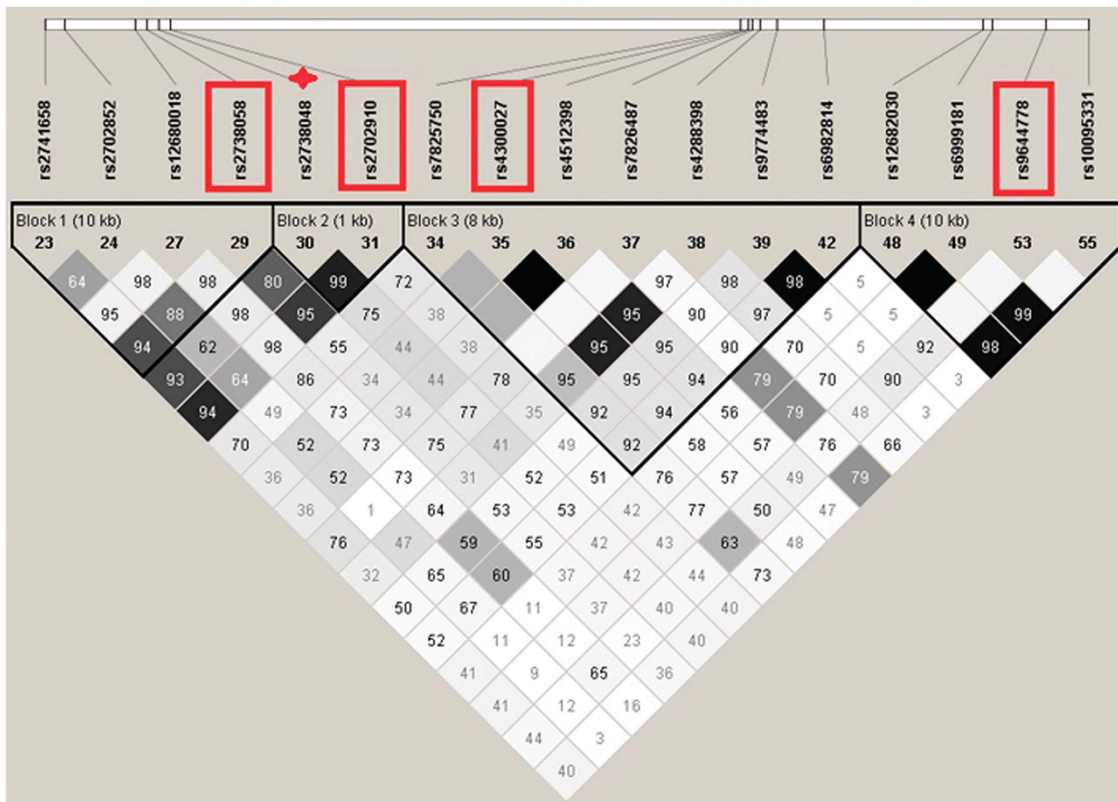


Figure 2. The LD maps of the 17 identified disease-associated SNPs in 902 healthy controls. The figure was generated using the Haploview software, and the LD was estimated using the r^2 and the blocks defined according to CIs.³⁸ The SNP marked with cross was the previous GWAS identified, rs2738048. The most strongly associated SNPs from each block, which were chosen as tag SNPs, are marked in the boxes.

experimentally supported regulatory elements, which improved the reliability.¹⁹

SNP annotations indicated that both rs2738058 and rs9644778 rather than rs2738048 were regulatory SNPs. Considering regulation types, by ChIP-seq technology, it was observed that rs2738058 affects proximal transcriptional regulation, whereas rs9644778 affects distal transcriptional regulation;²⁰ rs2738058 was located within different histone-marked regions in multiple cell lines, which was experimentally supported by the Broad Institute using ChIP-seq

technology from ENCODE data. In T-lymphocyte and leukemia cell lines, rs2738058 was located within H3K4me1, H3K9ac, H3K9ac and H3K27ac histone-marked regions. In lung epithelial cells, B lymphocytes, embryonic stem cells, liver carcinoma cells and myoblast cells, rs2738058 was located within H4K20me1 histone-marked regions. rs2738058 was located within H3K9me1 histone-marked regions in leukemia cells and keratinocyte cells. rs9644778 affects distal transcriptional regulation in breast adenocarcinoma MCF-7 cell lines using ChIA-PET (Genome Institute of Singapore, Singapore) tools.²⁰

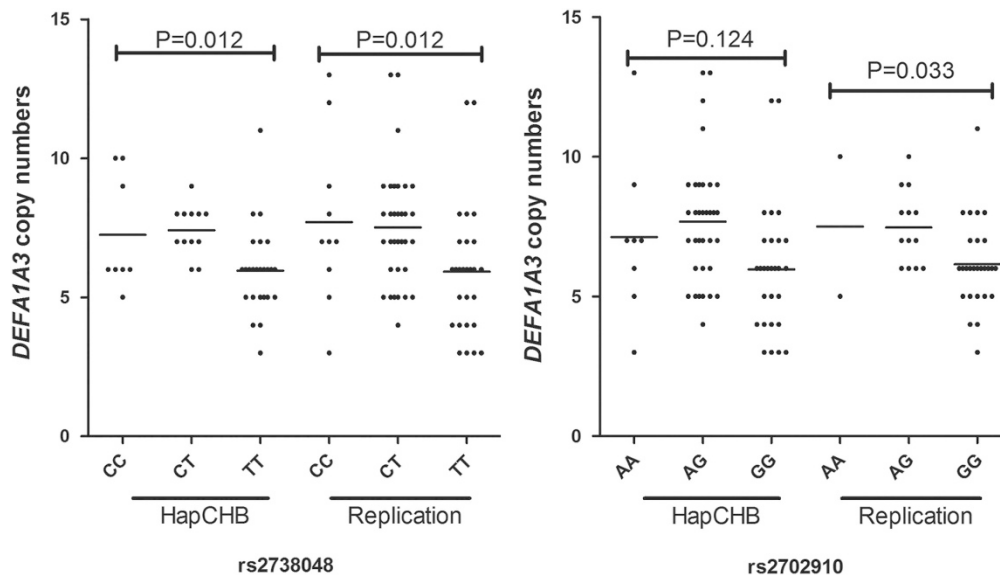


Figure 3. The correlation between *DEFA1A3* CNs and SNPs in HapMap CHB population. The replication cohort was analyzed by χ^2 test and the Pearson's correlation coefficient was calculated.

It was reported that the functional SNP most strongly supported by experimental evidence was possibly an SNP in LD with the reported association rather than the reported SNP itself.²¹ We further located the set of SNPs that were in strong LD ($r^2 \geq 0.8$) with leading SNPs in all HapMap populations, and annotated the SNPs using a scoring scheme integrated in Regulome DB. A total of 4 out of 7 SNPs and 6 out of 13 SNPs were in strong LD with regulatory SNPs rs2738058 and rs9644778. This was supported by more than one single experimental modality (for example, motif-based predictions, DNase I hypersensitivity peaks and ChIP-seq peaks).²⁰ Among the many predicted regulatory SNPs, rs2615787 (17 kb 5' of *DEFA4*, in LD with rs2738058, $r^2 = 0.89$ in Chinese Hans living in Beijing) had the highest Regulome DB score 2b (namely, containing predicted transcription factor-binding motifs supported by ChIP-seq, having protein-binding sequences as demonstrated by ChIP-seq, or located in the chromatin structures or the histone modification regions). rs2615787 overlapped with a region of predicted transcription factor-binding motifs (STAT,²² HSF2^{22,23}), protein (CTCF,²⁰ TAL1²⁴) binding, ChIP-seq peak (in 13 types of cell lines, that is, B lymphocyte-Gm12878 and Monocyte-Monocd14ro1746), and a DNase I hypersensitivity peak²⁰ (see Supplementary Table 4).

Supporting eQTL data surely will support a higher functional significance of the associated variants in addition to TF binding, matched TF motif, matched DNase footprint and DNase peak. Gene expression in Epstein-Barr virus-transformed lymphoblastoid B-cell lines from 210 unrelated HapMap samples was utilized to study the correlation between SNP genotypes and gene expression. Risk genotypes of both rs2738058 and rs4300027 showed significant correlation with low *DEFA3* mRNA expression ($P = 1.63 \times 10^{-4}$, $r = 0.265$ and $P = 8 \times 10^{-3}$, $r = 0.187$, respectively).

Genetic association of the variants with specific subphenotypes
Associations of the independent variants with clinical and demographic features of patients with IgAN were further analyzed in our cohort. As presented in Table 2, rs9644778 risk genotypes (CC+CA) showed significant association with early-onset (age ≤ 18) IgAN (17.5% vs 9.4%, $P = 4.69 \times 10^{-5}$, OR = 0.49, 95% CI 0.34–0.69) and marginally significant association with higher proportions of gross hematuria (CC+CA vs AA 35.2 vs 30.2%, $P = 7.3 \times 10^{-2}$).

DISCUSSION

DEFA gene cluster codes for small peptides that are important effector molecules in innate and adaptive immunity. In humans, there are two families of defensins, α -defensins and β -defensins. α -Defensins are mainly expressed in the neutrophils and in the paneth cells of the intestine, whereas β -defensins are expressed by the epithelial tissue.¹⁰ Elevated level of α -defensins was found in type 1 diabetic patients with nephropathy,¹¹ rheumatoid arthritis,¹² systemic lupus erythematosus¹³ and Sjögren's syndrome.¹⁴ CNVs and gene polymorphisms of *DEFA1* or/and *DEFA3* are found associated with multiple immune-related diseases such as Crohn's disease.¹⁷ All the above discoveries highlighted the critical role of α -defensins in autoimmunity, including mucosal innate immunity. SNPs together with CNVs may contribute to the making up of the IgAN susceptibility background.

For IgAN, it was suggested that innate, adaptive and mucosal immunity were involved in inherited defects in IgAN.² Recent GWASs have revealed that the *DEFA* locus at 8p23.1 was associated with susceptibility to IgAN in the Southern Chinese Han population.⁶ In the present study, we investigated 60 SNPs in 8p23.1 spanning 350 kb encompassing the *DEFA* gene cluster and its upstream region. A total of 17 (28.3%) identified SNPs were observed to significantly associate with the susceptibility to IgAN. Two independent signals, rs2738058 and rs9644778, discovered in *DEFA* locus were in low LD, and survived from stepwise analysis and conditional analysis. Multiple variants identified in *DEFA* locus suggested a likely true effect of *DEFA* SNP effects on IgAN susceptibility. Besides, our results also indicated that rs2738048 disease association effects may be partially inflated by *DEFA1A3* CNVs. Just like SNPs, CNVs are also an important source of genetic variation in the human genome. CNVs are expected to have an important role in the genetic basis of complex diseases such as autoimmune diseases and autism.^{25–27} *DEFA1A3* locus on human chromosome 8p23.1 is also a region that exhibits multiallelic CNVs. Future studies addressing the respective contributions of CNVs and SNPs in *DEFA* locus to IgAN are deserved. More recently, Black et al.²⁸ suggested that different structural haplotypes in *DEFA1A3* locus were associated with *DEFA1A3* CNVs. However, they mainly focused on a 4.1-kb region in high LD from a Caucasian population. The top-tagging SNP rs4300027 was also in low LD with ours, r^2 being between rs4300027 and rs2738048 ($r^2 = 0.16$), rs2738058 ($r^2 = 0.10$), rs2702910 ($r^2 = 0.06$) and rs9644778

Table 2. Associations between variants in *DEFA* gene cluster and severity of IgAN

Clinical parameters	rs2738058			rs9644778		
	GG+GA (n = 549)	AA (n = 645)	P-value	CC+CA (n = 635)	AA (n = 557)	P-value
Age onset (≤ 18 years, %)	66 (12.0%)	97 (15.1%)	0.13	111 (17.5%)	52 (9.4%)	4.69×10^{-5} 0.49 (0.34–0.69)
Gender (male, %)	295 (53.7%)	288 (44.7%)	0.57	349 (55.0%)	302 (54.2%)	0.55
SBP (mm Hg)	124.07 \pm 17.33	123.95 \pm 17.92	0.91	123.50 \pm 17.27	124.59 \pm 18.09	0.29
DBP (mm Hg)	78.38 \pm 12.36	78.86 \pm 13.22	0.52	78.35 \pm 12.53	78.95 \pm 13.19	0.43
Uric acid ($\mu\text{mol l}^{-1}$)	349.69 \pm 114.74	347.2 \pm 107.69	0.71	348.68 \pm 111.44	348.11 \pm 110.33	0.93
Scr ($\mu\text{mol l}^{-1}$)	119.13 (109.02–129.25)	113.67 (105.33–122.01)	0.24	117.28 (107.82–126.74)	115.02 (106.31–123.73)	0.30
24-h UTP (g per day)	2.16 (1.94–2.37)	2.22 (2.03–2.40)	0.74	2.21 (2.02–2.40)	2.17 (1.96–2.38)	0.95
Serum IgA (g l ⁻¹)	3.18 \pm 1.19	3.25 \pm 1.26	0.38	3.27 \pm 1.27	3.15 \pm 1.17	0.10
Gross hematuria (%)	182 (33.2%)	209 (32.5%)	0.80	223 (35.2%)	168 (30.2%)	0.07
eGFR (ml min ⁻¹ per 1.73 m ²)	83.71 (80.65–86.76)	87.34 (84.45–90.23)	0.06	86.86 (83.92–89.80)	84.29 (81.28–87.31)	0.20
Histological grading (mild lesions/severe lesions)	80/107	114/131	0.44	107/125	87/113	0.59

Abbreviations: DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; IgAN, IgA nephropathy; SBP, systolic blood pressure; Scr, serum creatinine; 24 hour UTP, 24 hours total urine protein. Associations between variants in *DEFA* gene cluster and severity of IgAN. GFR was estimated based on the Modified Diet in Renal Disease equation for the Chinese population. As none among the serum creatinine level, proteinuria or eGFR rate conformed to normal distribution, we performed a nonparametric Mann–Whitney *U*-test. Other parameters used Student's *t*-test for continuous variables or the χ^2 test for dichotomous variables. Histological grading: mild lesions refer to grades I and II; grades III, IV and V denote severe lesions. Associations with statistical significance or marginal significance were marked in bold.

($r^2 = 0.04$; LD relationships between SNPs were adopted based on the HapMap data using CHB samples). Thus, the structural haplotypes they observed may not be well tagged in our cohort and need further testing in multiple populations, because repeated data supported that CNVs of the region suffer from power of selection. In addition, both rs2738058 and rs9644778 are expected to be highly functional variants in *DEFA* gene cluster, which may act as regulatory factors for gene expression. From a GWAS of chronic periodontitis conducted in a general German population, rs2738058 was also observed to be one of the top association signals, further suggesting that rs2738058 may be one of the causal variants.²⁹ The risk genotypes (CC+CA) of rs9644778 were associated with gross hematuria marginally, which indicates that the SNPs in *DEFA* loci may not only influence disease susceptibility but also impact IgAN severity. Hematuria, a typical clinical presentation secondary to upper respiratory or gastrointestinal illness in IgAN patients, was speculated to be caused by the synergistic effect of innate, mucosal and adaptive immunity.^{30,31} Although limited by the small sample size, as more IgAN patients with detailed clinical data are needed to confirm the result, our findings support the notion that α -defensins are involved in the inherited defect in IgAN, via innate and mucosal immunity disorders. *Cis*-eQTL analysis showed that risk genotypes of both rs2738058 and rs4300027 showed significant correlations with lower gene expressions. When α -defensin concentration was low, human monocyte-derived dendritic cells elevated secretions of tumor-necrosis factor- α and interleukin-1 β , contributing to aggregation of inflammatory medium.³² The risk genotypes correlated with lower gene expressions may lead to low serum α -defensin concentration, which would provide a proinflammatory environment in glomeruli mesangial cells to amplify proinflammatory loop by upregulation of IgA receptors in mesangial cells.³³ In this case, more IgA may be deposited in mesangial cells, which is consistent with renal biopsy specimens. The data above showed the genetic, expression and clinical manifestation association of *DEFAs* and IgAN. However, the association in IgAN was relatively weak and precise mechanisms still need to be determined.

GWASs have identified numerous susceptibility loci for diseases or traits. One of the greatest challenges in the post-GWAS era is to clarify the functional variants within the identified loci. In this study, we investigated multiple SNPs spanning a 350-kb range in *DEFA* loci among 2096 CHB. We discovered two independent SNPs with potential regulatory functions and successfully replicated

previously GWAS-reported rs2738048. Genetic, expression and clinical manifestation association suggested that *DEFA* gene polymorphisms have potentially pathogenic roles in IgAN. The role of mucosal immunity in the pathogenesis of IgAN should be emphasized.

MATERIALS AND METHODS

Patients and controls

The cases and controls analyzed in this study were from our previous GWAS cohort,⁵ including 1194 renal biopsy-proved IgA nephropathy patients and 902 healthy controls of Chinese Han ancestry from Northern China. Participants in this cohort were geographically and ethnically matched and statistically analyzed in a previous report.⁵ The study got the approval from the Medical Ethics Committee of the Peking University First Hospital and written informed consent was obtained from all the patients.

SNP selection and genotyping

For genotyping, a customized Illumina (San Diego, CA, USA) Human 610-Quad BeadChip platform was applied and genotyping yield was over 99%.⁵ A region spanning 350 kb (chromosome 8, 6 750 000–7 100 000, GRCh37/hg19) encompassing the *DEFA* gene cluster (132 kb, GRCh37/hg19) and its upstream (*DEFA5*, 5' region, 200 kb, GRCh37/hg19) was selected, and a total of 60 SNPs within this region were included and analyzed in the current study.

Determination of *DEFA1A3* CNVs

The relative standard curve method of quantitative real-time PCR (7500 Real-time PCR System, Applied Biosystems, Foster City, CA, USA) was used to determine *DEFA1A3* CNVs, as described previously,³⁴ without modifications of primers and PCR conditions. For quality control, an internal control from HapMap CHB population (NA18570) was selected and randomly assigned in every plate for calibration. Seven serial 1:2 dilutions of genomic DNA of the calibrator were used to generate standard curves on each PCR plate and the minimal DNA content was 0.5 ng in the reaction. The CN count of each sample was determined by the average of three times of repeated amplification. For methodological validations of the *DEFA1A3* CNs determined by qPCR, we compared our results with the sequencing data in the database obtained from the Complete Genomics Institute (<http://www.completegenomics.com/>).³⁵ Our qPCR results were totally concordant with sequencing data in the database from Complete Genomics Institute. The resulting CN between our qPCR and sequencing data was consistent in triplicated qPCR assays.

Cis expression quantitative trait loci analysis

SNP–gene associations in eQTL were analyzed using data from Epstein–Barr virus-transformed lymphoblastoid cell lines of 210 unrelated HapMap samples, as previously reported.³⁶ Normalized mRNA data were retrieved from the database of the Gene Expression Variation (GENEVAR) project at the Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/humgen/genevar/>).

Bioinformatic analysis

Regulatory SNPs were analyzed by rSNPBase (<http://rsnp.psych.ac.cn/>).¹⁹ Annotations of variants, located in the transcription factor-binding motifs, protein-binding sequences, chromatin structures or the histone modification regions, were obtained from the Regulome DB database (<http://www.regulomedb.org>) in ENCODE project.³⁷

Statistical analysis

The identified 60 SNPs within *DEFA* gene cluster were consistent with Hardy–Weinberg equilibrium genotype frequency expectations ($P > 0.05$) using χ^2 goodness-of-fit test. The degrees of LD were estimated by CI method using Haploview4.1 (Cambridge, MA, USA). Allelic and genotypic associations were accessed with PLINK. A stepwise logistic regression analysis using forward method was applied to determine the independence of the SNP. Univariate linear regression was used to determine the correlations between genotypes of associated SNPs and CNs. Power was calculated by Power and Sample Size Calculations Software (<http://biostat.mc.vanderbilt.edu/>). In *cis*-eQTL analysis, Spearman's coefficient was calculated.

Means and s.d. were used to summarize normally distributed variables. Median and interquartile ranges were used for non-normally distributed variables. Categorical data were summarized by ratios and percentages. Student's *t*-test, nonparametric Mann–Whitney *U*-test and one-way analysis of variance were performed for comparison of continuous variables. χ^2 test or logistic regression analysis was used for categorical variables. A two-tailed *P*-value < 0.05 was considered statistically significant. Statistical analysis was performed with the SPSS 16.0 software package (SPSS Inc., Chicago, IL, USA).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank our collaborators of Ali G Gharavi from the Department of Medicine, Columbia University College of Physicians and Surgeons, New York City, NY, USA, for genotyping sorting and assistance. We also thank all the members of the laboratory for technical assistance and the patients and their families for their cooperation and for giving consent to participate in this study. This work was supported by grants from the National Natural Science Foundation of China (No. 81200524), The Foundation of Ministry of Education of China (20120001120008), Major State Basic Research Development Program of China (973 program, No. 2012CB517700), Research Fund of Beijing Municipal Science and Technology for the Outstanding PhD Program (20121000110) and Natural Science Fund of China to the Innovation Research Group (81321064).

DISCLAIMER

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

- 1 Tsukamoto Y, Wang H, Becker G, Chen HC, Han DS, Harris D *et al*. Report of the Asian Forum of Chronic Kidney Disease Initiative (AFCKDI) 2007. "Current status and perspective of CKD in Asia": diversity and specificity among Asian countries. *Clin Exp Nephrol* 2009; **13**: 249–256.
- 2 Kiryluk K, Novak J, Gharavi AG. Pathogenesis of immunoglobulin A nephropathy: recent insight from genetic studies. *Annu Rev Med* 2013; **64**: 339–356.
- 3 Kiryluk K, Julian BA, Wyatt RJ, Scolari F, Zhang H, Novak J *et al*. Genetic studies of IgA nephropathy: past, present, and future. *Pediatr Nephrol* 2010; **25**: 2257–2268.
- 4 Feehally J, Farrall M, Boland A, Gale DP, Gut I, Heath S *et al*. HLA has strongest association with IgA nephropathy in genome-wide analysis. *J Am Soc Nephrol* 2010; **21**: 1791–1797.
- 5 Gharavi AG, Kiryluk K, Choi M, Li Y, Hou P, Xie J *et al*. Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet* 2011; **43**: 321–327.
- 6 Yu XQ, Li M, Zhang H, Low HQ, Wei X, Wang JQ *et al*. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet* 2011; **44**: 178–182.
- 7 Julian BA, Quiggins PA, Thompson JS, Woodford SY, Gleason K, Wyatt RJ. Familial IgA nephropathy. Evidence of an inherited mechanism of disease. *N Engl J Med* 1985; **312**: 202–208.
- 8 Kiryluk K, Li Y, Sanna-Cherchi S, Rohanizadegan M, Suzuki H, Eitner F *et al*. Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis. *PLoS Genet* 2012; **8**: e1002765.
- 9 Kiryluk K, Li Y, Scolari F, Sanna-Cherchi S, Choi M, Verbitsky M *et al*. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet* 2014; **46**: 1187–1196.
- 10 Lehrer RI, Lu W. α -Defensins in human innate immunity. *Immunol Rev* 2012; **245**: 84–112.
- 11 Saraheimo M, Forsblom C, Pettersson-Fernholm K, Flyvbjerg A, Groop PH, Frystyk J. Increased levels of alpha-defensin (-1, -2 and -3) in type 1 diabetic patients with nephropathy. *Nephrol Dial Transplant* 2008; **23**: 914–918.
- 12 Bokarewa MI, Jin T, Tarkowski A. Intraarticular release and accumulation of defensins and bactericidal/permeability-increasing protein in patients with rheumatoid arthritis. *J Rheumatol* 2003; **30**: 1719–1724.
- 13 Stoecker ZM, Bezalel S, Chapnik N, Asher I, Froy O. High alpha-defensin levels in patients with systemic lupus erythematosus. *Immunology* 2009; **127**: 116–122.
- 14 Peluso G, De Santis M, Inzitari R, Fanali C, Cabras T, Messana I *et al*. Proteomic study of salivary peptides and proteins in patients with Sjögren's syndrome before and after pilocarpine treatment. *Arthritis Rheum* 2007; **56**: 2216–2222.
- 15 Ahn JK, Cha HS, Lee J, Jeon CH, Koh EM. Correlation of DEFA1 gene copy number variation with intestinal involvement in Behcet's disease. *J Korean Med Sci* 2012; **27**: 107–109.
- 16 Chen Q, Hakimi M, Wu S, Jin Y, Cheng B, Wang H *et al*. Increased genomic copy number of DEFA1/DEFA3 is associated with susceptibility to severe sepsis in Chinese Han population. *Anesthesiology* 2010; **112**: 1428–1434.
- 17 Jespersgaard C, Fode P, Dybdahl M, Vind I, Nielsen OH, Csillag C *et al*. Alpha-defensin DEFA1A3 gene copy number elevation in Danish Crohn's disease patients. *Dig Dis Sci* 2011; **56**: 3517–3524.
- 18 Zhou XJ, Cheng FJ, Lv JC, Luo H, Yu F, Chen M *et al*. Higher DEFB4 genomic copy number in SLE and ANCA-associated small vasculitis. *Rheumatology (Oxford)* 2012; **51**: 992–995.
- 19 Guo L, Du Y, Chang S, Zhang K, Wang J. rSNPBase: a database for curated regulatory SNPs. *Nucleic Acids Res* 2014; **42**: D1033–D1039.
- 20 Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C *et al*. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012; **489**: 91–100.
- 21 Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res* 2012; **22**: 1748–1759.
- 22 Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A *et al*. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006; **34**: D108–D110.
- 23 Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* 2011; **21**: 447–455.
- 24 Palii CG, Perez-Iratxeta C, Yao Z, Cao Y, Dai F, Davison J *et al*. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *Embo J* 2011; **30**: 494–509.
- 25 Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L *et al*. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 2007; **39**: 721–723.
- 26 Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B *et al*. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 2007; **80**: 1037–1054.
- 27 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al*. Strong association of de novo copy number mutations with autism. *Science* 2007; **316**: 445–449.
- 28 Black HA, Khan FF, Tyson J, Al AJ. Inferring mechanisms of copy number change from haplotype structures at the human DEFA1A3 locus. *BMC Genomics* 2014; **15**: 614.
- 29 Teumer A, Holtfreter B, Volker U, Petersmann A, Nauck M, Biffar R *et al*. Genome-wide association study of chronic periodontitis in a general German population. *J Clin Periodontol* 2013; **40**: 977–985.
- 30 Sanders JT, Wyatt RJ. IgA nephropathy and Henoch-Schönlein purpura nephritis. *Curr Opin Pediatr* 2008; **20**: 163–170.

- 31 Wyatt RJ, Kritchevsky SB, Woodford SY, Miller PM, Roy SR, Holland NH *et al*. IgA nephropathy: long-term prognosis for pediatric patients. *J Pediatr* 1995; **127**: 913–919.
- 32 Rodriguez-Garcia M, Oliva H, Climent N, Escribese MM, Garcia F, Moran TM *et al*. Impact of alpha-defensins1-3 on the maturation and differentiation of human monocyte-derived DCs. Concentration-dependent opposite dual effects. *Clin Immunol* 2009; **131**: 374–384.
- 33 Barratt J, Feehally J. IgA nephropathy. *J Am Soc Nephrol* 2005; **16**: 2088–2097.
- 34 Cheng FJ, Zhou XJ, Zhao YF, Zhao MH, Zhang H. Alpha-defensin DEFA1A3 gene copy number variation in Asians and its genetic association study in Chinese systemic lupus erythematosus patients. *Gene* 2013; **517**: 158–163.
- 35 Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG *et al*. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010; **327**: 78–81.
- 36 Zhou XJ, Lu XL, Lv JC, Yang HZ, Qin LX, Zhao MH *et al*. Genetic association of PRDM1-ATG5 intergenic region and autophagy with systemic lupus erythematosus in a Chinese population. *Ann Rheum Dis* 2011; **70**: 1330–1337.
- 37 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.
- 38 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.

Supplementary Information accompanies this paper on Genes and Immunity website (<http://www.nature.com/gene>)